



National Longitudinal Surveys

Search National I

NLS Home
NLS Publications ▼
NLS Data ▼
NLS Methods ▼
About NLS ▼
Contact NLS

NLSY97 Data Overview

The NLSY97 consists of a nationally representative sample of 8,984 men and women born during the years 1980 through 1984 and living in the United States at the time of the initial survey in 1997. Participants were ages 12 to 16 as of December 31, 1996. Interviews were conducted annually from 1997 to 2011 and biennially since then. The ongoing cohort has been surveyed 21 times as of date. Data are available from Round 1 (1997–98) through Round 20 (2021–22).

The NLSY97 collects extensive information on respondents' labor market behavior and educational experiences. The survey also includes data on the youths' family and community backgrounds to help researchers assess the impact of schooling and other environmental factors on these labor market entrants. Data from the NLSY97 also aid in determining how youths' experiences relate to establishing careers, participating in government programs, and forming families. Finally, information from the NLSY97 allows researchers to compare the progress of this cohort with that of other [NLS cohorts](#). (See [Cross-Cohort NLSY79/97 Data Harmonization](#)).

On This Page

- » [Topical Guide to the Data](#)
- » [Introduction to the Sample](#)
- » [Using and Understanding the Data](#)
- » [Other Documentation](#)
- » [Accessing Data](#)
- » [Latest News Release](#)
- » [Cross-Cohort NLSY79/97 Overview](#)

Topical Guide to the Data

The topical guide presents detailed information on major subject areas and variables in the NLSY97 survey. The [Asterisk Tables](#) provide summary tables of variables selected from Round 1 to the latest round. Below are the topics gathered in the survey.

- » [Education, Training & Achievement Scores](#)
- » [Employment](#)
- » [Household, Geography & Contextual Variables](#)
- » [Parents, Family Process & Childhood](#)
- » [Dating, Marriage & Cohabitation; Sexual Activity, Pregnancy & Fertility; Children](#)
- » [Income, Assets & Program Participation](#)
- » [Health, Conditions & Practices](#)
- » [Attitudes, Expectations, Non-Cognitive Tests, Activities](#)
- » [Crime & Substance Use](#)

Introduction to the Sample

The NLSY79 sample was selected to represent the civilian, noninstitutional population of the United States within the eligible age range—12 to 16 years of age as of December 31, 1996—with oversamples of Hispanics and non-Hispanic blacks. The majority of the oldest cohort members were still in secondary school during the first survey round and the youngest respondents had not yet entered the labor market. Eligible youths and their parents were administered a questionnaire that covered a range of topics, including work experience, education, work-related attitudes, and other labor force and human capital issues.

At the time of initial survey in 1997, 8,984 respondents were interviewed. Men accounted for 51 percent (4,385) and Women accounted for 49 percent (4,385). The survey included 51.9 percent (4,665) non-Black/non-Hispanic, 26 percent (2,335) Black non-Hispanic, 21.2 percent (1,901) Hispanic or Latino, and 0.9 percent Mixed raced respondents. To select the sample, interviewers screened 75,291 households in 147 primary sampling units that did not overlap. (A primary sampling unit is a metropolitan area or, in nonmetropolitan areas, a single county or group of counties.) Two samples were drawn—a cross-sectional sample representing the U.S. population born in the years 1980 through 1984 and supplemental samples of the black and Hispanic population born in those years. The supplemental samples facilitate more reliable statistical analyses of these groups. The table below provides an overview of those who responded to the survey in Round 1 and in the latest Round 19. [View the profile of the samples for all Rounds](#)

NLSY97 Sample Sizes by Subsample, Race/Ethnicity and Gender

Survey Year, Gender	Total Sample	Cross-sect. total	Cross-sectional Sample				Supp. total	Supplemental Sample		
			Non-black, non-Hisp.	Black, non- Hisp.	Hispanic or Latino	Mixed race		Black, non- Hisp.	Hispanic or Latino	Mixed

Survey Year, Gender	Total Sample	Cross-sectional Sample					Supplemental Sample			
		Cross-sect. total	Non-black, non-Hisp.	Black, non- Hisp.	Hispanic or Latino	Mixed race	Supp. total	Black, non- Hisp.	Hispanic or Latino	Mixed
Round 1										
Male	4599	3459	2413	537	469	40	1140	632	508	--
Female	4385	3289	2252	544	452	41	1096	622	472	2
Total	8984	6748	4665	1081	921	81	2236	1254	980	2
Round 20										
Male	3288	2456	1700	405	325	26	832	480	352	--
Female	3425	2511	1675	454	352	30	914	524	390	--

The [Technical Sampling Report](#) gives detailed descriptions of the NLSY79 sample. Learn more about the [NLSY79 Sample Design & Screening Process](#), [Interview Methods](#), [Interview Methods](#), [Sample Retention and Reasons for Noninterview](#), and [Data Confidentiality](#).

Using and Understanding the Data

Information on survey instruments, variable types, the interviewing process, item nonresponse, sample weights and design effects, data documentation, and how to access the data are available on the [NLSY97 User Guides and Documentation](#) section.

Other Documentation

- [Codebook Supplement](#)
- [Geocode Codebook Supplement](#)
- [NLSY97 Questionnaires](#)
- [Errata Report](#)
- [Tutorials](#)
- [Technical Sampling Report](#)
- [NLSY97 School Surveys](#)

Accessing Data

View [Accessing Data](#) for detailed information on accessing the NLSY97 public-use and the confidential files. The public-use NLS data are available on the [Investigator](#). The links below provide answers to frequently asked questions about requesting the NLSY97 Geocode, Zip-Code and Census Tract files, and the School Survey data.

- [How to Request Access to the NLSY79, NLSY79 Young Adult and NLSY97 Geocode data?](#)
- [How to Request Access to the NLSY Zip Code and Census Tract Files and NLSY97 School Survey Data?](#)

Latest News Release

- [March 29, 2022: Labor Market Experience, Education, Partner Status, and Health for those Born 1980-1984 \[NLSY97 Round 19\] \(HTML\) \(PDF\) \(Supplemental Tables\)](#)
- [View Archived NLSY97 News Releases](#)
- [View News Releases Tables](#)

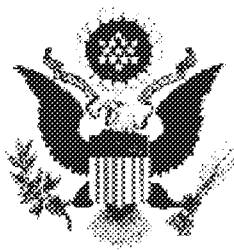
Cross-Cohort NLSY79/97 Overview

An NLSY79/NLSY97 cross-cohort beta release is now available on Investigator. This dataset harmonizes NLS data across two cohorts, allowing users added opportunities to perform cross-cohort comparisons. Initially the dataset includes interview date, reason for noninterview, age, marital status, highest grade attended, highest grade completed, and employment status, with possible expansion to over 100 variables. Fixed background variables are also included, such as sex, race, and AFQT score. See [Cross-Cohort Harmonization Overview](#) document for more information. To access the data, go to <https://www.nlsinfo.org/investigator> and select "NLS Cross Cohort Beta."

NLSY97 COVID-19 Supplement

In an effort to document the effects of the COVID-19 pandemic on the employment and health of Americans, the Bureau of Labor Statistics funded a short NLSY97 COVID-19 Supplement, fielded from February through May 2021. In total, 8,490 NLSY97 respondents were invited to participate in the NLSY97 COVID-19 Supplement and 5,616 respondents completed the survey. Respondents received postal and/or email invitations to complete a short (about 12 minute) web questionnaire about their experiences during the pandemic. The survey asked questions about how employment, health, and childcare have been affected by the pandemic. See the [NLSY97 COVID-19 Supplement](#) documentation for more information. To access the COVID-19 Supplement data, select the NLSY97 Cohort at <https://www.nlsinfo.org/investigator>.

Last Modified Date: March 20, 2024



PRESIDENT'S ADVISORY COMMISSION ON
Consumer Protection AND
Quality IN THE
Health Care Industry

Appendix A
Consumer Bill of Rights and Responsibilities

Table of Contents

Executive Summary

Preamble

Chapter One: Information Disclosure

Chapter Two: Choice of Providers and Plans

Chapter Three: Access to Emergency Services

Chapter Four: Participation in Treatment Decisions

Chapter Five: Respect and Nondiscrimination

Chapter Six: Confidentiality of Health Information

Chapter Seven: Complaints and Appeals

Chapter Eight: Consumer Responsibilities

Consumer Bill of Rights and Responsibilities
Executive Summary

The Advisory Commission on Consumer Protection and Quality in the Health Care Industry was appointed by President Clinton on March 26, 1997, to "advise the President on changes occurring in the health care system and recommend measures as may be necessary to promote and assure health care quality and value, and protect consumers and workers in the health care system." As part of its work, the President asked the Commission to draft a "consumer bill of rights."

The Commission includes 34 members and is co-chaired by The Honorable Alexis M. Herman, Secretary of Labor, and The Honorable Donna E. Shalala, Secretary of Health and Human Services. Its members include individuals from a wide variety of backgrounds including consumers, business, labor, health care providers, health plans, State and local governments, and health care quality experts. The Commission has four Subcommittees: Consumer Rights, Protections, and Responsibilities; Quality Measurement; Creating a Quality Improvement Environment; and Roles and Responsibilities of Public and Private Purchasers and Quality Oversight Organizations. The Commission and its Subcommittees meet monthly in public.

Following is a summary of the eight areas of consumer rights and responsibilities adopted by the President's Advisory Commission on Consumer Protection and Quality in the Health Care Industry:

I. Information Disclosure

Consumers have the right to receive accurate, easily understood information and some require assistance in making informed health care decisions about their health plans, professionals, and facilities.

This information should include:

- **Health plans:** Covered benefits, cost-sharing, and procedures for resolving complaints; licensure, certification, and accreditation status; comparable measures of quality and consumer satisfaction; provider network composition; the procedures that govern access to specialists and emergency services; and care management information.
- **Health professionals:** Education and board certification and recertification; years of practice; experience performing certain procedures; and comparable measures of quality and consumer satisfaction.
- **Health care facilities:** Experience in performing certain procedures and services; accreditation status; comparable measures of quality and worker and consumer satisfaction; procedures for resolving complaints; and community benefits provided.

Consumer assistance programs must be carefully structured to promote consumer confidence and to work cooperatively with health plans, providers, payers and regulators. Sponsorship that assures accountability to the interests of consumers and stable, adequate funding are desirable characteristics of such programs.

II. Choice of Providers and Plans

Consumers have the right to a choice of health care providers that is sufficient to ensure access to appropriate high-quality health care.

To ensure such choice, health plans should provide the following:

Provider Network Adequacy: All health plan networks should provide access to sufficient numbers and types of providers to assure that all covered services will be accessible without unreasonable delay -- including access to emergency services 24 hours a day and seven days a week. If a health plan has an insufficient number or type of providers to provide a covered benefit with the appropriate degree of specialization, the plan should ensure that the consumer obtains the benefit outside the network at no greater cost than if the benefit were obtained from participating providers. Plans also should establish and maintain adequate arrangements to ensure reasonable proximity of providers to the business or personal residence of their members.

Access to Qualified Specialists for Women's Health Services: Women should be able to choose a qualified provider offered by a plan -- such as gynecologists, certified nurse midwives, and other qualified health care providers -- for the provision of covered care necessary to provide routine and preventative women's health care services.

Access to Specialists: Consumers with complex or serious medical conditions who require frequent specialty care should have direct access to a qualified specialist of their choice within a plan's network of providers. Authorizations, when required, should be for an adequate number of direct access visits under an approved treatment plan.

Transitional Care: Consumers who are undergoing a course of treatment for a chronic or disabling condition (or who are in the second or third trimester of a pregnancy) at the time they involuntarily change health plans or at a time when a provider is terminated by a plan for other than cause should be able to continue seeing their current specialty providers for up to 90 days (or through completion of postpartum care) to allow for transition of care. Providers who continue to treat such patients must accept the plan's rates as payment in full, provide all necessary information to the plan for quality assurance purposes, and promptly transfer all medical records with patient authorization during the transition period.

Public and private group purchasers should, wherever feasible, offer consumers a choice of high-quality health insurance products. Small employers should be provided with greater assistance in offering their workers and their families a choice of health plans and products.

III. Access to Emergency Services

Consumers have the right to access emergency health care services when and where the need arises. Health plans should provide payment when a consumer presents to an emergency department with acute symptoms of sufficient severity -- including severe pain -- such that a "prudent layperson" could reasonably expect the absence of medical attention to result in placing that consumer's health in serious jeopardy, serious impairment to bodily functions, or serious dysfunction of any bodily organ or part.

To ensure this right:

- Health plans should educate their members about the availability, location, and appropriate use of emergency and other medical services; cost-sharing provisions for emergency services; and the availability of care outside an emergency department.
- Health plans using a defined network of providers should cover emergency department screening and stabilization services both in network and out of network without prior authorization for use consistent with the prudent layperson standard. Non-network providers and facilities should not bill patients for any charges in excess of health plans' routine payment arrangements.
- Emergency department personnel should contact a patient's primary care provider or health plan, as appropriate, as quickly as possible to discuss follow-up and post-stabilization care and promote continuity of care.

IV. Participation in Treatment Decisions

Consumers have the right and responsibility to fully participate in all decisions related to their health care. Consumers who are unable to fully participate in treatment decisions have the right to be represented by parents, guardians, family members, or other conservators.

In order to ensure consumers' right and ability to participate in treatment decisions, health care professionals should:

- Provide patients with easily understood information and opportunity to decide among treatment options consistent with the informed consent process. Specifically,
 - Discuss all treatment options with a patient in a culturally competent manner, including the option of no treatment at all.
 - Ensure that persons with disabilities have effective communications with members of the health system in making such decisions.
 - Discuss all current treatments a consumer may be undergoing, including those alternative treatments that are self-administered.

- Discuss all risks, benefits, and consequences to treatment or nontreatment.
- Give patients the opportunity to refuse treatment and to express preferences about future treatment decisions.
- Discuss the use of advance directives -- both living wills and durable powers of attorney for health care -- with patients and their designated family members.
- Abide by the decisions made by their patients and/or their designated representatives consistent with the informed consent process.

To facilitate greater communication between patients and providers, health care providers, facilities, and plans should:

- Disclose to consumers factors -- such as methods of compensation, ownership of or interest in health care facilities, or matters of conscience -- that could influence advice or treatment decisions.
- Ensure that provider contracts do not contain any so-called "gag clauses" or other contractual mechanisms that restrict health care providers' ability to communicate with and advise patients about medically necessary treatment options.
- Be prohibited from penalizing or seeking retribution against health care professionals or other health workers for advocating on behalf of their patients.

V. Respect and Nondiscrimination

Consumers have the right to considerate, respectful care from all members of the health care system at all times and under all circumstances. An environment of mutual respect is essential to maintain a quality health care system.

Consumers must not be discriminated against in the delivery of health care services consistent with the benefits covered in their policy or as required by law based on race, ethnicity, national origin, religion, sex, age, mental or physical disability, sexual orientation, genetic information, or source of payment.

Consumers who are eligible for coverage under the terms and conditions of a health plan or program or as required by law must not be discriminated against in marketing and enrollment practices based on race, ethnicity, national origin, religion, sex, age, mental or physical disability, sexual orientation, genetic information, or source of payment.

VI. Confidentiality of Health Information

Consumers have the right to communicate with health care providers in confidence and to have the confidentiality of their individually identifiable health care information protected. Consumers also have the right to review and copy their own medical records and request amendments to their records.

In order to ensure this right:

- With very few exceptions, individually identifiable health care information can be used without written consent for health purposes only, including the provision of health care, payment for services, peer review, health promotion, disease management, and quality assurance.
- In addition, disclosure of individually identifiable health care information without written consent should be permitted in very limited circumstances where there is a clear legal basis for doing so. Such reasons include: medical or health care research for which a institutional review board has determined anonymous records will not suffice, investigation of health care fraud, and public health reporting.
- To the maximum feasible extent in all situations, nonidentifiable health care information should be used unless the individual has consented to the disclosure of individually identifiable information.

When disclosure is required, no greater amount of information should be disclosed than is necessary to achieve the specific purpose of the disclosure.

VII. Complaints and Appeals

All consumers have the right to a fair and efficient process for resolving differences with their health plans, health care providers, and the institutions that serve them, including a rigorous system of internal review and an independent system of external review.

Internal appeals systems should include:

- Timely written notification of a decision to deny, reduce, or terminate services or deny payment for services. Such notification should include an explanation of the reasons for the decisions and the procedures available for appealing them.
- Resolution of all appeals in a timely manner with expedited consideration for decisions involving emergency or urgent care consistent with time frames consistent with those required by Medicare (i.e., 72 hours).
- A claim review process conducted by health care professionals who are appropriately credentialed with respect to the treatment involved. Reviews should be conducted by individuals who were not involved in the initial decision.
- Written notification of the final determination by the plan of an internal appeal that includes information on the reason for the determination and how a consumer can appeal that decision to an external entity.
- Reasonable processes for resolving consumer complaints about such issues as waiting times, operating hours, the demeanor of health care personnel, and the adequacy of facilities.

External appeals systems should:

- Be available only after consumers have exhausted all internal processes (except in cases of urgently needed care).
- Apply to any decision by a health plan to deny, reduce, or terminate coverage or deny payment for services based on a determination that the treatment is either experimental or investigational in nature; apply when such a decision is based on a determination that such services are not medically necessary and the amount exceeds a significant threshold or the patient's life or health is jeopardized.²
- Be conducted by health care professionals who are appropriately credentialed with respect to the treatment involved and subject to conflict-of-interest prohibitions. Reviews should be conducted by individuals who were not involved in the initial decision.
- Follow a standard of review that promotes evidence-based decisionmaking and relies on objective evidence.
- Resolve all appeals in a timely manner with expedited consideration for decisions involving emergency or urgent care consistent with time frames consistent with those required by Medicare (i.e., 72 hours).

VIII. Consumer Responsibilities

In a health care system that protects consumers' rights, it is reasonable to expect and encourage consumers to assume reasonable responsibilities. Greater individual involvement by consumers in their care increases the likelihood of achieving the best outcomes and helps support a quality improvement, cost-conscious environment. Such responsibilities include:

- Take responsibility for maximizing healthy habits, such as exercising, not smoking, and eating a healthy diet.
- Become involved in specific health care decisions.

- Work collaboratively with health care providers in developing and carrying out agreed-upon treatment plans.
- Disclose relevant information and clearly communicate wants and needs.
- Use the health plan's internal complaint and appeal processes to address concerns that may arise.
- Avoid knowingly spreading disease.
- Recognize the reality of risks and limits of the science of medical care and the human fallibility of the health care professional.
- Be aware of a health care provider's obligation to be reasonably efficient and equitable in providing care to other patients and the community.
- Become knowledgeable about his or her health plan coverage and health plan options (when available) including all covered benefits, limitations, and exclusions, rules regarding use of network providers, coverage and referral rules, appropriate processes to secure additional information, and the process to appeal coverage decisions.
- Show respect for other patients and health workers.
- Make a good-faith effort to meet financial obligations.
- Abide by administrative and operational procedures of health plans, health care providers, and Government health benefit programs.
- Report wrongdoing and fraud to appropriate resources or legal authorities.

Consumer Bill of Rights and Responsibilities

Preamble

American consumers and their families are experiencing an historic transition of the U.S. system of health care financing and delivery. In establishing the Advisory Commission on Consumer Protection and Quality in the Health Care Industry, President Clinton asked that it advise him "on changes occurring in the health care system and recommend such measures as may be necessary to promote and assure health care quality and value, and protect consumers and workers in the health care system." As part of that effort, the President has asked the Commission to draft a Consumer Bill of Rights and Responsibilities.

This Commission includes 34 members from a wide variety of backgrounds including consumers, business, labor, health care providers, health plans, State and local governments, and health care quality experts. We hope our diversity of interests and backgrounds will make our recommendations more valuable to those who consider them.

This is an appropriate time to reexamine and reconsider the methods by which our Nation and the health care industry establish and protect the rights and identify the responsibilities of those people who use the health care system. The Commission believes it is essential to preserve those elements of the emerging system that have a positive impact on the quality of care as well as the cost and availability of health insurance coverage.

Development of a Consumer Bill of Rights and Responsibilities is an important step forward for all those involved in the health care system. Consumers, health care professionals, administrators of health care facilities, and those who operate health plans will benefit from a clear set of unifying standards. The Consumer Bill of Rights and Responsibilities can help to establish a stronger relationship of trust among consumers, health care professionals, health care institutions, and health plans by helping to sort out the shared responsibilities of each of these participants in a system that promotes quality improvement.

The work of this Commission builds on the efforts of many others. The Commission reviewed dozens of proposals prepared and released by a variety of organizations that have addressed the rights, responsibilities, and protection of consumers. We have heard public testimony from dozens of individuals and organizations. We are grateful for their contributions.

The Consumer Bill of Rights and Responsibilities charts a course for the continued enhancement of health systems and processes that serve to protect consumers and ensure quality. While the rights and responsibilities

included in this report are intended to apply to all consumers and participants in the health care system, the Commission recognizes that the strength of these protections will grow over time as the capabilities of the health care industry become more sophisticated. Certain portions of the industry will require additional time to make these adjustments, but the Commission intends that the bulk of its recommendations be put in place within the next 3 years.

The Consumer Bill of Rights and Responsibilities was first drafted by the Subcommittee on Consumer Rights, Protections, and Responsibilities. The Subcommittee met in open session on seven separate occasions, and the Commission met six times during that same time period. The Subcommittee considered background papers on each topic, heard public testimony on most topics, and considered two or three drafts of each chapter. At each point in that process, the Subcommittee briefed the full Commission on its work and received feedback on those issues. The Commission also has considered draft chapters and revised drafts reflecting the input of its members. Throughout this process, the Subcommittee and the Commission have operated on a consensus basis that has allowed any member to place an issue before the respective body for consideration. The list of issues was refined to reflect the discussions of the Subcommittee and the Commission. The final product reflects the areas of overall agreement expressed by Commission members.

Objectives of a Consumer Bill of Rights and Responsibilities

The Consumer Bill of Rights and Responsibilities is intended to accomplish three major goals.

First, to strengthen consumer confidence by assuring the health care system is fair and responsive to consumers' needs, provides consumers with credible and effective mechanisms to address their concerns, and encourages consumers to take an active role in improving and assuring their health.

Second, to reaffirm the importance of a strong relationship between patients and their health care professionals.

Third, to reaffirm the critical role consumers play in safeguarding their own health by establishing both rights and responsibilities for all participants in improving health status.

Guiding Principles for the Consumer Bill of Rights and Responsibilities

The work of the Commission was guided by the following principles:

All consumers are created equal. The work of this Commission in establishing a Bill of Rights and Responsibilities must apply to all consumers. This includes all beneficiaries of such public programs as Medicare, Medicaid, the Department of Veterans Affairs, and the Department of Defense, as well as Federal, State, and local government employees. It also includes all those who have private insurance, including those who purchase their own insurance, those who work for companies that have self-funded health plans, and those who work for companies that purchase insurance for their employees and dependents. And, finally, to the extent possible, these rights should be accorded to those who have no health insurance but use the health care system.

Quality comes first. The first question we asked ourselves in each circumstance was: Will this improve the quality of care and of the system that delivers that care? Sometimes this led us to reject policy options that we believe could hinder the progress our Nation has made toward a health care system that is focused on improving quality through accountable organized systems.

Preserve what works. There are elements of managed care and of indemnity coverage that must be changed to protect the rights of consumers. But there also are elements of each system that have improved quality and expanded access. We have tried to make sure that we preserve what works while we address areas that can and should be improved.

Costs matter. Although a comprehensive cost-analysis was not performed for this Bill of Rights and Responsibilities, the Commission has sought to balance the need for stronger consumer rights with the need to keep coverage affordable. We recognize that, in some circumstances, rights may create

additional costs for employers; health plans; Federal, State, and local governments; and consumers. We also recognize that ultimately consumers can bear these costs in the form of lower wages, higher prices, higher taxes, or reduced benefits in other areas. The Commission believes some components of the Bill of Rights may also enhance the efficiency and effectiveness of the health care marketplace. While these efficiencies cannot be well calculated, they may help to offset some cost increases. The Commission has attempted to weigh these factors carefully and support recommendations that may prompt additional spending in cases where such spending may represent an investment in higher quality health care and better health outcomes.

Goals for Consumer Protection in a Quality-Focused Health Care System

A Consumer Bill of Rights and Responsibilities is, by its nature, a snapshot of what is needed at a particular time. The rights enumerated in this report are intended to move the health care system in a direction that is consistent with a system of health care delivery that is focused on obtaining the highest quality and best outcome for consumers and their families. In that light, the Commission has identified a series of goals for the continued reform of the American health care system that will maximize consumer rights in a system that focuses on quality.

Health coverage is the best consumer protection. A health care system that leaves more than 41 million Americans without health coverage cannot adequately protect the rights of consumers and their families. The fact that so many Americans live day in and day out without the security that health coverage provides is intolerable. Recent trends reported by the U.S. Census Bureau that the number of uninsured Americans rose by one million between 1996 and 1997 are cause for great concern. Moreover, the continued existence of a large group of Americans without health insurance increases the costs paid by those who have insurance as uncovered expenses are shifted to other purchasers. Efforts by Federal and State governments to expand the number of children who are insured are encouraging and should be strengthened. Similar efforts should be extended to other segments of the population so that all Americans are covered.

Consumers faced with catastrophic illness require assistance. Each year, an estimated 1,500 to 2,500 Americans lose their private health insurance coverage because their medical expenses exceed a lifetime limit included in their health insurance policy. Many of these consumers must exhaust their family savings before becoming eligible for Medicaid or other forms of public assistance. This creates a tremendous hardship on these individuals and their families. Employers, health plans, and others should seriously consider taking steps to ease this burden by (1) eliminating or increasing lifetime limits, (2) expanding the use of high-risk pools to provide immediate coverage at the time consumers reach a lifetime limit, or (3) offering supplemental coverage for workers who wish to increase their limits.

Coverage must be made affordable for all consumers, employers, and other purchasers. The recent moderation in health care costs is promising and has been a contributing factor in the slowing of insurance coverage losses. Employers, health plans, and Federal and State governments should be applauded for their efforts to make coverage more affordable for more Americans. Recent projections for 1998 are less favorable. History makes clear that we cannot assume that costs will remain under control without continued cost containment.

Vulnerable groups require special attention. Many consumers are, for reasons beyond their control, more vulnerable than others to losing their coverage or experiencing significant gaps in their coverage. Individuals with mental or physical disabilities, low-income individuals, children, non-English-speaking consumers, and others require considerable attention by decisionmakers at all levels of the system. Enactment of the Americans with Disabilities Act of 1990, the Health Insurance Portability and Accountability Act of 1996, and the Mental Health Parity Act of 1996 were important steps to protect these consumers. Further steps can and should be taken.

Small purchasers need assistance. The owners of small businesses, the self-employed, and those who purchase insurance in the individual market continue to have great difficulty finding and maintaining affordable health care coverage. For a variety of reasons, insurance premiums are higher for small firms relative to the benefits they are able to purchase, and some small firms are unable to purchase insurance at all. In its final report, the Commission intends to offer several recommendations to help ameliorate some of these effects, including voluntary approaches for expanding insurance pools and for adjusting payment systems to reflect the greater risk inherent in small group and individual markets.

Consumer participation in clinical research. The national investment in clinical research has led to breakthrough advances in diagnosis, prevention, and treatment of illness and disability that have lengthened and improved the quality of life for millions of consumers while also achieving significant cost savings to the health care industry. Consumer participation in clinical research through their inclusion in clinical trials is vitally important not only to continued advancement and innovation in medical care but to the often life-threatening nature of the conditions affecting such consumers. The Commission encourages the ongoing efforts by researchers, health plans, employers, public purchasers, and others to resolve impediments to consumer participation in clinical trials and urges participants to reach agreement on an appropriate sharing of costs and responsibilities related to such trials.

The Commission does not, in this report, speak to the issues of implementation or enforcement of the Consumer Bill of Rights and Responsibilities. The rights enumerated in this report can be achieved in several ways including voluntary actions by health plans, purchasers, facilities, and providers; the effects of market forces; accreditation processes; as well as State or Federal legislation or regulation. In its final report to the President, the Commission intends to speak to the optimal methods for implementing and enforcing these rights through one or more of these approaches.

Finally, the Commission believes that the American people should have access to health care that is of high quality, evidence-based, safe, free of errors, and is available to all Americans regardless of ability to pay. Progress, over time, will require changes that must be made prudently, realistically, and with due regard to the needs of all stakeholders in the system. This Consumer Bill of Rights and Responsibilities specifies improvements that we believe are achievable now and in the next several years. It acquires even more meaning in the context of a broader overarching commitment to ensure that full access to high-quality health care will eventually be available to all Americans.

Consumer Bill of Rights and Responsibilities

Chapter One

Information Disclosure

Statement of the Right **Consumers have the right to receive accurate, easily understood information and some require assistance in making informed health care decisions about their health plans, professionals and facilities.**

This information should include:

- **Health plans:** Covered benefits, cost-sharing, and procedures for resolving complaints; licensure, certification, and accreditation status; comparable measures of quality and consumer satisfaction; provider network composition; the procedures that govern access to specialists and emergency services; and care management information.
- **Health professionals:** Education and board certification and recertification; years of practice; experience performing certain procedures; and comparable measures of quality and consumer satisfaction.
- **Health care facilities:** Experience in performing certain procedures and services; accreditation status; comparable measures of quality and worker and consumer satisfaction; procedures for resolving

complaints; and community benefits provided.

Consumer assistance programs must be carefully structured to promote consumer confidence and to work cooperatively with health plans, providers, payers, and regulators. Sponsorship that assures accountability to the interests of consumers and stable, adequate funding are desirable characteristics of such programs.

Rationale

Value-based purchasing allows consumers to obtain greater value for their health care dollar by seeking higher quality care at the best price. To do this, consumers need accurate, reliable information that will allow them to assess differences in the quality and cost of health benefits plans, the health care providers who treat them, and the facilities and institutions that house them. Active and informed decisionmaking by consumers will improve the performance of the health care system, as providers seek to enhance their quality and reduce their costs in order to be more attractive to value-seeking consumers.

A more basic reason for providing consumers with information is an ethical one. Health plans, facilities, and professionals have an ethical obligation to inform consumers about how their actions can affect the consumer's life and health. Medical ethicists ground this obligation in the principle of respect for individual autonomy and individuals' right to make choices about how they receive medical care (Beauchamp and Childress, 1994).

This chapter provides a description of the types of information on health plans, health professionals, and health care facilities that should be made available to consumers either routinely or upon request. The Commission recognizes that much work remains to be done if all this information is to be readily available and understandable to consumers, specifically:

- **Detailed explanation is needed for certain types of information.** Some types of information are straightforward and require no further definition (e.g., the names, board certification status, and geographic location of primary care providers in a plan's network). Other types of information would benefit from the development of more detailed explanation, such as the care management information on clinical protocols, practice guidelines, and preauthorization and utilization review standards and procedures.
- **Standardized measures are needed for comparative purposes.** For the information intended to support consumer decisions regarding the choice of a health benefits plan, or choice of an individual provider or facility, standardized definitions will be needed to allow for "apples to apples" comparisons.
- **Ongoing development and promulgation of standardized measurement sets and instruments are needed for assessing satisfaction and quality.** The Commission believes that some of the most important types of information a consumer has a right to receive fall into the categories of consumer satisfaction ratings and clinical quality performance measures for health plans, health care professionals, and facilities. For all consumers to exercise this right, processes must be put in place to create standardized performance measures. In its final report, the Commission intends to address how such a process might be established so as to build on existing efforts, encourage ongoing innovation in quality measurement, and provide the best possible information to consumers at any given time to encourage quality improvement through market-based decisions.
- **Useful and appropriate reporting formats and processes are needed for consumers.** Although the Commission believes that consumers should have access to pertinent information, it recognizes that caution must be taken to provide information to consumers in useful formats (e.g., summary and detailed reports, printed copy, and Internet), at appropriate times (i.e., decision points), with assistance for vulnerable groups (i.e., those who are hearing impaired or non-English speaking). These issues also will be addressed in the Commission's final report.

Consumers should be able to obtain other information upon request as outlined below. Plans, providers, and facilities should inform consumers that such information is available and describe how it can be obtained.

Health Plan Information

Many consumers face a choice of health plans such as an indemnity plan, an HMO, a point-of-service plan, or a preferred provider organization. Consumers' choice of a health plan has a significant impact on consumers' ability to make other choices about facilities, health professionals, and treatment options. Even in cases where consumers do not have a choice of plans, they require information on the plan in which they are enrolled to use the available services effectively.

To the extent that a right to information creates disclosure requirements for health plans, these requirements should apply equally to all types of plans (including indemnity, HMO, PPO, and POS) regardless of sponsor (e.g., such government programs as CHAMPUS, VA, FEHBP, Medicare, and Medicaid and private plans including fully funded, partially self-funded, or fully self-funded plans). If the specific information required for disclosure does not exist, or is unavailable, the consumer should be informed.

The primary responsibility of providing consumers with health plan information falls upon the plans themselves. In the case of self-insured plans, this responsibility will rest with the plan sponsor unless it is delegated or contracted to a third-party administrator.

Within the category of health plan information, one can discern four principal subcategories of information: (1) benefits, cost-sharing, and dispute resolution; (2) health plan characteristics and performance information; (3) network characteristics; and (4) care management information.

A. Benefits, Cost-Sharing, and Dispute Resolution. Consumers should receive the following information about a health benefits plan:

- A general summary of all covered benefits, including:
 - General limits on coverage, including any annual or lifetime limits, as well as limits for specific conditions.
 - Whether preventative services are covered.
 - Whether a drug formulary is used and, if so, how decisions are made pertaining to inclusion of drugs, particularly new drugs (including a process to consider exceptions).
 - How drugs, devices, and procedures are deemed experimental.
- Enrollee cost-sharing, including employee or beneficiary premium contributions, deductibles, copayments, and coinsurance.
- Type and extent of dispute resolution procedures available in the event of a dispute.

B. Health Plan Characteristics and Performance Information. Consumers joining or considering whether or not to join a health plan should receive information about:

- State licensure status, Federal certification, and private accreditation status (including publicly available reports).
- Consumer satisfaction measures.
- Clinical quality performance measures.
- Service performance measures (e.g., waiting time to obtain an appointment with primary care providers and specialists).
- Disenrollment rates (adjusted for involuntary disenrollment and other relevant factors).

Additional information that should be made available *upon request* includes:

- Number of years in existence.
- Corporate form of the plan (i.e., public or private; gateway.html or for-profit ownership and management).
- Whether the plan meets requirements (State and Federal) for fiscal solvency.
- Whether the plan meets standards (State, Federal, and private accreditation) that assure confidentiality of medical records and orderly transfer to caregivers.

C. Network Characteristics. It is important to provide consumers with information about the characteristics of the network and the procedures that govern its use. Consumers should receive:

- Aggregate information on the numbers, types, board certification status, and geographic distribution of primary care providers and specialists.
- Detailed list of names, board certification status, and geographic location of all contracting primary care providers; whether they are accepting new patients; language(s) spoken and availability of interpreter services; and whether facilities are accessible to people with disabilities.
- Provider compensation methods, including base payment (e.g., capitation, salary, fee schedule) and additional financial incentives (e.g., bonus, withholds, etc.).
- Rules regarding coverage of out-of-network services, and applicable rates of cost-sharing.
- Information about circumstances under which primary care referral is required to access specialty care.
- Information about what options exist for 24-hour coverage and whether enrollees have access to urgent care centers.

Additional information that should be made available *upon request* includes:

- Detailed list of names, board certification status, and geographic location of all contracting specialists and specialty care centers; whether they are accepting new patients; language(s) spoken and availability of interpreter services; and whether facilities are accessible to people with disabilities.
- Detailed list of names, accreditation status, and geographic location of hospitals, home health agencies, rehabilitation and long-term care facilities; whether they are accepting new patients; language(s) spoken and availability of interpreter services; and whether they are accessible to people with disabilities.

D. Care Management Information. Information in this category that should be available *upon request* includes:

- Preauthorization and utilization review procedures followed.
- Use of clinical protocols, practice guidelines, and utilization review standards pertinent to a patient's clinical circumstances.
- Whether the plan has special disease management programs or programs for persons with disabilities. (This information should indicate whether these programs are voluntary or mandatory or if a significant benefit differential results.)
- Whether a specific prescription drug is included in a formulary and procedures for considering requests for patient-specific waivers.
- Qualifications of reviewers at the primary and appeals levels.

Health Professional Information

All consumers should receive information on:

- Whether the health professional's ownership or affiliation arrangement with a provider group or institution would make it more likely that a consumer would be referred to particular specialists or facility or receive a particular service.
- How the provider is compensated, including base payment method (e.g., capitation, salary, fee schedule) and types of additional financial incentives (e.g., bonus, withholds).

Consumers should receive *upon request* the following information on health professionals:

- Education, board certification, and recertification status.
- Names of hospitals where physicians have admitting privileges.
- Years of practice as a physician and as a specialist if so identified.
- Experience with performing certain medical or surgical procedures (e.g., volume of care/services delivered), adjusted for case mix and severity.
- Consumer satisfaction measures.
- Clinical quality performance measures.
- Service performance measures.

- Accreditation status (if applicable).
- Corporate form of the practice (i.e., public or private, gateway.html or for-profit, ownership and management, sole proprietorship or group practice).
- The availability of translation or interpretation services for non-English speakers and people with communication disabilities.
- Any cancellation, suspension, or exclusion from participation in Federal programs or sanctions from Federal agencies; any suspension or revocation of medical licensure, Federal controlled substance license, or hospital privileges.

Health Care Facility Information

Consumers should receive the following information from a health care facility:

- Corporate form of the facility (i.e., public or private; gateway.html or for-profit; ownership and management; affiliation with other corporate entities).
- Accreditation status.
- Whether specialty programs meet guidelines established by specialty societies or other appropriate bodies (e.g., whether a cancer treatment center has been approved by the American College of Surgeons, the Association of Community Cancer Centers, or the National Cancer Institute).
- The volume of certain procedures performed at each facility.
- Consumer satisfaction measures.
- Clinical quality performance measures.
- Service performance measures.
- Procedures for registering a complaint and achieving resolution of that complaint.
- The availability of translation or interpretation services for non-English speakers and people with communication disabilities.
- Numbers and credentials of providers of direct patient care (e.g., registered nurses, other licensed providers, and other caregivers).
- Whether the facility's affiliation with a provider network would make it more likely that a consumer would be referred to health professionals or other organizations in that network.
- Whether the facility has been excluded from any Federal health programs (i.e., Medicare or Medicaid).

Consumer Assistance Programs

Initial results indicate that consumer assistance programs support consumer needs for information on health plans, providers, and facilities. A loose patchwork of consumer assistance services currently exists in the public and private sectors. In the public sector, 14 State or locally based Medicaid programs now have established ombudsmen programs to assist beneficiaries with information needs. Some Medicare beneficiaries and people with chronic health problems have access to consumer assistance services through Information, Counseling, and Assistance (ICA) programs, long-term care ombudsmen programs, and protection and advocacy programs.

In the private sector, health plans often provide consumers with assistance services through customer and member service departments (Oxford Health Plans, 1997; Harvard Pilgrim Health Plan, 1997). Large group purchasers and labor unions often provide their employees with consumer assistance by organizing information on plans, educating employees about their rights, and intervening when employees have complaints about their plans (Darling, 1997).

While there are a number of sources that provide assistance to consumers, most programs target specific subpopulations and have limited funds, and hence provide a limited range of services. There are reasons to believe that consumers and other stakeholders would benefit from greater availability of consumer assistance programs that:

- **Inspire confidence.** Consumers want to know that they will be treated fairly.
- **Provide a safety valve.** Even in the best of systems, there will be individuals who fall through the cracks. Assistance programs provide a resource that can help such individuals resolve problems quickly and

- efficiently, often bridging communication failures between the consumer and the provider or health plan.
- **Foster collaboration.** Assistance programs should work with the array of available resources to best meet the needs of consumers.

The challenge to crafting assistance programs for health care consumers is to ensure that such programs are not duplicative, but rather that they supplement and complement existing resources.

With regard to consumer assistance, the Commission has not addressed issues of implementation. Specifically, this is not an endorsement or a requirement for any particular form of consumer assistance programs, but lays out desirable characteristics of such programs.

Implications of the Right

Obtaining the information listed above and making it available to consumers will not, by itself, equip consumers with the knowledge and abilities required to act on this information. Discussed below are some basic considerations in making this information useful to consumers and the implications of this for key segments of the health care industry.

- **Information Should Be Useful to Consumers and Cost Effective to Obtain.** Edgman-Levitan and Cleary (1996) have documented that consumers are able to evaluate critical information about quality. However, research on how consumers use information to make decisions suggests that too much information can be overwhelming. In its 1988 assessment of methods for communicating the quality of medical care to consumers, the Office of Technology Assessment's Expert Advisory Panel concluded that "limiting information to only a few indicators of quality will probably be necessary [because] people can consider only a few items at any one time. Information is processed as a unit or chunk -- a person's processing capacity has been estimated as being anywhere from four to seven chunks" (OTA, 1988). Ongoing research must be conducted to determine what is the most effective subset of information that consumers can use. Finally, while consumers clearly have a right to information, it must be understood that there are costs associated with collecting and distributing it. While providing information to consumers generates significant benefits for both the consumers and the health system as a whole, it is not necessarily inexpensive. Recognizing these costs, however, is not an argument for a "bare bones" approach to information disclosure. The failure to provide information also has costs. Well-informed consumers are the bedrock of an efficiently operating market. Without meaningful information, consumers are more likely to make choices that can result in less than optimal outcomes for themselves and there is less incentive for participants to strive for excellence. The challenge is to develop coordinated approaches to information collection and dissemination that will provide consumers the information they need to make decisions without imposing severe burdens on plans and providers.

Investments in Clinical Information Systems and Workforce Education and Training Will Be Needed. Greater investment in automated information systems will be necessary for health plans and providers to satisfy these information disclosure requirements, especially ones pertaining to product, facility, and provider performance and quality. The Commission is currently assessing barriers or impediments to investment in clinical information systems (e.g., inadequate data collection standards; confidentiality concerns; magnitude of capital investments required) and plans to speak to this issue in its final report. Responding to these increased information demands also has implications for the training and education of the health care workforce. There will be greater demand by health care organizations for individuals with particular technical and analytic skills (e.g., computer programming, engineering, data auditing, and statistics). Ongoing training and continuing education programs for practitioners and other workers whose work involves recording, compiling, or manipulating clinical and administrative data will also be needed to assure the completeness and accuracy of data and adherence to confidentiality safeguards.

References and Selected Reading

Beauchamp TL, Childress JF. *Principles of Biomedical Ethics* (New York): Oxford University Press; 1994.

Darling H. Xerox Corporation. testimony before the Advisory Commission on Consumer Protection and Quality in the Health Care Industry, May, 1997.

Edgman-Levitan S, Cleary S and P. "What Information Do Consumers Want and Need: What Do We Know About How They Judge Quality and Accountability?" in: Jones SB, Lewin ME, editors, *Improving the Medicare Market: Adding Choice and Protections*, Washington, DC: National Academy Press; 1996.

Harvard Pilgrim Health Care. telephone conversation with Janice Boyce, Public Affairs Department, August 25, 1997.

Office of Technology Assessment. *The Quality of Medical Care: Information for Consumers*, No. OTA-H-386. Washington, DC; 1988.

Oxford Health Plans. telephone conversation with Scott Schwartz, Assistant General Counsel, August 25, 1997.

Sofaer S. "How Will We Know If We Got It Right? Aims, Benefits and Risks of Consumer Information Initiatives," *The Joint Commission Journal on Quality Improvement*, May 1997; 23(5):258-264.

Consumer Bill of Rights and Responsibilities

Chapter Two

Choice of Providers and Plans

Statement of the Right **Consumers have the right to a choice of health care providers that is sufficient to ensure access to appropriate high-quality health care.**

To ensure such choice, health plans should provide the following:

Provider Network Adequacy: All health plan networks should provide access to sufficient numbers and types of providers to assure that all covered services will be accessible without unreasonable delay -- including access to emergency services 24 hours a day and seven days a week. If a health plan has an insufficient number or type of providers to provide a covered benefit with the appropriate degree of specialization, the plan should ensure that the consumer obtains the benefit outside the network at no greater cost than if the benefit were obtained from participating providers. Plans also should establish and maintain adequate arrangements to ensure reasonable proximity of providers to the business or personal residence of their members.

Access to Qualified Specialists for Women's Health Services: Women should be able to choose a qualified provider offered by a plan -- such as gynecologists, certified nurse midwives, and other qualified health care providers -- for the provision of covered care necessary to provide routine and preventative women's health care services.

Access to Specialists: Consumers with complex or serious medical conditions who require frequent specialty care should have direct access to a qualified specialist of their choice within a plan's network of providers. Authorizations, when required, should be for an adequate number of direct access visits under an approved treatment plan.

Transitional Care: Consumers who are undergoing a course of treatment for a chronic or disabling condition (or who are in the second or third trimester of a pregnancy) at the time they involuntarily change health plans or at a time when a provider is terminated by a plan for other than cause should be able to continue seeing their current specialty providers for up to 90 days (or through completion of postpartum care) to allow for transition of care. Providers who continue to treat such patients must accept the plan's

rates as payment in full, provide all necessary information to the plan for quality assurance purposes, and promptly transfer all medical records with patient authorization during the transition period.

Public and private group purchasers should, wherever feasible, offer consumers a choice of high-quality health insurance products. Small employers should be provided with greater assistance in offering their workers and their families a choice of health plans and products.

Rationale

The ability of consumers to exercise choice in the health care marketplace is associated with several desirable characteristics of a health care system.

- First, choice is associated with increased consumer satisfaction. In a survey of consumers receiving health care in both indemnity and managed care plans, individuals with a choice of health products report greater satisfaction with their plan and tend to rate both their health insurance product and their individual physicians of higher quality (Davis and Schoen, 1997).
- Second, the ability of consumers to choose among competing products is a hallmark of a healthy marketplace. Individual consumers are responsible for 34 percent of all direct expenditures for health care in the United States (Cowan et al., 1996). As the science of measuring and generating accurate and valid information on the quality of health plans, providers and facilities advances, consumers can wield their purchasing power to create incentives in the marketplace for improvements in health care quality.
- Third, consumers who have a role in the selection of their caregivers are likely to have greater confidence in those practitioners and are, therefore, more likely to seek appropriate care in a more timely fashion and follow agreed-upon care regimens.
- Fourth, having a choice of providers allows consumers to take action to preserve continuity of care within the health care system by selecting products and providers that allow them to continue provider relationships when continuity of care is especially important (e.g., prenatal care, care of individuals with complex chronic or disabling conditions).

Thus, a health care marketplace that promotes satisfied consumers, continuity of care, and continuous improvements in quality requires that an array of choices be available to consumers. Without consumers' ability to have and exercise choice, greater activities may need to be undertaken by group purchasers and regulators to ensure that the health care marketplace responds appropriately to consumers' health care needs.

Consumer Choice of Health Plans or Products

During the last decade, there has been a marked increase in the number and types of health insurance products available in most geographic markets. Prior to the widespread development of managed care plans, most Americans had limited choice of health insurance products. Indemnity products dominated the market with HMO and PPO products available primarily in certain metropolitan areas. The past 10 years have seen a significant increase of insurance products with the expansion of many health plans into new geographic markets and the development of multiple insurance product lines by indemnity insurers and managed care organizations. As a result, with the exception of sparsely populated areas, most communities now have available HMO, POS, PPO, and indemnity products offering consumers a variety of options in terms of benefits, premiums, copayments, and health care delivery systems.

At the same time, there has been a steady migration from traditional indemnity plans to various managed care products in both the public and private markets. Between 1991 and 1995, the percentage of American workers enrolled in indemnity plans decreased from 59 percent to 35 percent (EBRI, 1997). In 1997, more than 5 million Medicare beneficiaries were enrolled in 336 managed care plans, an increase of more than 100 percent since 1993. Under Medicaid, 13 million, or 35 percent, of all beneficiaries have been enrolled in managed care plans, an increase of more than 170 percent since 1993. The Balanced Budget Act of 1997 will increase those trends by expanding the types of products available to beneficiaries of those two public programs.

Although there is greater choice of health insurance products available in most markets, it is important to note that this choice often is exercised at the level of the group purchaser instead of by individual consumers. Between 1988 and 1997, health plan offerings by moderate- and large-sized employers declined (Gabel, 1997). Those offering three or more plans declined from 35 percent to 32 percent, while those offering only one plan climbed from 41 percent to 44 percent over that period. Notably, the percentage of employees in firms with 200 or more workers who were offered coverage of PPOs and POS plans increased from 12 percent in 1988 to 58 percent in 1997 (Gabel, 1997).

There also is evidence of variation in consumer preferences for various product characteristics. In the Kaiser-AHCPR survey (1996), 70 percent of survey respondents would prefer a high-cost product with a wide range of benefits over a low-cost product with a more limited range of benefits (26 percent). Respondents were more divided over other health product decisions. Fifty-three percent said they would pay more for unrestricted choice of physicians, while 43 percent would opt for a lower-cost product that limited choice to a list of physicians. Forty-six percent would pay more to have direct access to any specialist, whereas more than half (51 percent) would choose a lower-cost plan that requires a visit to the family physician for a referral (Robinson and Brodie, 1997).

The Commission is troubled by the limited choice of insurance products made available to many consumers through their employer group purchasers. Some of the reduction in choice of plan and product has resulted from conscious decisions by employers to select high-quality products at the best price in the market. In other instances, employers may be seeking to minimize administrative costs associated with multiple offerings. Affording consumers greater choice of plans would allow consumers to select the product that best meets their individual preferences and would encourage health plans to be responsive to consumers' expressed needs. However, the Commission recognizes that, for many consumers, the availability of one plan is better than no plan at all.

The Commission was unable to achieve consensus on creating a "right" to a consumer choice of health plan or product but it is determined to find ways to encourage and assist employers and other group purchasers in providing consumers with a meaningful choice of health plans and products. Consumer choice of health plans is important and should be provided whenever possible and in a way that is affordable both to employers and consumers. In its final report, the Commission will address policy options to provide greater choice of health plans and products, including encouraging the development of purchasing coalitions and alliances to assist small employers who encounter the greatest difficulty in offering multiple options.

Consumer Choice of Physicians and Other Health Care Providers

The shift from indemnity coverage to managed care arrangements can affect consumers' choice of physicians and other health care providers. In a 1995 study, 41 percent of managed care enrollees who changed health plans over the prior 3 years also changed physicians (Davis et al., 1995). However, nearly all covered workers can now choose a health plan that covers non-network providers. In some cases, however, the additional cost of these products or of the option to go out of network effectively puts such choice out of the reach of some consumers.

It also is clear that consumers value some degree of choice of physicians. The 1997 Kaiser/Commonwealth National Health Insurance Survey found that respondents with a choice of physicians registered the highest level of satisfaction with their plans (Davis and Schoen, 1997). A Kaiser-AHCPR survey of consumers identified four reasons why consumers prefer a greater choice of physicians and other health care professionals:

- "So you can see whatever doctor you think is best qualified to treat a particular medical problem" (43 percent);
- "So you can change doctors if you become dissatisfied with the one you're seeing" (24 percent);
- "So you can continue seeing your regular doctor" (20 percent); and,
- "So it's easier to see someone else if your doctor is not available for an appointment" (9 percent).

The most frequently cited reasons speak to consumers' desire to use choice of physicians as a way to obtain quality care. The third is directed toward maintaining relationships with physicians with whom consumers have

an existing relationship. In other words, 63 percent of consumers surveyed wanted a choice of physicians so that they can develop and maintain a relationship with a physician they trust to provide them high-quality care.

Therefore, it is important for all health plans and products to maintain an adequate network of physicians and other health care providers, to provide for continuity of care when consumers change plans, and to allow consumers with special health care needs to have adequate choice of physicians and other health care providers. This can lead to higher consumer satisfaction with providers and their health plans without undermining the efforts of provider groups and health plans to develop organized delivery systems.

The Commission's recommendations seek to build on these trends toward providing greater choice by taking several steps to ensure (1) network adequacy; (2) greater access for women to qualified specialists for women's health services; (3) ease of access to specialists for consumers with complex and serious conditions; and (4) greater continuity of care for consumers who enroll in new health plans or see their provider dropped from a plan for other than cause.

Provider Network Adequacy

When appropriately structured, a plan using a network of providers can improve the quality and coordination of care delivered to consumers through careful selection and credentialing of providers and through coordination of care by primary care physicians and those with specialty training. The National Association of Insurance Commissioners (NAIC, 1996) has developed standards for provider network adequacy that have been adopted by several States. The Commission believes universal adoption of these standards will improve both the quality of care and consumers' satisfaction with their health plans and their care. Because of its strong desire to maintain the integrity of health plan networks, the Commission has rejected approaches to mandate the inclusion of providers into networks (i.e., "any willing provider" laws) or to require plans to allow enrollees to go out of plan networks at will (i.e., "freedom of choice" laws).

Access to Specialists

Consumers with ongoing health needs often require regular access to physicians and other health care professionals who are specially trained to serve those needs (Bernstein, Dial, and Smith, 1995). This is especially true of those consumers who have disabling or terminal conditions. In such cases, the traditional "gatekeeper" approach used by some health plans can be an impediment to access to quality care and result in unnecessary inconvenience to consumers. The Commission's recommendations are designed to promote consumers' access to appropriately trained specialists while maintaining the integrity of network models of care. Consumers with complex and serious medical conditions who require frequent specialty care should have direct access to a qualified specialist of their choice within a plan's network of providers. Authorizations, when required, should be for an adequate number of direct access visits under an approved treatment plan.

Access to Qualified Specialists for Women's Health Services

Morbidity and mortality associated with breast cancer, cervical cancer, ovarian cancer, and sexually transmitted diseases in women can be significantly reduced through the provision of preventive and routine gynecological services. The U.S. Preventive Services Task Force has issued recommendations pertaining to the provision of Pap smears, mammograms, and other preventive services for women. Women should be able to choose a qualified provider offered by a plan -- including gynecologists, certified nurse midwives, and other qualified health care providers offered by a plan -- for the provision of routine and preventive women's health care services.

Transitional Care

Finally, consumers who are undergoing an extensive course of treatment (e.g., chemotherapy or prenatal care) at the time they join a new health plan should be able to continue to see their current providers for a period of up to 90 days (or through completion of postpartum care). Similarly, such consumers should be able to continue to see

a provider who is terminated from a plan's network for reasons other than cause. Sudden interruption of care can compromise the quality of care and patient outcomes. Continuity of care has been shown to increase the likelihood that patients receive appropriate preventive services (O'Malley et al., 1997). Appropriately transitioning of care can protect the quality of that care and improve consumers' satisfaction with a new health plan or product. The Commission's recommendations are designed to ease the impact of these transitions from one health insurance product to another and changes in the composition of health plan networks while maintaining the integrity of network models of care. Consumers who are undergoing a course of treatment for a chronic or disabling condition (or who are in the second or third trimester of a pregnancy) at the time they involuntarily change health plans or at a time when a provider is terminated by a plan for other than cause should be able to continue seeing their current specialty providers for up to 90 days (or through completion of postpartum care) to allow for transition of care.

Implications of the Right

Health plans will need to comply with network adequacy standards. Because these changes are primarily to be carried out within existing networks, there should not be a significant increase in costs to health plans or enrollees. Many licensed plans already meet these requirements as laid down by the National Association of Insurance Commissioners (NAIC) in its Managed Care Plan Network Adequacy Model Act. Plans also will need to develop processes to comply with requirements regarding continuity of care and ease of access to specialists within their network of providers.

Consumers will need to exercise their right to choice by using good judgment and providing direct feedback to plans about their level of satisfaction with the network provided for them.

Quality Oversight Organizations will need to incorporate network adequacy standards into their review activities.

References and Selected Reading

Bernstein AB, Dial TH, Smith MD. "Women's Reproductive Health Services in Health Maintenance Organizations." *West J. Med* 1995; 163[suppl]:15-18.

Cowan, CA, Braden, BR, McDonnell, PA, et al. "Business, Households and Government: Health Spending, 1994." *Health Care Finance Rev*; Summer 1996; 17(4):157-178.

Davis K, Collins KS, Schoen C, et al. "Choice Matters: Enrollees' Views of Their Health Plans." *Health Affairs*; Summer 1995; 99-112.

Davis K, Schoen C. testimony before the Advisory Commission on Consumer Protection and Quality in the Health Care Industry; June 25, 1997.

Employee Benefits Research Institute. *EBRI Databook on Employee Benefits*, Washington DC, 1997.

Gabel JR (KPMG Peat Marwick LLP). testimony before the Advisory Commission on Consumer Protection and Quality in the Health Care Industry; June 23, 1997.

Kaiser Family Foundation and the Agency for Health Care Policy and Research (AHCPR). *Americans as Health Care Consumers: The Role of Quality Information*. Princeton Survey Research Associates; October 1996.

KPMG Peat Marwick, Health Insurance Association of America. *Sourcebook of Health Insurance Data*. Washington, DC; 1996.

O'Malley AS, Mandelblatt J, Gold K, Cagney KA, Kerner J. "Continuity of Care and the Use of Breast and Cervical Cancer Screening Services in a Multiethnic Community." *Arch Intern Med* 1997; 157(13):1462-70.

National Association of Insurance Commissioners. *Managed Care Plan Network Adequacy Model Act*. Model Regulation Service; October 1996.

Robinson S, Brodie M. "Understanding the Quality Challenge for Health Consumers: The Kaiser/AHCPR Survey." *Journal on Quality Improvement*, May 1997; 23(5):239-244.

Consumer Bill of Rights and Responsibilities

Chapter Three

Access to Emergency Services

Statement of the Right Consumers have the right to access emergency health care services when and where the need arises. Health plans should provide payment when a consumer presents to an emergency department with acute symptoms of sufficient severity -- including severe pain -- such that a "prudent layperson" could reasonably expect the absence of medical attention to result in placing that consumer's health in serious jeopardy, serious impairment to bodily functions, or serious dysfunction of any bodily organ or part.

To ensure this right:

- Health plans should educate their members about the availability, location, and appropriate use of emergency and other medical services; cost-sharing provisions for emergency services; and the availability of care outside an emergency department.
- Health plans using a defined network of providers should cover emergency department screening and stabilization services both in network and out of network without prior authorization for use consistent with the prudent layperson standard. Non-network providers and facilities should not bill patients for any charges in excess of health plans' routine payment arrangements.
- Emergency department personnel should contact a patient's primary care provider or health plan, as appropriate, as quickly as possible to discuss follow-up and post-stabilization care and promote continuity of care.

Rationale

In 1995, Americans paid an estimated 96.5 million visits to emergency departments, nearly 37 visits per 100 persons (Stussman, 1997). By tradition, emergency departments (EDs) have handled a spectrum of illness, but have had the primary mission of treating those with acutely serious, even life-threatening, medical conditions. Emergency services can be defined as services that are needed or appear to be needed immediately because of injury or sudden illness that threatens serious impairment of any bodily function, and/or serious dysfunction of any bodily organ or part.

Patients go to the emergency department with nonurgent problems for various reasons. Economic and geographic barriers to other forms of care, the lack of a regular provider, and other factors can and do prompt patients to turn to the emergency department for primary and other nonurgent care. Apart from lack of health insurance coverage, nonfinancial barriers to primary care encourage patients to seek evaluation and treatment in the ED. These include problems with work schedules, access to transportation, and concerns about personal safety (Rask, Williams, Parker, et al., 1994). Physician offices and primary care clinics often have limited hours of operation, while EDs are open 24 hours a day. Medicaid beneficiaries, who have a history of limited access to regular providers, have particularly strong relationships with EDs as the provider of first and last resort. Nonurgent visits to the ED can be costly, contribute to overcrowded waiting rooms, divert resources away from other hospital-based care, and compromise the coordination and continuity of care.

But drawing the line between urgent and nonurgent use of the ED is not an easy decision for providers, health plans, and consumers. Criteria -- both prospective and retrospective -- for appropriate ED use are in many ways inadequate. By one criterion, a patient's ED visit might be deemed appropriate, and by another, not so (Lowe and

Bindman, 1997). Health care professionals do not agree among themselves about the need for urgent care among emergency department patients (Gill, Reese, and Diamond, 1996). In a survey of 56 hospital EDs, 5.5 percent of patients initially classified by triage nurses as nonurgent were later admitted to the hospital from the ED (Young, Wagner, Kellerman, et al., 1996). Studies estimate that those presenting with nonurgent problems to the ED range from 6.3 percent (Cunningham, Clancy, and Cohen, et al., 1995) to 54.2 percent (Stussman, 1997) of ED visits.

To better manage care and costs in the ED setting, indemnity and managed care plans use a range of tools that includes requirements for prior authorization and imposition of higher cost-sharing for use of out-of-network emergency departments. A 1989 survey of HMO medical directors found coverage policies for ED use across the HMO industry to be fairly uniform (Kerr, 1989). Unless the condition is life-threatening, patients must obtain prior authorization before seeking emergency care services in 80 percent of the responding HMOs, and 38 percent limited their coverage to the EDs of selected network hospitals. A study undertaken by the Center for Health Policy Studies shows that private indemnity insurers have adopted many of these same practices in their fee-for-service arrangements (PPRC, 1996).

A growing set of State and Federal laws and regulations clarify and protect consumers' access to appropriate emergency services. The Emergency Medical Treatment and Labor Act (EMTALA) requires all Medicare participating hospitals to evaluate whether a patient has an emergency medical condition and, if so, to stabilize the patient. The Balanced Budget Act of 1997 requires health plans participating in Medicare or Medicaid to reimburse for emergency services using a "prudent layperson" standard. Numerous States also have adopted this standard for access to emergency services. The Commission's recommendation seeks to create uniformity in all States.

Implications of the Right

Health care providers. Health care providers will need to work to educate consumers about the appropriate use of emergency department services while working to increase the hours and locations of primary care clinics and other facilities to ease access to such services outside of emergency departments. Emergency department personnel need to make strong efforts to ensure the continuity of care of emergency patients by communicating with patients' primary care providers. Efforts should be made to assist consumers with language, communication, or other barriers.

Health plans. Health plans need to expand consumer education efforts and, when it is within their control, expand hours and location of primary care facilities to facilitate access to such services outside of emergency departments. Plans need to ensure that their coverage and payment policies are consistent with the "prudent layperson" standard.

Consumers. Consumers need to become more familiar with the location and hours of nonemergency care settings and strive to make greater use of such facilities when appropriate. Consumers should communicate with their providers and plans to understand any restrictions on their access to emergency services.

References and Selected Reading

Cunningham PJ, Clancy CM, Cohen WJ. "The Use of Hospital Emergency Departments for Nonurgent Health Problems: A National Perspective." *Med Care Res Rev* 1995; 52(4):453-474.

Gill JM, Reese CL, Diamond JJ. "Disagreement among Health Care Professionals about the Urgent Care Needs of Emergency Department Patients." *Ann Emerg Med* 1996; 28(5):474-479.

Kerr HD. "Access to Emergency Departments: A Survey of HMO Policies." *Ann Emerg Med* 1989; 18(3):274-277.

Lowe RA, Bindman AB. "Judging Who Needs Emergency Department Care: A Prerequisite for Policy-Making." *Am J Emerg Med* 1997; 15(2):133-136.

Physician Payment Review Commission. *Managing Medicare's Fee-for-Service Program. 1996 Annual Report to Congress*. Washington, DC: PPRC; 1996; 196-199.

Rask KJ, Williams MV, Parker RM, et al. "Obstacles Predicting Lack of a Regular Provider and Delays in Seeking Care for Patients at an Urban Public Hospital." *JAMA* 1994; 271:1931-1933.

Stussman BJ. *National Hospital Ambulatory Medical Care Survey: 1995 Emergency Department Survey. Advance Data from Vital and Health Statistics* (no. 285). Hyattsville, MD: National Center for Health Statistics; 1997.

Young GP, Wagner MB, Kellermann AL, et al. "Ambulatory Visits to Hospital Emergency Departments: Patterns and Reasons for Use." *JAMA* 1996; 276:460-465.

Consumer Bill of Rights and Responsibilities

Chapter Four

Participation in Treatment Decisions

Statement of the Right Consumers have the right and responsibility to fully participate in all decisions related to their health care. Consumers who are unable to fully participate in treatment decisions have the right to be represented by parents, guardians, family members, or other conservators.

In order to ensure consumers' right and ability to participate in treatment decisions, health care professionals should:

- Provide patients with easily understood information and opportunity to decide among treatment options consistent with the informed consent process. Specifically,
 - Discuss all treatment options with a patient in a culturally competent manner, including the option of no treatment at all.
 - Ensure that persons with disabilities have effective communications with members of the health system in making such decisions.
 - Discuss all current treatments a consumer may be undergoing, including those alternative treatments that are self-administered.
- Discuss all risks, benefits, and consequences to treatment or nontreatment.
- Give patients the opportunity to refuse treatment and to express preferences about future treatment decisions.
- Discuss the use of advance directives -- both living wills and durable powers of attorney for health care -- with patients and their designated family members.
- Abide by the decisions made by their patients and/or their designated representatives consistent with the informed consent process.

To facilitate greater communication between patients and providers, health care providers, facilities, and plans should:

- Disclose to consumers factors -- such as methods of compensation, ownership of or interest in health care facilities, or matters of conscience -- that could influence advice or treatment decisions.
- Ensure that provider contracts do not contain any so-called "gag clauses" or other contractual mechanisms that restrict health care providers' ability to communicate with and advise patients about medically necessary treatment options.
- Be prohibited from penalizing or seeking retribution against health care professionals or other health workers for advocating on behalf of their patients.

Rationale

Consumers depend on health care professionals to provide them with expert consultation and advice on how to stay healthy or how to cure or palliate their health and medical problems. Unlike many other consumer transactions, the asymmetry of information between consumer and health care provider often is great. Decisionmaking also often occurs at a time of illness, which can undermine the patient's ability to act most effectively in his or her own interest.

Patient and Provider Communication

Relationships between consumers and health care professionals are most rewarding and likely to result in positive outcomes when they are characterized by open communication and active participation of patients in the treatment process. Patient participation in treatment is an essential part of compliance, and compliance improves the effectiveness of care and treatment.

The benefits of patient participation go beyond just the anticipated therapeutic effect of the intervention (Czajkowski and Chesney, 1990). For example, the Coronary Drug Project Research Group (1980), which studied the efficacy and safety of several lipid-lowering drugs, found that even among patients who only took placebos, good adherers had a much lower 5-year mortality rate (15 percent) than did poor adherers (24.6 percent).

Patient participation in treatment decision making also leads to improved satisfaction with care and better quality of life. For example, in a study of patients with early breast cancer, it was found that those who believed they were more responsible for treatment decisions and had more choice of treatment reported higher quality of life than those who perceived themselves as less in control of the treatment decisions (Street and Voigt, 1997).

To participate in decisionmaking about their care, consumers must have complete information about treatment options -- including the alternative of no intervention -- as well as the risks, benefits, and consequences of such options. Yet evidence suggests that clinical practice often falls short of these expectations. A 1988 study of hospitalized patients found that physicians discussed test or treatment rationale in only 43 percent of cases and alternatives in 12 percent of cases (Wu and Pearlman, 1988). Physicians shared with patients information about benefits in 34 percent of cases and risks in 14 percent of cases.

The continued development of communications technologies to help consumers more fully understand their treatment options and to evaluate the potential risks and benefits of treatments should be encouraged, for example, the use of videos to help men with prostate cancer evaluate the risks and benefits of surgery versus a "watchful waiting" strategy (Wennberg, 1995) and to help men with benign prostatic hypertrophy sort out options for treatment (Wagner et al., 1995).

Increasingly, effective communication between providers and patients demands some degree of cultural competence. By the year 2000, nearly one-quarter of the U.S. population will be members of racial or ethnic "minority" groups; this will grow to 47.5 percent by the middle of the next century. Cultural competence refers to the "demonstrated awareness and integration of three population-specific issues: health-related beliefs and cultural values, disease incidence and prevalence, and treatment efficacy" (Lavizzo-Mourey and Mackenzie, 1996). Effective communication for people with communication disabilities may require health care providers to provide auxiliary aids and services and remove certain communication barriers.

It also is imperative that providers be aware of and comply with their patients' decisions with respect to advance directives. Once a patient makes a decision, the health care team should respect this treatment choice. Yet there is clear evidence that this is not happening in far too many instances. Teno et al. (1995) studied 4,301 patients hospitalized in 6 hospitals and found that physicians often were unaware of their patients' wishes. In 47 percent of cases, physicians reported that they did not know of their patients' expressed desire for a "do not resuscitate" order. In another study focusing on nursing home residents transferred to hospitals, Davis, Southerland, Garrett, et al. (1991) found that medical treatment was consistent with advance directives in 75 percent of the 96 cases studied.

Organizational and Contractual Issues

There are a variety of organizational and contractual factors that also may influence communication between patients and providers. These include financial arrangements and contractual restrictions or sanctions that may inhibit the free exchange of information.

Much attention has focused in recent years on the potential effects of providers' financial incentives on treatment. Methods of compensating physicians can be a powerful mechanism to change provider practice, either to improve the quality of care provided to consumers or to reduce the costs of that care. But poorly designed compensation arrangements also can result in inappropriate use (including both overuse and underuse) and barriers to care.

All methods of compensating physicians and other health care providers create some form of incentive for behavior. Various approaches are used to offset the potential adverse effects of compensation arrangements. For example, fee-for-service systems may use utilization review mechanisms to temper incentives toward overutilization of health care services. Capitation systems may incorporate measures of quality and consumer satisfaction to minimize incentives toward overutilization. Similarly, salaried arrangements may use bonuses to encourage higher provider productivity and exemplary performance.

In 1996, the Health Care Financing Administration promulgated rules concerning the use of certain types of financial arrangements on behalf of health plans serving Medicare or Medicaid beneficiaries. These rules stipulate that compensation arrangements "may not include any direct or indirect payments to physicians or groups as an inducement to limit or reduce necessary services furnished to an individual enrollee who is covered under the managed care organization's contract." These regulations also require disclosure of information about arrangements that transfer substantial financial risk to the health care provider. If the compensation methods used places the physician or physician group at substantial financial risk, then the health plan must survey enrollees about access and satisfaction with the quality of services, and institute adequate and appropriate stop-loss protections.

In addition to financial incentives, contract rules that restrict providers' ability to advise patients about medically necessary treatment options have been the subject of much concern. Health care providers must be able to advocate for their patients without constraint or fear of reprisal. A report by the General Accounting Office (GAO, 1997) reported: "Of the 529 HMOs in our study, none used contract clauses that specifically restricted physicians from discussing all appropriate medical options with their patients. Two-thirds of responding plans and 60 percent of the contracts submitted had a nondisparagement, nonsolicitation, or confidentiality clause that some physicians might interpret as limiting communication about all treatment options. However, contracts with such business clauses often contained anti-gag language stating that the physician should not misconstrue the contract of a specific provision as restricting medical advice to patients or that the physician should foster open communication." As of mid-1997, 25 States had prohibited the use of such clauses in managed care contracts with physicians and legislation was pending in 23 other States (Health Policy Tracking Service, 1997). In December 1996, HCFA banned the use of gag rules under the Medicare program and in February 1997, HCFA took similar action regarding health plans' participating in Medicaid.

Implications of the Right

Consumers must take a more active part in the treatment decision process. Information can be empowering, but navigating the health care system requires patient effort, from completing advance directives to preparing questions for an office visit. This requires that the consumer ask questions, understand and give informed consent, and become a full partner in treatment decisions with his or her health care provider.

Health care providers also have the central role in ensuring the patient's participation in treatment decisions, including compliance with informed consent. They will need to improve their skills in providing information about the medical and scientific evidence underlying different treatment options to patients and their families; strive to overcome cultural and language and communication barriers; and keep abreast

of the latest and best available treatment options. At the same time, they will need to do a better job of listening to their patients and following their decisions, including the decision to forgo treatment or certain types of treatment. Health care providers should assume this responsibility well before a patient reaches a hospital door. To hold the trust of patients, providers will need to disclose financial incentives that may introduce bias into treatment decisionmaking and to avoid such incentives when the balance is tipped against the patient. To be above any potential bias, providers must avoid self-referral arrangements that can cloud their professional judgment. And, finally, health care providers are and should be the most effective advocates for their patients' rights.

Health care facilities and plans must create and maintain an environment supportive of consumer participation in treatment decisions. In the office practice, this means ensuring adequate visit time for patients and providing support for shared decisionmaking programs when questions about care linger, arise after hours, or require further explanation. Health plans can play a significant role in educating patients on how to get the most out of their visit with a health care provider. They can arrange for translator services for patients and continuing education courses for providers to assure cultural and language competency. By statute, health plans and hospitals have obligations to educate the public about the use of advance directives. As importantly, once advance directives are signed, these documents must become part of the patient's health record and must move with the patient from care setting to care setting. In establishing provider compensation arrangements, health plans and facilities must be vigilant in guarding against the unintended, negative consequences of financial incentives by implementing programs to monitor quality of care and patient satisfaction. The nature of these incentives ought to be disclosed to patients and providers. In contracting with health care providers, plans and facilities should not restrict the provider's ability to discuss treatment options with the patient and not take reprisal upon the health care provider who serves as patient advocate.

References and Selected Reading

Coronary Drug Project Research Group. "Influence of Adherence to Treatment and Response of Cholesterol on Mortality in the Coronary Drug Project." *N Engl J Med* 1980; 18:1038-1041.

Czajkowski SM, Chesney MA. "Adherence and the Placebo Effect." in Schumacher, SK, Schron EB, Ockene JK, et al., editors. *The Handbook of Health Behavior Change*. New York: Springer Publishing Company; 1990.

Davis M, Southerland LI, Garrett JM, et al. "A Prospective Study of Advance Directives for Life-Sustaining Care." *N Engl J Med* 1991; 324:882-8.

General Accounting Office. *Report to Congressional Requesters. Managed Care: Explicit Gag Clauses Not Found in HMO Contracts, But Physician Concerns Remain*. August 1997.

Health Policy Tracking Service. *Issue Brief: Bans on Gag Clauses*. April 1997.

Lavizzo-Mourey R, Mackenzie ER. "Cultural Competence: Essential Measurements of Quality for Managed Care Organizations." *Ann Intern Med* 1996; 124(10):919-921.

Street RL, Voigt B. "Patient Participation in Deciding Breast Cancer Treatment and Subsequent Quality of Life." *Med Decis Making* 1997; 17:298-306.

Teno JM, Hakim RB, Knaus WA, et al. "Preferences for Cardiopulmonary Resuscitation; Physician-Patient Agreement and Hospital Resource Use. The SUPPORT Investigators." *J Gen Intern Med* 1995; 10(4):179-86.

Wagner EH, Barrett P, Barry MJ, Barlow W, Fowler FJ. "The Effect of a Shared Decision Making Program on Rates of Surgery for Benign Prostatic Hyperplasia; Pilot Results." *Med Care* 1995; 33:765-770.

Wennberg J (Producer). *Treatment Choices for Prostate Cancer* [Film]. Hanover, NH: Foundation for Informed Decision Making, 1995.

Consumer Bill of Rights and Responsibilities

Chapter Five

Respect and Nondiscrimination

Statement of the Right Consumers have the right to considerate, respectful care from all members of the health care system at all times and under all circumstances. An environment of mutual respect is essential to maintain a quality health care system.

Consumers must not be discriminated against in the delivery of health care services consistent with the benefits covered in their policy or as required by law based on race, ethnicity, national origin, religion, sex, age, mental or physical disability, sexual orientation, genetic information, or source of payment.

Consumers who are eligible for coverage under the terms and conditions of a health plan or program or as required by law must not be discriminated against in marketing and enrollment practices based on race, ethnicity, national origin, religion, sex, age, mental or physical disability, sexual orientation, genetic information, or source of payment.

Rationale

Consumers want to be treated with respect and they want to be treated fairly. An environment of mutual respect is essential to maintain a quality health care system. Incidences of discrimination -- real and perceived -- mar the relationship between consumers and their health care professionals, plans, and institutions. Multiple consumer surveys (Levinson et al., 1997; Davis et al., 1995; Edgeman-Levitan and Cleary, 1996) have found that many consumers' complaints about the current health care system have their root in the perception that people believe they are not being treated with respect.

Respect has been defined as recognizing a "person's capacities and perspectives, including his or her right to hold certain views, to make certain choices, and to take certain actions based on personal values and beliefs" (Faden and Beauchamp, 1986). Manifestations of disrespect in the health care setting described by consumers in recent research (Levinson et al., 1997) and interviews include: poor communication with their doctor, feeling rushed or ignored, lack of dignity during examinations, experiencing extensive waiting room delays, receiving inadequate explanations or advice, having inadequate time with the doctor during routine visits, feeling that complaints are not taken seriously by providers, and feeling that providers are more concerned with holding down the cost of medical care than with giving the best medical care. Conversely, consumers defined respectful treatment as that which takes into consideration the values, preferences, and expressed needs of the patient. In addition, consumers wanted providers to communicate well, to be respectful of the patient's time, and to give emotional support to alleviate the patient's fear and anxiety.

In order to extend consumers the respect they deserve, members of the health care industry should strive to:

- Provide consumers with assurances that disrespect or discrimination of any kind is intolerable.
- Provide consumers with information regarding existing laws prohibiting disrespectful or discriminatory treatment.
- Provide consumers with an appropriate amount of time to fully discuss their concerns and questions.
- Provide consumers with reasonable assistance to overcome language (including limited English proficiency), cultural, physical or communication barriers.
- Provide consumers with a timely notice and explanation of changes in fees or billing practices.
- Avoid lengthy delays in seeing a patient; when delays occur, explain why they occurred and, if appropriate, apologize for such delays.

A key element of respectful and fair treatment is protection against discrimination in the delivery of health care services, and in marketing and enrollment, for those eligible for coverage under the terms and conditions of a health plan or program, based on race, ethnicity, national origin, religion, sex, age, mental or physical disability, sexual orientation, genetic information, or source of payment.

Sex. Disparities in medical treatment based on sex have been documented in a number of areas, including: diagnosis and treatment of coronary artery disease (Beery, 1995), kidney transplantation and dialysis, heart transplantation, cardiac catheterization, and diagnosis of lung cancer (AMA Council on Ethical and Judicial Affairs, 1991). Researchers have found that women are less likely to have diagnostic testing, even when functional disability and risk are higher. Women's complaints are seen as less urgent, and fewer referrals follow as a result of this belief (Tobin et al., 1987). Disparities have also been found in the quality of the doctor-patient relationship. For example, one-quarter of women (compared with 12 percent of men) reported that they have been "talked down to" or "treated like a child by a physician," and 17 percent of women (compared with 7 percent of men) had been told that a medical condition they experienced was "all in their head" (The Commonwealth Fund, 1993; Horton, 1995).

Race, ethnicity, national origin, and religion. Discrimination on the basis of race, ethnicity, national origin, or religion in the provision of health care has also been well documented. There is evidence of disparities in the quality of care, access to health care (because of language or geographic barriers), and the amount of care given to minorities as compared with others (Kahn et al., 1994; Giles et al., 1995; Rosenbaum et al., 1997; Smollar, 1988). In the case of facilities or individuals who accept Federal funds, Federal civil rights statutes prohibit the denial of services; the provision of a different service or services in a different manner from those provided to others; and the segregation of or separate treatment of individuals in any matter related to receiving services (Office of Civil Rights, 1990).

Age. Discrimination against consumers based on their age also occurs in the health care industry including: less aggressive treatment for elderly women with breast cancer and lower than average referral rates for mental health services in older people (Nattinger et al., 1992; Osteen et al., 1992; Ayanian et al., 1993). The Age Discrimination Act of 1972 also prohibits discrimination based on age by any institution or health care provider who accepts Federal funds.

Sexual orientation. Gay and lesbian patients have received reduced care or have been denied care because of their sexual orientation (AAPHR, 1994). Discrimination against gay/lesbian consumers has sometimes been compounded by fears of HIV.

Disability status. There is an extensive history of discrimination against people with disabilities and chronic illnesses that has led to action by Federal and State Government. The Americans with Disabilities Act of 1990 (ADA) prohibits discrimination against individuals with real or perceived disabilities in employment, public services, public accommodations, communications, and employer-provided health insurance. The Health Insurance Portability and Accountability Act of 1996 prohibits the exclusion of an individual from the group insurance market for more than 12 months based on a preexisting medical condition. The Mental Health Parity Act of 1996 prohibits differential lifetime or annual caps on coverage for physical and mental illnesses in certain situations.

Despite passage of these landmark laws, not all Americans living with disabilities or adverse medical conditions have access to health coverage at a cost they believe is fair or affordable. This is particularly true for consumers attempting to purchase coverage in the individual insurance market. Research into further refinements in the insurance market is needed to assist these individuals. The Commission strongly urges insurers, public and private purchasers, State and Federal Governments, and others to explore all policy options to make health coverage available and affordable to Americans who wish to obtain it, especially those who are living with mental or physical disabilities and chronic illnesses.

Finally, despite recent improvements, many health care facilities remain inaccessible to individuals with disabilities (Savage, 1997). The Commission believes that elimination of physical and communication

barriers in health care facilities should be a higher priority for government agencies charged with enforcing the ADA.

Source of payment. The health care system currently is undergoing an historic transformation in which low-income Medicaid beneficiaries are being enrolled into private health plans. While this is a positive development in terms of access for traditionally vulnerable populations to high-quality care, it is almost certain to create additional tensions that could be manifest in discrimination. Providers who agree to accept Medicaid beneficiaries must provide equal access, care, and waiting times to those patients. It will be vitally important for State and Federal agencies to closely monitor the provision of care to Medicaid beneficiaries as they move into new health plans.

Implications of the Right

Consumers will need to be vigilant in reporting instances of discrimination based on the factors discussed in this chapter. Consumers also must extend the same level of respect to health care providers and others in the health care system that they demand of same. An environment of mutual respect is essential to a healthy relationship between consumers and those who care for them.

Health care professionals and other health workers have the most direct contact with patients and, therefore, have the greatest responsibility to treat health care consumers with respect and to ensure that they do not discriminate. Providers have a responsibility to listen to patients and take their concerns and complaints seriously. Providers also have a responsibility to monitor their treatment of patients to assure they are treated with respect and nondiscrimination and to correct problems when they occur.

Health care facilities that renovate existing facilities or construct new ones must meet a high standard of access in order to avoid discriminating against persons with disabilities. While there is no ADA requirement to "retrofit" existing facilities to make them accessible, there is a responsibility to remove "readily achievable" physical and communication barriers. All health care providers should assess the level of access in their medical facilities and take steps to provide effective communication and unimpeded physical access to the maximum extent possible.

Health plans will need to examine the standards and incentives that exist within their systems that may inadvertently discourage providers from attending to the interpersonal aspects of health care quality that can be manifest as disrespect. Consumers enrolled in health plans with defined networks of providers should have access to their plans' participating providers, without regard to the source of their coverage (e.g., Medicare, Medicaid, employer-sponsored plan).

Quality oversight organizations should utilize tools that allow accurate measurement of dimensions of health care quality that reflect consumer concerns about being treated with respect. Public disclosure of these findings, together with measurements of clinical quality of care, cost, benefit, and other salient information can allow consumers to determine the relative importance they place on such information and make their purchasing decisions accordingly.

Health care worker education and training programs need to recognize and act upon the need for improvements in communication skills by providers. Receiving inadequate explanations and advice, having inadequate time to receive answers to questions, and failure to attend to the need for emotional support can have adverse consequences on health outcomes (Bame et al., 1993; Patterson et al., 1991; Juncos, 1990). Similarly, education and training programs need to develop and implement course content addressing the significance of cultural attitudes on the effectiveness of health care and the importance of being sensitive to the varying needs of people with disabilities, including those with sensory or cognitive disabilities, who often require auxiliary aids or extra time and plain-language explanation to ensure effective communication. Health plans, hospitals, and other large institutional providers are encouraged to have on-site interpreters for any language population that exceeds a specified standard (e.g., 5 percent or

more) and telephone interpreter services for other language minorities. Written material provided to patients should also be translated for the larger linguistic groups.

References and Selected Reading

AMA Council on Ethical and Judicial Affairs. *Genetic Disparities in Clinical Decision-Making: Council Report*. January 1991; 15(4):25-35.

American Associations of Physicians for Human Rights. *Anti-Gay Discrimination in Medicine: Results of a National Survey of Lesbian, Gay and Bisexual Physicians*. San Francisco; 1994.

Ayanian JZ, Kohler BA, Abe T, et al. "The Relation Between Health Insurance Coverage and Clinical Outcomes among Women with Breast Cancer." *N Eng J Med* 1992; 326:1102-1107.

Bame S, Petersen N, et al. "Variation in Hemodialysis Patient Compliance According to Demographic Characteristics." *Soc Sci Med*, Oct. 1993; 37(8):1035-1043.

Beery TA. "Diagnosis and Treatment of Cardiac Disease: Gender Bias in the Diagnosis and Treatment of Coronary Artery Disease." *Heart & Lung*. November 1995; 24(4):427-435.

The Commonwealth Fund. *The Commonwealth Fund Survey of Women's Health*. New York: July 1993.

Davis K, Collins KS, et al. "Choice Matters: Enrollees' Views of Their Health Plans." *Health Affairs*, Summer 1995; 9-112.

Edgeman-Levitan S, Cleary PD. "What Information Do Consumers Want and Need? A Synthesis of Research to Date, Plus Interviews with Health Plan Managers and Consumer Advocates." *Health Affairs*, Winter 1996; 15(4):42-56.

Faden R, Beauchamp T. *A History and Theory of Informed Consent*. New York: Oxford University Press, 1986; 8-9.

Giles WH, Anda RF, Casper ML, et al. "Race and Sex Differences in Rates of Invasive Cardiac Procedures in US Hospitals." *Arch Intern Med* 1995; 155.

Horton JA, ed. *The Women's Health Data Book: A Profile of Women's Health in the United States*. Washington, DC: Elsevier; 1995.

Kahn KL, Pearson ML, et al. "Health Care for Black and Poor Hospitalized Medicare Patients." *JAMA* 1994; 27(1).

Juncos LI. "Patient Compliance and Angiotensin Converting Enzyme Inhibitors in Hypertension." *J Cardiovascular Pharmacol* 1990; 15(3):S22-S25.

Levinson W, Roter DL, et al. "Physician-Patient Communication: The Relationship with Malpractice Claims among Primary Care Physicians and Surgeons." *JAMA* 1997; 277(7):553-559.

Nattinger AB, Gottlieb MS, Veum J, et al. "Geographic Variation in the Use of Breast-Conserving Treatments for Breast Cancer." *N Eng J Med* 1992; 326:1102-1107.

Office of Civil Rights. *Fact Sheet*. U.S. Department of Health and Human Services; 1990.

Osteen RT, Steele GD Jr, Menck HR, et al. "Regional Differences in Surgical Management of Breast Cancer." *CA Cancer J Clin* 1992; 42:39-43.

Patterson R, Greenberger PA, et al. "Potentially Fatal Asthma; the Problem of Noncompliance." *Ann Allergy*, Aug 1991; 67(2):138-142.

Rosenbaum S, Serrano R, et al. "Civil Rights in a Changing Health Care System," *Health Affairs* Jan-Feb 1997.

Savage E. U.S. Department of Justice. verbal communication; September 1997.

Smollar D. "Success of Indochinese Students May Vary with Ethnic Factors." *Los Angeles Times*, Feb 16, 1988.

Tobin JN, et al. *Ann Internal Med.* 1987;107.

Consumer Bill of Rights and Responsibilities

Chapter Six

Confidentiality of Health Information

Statement of the Right Consumers have the right to communicate with health care providers in confidence and to have the confidentiality of their individually identifiable health care information protected. Consumers also have the right to review and copy their own medical records and request amendments to their records.

In order to ensure this right:

- With very few exceptions, individually identifiable health care information can be used without written consent for health purposes only, including the provision of health care, payment for services, peer review, health promotion, disease management, and quality assurance.
- In addition, disclosure of individually identifiable health care information without written consent should be permitted in very limited circumstances where there is a clear legal basis for doing so. Such reasons include: medical or health care research for which an institutional review board has determined anonymous records will not suffice, investigation of health care fraud, and public health reporting.
- To the maximum feasible extent in all situations, nonidentifiable health care information should be used unless the individual has consented to the disclosure of individually identifiable information. When disclosure is required, no greater amount of information should be disclosed than is necessary to achieve the specific purpose of the disclosure.

Rationale

The legal right to confidentiality of health care information and its essential role in the delivery of quality health care has been recognized by the United States Supreme Court, lower Federal and State courts, and Federal and State lawmakers. Similarly, a health care provider's obligation to protect the confidentiality of health information is universally recognized. The assurance that consumers' health information will remain confidential is "fundamental to effective diagnosis, treatment and healing" (Shalala, 1997).

At the same time, the quality of the health care system also depends on the regular exchange of information between providers, employers, plans, public health authorities, researchers, and other users. The changing structure of the health care system and rapid advances in information technology and medical and health care research have increased the demand for and supply of health information among traditional users such as the treating physician, and new users, such as large networks of providers, information management companies, quality and utilization review committees, and independently contracted service providers. Concerns have been raised that, under the current system of information exchange, various entities can access individually identifiable information without sufficient security safeguards and consent requirements.

Other activities undertaken to improve quality and efficiency may present new risks to the confidentiality of health information. For example, quality oversight activities by plans, providers, accreditation bodies, and regulatory agencies require detailed information about the treatment and benefit status of individual consumers. The growing role of employers in workforce health issues has also contributed to the confidentiality debate.

Congress has made repeated attempts to enact a comprehensive Federal confidentiality law but has, to date, been unsuccessful. The web of protections at the Federal and State level that has evolved in the absence of a comprehensive law leaves many aspects of health information unevenly protected. Specialized Federal protections already exist through statutes that address substance abuse, Medicaid beneficiaries, public health, research, government records, and those living with disabilities.

Several States have enacted comprehensive laws and an effort is currently under way at the National Association of Insurance Commissioners to draft a Protected Health Information Model Act for States. Other safeguards have evolved outside of the legislative arena. Accreditation bodies have incorporated requirements for confidentiality policies and patient consent (JCAHO 1996; NCQA 1997; URAC 1996) and continue to collaborate on security and confidentiality issues (JCAHO/NCQA Joint Session, 1997).

The Health Insurance Portability and Accountability Act of 1996 (HIPAA) required the Secretary of Health and Human Services to submit to the Congress detailed recommendations on: (1) the rights that an individual who is a subject of individually identifiable information should have; (2) the procedures that should be established for the exercise of such rights; and (3) the uses and disclosures of such information that should be authorized or required (Public Law 104-191). On September 11, Health and Human Services Secretary Donna Shalala presented those proposals to the Congress (Shalala, 1997). Under the terms of HIPAA, if Congress fails to enact Federal confidentiality legislation by August 1999, the Secretary of HHS is required to promulgate regulations setting confidentiality standards.

The Secretary recommends a comprehensive Federal confidentiality law that would apply "floor preemption," meaning that the law would require that all States comply with a minimum set of confidentiality requirements but would not preempt stronger State laws.

Section 262 of HIPAA also requires the Secretary of HHS to adopt standards by February 1998 for electronic transmission of financial and administrative health care transactions (including information about claims, eligibility, payment, and injury), unique health identifiers (for individuals, employers, plans, and providers), and security.

The Commission believes that it is essential to establish a comprehensive confidentiality framework and encourages the Congress to move forward expeditiously.

Implications of the Right

Health plans, health providers, employers, and other group purchasers should examine existing confidentiality protections to safeguard against improper use or release of individually identifiable information. The Commission does not intend to impede employers or providers from complying with duties established by law. Health providers, facilities, and plans should develop procedures to ensure that when sensitive services (e.g., mental health, substance abuse, reproductive services, or treatment of sexually transmitted diseases) are involved, standard administrative techniques do not inadvertently disclose information to individuals other than the patient. This is not intended to create two standards of nondisclosure -- one for sensitive medical conditions and another for all others. It is merely a recognition that there may be high level concern about confidentiality with certain medical conditions by some patients.

Law enforcement officers, researchers, and public health agencies should examine their existing policies to ensure that they access individually identifiable information only when absolutely necessary and provide proper safeguards to assure confidentiality.

Consumers should become more aware of the content of their health records and pay particular attention to requests by providers, plans, employers, or others to gain access to those records.

References and Selected Reading

Hurwit C. Citizen Action. testimony before the President's Advisory Commission on Consumer Protection and Quality in the Health Care Industry. May 13, 1997.

Joint Commission on Accreditation of Healthcare Organizations, *Comprehensive Accreditation Manual for Health Care Networks*; 1996.

Joint Commission on Accreditation of Healthcare Organizations and National Committee for Quality Assurance. *Joint Session on Security and Confidentiality of Patient Medical Information*. Washington, DC; 1997.

Lowrance W. *Privacy and Health Research: A Report to the U.S. Secretary of Health and Human Services*; May 1997.

National Association of Insurance Commissioners. "Insurance Information and Privacy Protection Model Act" (October 1992); "Quality Assessment and Improvement Model Act" (July 1996); "Utilization Review Model Act" (October 1996).

National Committee for Quality Assurance (NCQA). "Draft Standards for Accreditation;" 1997.

Public Law No. 104-191, "The Health Insurance Portability and Accountability Act of 1996."

Pyles JC, on behalf of the National Coalition for Patient Rights. "The Right to Medical Privacy: An Indispensable Element of Quality Health Care." Washington, DC; 1997.

Shalala, Donna E. Secretary of Health and Human Services. "Confidentiality of Individually Identifiable Health Information: Recommendations Pursuant to Section 264 of the Health Insurance Portability and Accountability Act of 1996." Submitted to The Committee on Labor and Human Resources and the Committee on Finance of the Senate, and The Committee on Commerce and the Committee on Ways and Means of the House of Representatives. September 11, 1997.

URAC National Network Accreditation Standards (April 1996).

Consumer Bill of Rights and Responsibilities

Chapter Seven

Complaints and Appeals

Statement of the Right All consumers have the right to a fair and efficient process for resolving differences with their health plans, health care providers, and the institutions that serve them, including a rigorous system of internal review and an independent system of external review.

Internal appeals systems should include:

- Timely written notification of a decision to deny, reduce, or terminate services or deny payment for services. Such notification should include an explanation of the reasons for the decisions and the procedures available for appealing them.
- Resolution of all appeals in a timely manner with expedited consideration for decisions involving emergency or urgent care consistent with time frames consistent with those required by Medicare (i.e., 72 hours).
- A claim review process conducted by health care professionals who are appropriately credentialed with respect to the treatment involved. Reviews should be conducted by individuals who were not involved in the initial decision.
- Written notification of the final determination by the plan of an internal appeal that includes information on the reason for the determination and how a consumer can appeal that decision to an external entity.

- Reasonable processes for resolving consumer complaints about such issues as waiting times, operating hours, the demeanor of health care personnel, and the adequacy of facilities.

External appeals systems should:

- Be available only after consumers have exhausted all internal processes (except in cases of urgently needed care).
- Apply to any decision by a health plan to deny, reduce, or terminate coverage or deny payment for services based on a determination that the treatment is either experimental or investigational in nature; apply when such a decision is based on a determination that such services are not medically necessary and the amount exceeds a significant threshold or the patient's life or health is jeopardized.⁶
- Be conducted by health care professionals who are appropriately credentialed with respect to the treatment involved and subject to conflict-of-interest prohibitions. Reviews should be conducted by individuals who were not involved in the initial decision.
- Follow a standard of review that promotes evidence-based decisionmaking and relies on objective evidence.
- Resolve all appeals in a timely manner with expedited consideration for decisions involving emergency or urgent care consistent with time frames consistent with those required by Medicare (i.e., 72 hours).

Rationale

Health care consumers, like other purchasers, have concerns about the service they receive. Unlike other consumers, however, health care consumers have special interests at stake -- the length and quality of their lives. How consumer complaints are addressed has a significant impact on the quality of health services provided and on the satisfaction of consumers with the individuals and institutions that provide them.

Fair and efficient procedures for resolving consumer complaints about their health care serve many purposes. First and foremost, enhanced internal and external review processes will assist consumers in obtaining access to appropriate services in a timely fashion, thus maximizing the likelihood of positive health outcomes. Second, they can be used to bridge communication gaps between consumers and their health plans and providers, and to provide useful information to all parties regarding effective treatment and consumer needs. Third, the opportunity for consumers to be heard by people whose decisions significantly touch their lives evidences respect for the dignity of consumers as individuals and engenders their respect for the integrity of the institutions that serve them.

Properly structured complaint resolution processes should promote the resolution of consumer concerns as well as support and enhance the overall goal of improving the quality of health care. Internal and external complaint and appeal processes should be:

- Timely.
- Fair to all parties.
- Administratively simple.
- Objective and credible.
- Accessible and understandable to consumers.
- Cost and resource efficient.
- Subject to quality review.

Internal and external complaint and appeal processes should not interfere with communication between consumers and their health care providers. For example, in instances where consumers and their providers agree that a service should be reduced or terminated, no written notification of such decisions is needed. Additionally, health care providers who participate in the complaint and appeal processes on behalf of patients should be free from discrimination or retaliation. Likewise, consumers who file a complaint against a provider or plan should be free from discrimination or retaliation.

For the purposes of this chapter, the following definitions are used for the terms "complaints" and "appeals":

Complaint. A "complaint" is any expression of dissatisfaction to a health plan, provider, or facility by a consumer made orally or in writing. This includes concerns about the operations of providers, insurers, or health plans, such as waiting times, the demeanor of health care personnel, the adequacy of facilities or the respect paid to consumers, and claims regarding the right of the consumer to receive services or receive payment for services previously rendered, including the organization's refusal to provide services the consumer believes he or she is entitled to.

Appeal. An "appeal" is a consumer's request for a health plan, facility, or provider or other body to change an initial decision. An appeal process is a procedure for reconsideration of a specific determination made by a health provider, facility or plan.

Current Resolution Processes

Currently, many different procedures are used by group purchasers, health plans, and provider organizations to respond to consumer complaints. Licensed health plans are subject to numerous State and Federal laws, and many also comply with the standards of private accrediting bodies (e.g., NCQA, 1997; JCAHO, 1996; AAHCC/URAC, 1996). Virtually all private and public health plans provide consumers with some form of complaint resolution process. The Commission does not intend by these recommendations to weaken existing consumer protections. These include:

State Licensed Insurance Products. States traditionally have regulated the benefit structure, solvency, rates, and claims process of indemnity insurance companies doing business in the State. Some State insurance regulations require health insurers doing business in the State to provide certain complaint procedures to enrollees (Abraham, 1990). In addition, all 50 States have laws licensing or governing HMOs doing business in the State separate from their laws regulating indemnity insurance products. Many States' laws are based on the model HMO law drafted by the National Association of Insurance Commissioners (NAIC, 1996), which requires HMOs to establish complaint procedures approved by the State's insurance commissioner. An estimated 30 States have some specified complaint procedures that HMOs must follow and at least 7 States now require an expedited appeal for denials of urgently needed care.

ERISA Plans. All employers offering health benefits to their employees through managed care organizations or traditional indemnity insurers must comply with requirements of the Employee Retirement Income Security Act. ERISA requires private employer-provided health benefit plans to disclose certain information to plan participants, to report information to the Federal government, and to pay benefits that are promised under the plan. ERISA regulations generally require employer health plans to approve or deny claims within 90 days and to approve or deny appeals of claims denials within 60 days. Although ERISA health plans are required to establish and disclose complaint and appeals procedures to participants, and to notify participants of claims denials, the plans are not required to provide a particular complaint procedure (Butler and Polzer, 1996). An internal reconsideration of denied claims is stipulated but appeals may be decided by the same plan administrators that initially denied the claim. Determinations must be in writing and state specific reasons for the decision.

Medicare. Under the Medicare fee-for-service system, fiscal intermediaries and carriers must provide a two-step internal review and notification of their final decision before a beneficiary is entitled to seek reconsideration from the Social Security Administration's payment division and the Health Care Financing Administration (Kinney, 1996). Medicare provides a graded appeal process that includes a hearing before an administrative law judge and administrative appeals council review for claims under Part A (hospital coverage) if the amount in controversy is more than \$100; and under Part B (physical and outpatient coverage) if the claims are more than \$500. Claims under Part A and Part B for more than \$1,000 are entitled to judicial review.

HMOs that participate in Medicare are required to provide meaningful internal procedures for resolving complaints about the quality of care, untimely provision of care, or the improper demeanor of health care personnel (Stayn, 1994). HMO decisions to deny coverage for certain treatment, referral outside a plan, or reimbursement for emergency or out-of-area care are subject to an external review and administrative appeal. HCFA has contracted with a private organization, the Center for Health Dispute Resolution, to perform these reconsiderations (Richardson, Phillips, and Conley, 1993). After external review, a Medicare beneficiary enrolled in an HMO who is "dissatisfied by reason of his failure to receive any health service to which he believes he is entitled and at no greater charge than he believes he is required to pay" has a right to Social Security administrative review for controversies more than \$100 and judicial review for controversies more than \$1,000.

Medicaid. The Federal Medicaid statute requires State agencies to provide beneficiaries with a fair hearing and an administrative appeal when their eligibility or requests for services are denied or not acted upon within reasonable time. These State agency determinations can be challenged in State court under State administrative procedure acts or in Federal court. In addition, HMOs that contract to serve Medicaid beneficiaries must establish an internal complaint procedure that will resolve disputes promptly. These internal procedures are subject to review and approval by the State. Medicaid HMO enrollees have the same rights to administrative appeal as do fee-for-service enrollees and no recommendations are made concerning the changing of such rights.

Federal Employees Health Benefit Program. Federal employees and their dependents receive coverage through private insurance carriers, including more than 300 HMOs. Under the FEHBP complaint resolution process, enrollees may bring disputes concerning benefits or services to the Office of Personnel Management for review after asking the plan to reconsider its initial denial and failing to receive a satisfactory reply. OPM seeks to determine whether the enrollee or family member is entitled to the services or supply under the terms of the contract.

Other Approaches. The federal HMO Act requires that to be a "federally qualified HMO," a plan must provide meaningful procedures for hearing and resolving complaints between subscribers and the plan. The written procedures must be easily understood and provided upon request. HMOs are not required to comply with the Act's requirements but may do so to obtain favored status. Other approaches to complaint resolution exist in the Department of Defense's health programs, including the Civilian Health and Medical Program of the Uniformed Services (CHAMPUS).

Implications of the Right

Assuring that all consumers have access to both internal and external processes that satisfy the requirements of this right will require action on virtually every level of the health care industry.

Enhancing Internal Review Systems. Health plans will need to examine their existing internal review systems to assure that consumers receive a timely, understandable notice of decisions to deny, reduce, or terminate treatment or pay claims; notice of the reasons for that determination and of the complaint and appeals procedures available to them; and expedited processes for certain types of cases. While there do not appear to be reliable data indicating how many health plans currently provide internal complaint procedures, most apparently do. Thus, implementation of a general right to file internal complaints, to appeal within a health plan, and to receive a response will not require a majority of health plans to change their current practices significantly. It will be important for quality oversight organizations (State licensure programs, Federal certification programs, and private accrediting bodies) to assure that their standards and review processes adequately address internal complaint and appeal processes of health plans.

Establishing Independent External Appeals Systems. Additional analysis must be done to identify the most effective and efficient methods of establishing the independent external appeals function. Issues to be considered include: mechanisms for financing the external review system; sponsorship of the external review function; design of review processes to assure evidence-based decisionmaking; qualifications of reviewers; consumer cost-sharing responsibilities (e.g., filing fees); and methods of overseeing and

holding external appeals entities accountable. It will also be important to establish an ongoing evaluation mechanism to assess the impact of the external appeals process on access to appropriate services, rates of consumer disputes, litigation rates, consumer satisfaction, and costs. The evaluation mechanism should also assess the impact of certain design characteristics on the effectiveness and efficiency of the external appeals process.

References and Selected Reading

Abraham KS. *Insurance Law and Regulation* 92-139; 1990.

American Accreditation Health Care Commission/Utilization Review Accreditation Commission (AAHCC/URAC). *National Network Accreditation Standards*. Washington, DC; April 1996.

American Association of Health Plans. *Health Plans Announce Policies on Appeals Rights and Emergency Care Coverage*. Washington, DC; January 1997.

American Bar Association, Commission on Legal Problems of the Elderly, *Roundtable on the Resolution of Consumer Disputes in Managed Care*. Washington, DC; 1997.

American Hospital Association: State Issues Forum. *Designing Consumer Protection Standards*. Chicago; 1996.

American Psychiatric Association. *Principles for the Provision of Mental Health and Substance Abuse Treatment Service: A Bill of Rights*. Washington, DC; 1997.

Atkins G, Bass L and K. *ERISA Preemption: The Key to Market Innovation in Health Care*. New York: Corporate Health Care Coalition; 1995.

Butler P, Polzer K. *Private Sector Health Coverage: Variation in Consumer Protections under ERISA and State Law*. Washington, DC: National Health Policy Forum, George Washington University; June 1996.

Citizen Action. *Campaign for Health Security Managed Care Principles*. Washington, DC; January 1997.

Committee on Choice and Managed Care. *Assuring Public Accountability and Information, Improving the Medicare Market: Adding Choice and Protections*. Washington, DC: Institute of Medicine; 1996.

Dame L, Wolfe SM. *Serious Problems for Older Americans in Health Maintenance Organizations*. Public Citizen's Health Research Group; May 1995.

Families USA. *HMO Consumers at Risk: States to the Rescue*. Washington, DC; 1996. Update March 1997.

Joint Commission on Accreditation of Healthcare Organizations, 1996. *Accreditation Manual for Hospitals*. Chicago; 1996.

Kinney ED. "Resolving Grievances in a Managed Care Environment." *Health Matrix* 1996 winter; 6:147-165.

Kinney ED. "Protecting Consumers and Providers under Health Reform: An Overview of the Major Administrative Law Issues." *Health Matrix* 1995; 5:83-138.

Medicare Rights Center. *Medicare Appeals and Grievances: Strategies for System Simplification and Informal Consumer Decisionmaking*. New York; 1996.

National Association of Insurance Commissioners (NAIC). *Health Carriers Grievance Procedure Model Act*. October 1996.

National Committee on Quality Assurance (NCQA). "Standards for Members' Rights and Responsibilities." *Surveyor Guidelines*; April 1, 1997.

Physician Payment Review Commission. *Annual Report to Congress 1997*. Washington, DC; 1997.

Public Policy and Information Fund of New York. *The Managed Care Consumers' Bill of Rights*. New York; October 1995.

Richardson DA, Phillips J, Conley D Jr. *A Study of Coverage Denial Disputes between Medicare Beneficiaries and HMOs*. Pittsford (NY): Center for Health Dispute Resolution, Network Design Group, Inc.; September 1993.

Rodwin MA. "Managed Care and Consumer Protection: What are the Issues?" *Seton Hall Law Review* Winter 1996; 26:1007-1054.

Stayn SJ. "Securing Access to Care in Health Maintenance Organizations: Toward a Uniform Model of Grievance and Appeals Procedures." *Columbia Law Review* June 1994; 94:1674-1720.

Consumer Bill of Rights and Responsibilities

Chapter Eight

Consumer Responsibilities

Statement of Responsibilities In a health care system that protects consumers' rights, it is reasonable to expect and encourage consumers to assume reasonable responsibilities. Greater individual involvement by consumers in their care increases the likelihood of achieving the best outcomes and helps support a quality improvement, cost-conscious environment. Such responsibilities include:

- Take responsibility for maximizing healthy habits, such as exercising, not smoking, and eating a healthy diet.
- Become involved in specific health care decisions.
- Work collaboratively with health care providers in developing and carrying out agreed-upon treatment plans.
- Disclose relevant information and clearly communicate wants and needs.
- Use the health plan's internal complaint and appeal processes to address concerns that may arise.
- Avoid knowingly spreading disease.
- Recognize the reality of risks and limits of the science of medical care and the human fallibility of the health care professional.
- Be aware of a health care provider's obligation to be reasonably efficient and equitable in providing care to other patients and the community.
- Become knowledgeable about his or her health plan coverage and health plan options (when available) including all covered benefits, limitations, and exclusions, rules regarding use of network providers, coverage and referral rules, appropriate processes to secure additional information, and the process to appeal coverage decisions.
- Show respect for other patients and health workers.
- Make a good-faith effort to meet financial obligations.
- Abide by administrative and operational procedures of health plans, health care providers, and Government health benefit programs.
- Report wrongdoing and fraud to appropriate resources or legal authorities.

Rationale

In providing consumers with a set of rights and protections, the Commission believes that individual consumers must assume certain responsibilities. Responsibilities create benefits not only for individual consumers and their families but also for the health care system and society as a whole. Improved health status reduces medical costs

for the patient, the payer, and society.

The Commission, however, does not intend to create a link between an individual's conduct in meeting his or her responsibilities and the obligations of plans and providers to provide covered services.

Increased patient responsibility can improve consumers' sense of self-worth. For example, increased responsibility among individuals living with disabilities has resulted in increased independence for that population (Rodwin, 1994; National Health Council, 1995). In fact, this is the principle behind the independent living movement, where people with disabilities live in their homes with personal assistant services rather than in institutions. Individuals report that increased responsibility for their health has led to improved self-esteem and a greater sense of empowerment.

Promoting consumer responsibility is an essential component of the effort toward involving consumers directly in decisionmaking about their health and medical care. Consumers often perceive that the medical professionals who care for them are acting in a condescending or paternalistic manner. They resent being put in a position of dependence and being treated as if they are infantile and object to the presumption that they are incapable of making choices themselves (Rodwin, 1994).

While the Commission believes that consumers must assume certain responsibilities, it also recognizes that reasonable accommodations must be made for numerous consumers with disabilities. For example, some individuals with physical and mental disabilities require assistance with self care; for some individuals with mental disabilities, noncompliance with treatment regimes is a manifestation of their disability; and some individuals with mental and physical disabilities are unable -- due to their disability -- to clearly communicate their wants and needs and, therefore, rely on the assistance of a designated representative. In each case, the health care system must recognize these issues and accommodate these needs. The Commission also recognizes that there are many other factors, such as occupational hazards, language, and income status, that may pose significant barriers to consumers meeting these responsibilities.

Consumers who are able should take the opportunity to educate themselves with respect to the specifics of their benefit coverage and to learn how to access the health care and services available to them as a result of that coverage. This includes:

- Reading and understanding written information that explains benefit coverage.
- Reading and understanding information that describes health plan processes and procedures to follow when seeking care by a physician, hospital, or other provider.
- Seeking information or clarification of information from the health plan as necessary.
- Using the health plan's processes for addressing complaints or grievances when disputes with providers or health plan procedures arise.

Consumer responsibility is particularly relevant to the broad right to information established in this Consumer Bill of Rights and Responsibilities (see [Chapter One](#)). The Right to Information requires the disclosure of information to consumers either directly or upon request on such things as benefits, cost-sharing, complaints and appeals processes, licensure, accreditation, and performance measures. The Right to Information will improve health outcomes only to the extent that consumers have a choice of health plans and use that information in exercising the choice.

Although there is significant value in promoting the consumers' participation in their own health care by increasing their level of responsibility, it is important to set limits on the amount of responsibility expected. The patient's responsibility to comply with medical advice is limited by the principle of informed consent (Benjamin, 1985). The patient retains the right to choose whether to follow medical advice or not, as long as he or she is willing to accept the health outcome consequences that may result from noncompliance, and the noncompliance does not adversely affect the public (Brock and Wartman, 1994).

Consumers do not have a duty to be subjected to a treatment regime they have good reason to avoid -- for instance, one whose negative side effects outweigh its benefits (Mayer, 1992), or when excessive medication in

an institutional setting is used to "control" residents. Most consumer responsibilities do not extend to those who are incompetent to make decisions, including infants, those who are judged to be mentally incompetent, and comatose patients (Emson, 1995; Mayer, 1992; National Health Council, 1995).

In addition, certain high-risk behaviors (smoking, use of smokeless tobacco, illegal drug use) are addictive and cannot be considered fully under the volitional control of the individual consumer. Caution must be used to avoid "blaming the victim." For example, Bayer (1996) notes that during the history of the AIDS epidemic, "the emphasis on personal responsibility was often associated with condemnation of those whose sexual or drug-using behavior had exposed them to HIV, as well as with calls for invasion of privacy and deprivations of liberty."

Compliance with agreed-upon treatment protocols is a particularly important consumer responsibility. Noncompliance with the taking of medication has particular implications for the health status of consumers. Noncompliance includes taking too much medication, taking medication not prescribed, not taking medication prescribed, altering the prescribed dosage, or altering the time between doses.

Finally, it is important to recognize that while consumers should seek to assume the responsibilities discussed in this report, many factors influence consumers' acceptance of medical advice. Some are related to the health care system itself and others are related to the patient's individual psychology. Imanaka, Araki, et al. (1993) identified patient dissatisfaction with their health care providers and plans as a primary cause of patient noncompliance. Several studies have identified inadequate provider-consumer communication as a contributing factor (Imanaka, 1993; Ross, 1991; Donovan and Blake, 1992; Sluijs, Kok, et al., 1993). This leads to situations where:

- The patient and the prescriber have a different understanding of what the patient is supposed to do.
- The patient lacks information or understanding about the disease, pathology, or symptoms.
- The patient does not understand the correct purpose of the intervention.
- The patient and the health care provider have insufficient time to discuss the full range of issues concerning compliance.

Noncompliant patients also may have underlying psychiatric disorders. Yellowless and Ruffin (1989) found that 40 percent of patients who experience a life-threatening asthma episode have psychiatric disorders. Patients often are trying to balance the requirements of their prescribed medical regimen with other aspects of their life (Donovan and Blake, 1992). Finally, some patients choose not to comply with medical instructions as a way of expressing their attempts to cope with their disease; as a reaction to the way they have been treated by doctors; or as a way of fighting the system by breaking its "symbolic" rules (Ross, 1991).

Implications of the Responsibilities

Consumers will have to play an active role in the treatment and management of their health. Consumers will need to ask more questions of their health care providers, insurers, and institutions. They will need to express their wishes and desires clearly to those who care for them and to their family members in the event of incapacity; this should be done *before* an incapacity occurs. They will need to make sure that they understand a treatment regimen that is prescribed for them before they agree to follow it. Once they have made such an agreement, consumers will need to make every effort to comply and, if they cannot, to notify their provider of their desire or need to change that regimen. Consumers will need to recognize the financial and societal impact of their health care decisions and their health care choices should reflect this consideration.

Health care providers will need to communicate more clearly with their patients and their patients' families about diagnoses, treatment options, and treatment protocols. They will need to make greater efforts to ensure that those matters are clearly understood and agreed to. They will need to work with their patients to ensure that treatment regimens are possible to follow and that changes in treatment are made when possible to meet patients' needs or demands.

Health plans will need to consider ways to encourage greater communication between consumers and health care professionals, including incentives for such communication and acceptance of treatment regimens.

References and Selected Reading

- Adams ML. "The Public Health Impact and Economic Cost of Smoking in Connecticut -- 1989." *Conn Med* April 1994; 58(4):195-8.
- American Hospital Association. *Designing Consumer Protection Standards: State Issues Forum*. Washington, DC; October 1996.
- American Medical Association. *Code of Medical Ethics 1996-1997*. Chicago, Illinois; 1997.
- Arras JD. "Noncompliance in AIDS Research." *Hastings Center Report* September/October 1990; 24-32.
- Bame S, Petersen N, et al. "Variation in Hemodialysis Patient Compliance According to Demographic Characteristics". *Soc Sci Med* October 1993; 37(8):1035-43.
- Bayer R. "AIDS Prevention: Sexual Ethics and Responsibility." *New Eng J of Med* June 6, 1996; 334(23):1540-42.
- Beers MK, Sliwowski, et al. "Compliance with Medication Orders among the Elderly after Hospital Discharge." *Hosp Formul* July 1992; 27:720-724.
- Benjamin J. "Lay Obligations in Professional Relations." *J Med Philos* 1985; 10(1):85-103.
- Brock, DW, Wartman SA. "When Competent Patients Make Irrational Choices." In: Beauchamp T, Walters L, editors. *Contemporary Issues in Bioethics*. 4th ed. Belmont, CA: Wadsworth Publishing; 1994.
- Brody H. "Patients' Responsibilities." In: Reich W, editor. *Encyclopedia of Bioethics*. Revised ed. New York, NY: Simon and Schuster Macmillian; 1995.
- Cebrun AJ. "Consumer Responsibility in the HMO Environment." *J Health Care Poor Underserved* 1994; 5(3):179-181.
- Col N, Fanale JE, Kronhom P. "The Role of Medication Noncompliance and Adverse Drug Reactions in Hospitalizations of the Elderly." *Arch Intern Med* 1990; (150):841-845.
- Donovan JL, Blake DR. "Patient Non-compliance: Deviance or reasoned decision-making?" *Soc Sci Med* 1992; (34)5:507-513.
- Eddy DE. "The Individual vs Society: Resolving the Conflict." *JAMA* May 8, 1991; (265)18:2399-2406.
- Emson HE. "Rights, Duties, and Limits of Autonomy." *Cambridge Q Healthc Ethics* 1995; (4):7-11.
- Harris J, Holm S. "Is There a Moral Obligation Not to Infect Others?" *BMJ* November 1995; 311(4):1215-1217.
- Imanaka Y, Araki S, et al. "Effects of Patient Health Beliefs and Satisfaction on Compliance with Medication Regimens in Ambulatory Care at General Hospitals." *Nippon Eiseigaku Zasshi* June 1993; 48(2):601-11.
- JCAHO. *1996 Comprehensive Accreditation Manual for Health Care Networks*.
- Jensen GJ, Lorish CD. "Promoting Patient Cooperation with Exercise Programs: Linking Research, Theory, and Practice." *Arth Care and Res* Dec 1994; 7(4):181-189.
- Juncos LI. "Patient Compliance and Angiotensin Converting Enzyme Inhibitors in Hypertension." *J Cardiovascular Pharmacol* 1990; (15)Suppl 3:S22-5.
- Macdonald M, Meyer KC, Essig B. *Health Law Guidebook Series*. New York: Matthew Bender, 1996.

Mayer M. "Patients' Duties." *J of Med and Phil* 1992; (17):541-555.

National Highway Traffic Safety Administration. Web page, 1997.

National Health Council. *Principles of Patient Rights & Responsibilities*. Washington, DC: January 1995.

NCQA. "Standards for Member's Rights and Responsibilities." *1997 Surveyor Guidelines*.

Public Citizen Health Research Group Health Letter. June 1995; 11(6).

Rodwin M. "Patient Accountability and Quality of Care: Lessons from Medical Consumerism and the Patients' Rights, Women's Health and Disability Rights Movements." *Am J Law Med* 1994; (20)147.

Ross FM. "Patient Compliance: Whose Responsibility?" *Soc Sci Med* 1991; 32(1):89-94.

Sharkness CM, Snow DA. "The Patient's View of Hypertension and Compliance." *Am J Prev Med* 1992; 8(3):141-146.

Shiffman S. "Refining Models of Dependence: Variations across Persons and Situations." *Br J Addict* May 1991; 86(5):611-15.

Sluijs EM, Kok GJ, et al. "Correlates of Exercise Compliance in Physical Therapy." *Phys Ther* 1993; 73(11).

Sullivan SD, Kreling DH, Hazlet TK. "Noncompliance with Medication Regimens and Subsequent Hospitalizations; A Literature Analysis and Cost of Hospitalization Estimation." *J Res Pharmac Econom* 1990; 2:19-33.

TenHave, Loughlin. "Responsibilities and Rationalities: Should the Patient be Blamed?" *Health Care Anal* 1994; 2(12):112-117.

Tuohey JF. "Moving from Autonomy to Responsibility in HIV-Related Healthcare." *Camb Q Healthc Ethics* 1995; 4(4):64-70.

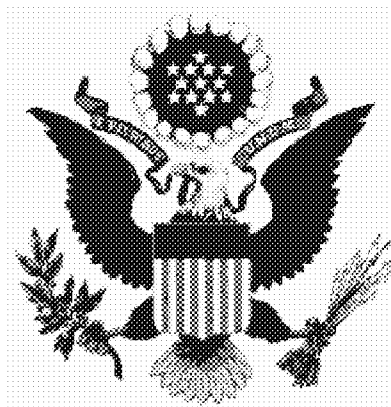
Weiss KB, Wagner DK. "Changing Patterns of Asthma Mortality: Identifying Target Populations at High Risk." *JAMA* 1990; 95(264):1683-7.

Yellowless PM, Ruffan RE. "Psychological Defenses and Coping Styles in Patients Following a Life-Threatening Attack of Asthma." *Chest* 1989; 95:1298-1303.

-
1. The term "health plans" is used throughout this report and refers broadly to indemnity insurers, managed care organizations (including health maintenance organizations and preferred provider organizations), self-funded employer-sponsored plans, Taft-Hartley trusts, church plans, association plans, State and local government employee programs, and public insurance programs (i.e. Medicare and Medicaid).
 2. The right to external appeals does not apply to denials, reductions, or terminations of coverage or denials of payment for services that are specifically excluded from the consumer's coverage as established by contract.
 3. The Commission examined proposals by organizations including: the American Association of Health Plans, the American Association of Retired Persons, the American Hospital Association, the American Medical Association, the Campaign for Health Security, Citizen Action, Families USA, the Health Insurance Association of America, HIP Health Plans, the Health Policy Tracking Service, Kaiser Permanente, Kaiser/Group Health, the Midwest Bioethics Center, the National Association of Insurance Commissioners, the National Committee on Quality Assurance, the National Health Council, the Public

Policy and Education Fund of New York, the Service Employees International Union, the Utilization Review Accreditation Committee, and many others.

4. The term "health plan" is used throughout this report and refers broadly to indemnity insurers, managed care organizations (including health maintenance organizations and preferred provider organizations), self-funded employer-sponsored plans, Taft-Hartley trusts, church plans, association plans, State and local government employee programs, and public insurance programs (i.e., Medicare and Medicaid).
5. In the context of this chapter, health care information is defined as "any information, whether oral or recorded, in any form or medium, that is created or received by a health care provider, health plan, public health authority, employer, life insurer, school, university, health care clearinghouse; and relates to the past, present, or future physical or mental health or condition of an individual, the provision of health care to an individual, or the past, present, or future payment for the provision of health care to an individual."
6. The right to external appeals does not apply to denials, reductions, or terminations of coverage or denials of payment for services that are specifically excluded from the consumer's coverage as established by contract.



[\[About the Commission | Charter | Commission Membership | Press Releases | Meetings \]](#)

Last Revised: Friday, July 17, 1998

This site displays a prototype of a "Web 2.0" version of the daily Federal Register. It is not an official legal edition of the Federal Register, and does not replace the official print version or the official electronic version on GPO's govinfo.gov.

The documents posted on this site are XML renditions of published Federal Register documents. Each document posted on the site includes a link to the corresponding official PDF file on govinfo.gov. This prototype edition of the daily Federal Register on FederalRegister.gov will remain an unofficial informational resource until the Administrative Committee of the Federal Register (ACFR) issues a regulation granting it official legal status. For complete information about, and access to, our official publications and services, go to About the Federal Register on NARA's archives.gov.

The OFR/GPO partnership is committed to presenting accurate and reliable regulatory information on FederalRegister.gov with the objective of establishing the XML-based Federal Register as an ACFR-sanctioned publication in the future. While every effort has been made to ensure that the material on FederalRegister.gov is accurately displayed, consistent with the official SGML-based PDF version on govinfo.gov, those relying on it for legal research should verify their results against an official edition of the Federal Register. Until the ACFR grants it official status, the XML rendition of the daily Federal Register on FederalRegister.gov does not provide legal notice to the public or judicial notice to the courts.

Consultation and Coordination With Indian Tribal Governments

A Presidential Document by the Executive Office of the President on 11/09/2000

DOCUMENT DETAILS

Printed version:

PDF (<https://www.govinfo.gov/content/pkg/FR-2000-11-09/pdf/00-29003.pdf>)

Publication Date:

11/09/2000 (/documents/2000/11/09)

Agency:

Executive Office of the President (<https://www.federalregister.gov/agencies/executive-office-of-the-president>)

Document Type:

Presidential Document

Presidential Document Type:

Executive Order

E.O. Citation:

E.O. 13175 of Nov 6, 2000

E.O. Notes:

Revokes: EO 13084 (/executive-order/13084), May 14, 1998

See: EO 12866 (/executive-order/12866), September 30, 1993; EO 12988 (/executive-order/12988), February 5, 1996; EO 13132 (/executive-order/13132), August 4, 1999; EO 13336 (/executive-order/13336), April 30, 2004; Memorandum of April 29, 1994; EO 14096 (/executive-order/14096), April 21, 2023

Document Citation:

65 FR 67249

Page:

67249-67252 (4 pages)

Document Number:

DOCUMENT STATISTICS

Page views:

52,822

as of 04/16/2024 at 10:15 am EDT

DOCUMENT STATISTICS

PUBLISHED DOCUMENT

Executive Order 13175 (/executive-order/13175) of November 6, 2000

Consultation and Coordination With Indian Tribal

Governments

By the authority vested in me as President by the Constitution and the laws of the United States of America, and in order to establish regular and meaningful consultation and collaboration with tribal officials in the development of Federal policies that have tribal implications, to strengthen the United States government-to-government relationships with Indian tribes, and to reduce the imposition of unfunded mandates upon Indian tribes; it is hereby ordered as follows:

Section 1. Definitions. For purposes of this order:

(a) “Policies that have tribal implications” refers to regulations, legislative comments or proposed legislation, and other policy statements or actions that have substantial direct effects on one or more Indian tribes, on the relationship between the Federal Government and Indian tribes, or on the distribution of power and responsibilities between the Federal Government and Indian tribes.

(b) “Indian tribe” means an Indian or Alaska Native tribe, band, nation, pueblo, village, or community that the Secretary of the Interior acknowledges to exist as an Indian tribe pursuant to the Federally Recognized Indian Tribe List Act of 1994, 25 U.S.C. 479a (<https://www.govinfo.gov/link/uscode/25/479a>).

(c) “Agency” means any authority of the United States that is an “agency” under 44 U.S.C. 3502(1) (<https://www.govinfo.gov/link/uscode/44/3502>), other than those considered to be independent regulatory agencies, as defined in 44 U.S.C. 3502(5) (<https://www.govinfo.gov/link/uscode/44/3502>).

(d) “Tribal officials” means elected or duly appointed officials of Indian tribal governments or authorized intertribal organizations.

Sec. 2. Fundamental Principles. In formulating or implementing policies that have tribal implications, agencies shall be guided by the following fundamental principles:

(a) The United States has a unique legal relationship with Indian tribal governments as set forth in the Constitution of the United States, treaties, statutes, Executive Orders, and court decisions. Since the formation of the Union, the United States has recognized Indian tribes as domestic dependent nations under its protection. The Federal Government has enacted numerous statutes and promulgated numerous regulations that establish and define a trust relationship with Indian tribes.

(b) Our Nation, under the law of the United States, in accordance with treaties, statutes, Executive Orders, and judicial decisions, has recognized the right of Indian tribes to self-government. As domestic dependent nations, Indian tribes exercise inherent sovereign powers over their members and territory. The United States continues to work with Indian tribes on a government-to-government basis to address issues concerning Indian tribal self-government, tribal trust resources, and Indian tribal treaty and other rights.

(c) The United States recognizes the right of Indian tribes to self-government and supports tribal sovereignty and self-determination.

Sec. 3. *Policymaking Criteria.* In addition to adhering to the fundamental principles set forth in section 2, agencies shall adhere, to the extent permitted by law, to the following criteria when formulating and implementing policies that have tribal implications: □

□ Start Printed
Page 67250

(a) Agencies shall respect Indian tribal self-government and sovereignty, honor tribal treaty and other rights, and strive to meet the responsibilities that arise from the unique legal relationship between the Federal Government and Indian tribal governments.

(b) With respect to Federal statutes and regulations administered by Indian tribal governments, the Federal Government shall grant Indian tribal governments the maximum administrative discretion possible.

(c) When undertaking to formulate and implement policies that have tribal implications, agencies shall:

(1) encourage Indian tribes to develop their own policies to achieve program objectives;

(2) where possible, defer to Indian tribes to establish standards; and

(3) in determining whether to establish Federal standards, consult with tribal officials as to the need for Federal standards and any alternatives that would limit the scope of Federal standards or otherwise preserve the prerogatives and authority of Indian tribes.

Sec. 4. *Special Requirements for Legislative Proposals.* Agencies shall not submit to the Congress legislation that would be inconsistent with the policymaking criteria in Section 3.

Sec. 5. *Consultation.* (a) Each agency shall have an accountable process to ensure meaningful and timely input by tribal officials in the development of regulatory policies that have tribal implications. Within 30 days after the effective date of this order, the head of each agency shall designate an official with principal responsibility for the agency's implementation of this order. Within 60 days of the effective date of this order, the designated official shall submit to the Office of Management and Budget (OMB) a description of the agency's consultation process.

(b) To the extent practicable and permitted by law, no agency shall promulgate any regulation that has tribal implications, that imposes substantial direct compliance costs on Indian tribal governments, and that is not required by statute, unless:

(1) funds necessary to pay the direct costs incurred by the Indian tribal government or the tribe in complying with the regulation are provided by the Federal Government; or

(2) the agency, prior to the formal promulgation of the regulation,

(A) consulted with tribal officials early in the process of developing the proposed regulation;

(B) in a separately identified portion of the preamble to the regulation as it is to be issued in the **Federal Register**, provides to the Director of OMB a tribal summary impact statement, which consists of a description of the extent of the agency's prior consultation with tribal officials, a summary of the nature of their concerns and the agency's position supporting the need to issue the regulation, and a statement of the extent to which the concerns of tribal officials have been met; and

(C) makes available to the Director of OMB any written communications submitted to the agency by tribal officials.

(c) To the extent practicable and permitted by law, no agency shall promulgate any regulation that has tribal implications and that preempts tribal law unless the agency, prior to the formal promulgation of the regulation,

(1) consulted with tribal officials early in the process of developing the proposed regulation;

(2) in a separately identified portion of the preamble to the regulation as it is to be issued in the **Federal Register**, provides to the Director of OMB a tribal summary impact statement, which consists of a description of the extent of the agency's prior consultation with tribal officials, a summary of the nature of their concerns and the agency's position supporting the need to issue the regulation, and a statement of the extent to which the concerns of tribal officials have been met; and

Start Printed
Page 67251

(3) makes available to the Director of OMB any written communications submitted to the agency by tribal officials.

(d) On issues relating to tribal self-government, tribal trust resources, or Indian tribal treaty and other rights, each agency should explore and, where appropriate, use consensual mechanisms for developing regulations, including negotiated rulemaking.

Sec. 6. Increasing Flexibility for Indian Tribal Waivers.

(a) Agencies shall review the processes under which Indian tribes apply for waivers of statutory and regulatory requirements and take appropriate steps to streamline those processes.

(b) Each agency shall, to the extent practicable and permitted by law, consider any application by an Indian tribe for a waiver of statutory or regulatory requirements in connection with any program administered by the agency with a general view toward increasing opportunities for utilizing flexible policy approaches at the Indian tribal level in cases in which the proposed waiver is consistent with the applicable Federal policy objectives and is otherwise appropriate.

(c) Each agency shall, to the extent practicable and permitted by law, render a decision upon a complete application for a waiver within 120 days of receipt of such application by the agency, or as otherwise provided by law or regulation. If the application for waiver is not granted, the agency shall provide the applicant with timely written notice of the decision and the reasons therefor.

(d) This section applies only to statutory or regulatory requirements that are discretionary and subject to waiver by the agency.

Sec. 7. Accountability.

(a) In transmitting any draft final regulation that has tribal implications to OMB pursuant to Executive Order 12866 of September 30, 1993, each agency shall include a certification from the official designated to ensure compliance with this order stating that the requirements of this order have been met in a meaningful and timely manner.

(b) In transmitting proposed legislation that has tribal implications to OMB, each agency shall include a certification from the official designated to ensure compliance with this order that all relevant requirements of this order have been met.

(c) Within 180 days after the effective date of this order the Director of OMB and the Assistant to the President for Intergovernmental Affairs shall confer with tribal officials to ensure that this order is being properly and effectively implemented.

Sec. 8. Independent Agencies. Independent regulatory agencies are encouraged to comply with the provisions of this order.

Sec. 9. General Provisions. (a) This order shall supplement but not supersede the requirements contained in Executive Order 12866 (Regulatory Planning and Review), Executive Order 12988 (/executive-order/12988) (Civil Justice Reform), OMB Circular A-19, and the Executive Memorandum of April 29, 1994, on Government-to-Government Relations with Native American Tribal Governments.

(b) This order shall complement the consultation and waiver provisions in sections 6 and 7 of Executive Order 13132 (/executive-order/13132) (Federalism).

(c) Executive Order 13084 (/executive-order/13084) (Consultation and Coordination with Indian Tribal Governments) is revoked at the time this order takes effect.

(d) This order shall be effective 60 days after the date of this order.

□

□ Start Printed
Page 67252

Sec. 10. Judicial Review. This order is intended only to improve the internal management of the executive branch, and is not intended to create any right, benefit, or trust responsibility, substantive or procedural, enforceable at law by a party against the United States, its agencies, or any person.

wj THE WHITE HOUSE, November 6, 2000. Filed 11-8-00; 8:45 am]
[FR Doc. 00-29003

Billing code 3195-01-P

PUBLISHED DOCUMENT



Spinoffs: A review and synthesis

Steven Klepper

Department of Social & Decision Sciences, Carnegie Mellon University, Pittsburgh, USA

Correspondence: Steven Klepper, Department of Social & Decision Sciences, Carnegie Mellon University, 208 Porter Hall, Pittsburgh, PA 15213, USA.
 Tel: 412 268 3235;
 Fax: 412 268 6938;
 E-mail: sk3f@andrew.cmu.edu

Abstract

In the past 10 years, various stylized facts have accumulated regarding the rate at which firms spawn spinoffs, the performance of spinoffs, and the effect of spinoffs on their parents. We review the evidence and use it to reflect on the various theories that have been developed to explain spinoffs. Many questions remain about spinoffs, and we discuss research opportunities associated with these questions. We also discuss implications of the accumulating evidence and theories for public policies bearing on spinoffs.

European Management Review (2009) 6, 159–171. doi:10.1057/emr.2009.18

Keywords: spinoff; learning; agglomeration



Introduction

A cornerstone of neoclassical models of industrial competition is the notion of a queue of potential entrants, ready to enter an industry whenever incumbent firms earn positive economic profits. In many models, the mere prospect of such entrants is sufficient to drive economic profits to zero, and models are commonly closed by assuming all firms earn zero economic profits in equilibrium. But exactly where do these potential entrants come from that drive economic profits to zero? To exact the fear of incumbents tempted to try to price above marginal (and average) cost, potential entrants must have the same capabilities as the incumbents. Indeed, even the incumbents must have the same capabilities to bring about a zero-profit equilibrium. In contrast to the neoclassical tradition, business strategists assume that firms have different capabilities. If properly managed, unique capabilities can give rise to persistent economic profits. But where do these capabilities come from? Do firms have them when they enter industries? If so, then firms will not be alike, and ‘free entry,’ such as it is, will not drive out firm economic profits.

In light of the importance of entry in models of industrial competition, it is surprising how little industry economists and strategists know about where entrants come from and how their backgrounds affect their fates. In recent years, one class of entrants has begun to receive special attention: entrants founded by employees of incumbent firms, which have been referred to as spinoffs or spinouts (sometimes with a hyphen after the spin). Such firms have spurred the evolution of a number of new industries, displacing the early leaders that diversified from related industries (cf. Tilton, 1971; Christensen, 1993). Spinoffs have also

played a key role in the formation and growth of well known industry clusters, including automobile producers in Detroit, tire producers in Akron, Ohio, and semiconductor producers in Silicon Valley (Buenstorf and Klepper, 2009; Klepper, 2009b). The best-performing spinoffs are those founded by employees in the same industry (called intra-industry spinoffs), suggesting that spinoffs are hardly all alike. Even among intra-industry spinoffs, those founded by employees of better performing firms tend to perform better, suggesting that the capabilities of intra-industry spinoffs also fundamentally differ according to their heritage.

But what do we know about the impetus for spinoffs and the factors underlying their performance? Until roughly 10 years ago, remarkably little. Apart from Garvin’s (1983) article drawing attention to spinoffs and offering anecdotal evidence, the only other analysis of spinoffs was an unpublished paper by Brittain and Freeman (1986) on intra-industry spinoffs in the semiconductor industry in Silicon Valley. Much has changed in the past 10 years. As discussed in Klepper and Thompson (2009), a number of studies of spinoffs in high-tech industries have examined the rate at which incumbent firms spawn spinoffs and the factors underlying the performance of spinoffs. Other studies examine the incidence and performance of spinoffs and other startups in large populations, such as for the entire country of Denmark. Spinoffs have also been studied in the context of industry clustering, and some studies have even explored the effects of spinoffs on their (unwitting) ‘parents.’

The purpose of this paper is to step back and consider the nature of the evidence that has been accumulating

about spinoffs and reflect on what it tells us about the spinoff process and its effect on social welfare. As spinoffs have drawn increasing attention, new theories have been proposed for why they occur. After reflecting on the accumulating evidence about spinoffs, the theories will be put to the test by assessing how well they explain the accumulating stylized facts. The reader can make his or her own judgment about which theories agree best with the evidence. But probably all will agree that no theory explains all the patterns and some of the patterns are particularly provocative concerning what might be going on. Much remains to be understood about spinoffs.

Spinoffs no doubt occur for many reasons. Some are voluntary and are engineered by the parent firm. Some are motivated purely by an employee's desire to be his or her own boss, some to satisfy career aspirations, and yet others because of the failure or imminent failure of an individual's employer. These kind of spinoffs tend not to raise deep policy or theoretical questions, and implicitly throughout we will abstract from such spinoffs and concentrate on spinoffs resulting from the interaction of employers and employees. Intra-industry spinoffs raise particular policy concerns regarding the use of intellectual property and nearly every study finds that intra-industry spinoffs perform better than other kinds of spinoffs. Consequently, the focus of the paper will be on intra-industry spinoffs. Accordingly, the descriptor intra-industry will be dropped and spinoffs will be used subsequently to refer to intra-industry spinoffs. Other types of spinoffs will be referred to as other employee startups, and the more generic term of startups will be used to refer to all kinds of new firms, including both spinoffs and other employee startups.

The paper is organized as follows. In the next section, we review the role spinoffs played in one industry, automobiles, in the US and Europe. In the subsequent section, we present various stylized facts that have been accumulating regarding spinoffs. In the following section, we discuss the impetus behind the leading spinoffs in the semiconductor industry, which helps put the stylized facts in perspective. In the penultimate section, we consider the alternative theories of spinoffs and their consistency with the accumulating evidence. In the final section, we discuss open questions and research opportunities and also reflect on public policy from the vantage point of the theories that accord most closely with the evidence.

Spinoffs in the automobile industry

As will be discussed in the next section, spinoffs have been studied in the context of a number of industries, both in the US and around the world. These studies tend to concentrate on high-tech manufacturing industries, where spinoffs have been especially prominent. They painstakingly identify every entrant into an industry and trace their pre-entry histories. Finding data sources that provide the requisite information is challenging, particularly regarding the pre-entry histories of entrants. Perhaps the most studied industry in the world is the automobile industry. Fortunately, hobbyists are interested in old cars and various directories have been compiled that chronicle the origins of automobile producers and the cars they produced over their lifetimes. This has made it possible to study

automobile spinoffs around the world. To provide a sense of industry studies of spinoffs, three studies of automobile spinoffs in the US (Klepper, 2007), Britain (Boschma and Wenting, 2007), and Germany (von Rhein, 2008) are reviewed.

Findings for the US spurred the other two studies, and so we begin by reviewing the role of spinoffs in the evolution of the US automobile industry. The US industry began in 1895, and Table 1 reports the leading US automobile producers every 5 years from 1900 to 1925.¹ Two patterns are evident immediately from Table 1: there was a great deal of turnover early on in the leaders of the industry, and over time the industry became heavily concentrated around Detroit, MI. Both of these developments were fundamentally related to spinoffs.

The initial leaders in 1900 were a mixture of diversifiers and new firms with backgrounds in related industries, especially bicycles, carriages and wagons, and engines. Five years later in 1905 only one of the leaders was left, and not for long. The new leader of the industry was Olds Motor Works. Prior to diversifying into autos, Olds was a leading engine producer located roughly 100 miles from Detroit in Lansing, MI. It set up its automobile operations in both Lansing and Detroit and introduced the first great car in the industry, the Curved-Dash Runabout, in 1901. Two of the other leading firms in 1905, Maxwell-Briscoe and Reo, were spinoffs descended from Olds. Reo was founded by Ransom Olds, the head of Olds Motor Works, after a clash with his major stockholder that led to his departure. Ford Motor Company, which was also located in Detroit, was Henry Ford's second startup and thus qualified as a spinoff as well. By 1910 eight new firms show up in the list of leaders, including five firms in the Detroit area, all of which were spinoffs. Three more ascended to the ranks of leaders in 1915, all in the Detroit area and all spinoffs, and collectively the share of industry output accounted by firms in the Detroit area reached 83%. After 1915 the leaders did not turn over as much and Detroit maintained its dominance of the US industry into the present, largely due to its spinoffs.

Table 2 lists the seven firms in the industry with three or more spinoffs. All but Northern produced a make of automobile that made it onto the annual list of 15–18 leading makes compiled by Bailey (1971), reflecting the greater fertility of the leading automobile producers. These firms also accounted for most of the spinoffs that also produced a leading make, reflecting a tendency for superior firms to spawn superior spinoffs. All seven firms were also located in the Detroit area, with all but Maxwell-Briscoe entering there,² reflecting a higher spinoff rate of firms located in the Detroit area.

Klepper (2007) analyzed statistically the rate at which firms spawned spinoffs and the length of time all firms, including spinoffs, produced automobiles. Breaking up each firm's history into annual intervals and pooling the observations, he estimated an ordered logit of the annual number of firm spinoffs. The analysis confirms the impression from Table 2 that better firms had a higher rate of spinoffs as did firms located in the Detroit area. He estimated a model of the annual hazard of firm exit to analyze the performance of automobile producers. Spinoffs had lower hazards of exit than other startups, and consistent with Table 2 spinoffs descended from the leading

Table 1 Market shares of leading US automobile firms, 1900–1925

	<i>Entry year</i>	<i>Entry location</i>	1900	1905	1910	1915	1920	1925
<i>Early entrants</i>								
Pope	1895	Hartford, CT	36					
Stanley	1896	Watertown, MA		2				
Locomobile	1899	Bridgeport, CT	18					
Knox	1900	Springfield, MA	0.3					
Packard	1900	Warren, OH/Detroit, MI		2	2			1
H.H. Franklin	1900	Syracuse, NY		4				
White Sewing Machine	1901	Cleveland, OH	0.02	4				
Olds/GM	1901	Detroit/Lansing, MI		26		1	2	1
Cadillac/GM	1902	Detroit, MI		16	6	2	1	1
Jeffery/Nash	1902	Kenosha, WI		16			2	3
<i>Later entrants</i>								
Studebaker	1902	South Bend, IN			8	5	3	4
Anderson/Union	1902	Anderson, IN			2			
Ford	1903	Detroit, MI		7	18	56	22	44
Maxwell-Briscoe/Maxwell/Chrysler	1903	Tarrytown, NY/Detroit, MI		3	6	5	2	4
Buick/GM	1903	Flint, MI		3	17	5	6	5
Willys	1903	Terre Haute, IN			9	10	6	6
Reo	1904	Lansing, MI		4	4	2		
Stoddard	1904	Dayton, OH		1				
E.R. Thomas-Detroit/Chrysler	1906	Detroit, MI			4	1		
Brush	1907	Detroit, MI			6			
Oakland/GM	1907	Pontiac, MI			2	1	2	1
Hupp	1909	Detroit, MI			3	1	1	3
Hudson	1909	Detroit, MI			3	1	2	7
Paige-Detroit	1909	Detroit, MI						1
Chevrolet/GM	1911	Flint, MI				1	6	12
Saxon	1913	Detroit, MI				2		
Chandler	1913	Cleveland, OH					2	
Dodge Brothers/Chrysler	1914	Detroit, MI				5	7	5
Dort	1915	Flint, MI					1	
Durant	1921	New York, NY					3	
Detroit-area firms			0	58	65	83	52	85

Table 2 Automobile firms with three or more spinoffs

<i>Firm</i>	<i>Years produced through 1924</i>	<i>Number of spinoffs through 1924</i>	<i>Number of spinoffs through 1924 that produced a leading automobile make</i>	<i>Did firm ever produce a leading automobile make through 1924?</i>
Olds	1901–1908	7	3	Yes
Buick/GM	1903–1924	7	2	Yes
Cadillac	1902–1908	4	3	Yes
Ford	1903–1924	4	2	Yes
Maxwell-Briscoe/Maxwell	1904–1924	4		Yes
Northern	1902–1910	3	1	No
Hupp	1909–1924	3		Yes

producers had even lower hazards. Furthermore, his findings suggested that the agglomeration of the industry around Detroit was largely driven by the spinoff process.

Spurred by Klepper's (2007) findings, Boschma and Wenting (2007) estimated a hazard of exit model for British

automobile producers in which diversifiers from related industries, spinoffs, and other startups were distinguished. Similar to Klepper (2007), they found that spinoffs survived longer than other startups, which was especially true for spinoffs of parents that themselves were superior

performers (measured in terms of the number of years they produced automobiles). Their findings also suggested that similar to Detroit, spinoffs were a key contributor to the agglomeration of the British industry around Birmingham–Coventry.

Cantner *et al.* (2006) also estimated a model of the hazard of firm exit to investigate the effect of pre-entry experience on the performance of German automobile producers. Von Rhein (2008) further broke down experienced entrants according to whether they were spinoffs and the number of years their parents produced automobiles.³ Similar to the US and Britain, her findings indicate that spinoffs survived longer than other startups, with the longevity of spinoffs and their parents positively related.

Thus, judging from the automobile industry in various countries, spinoffs performed distinctively, especially those descended from superior firms, and played a key role in the formation of regional clusters. We now consider the extent to which these patterns hold in other industries as well.

General patterns

As reflected in Klepper and Thompson (2009), in addition to automobiles, the formation and performance of spinoffs has been studied in various high-tech manufacturing industries during their early years of evolution. With the exception of the automobile and tire industries (Buenstorf and Klepper, 2009a, 2009b), the industries studied are modern, including semiconductors (Brittain and Freeman, 1986; Klepper *et al.*, 2009; Klepper, 2009a, b), biotechnology (Mitton, 1990; Stuart and Sorenson, 2003), disk drives (Agarwal *et al.*, 2004; Franco and Filson, 2006; McKendrick *et al.*, 2009), lasers (Sleeper, 1998; Klepper and Sleeper, 2005; Sherer, 2006; Buenstorf, 2007), and medical devices (Chatterji, 2008). Spinoffs have also been studied in Silicon Valley law firms (Phillips, 2002), Australian and New Zealand wine producers (Roberts *et al.*, 2007), Israeli Information Communication Technology (ICT) firms (Ellis *et al.*, 2008), and the world fashion design industry (Wenting, 2008). The main issues examined in the studies include the rate at which firms spawn spinoffs, the performance of spinoffs, the effects of spinoffs on their parents, and the role of spinoffs in the formation and growth of clusters. Also touched on is the extent to which knowledge is transferred from ‘parents’ to spinoffs.

Various studies have been conducted of startups generally that are also informative about spinoffs. Denmark has a data base matching employers and employees that has been used in various studies to analyze employee startups of all kinds (Eriksson and Kuhn, 2006; Dahl and Reichstein, 2007; Sørensen, 2007; Sørensen and Phillips, 2008). Wagner (2004) uses data from a survey to study nascent entrepreneurs (those intending to start a firm) in Germany. Gompers *et al.* (2005) use data concerning startups funded by venture capitalists (VCs) to study employee startups of publicly traded firms in the US that received VC support. Elfenbein *et al.* (2008) use data from biannual surveys of scientists and engineers trained in the US to study the rate at which they leave their employers to found their own firms and the performance of their startups. Last, Burton *et al.* (2002) use survey data on 173 high-tech startups in

Silicon Valley to study the strategy of these startups and whether they were able initially to raise external finance.

It is important to recognize that studies use different definitions and procedures to identify spinoffs and their parents, in large part due to differences in the available data. If a firm is incorporated and has employees, then identifying the founder or founders requires information about who organized the firm. Announcements about startups and historical accounts of the formation of firms generally identify a main founder or organizer of a firm. Industry studies typically have access to this kind of information for each entrant, which they use to identify spinoffs, with the prior employer of the main founder designated as the spinoff’s parent. In some studies, if multiple individuals from different firms are involved in founding a spinoff, then all of the firms are designated as parents of the spinoff.

In contrast to industry studies, studies of startups in large populations generally do not have detailed information on the founders of each firm. Consequently, they have to rely on different criteria to define founders and parents of startups and thus to identify spinoffs. For example, Eriksson and Kuhn (2006) define employee startups as new firms with between 2 and 10 initial employees in which at least 50% of the employees previously worked at the same firm (designated the parent) and constituted less than 50% of the workforce of that firm. Alternatively, for incorporated firms Sørensen (2007) defines as founders all the initial employees if there were three or less and all those with a title of director or top manager if there were more than three employees, with the prior employers of all the designated founders defined as the firm’s parents.⁴ Similarly, Gompers *et al.* (2005) include all the initial executives of their VC-backed startups as founders and all their prior employers as parents of the firm.

Studies not only differ regarding how they define spinoffs (and other employee startups) and parents, but they also differ regarding how they measure firm performance. Despite all these differences, though, Klepper and Thompson (2009) note that a number of common findings have been emerging from the various studies. They synthesize five ‘stylized facts,’ where a stylized fact is a pattern that holds in most if not all studies that examine the relevant issues. Below, these five stylized facts are elaborated and four additional stylized facts are presented.

The first three stylized facts pertain to the rate at which firms spawn spinoffs:

- (1) In autos, tires, semiconductors, disk drives, and lasers, better-performing firms, measured by longevity, peak market share, early entry, product quality, and/or product scope, have higher spinoff rates.^{5,6} Ellis *et al.* (2008) also found that Israeli ICT firms forged in more competitive environments have higher spinoff rates and Gompers *et al.* (2005) found that firms that were less diversified and received VC funding themselves spawned more VC-backed startups. In terms of the number of spinoffs per employee, Wagner (2004), Gompers *et al.* (2005), Eriksson and Kuhn (2006), Sørensen (2007), and Elfenbein *et al.* (2008) found that larger firms spawn less startups per employee than smaller firms, which seemed to hold as well for spinoffs in automobiles and semiconductors (Klepper, 2009a).



- (2) In autos, biotechnology, lasers, and semiconductors, firms acquired by non-industry incumbents have higher intra-industry spinoff rates around the time of their acquisition, while in autos and lasers (but not biotechnology or semiconductors) firms acquired by industry incumbents also have comparably higher intra-industry spinoff rates around the time of their acquisition. Relatedly, in semiconductors firms that hired a CEO from outside the company have higher intra-industry spinoff rates, which accords with findings from Eriksson and Kuhn (2006) that employee startups of all kinds, including spinoffs, are more likely in firms whose CEO has recently changed.
- (3) In autos, lasers, semiconductors, and law firms (but not disk drives or tires), the rate at which firms spawn intra-industry spinoffs tends to rise initially with age to around age 15 and then decline.⁷

The next three stylized facts pertain to the performance of spinoffs:

- (4) In autos, disk drives, lasers, medical devices, tires, fashion design, and wine, the performance of spinoffs in terms of longevity, peak market share, scope, years to first VC funding, and pre-money valuation is superior to other startups and is comparable if not superior to diversifiers from related industries. Eriksson and Kuhn (2006) and Dahl and Reichstein (2007) found similar patterns among all employee startups in Denmark.
- (5) In autos, tires, semiconductors, disk drives, law firms, and fashion design (but not lasers), the better the performance of parent firms based on their longevity, market share, and/or quality of technology then the better the performance of their spinoffs.⁸ Consistent with this, Burton *et al.* (2002) found that founders of Silicon Valley startups were more likely initially to raise external finance the greater the number of firms spawned by their parents, which they interpreted as a measure of the parent's prominence/performance. However, in the broad populations of startups analyzed by Sørensen and Phillips (2008) and Efenbein *et al.* (2008), entrepreneurial income of startup founders was lower the greater the size of the founder's prior employer.
- (6) In lasers and semiconductors, initially spinoffs tended to produce types of products that were a subset of those produced by their parents.

The next stylized fact describes the effect of spinoffs on their parents:

- (7) In law firms and disk drives, firms that spawned spinoffs subsequently experienced an initial rise in their hazard of exit (law firms) and decline in their technological standing relative to the frontier (disk drives). In both industries the adverse consequences dissipated over time and in the case of disk drives even reversed, with both the initial decline and subsequent improvement in the parent's technology relative to the frontier greater the better the technology of its spinoff.

The last two stylized facts pertain to aspects about spinoffs and clusters:

- (8) In autos, semiconductors, and lasers, firms located in industry clusters spawned spinoffs at a higher rate.

Gompers *et al.* (2005) also found that publicly traded firms based in Silicon Valley and Massachusetts, the two main centers of high-tech firms in the US, spawned more VC-backed companies.

- (9) In autos, tires, and semiconductors, spinoffs disproportionately entered in the centers of the industries (Detroit, Akron, and Silicon Valley respectively) and the joint market share of spinoffs and their parents exceeded the pre-spinoff market share of their parents, suggesting that spinoffs were more than a zero-sum phenomenon.

Case studies of the formation of spinoffs

In addition to the stylized facts, case studies of the circumstances in parent firms leading up to spinoffs can also provide insights into the forces governing their formation and performance. Case studies are difficult to assemble as they require detailed information regarding the inner workings of firms. Not surprisingly, they generally are available only for leading firms. While this means they are not necessarily representative of all spinoffs, it is the leading spinoffs that have the greatest social impact and thus are of the most interest.

Klepper and Thompson (2009) review case studies of the leading automobile spinoffs in Detroit, the leading semiconductor spinoffs in Silicon Valley, and representative laser spinoffs. The themes struck by the cases were similar across the three industries and are well illustrated by the experiences of the semiconductor spinoffs. They are listed in Table 3 along with information about their entry year, parent, impetus, and source of finance.

The spinoffs entered in the period 1957–1986, which covers the early years of the industry. It is useful to review the early challenges faced by semiconductor firms to appreciate the circumstances motivating the (leading) spinoffs. Semiconductor firms continually had to make difficult choices about which technologies to develop. Initially it was unclear whether germanium or silicon would be the best material for semiconductor devices. When integrated circuits (ICs) were developed, they were initially inferior to circuits composed of discrete devices and potentially infringed upon the markets of semiconductor customers. Metal Oxide Semiconductor (MOS) devices were slower than early, bipolar ICs and were unstable and difficult to make. Eventually, though, manufacturing problems were overcome and MOS devices proved to be superior for many applications because they enabled many more transistors to be packed onto chips. Similarly, Complementary Metal Oxide Semiconductor (CMOS) devices were extremely slow, but their low power needs ultimately facilitated even denser chips. Application-specific ICs (ASICs) initially were not economical but MOS technology eventually changed that. Linear, or analog, devices, which are used for amplification and other non-digital applications, have always posed distinct technical and market challenges.

Initial technical and market uncertainties over these technologies led to conflicts within firms and the departure of top employees to found their own firms. The first semiconductor producer in Silicon Valley was Shockley Semiconductor Laboratory, which was founded by William Shockley, who shared the Nobel Prize for the invention

Table 3 Origins of leading spinoffs of Silicon Valley producers

<i>Spinoff</i>	<i>Year</i>	<i>Parent</i>	<i>Reasons</i>	<i>Finance</i>
Fairchild	1957	Shockley Semiconductor Laboratories	Strategic disagreement (silicon transistors), management conflict	Fairchild Camera and Instrument
Amelco	1961	Fairchild	Strategic disagreement (ICs)	Teledyne
Signetics	1961	Fairchild	Strategic disagreement (ICs), management conflict	Investment banks
Electronic Arrays	1967	GME	Management conflict after acquisition by non-semiconductor firm	N.A.
Intersil	1967	Union Carbide	Compensation practices (stock options), management conflict with non-semiconductor parent	SSIH and Olivetti
National	1967	Fairchild	Compensation practices (stock options), management conflict with non-semiconductor parent	National Semiconductor
Intel	1968	Fairchild	Management conflict, technical frustration (MOS)	Venture capital
AMD	1969	Fairchild	Management conflict after CEO hired from outside firm	Minimal capital (\$100,000)
Zilog	1974	Intel	Personal tensions	Exxon
VLSI	1979	Synertek	Management conflict after acquisition by non-semiconductor firm	Venture capital
LSI Logic	1980	Fairchild	Management conflict after acquisition by non-semiconductor firm	Venture capital
Linear	1981	National	Strategic disagreement (linear circuits)	Venture capital
Cypress	1982	AMD	Strategic disagreement (CMOS)	Venture capital

N.A. – not available.

of the transistor. Shockley was a keen judge of talent but a dysfunctional manager. After he abandoned his original goal of producing silicon transistors in favor of a new device he invented that proved difficult to manufacture, eight of his talented employees left to form Fairchild Semiconductor after failing to convince Shockley to revert to his original plan (Lécuyer, 2006: 131–139). Fairchild pioneered the development and production of ICs in the early 1960s, but did not pursue ICs aggressively at first due to their initial inferior performance and fear of infringing on the markets of their customers. Some of the founders and top researchers that were involved in the development of ICs thought they had greater potential than Fairchild

and left to form two spinoffs, Amelco, and Signetics (Lécuyer, 2006: 213–218; Sporck, 2001: 70). Fairchild maintained separate facilities for R&D and manufacturing, which caused strains when manufacturing proved unable to produce devices developed by the R&D group (Bassett, 2002: 172–173). As a consequence, Fairchild was unable to develop MOS products despite being the industry leader in MOS research, which led Robert Noyce and Gordon Moore, the leaders of Fairchild Semiconductor, to leave to found Intel. National Semiconductor produced linear devices but the head of its linear division felt National treated linear devices as a means of getting into other businesses rather than an attractive business of its own and left to form



Linear Technology (Wilson, 2004). Cypress was formed by T.J. Rodgers, a top manager at AMD, after AMD and other established firms were not interested in developing CMOS devices (Gilder, 1989: 143).

Another significant challenge that semiconductor firms faced early on was how to reward innovators and structure their organizations to harness scientific and technical advances for commercial benefit (Moore and Davis, 2004). At first it was unclear to some firms how important stock options would prove to be in rewarding top individuals. It was also challenging to figure out how to get manufacturing and marketing personnel to cooperate with R&D in transferring new devices developed by R&D into large-scale production. The result was managerial conflicts that led to exodus of top individuals to form their own firms. Some firms were either financed or acquired by non-semiconductor firms, which seemed to fan managerial tensions and the formation of spinoffs. Bringing in a new CEO from outside the firm seemed to operate similarly.

A number of the leading spinoffs resulted from these kind of managerial tensions. For example, Electronic Arrays was formed after its parent, General Microelectronics, was acquired by Philco, which canceled stock options and moved the company from Silicon Valley to Philadelphia (Lécuyer, 2006: 263). It was not long before Philco exited the industry. Similarly, Intersil was formed after its parent, Union Carbide, refused to give its leader stock options (Lécuyer, 2006: 263–264). Four years later Union Carbide exited the industry. Fairchild Semiconductor was controlled by a Fairchild Camera and Instrument, a Long Island defense contractor that had financed its entry into semiconductors. When its parent refused to grant more than meager stock options to its top employees, Fairchild's production chief left to found National (Sporck, 2001: 207–214; Lécuyer, 2006: 259–261).

Other spinoffs occurred after new management was brought in from outside the firm. For example, Les Hogan was hired away from Motorola, a leading semiconductor producer, to manage Fairchild Camera and Instrument after it floundered. Soon after, Fairchild Semiconductor's head of marketing clashed with Hogan and left to found AMD (Sporck, 2001: 152–157). Fairchild continued to flounder and was later acquired by Schlumberger, a French firm without experience in semiconductors. When Schlumberger brought in its own management, Fairchild's CEO, Lester Corrigan, left to found LSI Logic to produce ASICs, a market Fairchild had pursued earlier but then abandoned. Fairchild subsequently declined and was sold to National (Walker, 1992: 54–57). Similarly, after Synertek was acquired by Honeywell, a computer manufacturer that also brought in its own management to run the company (Walker, 1992: 184–186, 195–197), a co-founder of Synertek left to found VLSI to produce ASICs. Seven years later Honeywell sold Synertek and exited the industry.

The common theme underlying the spinoffs was some kind of disagreement among top managers concerning either the strategic direction of the firm or its management structure. Those that left to found spinoffs often continued initiatives, such as the initial development of ICs in the case of Amelco and Signetics or the production of linear devices in the case of Linear Technologies, that had been

started by their parents but not pursued aggressively enough to suit their tastes. In other cases, they left after new management was brought in and moved the company away from its historical strategic focus and/or organizational structure. Klepper and Thompson (2009) reported similar motivations for spinoffs in automobiles and lasers and Lindholm-Dahlstrand (2001) found that the most important reason for spinoffs in Swedish technology-based firms was the refusal of the parent firm to commercialize internally generated ideas.

The semiconductor spinoffs in Table 3 include many of the most successful firms in the industry, such as Intel, National, and AMD. They advanced the development of some of the most important innovations in the industry, including silicon transistors, ICs, MOS devices, CMOS devices, and AISECs, after their parents were unwilling, unable, or slow to pursue them. Indeed, many observers have trumpeted the role of the leading spinoffs in the success of the US semiconductor industry and the emergence and growth of Silicon Valley (Saxenian, 1994: 112; Sporck, 2001: 268–271; Moore and Davis, 2004). As such, it is instructive to consider how they were financed. Table 3 indicates the spinoffs were financed predominantly by downstream firms and VCs. Many of the VCs were themselves past employees of successful semiconductor firms and thus spinoffs of their own. Both downstream firms and VCs had their own knowledge about the industry that they drew upon to make financing decisions regarding spinoffs. Thus, even when parents faltered or were slow to pursue attractive opportunities, spinoffs were able to find support and compensate for the deficiencies of their parents.

Theories meet the evidence

In this section we first review the principal theories of why spinoffs occur and the extent to which they can explain the accumulating evidence.

A useful starting point is Pakes & Nitzan (1983). They consider the case where an entrepreneur could benefit by hiring a scientist to work on a project that one period later yields an uncertain return. In the second period, the scientist could work for the entrepreneur, another firm, or set up his own firm and earn part of the return, α , that the entrepreneur could earn if he undertook the project alone. The scientist's profits would be a fraction of the entrepreneur's profits assuming the scientist must incur additional costs to undertake the project than the entrepreneur (i.e., the entrepreneur has complementary assets that are not used to capacity) and/or the profits of duopolists are less than those of a monopolist. They show that the entrepreneur could offer the scientist a stock option equal to α times the entrepreneur's return plus a suitable fixed payment in the first period (to compensate for the scientist's alternative wage) to induce the scientist to work for the firm in period 2. So in this framework, the only way a spinoff occurs is if the scientist could make more from the project than the entrepreneur alone. They consider two ways this could occur. First, if the project was far afield of the entrepreneur's main line of business, it might be less costly for a specialist to manage the project than the entrepreneur.

Second, if a third party might also undertake the project, the scientist could earn back some of the profits the entrepreneur would lose to the third party, benefiting both the entrepreneur and scientist. Both conditions are more likely to be satisfied the less significant the entrepreneur's complementary assets for the project.

One implication of this theory is that scientists would be more likely to undertake projects further afield from the entrepreneur's specialty, which would both be more difficult to manage and involve less additional costs for the scientist. Such a theory is better suited to explain employee startups outside of the parent's (entrepreneur's) main business than spinoffs. It also implies that spinoffs would initially produce different types of products than their parents, which is not consistent with stylized fact 6 or the case studies. The theory also does not provide much help in explaining the other stylized facts.

Wiggins (1995) introduces a twist on the Pakes–Nitzan framework. He considers the case where in period 2 the parent could understate its profits and not pay the employee his promised profits. Clearly, this is more tempting the larger the period 2 profits. In Wiggins's framework, the period 2 profits are larger the greater the effort the scientist has to put forth in period 1 and the riskier the project. Wiggins associates these conditions with projects involving (potential) path breaking innovations and new lines of business. Anton and Yao (1995) reach a similar conclusion in a related model in which employees can profit by behaving opportunistically and misrepresenting their role in the development of a profitable idea (Klepper, 2001). Thus, both theories suggest spinoffs will initially produce different types of products than their parents, which as already noted does not accord well with the evidence. Similar to Pakes and Nitzan, neither of the theories provides any insight into the other stylized facts.

Cassiman and Ueda (2006), Hellmann (2007), and Gambardella and Panico (2008) also develop theories that predict that spinoffs will pursue ideas less related to their parents' main activities. In Cassiman and Ueda's model, firms have a limited capacity to develop new ideas. They undertake the most profitable to them, which are the ones that fit closest to their main activities, and leave the rest for their employees to develop in spinoffs. While this does not accord well with stylized fact 6, Cassiman and Ueda can explain a number of the other stylized facts and aspects of the case studies. They assume that better firms generate more and better ideas, which implies that better firms pass up more good ideas than other firms and hence spawn more and better-performing spinoffs, consistent with stylized facts 1 and 5. Furthermore, if other kinds of startups are not based on internally generated ideas within a prior employer or at least ones not as good as those passed up by incumbent firms in an industry, spinoffs will outperform other kinds of startups, consistent with stylized fact 4.

Gambardella and Panico's (2008) model has similar implications. It assumes that asymmetric information about project inputs and outputs makes performance-based contracts infeasible. When the firm has few complementary resources related to ideas generated internally, it is better off delegating to the best workers the decision about how the ideas should be developed. Consequently, better

workers develop ideas in spinoffs that are less related to the firm's core activities. Spinoffs thus have above average founders and outperform other startups, consistent with stylized fact 4. Furthermore, firms with better employees, which are better performers, are more likely to spawn spinoffs, consistent with stylized fact 1.

In Hellmann (2007), employees control the kinds of ideas they generate in the firm through how they allocate their efforts. Some may want to pursue ideas less related to their employer's main line of activity, which are less profitable to the employer. To discourage this, employers pledge never to develop such ideas internally, which can be a credible threat if the worker must incur additional costs relative to the employer to develop the ideas. Some employees may still prefer to indulge their own preferences, which leads to employees leaving to form spinoffs. Once again, though, this implies that spinoffs will initially produce different products than their parents, which is not consistent with stylized fact 6 or the case studies. The theory does not address the other stylized facts.

A different approach is taken by Franco and Filson (2006) to explain spinoffs. Historically, many individuals worked as apprentices, at reduced wages, in order to receive sufficient training to start their own firms. In Franco and Filson's modernization of this theory, employees accept lower wages to work at better firms because they have the prospect of learning and capitalizing on the know-how of their parents by starting their own firms. Firms differ in the quality of their know-how, which determines their performance. If employees learn their employer's know-how and it is of sufficient quality, they start their own firm. Spinoffs have the same know-how as their parents and hence do the same things as their parents, which is consistent with spinoffs initially producing similar products to their parents (stylized fact 6). Furthermore, the better the employer's know-how then the more likely its employees will leave to found spinoffs, consistent with stylized fact 1. Similarly, the better the firm's know-how then the better its performance and that of its spinoffs, consistent with stylized fact 5. Finally, assuming non-spinoff startups do not have distinctive knowledge to exploit from their employers, spinoffs will perform better than other startups, consistent with stylized fact 4.

Franco and Filson apply their theory to the disk drive industry. Although they find that firms with better technology spawned more spinoffs, the spinoffs of such firms actually performed worse even though overall firms that survived longer spawned better spinoffs⁸. This questions the conduit featured in their theory as the impetus for spinoffs. Furthermore, a number of spinoffs pioneered new, smaller disk drives their parents declined to pursue (Christensen, 1993), which is not consistent with spinoffs doing the same things as their parents.

Klepper and Sleeper (2005) also develop a theory of spinoffs based on learning. In their model, firms and their top employees learn about how to develop variants of their initial products. Employees have greater incentives to pursue the variants than their employer whenever the variants cannibalize the employer's profits. If employers were sure they had employees ready to found their own firms, they would preempt them, but otherwise it is best to gamble that spinoffs won't occur. If they do, they will



develop variant's of their parent's products, consistent with stylized fact 6. If better firms are more likely to develop attractive variants of their initial products, then they will be more likely to spawn spinoffs, consistent with stylized fact 1.

Recently, a new set of spinoff theories have been developed that are based on the idea that employers have a limited ability to recognize the best ideas and/or the best employees with the best ideas, leading them to offer similar terms to all employees to develop their ideas internally (Cabral and Wang, 2008; Chatterjee and Rossi-Hansberg, 2008; Klepper and Thompson, 2009). The result is that better employees are more likely to develop their ideas in their own firms. If the performance of employers is based on the quality of their employees, as assumed in Cabral and Wang (2008) and Klepper and Thompson (2009), then better firms spawn more and better spinoffs and on average spinoffs outperform other startups, consistent with stylized facts 1, 4, and 5. The departure of these better employees also lowers the performance of their parents, consistent with the first part of stylized fact 7 concerning the initial impact of spinoffs on their parents.

In Klepper and Thompson (2009), employees exchange their ideas about what their employer should do, and the ideas pursued by the firm are a weighted average of the ideas proposed by their employees, where their weights reflect their influence in the firm. Therefore, the more influential an employee then the closer the firm's actions to what it thinks it should do. Firms are unable to recognize the best ideas and thus underweight them in their choice of what to do, leading to disagreements between the firm and its best employees about what it should do. If a disagreement is sufficiently large, then the employee leaves to found his own firm. At first, all employees have the same priors about what the firm should do, hence there are no disagreements. Then they receive different signals about what the firm should do, and disagreements materialize. But with enough signals, the views of all employees converge on the best ideas, and eventually all agree on the best course of action and no disagreements occur. This can explain stylized fact 4 in which the probability of spinoffs first rises and then falls as firms age. It is assumed that when a firm is acquired or a new CEO is brought in from the outside, then every incumbent employee's influence on decision making is reduced, which raises the probability of a spinoff, consistent with stylized fact 4. Last, employees learn from each other, which influences the founders of spinoffs in their choice of what ideas to pursue, causing them to overlap with their parents in terms of their initial activities, consistent with stylized fact 6. At the same time, spinoffs differentiate themselves from their parents by giving greater weight to the ideas of their founders. These ideas are underweighted by their parents, which can explain why spinoffs are not a zero-sum phenomenon, consistent with stylized fact 9.

In summary, the early theories relied on a contracting perspective to explain spinoffs. In many ways, this is a natural place to begin. If incumbent firms have underutilized complementary assets that make it less expensive for ideas to be developed in incumbents than spinoffs, then some kind of contracting problem would seem to be required to get spinoffs to occur. But theories in which the nature of contracting problems dictates the kind of ideas developed

in spinoffs have a hard time explaining the accumulating evidence, particularly the tendency for spinoffs to outperform other startups and the positive relationship between parent and spinoff performance. A natural way to explain these patterns is to allow for employees learning about their employer's distinctive knowledge. A less obvious way to do it is to posit a limit on a firm's capacity to develop new ideas. Both approaches can account for a number of the patterns regarding the exemplary performance of spinoffs and the influence of firm quality on the rate at which firms spawn spinoffs and the performance of their spinoffs. But these theories tend to predict either no overlap or complete overlap between the activities of parents and spinoffs and do not map well into the case study evidence suggesting that in many instances spinoffs result from disagreements over ideas that originated in their parent firms.

These patterns and some of the other stylized facts can be explained by the latest round of theories that feature the difficulty of firms assessing the quality of their employees and their ideas, leading employees with better ideas to develop them in spinoffs. None of the theories, however, can address why firms located in clusters spawn spinoffs at higher rates (stylized fact 8), nor can they explain why parents ultimately might improve their performance after spawning spinoffs (stylized fact 7). They also cannot explain why aspects of firms such as the competitive environment in which they were forged or whether they received VC funding should affect the rate at which they spawn spinoffs (stylized fact 1), nor why larger firms should spawn less spinoffs per employee (stylized fact 1).

Discussion

Numerous questions remain about spinoffs, each of which defines a research opportunity. The main unanswered questions are discussed below. While questions abound, policy has to be made, and following the discussion of open questions we use the state of the literature to reflect on key policy issues.

A major question concerns whether employees that found spinoffs exploit knowledge they learned from their parents, and if so, what kinds of knowledge. The positive correlation between the performance of spinoffs and parents found in many of the industry studies is suggestive of some kind of learning mechanism, with better firms having more knowledge to be learned by their employees. Alternatively, might these patterns simply reflect that better firms have better employees, and better employees are more likely to start their own firms and to perform better?.

A fundamental question concerns what firms need to know at the outset to be able to compete effectively in an industry and to what extent can employees learn this as a byproduct of their employment? Moore and Davis (2004) provide a rare insight into these issues in the context of the semiconductor industry during its early evolution, reflecting the experience of Gordon Moore as the co-founder of two of the most important firms in the history of the industry, Fairchild Semiconductor and Intel. The title of his paper, 'Learning the Silicon Valley Way,' reflects the importance that is attached to employees learning how to structure organizations and incentives to compete in the early semiconductor industry. This is an intriguing study.

We need more looks inside organizations like this to guide our theorizing about spinoffs.

Learning conjures up other important issues as well. Do firms have different entrepreneurial cultures that not only are imparted to employees but also influence their inclination to found spinoffs? Higgins (2005) develops this idea to explain the differential rate at which up-and-coming employees from health care companies were tapped for leadership positions in new biotechnology companies. This is not the same as employees founding companies, but Ellis *et al.* (2008) do apply the same idea to explain differential rates of spinoffs from Israeli ICT companies. Related to the culture question is whether firm size fundamentally influences a firm's entrepreneurial culture. Could that explain why large-scale, non-industry studies find that employees of larger firms are less likely than those of smaller firms to found startups (Wagner, 2004; Sørensen, 2007; Elfenbein *et al.*, 2008)? Some studies also find that the performance of startups is better for those founded by employees of smaller firms (Sørensen, 2007; Elfenbein *et al.*, 2008), which is opposite to the industry studies, which find that the largest firms spawn the best-performing spinoffs. Is that because the industry studies generally focus on young, growing industries, so the leading firms have not yet fallen prey to the bureaucratic inertia that stifles initiative? Clearly, much remains to be learned here.

The flip side of learning is spillovers. If founders of spinoffs benefit from knowledge they learned while working at their parents, a positive externality arises. At the same time, if the parents are hurt in the process, an offsetting negative externality occurs. The potential harm experienced by parents is one of the main rationales used to defend practices such as requiring employees to sign non-compete covenants, restricting their ability to found their own firms in the same industry as their parent (in states that allow such covenants to be enforced – more on this below). Exactly what these negative externalities entail is a critical question. McKendrick *et al.* (2009) directly engage one dimension of this question and Phillips (2002) another, but much remains to be explored here.

Another question related to learning is the extent to which it can be anticipated by employees and influences where they choose to work and the wage and other conditions they will accept. Franco and Filson (2006) assume that potential spinoff founders accept lower wages to work for better firms, which forms a critical component of their welfare assessment of spinoffs. Møen (2005) presents some intriguing evidence supporting such a link, but again much more remains to be learned about this important question. Indeed, key to forming policy related to spinoffs, as discussed below, is the extent to which spinoffs harm their parents and possibly create a disincentive for their parents to engage in privately and socially productive activity like R&D. We dearly need to know more about this. States differ in their law on the enforcement of non-compete covenants (Gilson, 1999), which might provide a natural experiment to explore these issues. This natural variation has been used to analyze how much laws influence the formation of startups (Stuart and Sorenson, 2003; Samila and Sorenson, 2009) but not the harm incurred by parents.

Some of the other stylized facts also seem to define research opportunities. Virtually every study that looks

at the influence of acquisitions, changes in the CEO, initial public offerings (IPOs) and other forms of leadership changes finds they have a marked influence on the formation of spinoffs. Yet only one theory attempts to explain this. A number of the leading semiconductor spinoffs that were reviewed seem to have been precipitated by acquisitions or changes in the CEO, but case studies are few and far between. This seems like a ripe opportunity for further research. What changes inside firms regarding decision making, the choice of projects, the ideas to develop, etc. that raises the probability of employees leaving to start their own firms? Should this inform theorizing about spinoffs?

A kind of flip side to leadership changes influencing spinoffs is the phenomenon of corporate-sponsored spinoffs. From the outset, voluntary spinoffs were excluded from the analysis under the presumption that they have a different motivation from involuntary spinoffs. Unless otherwise indicated, it was assumed that empirical studies were largely about involuntary spinoffs, and in some cases explicit steps were taken to insure this (for example, see Eriksson and Kuhn, 2006). But surely there is a continuum operating in which parents have different degrees of involvement in their spinoffs. We know little about this except for a few studies of corporate sponsored spinoffs, the most intriguing of which are about Xerox's spinoffs (Chesbrough and Rosenbloom, 2002; Chesbrough, 2003). It would seem that much could be learned by studying which ideas firms chose not to pursue internally but nonetheless thought sufficiently worthwhile to pursue in a separately constituted venture. Furthermore, Chesbrough's findings regarding how Xerox's involvement in its corporate spinoffs adversely influenced their performance is provocative about the internal functioning of firms regarding the development of ideas. Among the spinoff theories, only Hellmann (2007) addresses corporate as well as involuntary spinoffs while Gambardella and Panico (2008) engage the issues in motivating their theory of involuntary spinoffs. Surely much remains to be learned here that could inform theorizing.

Another set of intriguing patterns concerns the interplay between spinoffs and agglomerations. In a number of cases of extreme agglomerations, spinoffs show up prominently, as the example of the US automobile illustrated. Furthermore, the spinoff rate appears to be markedly higher for firms located in agglomerations. What do we learn from this about the motivation for spinoffs? Could this be due to a kind of demonstration effect, which Nanda and Sørensen (2008) frame as a peer effect, which presumably has a geographic dimension? Are there more opportunities to observe spinoffs in regions where there are more firms in an industry, which in turn encourages individuals to found spinoffs? If so, how does this operate – does it reduce uncertainty and counteract inherent risk aversion when it comes to starting firms? Surely a top employee venturing out on his or her own at middle age with a family to start a firm must be quite a risky endeavor. If agglomerations work to overcome risk aversion, society will benefit. But many other things could be going on, some of which may also involve externalities, such as agglomerated areas developing a better infrastructure to support the organization and finance of new firms. Again, much remains to be



learned here that presumably could help inform theorizing about spinoffs.

Like so many areas in economics and management, social policy regarding spinoffs needs to be made even in the face of great ignorance. Passions abound on both sides of the issues. Some scholars see spinoffs as parasites feeding off the innovative efforts of their parents, aided by 'vulture' capitalists that help them get started. Those who share this view fear the effects spinoffs may have on the ability and incentives of incumbent firms to innovate and thus to be able to compete with the likes of Japanese firms blessed with lifetime employment (Florida and Kenney, 1990: 79–97). Others, however, see spinoffs as the font of innovation, compensating for the deficiencies of their parents. Those who take this view fear that practices such as employee non-compete covenants could be used to restrict spinoffs and kill the geese that laid the golden eggs.

How do the theories and accumulating evidence regarding spinoffs inform this debate? Two policies have been singled out for examination: firms being allowed to enforce non-compete covenants that employees are asked to sign when they are hired⁹ and the use of the doctrine of inevitable disclosure of trade secrets¹⁰ to prevent spinoffs from being formed (cf. Gilson, 1999; Hyde, 2003). In the US, the former is a state law, and a minority of states, including most prominently California, have outlawed the enforcement of non-compete covenants. The latter involves federal law, although district courts differ in how they interpret the law.

While no definitive conclusions can be reached, the accumulating evidence and the emerging interpretation of the evidence suggests that spinoffs often compensate for deficiencies of their parents, and as such are socially beneficial.¹¹ It appears that without spinoffs, society would either not pursue or take longer to pursue worthwhile ideas that incumbent firms are reluctant to pursue even though they fall within their main areas. This is particularly evident from the case studies of the spinoffs in the semiconductor industry. Spinoffs expand the range of ideas pursued, which appears to have played an important role in the formation and growth of innovative clusters like Silicon Valley that certainly seem socially beneficial. Indeed, if startups generally and spinoffs in particular cannot fully appropriate the value of the ideas they pursue, then it is easy to see how spinoffs could be socially beneficial as long as they do not do exactly the same things as their parents.

The only question is whether spinoffs might undermine the incentives of parents to create intellectual property and whether this harm might outweigh their putative benefits. This is a hard question to answer since currently restrictions on spinoffs limit this harm. On a micro-level, stylized fact 7 suggests that while spinoffs might harm their parents in the short run, they might actually stimulate them to improve in the long run. On a more macro-level, if Silicon Valley and California are representative it would appear that not allowing employee non-compete covenants to be used to suppress spinoffs is socially beneficial. Although it is difficult to extrapolate from a single case, a few studies have attempted to analyze whether non-compete covenants in fact restrict employee mobility and the formation of startups, and the verdict appears to be yes (Stuart and Sorenson, 2003; Marx *et al.*, 2009; Samila and

Sorenson, 2009). If states wanted to be proactive, they could take steps to encourage spinoffs. This could include sponsoring forums that facilitate the formation of teams of founders or even providing financing to spinoffs, although the ability of governments to pick firms deserving of support certainly has a checkered past.

There is surely much to be learned about spinoffs, but we have come a long way in the past 10 years in generating discriminating evidence regarding spinoffs. No doubt the current theories will be refined if not overturned as more evidence accumulates about the origin of entrants and how their backgrounds affect their performance. As we back away from the fiction that all firms are created equal, hopefully we will develop a deeper appreciation of how firms come to be and the kinds of public policies that can promote the formation of better firms. Adam Smith's invisible hand may well allocate resources optimally (under certain conditions) given the set of firms in an economy, but it surely cannot guarantee the formation of the kind of firms that best promote society's welfare. Only as we come to understand the origins of entrants, and particularly spinoffs given their exemplary performance, are we likely to understand how to best structure policy governing the startup of new firms.

Acknowledgement

I thank Guido Buenstorf, Serguey Braguinsky, and two anonymous referees for helpful comments.

Notes

- 1 This table is taken from Klepper (2009a), which provides sources.
- 2 Maxwell-Briscoe was a spinoff of a Detroit firm but initially located in New York and later moved to Detroit where it spawned all of its spinoffs.
- 3 The location of German automobile producers was not examined.
- 4 Dahl and Reichstein (2007) use a similar approach to identify the founders of startups.
- 5 In the US laser industry, however, Klepper and Sleeper (2005) found that the rate at which firms spawned spinoffs initially producing a particular laser was unrelated to how many years the firms produced the laser.
- 6 For the disk drive industry, Agarwal *et al.* (2004) and Franco and Filson (2006) were also able to control for the effect of firm size on spawning rates, which did not change their findings. This suggests that better firms did not spawn more spinoffs merely because they were larger.
- 7 For biotechnology firms in San Diego, Mitton (1990) also found that the spawning rate rose initially through age 10, which was the oldest age among the firms in his sample.
- 8 In disk drives, periodic data from Christensen (1993) on the leaders of the industry coupled with the data in Franco and Filson (2006) on the parents and years survived of each spinoff indicate that spinoffs of the leading firms survived markedly longer than other spinoffs. However, both Agarwal *et al.* (2004) and Franco and Filson (2006) found that the survival of spinoffs was *negatively* related to the quality of the technology of their parent firm even though firms with better technology survived longer and spawned spinoffs at a higher rate.

- 9 These covenants prevent employees for some period of time from working for a competing firm, which includes forming a competing firm.
- 10 Firms are afforded protection of so-called trade secrets even if the underlying intellectual property is not patented or copyrighted as long as they actively work to keep secret the intellectual property. Spinoffs cannot exploit the trade secrets of their parents. Even when at first they are not predicated on their parent's trade secrets, some courts have upheld the view that down the line they inevitably will exploit their parent's trade secrets and on that basis have prevented them from forming.
- 11 See Hellmann and Perotti (2006) for a formal model underlying this idea.

References

- Agarwal, R., R. Echambadi, A. M. Franco and M. B. Sarkar, 2004, "Knowledge transfer through inheritance: Spinout generation, development, and survival". *Academy of Management Journal*, 47: 501-522.
- Anton, J. J. and D. A. Yao, 1995, "Start-ups, spin-offs, and internal projects". *Journal of Law, Economics, and Organization*, 11: 362-378.
- Bailey, L. S., 1971, *The American car since 1775*. New York: Automobile Quarterly, Inc.
- Bassett, R. K., 2002, *To the digital age*. Baltimore, MD: Johns Hopkins University Press.
- Boschma, R. A. and R. Wenting, 2007, "The spatial evolution of the British automobile industry". *Industrial and Corporate Change*, 16: 213-238.
- Brittain, J. W. and J. Freeman, 1986, "Entrepreneurship in the semiconductor industry". Working paper.
- Buenstorf, G., 2007, "Evolution on the shoulders of giants: Entrepreneurship and firm survival in the German laser industry". *Review of Industrial Organization*, 30: 179-202.
- Buenstorf, G. and S. Klepper, 2009, "Heritage and agglomeration: The Akron tyre cluster revisited". *Economic Journal*, 119: 705-733.
- Burton, M. D., J. P. Sorenson and C. Beckman, 2002, "Coming from good stock: Career histories and new venture formation". *Social Structure and Organizations Revised*, 19: 229-262.
- Cabral, L. and Z. Wang, 2008, "Spin-offs: Theory and evidence". Working paper.
- Cantner, U., K. Dressler and J. J. Krueger, 2006, "Firm survival in the German automobile industry". *Empirica*, 33: 49-60.
- Cassiman, B. and M. Ueda, 2006, "Optimal project rejection and new firm start-ups". *Management Science*, 52: 262-275.
- Chatterjee, S. and E. Rossi-Hansberg, 2008, "Spinoffs and the market for ideas". Working paper no. 08-26, FRB of Philadelphia.
- Chatterji, A. K., 2008, "Spawned with a silver spoon". *Strategic Management Journal*, 30: 185-206.
- Chesbrough, H., 2003, "The governance and performance of Xerox's technology spinoff companies". *Research Policy*, 32: 403-421.
- Chesbrough, H. and R. R. Rosenbloom, 2002, "The role of the business model in capturing value from innovation: Evidence from Xerox corporation's technology spinoff companies". *Industrial and Corporate Change*, 11: 529-555.
- Christensen, C. M., 1993, "The rigid disk drive industry: A history of commercial and technological turbulence". *Business History Review*, 67: 531-588.
- Dahl, M. S. and T. Reichstein, 2007, "Are you experienced? Prior experience and the survival of new organizations". *Industry and Innovation*, 14: 497-511.
- Elfenbein, D. W., B. H. Hamilton and T. R. Zenger, 2008, "The small firm effect and the entrepreneurial spawning of scientists and engineers". Working paper.
- Ellis, S., I. Drori and Z. Shapira, 2008, "Genealogy as evolutionary dynamics". Working paper.
- Eriksson, T. and J. M. Kuhn, 2006, "Firm spin-offs in Denmark, 1981-2000 - Patterns of entry and exit". *International Journal of Industrial Organization*, 24: 1021-1040.
- Florida, R. and M. Kenney, 1990, *The breakthrough illusion*. New York: Basic Books.
- Franco, A. M. and D. Filson, 2006, "Knowledge diffusion through employee mobility". *RAND Journal of Economics*, 37: 841-860.
- Gambardella, A. and C. Panico, 2008, "Designing governance mechanisms for knowledge-intensive activities". Working paper.
- Garvin, D. A., 1983, "Spin-offs and the new firm formation process". *California Management Review*, 25: 3-20.
- Gilder, G., 1989, *Microcosm*. New York: Simon and Schuster.
- Gilson, R. J., 1999, "The legal infrastructure of high technology industrial districts: Silicon Valley, route 128, and covenants not to compete". *New York University Law Review*, 74: 575-629.
- Gompers, P., J. Lerner and D. Scharfstein, 2005, "Entrepreneurial spawning: Public corporations and the genesis of new ventures, 1986-1999". *Journal of Finance*, 60: 577-614.
- Hellmann, T. F., 2007, "When do employees become entrepreneurs?". *Management Science*, 53: 919-933.
- Hellmann, T. F. and E. C. Perotti, 2006, "The circulation of ideas: Firms versus markets". CEPR Discussion paper no. 5469.
- Higgins, M. C., 2005, *Career imprints*. San Francisco: Jossey-Bass.
- Hyde, A., 2003, *Working in Silicon Valley: Economic and legal analysis of a high-velocity labor market*. Armonk, NY: M.E. Sharpe.
- Klepper, S., 2001, "Employee startups in high-tech industries". *Industrial and Corporate Change*, 10: 639-674.
- Klepper, S., 2007, "Disagreements, spinoffs, and the evolution of Detroit as the capital of the U.S. automobile industry". *Management Science*, 53: 616-631.
- Klepper, S., 2009a, "Silicon Valley - A chip off the old Detroit bloc". In D.B. Audretsch and R. Strom (eds.) *Entrepreneurship, growth, and public policy*. Cambridge, UK: Cambridge University Press, pp: 79-118.
- Klepper, S., 2009b, "The origin and growth of industry clusters: The making of Silicon Valley and Detroit". Working paper.
- Klepper, S. and S. D. Sleeper, 2005, "Entry by spinoffs". *Management Science*, 51: 1291-1306.
- Klepper, S. and P. Thompson, 2009, "Intra-industry spinoffs". Working paper.
- Klepper, S., J. Kowalski and F. Veloso, 2009, "Technological spillovers and the agglomeration of the semiconductor industry in Silicon Valley". Working paper.
- Lécuyer, C., 2006, *Making Silicon Valley*. Cambridge, MA: MIT Press.
- Lindholm-Dahlstrand, Å., 2001, "Entrepreneurial origin and spin-off performance: A comparison between corporate and university spin-offs". In P. Moncada-Paterno-Castello (ed.) *Corporate and research-based spinoffs: Drivers for knowledge-based innovation and entrepreneurship*. European Commission, IPTS Technical Report Series, EUR 19903 EN, 43-66.
- Marx, M., D. Strumsky and L. Fleming, 2009, "Noncompetes and inventor mobility: Specialists, stars, and the Michigan experiment". *Management Science*, 55: 875-889.
- McKendrick, D. G., J. B. Wade and J. Jaffee, forthcoming, "Good riddance? Spin-offs and the technological performance of parent firms". *Organization Science*.
- Mitton, D. G., 1990, "Bring on the clones: A longitudinal study of the proliferation, development, and growth of the biotech industry in San Diego". In N.C. Churchill, W.D. Bygrave, J.A. Hornaday, D.F. Muzyka, K.H. Vesper and W.E. Wetzel (eds.) *Frontiers of entrepreneurship research*. Wellesley, MA: Babson College, pp: 344-358.
- Moen, J., 2005, "Is mobility of technical personnel a source of R&D spillovers?". *Journal of Labor Economics*, 23: 81-114.
- Moore, G. and K. Davis, 2004, "Learning the Silicon Valley way". In T. Bresnahan and A. Gambardella (eds.) *Building high-tech clusters: Silicon Valley and beyond*. Cambridge, UK: Cambridge University Press, pp: 7-38.
- Nanda, R. and J. B. Sørensen, 2008, "Peer effects and entrepreneurship". Working paper 08-051, Harvard Business School.
- Pakes, A. and S. Nitzan, 1983, "Optimum contracts for research personnel, research employment, and the establishment of 'rival' enterprises". *Journal of Labor Economics*, 1: 345-365.
- Phillips, D. J., 2002, "A genealogical approach to organizational life chances: The parent-progeny transfer". *Administrative Sciences Quarterly*, 47: 474-506.
- Roberts, P., S. Klepper and S. Hayward, 2007, "Founder backgrounds and the evolution of firm size and scope". Working paper.

- Samila, S. and O. Sorenson, 2009, "Non-compete covenants: Incentives to innovate or impediments to growth". Working paper.
- Saxenian, A., 1994, *Regional advantage*. Cambridge, MA: Harvard University Press.
- Sherer, J., 2006, "Entrepreneurial innovation: Evidence from the American laser industry". Working paper.
- Sørensen, J. B., 2007, "Bureaucracy and entrepreneurship: Workplace effects on entrepreneurial entry". *Administrative Science Quarterly*, 52: 387–412.
- Sørensen, J. B. and D. J. Phillips, 2008, "Competence and commitment: Employer size and entrepreneurial commitment". Working paper.
- Sporck, C. E., 2001, *Spinoff*. Saranac Lake, NY: Saranac Publishing.
- Sleeper, S., 1998, "The role of firm capabilities in the evolution of the laser industry: The making of a high-tech market". Ph.D. Dissertation, Carnegie Mellon University.
- Stuart, T. E. and O. Sorenson, 2003, "Liquidity events and the geographic distribution of entrepreneurial activity". *Administrative Sciences Quarterly*, 48: 175–201.
- Tilton, J. E., 1971, *International diffusion of technology: The case of semiconductors*. Washington DC: The Brookings Institution.
- Von Rhein, K., 2008, "Heritage and firm survival – An analysis of German automobile spinoffs, 1886–1939". *Economics Bulletin*, 12: 1–8.
- Wagner, J., 2004, "Are young and small firms hothouses for nascent entrepreneurs? Evidence from German micro data". *Applied Economics Quarterly*, 50: 379–391.
- Walker, R., 1992, *Silicon destiny*. Milpitas, CA: C.M.C. Publications.
- Wenting, R., 2008, "Spinoff dynamics and the spatial formation of the fashion design industry, 1858–2005". *Journal of Economic Geography*, 8: 593–614.
- Wilson, D., 2004, "Linear technology: Enviaible position". Reed Electronics, www.reed-electronics.com/moversandshakers/article/CA6253.
- Wiggins, S. N., 1995, "Entrepreneurial enterprises, endogenous ownership, and the limits to firm size". *Economic Inquiry*, 33: 54–69.

CHAPTER 1

A SPECIAL FORM OF LEADERSHIP FOR INNOVATION?

Innovation leadership? It is passion; it is learning; it is humility in front of mistakes and errors - understanding that they are necessary elements to learn faster than the others - and it is the target setting . . . yes, stretched targets!

Pekka Ala-Pietilä

Former President of Nokia1

Many companies claim that innovation is one of their critical values and priorities. Stakeholders are reassured that management is vibrantly committed to innovation as a source of customer value, organic growth and job protection. However, the reality is often less bullish than the intent. R&D may be busier than ever developing new products, but how many can be called truly innovative? Projects are proliferating in most companies, for sure, but which ones will reinvent their category or take the company into a brand new market? Why don't product managers dare to go beyond renewing current products or providing line extensions? Which management teams have successfully crafted an innovation vision and built an effective innovation culture and process within their organizations?

If the innovation testimonials contained in so many annual reports were accurate, we would have thousands of examples of truly innovative companies, and the mystique of who does it well would be of little interest. Yet when we look beyond the message for the marks of an archetypically innovative company, only a dozen or so really stand out. These iconic companies, often cited by innovation pundits and the media as first in class, become fallible and begin to lose their 'magic innovation touch' when changes occur within the leadership ranks. Some examples:

- 3M struggled to integrate the Six Sigma credo of its former CEO, James McNerney, into its traditional innovation culture.
- Apple experienced a performance roller-coaster before the return of Steve Jobs as CEO.
- Intel struggled to diversify its product line fast enough to face the growing market of mobile devices.
- Procter & Gamble had sluggish organic growth before the appointment of A.G. Lafley as CEO.
- Corning witnessed each of its blockbuster markets flounder and is constantly trying to reinvent itself.
- Dell had to kick-start its growth again after its highly praised direct business model reached a plateau.
- Others like Sony, Pfizer, Nokia and Airbus were put on a pedestal for their innovativeness, and yet have gone through turbulent times in the past few years.

Why does this happen?

THE LEADERSHIP FACTOR

Some companies surprise the market with one brilliant innovative move - like Pilkington with its float-glass technology - and then fall back into an innovation dormancy. Others may have an innovative surge but are unable to sustain it in the long term. These innovative spells, when not triggered by pure serendipity, generally reflect a high degree of faith and determination on the part of the current executive team: faith in the competitive power of innovation; determination to turn it into a core capability. But CEOs and management teams change, as do market and competitive conditions. New leaders often bring with them new management and change priorities. Newly arrived CEOs may introduce management philosophies and processes that boost innovation, as A.G. Lafley did at P&G with the 'connect and develop' approach. Sometimes they launch new policies and tools to improve business performance that restrict their staff's traditional innovation freedom, as exemplified by McNerney's controversial introduction of a systematic Six Sigma process at 3M.² Unless innovation is deeply ingrained in the genes of the company, in both culture and process, it is liable to become a second-level priority when leadership changes.

Many Try . . . Few Keep at It!

At some stage, most companies will launch a company- or division-wide innovation improvement campaign. Some zealous management teams attack the problem with a top-down approach, launching a massive innovation change program throughout the company. The Centurion program initiated by Royal Philips Electronics' CEO Jan Timmer in the 1990s fits in this category. These efforts focus on restructuring the company's innovation process and organization. Some companies may gain benefits from a streamlined process, but it is paramount for the company culture to change, or behaviors will remain the same and innovative results will flounder.

By contrast, the majority of management teams approach innovation in a low-key, pragmatic way. They do not engage in a big public change program, but instead look for low-hanging fruits, fixing the deficient parts of their innovation process as they find them, step by step. This may improve performance initially, but without an overall innovation vision and model, company culture and behavior generally do not change, which prevents the full benefit of their efforts being realized.

Fewer companies manage the process well. One that has succeeded is the packaging giant Tetra Pak. The leadership team not only overhauled the company's innovation capabilities, which has improved and streamlined processes, but is also working hard at mobilizing staff. Using its company-wide leadership development and culture change programs, Tetra Pak continuously promotes the adoption of innovation initiatives. The company has also put in place innovation steering mechanisms that should promote innovation in the long term and safeguard against changes in top management.

Most management teams today do a reasonably good job of streamlining and formalizing their innovation process and adapting it to the imperatives of their industry. The determining factor for sustained innovation performance - or lack of it - seems to be the level and consistency of commitment to innovation at the top. Management attitudes to innovation create the 'collective innovation leadership' and this is generally ingrained in the company culture. This is why we propose that there is a specific and distinctive form of leadership for innovation, which not all leaders possess and which this book will illustrate.

Innovation Leadership

There is no shortage of books and articles describing the core characteristics of innovative organizations. Jones and Austin, for example, have compiled a list of five core characteristics of 'innovation leaders':³

- in-depth customer insight;
- leading-edge technical awareness;
- inspirational leadership;
- motivational organizational rewards;
- sharing knowledge.

But these 'differentiators of enhanced innovation performance,' as they call them, relate more to the collective management of innovative companies than to specific individuals. To date, there has been no formal attempt to paint a comprehensive portrait of 'innovation leaders' as defined in this book.

Based on empirical research, this book will analyze the profiles and attributes of various innovation leaders. The portrait will be impressionistic to include a great diversity of characters. Each brush stroke will add a dimension to our description of the special forms of leadership that foster innovation.

Defining Leadership

Professor Preston Bottger, who teaches organizational behavior at IMD business school in Lausanne, has coined a simple definition that conveys the full dimension of leadership:

Leaders do or cause to be done all that must be done and is now not being done to achieve what we say is important! They provide a sense of purpose, direction and focus. They build alignment and get commitment!⁴

When it is applied to innovation, this definition has several merits.

First, true leaders are action-oriented change agents; they don't just think and talk, they 'do or cause to be done . . .' Most companies state that innovation is important, but what do they really do other than invest money in R&D?

Second, this definition highlights three types of fundamental questions raised by most innovation drives:

- (1) Leaders provide a 'sense of purpose,' i.e. Why are we doing it? What are the benefits of a change in innovation? What are the penalties if we don't do it?
- (2) They propose a 'sense of direction,' i.e. Which way should we go? What innovation model should we adopt?
- (3) They introduce a 'sense of focus,' i.e. What are our priorities? Where should we concentrate our efforts?

Third, this definition stresses that if innovation is to become a corporate capability, it cannot be confined to a specialist function or a small group, for example to new business development or R&D. It has to permeate the entire organization, become a priority and then an expectation - with this kind of commitment the motivation will be there to make it happen.

Is There a Special Form of Leadership for Innovation?

I like to ask this question to executives who participate in my innovation courses, forcing them - unfairly, I admit - to give a simple yes or no answer. The answers are usually split. Those who come from R&D and register specifically for a course dedicated to innovation, tend to vote overwhelmingly 'yes.' Coming from the innovation functions of their business, they may not be able to articulate what innovation leadership entails, but they understand it instinctively. By contrast, executives attending a single session on innovation as part of a general management course seem to be more split in their responses, even though the 'yeses' usually prevail.

Those who answer 'no' typically argue that purpose, direction and focus are needed in all business endeavors, including innovation. Consequently, a true leader should be able to become an innovation leader if and when conditions require it. Executives who do not believe in a special form of leadership for innovation tend to refer intuitively to mental models of what leaders actually do. Some of the most popular leadership models support their claim that leadership is a universal trait that embraces innovation (refer to Appendix A for a reference to such models).

By contrast, managers who believe that innovation requires a special form of leadership maintain that if this weren't so, then most business leaders would excel at innovation if they paid attention to it. But as the evidence shows, this is not the case in many companies. Furthermore, few of the leadership icons celebrated by the media for their achievements in shareholder value creation, like Jack Welch at GE, could claim that innovation is their forte. Most would not qualify as innovation leaders and the opposite also seems to be true, i.e. not all innovation leaders are fully fledged business leaders. These arguments convince many managers that since innovation is different from most other business endeavors, it probably requires different attitudes and behaviors.

FACING THE INNOVATION IMPERATIVES

Before trying to characterize the unique traits of innovation leaders, let's look at some of the essential aspects of innovation, and reflect on the challenge they raise for business leaders. We shall focus on six of these innovation imperatives:

- the urge to do 'new things';
- an obsession with redefining customer value;
- the courage to take risks;
- an ability to manage risk;
- speed in spotting opportunities and in project execution;
- a shift in focus and mindset from business optimization to business creation.

Innovation Requires an Insatiable Urge to Try New Things

It goes without saying that innovation is about challenging the status quo and introducing new and, one hopes, better products, processes, services or management approaches. Innovation requires curiosity, experimentation and openness to change. Innovation leaders are those who constantly challenge the present state of affairs, encourage wild ideas and instigate trying new things in their companies.

Despite frequent management denials, many companies adopt an 'if-it-ain't-broke-don't-fix-it' stance. Therefore, innovation leaders must have the courage to foster a climate of experimentation and permanent change in their organizations.

It's no surprise that few mavericks and innovation champions exist in most top management teams. Career progression often favors managers who deliver results without making waves, not the revolutionaries. The creators of the 'organized chaos' so dear to innovation scholars⁵ often meet obstacles and resistance on their way to the top. To stimulate innovation, however, companies must promote 'challengers,' not just 'fixers.'

Innovation Requires an Obsession with Redefining Customer Value

Innovation has to do with adding value, and the way to add value is through leadership, argues Nick Shreiber, former CEO of Tetra Pak:

One can add value in many ways. The most important, perhaps, is through leadership - a very elusive concept! Just like good judgment, good leadership is hard to define, but you know it when you see it! Leadership can inspire an organization to reach goals it had never dreamed of, and will encourage each employee to reach his or her full potential in pursuit of their objectives. Inspired leadership will encourage new ideas through innovation and entrepreneurship and will provide the resources to implement them.⁶

In hindsight, highly successful innovators have generally established new standards of value in their industries. For a long time, value creation came primarily from leading-edge technology-based products or processes. Michelin redefined the notion of value in tires - as expressed in mileage life - with its radial tire technology, and Sony did something similar with its PlayStation game consoles. Nowadays, value creation can come from introducing radically new business models or management methods. It is no longer necessary to be a great technical innovator to qualify as an innovation leader. By radically changing the economics of the PC industry, not the product itself, Michael Dell can arguably be called an innovation leader:

People look at Dell and they see the customer-facing aspects of the direct-business model, the one-to-one relationships. What is not really understood is that behind these relationships lies the entire value chain: invention, development, design, manufacturing, logistics, service, delivery, and sales. The value created for our customers is a function of integrating all those things.⁷

Kim and Mauborgne suggest that redefining value starts with questioning current industry assumptions by asking four probing questions:

- Which of the factors that our industry takes for granted should be eliminated?
- Which factors should be reduced to well below the industry standard?
- Which factors should be raised well above the industry standard?
- Which factors that the industry has never offered should be created?⁸

Consciously or instinctively, innovation leaders challenge industry assumptions in order to unearth opportunities for a quantum jump in customer value. A strong customer orientation often fuels this urge to redefine value. Value creators, typically, have an insatiable curiosity about their customers' needs, empathy with their conscious or subconscious frustrations, and an instinct for what they might need or want in the future. As Akio Morita⁹ stressed in his story of Sony's legendary Walkman[®], this type of curiosity is not synonymous with a thirst for traditional market information. No market research, he argued, would have indicated a need for the Walkman[®]. Morita is referring, rather, to the kind of customer intimacy that comes from a deeply ingrained, instinctive curiosity. Sony's past advertising slogan - 'You dreamt it! Sony made it' - reflects the company's view of its innovation mission: To redefine value constantly by correctly guessing the customer's unarticulated desires, and applying its technological expertise to satisfy them.

The challenge for innovation leaders is to encourage this constant reappraisal of value factors despite the fact that, at times, such an attitude may prove highly destabilizing. Challenging the current ways of delivering value in your industry is very difficult when you are an established player and even more so when you are the market leader. As Harvard Business School professor Clayton Christensen convincingly demonstrated, introducing disruptive technologies and defying the status quo is much more natural for new entrants looking for ways to challenge incumbents. ¹⁰ This is why many innovations have originated with outsiders who forced their way into the market with radically new concepts.

The highly successful story of the no-frills, low-cost airlines - first pioneered by Southwest Airlines in the US, then by Ryanair and easyJet in Europe¹¹ - provides a good illustration of this rule. Their founders challenged every single prevailing assumption in the traditional airline industry¹² to come up with a revolutionary business model. This gave them unbeatably low costs and allowed them to redefine the notion of value for budget-conscious air travelers. Arguably, it would have been very difficult for any traditional airline to introduce such radical changes internally.

Innovation Requires the Courage to Take Risks

One of the most widely recognized drivers of innovation is management's willingness to take risks. It is hotly debated because risk taking is subject to all kinds of interpretations. In its classical definition, risk taking for innovation is related to the concept of entrepreneurship - being ready to bet one's resources on a new, and often untested, business proposition.

The challenge for innovation leaders is to live by this principle on a day-to-day basis and make the rest of the organization comply with it as well.¹³ Although many companies describe risk taking as one of their core values, they often fail to change their performance review and reward systems accordingly. Managers are rarely penalized for not taking risks, especially if they are meeting their targets. The right to fail comes up invariably in most innovation speeches, but it is not necessarily carried into practice.

Andy Grove,¹⁴ Intel's legendary former CEO, adds two very interesting dimensions to the risk taking imperative. First, he claims that innovation leaders must have the courage to focus, which means identifying unambiguously either the things they will not do or the things they will stop doing. Second, Grove believes that innovation leaders must have the courage to 'self-cannibalize,' i.e. to make their own business obsolete before others force obsolescence on them. As we know, it takes courage to kill one's own products before their full potential has been exploited and to replace them with higher-performance - but unproven - ones, as a venture capital partner suggests:

You have to decide you're going to eat your own business yourself, as opposed to having eToys or Amazon or somebody else doing it for you. This is a very different mindset from most companies that are trying to protect what they've got, as opposed to cannibalizing.¹⁵

It is this policy, coupled with management's belief in the now famous Moore's 'law'¹⁶ that enabled Intel to stay at the top of its industry for so long. Whereas the willingness to take entrepreneurial risk applies to all managerial echelons, Grove's observations apply only to the highest level of innovation leaders, the CEO and his/her key executives.

Innovation Requires an Ability to Manage Risk

The debate about acceptable levels of risk in an innovation project often pits risk takers (usually the project champions) against those who shrink from taking risks (typically senior managers). Innovators often complain that the controlling attitude of their top managers hides a fundamental aversion to risk, while the more conservative proponents of risk management accuse risk takers of being irresponsible. This debate is fruitless because both arguments are right. Innovation is as much about good risk management as it is about risk taking.

The challenge for innovation leaders, therefore, is to strike a balance between single-minded, enterprising risk taking and pragmatic, cautious risk management. The first attitude is necessary for pushing ahead and brushing away objections. In a sense, frontline innovation champions should be so determined and persistent that they could be accused of being both blind and stubborn. Innovation leaders, by contrast, carry the burden of ensuring that all the known risk factors have been identified at each stage and properly managed - a precarious balance, as this needs to be done without discouraging innovators and entrepreneurs.

A dilemma arises whenever the CEO or business unit head is simultaneously the champion of a particular project and the leader who is supposedly responsible for containing risk. No one will dare oppose his/her hierarchical head by spotlighting dangerous risk factors on the boss's favorite project. The story of Philips' ill-fated CDi¹⁷ illustrates that danger. It was well known within Philips that its determined CEO, Jan Timmer, had adopted the CDi as his pet project, as he had successfully championed the CD-Audio years earlier. Many in the company argue today that the CDi concept had inherent flaws and that its proponents blindly underestimated the competing PC-based technology, CD-ROM. Very few dared to openly challenge the notoriously tough CEO, and finally, after a few years and huge losses, Philips abandoned the project.

A similar story can be told about the energetic pursuit of the market for genetically modified organisms (GMOs) at Monsanto. Its CEO, Robert Shapiro, was consumed by the vision of Monsanto becoming a life sciences powerhouse on the strength of its genetic engineering technology. And he was convinced that realizing his vision meant betting the company's future on GMOs and promoting them aggressively worldwide. But experts are likely to point out that, after the controversy over the company's commitment to GMOs erupted in the media, Monsanto's top management failed to grasp the power of the arguments of GMOs' detractors. It is hard to be a visionary, risk taking innovation champion, while at the same time being a cautious risk analyzer and container. This balance is the challenge of innovation leaders.

Innovation Requires Speed in Spotting Opportunities and in Project Execution

Silicon Valley innovators and entrepreneurs have known for a long time that the best idea or the best technology does not necessarily win - the winner is the one that is implemented first.¹⁸ Whoever comes first learns fastest. Success with new products comes from launching first, then learning fast to correct mistakes before others have prepared their response, and relaunching a superior product as competitors start coming in. In the words of Matt Hobart, a 28-year-old Silicon Valley entrepreneur:

If you have an idea, it's safe to assume that four or five people have the same idea. But it's not the person with the best idea who wins. It's the person who can execute quickly.¹⁹

That kind of speed requires three unique skills:

- (1) the ability to search continuously for opportunities;
- (2) management decisiveness at all stages in the process; and
- (3) speed in execution, typically achieved through a pragmatic reliance on external and internal resources, and, of course, highly effective teams.²⁰

Innovation leaders instinctively create an environment that values the search for opportunities and the generation of ideas to exploit them. They typically encourage people to flag opportunities early and make their ideas bubble freely upward for discussion. The challenge lies in the decision process. On what grounds should the project go ahead? What criteria should be met at each stage? When and on what basis should the plug be pulled? As the champions of risk taking entrepreneurs, innovation leaders are bound to allow their staff both a fair amount of freedom to experiment and the necessary resources. Finding an acceptable balance is a challenge, and so is the need to decide fast, whatever the decision. In Silicon Valley, innovators usually get the same advice from venture capitalists: If you are going to fail, at least fail fast and fail better!

Innovation Requires a Shift in Focus and Mindset: From Optimizing Business to Creating Business

Business unit heads are generally responsible for new product development in their fields and innovation is generally pursued to protect and grow the current business, seldom to create new businesses. This is why most companies struggle to exceed the growth rate of their industry. How can Unilever or Nestlé grow in the mature food industry except by creating entirely new, and hence faster-

growing, product categories? Now that the second-generation mobile phone market is nearing saturation, the same question applies to Nokia and Motorola. Creating new businesses is completely different from tweaking product lines to introduce extensions.

So, innovation leaders face a double challenge. The first is to strike the right balance between running the current business and growing new businesses, or as Professor Derek Abell puts it, between mastering the present and preempting the future.²¹ The sudden shift in what financial markets demand in the way of share performance - from growth potential yesterday to profitability today - makes finding the right balance a tough task. The challenge is for companies to avoid the tyranny of success and learn to 'organize both incremental and disruptive innovative activities.'²²

The second challenge for innovation leaders is sensing untapped market needs and choosing promising areas to pursue. Here, innovation leaders must have the ability to shape a vision that will guide them toward new business opportunities.

We have so far talked about innovation and its imperatives in generalities, as if innovation was a uniform process without any 'subspecies.' The reality is more complex and, as we have all observed, there are many different types of innovations. As a consequence, it is legitimate to ask whether different styles of leadership are required to handle the different types of innovation. This is what this book is about. But before attempting to define and characterize various types of innovation leaders, we will first establish a broad typology of innovations.

DEFINING AND CHARACTERIZING INNOVATION

Even though everyone talks about innovation, there is still confusion as to what the word really means and entails in the business world. 3M distinguishes between research - transformation of money into knowledge - and innovation - transformation of knowledge into money. The Organization for Economic Cooperation and Development (OECD) proposed the following general definition of innovation:

. . . an iterative process initiated by the perception of a new market and/or new service opportunity for a technology-based invention which leads to development, production and marketing tasks striving for the commercial success of the invention.²³

Although this definition is slanted toward technology- and product-based inventions - by no means the only types of innovation - it has the merit of considering innovation as a wide-ranging business undertaking.

Defining the Processes in Innovation

Another way to define innovation is to refer to its processes, grouped around easy-to-remember 'i' words. The following series can help define what innovation covers: Innovation is the combination of two processes - invention and implementation.

Invention is itself the result of immersion in the market to identify unmet needs, or immersion in the problem at hand. This is followed by a phase of imagination to envision the potential benefits of

addressing that opportunity, ideation to develop and select attractive new concepts to meet the identified need, and initiation of a concrete project or venture.

Implementation, in turn, consists of an incubation phase to develop and test the new product or service, followed by an industrialization process to make it and deliver it in large quantities. This is followed by an introduction phase with an initial launch, followed by roll-out and full deployment, complemented at each customer site by a phase of installation and integration to ensure that the new product or service is adopted and integrated into the customer's organization and processes. This simplified typology will lead us to explore different types and styles of innovation leaders.

Innovation observers and scholars have long pointed to the existence of two very different patterns of innovation generation and diffusion within a company: 'bottom up' and 'top down' (see Figure 1.1). This distinction has a direct bearing on our topic because, as we see in the following chapters, each mode requires a different type of focus on the part of innovation leaders.

Figure 1.1 The two modes of innovation

In the bottom-up mode, innovative ideas originate spontaneously from people at the operational level, whatever their function. These ideas get developed out in the open and the resulting projects flow upward for management funding and support. This type of innovation is driven by the commitment and dedication of internal entrepreneurs who feel encouraged and empowered by management. The main driver of bottom-up innovation is the entrepreneurial culture of the organization, which encourages individual initiatives, experimentation and risk taking.

Top-down innovation, by contrast, is initiated by management in response to an ambition or the vision of an attractive business opportunity. The big idea that generally results from that vision flows downward to the teams that are then mobilized for its implementation. The main driver of top-down innovation is the organized process by which an innovation vision is made 'actionable' by management and ultimately implemented.

In truly innovative companies, both modes can coexist because they are complementary. The most promising ideas from those generated in a bottom-up mode may be appropriated higher up by management and turned into top-down projects with strong management involvement and guidance. Similarly, a top-down initiative may be launched by management, but handed over to the staff with the mandate to generate creative ways to implement it bottom up.

Nevertheless, some companies are known for using one of the two modes as their 'default' innovation pattern. For example, 3M was long qualified as an archetypical bottom-up innovator, at least until the arrival in 2001 of its CEO James McNerney who tried to rebalance its focus toward more top-down innovation. In contrast, Japanese technology companies like Canon are said to be more inclined to innovate in a top-down mode with strong management involvement.

Professor Eric Mankin from Babson College highlights that the two innovation modes differ on at least three criteria:

- (1) the number of initiatives;
- (2) the way results are generated; and
- (3) the level of iteration.

Table 1.1 Best Buy vs. GE

Table 1.1 highlights how Mankin contrasts the approaches of retailer Best Buy (a declared bottom-up innovator) and GE (a proponent of top-down innovation) on these three criteria.²⁴

But these two innovation modes differ also in their leadership focus and requirements. By nature, bottom-up innovation occurs spontaneously - i.e. without direct management intervention - in the right kind of culture or climate. The main role of leaders in encouraging bottom-up innovation is to proactively develop a highly supportive culture.

Top-down innovation, by contrast, is steered by management. Making the vision a reality is what top-down innovation leaders excel at doing.

INNOVATION LEADERS: A DIFFERENT BREED?

Defining Innovation Leaders

In summary, innovation leaders can be defined as those senior executives who promote an innovation agenda in their company. Whatever their function or position, they instigate, sponsor and steer innovation in their organization. Through personal conviction or competitive necessity, they are obsessed with providing superior new value to customers. Even in the face of resistance from their top management colleagues, these executives stand up for innovators and challengers of the status quo. They know how to mobilize their staff behind concrete initiatives and they do not hesitate to personally coach innovation project teams.

Many times in innovation literature, they are named 'champions,' 'sponsors' or 'promoters.'²⁵ Whatever they are called, true innovation leaders tend to share the same determination and are not afraid to risk their credibility with top management in case of failure. Lewis Lehr, the highly charismatic former CEO of 3M, described the behavior of an innovation leader very convincingly when he said, 'We learned to follow the fellows who follow a dream!'²⁶

The ideal place for an innovation leader is, obviously, at the head of the company or one of its businesses. The archetype is the CEO of the company he/she has helped create. Famous names spring to mind: Edwin Land at Polaroid, Robert Noyce at Intel, Steve Jobs at Apple and Bill Gates at Microsoft and, more recently, John Chambers at Cisco, Jeff Bezos at Amazon or Larry Page and Sergey Brin at Google. But charismatic entrepreneurs are not the only innovation leaders worth considering. Innovation leaders can be found at various management levels in different types of companies. They also come from different parts or functions of the organization, with a particular emphasis on marketing and R&D.²⁷ With or without top management blessing, they are committed to keeping alive the company's innovation legacy - if it exists - or, more often, restoring it. Depending on their personal orientation, they

see themselves as the linchpins of their company's innovation process and/or the evangelists of an innovation and entrepreneurship culture.

Innovation leaders use a variety of levers to improve their company's innovation process and forge a strong innovation culture. They seem generally to share a number of distinctive leadership characteristics, particularly when compared with other excellent but more traditional business leaders.

The Need for a Network of Innovation Leaders

Marvin Bower, McKinsey's legendary managing partner and leadership guru, maintains that '... a business should be run by a network of leaders positioned right through the organization.'²⁸ This belief probably applies even more to innovation leaders than to any other types. Indeed, innovation is never the result of a single person's efforts, either at the project level or at the sponsoring level. As the well-known saying goes, 'It takes only one "no" coming after nine "yeses" to kill a project.' Innovation is in danger if it lies in the hands of an isolated leader in the top management team, whatever his/her charisma. The first role of an innovation leader is, therefore, to breed or attract others to take on leadership roles, propagate innovation values and support concrete projects.

It is relatively easy for a lone innovation leader to build a team of subordinates sharing similar values and behaviors for two reasons. First, people tend to be attracted to like-minded people. And second, unless they are authoritarian, innovation leaders usually exude a high level of openness and communicate enthusiasm, to say nothing of passion. Working for them is exciting!

The situation is more complex at the top management level. Lone innovation leaders, unless they occupy the top job themselves, may be unable to influence the profile and behavior of their top management colleagues. They need to muster CEO support to be effective. If they show growth and results, they can hope to propagate their values through sheer emulation. When they have established a reputed stable of talent in their organizations, they transfer some of their best and most motivated staff into other divisions, in the hope of initiating a bottom-up movement of contagion.

MAPPING OUR JOURNEY

Defining and Characterizing Innovation Leaders

As we have established that there is a special form of leadership needed for innovation, Chapter 2 will further paint the portrait of innovation leaders by characterizing what differentiates this subset from other types of leaders - behavior, common personality traits, instincts and actions. As there is a broad universe of innovation leaders, Chapter 2 will classify innovation leaders according to their focus on a particular aspect of the innovation process, i.e. the front end vs. the back end and show that they naturally tend to adopt a preferred mode of innovation, i.e. bottom up or top down.

Bottom-up innovation and what leaders can do to encourage and sustain it will be the main theme of Chapter 3. Bottom-up innovation is the embodiment of the company's innovation culture, which often reflects the history of the organization and the legacy of its founders or charismatic leaders. This does not mean that bottom-up innovation is limited to companies that have kept their historic innovation heritage intact. In fact, through their attitudes, policies and processes, leaders can exert a strong influence on at least four direct enablers of innovation, i.e. the company's organizational creativity; the

systematic deployment of teams of complementary champions; the encouragement of customer intimacy practices; and the promotion of a 'can-do' climate.

Chapter 4 will explore the characteristics of top-down innovation and highlight how leaders reinvent their business, introduce disruptive technologies or steer their company into new market space. Top-down innovation usually stems from management's realization that changes in the market environment or technology offer big opportunities to disrupt an established industry. Innovation leaders mobilize their organization to seize that opportunity. They make sure that the big initial idea is turned into an actionable vision, i.e. one that leads to concrete implementation roadmaps and a seamless process.

Chapter 5 will focus on one of the role models for innovation leadership in companies, i.e. the chief technology officer (CTO) or chief research officer (CRO), sometimes called chief innovation officer (CIO). It will also examine the extent to which the role of these technical executives is changing, in terms of visibility within the senior management group, and it will highlight the CTO/CIO's new leadership challenges:

- (1) instilling a vision and sense of purpose for the role of science and technology;
- (2) providing a sense of direction for investments in science and technology;
- (3) enforcing a sense of focus on the technologies to be developed vs. those to be outsourced; and
- (4) becoming corporate entrepreneurs to turn technology into new businesses.

The Leadership Imperative of Innovation Strategies

The first part of the book is based, implicitly, on the assumption that innovation is a generic process that proceeds in a fairly similar fashion, whatever the circumstances and the company. Innovation leaders, it implies, display common characteristics and the differences among them pertain mainly to their natural emphasis - on the front end vs. the back end - and their preferred mode of intervention - top down vs. bottom up. The reality is arguably more complex and we all know that innovation takes on the most varied forms. It is therefore safe to assume that different innovation leadership styles may be needed for different types of innovation.

Chapter 6 will outline four different innovation thrusts, based on the development of:

- (1) new/improved products, processes or service offerings;
- (2) totally new product categories or service offerings;
- (3) totally new business systems or models; and
- (4) new/improved customer solutions.

These four thrusts share one common trait, i.e. an almost obsessive quest for a unique customer value proposition. However, each requires a distinct emphasis in terms of process, structure, culture and people. CEOs ought to map whether and how their senior officers meet some of the innovation

leadership traits required by their innovation strategy. The following four chapters will illustrate each of these aspects with an example and characterize their specific leadership imperatives.

The incremental development of new/improved products or services is the most prevalent type of innovation axiom, probably accounting for the bulk of R&D expenditures in most companies. The leadership imperatives of this type of thrust will be illustrated in Chapter 7 by the transformation of Medtronic from a renowned but weakening competitor in the industry it created - cardiac pacemakers - to a 'born-again' innovator and market leader. This story features a strong leader willing to confront a lenient but complacent culture and introduce a sense of urgency and a high degree of process discipline. This example also highlights the role of top management in supporting the new culture and its courageous and sometimes unpopular champion.

One can compare the leaders who focus on the incremental development of new/improved products to tough sports coaches, very demanding with their team but able to motivate them to give their best to win. Their emphasis is on challenging, setting goals and measuring.

The creation of a totally new product category through radical innovation is a less frequently adopted strategy. Few senior management teams feel comfortable taking a very long-term payback perspective and tolerating the uncertainty of moving into a completely new market space. This is nevertheless what Tetra Pak did when it decided to develop a retortable carton alternative to the ubiquitous metal can used for more than a century by the food industry. This example will be outlined in Chapter 8. It highlights the importance of management's initial vision; its persistence through the unavoidable ups and downs of a risky project; its dogged determination to remain faithful to its initial value proposition; and its willingness to steer and run such projects with a strong business focus.

Innovation leaders who concentrate on the development of totally new product categories or service offerings have many of the leadership characteristics of no-nonsense sponsors. They tend to be very supportive of their teams, but if they are visionaries, they also know how to keep their feet on the ground. They know how to make their teams confront and systematically address each obstacle in their way, in order to reduce risk. Their emphasis is on nurturing, challenging and empowering.

The generally long time frame of these innovation projects and their multi-functional emphasis often make it difficult for a single senior manager to steer such projects from beginning to end. Collective leadership by a team of senior managers is a key requirement. This means that various types of leaders will have to step in and out during the life of the project, while maintaining as much continuity as possible in what can be called an uninterrupted chain of leadership.

The creation of a totally new business system, together with selected internal or external partners, will be covered in Chapter 9. Most often it is accompanied by the introduction of a radically new business model, capable of deeply transforming an existing industry or creating a totally new one. TiVo, the iconic US proponent of view-on-demand TV, presents a good example of a 'business system' with its various

components: hardware, software and service. The TiVo story highlights the critical importance of specific leadership skills for handling this type of innovation.

The leaders capable of pulling off such system businesses or, more generally, business model innovations, have skills similar to those of pragmatic architects. They are capable of devising complex constructions and leading teams of different organizations to implement them, down to the finest details. Their emphasis is on visioning, partnering and master-planning.

Chapter 10 will focus on the development of incrementally new products that aim to offer customers a richer experience, because they provide a more comprehensive solution to their problems or needs than traditional products. These new 'solution-products' often consist of different elements, for example a product and the consumables that go with it, or a product and its customized delivery device. They may be provided by complementary partners, working together under different types of arrangements. We will illustrate this phenomenon by looking at what is happening in the home coffee business with the introduction of single-serve systems, notably by Sara Lee and its partner Philips (Senseo).

Leading such developments requires a deep understanding of what makes a good customer experience and the willingness to reach out to complementary partners that will share the same objective and deliver that experience in a repeatable fashion. This type of innovation thrust shares some of the characteristics of system business innovations, but is a lot less complex to orchestrate. To pull it off, however, leaders must have skills similar to those of orchestra conductors with their emphasis on interpreting, orchestrating and integrating the necessary input.

Developing a Cadre of Innovation Leaders

The concept of a 'chain of leadership,' introduced in Chapter 8, stresses the importance of having a number of innovation leaders willing to play complementary roles in the course of an innovation project. As Chapter 11 points out, this will happen only when the company has developed an innovation leadership culture, i.e. a set of management values and behaviors that foster the emergence and empowerment of a cadre of innovation leaders. Few large companies exhibit a visibly strong innovation leadership culture, at least such as the one prevailing in Logitech, the American and Swiss digital accessories company. Logitech has managed to grow profitably while maintaining the innovation spirit of its start-up era. Logitech's culture has developed through the combination of five critical elements:

- (1) A strong innovation legacy, rooted in the company's creation history and shaped by its defining moments, innovation achievements and threats.
- (2) A deeply competitive industrial and market environment, highlighting the critical importance of innovation as a survival process.
- (3) The visibility and influence of its major innovation role models, notably its founder and the CEO he chose to replace him.
- (4) The company's embedded values and its current management attitudes, policies and processes.
- (5) A great degree of attention to managing innovation as a process mixing creativity and discipline.

To conclude, Chapter 12 will address some of the key concerns of senior managers wishing to build a cadre of innovation leaders. It will avoid discussing whether leadership is an innate or developed talent, and whether you hire on attitudes and train for skills or the reverse because the answer to the two questions is, obviously: Both. Instead, we shall focus on what leaders of innovative companies do to: (1) assess; (2) attract, select and hire; (3) develop and deploy; and (4) retain talented individuals to lead their innovation efforts.

ENDNOTES

- 1 Ala-Pietilä, P. (2001). 'Leadership and Innovation.' Videotaped interview by J.-P. Deschamps, IMD, Lausanne, reproduced with permission. After stepping down as Nokia's president, Pekka Ala-Pietilä co-founded Blyk.com, a new mobile network for 16- to 24-year-olds that is funded by advertising.
- 2 Byrne, John and Hindo, Brian (2007). '3M's Innovation Crisis: How Six-Sigma Almost Smothered its Idea Culture.' Business Week Cover Stories podcast, June 11.
- 3 Jones, T. and Austin, S. (2002). Innovation Leadership. London, Datamonitor Plc.
- 4 Bottger, P. (2008). 'The Leadership Imperative.' Leading in the Top Team: The CXO Challenge. Ed. Preston Bottger. Cambridge, Cambridge University Press.
- 5 Peters, T. (1987). Thriving on Chaos. New York, Alfred A. Knopf. Quinn, J.B. (1985). 'Managing Innovation: Controlled Chaos.' Harvard Business Review, May-June: 73-83.
- 6 Shreiber, N. (2002). 2001 MBA graduation speech. Published in IMD's Webletter@imd.ch, January.
- 7 Technology Review (2001). 'Direct from Dell: At 19, he revolutionized the selling of PCs. AT 36 he's ready to take on HP, Sun, EMC and Cisco.' Q&A with Michael Dell. MIT Enterprise, July. <http://www.technologyreview.com/articles/qa0701.asp>.
- 8 Kim, W.C. and Mauborgne, R. (1997). 'Value Innovation: The Strategic Logic of High Growth.' Harvard Business Review, January-February: 103-112.
- 9 Morita, A. (1987). Made in Japan. London, William Collins, pp. 79-83.
- 10 Christensen, C.M. (1997). The Innovator's Dilemma. Boston, Harvard Business School Press.
- 11 Kumar, N. and Rogers, B. (2000). 'easyJet: The Web's Favorite Airline.' IMD Case No. IMD-3-0873.
- 12 They introduced point-to-point connections using low-cost airports, with no possibility of adding connecting flights; direct sales (90% via the internet), bypassing travel agents; sales staff paid on commission; no fixed prices for flights (extensive use of 'yield management methods'); no tickets; one type of aircraft; no business class; no meals; etc. - but high punctuality levels and unbeatable prices.
- 13 Perel, M. (2002). 'Corporate Courage: Breaking the Barrier to Innovation.' Industrial Research Institute, May-June, pp. 9-17.
- 14 Grove, A.S. (1996). Only the Paranoid Survive. New York, Bantam Doubleday.
- 15 Dunlevie, B. Benchmark Capital Management Co. Quoted in Stross, R.E. (2000). eBoys: The First Inside Account of Venture Capitalists at Work. New York, Crown Publishing Group.

16 Gordon Moore, one of Intel's founders, predicted in 1965 that the number of transistors inserted on a silicon chip would double every 18 to 24 months. As Intel tried to follow that 'law,' it became a kind of self-fulfilling prophecy.

17 CDi: Compact Disk interactive, a precursor of CD-ROMs and DVDs, introduced by Philips toward the end of the 1980s and abandoned in the early 1990s owing to the growing success of CD-ROMs.

18 Rogers, E.M. and Larsen, J.K. (1984). *Silicon Valley Fever: Growth of High Technology Culture*. New York, Basic Books.

19 Ratnesar, R. and Stein, J. (1999). 'This Week's Model.' *Time*, 19 September.

20 Cooper, R.G. (2003). 'Profitable Product Innovation: The Critical Success Factors.' *International Handbook on Innovation*. Ed. L.V. Shavinina. Amsterdam, Elsevier Science, pp. 139-157.

21 Abell, D.F. (1993). *Managing with Dual Strategies: Mastering the Present, Preempting the Future*. New York, Free Press.

22 Katz, R. (2003). 'Managing Technological Innovation in Business Organizations.' *International Handbook on Innovation*. Ed. L.V. Shavinina. Amsterdam, Elsevier Science, pp. 775-789.

23 Garcia, R. and Calantone, R. (2002). 'A Critical Look at Technological Innovation Typology and Innovativeness Terminology: A Literature Review.' *The Journal of Product Innovation Management* 19(2): 110-132.

24 Mankin, E. (2005). 'Top-down Innovation at GE.' *Babson Insight*.
<http://www.babsoninsight.com/contentmger/showdetails.php/id/847> accessed 23 November 2007.

25 Hauschildt, J. (2003). 'Promoters and Champions in Innovations: Development of a Research Paradigm.' *International Handbook on Innovation*. Ed. L.V. Shavinina. Amsterdam, Elsevier Science. In his chapter, Hauschildt distinguishes between three types of promoters - 'promoters by expertise,' 'power promoters' and 'process promoters' - who need to work as a 'troika' to stimulate and support innovation.

26 Lehr, L.W. (1979). 'Stimulating Technological Innovation: The Role of Top Management.' *Research Management*, November: 23-25.

27 Although entrepreneurial CEOs of start-ups and large company founders like Steve Jobs and Jeff Bezos undoubtedly qualify as innovation leaders - more instinctively than consciously for the former - they will not be the only subjects of this book. I will focus also on members of the top management team and senior executives of large established companies.

28 Bower, M. (1997). 'Developing Leaders in a Business.' *The McKinsey Quarterly* 4: 4-17.

Metrics and Tools for Measuring Construction Productivity: Technical and Empirical Considerations

Allison L. Huang, Robert E. Chapman, and David T. Butry





U.S. Department of Commerce
National Institute of Standards and
Technology

Office of Applied Economics
Building and Fire Research Laboratory
Gaithersburg, Maryland 20899

Metrics and Tools for Measuring Construction Productivity: Technical and Empirical Considerations

Allison L. Huang, Robert E. Chapman, and David T. Butry

Sponsored by:

National Institute of Standards and Technology
Building and Fire Research Laboratory

September 2009



U.S. DEPARTMENT OF COMMERCE

Dr. Gary Locke, Secretary

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY

Dr. Patrick D. Gallagher, Deputy Director

Abstract

Although the construction industry is a major component of the U.S. economy, it has experienced a “perceived” prolonged period of decline in productivity. Due to the critical lack of measurement methods, however, the magnitude of the productivity problem in the construction industry is largely unknown. The measurement problem is exacerbated by the fact that the construction industry is composed of four sectors that differ significantly in the outputs produced, firm size, and use of technology. The four sectors, which taken together define the construction industry, are residential, commercial/institutional, industrial, and infrastructure.

This report describes efforts underway that focus on the measurement of construction productivity at three levels—task, project, and industry—and how such measurements can be developed. This report analyzes the measurement challenges associated with the development of meaningful measures of construction productivity at the task, project, and industry levels and establishes a framework for addressing those challenges.

Specifically, this report identifies the metrics, tools, and data needed to move forward in collaboration with key construction industry stakeholders. Once produced, these metrics, tools and data will help construction industry stakeholders make more cost-effective investments in productivity enhancing technologies and improved life-cycle construction processes; they will also provide stakeholders with new measurement and evaluation capabilities. Finally, this report lays the foundation for future research and for establishing key industry collaborations that will enable more meaningful measures of construction productivity to be produced at the task, project, and industry levels.

Keywords

Building economics; construction; economic analysis; information technology; labor productivity; metrics; performance measurement; productivity

Preface

This study was conducted by the Office of Applied Economics in the Building and Fire Research Laboratory at the National Institute of Standards and Technology. This report analyzes the measurement challenges associated with the development of meaningful measures of construction productivity at the task, project, and industry levels and establishes a framework for addressing those challenges. The intended audience is the National Institute of Standards and Technology, as well as, other government agencies that compile and publish construction-related statistics, private sector organizations concerned about the perceived decline in construction productivity, and standards development organizations that produce standards used by the construction industry.

Disclaimer

Certain trade names and company products are mentioned in the text in order to adequately specify the technical procedures and equipment used. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products are necessarily the best available for the purpose.

Disclaimer Regarding Non-Metric Units

The policy of the National Institute of Standards and Technology is to use metric units in all of its published materials. Because this report is intended for the U.S. construction industry that uses U.S. customary units, it is more practical and less confusing to use U.S. customary units rather than metric units. Measurement values in this report are therefore stated in U.S. customary units first, followed by the corresponding values in metric units within parentheses.

Cover Photograph Credits

DigitalVision Construction in Action clip gallery image used in compliance with DigitalVision's royalty free digital stock photography use policy.

Acknowledgements

The authors wish to thank all those who contributed so many excellent ideas and suggestions for this report. They include Mr. Mark E. Palmer, Leader of the Computer-Integrated Building Processes Group in the Building and Fire Research Laboratory (BFRL) at the National Institute of Standards and Technology (NIST), for his technical guidance, suggestions, and support. Special appreciation is extended to Mr. Douglas Thomas of BFRL's Office of Applied Economics (OAE) for his thorough review and many insights and to Ms. Carmen Pardo for her assistance in preparing the manuscript for review and publication. Special appreciation is also extended to Professor Paul M. Teicholz, Professor Emeritus in Civil and Environmental Engineering at Stanford University, for the many stimulating exchanges that have truly improved this report. The report has also benefited from the review and technical comments provided by Dr. Nicos S. Martys and Ms. Geraldine Cheok of BFRL's Materials and Construction Research Division.

Table of Contents

Abstract	iii
Preface	v
Acknowledgements	vii
Acronyms and Abbreviations	xiii
Acronyms and Abbreviations	xiii
1 Introduction	1
1.1 Background	1
1.2 Purpose.....	3
1.3 Scope and Approach.....	3
2 Construction: An Engine for Economic Growth	7
2.1 Value of Construction Put in Place	7
2.2 Overview of the Construction Industry Supply Chain	11
2.3 Research and Innovation in the Construction Industry	18
3 Productivity and Competitiveness: A Survey of the Literature	21
3.1 Three Dimensions of Construction Productivity: Task, Project, and Industry ..	21
3.2 Factors Affecting Construction Productivity	22
3.2.1 Skilled Labor Availability.....	22
3.2.2 Technology Utilization	27
3.2.3 Offsite Fabrication and Modularization.....	29
3.2.4 Use of Industry Best Practices	30
3.3 Task Level Productivity Metrics	31
3.3.1 Task Level Productivity Measures.....	31
3.3.2 Task Level Productivity Estimates	33
3.4 Project Level Productivity Metrics	35
3.4.1 Project Level productivity Measures	35
3.4.2 Project Level Productivity Estimates.....	36
3.5 Industry Level Productivity Metrics	38
3.5.1 Industry Level Productivity Measures	38
3.5.2 Industry Level Productivity Estimates.....	39
3.6 Reconciling Industry-Level Productivity Estimates with Task-Level Productivity Estimates	41
3.6.1 Quality of Industry-Level Productivity Estimates	41
3.6.1.1 Appropriateness of Output Measure.....	41
3.6.1.2 Lack of Output Deflators.....	42
3.6.1.3 Quality of Input Measures	44
3.6.2 Changes in Output Mix.....	44
3.6.3 Task-Level Productivity Does Not Completely Reflect Industry-Level Productivity	47
3.6.4 Different Definitions of Productivity Measures and Different Definitions of the Construction Industry	48

3.7	Conclusions and Observations	49
4	Challenges and Opportunities in Construction Productivity Measurement Using National Statistics.....	51
4.1	Discussion of the Teicholz Diagram	51
4.3	Approaches for Measuring Construction Productivity.....	54
4.3.1	Focus on Building Types, Gross Output.....	55
4.3.2	Focus on Building Types, Value Added.....	57
4.3.3	Focus on Infrastructure Type, Value Added.....	59
4.3.4	Focus on Specialty Trades.....	59
4.3.5	Focus on Residential Building Construction	60
5	Summary and Recommendations for Future Research.....	63
5.1	Summary	63
5.2	Recommendations for Future Research	63
5.2.1	Standard Practice for Measuring Task-Level and Project-Level Productivity	63
5.2.2	Database of Project-Level Productivity Measures for Capital Facilities..	64
5.2.4	Industry-Level Productivity Metrics.....	65
	References.....	67
	Appendix A Metrics of Productivity.....	75
A.1	Industry-Level Productivity Measures.....	75
A.1.1	Labor Productivity	75
A.1.2	Multifactor Productivity.....	76
A.1.3	Value-Added Function, Choice of Output Measure, and the Role of Intermediate Inputs.....	78
A.1.4	Output Measures	80
A.1.4.1	Gross Output versus Value Added.....	80
A.1.4.2	Price Deflators.....	81
A.1.5	Labor Input.....	83
A.1.5.1	Hours (Production Workers).....	83
A.1.5.2	Conversion from Hours Paid to Hours Worked.....	83
A.1.5.3	Hours (Nonproduction Workers).....	84
A.1.5.4	Self-Employment.....	85
A.1.5.5	Labor Costs.....	86
A.1.5.6	Labor Quality	86
A.1.6	Capital Input.....	88
A.1.7	Intermediate Inputs	90
A.2	Aggregation Methods.....	92
A.3	Overview of BLS Productivity Program.....	93
A.4	Classification Issues	93
	Appendix B Productivity and Competitiveness: An Annotated Bibliography	97
B.1	Construction Productivity and Related Issues.....	97
B.2	Construction Data.....	108

B.3 Productivity Data and Measurement.....	109
Appendix C Sources of Construction Data Related to Productivity and Their Availability.....	117
C.1 Sources of Construction Data Related to Productivity.....	117
C.2 Classification Systems, Variables, and Coverage.....	125
Appendix D Glossary of Selected Terms	161

List of Figures

Figure 2.1 2008 Breakdown of \$1072 Billion Construction Market.....	11
Figure 2.2 Impacts of Construction Industry Supply Chain in 2008.....	12
Figure 2.3 Value of Construction Work and Value of Business done by Size of Establishment.....	17
Figure 3.1 Annual Average Number of Employed Persons in the Construction Industry by Age Groups, 2000 through 2008.....	23
Figure 3.2 Percentages of Employed Persons in the Construction Industry by Age Groups, 2008 through 2008.....	24
Figure 3.3 Percentage of Private Construction Workers with Union Membership and Percentage of Private Construction Workers under Collective Bargaining Agreement, 1973 through 2008.....	25
Figure 3.4 Construction Industry Production Worker Average Weekly Hourly Wage as a Percentage of Total Private Sector Production Worker Average Hourly Wage.....	26
Figure 3.5 Ratio of Total Installed Cost to Work Hour (Normalized to the Value of Year 2002).....	37
Figure 3.6 Changes in Construction Output Mix (1964-2007).....	46
Figure 3.7 Dollar Amount Single-Family House Construction as a percentage of Residential, Commercial, Industrial, and Institutional Construction (1964-2002).....	47
Figure 4.1 Construction Labor Productivity and Non-Farm Business Labor Productivity Comparison.....	51

List of Tables

Table 2.1 Value of Construction Put in Place in Millions of Constant 2008 Dollars.....	9
Table 2.2 Value of Construction Put in Place: Sector Totals and Sum Total in Millions of Constant 2008 Dollars.....	10
Table 2.3 Number of Establishments by Size of Establishment in the Construction Industry (2002).....	16
Table 2.4 Percentage and Cumulative Percentage of Value of Business Done by Size of Establishment.....	18
Table 3.1 Compounded Annual Rate of Change in Labor and Multifactor Productivity for Activities by Division from 1976 to 1998.....	34
Table 3.2 Changes in Equipment and Material Technology versus Changes in Labor Productivity.....	34
Table C.1 Standard Industry Classification (SIC).....	126

Table C.2 North American Industry Classification System (NAICS) 1997.....	127
Table C.3 North American Industry Classification System (NAICS) 2007.....	128
Table C.4 Current Employment Statistics Survey with SIC Classification (Bureau of Labor Statistics).....	130
Table C.5 Current Employment Statistics Survey with NAICS Classification (Bureau of Labor Statistics).....	132
Table C.6 GDP by Industry (Bureau of Economic Analysis).....	136
Table C.7 Producer Price Indices (Bureau of Labor Statistics).....	137
Table C.8 Producer Price Indices for Materials and Supply Inputs to Construction Industries (Bureau of Labor Statistics).....	138
Table C.9 National Income and Product Accounts Private Fixed Investment by Structure Type (Bureau of Economic Analysis).....	139
Table C.10 National Income and Product Accounts Investment in Government Fixed Assets (Bureau of Economic Analysis).....	141
Table C.11 Value of Construction Put in Place (C30 Data), Old Classification System (Census Bureau).....	144
Table C.12 Value of Construction Put in Place (C30 Data), New Classification System (Census Bureau).....	146
Table C.13 Industry Series of the 1992 Economic Census.....	148
Table C.14 Industry Series of the 1997 Economic Census.....	149
Table C.15 Industry Series of the 2002 Economic Census.....	151
Table C.16 Variables Reported by Economic Census.....	152
Table C.17 Source Data and Price Indices for BEA's Annual Estimates of Private Fixed Investment in Structures by Type.....	155
Table C.18 Web Links to Key Sources.....	159

Acronyms and Abbreviations

AAR	Additions, Alterations, and Reconstruction
BEA	Bureau of Economic Analysis
BFRL	Building and Fire Research Laboratory
BIM	Building Information Modeling
BLS	Bureau of Labor Statistics
CERF	Civil Engineering Research Foundation
CES	Current Employment Statistics
CII	Construction Industry Institute
CPS	Current Population Survey
CURT	Construction Users Roundtable
ECI	Employment Cost Index
GDP	Gross Domestic Product
GPS	Global Positioning Satellite
HWS	Hours-at-Work Survey
ICS	Industry Classification System
KLEMS	Capital, Labor, Energy, Materials, and Services
M&R	Maintenance and Repair
NAICS	North American Industry Classification System
NCS	National Compensation Survey
NIPA	National Income and Product Accounts
NIST	National Institute of Standards and Technology
NRC	National Research Council
NSF	National Science Foundation

OAE	Office of Applied Economics
OECD	Organization of Economic Co-Operation and Development
PPI	Producer Price Index
PPMOF	Prefabrication, preassembly, modularization, and offsite fabrication
R&D	Research and Development
RFID	Radio-Frequency Identification
SIC	Standard Industry Classification
SOC	Standard Occupational Classification
TFP	Total Factor Productivity

1 Introduction

1.1 Background

Although the construction industry is a major component of the U.S. economy, it has experienced a “perceived” prolonged period of decline in productivity. Due to the critical lack of measurement methods, however, the magnitude of the productivity problem in the construction industry is largely unknown. Construction productivity is a highly important topic. An analysis of articles published in the American Society of Civil Engineers’ Journal of Construction Engineering and Management (JCEM) during 1985-2002 indicates that productivity is a second highest ranked topic, in terms of number of articles.¹ The measurement problem is exacerbated by the fact that the construction industry is composed of four sectors that differ significantly in the outputs produced, firm size, and use of technology. The four sectors, which taken together define the construction industry, are residential, commercial/institutional, industrial, and infrastructure.

To address these challenges, the National Institute of Standards and Technology (NIST) requested that the National Research Council (NRC) appoint an *ad hoc* committee of experts to provide advice for advancing the competitiveness and productivity of the U.S. construction industry. The committee’s specific task was to plan and conduct a workshop to identify and prioritize technologies, processes, and deployment activities that have the greatest potential to advance significantly the productivity and competitiveness of the capital facilities sector of the U.S. construction industry over the next 20 years.²

To assist the committee in its planning for the workshop, NIST prepared a white paper³ describing efforts underway that focus on the measurement of construction productivity at three levels: task, project, and industry.⁴ The NIST white paper discussed how such

¹ Osama Abudayyeh, Amber Dibert-De Young, and Edward Jaselskis, “Analysis of Trends in Construction Research: 1985-2002,” Journal of Construction Engineering and Management May/June (2004): 433-439.

² The capital facilities sector includes commercial/institutional buildings (including high-rise and multifamily residential), industrial, and infrastructure projects. It does not include single-family and low-rise residential projects.

³ Robert E. Chapman and David T. Butry, *Measuring and Improving the Productivity of the U.S. Construction Industry: Issues, Challenges, and Opportunities*, NIST White Paper (Gaithersburg, MD: National Institute of Standards and Technology May 2008).

⁴ Tasks refer to specific construction activities such as concrete placement or structural steel erection. Projects are the collection of tasks required for the construction of a new facility (e.g., the construction of a new commercial office building, bridge, or power plant) or renovation (i.e., additions, alterations, and major replacements) of an existing constructed facility. Industry measures are based on the North

measurements can be developed, how they are related to the use of information and automation technologies and construction processes over the life of the project, and how to build on several ongoing collaborative efforts aimed at improving the efficiency, competitiveness, and innovation of the construction industry.

NIST briefed the NRC committee in July 2008 on the Building and Fire Research Laboratory's overall research program, its Measurement Science for Advanced Infrastructure Delivery goal that focuses on metrics and tools for construction productivity, and the contents of the white paper. Members of the NRC committee discussed on-going productivity-related research with NIST and asked for recommendations of researchers who might be willing to prepare white papers that would be presented as part of a major workshop planned for November 2008. As a result of NIST's input and input from other subject-matter experts, three white papers were commissioned. The three white papers were presented at the November workshop, which was attended by the NRC committee members, several key NIST staff, and approximately 50 additional experts. At the end of the workshop, the participants identified a range of activities that could improve construction productivity. From among these, the committee identified five that could lead to breakthrough improvements in construction efficiency and productivity in 2 to 10 years. These activities are highlighted in the NRC report which states "If implemented throughout the capital facilities sector, these activities could significantly advance construction efficiency and improve the quality, timeliness, cost-effectiveness, and sustainability of construction projects."⁵ The five activities, entitled "Opportunities for Breakthrough Improvements," are:

1. Widespread deployment and use of interoperable⁶ technology applications, also called Building Information Modeling (BIM);
2. Improved job-site efficiency through more effective interfacing of people, processes, materials, equipment, and information;
3. Greater use of prefabrication, preassembly, modularization, and off-site fabrication techniques and processes;

American Industrial Classification System (NAICS) codes for the construction sector and represent the total portfolio of projects.

⁵ National Research Council. *Advancing the Competitiveness and Efficiency of the U.S. Construction Industry*. (Washington, DC: National Academies Press, October 2009).

⁶ Interoperability is the ability to manage and communicate electronic data among owners, clients, contractors, and suppliers, and across a project's design, engineering, operations, project management, construction, financial, and legal units.

4. Innovative, widespread use of demonstration installations; and
5. Effective performance measurement to drive efficiency and support innovation.⁷

Although the focus of this report is on effective performance measurement (activity 5), it also provides limited coverage of activities 1 through 4. This is due in part to the treatment of those activities in the NIST white paper and the expansion of that treatment in various sections of this report.

1.2 Purpose

The purpose of this report is to analyze the measurement challenges associated with the development of meaningful measures of construction productivity at the task, project, and industry levels and establish a framework for addressing those challenges. Measuring construction productivity is challenging because on the one hand construction industry stakeholders, such as building owners and managers, want easy answers to complicated questions that are made available through task-level metrics, while, on the other hand, industry leaders, policy makers at the federal and state levels, construction industry researchers/academics, and industry specialists demand complicated data-intensive metrics to assess national and industry-wide trends and challenges facing this critical sector of the U.S. economy. To address these challenges, this report identifies the metrics, tools, and data needed to move forward in collaboration with key construction industry stakeholders. Once produced, these metrics, tools, and data will help construction industry stakeholders make more cost-effective investments in productivity enhancing technologies and improved life-cycle construction processes; they will also provide stakeholders with new measurement and evaluation capabilities.

1.3 Scope and Approach

This report contains four chapters and three appendices in addition to the Introduction. Chapters 2 through 4 are the core components of the report. These chapters lay the foundation for future research and for establishing key industry collaborations that will enable more meaningful measures of construction productivity to be produced at the task, project, and industry levels.

Chapter 2 provides a snapshot of the U.S. construction industry. As such, it provides the context within which the scope and size of the construction productivity measurement problem is defined. The chapter contains three sections. Section 2.1 presents information on the value of construction put in place to show the size of the construction industry and each of its four sectors. The four sectors, which taken together define the

⁷ National Research Council, *Advancing the Competitiveness and Efficiency of the U.S. Construction Industry*. *Op. cit.*

construction industry, are residential, commercial/institutional, industrial, and infrastructure. Section 2.2 uses information on the construction supply chain to highlight the critical importance of manufactured products (materials, components, and systems). Section 2.3 places special emphasis on the role of research and innovation in the construction industry.

Chapter 3 provides a survey of the literature on productivity and competitiveness. The chapter contains seven sections. Section 3.1 describes the three dimensions of construction productivity—task, project, and industry. Section 3.2 discusses the factors affecting construction productivity. Sections 3.3 through 3.5 describe existing productivity measures and present estimates of construction productivity measures at the task, project, and industry levels, respectively. Section 3.6 discusses the divergence between task-level and industry-level productivity estimates and presents possible explanations for the divergence. Section 3.7 synthesizes a number of conclusions and observations from the literature survey.

Chapter 4 analyzes the challenges and opportunities for using national statistics in construction productivity measurement. The chapter is divided into two sections. Section 4.1 discusses the widely-referenced productivity comparison diagram produced by Paul Teicholz. The discussion focuses on the productivity calculations by Teicholz with particular emphasis on the data challenges associated with construction productivity measurement. Section 4.2 examines the types of data that are available in national statistics and suggests ways in which they would enable the development of meaningful productivity measures for the construction industry.

Chapter 5 concludes with a summary and a discussion of topics for future research.

Appendix A presents a mathematically-oriented discussion of productivity metrics. The metrics described in Appendix A are largely based on the Bureau of Labor Statistics productivity methodology. Both single factor labor productivity and multifactor productivity methodologies are presented and discussed.

Appendix B presents an annotated bibliography on productivity and competitiveness. The annotated bibliography consists of three sections. Section B.1 focuses on documents with particular emphasis on productivity measurement and other related issues in the construction industry. Some of the measurement issues covered are deflators, quality adjustments of output, and the definition of what constitutes the construction industry. Section B.2 focuses on construction data-related documents. Section B.3 focuses on documents that treat the general topic of productivity methods and measurement.

Appendix C identifies sources of construction productivity data and discusses their availability. Appendix C contains two sections. Section C.1 provides a description of data sources that may be relevant to construction productivity measurement. Section C.2

describes classification systems, variables, and availability. The section concludes with a series of tables cross-referencing key sources of data and their availability.

Appendix D is a glossary of terms used in economics and construction.

2 Construction: An Engine for Economic Growth

Construction is an engine of growth for the U.S. economy. Investment in plant and facilities, in the form of construction activity, provides the basis for the production of products and the delivery of services. Investment in infrastructure promotes the smooth flow of goods and services and the movement of individuals. Investment in housing accommodates new households and allows existing households to expand or improve their housing. It is clear that construction activities affect nearly every aspect of the U.S. economy and that the industry is vital to the continued growth of the U.S. economy.

This chapter provides a snapshot of the U.S. construction industry. As such, it provides the context within which the scope and size of the construction productivity measurement problem is defined. The chapter contains three sections. Section 2.1 presents information on the value of construction put in place to show the size of the construction industry and each of its four sectors. The four sectors, which taken together define the construction industry, are residential, commercial/institutional, industrial, and infrastructure. Section 2.2 uses information on the construction supply chain to highlight the critical importance of manufactured products (materials, components, and systems). Section 2.3 places special emphasis on the role of research and innovation in the construction industry.

2.1 Value of Construction Put in Place

This section provides information on a key indicator of construction activity; the value of construction put in place. Data published by the U.S. Bureau of the Census are used to establish the composition of construction expenditures by type of construction/function (e.g., non-residential/office building). These expenditures are then assigned to four key construction industry sectors. The reference document used throughout this section is the Current Construction Reports series C30 publication Value of Construction Put in Place.

The data presented in the C30 report are summarized in Table 2.1. To facilitate comparisons between this report and the C30 report, Table 2.1 uses the same row and column headings as are used in the C30 report. Table 2.1 records annual values in millions of constant 2008 dollars for the years 2002 through 2008.⁸

⁸ Inflation reduces the purchasing power of the dollar over time; deflation increases it. When amounts are stated in actual prices as of the year in which they occur, they are said to be in *current dollars*. Current dollars are dollars of any one year's purchasing power, inclusive of inflation/deflation. That is, they reflect changes in purchasing power of the dollar from year to year. In contrast, *constant dollars* are dollars of uniform purchasing power, exclusive of inflation/deflation. Constant dollars indicate what the same good or service would cost at different times if there were no change in the general price level to change the purchasing power of the dollar. For additional information on conducting economic analyses using either constant dollars or current dollars, see Sieglinde K. Fuller and Stephen R. Petersen, *Life-Cycle Costing*

Reference to Table 2.1 reveals that total construction expenditures in real terms increased gradually from 2002 (\$1015 billion) to 2006 (\$1247 billion) and then declined in 2007 (\$1195 billion) and 2008 (\$1072 billion). Table 2.1 is organized to allow for in-depth analyses of the components/subcomponents of total construction expenditures. To facilitate such analyses, the data presented in Table 2.1 are initially divided into two parts: (1) private construction; and (2) public construction.

Private construction contains two major components—residential buildings and non-residential buildings—plus a number of subcomponents. Both the two major components and the subcomponents are shown as headings in the first column of Table 2.1.

The residential buildings component includes new private housing and improvements. New private housing includes new houses and town houses (single family) and apartments and condominiums (multifamily). The value of improvements put in place is a direct measure of the value of residential additions and alterations activities.

The non-residential buildings component includes manufacturing (industrial), office buildings, lodging, and commercial. Also falling under the non-residential buildings component are religious, educational, health care, and public safety.

Rounding out the private construction component are farm non-residential, public utilities, and “all other private.” These are generally of a non-residential nature, but are not part of non-residential buildings. Farm non-residential construction includes structures such as barns, storage houses, and fences. Land improvements such as leveling, terracing, ponds, and roads are also a part of this subcomponent. Privately owned public utilities construction is categorized by industry rather than function of the building or structure. This subcomponent includes expenditures made by utilities for telecommunications, railroads, petroleum pipelines, electric light and power, and natural gas. “All other private” includes privately owned streets and bridges, sewer and water facilities, airfields, and similar construction.

For public construction, there are two major components—residential and non-residential. Both the two major components and the various subcomponents are shown as headings in the first column of Table 2.1. The non-residential building component contains subcomponents similar to those for private construction, with educational buildings being the largest subcomponent. Expenditures for the non-building

Manual for the Federal Energy Management Program. NIST Handbook 135. (Gaithersburg, MD: National Institute of Standards and Technology, 1996).

subcomponents overwhelmingly consist of outlays for highways and streets, with sewer systems being a distant second subcomponent.

Table 2.1 Value of Construction Put in Place in Millions of Constant 2008 Dollars⁹

Type of Construction	Millions of Constant Dollars (2008)						
	2002	2003	2004	2005	2006	2007	2008
Total Construction	1,014,728	1,043,163	1,130,154	1,215,644	1,246,914	1,194,869	1,072,132
Total Private Construction	759,287	790,267	879,195	957,501	974,170	894,697	766,170
Residential	474,763	521,917	607,385	674,571	655,447	512,184	350,078
New Housing Units	357,651	404,502	475,856	530,052	500,665	367,740	229,934
New single family	318,214	363,412	430,329	477,911	444,273	316,902	185,776
New multi-family	39,437	41,090	45,527	52,141	56,392	50,839	44,158
Improvements	117,112	117,415	131,529	144,519	154,782	144,444	120,144
Nonresidential	284,524	268,351	271,811	282,930	318,723	382,513	416,092
Lodging	12,527	11,619	13,657	13,963	18,822	28,536	35,379
Office	42,242	35,781	37,475	41,094	48,785	55,881	57,084
Commercial	70,620	67,288	72,028	73,404	78,355	89,155	81,495
Health Care	26,854	28,337	29,944	31,414	34,192	36,954	39,101
Educational	15,689	15,708	14,476	14,098	14,780	17,332	18,585
Religious	9,975	10,015	9,293	8,505	8,266	7,811	7,097
Public Safety	260	216	329	450	447	618	650
Amusement and Recreation	8,950	9,105	9,611	8,276	9,960	10,584	10,316
Transportation	8,106	7,685	7,797	7,854	9,242	9,355	9,896
Communication	22,002	16,915	17,630	20,776	23,695	28,543	25,496
Power	39,025	39,338	31,184	28,998	33,282	49,184	68,702
Sewage and Waste Disposal	294	325	377	265	326	424	548
Water Supply	475	460	462	359	509	536	696
Manufacturing	27,220	25,080	26,975	32,947	37,471	47,042	60,784
Other	286	476	573	528	591	558	263
Total Public Construction	255,441	252,896	250,958	258,143	272,744	300,172	305,962
Residential	6,300	6,103	6,278	6,182	6,496	7,499	7,330
Nonresidential	249,141	246,792	244,681	251,961	266,248	292,674	298,632
Office	10,750	10,343	10,856	9,356	9,085	11,884	13,222
Commercial	4,203	4,709	4,402	4,033	3,572	3,974	3,447
Health care	5,626	5,982	6,738	6,543	6,895	8,493	8,598
Educational	72,709	71,251	70,152	73,751	75,921	83,142	85,496
Public safety	9,108	8,163	7,671	7,613	7,850	9,975	12,286
Amusement and recreation	11,790	10,608	9,418	8,520	10,367	11,442	11,172
Transportation	22,747	21,228	20,766	19,764	20,623	23,746	24,057
Power	5,022	9,163	9,158	10,099	9,174	12,398	11,457
Highway and street	68,636	66,667	66,442	70,323	76,431	79,176	81,592
Sewage and waste disposal	19,138	19,078	20,058	21,637	24,436	25,403	24,596
Water supply	14,415	14,159	13,922	15,106	15,467	15,869	16,255
Conservation and development	4,208	4,322	4,410	4,765	5,390	5,353	5,350
Other	790	1,120	688	450	1,037	1,818	1,104

Source: Census C30 Report. Individual entries may not sum to totals due to independent rounding.

To get the sector totals, each subcomponent was assigned to a sector and summed. The sector totals and the overall total are recorded in Table 2.2. Reference to the table reveals

⁹ Value of construction put in place is reported in current dollars by the Census Bureau. Constant 2008 dollars are obtained using consumer price indices.

that sector totals vary considerably, with residential normally being the largest and industrial the smallest.

Table 2.2 reveals that the commercial/institutional, industrial, and infrastructure sectors grew more or less consistently in real terms over the entire seven-year period. In real terms, expenditures in the commercial sector grew from \$301.8 billion in 2002 to \$384.4 billion in 2008, an increase of almost 30 %. Real expenditures for two of the four sectors, industrial and infrastructure, were essentially constant between 2002 and 2005 and then increased sharply between 2006 and 2008. Real expenditures for the industrial sector grew from \$27.4 billion in 2002 to \$61.3 billion in 2008, an increase of almost 125 %. Over the 2002 to 2008 period, real expenditures for infrastructure increased by slightly more than 30 %. Real expenditures for the residential sector exhibited a cyclical pattern that highlights the magnitude of the current housing crisis. Real expenditures for the residential sector first increased sharply, from \$481.1 billion in 2002 to \$680.8 billion in 2005, declined gradually in 2006 (to \$661.9 billion), and then fell precipitously in 2007 (to \$519.7 billion) and 2008 (to \$357.4 billion).

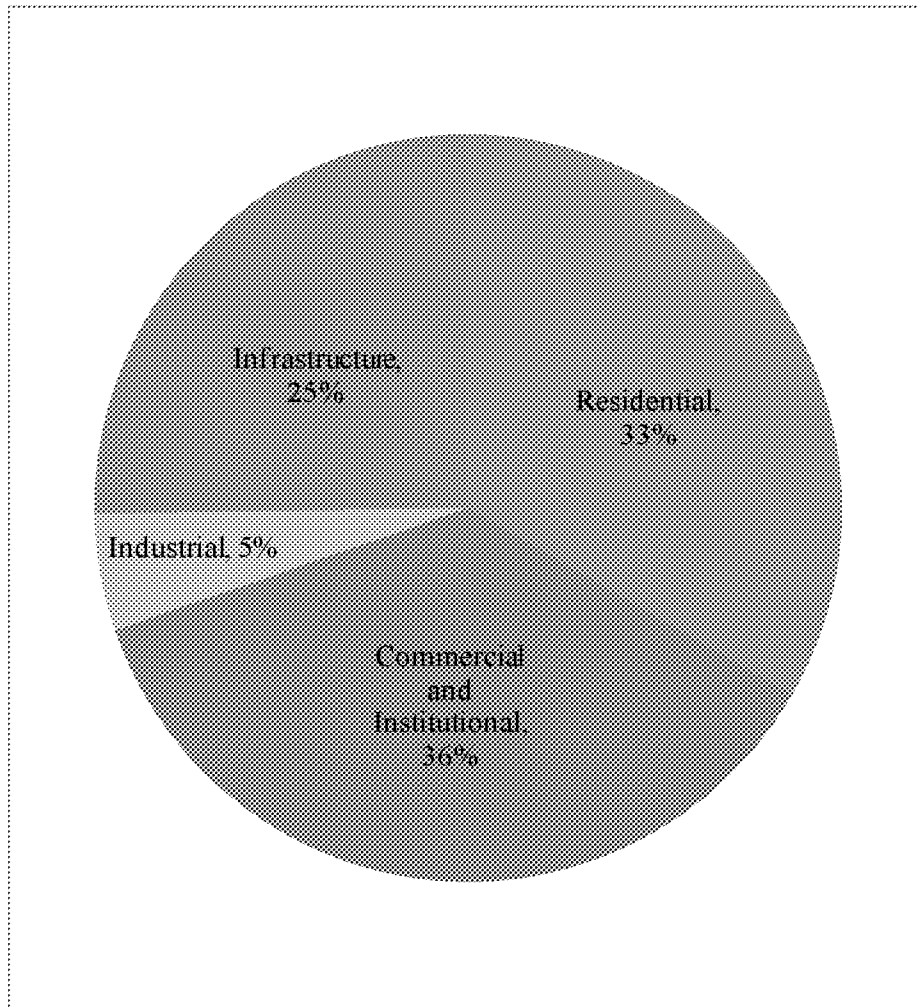
Table 2.2 Value of Construction Put in Place: Sector Totals and Sum Total in Millions of Constant 2008 Dollars

Type of Construction	Millions of Constant Dollars						
	2002	2003	2004	2005	2006	2007	2008
Residential	481,063	528,020	613,663	680,753	661,944	519,684	357,408
Commercial/Institutional	301,784	290,052	296,490	301,233	327,855	377,068	384,394
Manufacturing	27,438	25,167	27,136	33,117	37,913	47,475	61,269
Public Works	204,443	199,924	192,868	200,543	219,204	250,642	269,062
TOTAL	1,014,728	1,043,163	1,130,154	1,215,644	1,246,914	1,194,869	1,072,132

Source: Census C30 Report. Note that due to rounding the values entered in the “Total” row in Table 2.2, differ slightly from the values entered in the “Total Construction” row in Table 2.1.

The data contained in Table 2.2 provide the basis for calculating each sector’s relative share of total construction expenditures. Each sector’s relative share of total construction expenditures is shown graphically in pie chart form in Figure 2.1. It was constructed using 2008 data from Table 2.2. Figure 2.1 reveals that in 2008 the commercial sector accounted for 36 % of total construction expenditures, followed by the residential sector with 33 % of total construction expenditures. Over the longer term, the commercial/institutional sector’s relative share of total construction expenditures is usually exceeded by the residential sector, which normally constitutes about 45 % of the total. However, due to the current housing crisis, their relative shares were reversed. Historically, the commercial sector’s relative share tends to exceed the combined total for the industrial and infrastructure sectors.

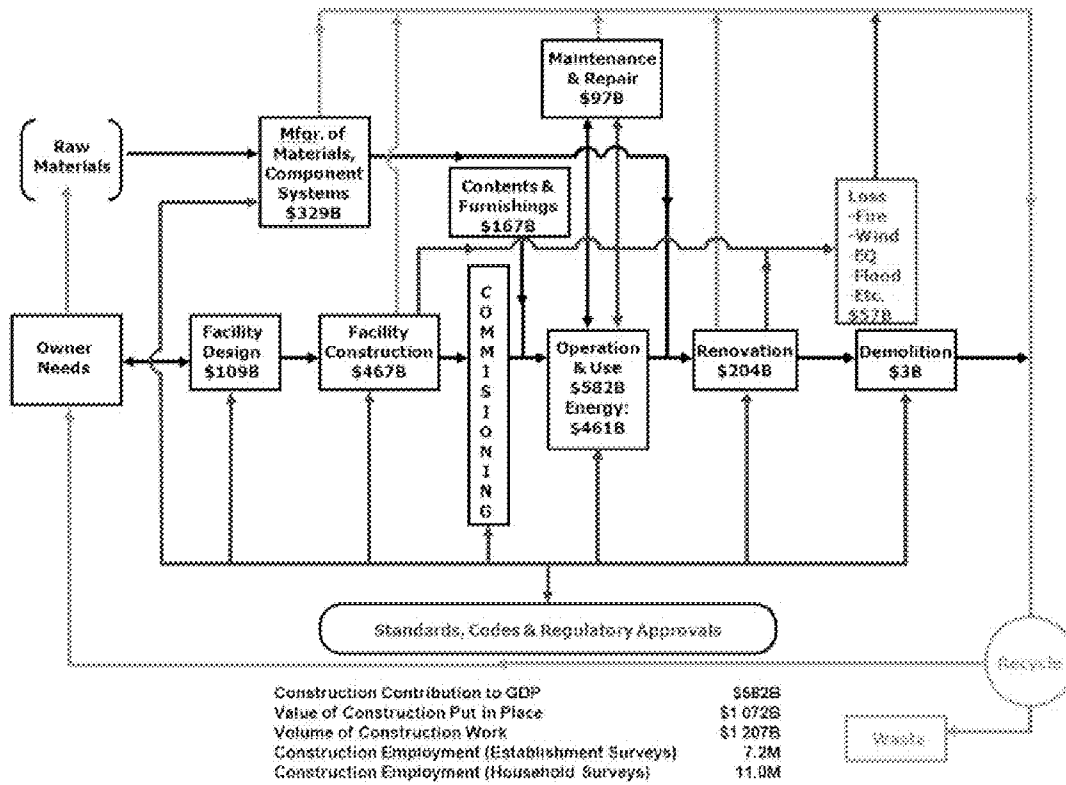
Figure 2.1 2008 Breakdown of \$1072 Billion Construction Market



2.2 Overview of the Construction Industry Supply Chain

A total industry supply chain for construction gives a more complete representation of construction work in the United States. Complete data is not gathered on an annual basis; however, there is sufficient data in the 1997 and 2002 Census of the Construction Industry reports to extrapolate construction data that is gathered on an annual basis. Using the Census Bureau's C30 annual figures for construction put in place along with Census data from 1997 and 2002, one can calculate values for five components of the construction industry: facility design; facility construction; renovation; maintenance and repair; and a value for materials, components, supplies, and fuels. Other components of the construction supply chain include contents and furnishings, operation and use, demolition, and losses. Each of these components is labeled in Figure 2.2, which records both the linkages between supply chain components and their estimated values.

Figure 2.2 Impacts of Construction Industry Supply Chain in 2008



In 2008, the construction industry's contribution to gross domestic product (GDP) was \$582 billion (see Figure 2.2), or 4.1 % of GDP.¹⁰ In 2008, the value of construction put in place was \$1072 billion (\$750 billion for new construction, \$323 billion for additions, alterations, and reconstruction (AAR)).¹¹ Table 2.2 reveals that the value of construction put in place declined by 6.8 % from 2007 to 2008. This decline was caused by a 34.3 % decline in new residential construction and a 13.6 % decline in residential renovations (see Table 2.1). The total of these two declines resulted in a -28.6 % change in the value of residential construction put in place. The remaining sectors of construction, commercial/institutional, industrial, and infrastructure, grew by 5.9 %, 34.0 %, and 11.5 % respectively. Overall, new construction declined by 9.4 % while renovations declined by 0.2 %.

Maintenance and repair activities are an integral part of the construction industry. Expenditures for maintenance and repair (M&R) amounted to \$134 billion in 2008.¹² Thus, the total volume of construction work in 2008—equal to the value of construction put in place plus expenditures for maintenance and repair—was \$1207 billion. It is important to note that expenditures for maintenance and repairs declined by 9.4 % from 2007 to 2008.

Approximately 30 % of the volume of construction work—\$329 billion—was due to the demand for manufactured products (materials, components, and systems).¹³ Note that expenditures for manufactured products are derived as percentages of expenditures for facility design services, new construction, AAR, and M&R. Thus, expenditures on manufactured products are tied to the volume of construction work done. Consequently, these expenditures decreased by 7.1 % from 2007 to 2008.

Figure 2.2 is organized so that expenditures are not double counted. Since expenditures for manufactured products (materials, components, and systems) are derived as percentages of expenditures for facility design services, new construction, AAR, and M&R, the values for the latter items are reduced by the appropriate percentage. Facility design services is also a derived

¹⁰ Bureau of Economic Analysis, "Gross-Domestic-Product-(GDP)-by-Industry Data." *Industry Economic Accounts* (Washington, DC: Bureau of Economic Analysis), http://www.bea.gov/bea/dn2/gdpbyind_data.htm (accessed July 2009).

¹¹ United States Census Bureau: Manufacturing and Construction Division, "Annual Value of Construction Put in Place." *Current Construction Report (CCR) C30* (Washington, DC: United States Census Bureau, July 2009), <http://www.census.gov/const/C30/total.pdf> (accessed July 2009).

¹² The value for maintenance and repair is calculated by using the ratio of maintenance and repair to new construction put in place from the 1997 census and multiplying it by the current value for new construction put in place.

¹³ The value of manufactured products, materials, components, and systems is calculated using ratios from the 2002 census. United States Census Bureau. "2002 Economic Census: Construction Subject Series." *Industry General Summary: 2002*. EC02-23SG-1 (Washington, DC: U.S. Census Bureau, October 2005).

calculation; it is derived based on data from the 2002 Census of the Construction Industry for architectural services, surveying services, and engineering services. The total thus derived for facility design services is allocated according to the percentage shares between the value of new construction and AAR put in place, also from the 2002 Census of the Construction Industry.

Four components recorded in Figure 2.2 are of particular importance in understanding how the double counting of expenditures is avoided; they are: (1) facility design; (2) facility construction; (3) renovation; and (4) maintenance and repair. The value of facility design recorded in Figure 2.2, \$109 billion, equals the sum of architectural services (\$32.0 billion), surveying services (\$5.4 billion), and engineering services (\$73.7 billion) for a total of \$111.2 billion¹⁴ less manufactured products associated with these services (\$2.2 billion). The value for facility construction in Figure 2.2, \$467 billion, equals the value of new construction put in place (\$749.7 billion) less new construction-related facility design services (\$79.7 billion) and new construction-related manufactured products (\$202.7 billion). The value for renovation recorded in Figure 2.2, \$204 billion, equals the value of AAR (\$323.4 billion) less AAR-related facility design services (\$31.6 billion) and AAR-related manufactured products (\$87.5 billion). The value for maintenance and repair recorded in Figure 2.2, \$97 billion, equals M&R expenditures (\$133.6 billion) less M&R-related manufactured products (\$36.4 billion). Thus, the value of manufactured products (materials, components, and systems) recorded in Figure 2.2, \$329 billion, equals the sum of manufactured products associated with: (1) facility design services (\$2.2 billion); (2) new construction (\$202.7 billion); (3) AAR (\$87.5 billion); and (4) M&R (\$36.4 billion).

The large value of manufactured products that appear in the construction industry supply chain is noteworthy because any productivity improvements associated with those products is not captured in productivity calculations for the construction industry. Construction activities often involve on-site assembly of manufactured products, which would be captured in part by construction productivity calculations. However, recent trends have emphasized the increased use of pre-assembly and off-site fabrication, particularly for many industrial applications.¹⁵ This trend poses a serious measurement challenge for the industry. Consequently, it is discussed in detail in Chapter 3.

¹⁴ The value of facility design services is allocated according to the percentage shares between the value of new construction and AAR put in place. Thus, \$79.7 billion is for new construction-related facility design services and \$31.6 billion is for AAR-related facility design services.

¹⁵ Construction Users Roundtable, "Pre-Assembly Perks: Discover Why Modularization Works," *The Voice*. (Fall 2007), pp 28-31.

Construction also has a major impact on U.S. employment. In 2008, 11.0 million persons were employed in the construction industry.¹⁶ This translates into 7.6 % of the total U.S. workforce. During the 2007 to 2008 period, the construction industry shed 882 000 jobs representing 7.4 % of all construction jobs, according to the Current Population Survey. This loss was the most severe among all industries in terms of percent lost and number of jobs lost. No other industry exceeded a loss of more than 3 % of employment or more than 400 000 jobs.

The composition of the construction workforce differs from much of the U.S. workforce due to the large number of self-employed workers (sole proprietorships and partnerships). Within the construction industry, there are 1.8 million self-employed workers. In contrast, manufacturing, which employs 15.9 million workers, has only 308 thousand self-employed workers.¹⁷ The large number of self-employed workers both reduces the size of the average firm and increases fragmentation within the construction industry. Table 2.3 shows number of establishments in construction industry by size of establishment.¹⁸ Nonemployers, which are businesses without paid employees that are subject to federal income tax, constitute about 2 million establishments and represents 74.46 % of all establishments in the construction industry. Establishments with 1 to 4 employees constitute another 15.17 % of all establishments. Nonemployers, together with establishments with 1 to 4 employees, represent nearly 90 % of all establishments. Figure 2.3 shows value of construction work and value of business done by size of establishment. Value of construction work is defined as receipts, billings, or sales for construction work. Value of business done is the sum of value of construction work and other business receipts.¹⁹ For nonemployers, only receipts data are available, and this variable is labeled “value of business done” in Figure 2.3. Figure 2.3 shows that value of construction work or value of business done is much more evenly distributed among size categories. Table 2.4 shows percentage and cumulative percentage of value of business done in each size category. Nonemployers and establishments with 1 to 4 employees each perform about 9 % of total value of business done. In other words, establishments with 5 or more employees, which constitute 10 % of all establishments, perform 82 % of total value of business done. The prevalence of self-employed workers and small-sized establishments complicates the adoption of new technologies and practices. Construction employment is affected by both the weather and the business cycle. Thus, year-to-year changes in employment can be substantial, resulting in layoffs and hiring surges. The cyclical nature of construction employment produces shortages in many highly-

¹⁶ United States Bureau of Labor Statistics, “Household Data: Employed Persons in Nonagricultural Industries by Sex and Class of Worker.” *Current Population Survey* (Washington, DC: Bureau of Labor Statistics), <http://www.bls.gov/cps/cpsaat16.pdf> (accessed July 2009).

¹⁷ *Ibid.*

¹⁸ Nonemployer Statistics.

¹⁹ 2002 Economic Census.

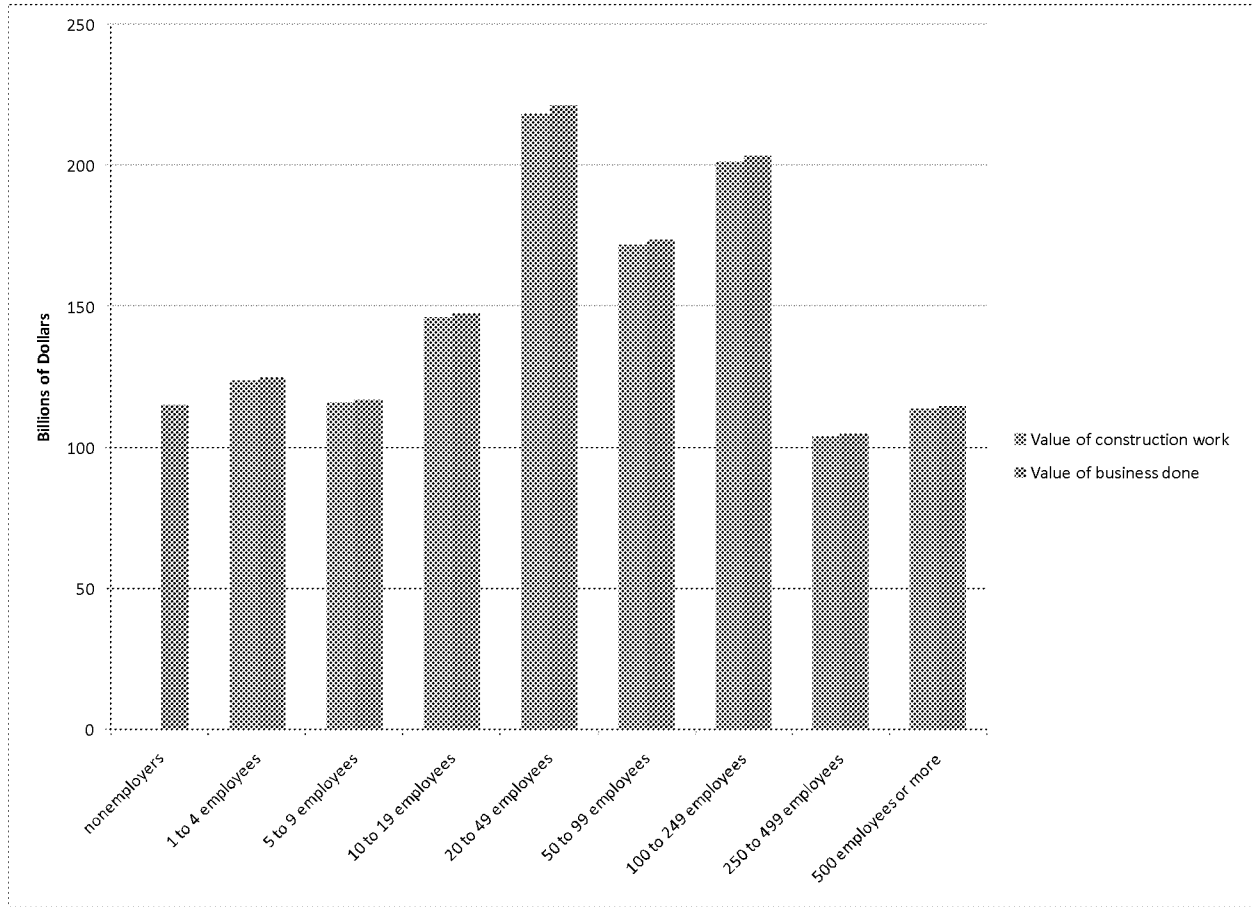
skilled trades. These shortages adversely impact productivity in the construction industry. Finally, declining construction productivity is exacerbated by the influx of unskilled labor from abroad, many of whom find their first employment opportunity in the construction industry.

Table 2.3 Number of Establishments by Size of Establishment in the Construction Industry (2002)

	Number of establishments	Percentage of total number of establishments
nonemployers	2 071 317	74.5 %
1 to 4 employees	421 959	15.2 %
5 to 9 employees	140 498	5.1 %
10 to 19 employees	78 917	2.8 %
20 to 49 employees	46 625	1.7 %
50 to 99 employees	13 649	0.5 %
100 to 249 employees	6640	0.2 %
250 to 499 employees	1434	0.05 %
500 employees or more	585	0.02 %

Source: 2002 Nonemployer Statistics and 2002 Economic Census.

Figure 2.3 Value of Construction Work and Value of Business done by Size of Establishment



Source: 2002 Nonemployer Statistics and 2002 Economic Census.

Table 2.4 Percentage and Cumulative Percentage of Value of Business Done by Size of Establishment

	Percentage of value of business done	Cumulative percentage of value of business done
nonemployers	8.7 %	8.7 %
1 to 4 employees	9.4 %	18.1 %
5 to 9 employees	8.9 %	27.0 %
10 to 19 employees	11.2 %	38.2 %
20 to 49 employees	16.7 %	54.9 %
50 to 99 employees	13.2 %	68.0 %
100 to 249 employees	15.4 %	83.4 %
250 to 499 employees	7.9 %	91.3 %
500 employees or more	8.7 %	100 %

Source: 2002 Nonemployer Statistics and 2002 Economic Census.

2.3 Research and Innovation in the Construction Industry

Given the demonstrated large impact of construction on the nation’s macroeconomic objectives, effective construction research becomes critical to the economy. Key drivers for change in construction research are sustainability; competition due to globalization and offshoring; homeland security and disaster resilience; infrastructure renewal; demand for better, faster, and less costly construction; and information technology.

The problem is that the U.S. construction industry invests little in research relative to its significant GDP contribution to the economy. A landmark study co-sponsored by the Civil Engineering Research Foundation (CERF) and the National Science Foundation (NSF) involved a nationwide survey of civil engineering-related research and development (R&D). The study, later published by CERF,²⁰ is especially noteworthy because it includes R&D associated with each of the key construction industry stakeholders. The CERF study reported that all key construction industry stakeholders combined invested in R&D at a rate that corresponds to only 0.5 % of the value of construction put in place. This translates into approximately \$5.4 billion in 2008. A recently published NSF study covering companies performing industrial R&D provides a useful contrast.²¹ Private sector R&D investments in manufacturing totaled nearly \$167 billion in 2007. Total R&D investments in construction were even surpassed by segments of the manufacturing industry (e.g., \$9.8 billion for machinery, a mature segment of the industry).

²⁰ Civil Engineering Research Foundation, *A Nationwide Survey of Civil Engineering-Related R&D*. CERF Report #93-5006 (Reston, VA: American Society of Civil Engineers, 1994).

²¹ <http://www.nsf.gov/statistics/infbrief/nsf079316/> (accessed July 2009).

Underinvesting reduces the potential for research-inspired innovations that contribute to substantial national benefits—namely constructed facilities that are more user and environmentally friendly, affordable, productive, and that are easier, faster, and more life-cycle cost effective to build, operate, and maintain. Given the impact of construction spending on the economy’s health, and that construction research helps make construction workers more productive and the construction industry more globally competitive and profitable, construction research becomes a critical variable in generating economic growth.

Although the generally accepted perception of the construction industry views innovation as a rare occurrence, in actuality it occurs consistently throughout the industry. Construction innovation offers the potential for significant company, industry, and societal benefits. As the demand rises for increasingly complex facilities, and the traditional sources of skilled labor shrinks, many construction firms are looking for design and technology innovations to improve their products and services and reduce their costs. Owners and clients seek construction innovations to increase the technical feasibility of their proposed projects and improve the performance of the completed facility.

Slaughter’s paper on “Models of Construction Innovation” is especially instructive.²² In that paper, five models of construction innovation are presented as a basis for construction firms to plan and carry out activities to effectively use specific construction innovations. These models are based upon established theories in management and economics but are modified to reflect the special conditions associated with constructed facilities, such as their scale, complexity, durability, and organizational contexts. For the purposes of project incorporation, the five categories of innovation are differentiated with respect to their degree of change from current practices and their links to other components and systems. Based on these models of innovation, firms can evaluate what they must do to implement the innovations. This framework can provide firms with a means through which to reduce the perceived risks of using construction innovations, and thereby somewhat lower the barriers to those innovations throughout the industry.²³

²² E. Sarah Slaughter, “Models of Construction Innovation.” *Journal of Construction Engineering and Management*, Vol. 124 (May/June 1998), pp. 226-231.

²³ *Ibid.*

3 Productivity and Competitiveness: A Survey of the Literature

This chapter provides a survey of the literature on productivity and competitiveness. The chapter contains seven sections. Section 3.1 describes the three dimension of construction productivity—task, project, and industry. Section 3.2 discusses the factors affecting construction productivity. Sections 3.3 through 3.5 describe existing productivity measures and present estimates of construction productivity measures at the task, project, and industry levels, respectively. Section 3.6 discusses the divergence between task-level and industry-level productivity estimates and presents possible explanations for the divergence. Section 3.7 synthesizes a number of conclusions and observations from the literature survey.

3.1 Three Dimensions of Construction Productivity: Task, Project, and Industry

The nature of the construction process points to a need for measures of construction productivity at three levels: (1) task; (2) project; and (3) industry. *Tasks* refer to specific construction activities such as concrete placement or structural steel erection. *Projects* are the collection of tasks required for the construction of a new facility (e.g., the construction of a new commercial office building) or renovation (i.e., additions, alterations, and major replacements) of an existing constructed facility. *Industry* measures are based on the North American Industrial Classification System (NAICS) codes for the construction sector and represent the total portfolio of projects.

Producing measures of construction productivity at each level involves the development of both metrics (i.e., the definition of the appropriate measure [parameter] that forms the basis for the calculation) and tools (i.e., the means through which construction industry stakeholders can perform the calculation for the selected metrics). Once produced, these metrics and tools will help construction industry stakeholders make more cost-effective investments in productivity enhancing technologies and life-cycle construction processes; they will also provide stakeholders with new measurement and evaluation capabilities (e.g., enabling them to simulate key elements of the project delivery process).

The basic concept underlying construction industry productivity measures is a comparison of the output of a task, project, or industry with the corresponding factors of production (inputs) required to generate that output.²⁴ The output and inputs of production thus constitute the basic components of every productivity measure. Typically, productivity measures are formulated as a ratio of output to one or more inputs. If only one of the inputs is used, then the ratio is a single factor productivity measure. A common example of this type of measure is output per labor hour. If all of the inputs are used, then the ratio is a multifactor productivity measure.

²⁴ Stephen F. Weber and Barbara C. Lippiatt, *Productivity Measurement for the Construction Industry*. NBS Technical Note 1172 (Gaithersburg, MD: National Bureau of Standards, February 1983).

3.2 Factors Affecting Construction Productivity

Much has been published about the factors that affect construction productivity. Although a comprehensive treatment is beyond the scope of this study, several key factors are usually cited in the literature. These factors are: (1) skilled labor availability; (2) technology utilization; (3) offsite fabrication and modularization; and (4) use of industry best practices.

3.2.1 Skilled Labor Availability

One of the greatest challenges facing the construction industry is its ability to attract and retain qualified workers. This is underscored by the fact that shortages of skilled workers continue to plague the construction industry.²⁵ A 1996 survey by the Business Roundtable, for example, found that over 60 % of its members who responded to the survey reported shortages of skilled labor on construction projects. Furthermore, 75 % indicated that the trend had worsened during the past five years.²⁶ Nearly 90 % of chemical and petrochemical companies have experienced difficulty in recruiting skilled craft workers.²⁷ Craft worker shortages appear to be the most severe for electricians, pipefitters, and welders. But the survey results also suggest labor shortages among all other types of craft workers.²⁸ Most respondents believe this skilled labor shortage is driven more by a shrinking skilled workforce, and less by increasing demand.²⁹ Many industry practitioners have suggested the shortage of skilled labor is a result of aging construction workforce, with fewer young people entering the industry. Figure 3.1 shows the annual average number of employed persons in the construction industry by age groups from 1994 through 2008. The median age is plotted against the secondary axis for 2000 through 2008 with a clear upward trend.³⁰ The median age has risen from 38.7 years old in 2000 to 40.3 years old in 2008. Figure 3.2 plots the same data using percentages. The decline of young workers (34 years old or younger) in proportion is evident. Part of the decline in 2008 may be due to the economic downturn, as inexperienced workers, who tend to be younger, tend to be laid off first. Since experienced workers tend to be more productive, as the proportion of experienced workers increases, productivity is likely to increase.

²⁵ Construction Industry Institute, *The Shortage of Skilled Craft Workers in the U.S.* RS 182-1 (Austin, TX: Construction Industry Institute, 2003).

²⁶ The Business Roundtable, *Confronting the Skilled Construction Work Force Shortage—A Blueprint for the Future* (The Business Roundtable, October, 1997).

²⁷ *Ibid.*

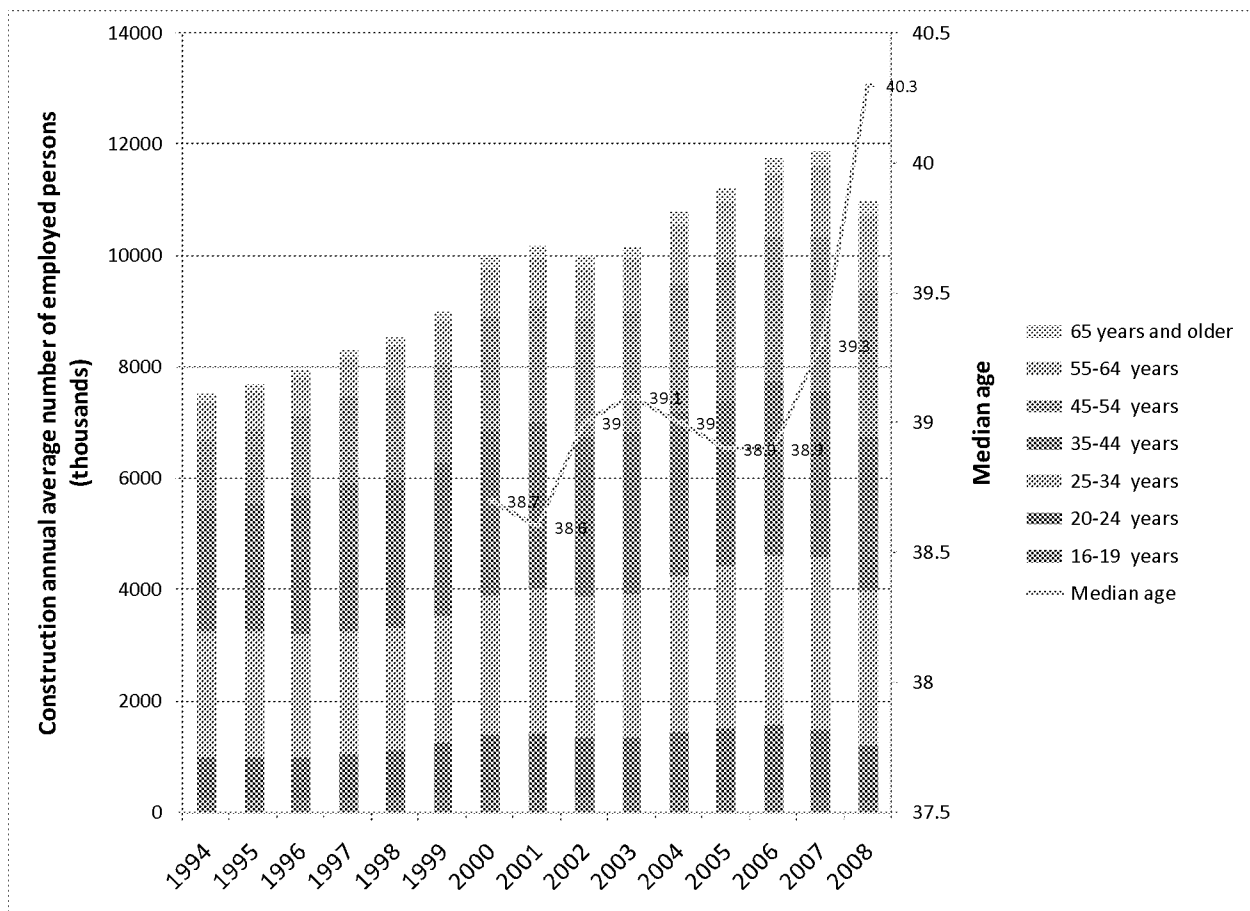
²⁸ *Ibid.*

²⁹ *Ibid.*

³⁰ Current Population Survey is the data source, and median age by industry is not available prior to 2000.

With an aging workforce, one concern for the viability of the construction industry is that skills and knowledge processed by experienced workers are not being passed onto younger generations because there is not enough new blood entering the industry. This challenge is compounded by the decline in training programs. Typically, training programs are funded by both owners and contractors through union and collective bargaining agreements. While open shop training programs exist, they tend to be rare.³¹ Figure 3.3 shows that the percentage of private construction workers that are union members and the percentage of private construction workers that are covered by collective bargaining agreements have declined since the 1970s. With the decline of union membership and collective bargaining agreements, training programs and the number of apprentices also have declined.

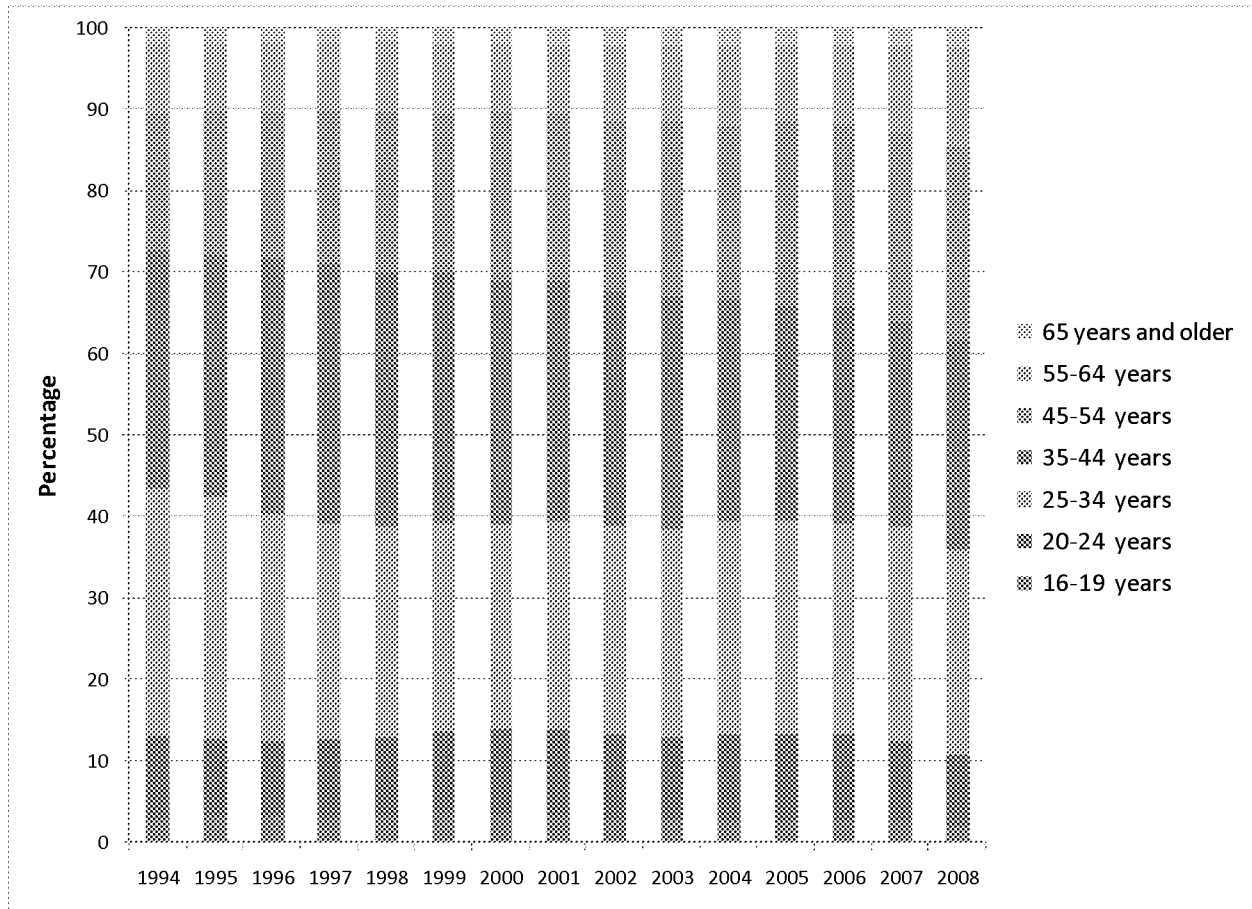
Figure 3.1 Annual Average Number of Employed Persons in the Construction Industry by Age Groups, 2000 through 2008³²



³¹ Construction Industry Institute, *Construction Industry Craft Training in the United States and Canada*. RS 231-1 (Austin, TX: Construction Industry Institute, 2007).

³² Current Population Survey. Median age data were not collected prior to 2000.

Figure 3.2 Percentages of Employed Persons in the Construction Industry by Age Groups, 2008 through 2008³³



Difficulty in staffing projects has resulted in increasing costs and schedule delays.³⁴ Skilled labor shortage might pose a greater challenge in years to come, as the Bureau of Labor Statistics has projected, prior to the current financial crisis, an annual 1 % increase of jobs in the construction sector by 2016, reaching a level of 8.5 million.³⁵ This increase in employment amounts to 10.2 % from 2006 to 2016. This projected growth in construction jobs is based on a projected output growth at a rate of 1.4 % per year to reach \$1.9 trillion by 2016.³⁶

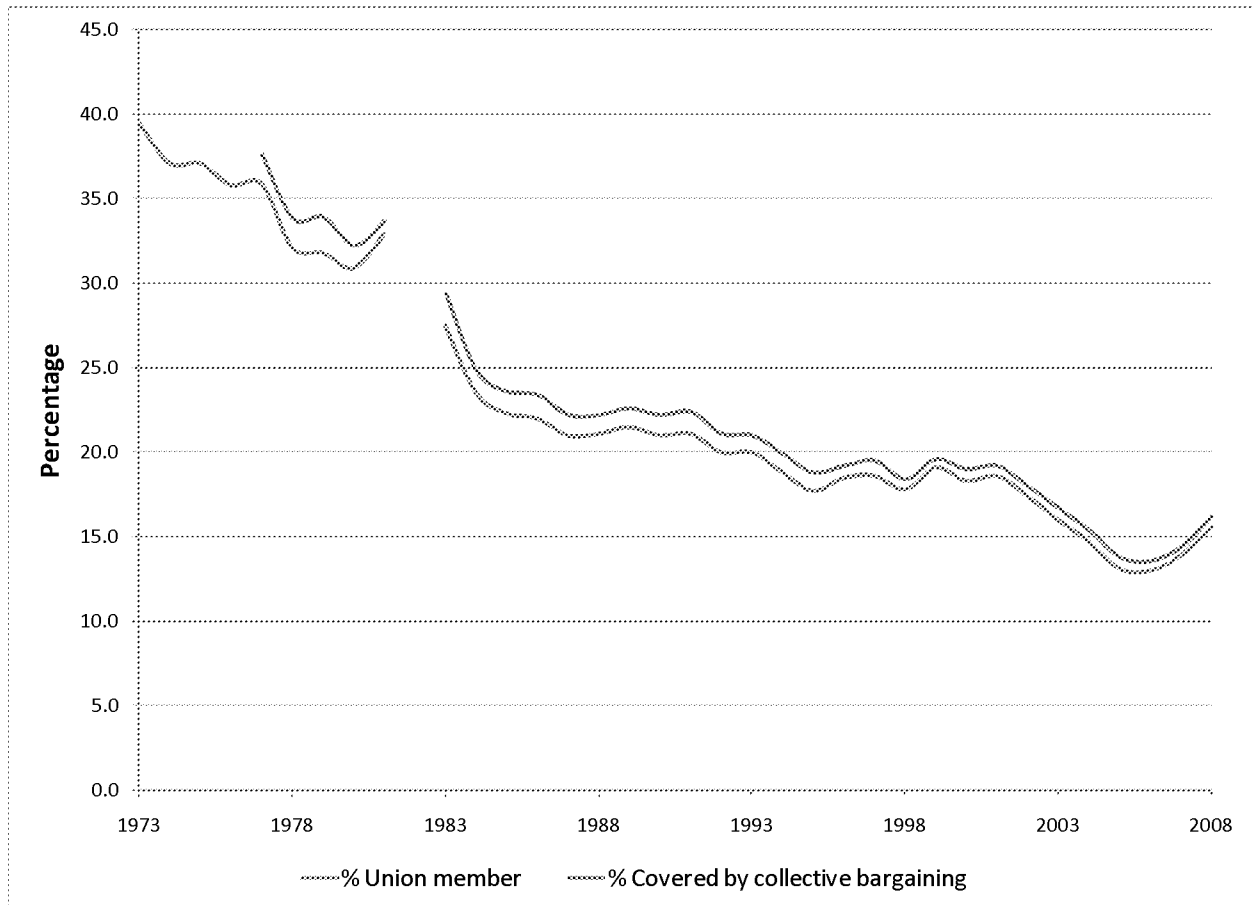
³³ Current population Survey.

³⁴ Construction Users Roundtable, *Confronting the Skilled Construction Workforce Shortage*. WP-401. (Cincinnati, OH: Construction Users Roundtable, June, 2004).

³⁵ Eric B. Figueroa and Rose A. Wood, “Industry Output and Employment Projections to 2016,” *Monthly Labor Review* November (2007): 53-85.

³⁶ *Ibid.*

Figure 3.3 Percentage of Private Construction Workers with Union Membership and Percentage of Private Construction Workers under Collective Bargaining Agreement, 1973 through 2008³⁷



Employers have attempted to identify the root causes and to develop strategies to overcome these shortages. Construction Industry Institute (CII) and others have funded research on the problem and generated potential solutions.³⁸ For instance, using the CII model plant, actual data from companies that had implemented training programs, and estimations of benefits from experts, CII estimated the return for each dollar invested in training to be between \$1.30 to \$3.00.³⁹ These benefits are in the form of increased productivity and reductions in turnover, absenteeism, and

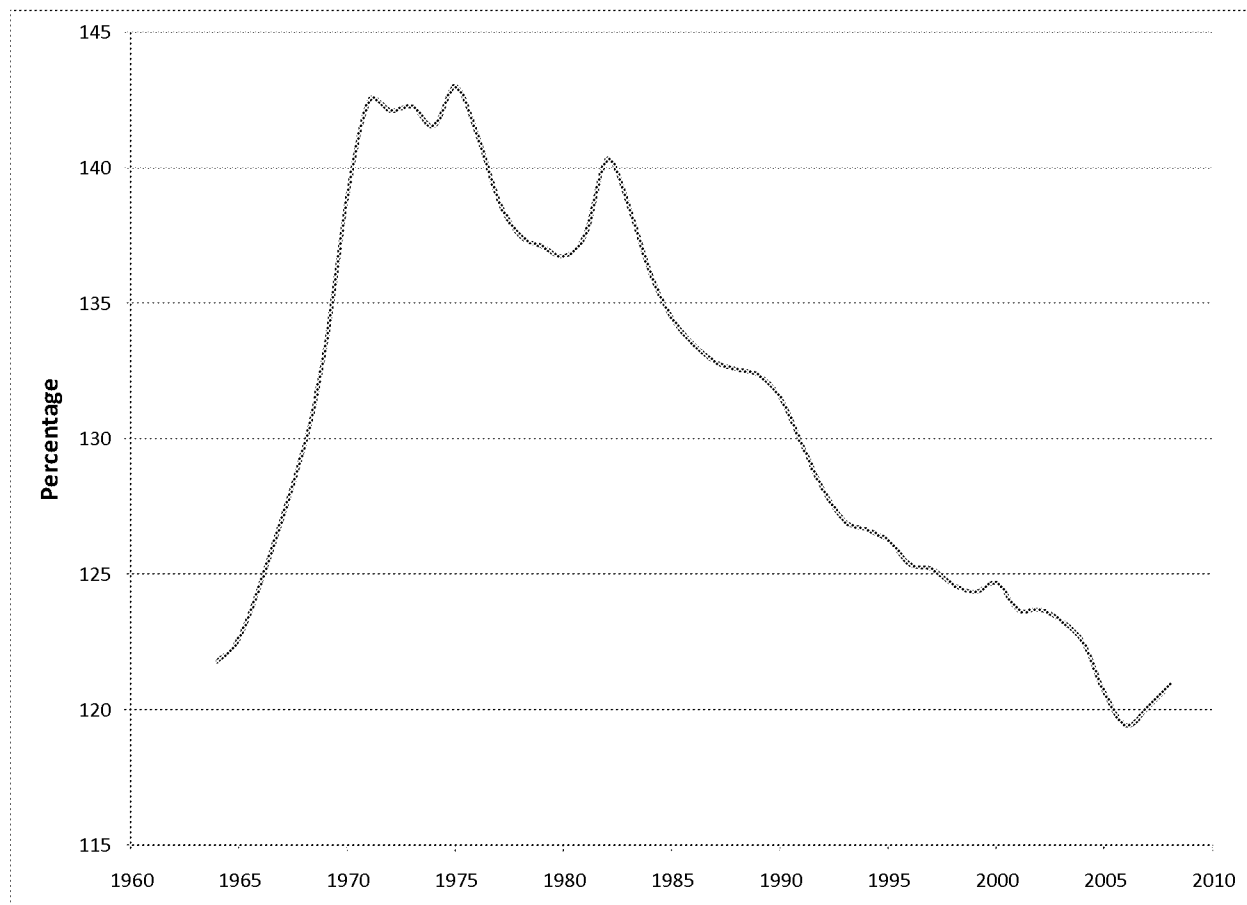
³⁷ Barry T. Hirsch and David A. Macpherson, "Union Membership and Coverage Database from the Current Population Survey: Note," *Industrial and Labor Relations Review*, 56(2003): 349-54.

³⁸ Construction Industry Institute, *The Shortage of Skilled Craft Workers in the U.S. Op. cit.*

³⁹ Construction Industry Institute, *Construction Industry Craft Training in the United States and Canada. Op. cit.*

rework.⁴⁰ Craft training benefits project financial performance by increasing the craft workers' average duration on a project and reducing turnover.⁴¹ Craft training also benefits individual workers by increasing their skills and knowledge, income, and job satisfaction. It is also essential for providing the skilled labor the industry needs. Despite this research and efforts to stem the problem, the construction industry's skilled worker pool continues to shrink. The decreasing number of young people entering the work force and the failure to recruit from non-traditional labor pools exacerbate this trend. Over the past 30 years, real wages of construction workers have declined relative to those of other workers (Figure 3.4). Poor industry image, tough working conditions, the industry's perceived poor safety record, and limited career development opportunities also have contributed to the decline in the number of people willing to enter and remain in the industry.

Figure 3.4 Construction Industry Production Worker Average Weekly Hourly Wage as a Percentage of Total Private Sector Production Worker Average Hourly Wage⁴²



⁴⁰ *Ibid.*

⁴¹ *Ibid.*

⁴² Source: Bureau of Labor Statistics.

3.2.2 Technology Utilization

Technology utilization impacts construction productivity in a number of ways. Historical changes in construction equipment have resulted in sustained improvements in task level labor productivity. Goodrum and Haas have shown, using commercially available cost estimation data, that these improvements stem from better control, amplification of human energy, increased functionality, better ergonomics, and better information processing and feedback.⁴³ Improved level of control refers to advances in machinery and hand tools with built-in capability to automatically adjust the level of power or other characteristics of the equipment. One example is a concrete vibrator that automatically adjusts the vibration frequency to match the concrete's slump. Better information processing and feedback refers to advances in heavy machinery that have the capability of performance monitoring and self-diagnosis systems. Overall, these technological advancements have enabled labor productivity to improve by 30 % to 45 %.⁴⁴ Goodrum *et al.* came to a similar conclusion regarding material characteristics that lead to reductions in unit weight and installation flexibility.⁴⁵ Reductions in unit weight enable ease of handling. Installation flexibility refers to the environmental conditions under which a material can be installed, such as temperature or moisture ranges. Comparing activities that experience such changes in materials with activities that did not, Goodrum *et al.* found labor productivity improved at least twice as much in activities with material improvements over the period of study (1977-2004).

Preliminary analyses of CII Benchmarking data covering information integration and automation technologies revealed significant task level productivity improvements.⁴⁶ Automation technologies focus on the degree to which individual work functions (e.g., supply management and project management) are automated. Integration technologies focus on the ability to exchange information between work functions and their associated databases (e.g., exchanges of information among supply management and project management functions). For the four trades examined—concrete, structural steel, electrical, and piping—labor productivity was about 30 % higher for projects with a high level of automation compared to projects with a low level of automation. The difference in labor productivity was about 45 % between projects with different levels of integration.

⁴³ Paul M. Goodrum and Carl T. Haas, "Long-Term Impact of Equipment Technology on Labor Productivity in the U.S. Construction Industry at the Activity Level," *Journal of Construction Engineering and Management* January/February (2004): 124-133.

⁴⁴ Construction Industry Institute, *Leveraging Technology to Improve Construction Productivity*. RS 240-1 (Austin, TX: Construction Industry Institute, October, 2008).

⁴⁵ Goodrum, Paul M., M. Yasin, and Z. Dong. "The Relationship Between Changes in Material Technology and Construction Productivity." Mimeo. (Lexington, KY: University of Kentucky).

⁴⁶ Construction Industry Institute, *Leveraging Technology to Improve Construction Productivity*. *Op. cit.*

A field test was conducted to examine how materials tracking and locating technologies can contribute to productivity. The use of Radio-frequency identification (RFID) tags and a Global Positioning Satellite (GPS) system were coupled to track materials in lay down areas in two CII member projects.⁴⁷ Improved materials tracking was shown to increase productivity at the workforce because material retrieval became efficient.

Previous paragraphs describe how technology can enhance productivity of individual tasks. Note that while technology can generally improve labor productivity, there is a cost associated with employing technology. Improvement in labor productivity is not an ultimate goal. For example, capital investment in technology can be increased to improve labor productivity, but this approach may not be the optimal solution when overall costs and benefits are considered.

Building Information Modeling is one technology that can enhance productivity of an entire project, from the planning phase to the decommissioning phase. Eastman *et al.* describe Building Information Modeling (BIM) as “a new approach to design, construction, and facility management in which a digital representation of the building process is used to facilitate the exchange and interoperability of information in digital format.”⁴⁸ A previous NIST study on interoperability estimated the cost of inadequate interoperability in the U.S. capital facilities industry to be \$15.6 billion per year,⁴⁹ and therefore enhanced interoperability has a great potential in efficiency gains. The National Research Council has identified the use of Building Information Modeling as a key activity that could lead to breakthrough improvements in construction productivity.⁵⁰

Concerns over the perceived decline in construction productivity have stimulated interest in ways to use technology and management practices to address this challenge. Current industry efforts aimed at the seamless flow of information in an interoperable design and construction environment seek to promote labor productivity both by enabling the project team to respond quickly and effectively to new requirements, changes in scope, site conditions, and delivery delays and by promoting the use of value adding processes and technologies. The CII Strategic Plan,⁵¹ the FIATECH Capital Projects Technology Roadmap,⁵² CURT’s efforts to address owner

⁴⁷ *Ibid.*

⁴⁸ Chuck Eastman, Paul Teicholz, Rafael Sacks, and Kathleen Liston, *BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers, and Contractors*, (Hoboken, New Jersey: John Wiley & Sons, Inc., 2008).

⁴⁹ National Institute of Standards and Technology, *Cost Analysis of Inadequate Interoperability in the U.S. Capital Facilities Industry*. NIST GCR-04-867. (Gathersburg, MD: National Institute of Standards and Technology, 2004).

⁵⁰ National Research Council, *Advancing the Competitiveness and Efficiency of the U.S. Construction Industry*, *Op. cit.*

⁵¹ Construction Industry Institute. *Strategic Plan* (Austin, TX: Construction Industry Institute, 2005).

issues associated with productivity improvement and cost reduction, the American Institute of Steel Construction CIS/2 protocol, the Hydraulics Institute's initiative on electronic data exchange, and ASTM's E 57 Committee are several noteworthy examples.

3.2.3 Offsite Fabrication and Modularization

Prefabrication,⁵³ preassembly,⁵⁴ modularization,⁵⁵ and offsite fabrication⁵⁶ (PPMOF) offer potential benefits in the increasingly competitive global marketplace. Owners want better facilities faster, at the lowest possible cost, and with increased safety. Both owners and contractors view PPMOF as a means to meet challenges of demanding schedules, adverse site conditions, and limited availability of skilled labor. Offsite fabrication and modularization can enable speedier delivery because offsite manufacturing of building components and onsite field preparation can proceed in parallel.⁵⁷ Costs can be reduced because moving part of the onsite construction work to a controlled environment offsite can reduce the impact of adverse site condition on the project and can enhance safety and productivity.⁵⁸ Additionally, offsite fabrication and modularization is a way to mitigate skilled labor shortage. Modularization has been used in the industrial sector for decades. As recent developments in modular construction have made this concept more versatile and applicable to the commercial sector,⁵⁹ increasing demand for modularization may emerge.

However, CII research shows that effective use of these methods requires careful consideration of their implications for engineering, transportation, coordination, and project organization.⁶⁰ To

⁵² FIATECH. Capital Projects Technology Roadmapping Initiative (Austin, TX: FIATECH, October 2004).

⁵³ Prefabrication: a manufacturing process, generally taking place at a specialized facility, in which various materials are joined to form a component part of a final installation.

⁵⁴ Preassembly: a process by which various materials, prefabricated components, and/or equipment are joined together at a remote location for subsequent installation as a sub-unit; generally focused on a system.

⁵⁵ Module: a major section of a plant/building resulting from a series of remote assembly operations and may include portions of many systems; usually the largest transportable unit or component of a facility.

⁵⁶ Offsite fabrication: the practice of preassembly or fabrication of components both off the site and onsite at a location other than at the final installation location.

⁵⁷ Construction Users Roundtable, "Pre-Assembly Perks: Discover Why Modularization Works." *Op. cit.*

⁵⁸ Charles M. Eastman and Rafael Sacks, "Relative Productivity in the AEC Industries in the United States for On-site and Off-site Activities," *Journal of Construction Engineering and Management* 134, no. 7 (2008): 517-526.

⁵⁹ Construction Users Roundtable, "Pre-Assembly Perks: Discover Why Modularization Works." *Op. cit.*

⁶⁰ Construction Industry Institute, *Prefabrication, Preassembly, Modularization, and Offsite Fabrication in Industrial Construction: A Framework for Decision-Making*. RS 171-1 (Austin, TX: Construction Industry Institute, 2002).

successfully incorporate offsite fabrication and modularization in projects, careful upfront planning and early decision making are essential.⁶¹ The use of PPMOF may also increase the level of details required in the design, it may increase the requirement for procurement logistics, and it may also limit the ability to inspect work in progress if the fabrication is done remotely.⁶²

Recent advances in design and information technologies, combined with increasing emphasis within the industry to address cost, schedule, and labor issues, have proven the use of PPMOF to be more viable than ever. In a recent Construction Users Roundtable (CURT) publication, CII Director Wayne Crew noted that the use of PPMOF has increased in the last 10 years, especially with new technologies such as building information modeling and internet design capabilities.⁶³ Future workforce shortages will likely encourage the use of PPMOF. PPMOF benefits such as reduced construction time, decreased costs, and increased safety have all contributed to its popularity, and while many companies in the oil and gas industries have used it for decades, others are realizing its full set of benefits. Widespread use of PPMOF has also been identified by the National Research Council as a key activity that could lead to breakthrough improvements in construction productivity.⁶⁴

3.2.4 Use of Industry Best Practices

Management practices affect productivity over the life cycle of a construction project in a number of ways, including planning, resource supply and control, and supply of information and feedback. Management practices that are inflexible or applied inappropriately can introduce inefficiencies that reduce productivity. A key opportunity for breakthrough improvement in productivity identified by the National Research Council is improved job-site efficiency through effective interfacing of people, processes, materials, equipment, and information.⁶⁵ To address issues associated with management of resources, organizations such as CII have developed a suite of best practices aimed at improving the project execution process.⁶⁶ These practices are directed at all phases of the project life cycle, from design, through procurement, fabrication,

⁶¹ Construction Users Roundtable, “Pre-Assembly Perks: Discover Why Modularization Works.” *Op. cit.*

⁶² Construction Industry Institute, *Prefabrication, Preassembly, Modularization, and Offsite Fabrication in Industrial Construction: A Framework for Decision-Making.* *Op. cit.*

⁶³ Construction Users Roundtable, “Pre-Assembly Perks: Discover Why Modularization Works.” *Op. cit.*

⁶⁴ National Research Council, *Advancing the Competitiveness and Efficiency of the U.S. Construction Industry.* *Op. cit.*

⁶⁵ *Ibid.*

⁶⁶ For a list of CII knowledge areas, practices, and information resources, see: http://www.construction-institute.org/source/Orders/CII_Matrix.cfm?section=orders&OrdersSection=Matrix

construction, commissioning, and operations and maintenance.⁶⁷ One example of a best practice is to incorporate maintainability as a project goal in the design process to enhance reliability and reduce total life-cycle costs.⁶⁸ Other examples of best practices include front-end planning, alignment during front-end planning, partnering, team building, project delivery and contract strategy, constructability, project risk assessment, change management, zero accident techniques, and planning for startup.⁶⁹ In-depth analyses of the value of best practices on cost and schedule control, as well as field rework have been performed.⁷⁰ Increasing use of best practices is associated with improved cost, schedule, and safety performance, for both owners and contractors. For owners, the potential cost benefits are estimated to be \$1.7 million to \$3.4 million, depending on industry group and project size. For contractors, the potential cost benefits can be \$7.2 million for the typical \$88 million heavy industrial project. Owners benefit most from schedule reductions, which can be as much as 16 % or 27 weeks for large projects. Finally, in terms of CII's zero accident best practice, the difference between a 4th quartile (lowest practice use) project to a 1st quartile (highest use) project amounts to potential savings of more than \$200 000 from lost workday cases avoided.⁷¹ Note also that the use of BIM can facilitate effective planning and management, which are the foundation for efficient processes that contribute to overall project success.

3.3 Task Level Productivity Metrics

3.3.1 Task Level Productivity Measures

Tasks refer to specific construction activities such as concrete placement or structural steel erection. Task-level metrics are widely used within the construction industry. Most task-level metrics are single factor measures and focus on labor productivity. For example, R.S. Means has published task level metrics for many years. Typical task-level metrics published by R.S. Means estimate how much a given output is produced by a designated crew in a normal 8-hour day.⁷²

⁶⁷ Construction Industry Institute, *Design for Maintainability: Improving Project Return on Investment*. RS142-1 (Austin, TX: Construction Industry Institute, 1999).

⁶⁸ *Ibid.*

⁶⁹ Youngcheol Kang, William O'Brien, Jiukun Dai, Stephen P. Mulva, Stephen R. Thomas, and Pin-Chao Liao, *Measuring Interoperability and Best Practices Impacts on Capital Project Productivity*. NIST GCR 09-925 (Gaithersburg, MD: National Institute of Standards and Technology, 2009).

⁷⁰ Construction Industry Institute, *Benchmarking and Metrics Value of Best Practices Report*. BMM 2003-4 (Austin, TX: Construction Industry Institute, 2003).

⁷¹ *Ibid.*

⁷² R.S. Means. *Building Construction Cost Data: 2009*. 67th Edition. (Kingston, MA: R.S. Means, 2008).

In this case, the denominator is the number of hours associated with a designated “crew day.” Thus, for a designated crew day, higher output is better. In this case higher output equates to higher task labor productivity. For some tasks, equipment may be involved, in such cases, R.S. Means provides estimates of output that is produced by a designated crew in an 8-hour day along with the equipment they use, and these measures can be considered multifactor.

The CII Benchmarking and Metrics Program uses a different metric to measure task labor productivity. CII fixes the output (e.g., cubic yards of concrete put in place) and measures the labor hours required to produce that output. In this case, the denominator is the fixed output and the numerator is the number of labor hours. Thus, for a given amount of output, lower labor hours is better. In this case, lower labor hours equates to higher task labor productivity.

Both the R.S. Means and the CII task labor productivity metrics include explicit measures of output and labor hours in the values reported. Such metrics are easy to understand and are widely used within the industry as a basic estimating tool. In addition to resorting to cost estimating guides, such as R.S. Means, some contractors collect output and labor hour information from their projects and these data become the basis for cost estimation for their future projects. To differentiate these metrics from alternative formulations, we use the term “raw metrics” to refer to these ratios of input and output. These metrics are raw in the sense that they include the units of measure and are based on unadjusted outputs and labor hours. For example, the relative prices for selected labor inputs and the given output may vary over time.

The CII Benchmarking and Metrics Program collects data on a project basis, where productivity is but one data element. The raw task level metrics produced by CII include not only the average productivity for that task—referred to as a baseline measure—but the full set of observed values. The observed raw task productivity values are then rank ordered into a distribution. Once this is done, the raw task productivity values can be assembled into quartiles. CII researchers can then examine the characteristics for a given task associated with projects in the best performing quartile and in the worst performing quartile.

A task productivity index is an alternative to the raw metrics discussed previously. An index is a dimensionless number, pegged to a reference data set, where the reference data set establishes the baseline value for one or more components of the index. An index can be a ratio of raw metrics. For example, the denominator could correspond to the baseline value for that task’s labor productivity (e.g., labor hours per cubic yard of concrete) and the numerator could be the value for a specific project. In that case, the computed value of the index shows how that project’s task productivity compares to the overall average of the reference data set. Alternatively, the numerator could correspond to an average value for a new data set of task productivity values collected at some future point in time. Thus, the index can be used to track how task productivity is changing over time.

An index can also incorporate additional information, such as the value of a deflator to help

control for changes in relative prices over time. Because the index is a dimensionless number, users can focus on the changes in the index value rather than the functional form of the metric underlying the index. If for example, the index value was pegged at 100.0 at time zero and higher values are better, then a future value of 102.5 indicates improvement in the amount of 2.5 %.

3.3.2 Task Level Productivity Estimates

Goodrum and Haas examined productivity measures for 200 construction activities over a 22-year period. The data sources were cost estimating guides.⁷³ They found that average activity productivity has increased. Table 3.1 lists compounded annual rate of change in labor and multifactor productivity for activities by division from 1976 to 1998. This table is reproduced from Goodrum *et al.* (2002). Labor productivity and multifactor productivity increased for all divisions. One exception is that labor productivity for electrical work has stayed the same. Furthermore, studies by Goodrum and Haas show that activities that experienced a significant change in equipment technology (i.e., hand tools and machinery) generally also witnessed substantially greater long-term productivity improvements. Activities that experienced a significant change in material technology in terms of modularization, reduction in unit weight, or installation flexibility, also experienced greater productivity improvements. These results are summarized in Table 3.2. The partial factor productivity used by Goodrum and Haas was defined as units of physical output divided by the sum of labor costs and fixed capital costs. These authors also conducted other related studies and reached similar conclusions.⁷⁴

⁷³ Paul M. Goodrum and Carl T. Haas, "Partial Factor Productivity and Equipment Technology Change at Activity Level in U.S. Construction Industry," *Journal of Construction Engineering and Management* 128 (2002): 463-472.

⁷⁴ Paul M. Goodrum, Carl T. Haas, and Robert W. Glover, "The Divergence in Aggregate and Activity Estimates of US Construction Productivity," *Construction Management and Economics* 20, no. 5 (2002): 415-423; Paul M. Goodrum and Carl T. Haas, "Long Term Impact of Equipment Technology on Labor Productivity in the U.S. Construction Industry at the Activity Level" *Op. cit.*; E. Allmon, C. T. Hass, J. D. Borcharding, and P. M. Goodrum, "U.S. Construction Labor Productivity Trends, 1970-1998," *Journal of Construction Engineering and Management* 126, no. 2 (2000): 97-104.

Table 3.1 Compounded Annual Rate of Change in Labor and Multifactor Productivity for Activities by Division from 1976 to 1998⁷⁵

Construction Division	Change in labor productivity 1976-1998 (compound annual rates)	Change in multifactor productivity 1976-1998 (compound annual rates)
Sitework	+2.8 %	+2.4 %
Doors and Windows	+1.6 %	+1.8 %
Metals	+1.5 %	+1.0 %
Finishes	+1.2 %	+1.6 %
Masonry	+1.2 %	+0.8 %
Concrete	+1.1 %	+1.4 %
Mechanical	+1.0 %	+1.4 %
Wood and Plastic	+0.3 %	+0.4 %
Moisture and Thermal Protection	+0.2 %	+0.6 %
Electrical	+0.0 %	+0.8 %

Table 3.2 Changes in Equipment and Material Technology versus Changes in Labor Productivity⁷⁶

Technology Characteristic		Change in Labor Productivity		
Equipment Technology Characteristic	No Change in Equipment Technology Characteristic	Change in Equipment Technology Characteristic	Δ	
Energy	3.6 %	39.8 %	36.2 %	
Control	14.9 %	16.6 %	31.7 %	
Functional Range	13.5 %	51.8 %	38.3 %	
Information Processing	21.0 %	56.4 %	35.4 %	
Material Technology Characteristic	No Change in Material Technology Characteristic	Change in Material Technology Characteristic	Δ	
Modularization	8.1 %	24.2 %	16.1 %	
Reduction in Unit Weight	10.4 %	48.6 %	38.2 %	
Installation Flexibility	8.7 %	23.1 %	14.4 %	

⁷⁵ Table is reproduced from Paul M. Goodrum, Carl T. Haas, and Robert W. Glover, “The Divergence in Aggregate and Activity Estimates of US Construction Productivity” *Op. cit.*

⁷⁶ Table is reproduced from Construction Industry Institute, *Leveraging Technology to Improve Construction Productivity*. Research Summary 240-1. October 2008.

3.4 Project Level Productivity Metrics

3.4.1 Project Level productivity Measures

Projects are the collection of tasks required for the construction of a new facility (e.g., the construction of a new commercial office building) or renovation (i.e., additions, alterations, and major replacements) of an existing constructed facility. Since a project is a collection of tasks, project level metrics are more complicated. The inputs and outputs for a given task, say concrete placement, differ from those of another task, say structural steel erection. Thus, it is not possible to aggregate the individual raw task productivity metrics into a project productivity metric unless adjustments are made.

One way to make these adjustments is to use a reference data set to calculate baseline values for each task. Information is still needed, however, to calculate a meaningful project level productivity metric. For instance, information yielding the task weight (share that it represents to the overall project) is required, as is an understanding of the task flows. Because some tasks are completed in parallel, while other in series, the composition of the task flows affects overall project productivity. Therefore, each component of the project productivity metric contains: (1) the task weight; (2) the raw task productivity baseline value in the denominator; (3) the raw task productivity value for that project in the numerator; and (4) a measure of the task mix (in parallel versus in series task flows). The project productivity index value is a function of the individual components.

The project level productivity metric just described is useful in measuring how an individual project compares to the overall average in the reference data set. In addition, data from all projects can be compiled into a distribution. Further analyses can then be conducted to identify characteristics associated with the best performing or worst performing projects.

A project level productivity index can also be used to track changes in project productivity over time. In this case, the reference data set corresponds to time zero. For each index component, the values for the task weights and the task baseline values appearing in the denominator are equal to values computed in the reference data set. The numerator in each index component then becomes the average value of the corresponding task productivity in the future data set. As noted earlier, an index can also include a deflator to adjust for changes in relative prices over time.

An alternative project level productivity index can be produced as follows. We can create an index which is the quotient of two ratios, in each ratio the numerator is the value of construction put in place and the denominator is the number of field work hours. As noted earlier, a reference data set can be used to fix a baseline value for the ratio of value put in place to field work hours. The baseline value for the ratio is then used as the denominator in the index calculation. How an individual project compares to the baseline is determined by inserting its ratio of value put in place to field work hours in the numerator of the index. Alternatively, this project level

productivity index can be used to track changes in productivity over time by following the process described in the previous paragraph.

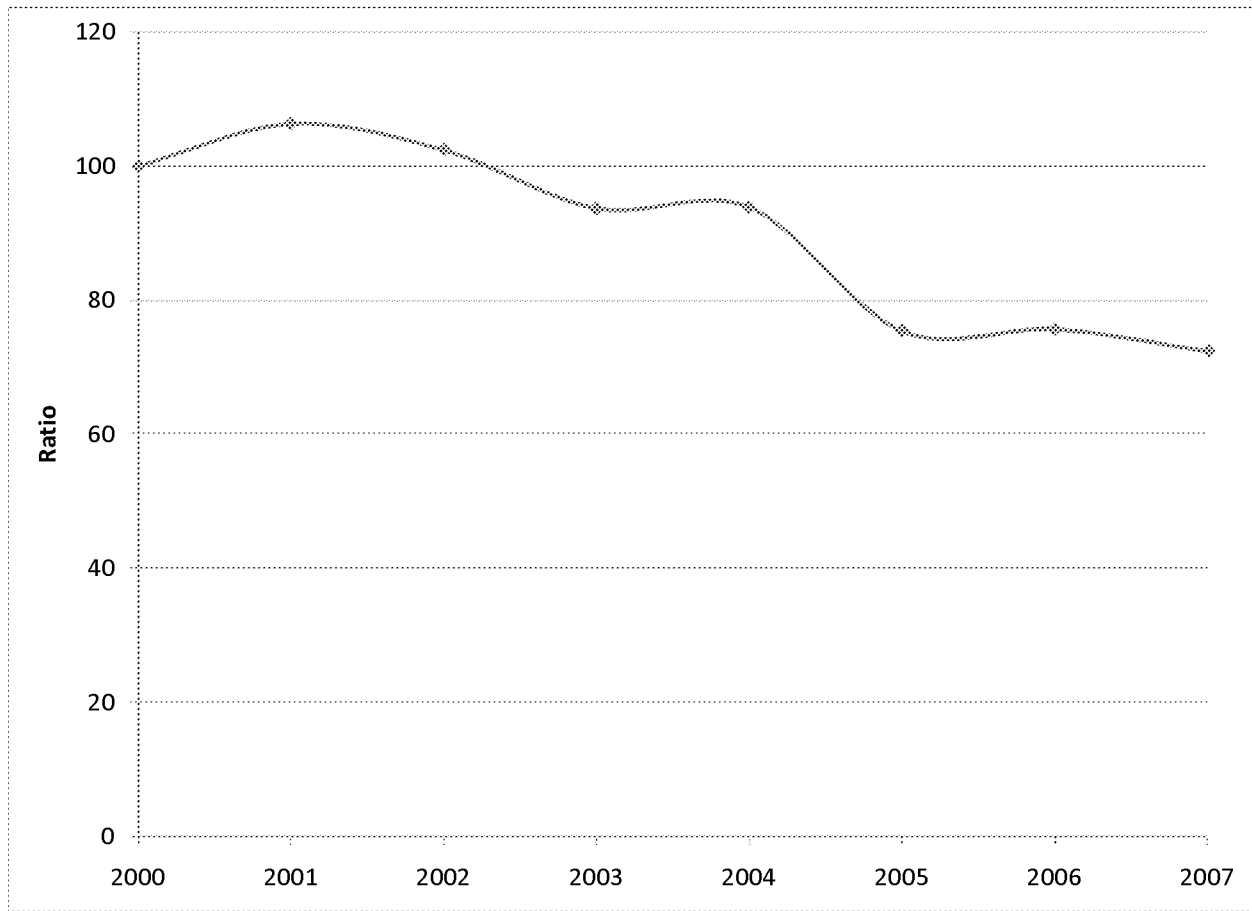
A related measure is cost per square footage data for a particular type of building. R.S. Means produces a square footage model that requires limited inputs, such as building type, exterior wall type, structural system, and square footage, and yields rough estimates for the overall cost of a project or its major components.

3.4.2 Project Level Productivity Estimates

Publicly available project-level productivity estimates tend to be rare. Construction firms collect data on project productivity for internal uses, such as cost estimation in bid preparation. As the information is pertinent to the competitiveness of the firms, it is not generally shared. The CII collects and compiles project-level data from its member organizations. The projects are predominantly industrial projects. The resultant dataset, Benchmarking and Metrics Productivity Database, is used to study project performance as influenced by factors such as technology and best practices. The CII studies are conducted such that information on individual projects remains confidential. Since the dataset contains projects of the member organizations, the dataset is considered to be representative of member organizations' projects, which tend to be more progressive in terms of project performance improvements, but not of the industry as a whole. Project level productivity measures can be calculated using this dataset. Figure 3.5 shows an index based on total installed cost per field work hour from 2000 through 2007.⁷⁷ The sample size varies from 16 in 2007 to 49 in 2004. The trend suggests a general decline over the seven-year span. However, it needs to be noted that changes in productivity may reflect changes in the composition of projects, in addition to changes in productivity. Each construction project is unique, and the mix of projects in each year is different. This is an intrinsic challenge in construction industry productivity analysis.

⁷⁷ Youngcheol Kang, William O'Brien, Jiukun Dai, Stephen P. Mulva, Stephen R. Thomas, and Pin-Chao Liao, *Measuring Interoperability and Best Practices Impacts on Capital Project Productivity*. *Op cit*.

Figure 3.5 Ratio of Total Installed Cost to Work Hour (Normalized to the Value of Year 2000)⁷⁸



⁷⁸ *Ibid.*

3.5 Industry Level Productivity Metrics

3.5.1 Industry Level Productivity Measures

At the industry level, productivity—the amount (or value) of output produced per unit of input—provides a measure of industrial efficiency. The Bureau of Labor Statistics (BLS) publishes two common measures of productivity: (single factor) labor productivity and multifactor productivity. Labor productivity is an output per hour measure. In the case of an industry producing multiple outputs, a Tornqvist index (weighted sum of the natural log of the ratio of output in different time periods) is used to chain multiple output indices together to form a single output measure (see Appendix A).

Increases in labor productivity may be due to increases in labor quality or labor efforts. However, it can also increase simply due to other factors such as technology or increased capital utilization, even when labor quality and worker efforts are held constant.

BLS measures multifactor productivity using output, labor, capital, and intermediate purchases input. A Tornqvist index is used to combine the inputs into a single measure of production. Multifactor productivity captures growth in output that is not explained by growth in these quantifiable inputs. In the growth accounting framework, multifactor productivity is calculated as a residual. Multifactor productivity growth can be attributed to factors such as management practices, best practices in the production process, etc. Because multifactor productivity is the part of output growth not explained by input growth, labor hours in multifactor productivity need to be quality adjusted. For instance, labor hours worked by workers with different skill levels need to be distinguished in multifactor productivity calculations. When an input quality increases, the input can be considered to have grown in quantity at the original quality level. In contrast, labor hours used in labor productivity calculations are simply the raw numbers of hours worked.

Multifactor productivity is often a preferred measure compared to labor productivity. This is because labor productivity measures are more prone to misinterpretation. Increases in labor productivity may reflect increases in the capital-labor ratio, rather than increases in labor quality and efforts. Additionally, a unit of production may achieve high levels of labor productivity, but the overall productivity may be compromised because the underlying capital-labor ratio may not be optimal. Similarly, low labor productivity might be efficient in the sense that low wages induce contractors to adopt more labor intensive practices and save on capital costs. Labor productivity measures are limited in the sense that they do not reveal a complete picture and are prone to misinterpretation. While labor productivity is often a less preferred measure of productivity compared to multifactor productivity, it is calculated with much more precision with

fewer assumptions.⁷⁹ Obviously, the data requirement for labor productivity calculation is also significantly less compared to multifactor productivity calculation.

3.5.2 Industry Level Productivity Estimates

There are no official productivity measures published by the BLS for the construction industry due to lack of suitable data. Productivity estimates of the construction industry do, however, exist in the literature. These estimates are produced by scholars in governmental agencies and academia. One highly referenced work is the productivity comparison diagram plotted by Teicholz.⁸⁰ In this diagram, constant contract dollars of new construction work per field work hour is shown to have trended downward over the past 40 years at an average compound rate of -0.6 % per year. In contrast, labor productivity of all non-farm industries (which includes the construction industry) has trended upward at an average compound rate of 1.8 % per year. Teicholz believes the reasons for the declines in labor productivity are due to lack of R&D spending, fragmentation within the industry, and declining real wage rates. He also notes that despite the fact that there has been a significant adoption of new information technology by the construction industry over the past 35 years, these applications tend not to be integrated with other systems and therefore do not permit improved collaboration by the project team.

Industry-level productivity estimates made by other scholars tend to show a similar trend. Allen, for instance, shows that construction productivity declined between 1968 and 1978 and argues that the biggest factor in the decline was the shift in the mix of output from large-scale commercial and industrial projects to residential construction and its associated lesser skill requirements.⁸¹ Stokes also argues that construction productivity declined between 1968 and 1978 and asserts that the major contributing factor to that decline was slower growth in capital per worker.⁸² This belief that construction productivity is declining is shared by industry observers such as the Business Roundtable.⁸³

Other scholars have analyzed productivity trends using more recent data. Harrison examined the period between 1961 through 2005 using data from the Bureau of Economic Analysis' National

⁷⁹ Jerome A. Mark, "Problems Encountered in Measuring Single- and Multifactor Productivity," *Monthly Labor Review* (1986): 3-11.

⁸⁰ Paul Teicholz, "Discussion of 'U.S. Construction Labor Productivity Trends, 1970-1998,'" *Journal of Construction Engineering and Management* 127 (2001): 427-428.

⁸¹ Steve G. Allen, "Why Construction Industry Productivity is Declining," *Review of Economics and Statistics* 67(1985): 661-669.

⁸² H. Kemble Stokes, Jr, "An Examination of the Productivity Decline in the Construction Industry," *Review of Economics and Statistics* 63 (1981): 495-502.

⁸³ The Business Roundtable, *CICE—The Next Five Years and Beyond* (New York, NY: The Business Roundtable, 1988).

Economic Accounts and Industry Economic Accounts.⁸⁴ He found the productivity growth was -2.43 % for 1961-1981, 0.13 % for 1981-1989, -1.18 % for 1989-2000, and -0.53 % for 2000-2005. Multifactor productivity calculated by Jorgenson *et al.* for the construction industry was -1.08 % for the period of 1977-2000.⁸⁵ Faruqui *et al.* examined productivity growth for selected business sectors between 1987 and 2000.⁸⁶ During the 1987 to 1996 period, construction experienced a slight increase in productivity, whereas between 1996 and 2000, construction experienced a sharp decline in productivity. Even during the 1987 to 1996 growth period, construction productivity improvements significantly lagged productivity improvements in manufacturing, services, and primary industries (i.e., agriculture, fishing, mining, and forestry). The general pattern of productivity decline is also found in other studies that used national statistics.⁸⁷

It should be noted that not everyone in the construction industry agrees that construction productivity is declining. For example, Young and Bernstein, in their McGraw-Hill SmartMarket Report, contend that the U.S. construction industry is making productivity improvements through innovation with new technologies, processes, and services.⁸⁸ Teicholz asserts, however, that a fragmented market with very small players makes application of these innovations less frequent than desired for a healthy increase in industry productivity. Another reason the Teicholz chart may show declining productivity is that it focuses on field work. For example, many of the improvements in construction productivity in the oil and gas industries over the past decade stem from the use of offsite fabrication facilities, where component production is well-controlled and highly-automated.⁸⁹ The debate about whether construction industry is declining, holding its own, or increasing cannot be easily resolved, because there are

⁸⁴ Centre for the Study of Living Standards, *Can Measurement Error Explain the Weakness of Productivity in the Canadian Construction Industry?* Research Report no. 2007-01 (Ontario: Centre for the Study of Living Standards, 2007).

⁸⁵ Dale W. Jorgenson, Mun S. Ho, and Kevin J. Stiroh, *Productivity Volume 3: Information Technology and the American Growth Resurgence* (Cambridge, Massachusetts and London, England: MIT Press, 2005).

⁸⁶ Umar Faruqui, Wulong Gu, Mustapha Kaci, Mirielle Laroche, and Jean-Pierre Maynard, "Differences in Productivity Growth: Canadian-U.S. Business Sectors, 1987-2000," *Monthly Labor Review* 126 April (2003): 16-29.

⁸⁷ Martin Neil Baily and Robert J. Gordon, "The Productivity Slowdown, Measurement Issues, and the Explosion of Computer Power," *Brookings Papers on Economic Activity* 1988 No. 2 (1988): 347-420; Wulong Gu and Mun S. Ho, "A Comparison of Industrial Productivity Growth in Canada and the United States," *American Economic Review* 90 (2000): 172-175; William Gullickson and Michael J. Harper, "Possible Measurement Bias in Aggregate Productivity Growth," *Monthly Labor Review* February (1999): 47-67.

⁸⁸ Norbert W. Young Jr. and Harvey M. Bernstein, "Key Trends in the Construction Industry—2006." *SmartMarket Report* (New York, NY: McGraw Hill Construction, July 2006).

⁸⁹ Construction Users Roundtable, "Pre-Assembly Perks: Discover Why Modularization Works." *Op. cit.*

no accurate industry-level measures of productivity for either the construction industry as a whole or its components (i.e., commercial, industrial, infrastructure, and residential).

3.6 Reconciling Industry-Level Productivity Estimates with Task-Level Productivity Estimates

The disparity between the conclusions of industry-level studies and task-level studies has been recognized.⁹⁰ There are many explanations for the observed difference in productivity trends.

3.6.1 Quality of Industry-Level Productivity Estimates

The national statistics offices in the U.S. collect tremendous amounts of data. Many elements of data required for productivity measurement exist. However, there are challenges associated with different classification systems and incomplete coverage. These challenges arise primarily because the existing data collection approaches are not designed specifically for productivity measurement.

Productivity measurement requires highly accurate measures of output, inputs, and deflators. This requirement applies to both labor productivity and multifactor productivity measurements. The requirement is particularly challenging for multifactor productivity measurement because multifactor productivity by definition requires more data and because multifactor productivity is the residual, the portion of growth in output not explained by growth of inputs. Since the data requirement is more limited for labor productivity, this discussion will focus mainly on labor productivity measurement in demonstrating the fundamental challenges in implementation.

3.6.1.1 Appropriateness of Output Measure

The appropriateness of the output measure is a major challenge in productivity measurement. In addition, Lawson *et al.* have noted the low quality of output data in the construction industry.⁹¹ The Economic Census is a major survey of industries. Because it is an establishment-based survey, it only surveys and reports data on establishments with payrolls, and a large number of workers in the construction industry are self-employed. Furthermore, because the Economic Census covers both general contractors and subcontractors, there is a significant amount of double counting. Double counting is a concern if the output measure is gross output, and it is not a concern if the output measure of interest is value added.

⁹⁰ Paul M. Goodrum, Carl T. Haas, and Robert W. Glover, “The Divergence in Aggregate and Activity Estimates of US Construction Productivity” *Op. cit.*

⁹¹ Ann M. Lawson, Brian C. Moyer, Sumiye Okubo, and Mark A. Planting, “Integrating Industry and National Economic Accounts, First Steps and Future Improvements,” In *A New Architecture for the U.S. National Accounts*, ed. Dale W. Jorgenson, J. Steven Landefeld, and William D. Nordhaus. (Chicago and London: University of Chicago Press, 2006).

Another source of output data is Value of Construction Put in Place from the C30 reports produced by the Census Bureau. Value Put in Place is collected at the project level. Compared to manufacturing, for which data are collected at the establishment level, there are more data collecting units for the construction industry. This is one reason for less accurate data for the construction industry.⁹² Another reason is the lack of annual data for benchmarking the value put in place data.⁹³ An additional complication associated with the use of value put in place as a data source for output measure is that the C30 reports document the total project costs, including architectural services, engineering services, construction services, and materials. These different types of costs are not distinguished in the reported summary statistics. Contract construction cost is reported separately from owner supplied materials and labor and architectural, engineering, and miscellaneous costs in the survey form. An output measure based on the contract construction cost or the total construction cost would be a gross output measure, with a boundary that approximates the construction industry.

3.6.1.2 Lack of Output Deflators

An important element in productivity measurement is the price deflators. Price deflators are needed to derive a quantity index of output. This is done by dividing the monetary value of construction in current dollars with an appropriate price deflator. Deflators are needed to strip away price changes due to inflation. In the case of construction, there is an additional challenge associated with the product not being uniform. Construction projects are heterogeneous even within a well-defined category, such as single family houses. Moreover, what is considered a typical new house in 1960 is very different from a typical new house in 2009. In addition to stripping away changes in prices due to inflation, there is also a need to define a uniform and time invariant “standard” house so that the output quantity index time series is meaningful. The nominal price of a new house in 2009 is higher than the price of a new house in 1960 for two reasons. The first reason is inflation. The second reason is that the house in 2009 is probably larger with more amenities. If the typical house in 1960 is chosen as the “standard house,” the quantity index is defined in terms of units of this “standard house.” The larger house with more amenities in 2009 is counted as more units of the “standard house,” while taking into account price increase due to inflation.

The construction industry has been known to be deficient in the area of price deflators. Two notable price deflators with long time series are associated with the residential sector. The Census Bureau publishes price indices for new one-family houses sold and for new one-family houses under construction using a hedonic regression model. The series are monthly from 1963 and from 1964, respectively. Using a similar approach, the Bureau of Economic Analysis

⁹² Eddy M. Rojas and Peerapong Aramvareekul, “Is Construction Labor Productivity Really Declining?” *Journal of Construction Engineering and Management* 129 (2003): 41-46.

⁹³ *Ibid.*

(BEA), in conjunction with the Census Bureau, has developed a price index for multifamily housing units.⁹⁴ This price index series extends back to 1978.

For many years, price deflators based on input data were used for nonresidential construction because appropriate deflators did not exist. While this practice of using input cost data still exists today, these input-based deflators are used to a lesser extent due to recent development of price deflators. Using a deflator based on inputs to deflate output biases multifactor productivity towards no change. In the growth accounting framework, multifactor productivity is the residual of output growth that is not explained by input growth. In other words, multifactor productivity is the ability to produce more output with the same inputs.⁹⁵ It is also the residual between output and input prices. It "represents the means by which a competitive position may be enhanced in the absence of input price reductions; the means by which the effects of input price increases may be mitigated; or the means by which payments to labor and to the owners of the capital may rise without increasing price."⁹⁶ If there is positive multifactor productivity growth, the prices of inputs should grow faster than the prices of outputs. In other words, when input cost data are used to deflate output, it is implicitly assumed that the relationship between inputs and outputs stays constant, which translates to constant productivity.⁹⁷ The use of input costs to deflate output has been cited as a reason for downward bias in productivity.⁹⁸ Note however, using a cost index to deflate output can affect labor productivity in different ways. That is, using a cost index to deflate output could bias output upward or downward.

Goodrum and Haas point out that a possible source of underestimation of output comes from decreases in real wage in construction, which is one component of value of construction put in place.⁹⁹ A properly constructed output price index takes into account decreases in real wage. To create an accurate output measure, a high quality price index is fundamental.

⁹⁴ Frank de Leeuw, "A Price Index for New Multifamily Housing," *Survey of Current Business* Feb (1993) 73(2): 33-42.

⁹⁵ Jorgenson, Dale W., "Productivity and Economic Growth," in *Fifty Years of Economic Measurement—the Jubilee of the Conference on Research in Income and Wealth*, ed. Ernst R. Berndt and Jack E. Triplett (Chicago and London: University of Chicago Press, 1990).

⁹⁶ William Gullickson, "Measurement of Productivity Growth in U.S. Manufacturing," *Monthly Labor Review*. July (1995): 13-35.

⁹⁷ Paul Pieper, "The Measurement of Construction Prices: Retrospect and Prospect." In *Fifty Years of Economic Measurement: The Jubilee of the Conference on Research in Income and Wealth*, Volume 54, ed. Ernst R. Berndt and Jack E. Triplett (Chicago: University of Chicago Press, 1990).

⁹⁸ Edwin R. Dean, "The Accuracy of the BLS Productivity Measures," *Monthly Labor Review* February (1999): 24-34.

⁹⁹ Paul M. Goodrum and Carl T. Haas, "Closure to 'U.S. Construction Labor Productivity Trends, 1970-1998,'" *Journal of Construction Engineering and Management* 127 (2001): 427-429.

The BLS has recently developed producer price indices for the nonresidential sector of the construction industry. These new producer price indices cover four types of new building construction and four types of specialty trades. These newly available producer price indices have been incorporated by the BEA in its estimates of investments in private structures.¹⁰⁰ More details on the BLS producer price indices are discussed in Appendix A.

3.6.1.3 Quality of Input Measures

The main source of labor hours data is the Current Employment Statistics (CES). CES reports total number of employees, number of production workers, and average weekly hours of production workers by NAICS code. A challenge for the construction industry is that the CES is an establishment survey, and the self-employed and unpaid family workers are not within the scope of the survey. Based on Current Population Survey data, about 15 % to 19 % of total work hours and total workforce are attributable to the self-employed and unpaid family workers. At the Bureau of Labor Statistics, the CPS is used to supplement the CES, for data on proprietors and unpaid family workers. One limitation of using the CPS to obtain information on the self-employed is the sample size.¹⁰¹ An implication is that the number of self-employed, the total number of workers, and the data work hours may not be accurate, particularly at the industry level or a sub-industry level. In CPS, the construction industry is not further categorized at a finer level. Coding of industries and reporting are more accurate in establishment level surveys compared to household surveys. For this reason, data from the CES are used as a primary source of data, and data from the CPS are used as a supplemental source of data in the BLS productivity programs.¹⁰² For materials flows, which are needed for multifactor productivity measurement, it has been noted that although the input-output framework tracks materials flows, the data outside of manufacturing tends to be incomplete.

3.6.2 Changes in Output Mix

Rojas and Aramvareekul point out that productivity changes can simply be due to changes in output mix.¹⁰³ Residential and commercial building construction is labor intensive, compared to industrial and heavy construction, which tends to be capital intensive.¹⁰⁴ They contend labor

¹⁰⁰ Paul R. Lally, "How BEA Accounts for Investment in Private Structures," *Survey on Current Business* February (2009): 9-15.

¹⁰¹ Edwin R. Dean, "The Accuracy of the BLS Productivity Measures," *Op. cit.*

¹⁰² Lucy P. Eldridge, Marilyn E. Manser, and Phyllis Flohr Otto, "Alternative Measures of Supervisory Employee Hours and Productivity Growth," *Monthly Labor Review* April (2004): 9-28.

¹⁰³ Eddy M. Rojas and Peerapong Aramvareekul, "Is Construction Labor Productivity Really Declining?" *Op. cit.*

¹⁰⁴ Based on 2002 Economic Census, capital labor ratio for construction of buildings was 0.06, and it was 0.15 for heavy and civil engineering construction.

productivity is lower for residential and commercial building construction than industrial and heavy construction. From 1964 through 2007, the output mix changed from 64 % residential and commercial and 36 % industrial and heavy construction to 76 % residential and commercial and 24 % industrial and heavy construction.¹⁰⁵ Figure 3.6 shows the general upward trend of residential and commercial construction as a fraction of total construction and the general downward trend of industrial and heavy construction as a percent of total construction. Following Rojas and Aramvareekul's argument, changing the output mix intrinsically translates into decline in measured "labor productivity."¹⁰⁶ This decline in labor productivity due to change in output mix is a result of shifting labor and capital usage, and it does not necessarily indicate lower labor quality or effort. Allen also argues that labor intensity associated with single-family house construction is higher and that the decline in construction productivity between 1968 and 1978 was partially due to the shift in output mix from large scale commercial, industrial, and institutional projects to single-family houses.¹⁰⁷ Figure 3.7 shows dollar amount of single-family house construction as a percent of residential, commercial, industrial, and institutional construction for the period of 1964 and 2002.¹⁰⁸ A relative increase in single-family house construction between 1968 and 1978 is evident. Between 1964 and 2002, we also observe a general upward trend. Estimates of industry-level labor productivity of the construction industry tend to show a declining trend. This decline could be partially explained by changes in the output mix.

Rojas and Aramvareekul also point out that the increase in labor productivity in manufacturing may be partially due to changes in output mix. If the changes in output mix are taken into account, the increase in labor productivity is smaller. This example illustrates the importance of accounting for all inputs in productivity measures. It also indicates the importance of focusing on homogenous products. When productivity is calculated for homogenous building or infrastructure types, the influence of changes in output mix is taken away.

¹⁰⁵ The raw data on which these values are based are originated from Census Bureau's Value of Construction Put in Place. The same dataset is used to generate Figure 3.6.

¹⁰⁶ Eddy M. Rojas and Peerapong Aramvareekul, "Is Construction Labor Productivity Really Declining?" *Op. cit.*

¹⁰⁷ Steve G. Allen, "Why Construction Industry Productivity is Declining," *Op. cit.*

¹⁰⁸ The data source is Census Bureau's Value of Construction Put in Place.

Figure 3.6 Changes in Construction Output Mix (1964-2007)

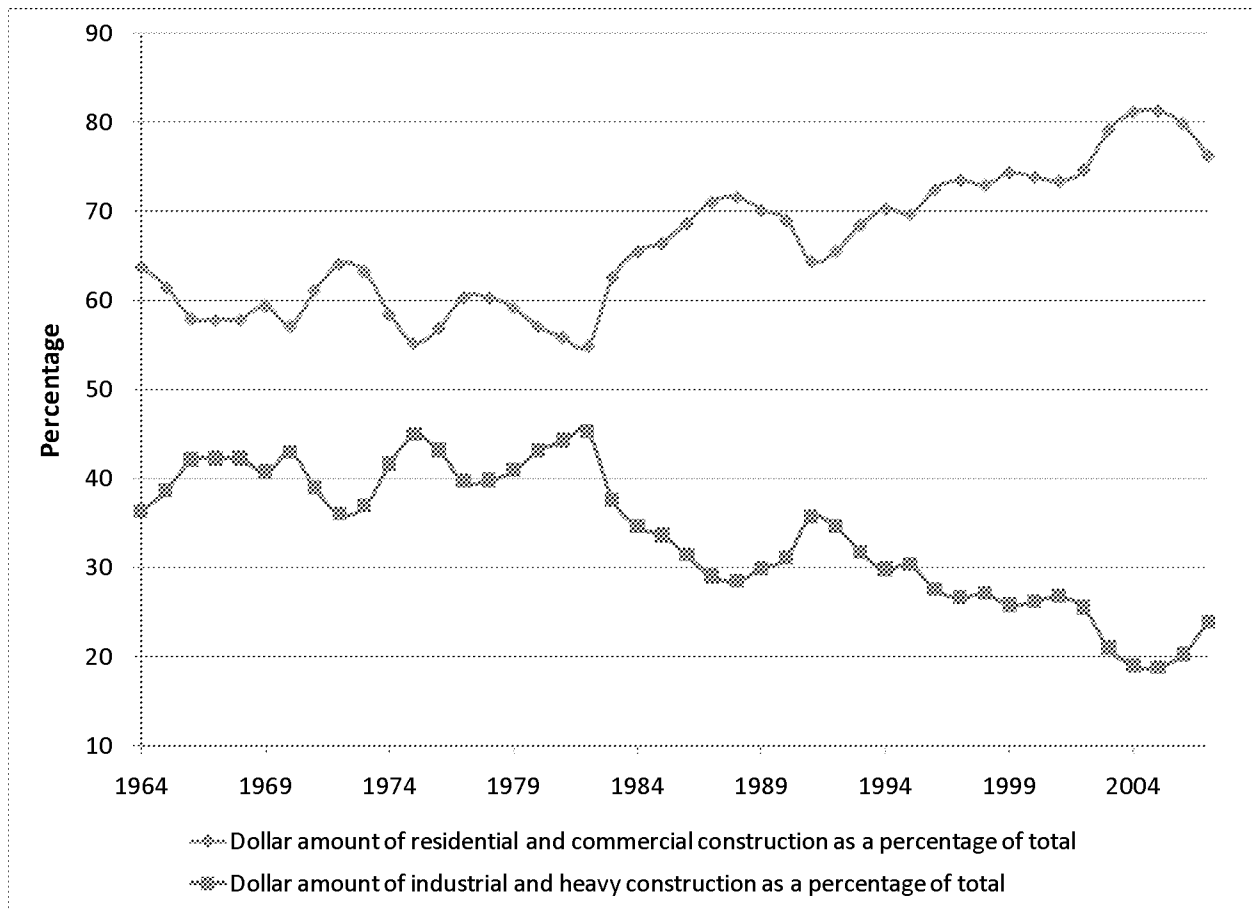
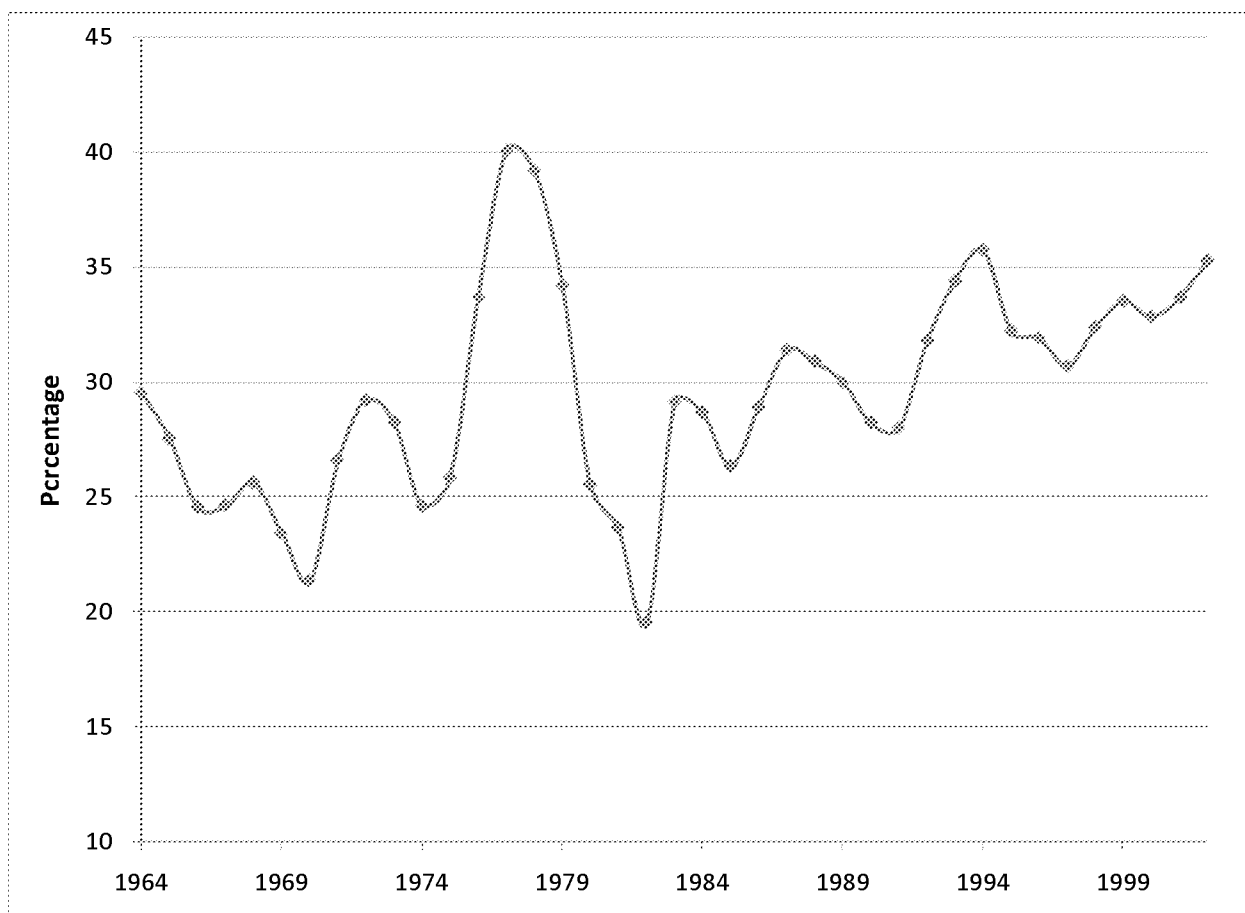


Figure 3.7 Dollar Amount Single-Family House Construction as a percentage of Residential, Commercial, Industrial, and Institutional Construction (1964-2002)



3.6.3 Task-Level Productivity Does Not Completely Reflect Industry-Level Productivity

Construction projects tend to be unique and are increasingly more complex. Task-level productivity does not capture project-level uniqueness and complexity. The trend of increasing project complexity could partly explain productivity decline at the industry level. High productivity at the task level also does not necessarily translate into high productivity at the project level. A project level success depends on managerial coordination and planning, which task-level productivity does not capture. For instance, idle time is not included in task-level productivity measurement, but it certainly can impede progress and productivity at the project level. Regulation is sometimes cited as one reason for low productivity in the construction industry. Regulation does not generally apply to task-level productivity, but it does affect project-level and therefore industry-level productivity.

3.6.4 Different Definitions of Productivity Measures and Different Definitions of the Construction Industry

When comparing productivity estimates, it is helpful to keep in mind the different definitions of productivity. In terms of labor productivity, two different output measures can be used. Labor productivity is often defined as output per hour. The output measure can be gross output or value added. The choice of the output measure is also related to the definition and scope of the construction industry. Industry practitioners tend to define labor productivity in concepts similar to the gross output based labor productivity.¹⁰⁹ In contrast, statistical offices, such as the BLS, tend to use value-added labor productivity measures. One difference between the labor productivity measures based on these two output measures can be seen by looking at prefabrication. Eastman and Sacks, for instance, have studied a number of similar on-site and off-site activities.¹¹⁰ They have observed that off-site activities tend to have higher productivity than their on-site counterparts. These authors therefore argue that construction productivity is underestimated and that the production of prefabricated materials ought to be included in the construction productivity measurement. Eastman and Sacks' concept of construction labor productivity involves the gross output measure. Offsite activities, such as prefabrication, are productivity enhancing. If they are incorporated in a construction project, then labor productivity based on gross output is expected to improve. On the other hand, labor productivity based on value added is not expected to change with the use of prefabrication.¹¹¹ This difference in the definitions of output and the implied scope of the construction industry can lead to different estimates and may be one reason for the divergent perceptions of industry productivity trends.

The most commonly calculated labor productivity measure is defined to be value added per hour. This definition, for instance, is used by the BLS. In task-level productivity studies, the definitions usually vary. For instance, Goodrum *et al.* defined task-level labor productivity in terms of physical units of output per hour.¹¹² This measure was shown to have increased during 1976-1988. Differences in estimates may be partly due to different definitions.

The rate of change in multifactor productivity in the growth accounting framework is defined to be the rate of change in output minus the weighted rates of change in capital, labor, and

¹⁰⁹ Centre for the Study of Living Standards, *Can Measurement Error Explain the Weakness of Productivity in the Canadian Construction Industry?* *Op. cit.*

¹¹⁰ Charles M. Eastman and Rafael Sacks, "Relative Productivity in the AEC Industries in the United States for On-site and Off-site Activities" *Op. cit.*

¹¹¹ Centre for the Study of Living Standards. *Can Measurement Error Explain the Weakness of Productivity in the Canadian Construction Industry?* *Op. cit.*

¹¹² Paul M. Goodrum, Carl T. Haas, and Robert W. Glover, "The Divergence in Aggregate and Activity Estimates of US Construction Productivity" *Op. cit.*

intermediate inputs. The weights are cost shares of the corresponding inputs. The multifactor productivity measure defined in Goodrum *et al.* for task-level productivity analysis is also different from the definition conventionally used at the industry level. It is defined as units of physical output divided by the deflated sum of labor cost and equipment cost. For multifactor productivity analysis, the productivity literature recommends the use of gross output (as opposed to value added) as the output measure along with symmetrical treatment of labor, capital, and intermediate inputs.¹¹³ While gross output measures are used in both formulations, task-level productivity defined by Goodrum *et al.* does not incorporate intermediate inputs. The presumed increase in prefabrication of materials could explain some of the increase in multifactor productivity and labor productivity defined in Goodrum *et al.*, which are not reflected in productivity measures at the industry level.¹¹⁴

3.7 Conclusions and Observations

Task-level productivity estimates tend to show improvement in construction productivity over time, while industry-level productivity estimates tend to suggest otherwise. Some industry practitioners believe the construction industry has witnessed enhancements in productivity, while others believe productivity has been lagging. This divergence in estimates and in perceptions highlights the challenges associated with productivity measurement of the construction industry. If we set aside the issue of data not collected for the purpose of productivity measurement, we find that there is an intrinsic difficulty in construction productivity measurement. Much of this difficulty lies in the heterogeneous nature of the industry. Construction building or infrastructure types are heterogeneous. Within each building or infrastructure type, there is also heterogeneity as each project is unique. Building processes are heterogeneous, as demonstrated by the diversity of contract work on which the North American Industry Classification System (NAICS) is based. Finally, there is heterogeneity in the composition of construction firms, with large operations taking advantage of economies of scale and scope and making more profits than small companies. The heterogeneity that exists in these multiple dimensions means that productivity may be improving or deteriorating for a particular segment of the industry, at a particular level of analysis. Changes in productivity at an aggregated level may simply be caused by changes in the composition of projects or firms involved, rather than reflecting productivity change per se. The next chapter will discuss possible approaches of disaggregating the industry to create productivity measures that are more meaningful.

¹¹³ Frank M. Gollop, “Accounting for Intermediate Input: The Link Between Sectoral and Aggregate Measures of Productivity Growth,” in *Measurement and Interpretation of Productivity*, National Research Council (Washington, D.C.: National Academy of Sciences, 1979); Dale W. Jorgenson, Mun S. Ho, and Kevin J. Stiroh, *Productivity Volume 3: Information Technology and the American Growth Resurgence* (Cambridge, Massachusetts and London, England: The MIT Press, 2005).

¹¹⁴ *Ibid.*

4 Challenges and Opportunities in Construction Productivity Measurement Using National Statistics

4.1 Discussion of the Teicholz Diagram

One highly referenced work in construction productivity is the productivity comparison diagram plotted by Teicholz.¹¹⁵ This following discussion focuses on the productivity calculations by Teicholz and highlights the data challenges associated with construction productivity measurement. In Teicholz's diagram, constant contract dollars of new construction work per field work hour is shown to have trended downward over the past 40 years at an average compound rate of -0.6 % per year. In contrast, labor productivity of all non-farm industries, which includes the construction industry, has trended upward at an average compound rate of 1.8 % per year.

Figure 4.1 Construction Labor Productivity and Non-Farm Business Labor Productivity Comparison¹¹⁶



¹¹⁵ Paul Teicholz, "Labor Productivity Declines in the Construction Industry: Causes and Remedies," *AECbytes Viewpoint*. Issue 4, April 14, 2004.

¹¹⁶ *Ibid.*

In the Teicholz calculations, the output measure for the construction industry, constant contract dollars of new construction work, is from the C30 Value of Construction Put in Place reports produced by the Census Bureau. Value Put in Place, as defined by the C30 survey, includes architectural design, engineering costs, construction management (since 1997), force-account construction, and secondary construction, in addition to total construction cost as defined by the Economic Census.¹¹⁷ The Census reports that about two thirds of Value Put in Place corresponds to the work performed by the construction industry as defined by the Economic Census.¹¹⁸ Defined as such, the output measure contains contributions of industries outside of the construction industry.

C30 reports published contract amounts in constant dollars, and these published figures were the output measure in the Teicholz calculations. The Census Bureau used an array of price and cost indices for the deflation. Possibly because of the lack of appropriate deflators, C30 reports currently only publish contract amounts in current dollars.

Labor hours data that Teicholz used are field work hours, and they came from the Bureau of Labor Statistics. The BLS collects and reports labor hours data under the Current Employment Statistics Survey. Field work hours are work hours of the so-called production workers. Executive and managerial personnel, professional and technical employees, and workers with routine office jobs are considered non-production workers and therefore are excluded. In the BLS productivity program, hours of both production workers and non-production workers are combined to form total hours, which are used in labor productivity calculations.¹¹⁹ The CES is an establishment survey that covers establishments with payrolls. The self-employed are not under the sampling universe of the CES. About 15 % to 19 % of workers in the construction industry are self-employed or unpaid family workers, and therefore non-negligible.¹²⁰ The labor hours data used in Teicholz productivity calculations come from the CES and do not include hours worked by the self employed. Excluding the self employed in the labor hours biases productivity measure upward.

In addition, the Census notes that there is “a significant amount of construction work done in the underground economy.”¹²¹ The existence of an underground economy might be more likely to

¹¹⁷ United States Census Bureau. Construction Statistics Data Users’ Conference. October 28, 1997. Washington, DC. Document issued March, 1999.

¹¹⁸ *Ibid.*

¹¹⁹ Appendix A discusses the estimation of non-production worker hours, as they are not collected under the Current Employment Statistics.

¹²⁰ Current Population Survey.

¹²¹ *Ibid.*

affect the labor input than the output measure when the output measure is based on projects. The labor input is underestimated if the labor in the underground economy is ignored.

Finally, changes in labor productivity may be a result of changes in the mix of outputs. Residential and commercial building construction is labor intensive, compared to industrial and heavy construction, which tends to be capital intensive. Rojas and Aramvareekul argue that Labor productivity is lower for residential and commercial building construction than industrial and heavy construction.¹²² From 1964 through 2007, the output mix changed from 64 % residential and commercial and 36 % industrial and heavy construction to 76 % residential and commercial and 24 % industrial and heavy construction.¹²³ Increasing residential and commercial construction in the overall output mix could translate into decline in “labor productivity.” The decline in construction labor productivity illustrated in the Teicholz diagram might be partly explained by change in output mix. This illustration indicates the importance of focusing on homogenous products in productivity measurement. A productivity measure at the industry level alone is not sufficiently informative.

4.2 Data Issues Associated with the Teicholz Diagram

The C30 survey form specifically asks for contract construction cost, owner supplied materials and labor, and total construction cost, which is the sum of the former two. Architectural, engineering, and miscellaneous costs is asked of the survey respondent separately, as well as estimated amount of all other capital expenditures. These separate cost estimate data are not reported in the published C30 reports. Therefore, data on contract construction cost and total construction cost are not readily available. The contract construction cost and the total construction cost both contain labor costs and material costs. Value of Construction Put in Place includes architectural design, engineering costs, construction management, force-account construction, secondary construction, and total construction cost. In contrast, an output measure based on the contract construction cost or the total construction cost would be a gross output measure, with a boundary that approximates the construction industry. An additional investigation is needed to examine how labor hours data from the Current Employment Statistics treat owner supplied labor to determine whether contract construction cost or the total construction cost is a better output measure.

One source of data challenge for productivity analysis in the construction industry is the lack of appropriate price deflators. The output data from the C30 reports used to be deflated using an array of price and cost indices, and the more recent data are no longer deflated and are reported only in current dollars. The Bureau of Economic Analysis (BEA), on the other hand, deflates output from the construction industry for GDP estimation. For the residential sector, price

¹²² Eddy M. Rojas and Peerapong Aramvareekul, “Is Construction Labor Productivity Really Declining?” *Op. cit.*

¹²³ Data source is Census Bureau’s Value of Construction Put in Place.

deflators based on hedonic regressions are used and are considered reliable. For the nonresidential sector, the newly developed Producer Price Indices (PPI) by the Bureau of Labor Statistics are used. Two limitations are that these PPIs have only been available recently, and they have been developed for only four types of building construction projects. For the historical data, indices based on input costs are used for deflation. For the more recent data, this practice still continues as the price deflators do not exist for all construction types.

The main source of data on labor hours is the CES, which does not cover the self-employed. The Current Population Survey is one data source where the self-employed are covered. The Current Population Survey is a household survey and covers the self-employed. Additionally, the CPS collects data on hours worked and weeks worked. Therefore, hours of the self-employed can be obtained from the CPS. CPS hours data are hours at work, while the CES hours data are hours paid. Ratios of hours at work to hours paid may be available from the National Compensation Survey to convert hours paid to hours worked. Additionally, the CPS data are reported for all workers in the construction industry, while the CES hours data are reported only for production workers. Official productivity measures published by the BLS are constructed using hours worked for both production and nonproduction workers. Therefore, the same definition needs to be used to calculate a labor productivity measure for the construction industry that is consistent with official labor productivity measures of other segments of the economy.

4.3 Approaches for Measuring Construction Productivity

Literature review reveals that there is no consensus on the trend of construction productivity. This document aims to examine what data are available in national statistics that would allow us to create meaningful productivity measures for the construction industry. This effort focuses on labor productivity, rather than multifactor productivity, because the data requirements for labor productivity measurement are more limited and therefore more feasible.

The construction industry is highly heterogeneous. As a result, a single industry-level productivity measure alone is not sufficiently informative. Changes in the industry-level productivity may be due to changes in the composition of projects and therefore may not reveal true productivity changes. There are two possible types of approaches in classifying the industry into a finer level. The first approach is to focus on products. That is, productivity measures can be developed for different building types or infrastructure types. For each building or infrastructure type, productivity measures can be created, and these productivity measures can serve as benchmarks for practitioners who engage in such projects. The second approach is focused on production units, akin to how the NAICS codes are structured. The Economic Classification Policy Committee decided that “as a matter of principle, an industry classification

system should be based on producing units rather than products or services.”¹²⁴ In 2002, the NAICS codes for the construction industry were structured such that there are three broad categories: (1) construction of buildings, (2) heavy and civil engineering construction, and (3) specialty trade contractors. This basic structure of categorization was the foundation for SIC codes of the construction industry that were in use until supplanted by the NAICS codes in 1997. Since much data from national statistics are based on the NAICS codes, creating productivity measures that follow NAICS structure is a natural possibility. Five possible specific approaches are described below.

4.3.1 Focus on Building Types, Gross Output

Output:

Focusing on building types, say an office building, an output measure could be square footage of a project. Square footage values are collected in the C30 survey, which is part of the Census Bureau’s Value Put in Place Program. The sampling frame of the C30 survey for private nonresidential and for state and local government construction is based on F.W. Dodge reports,¹²⁵ which is a compilation of construction projects, and the Dodge reports also contains the square footage values of projects. Dodge reports go back to 1967. Using square footage as an output measure avoids the problem of lack of good output deflators, and a long time-series is also available.

Labor Input:

Data for labor input can come from the Economic Census. The Economic Census is an establishment survey and it covers all large establishments and a sample of small establishments. These establishments include general contractors and specialty trade contractors. The Economic Census reports “number of construction workers.”¹²⁶ The Economic Census also asks establishments regarding percentages of their work, based on sales, shipments, receipts, or revenue, in various building and infrastructure types. Starting with general contractors, we can first focus on establishments that specialize in office building construction. Obtaining the number of construction workers that work in office building construction is straight forward for establishments that specialize in office building construction. For establishments that work in

¹²⁴ Yuskavage, Robert E. 2007. “Converting Historical Industry Time Series Data from SIC to NAICS” Paper prepared for the Federal Committee on Statistical Methodology 2007 Research Conference, November 5-7, Arlington, VA.

¹²⁵ United States Census Bureau. Construction Statistics Data Users’ Conference. October 28, 1997. Washington, DC. Document issued March, 1999. *Op. cit.*

¹²⁶ Number of construction workers is available from the Economic Census for 2007 and 1992. For 1997, annual payroll costs for construction workers are reported instead of number of construction workers. For 2002, both number of construction workers and number of leased construction workers are reported.

office building construction but do not specialize, extrapolation is needed. For example, we can draw values of number of construction workers per dollar revenue from the distribution based on establishments that specialize. Using data on percentage revenue on office building construction and total revenue, we can obtain an estimate for number of construction workers. The same procedure can be applied to specialty trade contractors to obtain number of construction workers in office building construction.

The number of construction workers can then be combined with average weekly hours of production workers from the Current Employment Statistics Survey to yield an estimate of annual hours, which is the labor input that can be used in labor productivity calculation. The Current Employment Statistics Survey is an establishment survey, and it contains monthly data for detailed classification, largely based on NAICS. In the case of office building construction, average weekly hours of production workers in commercial building construction can be used, and data for this variable is available monthly starting from January 1990.

Challenges:

In terms of the output measures, square footage and the value of construction put in place are both gross output measures. A labor productivity measure can be constructed using gross output measures, but we need to keep in mind that only part of the output is contributed by the construction industry. An additional caveat is that while square footage is a sensible proxy for gross output and using it as an output measure avoids the problem of deflators, it does not control for quality and complexity changes over time.

There is no data on the self-employed (proprietorships and partnerships) in terms of number of construction workers or hours devoted to different building types. The Economic Census is an establishment survey that covers only establishments with payrolls. The Economic Census does report monetary amount of subcontract work, but this value includes both labor and materials. In addition, the subcontract work reported by a general contractor may be performed by a specialty trade contractor who is also included in the Economic Census. It is not possible to estimate the amount of work performed by the self-employed in particular types of building construction work.

The Current Population Survey is one data source where the self-employed are covered. From the Current Population Survey, about 15 % to 19 % of workers in construction are self-employed or unpaid family workers. The self-employed therefore represent a non-negligible portion of the construction work force. While the Current Population Survey is a household survey and covers the self-employed, the construction industry is not further categorized at a finer level. Under the CPS occupation classification, there are 31 occupation types of construction trades. Some of the construction occupations may fall under categories such as installation, maintenance, and repair workers or management occupations. Using data from the CPS, the BLS compiles construction occupations with the most substantial percentage of self-employed workers. Additionally, the

CPS collects data on hours worked and weeks worked. Therefore, information on the occupations of the self-employed workers is available. What is missing is the project types the self-employed engage in.

Another data source on the self-employed is the Nonemployer Statistics, annually published by the Census Bureau, as previously discussed. But this data source also does not allow us to link the self-employed to different project types. An additional issue is the lack of labor data from the underground economy.

Under this approach, although square footage data is available monthly and average weekly hours are also available monthly, number of construction workers working in the office building construction is available only every five years. The C30 survey contains project cost information; however, the labor and materials costs are not distinguished from each other. Labor input cannot be extrapolated using C30 results. Another data source on establishments is the County Business Patterns, which is an annual data on number of employees, payroll, and number of establishments by NAICS codes. It also reports number of establishments by employment-size class for NAICS categories. However, this data source also does not contain information pertinent for productivity measurement with a focus on project types. As a result, productivity estimates can only be made every five years.

4.3.2 Focus on Building Types, Value Added

Output:

The previous approach suffers from the lack of labor data for the self-employed and the underground economy. An alternative approach is to confine the scope and focus on the output of establishments with payrolls and the labor input of establishments with payrolls. Economic Census surveys ask individual establishments about labor costs, materials costs, the amount of business done, and percentage dollar value of work done by different building types. “Value added for the construction industry is defined as the dollar value of business done less costs for construction work subcontracted to others and payments for materials, components, supplies, and fuels.”¹²⁷ Therefore information needed to calculate value added at the individual establishment level is available. Since the individual components of value added are collected at the establishment level only (rather than at the establishment level by building types), value added of new construction by building types needs to be extrapolated.

To estimate value added by building types, we can first obtain value added from establishments that specialize in, say, office building construction. We can then construct a distribution based on value added as a fraction of total revenue. Then for establishments that engage in office building construction but do not specialize, we can draw values from the distribution. This a sensible

¹²⁷ Economic Census 2002.

approach to the extent that value added as a fraction of revenue is similar for the same building type. The Economic Census forms ask respondents to report percent of construction work by building type and by types of construction (new construction; additions, alterations, or reconstruction; and maintenance). Using the percent of construction work and the randomly drawn fraction of value added divided by total revenue, we can derive value added by building type for each establishment. If there are not enough establishments that specialize in certain building or infrastructure types, then it may be sensible to combine several building or infrastructure types together.

Labor Input:

Labor input data is the same as in the first approach (4.3.2).

Challenges:

To use the value added approach, two deflators are needed. The so-called double deflation method involves first deflating gross output by an appropriate price deflator and second deflating intermediate inputs by an appropriate deflator. Subtracting the deflated intermediate inputs from the deflated gross output yields a deflated value added measure.

Currently, the Producer Price Index (PPI) published by the Bureau of Labor Statistics for the new office building construction is available monthly only from June 2006. PPI for new industrial building construction has been available monthly since June 2007. PPI for new warehouse building construction is available monthly from December of 2004. Finally, PPI for new school building construction is available from December 2005. The short data series of PPIs presents a limitation.

In terms of deflators for intermediate inputs, the BLS constructs producer price indices for material and supply inputs to the construction industries. There is an index for the overall inputs to construction industries, and under which there are indices for new construction and maintenance and repair construction. The breakout of new construction is: single-unit residential, multi-unit residential, nonresidential buildings, highway and street construction, and other heavy construction. The breakout for maintenance and repair is: residential and non-residential. Each of the indices is based on a list of NAICS industries that supply materials to the sector in question and their relative weights. The lists of NAICS industries and the relative weights come from the BEA's input-output tables. Since the input-output tables are based on NAICS codes, data to reconstruct deflators for intermediate inputs to specific project or building types, such as office building construction, are not readily available. A compromise might be simply to use the index for non-residential building construction as a proxy for office building construction. Another limitation is the infrequent data collection of the Economic Census.

4.3.3 Focus on Infrastructure Type, Value Added

Output:

This approach is similar in nature to the previous approach. Instead of focusing on a particular building type, we focus on an infrastructure type. Under the category “Heavy and Civil Engineering Construction,” we could focus on “highway and street construction,” which includes both general contractors and specialty trade contractors who work in this area. The output measure is value added by establishments that specialize in this area plus imputed value added by establishments that work in this area but do not specialize.

Labor Input:

From the Economic Census we can obtain number of construction workers. From Current Employment Statistics Survey, available monthly from January 1990, we can obtain number of production workers for “highway, street, and bridge construction,” as well as total number of employees and average weekly hours of production workers in “highway, street, and bridge construction.” From the Current Employment Statistics survey under SIC, we can obtain monthly data for the period of January 1988 through March 2003, although the category is “highway and street construction.”

Challenges:

Because the output measure is value added by establishments with payrolls, this approach avoids the problem of lack of labor data associated with self-employment. Producer Price Index is available for material and supply inputs to “highway and street construction” from the Bureau of Labor Statistics monthly from June 1986. Whether a producer price index is available for the output needs to be investigated.

Another challenge is that labor data from the Current Employment Statistics Survey are available for the combined highway, street, and bridge construction for the new series, while labor data is available for only highway and street construction in the old series. Assumptions are needed to use data from these different sources together.

4.3.4 Focus on Specialty Trades

Output:

Since much of existing data are classified under the NAICS system, it is natural to follow the NAICS system when creating productivity measures. PPIs have been developed for four specialty trades by the BLS: concrete contractors (nonresidential building), roofing contractors (nonresidential building), electrical contractors (nonresidential building), and plumbing/HVAC contractors (nonresidential building). These four types of specialty trades are also covered under

the Economic Census. Gross output obtained from the Economic Census can be deflated using the corresponding PPIs.

Labor Input:

The Economic Census contains information on the number of construction workers and the number of total employees. Current Employment Statistics Survey contains average weekly hours, number of all employees, and number of production workers for “poured concrete structure contractors,” “steel and precast concrete contractors,” “roofing contractors,” “electrical contractors,” and “plumbing and HVAC contractors.”

Challenges:

These PPIs for the specialty trades have become available monthly since December 2007. As a result, there is not a long enough time series to construct a productivity trend using these data. Specialty trades may subcontract work to other contractors, and some of which may be self-employed. This issue can be mitigated, as previously discussed, by focusing on value added by the establishments with payrolls and labor input by these establishments with payrolls. To obtain deflated value added, deflators of intermediate inputs are not currently available and therefore would also need to be constructed.

4.3.5 Focus on Residential Building Construction

Output:

Focusing on establishments with payrolls avoids the lack of data associated with the self-employed. Value added output measure can be derived using data from the Economic Census. The PPIs for “new one-family houses under construction” and “new one-family house sold” are available monthly for 1964 through 2007 and 1963 through 2007, respectively. These PPIs are published by the Census Bureau. The BEA and the Census also produce a PPI for multifamily housing. This latter PPI was introduced in 1991. The data series starts from 1958 in the BEA’s National Income and Product Accounts (NIPA) tables. The PPIs for the residential sector are derived using the hedonic regression approach and are considered high quality. In terms of deflators for intermediate inputs, deflators for material and supply inputs for single-unit residential construction and multi-unit residential construction from the BLS can be used.

Labor Input:

Number of all employees, number of production workers, and average weekly hours of production workers are available from the Current Employment Statistics Survey for construction of residential buildings and new single-family general contractors. These data are available monthly from January 1990, except the number of all employees for construction of residential buildings is available monthly from January 1985. Number of all employees is available for new multifamily general contractors from January 1990. Starting in January 2001,

the number of all employees is collected for residential specialty trade contractors, residential building foundation and exterior contractors, residential building equipment contractors, residential building finishing contractors, and other residential trade contractors.

Challenges:

Using data from Economic Census, we can produce a productivity trend using data from 1987, 1992, 1997, 2002, and 2007.

5 Summary and Recommendations for Future Research

5.1 Summary

Although the construction industry is a major component of the U.S. economy, it has experienced a “perceived” prolonged period of decline in productivity. Due to the critical lack of measurement methods, however, the magnitude of the productivity problem in the construction industry is largely unknown. The measurement problem is exacerbated by the fact that the construction industry is composed of four sectors that differ significantly in the outputs produced, firm size, and use of technology. The four sectors, which taken together define the construction industry, are residential, commercial/institutional, industrial, and infrastructure.

This report describes efforts underway that focus on the measurement of construction productivity at three levels—task, project, and industry—and how such measurements can be developed. This report analyzes the measurement challenges associated with the development of meaningful measures of construction productivity at the task, project, and industry levels and establishes a framework for addressing those challenges. Specifically, this report identifies the metrics, tools, and data needed to move forward in collaboration with key construction industry stakeholders. Once produced, these metrics, tools, and data will help construction industry stakeholders make more cost-effective investments in productivity enhancing technologies and improved life-cycle construction processes; they will also provide stakeholders with new measurement and evaluation capabilities. Finally, this report lays the foundation for future research and for establishing key industry collaborations that will enable more meaningful measures of construction productivity to be produced at the task, project, and industry levels.

5.2 Recommendations for Future Research

The background work for this report uncovered additional areas of research that might be of value to private-sector organizations and government agencies concerned with the measurement and analysis of construction productivity. These areas of research are concerned with: (1) the development of a standard practice for measuring task-level and project-level productivity; (2) the establishment of a database of project-level productivity measures for selected types of capital facilities; (3) factors affecting the use of prefabrication, preassembly, modularization, and off-site fabrication techniques and processes; and (4) industry-level productivity metrics.

5.2.1 Standard Practice for Measuring Task-Level and Project-Level Productivity

Standards are an efficient way of translating research results into practice. Improved metrics, if embodied in a voluntary industry consensus standard, will increase the rate of investment in productivity enhancing technologies, including information, communication, and automation and integration technologies, conveying benefits on individuals, businesses, and government in the form of lower costs of building services and products. Future research aimed at developing, in collaboration with ASTM International, a standard practice for measuring task-level and project-

level productivity will fill that need. Ideally, the standard practice will incorporate metrics that enable leading-indicators of performance to be calculated and used to identify areas for improvement during the construction phase. The Building Economics Subcommittee, ASTM E06.81, is the ideal venue for producing this standard practice. Furthermore, the 45 members of the ASTM E06.81 Subcommittee on Building Economics and over 600 ASTM E06 Committee members on Building Performance and Constructions provide an excellent user base as well as industry marketing spokespersons for such a standard practice. Over the longer term, the metrics defined in the standard practice can be embodied in supporting software products that will help implement the standard by various stakeholder groups.

5.2.2 Database of Project-Level Productivity Measures for Capital Facilities

Although there are a number of sources for task-level productivity data, no such sources exist for project-level productivity measures. A recent study, sponsored by NIST and conducted by the Construction Industry Institute (CII), discussed two promising approaches for reporting project-level productivity metrics.¹²⁸ Additional research on these two approaches in conjunction with CII's Benchmarking and Metrics Program and other key construction industry stakeholders, could result in a database of project-level productivity measures for selected types of capital facilities (e.g., industrial facilities, commercial and institutional buildings, and infrastructure projects). Such a database would offer a means for disseminating information on project-level productivity. The database would consist of both raw metrics (e.g., direct measures of inputs and outputs) and index-based metrics (e.g., a reference value of 100 pegged to some reference point in time). An advantage of index-based metrics is that they enable project-level productivity to be tracked over time and to spot trends. Ideally, the database would incorporate the capability to analyze how the use of industry best practices and automation and integration technologies affect project-level productivity.

5.2.3 Prefabrication, Preassembly, Modularization, and Off-Site Fabrication

Prefabrication, preassembly, modularization, and off-site fabrication (PPMOF) involve the fabrication or assembly of systems and components at off-site locations and manufacturing plants. Once completed, the systems or components are shipped to a construction job site for installation at the appropriate time. Both owners and contractors view PPMOF as a means to meet challenges of demanding schedules, adverse site conditions, and limited availability of skilled labor. PPMOF offers the promise—if used appropriately—of lower project costs, shorter schedules, improved quality, and more efficient use of labor and materials. However, various obstacles stand in the way of the widespread use of PPMOF, including building codes that hinder innovation as well as conventional design and construction practices. In addition to the obstacles referenced above, CII research shows that using PPMOF techniques and practices requires

¹²⁸ Youngcheol Kang, William O'Brien, Jiukun Dai, Stephen P. Mulva, Stephen R. Thomas, and Pin-Chao Liao, *Measuring Interoperability and Best Practices Impacts on Capital Project Productivity*. *Op. cit.*

careful consideration of their implications for engineering, transportation, coordination, and project organization.¹²⁹ Additional research is needed to better understand why the successful use of PPMOF techniques and practices in the industrial sector has not been duplicated in the commercial/institutional and infrastructure sectors. Ideally, this research would be conducted as part of a broad-based initiative to understand the pros and cons of PPMOF techniques and practices from a market-based perspective.

5.2.4 Industry-Level Productivity Metrics

Statistical agencies, such as the Census Bureau, the Bureau of Labor Statistics (BLS), and the Bureau of Economic Analysis (BEA), fulfill many needs of the nation through collection and compilation of high-quality data. For instance, an accurate account of the size of the GDP and sizes of contributions by various industry sectors are fundamental for a basic understanding of the nation's economy. Data on housing starts serve as an important indicator that captures the cyclic nature of the economy. To understand how industries make decisions on labor and capital utilizations, data on labor and capital investments are indispensable. To achieve goals like these, the national statistics offices have collected much of the data that are relevant to productivity measurement of the construction industry.

One key element in productivity measurement is output deflators. BLS has recently produced several producer price indices in the nonresidential sector, and this effort has enhanced the estimates of investments in BEA's National Income and Product Account tables. BEA and BLS may collaborate further to develop other nonresidential building construction indices, such as price indices for highways, hospitals, retail, communication, power, and lodging structures.¹³⁰ Efforts such as this improve the quality of existing statistics and have spillover benefits for productivity measurement.

There are currently no official productivity statistics on the construction industry due to the lack of adequate data. One challenge that stands out is the mismatch of classification systems between different data sources. One potential remedy may be to classify micro-level data in two or more classification systems designed for different purposes. For instance, the updated classification system of Value of Construction Put in Place is based on end use. For productivity measurement of the construction industry, a classification system based on building type is preferable. Information on both the building type and the end use is available, and therefore it is possible to classify the same data under two systems for different uses. The use of microdata for reclassification can be applied to datasets that have undergone classification changes as well. For instance, the Economic Census and the Current Employment Statistics both transitioned

¹²⁹ Construction Industry Institute, *Prefabrication, Preassembly, Modularization, and Offsite Fabrication in Industrial Construction: A Framework for Decision-Making*. *Op. cit.*

¹³⁰ Paul Lally, "How BEA Accounts for Investment in Private Structures." *Op. cit.*

from SIC to NAICS, and as a result, there is a major break in the time series. Reconstructing historic SIC-based data by reclassifying establishments surveyed is one approach.

Another remedy for reconciling the different classification systems under which output and labor input data are organized may be to ask additional questions on existing surveys. The Value of Construction Put in Place survey, for instance, is reported monthly. The main variable is construction costs, which includes costs of both materials and labor. The current Value of Construction Put in Place does not request owners to report material cost and labor cost separately. If information on labor costs and/or hours can also be collected in the same survey, then productivity measures can be developed for different project types and possibly for different geographic regions. Alternatively, the Current Employment Statistics Survey could ask respondents the types of projects they are currently working on. If monthly surveys such as the Value of Construction Put in Place or the Current Employment Statistics Survey can be slightly amended, a rich data set for productivity measurement could be within reach and can enable calculations of labor productivity by project types. Such a productivity measure could be used as a benchmark for owners and contractors engaged in specific project types. Another challenge is the under coverage of the self-employed in establishment-based surveys. A solution may be to develop a supplementary survey that aims to fill this gap.

References

- Abdowd, John M., Paul A. Lengermann, and Kevin L. McKinney. 2002. *The Measurement of Human Capital in the U.S. Economy*. Longitudinal Employer Household Dynamics Technical Paper No. TP-2002-09.
- Allen, Steve G. 1985. Why Construction Industry Productivity is Declining. *Review of Economics and Statistics* 67: 661-669.
- Allmon, E., C. T. Hass, J. D. Borcharding, and P. M. Goodrum. 2000. U.S. Construction Labor Productivity Trends, 1970-1998. *Journal of Construction Engineering and Management* 126, no. 2: 97-104.
- Baily, Martin Neil and Robert J. Gordon. 1988. The Productivity Slowdown, Measurement Issues, and the Explosion of Computer Power. *Brookings Papers on Economic Activity* 1988 No. 2: 347-420.
- Bayard, Kimberly N. and Shawn D. Klimek. 2004. Creating a Historical Bridge for Manufacturing Between the Standard Industrial Classification System and the North American Industry Classification System. *Proceedings of the American Statistical Association, Business and Economic Statistics Section* [CD-ROM], pp. 478-84.
- Black, John. 2002. *Oxford Dictionary of Economics*. Oxford and New York: Oxford University Press.
- Brundtland, G.H. (editor). 1987. *Our common future: Report of the UN Commission on Environment and Development*. <http://www.un-documents.net/wced-ocf.htm> (accessed September 3, 2009).
- Bureau of Economic Analysis. 2009. *Concepts and Methods of the U.S. Input-Output Accounts*. Washington D.C.: Bureau of Economic Analysis.
- Bureau of Economic Analysis. Gross-Domestic-Product-(GDP)-by-Industry Data. *Industry Economic Accounts*. Washington, D.C.: Bureau of Economic Analysis, http://www.bea.gov/bea/dn2/gdpbyind_data.htm (accessed July 2009).
- Bureau of Labor Statistics. Household Data: Employed Persons in Nonagricultural Industries by Sex and Class of Worker. *Current Population Survey* (Washington, DC: Bureau of Labor Statistics), <http://www.bls.gov/cps/cpsaat16.pdf> (accessed July 2009).
- Bureau of Labor Statistics. 1993. *Labor Composition and U.S. Productivity Growth, 1948-90*. Washington D.C.: Bureau of Labor Statistics.
- Bureau of Labor Statistics. 1997. *Handbook of Methods*. Washington D.C.: Bureau of Labor Statistics.

- Bureau of Labor Statistics. 2005. *Construction of Average Weekly Hours for Supervisory and Nonproduction Wage and Salary Workers in Detailed Industries*. Technical Note. Washington D.C.: Bureau of Labor Statistics.
- Bureau of Labor Statistics. 2006. *Construction of Employment and Hours for Self-employed and Other Nonfarm Workers and for all Farm Workers, Using Current Population Survey Data for Primary and Secondary Job*. Technical Note. Washington D.C.: Bureau of Labor Statistics.
- Bureau of Labor Statistics. 2007. *Technical Information About the BLS Multifactor Productivity Measures*. Washington D.C.: Bureau of Labor Statistics.
- Bureau of Labor Statistics. 2009. *Changes in the Composition of Labor for BLS Multifactor Productivity Measures, 2007*. Technical Report. Washington D.C.: Bureau of Labor Statistics.
- Bureau of Labor Statistics. 2009. *Multifactor Productivity Trends, 2007*. Washington D.C.: Bureau of Labor Statistics.
- Business Roundtable. 1988. *CICE—The Next Five Years and Beyond*. New York, NY: The Business Roundtable.
- Business Roundtable. 1997. *Confronting the Skilled Construction Work Force Shortage—A Blueprint for the Future*. The Business Roundtable.
- Centre for the Study of Living Standards. 2007. *Can Measurement Error Explain the Weakness of Productivity in the Canadian Construction Industry?* Research Report no. 2007-01. Ontario: Centre for the Study of Living Standards.
- Chapman, Robert E., and David T. Butry. 2008. *Measuring and Improving the Productivity of the U.S. Construction Industry: Issues, Challenges, and Opportunities*. NIST White Paper. Gaithersburg, MD: National Institute of Standards and Technology.
- Civil Engineering Research Foundation. 1994. *A Nationwide Survey of Civil Engineering-Related R&D*. CERF Report #93-5006. Reston, VA: American Society of Civil Engineers.
- Construction Industry Institute. 1999. *Design for Maintainability: Improving Project Return on Investment*. RS142-1. Austin, TX: Construction Industry Institute.
- Construction Industry Institute. 2002. *Prefabrication, Preassembly, Modularization, and Offsite Fabrication in Industrial Construction: A Framework for Decision-Making*. RS 171-1. Austin, TX: Construction Industry Institute.
- Construction Industry Institute. 2003. *Benchmarking and Metrics Value of Best Practices Report*. BMM 2003-4. Austin, TX: Construction Industry Institute.

- Construction Industry Institute. 2003. *The Shortage of Skilled Craft Workers in the U.S.* RS 182-1. Austin, TX: Construction Industry Institute.
- Construction Users Roundtable. 2004. *Confronting the Skilled Construction Workforce Shortage.* WP-401. Cincinnati, OH: Construction Users Roundtable.
- Construction Industry Institute. Strategic Plan (Austin, TX: Construction Industry Institute, 2005).
- Construction Industry Institute. 2007. *Construction Industry Craft Training in the United States and Canada.* RS 231-1. Austin, TX: Construction Industry Institute.
- Construction Industry Institute. 2008. *Leveraging Technology to Improve Construction Productivity.* RS 240-1. Austin, TX: Construction Industry Institute.
- Construction Users Roundtable. 2007. Pre-Assembly Perks: Discover Why Modularization Works. *The Voice*, Fall , pp 28-31.
- Dean, Edwin R. 1999. The Accuracy of the BLS Productivity Measures. *Monthly Labor Review* February: 24-34.
- Dean, Edwin R., Michael J. Harper, and Mark S. Sherwood. 1996. "Productivity Measurement with Changing-Weight Indices of Outputs and Inputs" In *Industry Productivity: International Comparison and Measurement Issues*. Paris: OECD.
- Dumagan, Jesus C. 2002. Comparing the superlative Tornqvist and Fisher ideal indices. *Economic Letters* 76: 251-258.
- de Leeuw, Frank. 1993. A Price Index for New Multifamily Housing. *Survey of Current Business* Feb 73(2): 33-42.
- Diewert, W. Erwin. 1976. Exact and Superlative Index Numbers. *Journal of Econometrics* vol 4, no. 4: 115-45.
- Diewert, W. Erwin. 1978. Superlative Index Numbers and Consistency in Aggregation. *Econometrica* July: 883-900.
- Diewert, W. Erwin. 2001. "Which (Old) Ideas on Productivity Measurement Are Ready to Use?" in *New Developments in Productivity Analysis*, ed. Charles Hulten, Edwin Dean, and Michael J. Harper. Conference on Research in Income and Wealth. Chicago: University of Chicago Press.
- Eastman, Chuck, Paul Teicholz, Rafael Sacks, and Kathleen Liston. 2008. *BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers, and Contractors*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Eastman, Charles M. and Rafael Sacks. 2008. Relative Productivity in the AEC Industries in the United States for On-site and Off-site Activities. *Journal of Construction Engineering and Management* 134, no. 7: 517-526.

Eldridge, Lucy P., Marilyn E. Manser, and Phyllis Flohr Otto. 2004. Alternative Measures of Supervisory Employee Hours and Productivity Growth. *Monthly Labor Review* April: 9-28.

Faruqui, Umar, Wulong Gu, Mustapha Kaci, Mirielle Laroche, and Jean-Pierre Maynard. 2003. Differences in Productivity Growth: Canadian-U.S. Business Sectors, 1987-2000. *Monthly Labor Review* 126 April: 16-29.

FIATECH. *Capital Projects Technology Roadmapping Initiative* (Austin, TX: FIATECH, October 2004).

Figuroa, Eric B. and Rose A. Wood. 2007. Industry Output and Employment Projections to 2016. *Monthly Labor Review* November: 53-85.

Fuller, Sieglinde K., and Stephen R. Petersen. 1996. *Life-Cycle Costing Manual for the Federal Energy Management Program*. NIST Handbook 135. Gaithersburg, MD: National Institute of Standards and Technology.

Gallaher, Michael P., Alan C. O'Connor, John L. Dettbarn, and Linda T. Gilday. 2004. *Cost Analysis of Inadequate Interoperability in the U.S. Capital Facilities Industry*. NIST GCR-04-867. Gaithersburg, MD: National Institute of Standards and Technology.

Gollop, Frank M. 1979. "Accounting for Intermediate Input: The Link Between Sectoral and Aggregate Measures of Productivity Growth," in *Measurement and Interpretation of Productivity*. National Research Council. Washington, D.C.: National Academy of Sciences.

Goodrum, Paul M., M. Yasin, and Z. Dong. *The Relationship Between Changes in Material Technology and Construction Productivity*. Mimeo. Lexington, KY: University of Kentucky.

Goodrum, Paul M. and Carl T. Haas. 2001. Closure to "U.S. Construction Labor Productivity Trends, 1970-1998." *Journal of Construction Engineering and Management* 127: 427-429.

Goodrum, Paul M. and Carl T. Haas. 2002. Partial Factor Productivity and Equipment Technology Change at Activity Level in U.S. Construction Industry. *Journal of Construction Engineering and Management* 128: 463-472.

Goodrum, Paul M., Carl T. Haas, and Robert W. Glover. 2002. The Divergence in Aggregate and Activity Estimates of US Construction Productivity. *Construction Management and Economics* 20, no. 5: 415-423.

- Goodrum, Paul M. and Carl T. Haas. 2004. Long-Term Impact of Equipment Technology on Labor Productivity in the U.S. Construction Industry at the Activity Level. *Journal of Construction Engineering and Management* January/February: 124-133.
- Gu, Wulong and Mun S. Ho. 2000. A Comparison of Industrial Productivity Growth in Canada and the United States. *American Economic Review* 90: 172-175
- Gullickson, William. 1995. Measurement of Productivity Growth in U.S. Manufacturing. *Monthly Labor Review*. July: 13-35.
- Gullickson, William and Michael J. Harper. 1999. Possible Measurement Bias in Aggregate Productivity Growth. *Monthly Labor Review* February: 47-67.
- Harper, Michael J. 1999. Estimating Capital Inputs for Productivity Measurement: An Overview of U.S. Concepts and Methods. *International Statistical Review* 67:327-337.
- Hirsch, Barry T. and David A. Macpherson. 2003. Union Membership and Coverage Database from the Current Population Survey: Note. *Industrial and Labor Relations Review*, 56: 349-54.
- Jorgenson, Dale W. 1990. "Productivity and Economic Growth," in *Fifty Years of Economic Measurement—the Jubilee of the Conference on Research in Income and Wealth*, ed. Ernst R. Berndt and Jack E. Triplett. Chicago and London: University of Chicago Press.
- Jorgenson, Dale W. 1995. *Productivity, Volume 1: Postwar U.S. Economic Growth*. Cambridge, MA and London: MIT Press.
- Jorgenson, Dale W., F. M. Gollop, and B. M. Fraumeni. 1999. *Productivity and U.S. Economic Growth*. Cambridge, MA and London: Harvard University Press.
- Jorgenson, Dale W., Mun S. Ho, and Kevin J. Stiroh. 2005. *Productivity Volume 3: Information Technology and the American Growth Resurgence*. Cambridge, Massachusetts and London, England: MIT Press.
- Kang, Youngcheol, William O'Brien, Jiukun Dai, Stephen P. Mulva, Stephen R. Thomas, and Pin-Chao Liao. 2009. *Measuring Interoperability and Best Practices Impacts on Capital Project Productivity*. NIST GCR 09-925. Gaithersburg, MD: National Institute of Standards and Technology.
- Klimek, Shawn D. and David R. Merrell. 2000. On Reclassifying Industries from the Standard Industrial Classification System to the North American Industry Classification System. Paper presented at the Second International Conference on Establishment Surveys, Buffalo, New York.
- Lally, Paul R. 2009. How BEA Accounts for Investment in Private Structures. *Survey on Current Business* February: 9-15.

Lawson, Ann M., Brian C. Moyer, Sumiye Okubo, and Mark A. Planting. 2006. "Integrating Industry and National Economic Accounts, First Steps and Future Improvements," In *A New Architecture for the U.S. National Accounts*, ed. Dale W. Jorgenson, J. Steven Landefeld, and William D. Nordhaus. Chicago and London: University of Chicago Press.

LEHD, *A Layman's Guide to the LEHD Human Capital Measures*. Longitudinal Employer-Household Dynamics Informational Document No. ID-2003-04. 2003

Mark, Jerome A. 1986. Problems Encountered in Measuring Single- and Multifactor Productivity. *Monthly Labor Review*: 3-11.

Morisi, Teresa L. 2003. Recent Changes in the National Current Employment Statistics Survey. *Monthly Labor Review* June: 3-13.

National Research Council. 1979. *Measurement and Interpretation of Productivity* (National Academy of Sciences. Washington D.C.: National Academy of Science.

National Research Council. 2009. *Advancing the Competitiveness and Efficiency of the U.S. Construction Industry*. Washington, DC: National Academies Press.

OECD (Organization for Economic Co-operation and Development). 2001. *Measuring Productivity—Measurement of Aggregate and Industry-Level Productivity Growth*. Paris: OECD.

OECD (Organization for Economic Co-operation and Development). Glossary of Statistical Terms. <http://stats.oecd.org/glossary/> (accessed September 2009).

Pieper, Paul. 1990. "The Measurement of Construction Prices: Retrospect and Prospect." In *Fifty Years of Economic Measurement: The Jubilee of the Conference on Research in Income and Wealth*, Volume 54, ed. Ernst R. Berndt and Jack E. Triplett. Chicago: University of Chicago Press.

Teicholz, Paul. 2001. Discussion of "U.S. Construction Labor Productivity Trends, 1970-1998." *Journal of Construction Engineering and Management* 127: 427-428.

Teicholz, Paul. 2004. Labor Productivity Declines in the Construction Industry: Causes and Remedies. *AECbytes Viewpoint*. Issue 4, April 14.

Rojas, Eddy M. and Peerapong Aramvareekul. 2003. Is Construction Labor Productivity Really Declining? *Journal of Construction Engineering and Management* 129: 41-46.

R.S. Means. 2008. *Building Construction Cost Data: 2009*. 67th Edition. Kingston, MA: R.S. Means.

Seskin, Eugene P. and Shelly Smith. "Improved Estimates of the National Income and Product Accounts: Results of the 2009 Comprehensive Revision" *Survey of Current Business* September (2009): 15-35.

Slaughter, E. Sarah. 1998. Models of Construction Innovation. *Journal of Construction Engineering and Management*. Vol. 124 May/June, pp. 226-231.

Stokes, H. Kemble Jr. 1981. An Examination of the Productivity Decline in the Construction Industry. *Review of Economics and Statistics* 63: 495-502.

United States Census Bureau: Manufacturing and Construction Division. Annual Value of Construction Put in Place. *Current Construction Report (CCR) C30* (Washington, DC: United States Census Bureau, July 3, 2007), <http://www.census.gov/const/C30/total.pdf> (accessed July 2009).

United States Census Bureau. 2005. 2002 Economic Census: Construction Subject Series. *Industry General Summary: 2002*. EC02-23SG-1. Washington, DC: U.S. Census Bureau, October.

United States Census Bureau. 1999. Construction Statistics Data Users' Conference. October 28, 1997. Washington, DC.

Weber, Stephen F. and Barbara C. Lippiatt. 1983. *Productivity Measurement for the Construction Industry*. NBS Technical Note 1172. Gaithersburg, MD: National Bureau of Standards, February.

Young, Norbert W. Jr. and Harvey M. Bernstein. 2006. Key Trends in the Construction Industry—2006. *SmartMarket Report*. New York, NY: McGraw Hill Construction, July.

Yuskavage, Robert E. 2007. Converting Historical Industry Time Series Data from SIC to NAICS. Paper prepared for the Federal Committee on Statistical Methodology 2007 Research Conference, November 5-7, Arlington, VA.

Appendix A Metrics of Productivity

Methodologies described in this section are largely based on the approach use by BLS in its productivity programs. BLS has been recognized as a world leader in productivity measurement, and its approach, based on work by Jorgenson and others, has been incorporated in the Organization of Economic Cooperation and Development (OECD) manual titled “Measuring Productivity,” which sets the international standard.¹³¹

A.1 Industry-Level Productivity Measures

A.1.1 Labor Productivity

Labor productivity is an output per hour measure. It is defined as $\frac{Q_t}{L_t} \div \frac{Q_0}{L_0}$, in the BLS formulation, where $\frac{Q_t}{Q_0}$ is the index of output in the current year (t) and $\frac{L_t}{L_0}$ is the index of labor input in the current year. Output Q_t can be a physical quantity measure or a deflated value of production, where output expressed in a monetary unit is divided by a price index, such as BLS producer price indices. Deflating value of production using appropriate producer price indices takes away changes in the value of output due to price changes. L_t is labor hours in year t . The construction industry produces different products—such as different building types and infrastructure types. To aggregate the different types of outputs, the output index can be calculated using a Tornqvist formula,¹³² in which quantities of different products are weighted using the value shares of the products:

$$\frac{Q_t}{Q_{t-1}} = \exp \left[\sum_{i=1}^n w_{i,t} \left(\ln \frac{q_{i,t}}{q_{i,t-1}} \right) \right]$$

where:

$\frac{Q_t}{Q_{t-1}}$ =the ratio of output in the current year (t) to previous year ($t-1$)

n =number of products,

$\ln \frac{q_{i,t}}{q_{i,t-1}}$ =the natural logarithm of the ratio of the quantity q of product i in current t year to the quantity in the previous year, and

$w_{i,t}$ =the average value share weight for product i

The average value share weight for product j is:

¹³¹ OECD (Organization for Economic Co-operation and Development), *Measuring Productivity—Measurements of Aggregate and Industry-Level Productivity Growth* (Paris: OECD, 2001).

¹³² Bureau of Labor Statistics, *Handbook of Methods* (Washington D.C.: Bureau of Labor Statistics, 1997).

$$w_{j,t} = \frac{S_{j,t} + S_{j,t-1}}{2}$$

Where: $S_{j,t} = p_{j,t} \times q_{j,t} \div (\sum_{i=1}^n p_{i,t} \times q_{i,t})$

And $p_{i,t}$ =price of product i at time t

The Tornqvist formula yields the ratio of output in year t to output in year $(t-1)$. The series of ratios can be chained to form the index of output used in the productivity formula. That is,

$$\frac{Q_t}{Q_0} = \frac{Q_t}{Q_{t-1}} \times \frac{Q_{t-1}}{Q_{t-2}} \times \dots \times \frac{Q_2}{Q_1} \times \frac{Q_1}{Q_0}$$

$q_{i,t}$ is generally calculated by dividing the value of output by the corresponding price index (BLS's producer price index). This approach is conceptually equivalent to indices based on physical quantities of output.

$\frac{L_t}{L_0}$ is calculated by dividing total labor hours in year t with total labor hours in the base year.

BLS does not distinguish between different types of labor hours in the output per hour measures.¹³³

Changes in labor productivity reflect changes in output that cannot be attributed to changes in the hours of labor in production.¹³⁴ Labor productivity reflects influences such as changes in capital input per labor unit, changes in technology, rates of capacity utilization, level of output, managerial skill, and effort and quality of labor. Changes in labor productivity cannot be solely attributed to changes in labor effort or quality.

A.1.2 Multifactor Productivity

Multifactor productivity (or total factor productivity) is the ability to produce more output with the same inputs.¹³⁵ It represents a shift in production function. Changes in multifactor productivity reflect changes in output that cannot be attributed to changes in capital inputs, labor inputs, and intermediate inputs. Changes in multifactor productivity reflect technological change, changes in capacity utilization, economies of scale, changes in managerial skills, changes in the organization of production, changes in the resource allocation, and measurement

¹³³ *Ibid.*

¹³⁴ Bureau of Labor Statistics, *Multifactor Productivity Trends, 2007* (Washington DC: Bureau of Labor Statistics, 2009).

¹³⁵ Dale W. Jorgenson, Mun S. Ho, and Kevin J. Stiroh, *Productivity Volume 3: Information Technology and the American Growth Resurgence. Op. cit.*

error.¹³⁶ Productivity represents the residual between output and inputs, and it also represents the residual between output prices and input prices.¹³⁷ It is the ability to mitigate input price increases without increasing the price of output. Or it is the ability to gain a competitive edge without input price reductions.¹³⁸

In the growth accounting framework, multifactor productivity growth is the growth in output minus weighted growth rates in capital, labor, and intermediate inputs. It is the residual, which is not accounted for by growth in labor and capital. The weights are the average value shares of the respective inputs in the value of the output between the two periods of consideration. In the equation form, the multifactor productivity growth is:

$$\frac{\dot{A}}{A} = \frac{\dot{Q}}{Q} - S_k * \frac{\dot{K}}{K} - S_l * \frac{\dot{L}}{L} - S_x * \frac{\dot{X}}{X}$$

where A is multifactor productivity, Q is output, K is capital, L is labor, X is intermediate input, S_k , S_l , and S_x are cost shares of capital, labor, and intermediate input, respectively, assuming competitive factor markets and constant returns to scale. That is,

$$S_i = \frac{p_i * i_i}{\sum(p_i * i_i)}$$

where p_i is the price of input i , and i is the quantity of input i .

A specific functional form of the production function must be chosen for implementation. The translog function is used because the assumptions associated with it are the least restrictive.¹³⁹ The Tornqvist index, which is consistent with the translog function, is used for aggregation. By being consistent, it is meant that “changes in output consistent with the very general translog production function are exactly measured by changes in Tornqvist indices.”¹⁴⁰ The

¹³⁶ National Research Council, *Measurement and Interpretation of Productivity* (National Academy of Sciences, Washington D.C.: National Academy of Science, 1979); Bureau of Labor Statistics, *Handbook of Methods* (Washington D.C.: Bureau of Labor Statistics, 1997); Jorgenson, Dale W., Mun S. Ho, and Kevin J. Stiroh, *Productivity Volume 3: Information Technology and the American Growth Resurgence* (Cambridge, MA and London, England: The MIT Press, 2005).

¹³⁷ William Gullickson, “Measurement of Productivity Growth in U.S. Manufacturing.” *Op. cit.*

¹³⁸ *Ibid.*

¹³⁹ Bureau of Labor Statistics, *Labor Composition and U.S. Productivity Growth, 1948-90* (Washington D.C.: Bureau of Labor Statistics, 1993).

¹⁴⁰ *Ibid.*

instantaneous growth rates are replaced by annual growth rates. For instance, $\frac{L}{L}$ is replaced by $\Delta \ln L = \ln L_t - \ln L_{t-1}$.¹⁴¹

The Tornqvist index of multifactor productivity growth is:

$$\Delta \ln A = \Delta \ln Q - \frac{1}{2} * (S_k(t) + S_k(t - 1)) * \Delta \ln K - \frac{1}{2} * (S_l(t) + S_l(t - 1)) * \Delta \ln L - \frac{1}{2} * (S_x(t) + S_x(t - 1)) * \Delta \ln X$$

Note that the output measure should not include any commodity taxes, because the producers do not receive these taxes.¹⁴² Intermediate input costs, on the other hand, should include commodity taxes because these taxes are paid by the producers.

For the multifactor productivity measures developed by the BLS, KLEMS inputs are used—capital, labor, energy, materials, and purchased business services. The construction of the productivity statistic using more input types is analogous to the case presented above.

A.1.3 Value-Added Function, Choice of Output Measure, and the Role of Intermediate Inputs

Note in the multifactor productivity formula, all outputs—capital, labor, and intermediate inputs—are treated symmetrically. When all outputs are treated symmetrically, substitution between any inputs is allowed. In contrast, when a value-added sub-function is assumed, intermediate inputs cannot be substituted with capital or labor inputs.¹⁴³ With the assumption of a value-added sub-function, the production function is written as:

$$Q = f(V(K, L, t), X)$$

This formulation assumes that the value added function is separable from intermediate inputs and that intermediate inputs do not influence the relative use of labor and capital. For instance, when the price of intermediate inputs decrease, a construction project may increase the use of intermediate inputs and reduce labor input. The assumption of a value-added function does not allow for such shifts in resource allocation. Additionally, this assumption implies that

¹⁴¹ *Ibid.*

¹⁴² W. Erwin Diewert, “Which (Old) Ideas on Productivity Measurement Are Ready to Use?” in *New Developments in Productivity Analysis*, ed. Charles Hulten, Edwin Dean, and Michael J. Harper. Conference on Research in Income and Wealth. (Chicago: University of Chicago Press, 2001).

¹⁴³ National Research Council, *Measurement and Interpretation of Productivity*. *Op. cit.*

productivity growth can only be accomplished through the value-added function. That is, intermediate inputs cannot be the medium or source of productivity growth.¹⁴⁴

Jorgenson *et al.* have developed an econometric method to test the existence of a value-added function.¹⁴⁵ They rejected the existence of a value-added function for construction industry, which was among the 40 of the 45 industries that were rejected for this assumption in their analysis. They observed that intermediate inputs constitute a large proportion of gross output for about 70 percent of the industries studied, suggesting for the use of gross output concept rather than value-added concept for productivity studies at the industry level.¹⁴⁶

Intermediate inputs are often substitutable with capital or labor in reality. For instance, at a construction site, putting together a door may involve cutting a door to fit certain dimensions, sanding the door, painting the door, and finally installing the door. In this process, all the activities are done by direct labor of the construction crew. The construction supervisor could also decide to purchase a pre-fabricated door. In this case, only the installation of the door involves direct labor, and the rest is accomplished through the purchase of an intermediate input.

Traditionally, construction is defined to include only activities at the work site. Labor productivity of construction associates value-added with labor input. Off-site prefabrication is considered manufacturing in Census classification. Comparable on-site and off-site activities, such as precast concrete, cast-in-place reinforced concrete, sheetrock installation, and elevators and escalators, were studied by Eastman and Sacks, and it was observed that onsite activities are less productive than the counterpart off-site activities.¹⁴⁷ The authors therefore argue that construction productivity is underestimated and that the production of prefabricated materials ought to be included in the construction productivity measurement. The observation by Eastman and Sacks and the theoretical and empirical work of Jorgenson *et al.* highlight the importance of treating intermediate inputs symmetrically in the productivity measurement. Including intermediate input in productivity measurement in this way recognizes the interdependence between sectors and makes it possible to evaluate the impact of contributions by other sectors, such as off-site prefabrication, on productivity growth.

¹⁴⁴ Frank M. Gollop, "Accounting for Intermediate Input: The Link Between Sectoral and Aggregate Measures of Productivity Growth." *Op. cit.*

¹⁴⁵ Jorgenson, Dale W., Mun S. Ho, and Kevin J. Stiroh, *Productivity Volume 3: Information Technology and the American Growth Resurgence.* *Op. cit.*

¹⁴⁶ *Ibid.*

¹⁴⁷ Charles M. Eastman and Rafael Sacks, "Relative Productivity in the AEC Industries in the United States for On-site and Off-site Activities," *Journal of Construction Engineering and Management* 134, no. 7 (2008): 517-526.

A.1.4 Output Measures

A.1.4.1 Gross Output versus Value Added

For multifactor productivity calculation, as discussed earlier, it is preferable to treat intermediate inputs symmetrically with capital and labor inputs. The proper measure of output is therefore gross output, rather than value added, where value added is defined to be gross output minus intermediate inputs.

Some researchers use value-added as the output measure, and incorporate only capital and labor inputs in their multifactor productivity measurement. This is the approach BLS uses for multifactor productivity for the two major sectors, namely, private business and private nonfarm business sectors. Value-added and gross output may be close in value at this level of aggregation.¹⁴⁸ But for disaggregated industries, gross output is preferred. And although it is more appropriate to use gross output, rather than value-added as the output measure, it might be preferable to use value-added as the output measure, for international productivity comparisons, as value-added data tend to be more available.¹⁴⁹

Which output measure is the preferred output measure in the labor productivity calculation is not clear, however, and there has been little coverage of this issue in the literature.¹⁵⁰

One data source for gross output is the C30 Value of Construction Put in Place, which includes architectural and engineering design, construction management, force-account construction, and secondary construction, in addition to construction services performed by the construction industry. Value of construction Put in Place is collected from the owners at the project level. Therefore, construction by the self-employed, homeowner construction, and construction done as a secondary source of revenue by nonconstruction establishments are covered.¹⁵¹ In contrast, these types of construction are not covered by an establishment-level survey such as the Economic Census. Contract construction cost is reported separately from owner supplied materials and labor and architectural, engineering, and miscellaneous costs in the survey form. Therefore, it is possible to obtain separate data on total construction cost. An output measure based on the contract construction cost or the total construction cost would be a gross output measure, with a boundary that approximates the construction industry. The classification of the

¹⁴⁸ Edwin R. Dean, Michael J. Harper, and Mark S. Sherwood, "Productivity Measurement with Changing-Weight Indices of Outputs and Inputs" In *Industry Productivity: International Comparison and Measurement Issues*, (Paris: OECD, 1996).

¹⁴⁹ *Ibid.*

¹⁵⁰ *Ibid.*

¹⁵¹ Bureau of Economic Analysis, *Concepts and Methods of the U.S. Input-Output Accounts* (Washington D.C.: Bureau of Economic Analysis, 2009).

Value of Construction Put in Place data is based on end usage, and the pre-1993 classification system was based on building and infrastructure types. Data classified under these two systems are generally not comparable, particularly at a disaggregated level. Value of Construction Put in Place data are collected monthly and are not deflated.

The only data source that allows the calculation of a value added output measure is the Economic Census. The Economic Census collects data on value of business done, costs for construction work subcontracted to others, and payments for materials, components, supplies, and fuels, which are components needed for calculating value added. One limitation of the Economic Census is that only establishments with payrolls are covered. The Economic Census is collected every five years with SIC/NAICS classification.

A.1.4.2 Price Deflators

An important element in productivity measurement is the price deflators. Price deflators are needed to derive a quantity index of output. This is done by dividing current dollars with an appropriate price deflator. Deflators are needed to strip away price changes due to inflation. Construction industry has been known to be deficient in this area, although many advances have been made in recent years.

Two notable price deflators with long time series are associated with the residential sector. The Census Bureau publishes price indices for new one-family houses sold and for new one-family houses under construction using the hedonic regression model. The series are monthly from 1963 and from 1964, respectively. Using a similar approach, the Bureau of Economic Analysis, in conjunction with the Census Bureau, has developed a price index for multifamily housing units.¹⁵² This price index series extends back to 1978.

The hedonic approach is a multiple regression approach, where the price is regressed on a number of characteristics that determine the price. The regression coefficients tell us how much each of the characteristics contribute to the price. If we choose a "typical" house in 1970 as the "standard house," we can figure out what the price of this house, if constructed new, would be in a later year. This is done by substituting the characteristics of the "standard house" in the regression model of, say 2009, to get the price of the house in 2009. Using this approach, we can get a price series through time for the same "standard house." If we then divide the monetary value of construction with the corresponding price, we can get a quantity index series through time. This quantity index series is in terms of the number of "standard houses" in each year. In this way, we keep the "product" quality constant. This quantity index can be an output measure in productivity studies. A bigger house in 2009 would cost more to construct than the standard house in 1970, for two reasons. One reason is that there is inflation. The other reason is that the house is bigger. The price series obtained from the above approach addresses both of these

¹⁵² Frank de Leeuw, "A Price Index for New Multifamily Housing." *Op. cit.*

issues. Inflation is accounted for because the regression model from year 2009 was used to "predict" the price of the "standard house." Houses in 2009 are bigger, and since we are artificially keeping the "product" constant, we are giving more credit to the output of 2009. That is, the actual number of houses constructed in 2009 is inflated to reflect the fact that the unit is the smaller "standard house."

In the previous section, two output data sources were discussed. The C30 Value of Construction Put in Place data are classified by project types or end usage while the Economic Census data are classified by SIC/NAICS. Two agencies currently create price deflators relevant to the construction industry. The BEA uses the C30 Value of Construction Put in Place data from the Census Bureau for its fixed investment and fixed assets data. In addition to construction spending defined by the Census Value of Construction Put in Place, BEA includes mining exploration, shafts, and wells, brokers' commissions on the sale of new and used structures, mobile structures, manufactured homes, and net purchases of used structures. The BEA reports quantity and price indices for categories of structures based on end usage, which correspond to C30 classification. Although the added categories are not of relevance in terms of construction output, the BEA's price deflators could potentially be used in conjunction with C30 Value of Construction Put in Place data to yield a constant-dollar output time series.

The other agency that produces price deflators is the BLS. BLS has recently developed producer price indices for the construction industry that are based on prototypes of buildings. The appropriate price deflators contain the contractor's profit, materials costs, and labor costs. Producer price indices are produced for the following new building construction categories: (1) New industrial building construction (NAICS 236211), (2) New warehouse building construction, (3) New school building construction (NAICS 236222), and (4) New office building construction (NAICS 236223). Producer price indices are also produced for four types of special trades in the nonresidential setting: (1) Concrete contractors (NAICS 23811X), (2) Roofing contractors (NAICS 23816X), (3) Electrical contractors (NAICS 23821X), and (4) Plumbing/HVAC contractors (NAICS 23822X).

The PPIs are available for new industrial building construction starting from June 2007. The PPIs for new warehouse building construction are available from December 2004. For new school building construction, the PPIs are available from December 2005. And for new office building construction, June 2006 was the first time the PPIs are available. All of the newly available PPIs for nonresidential structures have been incorporated in BEA's estimates of investments in private structures.¹⁵³ For nonresidential structure types for which PPIs are not available, the BEA combines an input cost index with an output cost index to capture some of the productivity and quality changes in the industry and the costs for a particular building type.¹⁵⁴

¹⁵³ Paul R. Lally, "How BEA Accounts for Investment in Private Structures." *Op. cit.*

¹⁵⁴ *Ibid.*

Using hospitals as an example, the BEA uses Census Bureau's single-family houses under construction index along with the Turner Construction Company building cost index. BEA and BLS may collaborate further to develop other nonresidential building construction indices, such as price indices for highways, hospitals, retail, communication, power, and lodging structures.¹⁵⁵ For the four special trades groups, the PPIs are available starting from December of 2007.

These PPIs correspond to NAICS categories, which is the basis of the Economic Census. However, the Economic Census is conducted every five years, and the PPIs have been available only since 2004. Coupling the output data from the Economic Census and the BLS PPIs would require longer time series than what is currently available.

A.1.5 Labor Input

A.1.5.1 Hours (Production Workers)

The BLS Current Employment Statistics (CES) program is used as the primary source of industry employment and hours data. The data are collected monthly and the employment levels are benchmarked yearly using data from State unemployment insurance programs, which covers about 98 % of all nonfarm employees.¹⁵⁶ The classification of industries in this survey is the NAICS system since 2003, and the historic data were classified by the SIC system. The CES reports the number of all employees, the number of production workers, the number of women workers, the average weekly hours of production workers, the average weekly earnings of production workers, and the average weekly hours of overtime of production workers. In the case of the construction industry, the production workers include “workers, up through the level of working supervisors, who are engaged directly in a construction project, either at the site or in shops or yards, at jobs ordinarily performed by members of construction trades.”¹⁵⁷ For non-production workers, which are executive and managerial personnel, professional and technical employees, and workers in routine office jobs, only employment data is available from this survey. Note that the numbers of jobs are counted, not persons in the CES program. The hours are hours paid, not hours at work. Work hours of non-production workers are not collected, and therefore would need to be estimated.

A.1.5.2 Conversion from Hours Paid to Hours Worked

One disadvantage of the CES data set is that hours paid, instead of hours worked are reported. Hours paid include vacation, paid sick leave, and holidays, in addition to hours worked. Hours at work includes paid time for traveling between jobs sites, coffee breaks, and machine

¹⁵⁵ *Ibid.*

¹⁵⁶ Bureau of Labor Statistics, *Handbook of Methods*. *Op. cit.*

¹⁵⁷ *Ibid.*

downtime.¹⁵⁸ While some of the hours at work do not increase productivity, others do. One such example is activities that motivate workers and reduce shirking. Paid leave is best considered a benefit.¹⁵⁹ BLS has been collecting annual establishment level data on actual hours worked for production and nonsupervisory workers (Hours-at-Work Survey) since 1981. Data from the Hours at Work Survey are used to derive ratios of hours at work to hours paid. This is done for 1-digit Standard Industrial Classification (SIC) industry groups on an annual basis. These ratios are then used to convert hours paid data from the establishment survey. Hours-at-Work Survey by itself, however, is not detailed enough to be used in industry-level productivity measurement. The BLS terminated the Hours at Work Survey in 2000, and replaced the HWS with National Compensation Survey. The Employment Cost Index (ECI) from the National Compensation Survey is used to convert hours paid to hours worked. The Hours at Work Survey had a few limitations. Eventually, because of stringent data reporting requirements, the response rate decreased to the point where not enough data were usable. The ECI was designed to capture the hourly cost of wages and benefits, including paid leave. Ratios of hours at work to hours paid can also be constructed using the ECI data. These ratios are calculated using the ECI data since 2001. For the years before 2001, the ratios are based on Hours at Work Survey. The HWS survey included production and nonsupervisory workers in nonagricultural establishments. The National Compensation Survey, on the other hand, covers all workers. Another advantage of the National Compensation Survey is that it contains a bigger sample. The sample size is 37000 occupations within 8500 private establishments whereas the HWS sampled fewer than 6000 establishments. The response rate associated with the NCS is also higher than that of the HWS.

A.1.5.3 Hours (Nonproduction Workers)

The BLS Current Employment Statistics (CES) Survey reports the number of all employees, the number of production workers, the number of women workers, the average weekly hours of production workers, the average weekly earnings of production workers, and the average weekly hours of overtime of production workers. What the CES survey does not collect is the average weekly hours of supervisory and professional workers. For the non-production workers, only employment data are available from this survey.

The Current Population Survey (CPS) collects data on hours worked. And it is used by the BLS in its productivity program to derive annual ratios of supervisory (or nonproduction) worker average weekly hours to nonsupervisory (or production) worker average weekly hours, and subsequently nonproduction worker hours. The CPS asks respondents for their occupation and

¹⁵⁸ Bureau of Labor Statistics, *Handbook of Methods*. *Op. cit.*

¹⁵⁹ Lucy P. Eldridge, Marilyn E. Manser, and Phyllis Flohr Otto, "Alternative Measures of Supervisory Employee Hours and Productivity Growth." *Op. cit.*

employment status. The information on occupation and employment status is used to sort the data into supervisory (nonproduction) and nonsupervisory (production) categories.¹⁶⁰

Ratios of supervisory (nonproduction) worker average weekly hours to nonsupervisory (production) worker average weekly hours are calculated. These ratios are multiplied by nonsupervisory (production) worker average weekly hours from the CES. Note the hours data from the CES are for hours paid (rather than hours worked), and therefore some discrepancy is introduced. The ratios between hours worked and hours paid are available at major sector level (from the National Compensation Survey), but not available at detailed industry level. The resultant number, supervisory (nonproduction) worker average weekly hours, is then multiplied by the number of supervisory workers to yield total supervisory worker weekly hours. Total supervisory worker hours are obtained by multiplying total supervisory worker weekly hours by 52. Total supervisory worker hours are then combined with total nonsupervisory worker hours from CES and total self-employed hours and unpaid family worker hours to yield total hours for an industry.

A.1.5.4 Self-Employment

The Current Population Survey (CPS) is used to supplement the CES, for data on proprietors and unpaid family workers. Self-employed individuals are not included in the CES. This is particularly a concern for the construction industry where a large proportion of the workers are self-employed. Starting in 1994, the CPS collects monthly data on employment and hours for primary job and all other jobs separately.¹⁶¹ In contrast, prior to 1994, CPS reports hours worked for all jobs a person holds, but only the primary job is recorded.

The CPS is based on Census Bureau's Industry Classification System (ICS). The CPS currently uses 2002 Census occupational classification and the 2007 Census industry classification. These are derived from the 2000 Standard Occupational Classification (SOC) and the 2007 North American Industry Classification System (NAICS). Crosswalks are available on BLS's website to link census classification systems with SOC and NAICS. One limitation of using the Current Population Survey to obtain information on the self employed is the sample size. Coding of industries and reporting are more accurate in establishment level surveys compared to household surveys. For this reason, data from the Current Employment Statistics program should be used

¹⁶⁰ Bureau of Labor Statistics, *Construction of Average Weekly Hours for Supervisory and Nonproduction Wage and Salary Workers in Detailed Industries*, Technical Note. (Washington D.C.: Bureau of Labor Statistics, 2005).

¹⁶¹ Bureau of Labor Statistics, *Construction of Employment and Hours for Self-employed and Other Nonfarm Workers and for all Farm Workers, Using Current Population Survey Data for Primary and Secondary Jobs*, Technical Note. (Washington D.C.: Bureau of Labor Statistics, 2006).

as a primary source of data, and data from the Current Population Survey should be used as a supplemental source of data.¹⁶²

In a BLS study, self-employed were excluded from the analysis of productivity. The reason was that proprietors' incomes include both returns from labor and capital.¹⁶³ It is difficult to separate the income into these two components. The more recent BLS approach¹⁶⁴ is similar to the Jorgenson *et al.* approach.¹⁶⁵ First, it is assumed that the self-employed and unpaid family workers are paid the same hourly wages as employees with similar characteristics in the same sector. It is also assumed that the noncorporate rate of return is the same as after-tax rates of return for corporate businesses. These two rates are then adjusted proportionately such that the reported proprietor's income matches with the sum of labor income and noncorporate income.

A.1.5.5 Labor Costs

In labor productivity calculations, labor input is simply expressed in hours. For multifactor productivity calculations, various inputs are combined using corresponding costs as weights. Since the price of labor includes both wage and benefits, from the producer's point of view, both wage and benefits should be included in labor costs. BLS includes in the labor compensation wages, salaries, supplemental payments, including employer's contribution to social security, unemployment insurance taxes, and payments for health insurance and pension plans.¹⁶⁶ Supplemental payments also include paid leave, such as vacation and holiday leave. The labor compensation data come from the National Income and Product Account (NIPA) developed by the BEA and is based on establishment-level data. Labor compensation for proprietors and unpaid family workers needs to be estimated. Real compensation per hour are calculated using Consumer Price Index.¹⁶⁷ Labor costs also include a portion of noncorporate income.¹⁶⁸

A.1.5.6 Labor Quality

Levels of labor quality are not distinguished in labor productivity calculations. However, accounting for labor quality is important in multifactor productivity. Jorgenson presented

¹⁶² Lucy P. Eldridge, Marilyn E. Manser, and Phyllis Flohr Otto, "Alternative Measures of Supervisory Employee Hours and Productivity Growth," *Op. cit.*

¹⁶³ Bureau of Labor Statistics, *Labor Composition and U.S. Productivity Growth, 1948-90*, *Op. cit.*

¹⁶⁴ Bureau of Labor Statistics, *Technical Information About the BLS Multifactor Productivity*, *Op. cit.*

¹⁶⁵ Dale W. Jorgenson, *Productivity, Volume 1: Postwar U.S. Economic Growth* (Cambridge, MA and London: MIT Press, 1995).

¹⁶⁶ Bureau of Labor Statistics, *Handbook of Methods*, *Op. cit.*

¹⁶⁷ *Ibid.*

¹⁶⁸ *Ibid.*

evidence suggesting that the assumption that labor and capital inputs are homogenous is not valid.¹⁶⁹ For instance, for the period between 1947 and 1985, more than a third of labor input growth is from the growth of labor quality. It is important to take into account varying levels of labor quality. Jorgenson *et al.*, for instance, categorizes labor into two gender, eight age, five education, two employment status (employed and self-employed), and ten occupation categories.¹⁷⁰ In their approach, labor input growth is a function of growth in labor hours, as well as growth in labor quality. For each industry, a price matrix is established using labor categorizations. Characteristics of labor input include gender (2 groups), age (8 groups), employment class (2 groups), occupation (10 groups), education (5 groups), and industry (51 groups). This matrix contains 81 600 cells, which is the product of the numbers of groups for each labor input characteristic. Each cell of this price matrix would be populated with the corresponding labor compensation of the particular labor category. Similarly, a quantity matrix is established with each cell of the matrix being populated with hours worked by labor in a particular category. With these two matrices, labor input can be obtained by summing labor inputs of various categories with the corresponding weights, where the weights are the average value shares of the two periods in consideration.

BLS cross classifies the hours of workers by different schooling levels, gender, and age for its multifactor productivity measures, only at the major sector level. In contrast, for multifactor productivity measures at a less aggregated level, such as the manufacturing industries, labor input is simply a sum of all hours. For a few detailed industries, the number of employees is used as the labor input measure.¹⁷¹ Similarly, labor productivity is an output per hour measure and is calculated assuming all hours are homogenous.

BLS labor classification used to include experience. The recent removal of experience from labor classification is consistent with Jorgenson *et al.*¹⁷², who assume that experience is implicitly included through data on education and age.¹⁷³ The hours at work by the different types of workers are weighted and aggregated using an annually chained (Tornqvist) index. An earlier approach BLS used regarding labor costs was to estimate earnings using data from

¹⁶⁹ Dale W. Jorgenson, "Productivity and Economic Growth," in *Fifty Years of Economic Measurement—the Jubilee of the Conference on Research in Income and Wealth*. *Op. cit.*

¹⁷⁰ Dale W. Jorgenson, F. M. Gollop, and B. M. Fraumeni, *Productivity and U.S. Economic Growth*. *Op. cit.*

¹⁷¹ Edwin R. Dean, Michael J. Harper, and Mark S. Sherwood, "Productivity Measurement with Changing-Weight Indices of Outputs and Inputs," *Op. cit.*

¹⁷² Dale W. Jorgenson, Mun S. Ho, and Kevin J. Stiroh, *Productivity Volume 3: Information Technology and the American Growth Resurgence*. *Op. cit.*

¹⁷³ Bureau of Labor Statistics, "Changes in the Composition of Labor for BLS Multifactor Productivity Measures, 2007." Technical Report. (Washington D.C.: Bureau of Labor Statistics, 2009).

education, estimated experience, and other characteristics. The recent approach is to simplify the procedure by using actual earnings.

An alternative approach utilized the Longitudinal Employer-Household Dynamics database to estimate human capital.¹⁷⁴ Dale Jorgenson commented that this methodology is “taking productivity analysis to the next level.”

A.1.6 Capital Input

This section is based largely on Harper.¹⁷⁵ Capital could include anything that is costly to obtain at the present, but it earns return in the future.¹⁷⁶ Capital therefore could include equipment, structures, land, inventories, financial assets, human capital, and intangibles such as software development, advertising costs, or organizational efforts.¹⁷⁷ In productivity studies, only equipment, structures, land, and inventories are accounted for as capital inputs. Although data on financial assets exist, they are not included in capital input calculations because it is difficult to link decisions about financial assets with production decisions.¹⁷⁸ While intangible assets play direct roles in production, they are excluded because it is hard to quantify their service flows.¹⁷⁹

Property income of capital is defined as nominal revenues minus expenses for variable inputs. It represents the return of the capital to the investor who made the capital investment. It also represents the nominal cost paid by the production manager to the investor for the use of the capital. Property income of capital is readily available in firm’s accounting records.¹⁸⁰

To construct capital input, the first step is to use the perpetual inventory method to convert investment data into capital stocks. Capital stocks then are combined with property income data

¹⁷⁴ LEHD, “A Layman’s Guide to the LEHD Human Capital Measures” Longitudinal Employer-Household Dynamics Informational Document No. ID-2003-04. 2003; John M. Abdowd, Paul A. Lengermann, and Kevin L. McKinney. “The Measurement of Human Capital in the U.S. Economy,” Longitudinal Employer Household Dynamics. Technical Paper No. TP-2002-09. 2002

¹⁷⁵ Michael J. Harper, “Estimating Capital Inputs for Productivity Measurement: An Overview of U.S. Concepts and Methods,” *International Statistical Review* 67(1999):327-337. 1999.

¹⁷⁶ *Ibid.*

¹⁷⁷ *Ibid.*

¹⁷⁸ *Ibid.*

¹⁷⁹ *Ibid.*

¹⁸⁰ *Ibid.*

to derive rental prices. These rental prices are then used as weights to aggregate capital services from different assets into a capital input index.¹⁸¹

The BEA uses BLS producer price indices as a basis to deflate nominal investments to yield real investments. The price indices used incorporate quality change such that investments of higher quality are treated as being higher in quantity, while the quality is kept constant. BLS then uses an age/efficiency function to weight real investments by age/efficiency, and weighted real investments are aggregated by asset types. The productive capital stock $K_{i,t}$ at time t for the i^{th} type of capital asset is a sum of past investments, $I(i, t - \tau)$, of asset type i and age τ , weighted by the age/efficiency function, S_τ :

$$K_{i,t} = \sum_{\tau=0}^{\infty} S_\tau I(i, t - \tau)$$

The age/efficiency function, S_τ , used in BLS productivity program is of the following functional form:

$$S_\tau = \frac{L - \tau}{L - B\tau}$$

Where L is service life of the asset, τ is the age of the asset, and B is a parameter. B is assumed to be 0.5 for equipment and 0.75 for structures. The age/efficiency profile is based on empirical evidence when it is available. However, such information tends to be limited. Note that the vintage aggregation is based on efficiency of the asset rather than its value.

Property income, Ψ_t , is the total rent from different assets at time t . That is,

$$\Psi_t = \sum_i c_{i,t} K_{i,t}$$

where $K_{i,t}$ is the productive stock of the i^{th} asset and $c_{i,t}$ is the rental price. The rental price can be written as the following equation if the price of the asset is assumed to be the discounted sum of all future rents.

$$c_{i,t} = p_{i,t} r_t + p_{i,t} \delta_{i,t} - (p_{i,t} - p_{i,t-1})$$

where $\delta_{i,t}$ is the rate of depreciation, r_t is the discount rate, and $p_{i,t}$ is the price. The rate of depreciation is derived from the age/price profile that corresponds to the age/efficiency profile used to aggregate assets of different vintages earlier. With data on property income, productive capital stock, the rate of depreciation, the two equations above are used to estimate the rate of return, r_t and thus $c_{i,t}$, the rental price. The rental price is implicit and needs to be estimated

¹⁸¹ Dale W. Jorgenson, Mun S. Ho, and Kevin J. Stiroh, Productivity Volume 3: Information Technology and the American Growth Resurgence, *Op. cit.*

because companies often own the capital they use and there is no actual transactions that take place.

Property taxes for specific asset types are then added to the rental price, and a multiplier is created to take into account corporate income tax, depreciation deductions, and credits. This procedure takes into account the different effective tax rates on different types of assets. For instance, the effective cost for equipment use is lower than that of structures due to investment tax credit for equipment and possible depreciation deductions for equipment over very short periods of time.

Productive stocks of different asset types are then aggregated for each industry, using the rental price shares as weights. These aggregated productive stocks then constitute capital input for the industry. For the productivity measurement at the major sector level, such as the private business sector, then the industry-level capital inputs are aggregated using relative capital income as weights.

Similar to labor, capital quality is accounted for in capital input calculation in the framework of Jorgenson *et al.*¹⁸² Capital is broken down by class of asset and legal form of organization. Capital stock at any time point is the sum of weighted past investments. The weights represent relative efficiencies of capital due to age differences. The cross-classification, however, was not done by industry.

BLS classifies capital assets into 42 types for equipment, 21 types for nonresidential structures, 9 types for residential structures, 3 types for inventories (by stages of processing), and land.¹⁸³ Notably, information processing equipment and software is included under the equipment category.

A.1.7 Intermediate Inputs

Intermediate inputs include energy, materials, and purchased business services inputs. These data are available from BEA's input output tables. The role of intermediate inputs becomes more important when the focus is on a more disaggregated industry level.¹⁸⁴ Intermediate inputs are constructed only for manufacturing industries in the BLS productivity program as manufacturing as a whole and 18 3-digit NAICS manufacturing industries are the only industries for which KLEMS productivity measures are published.¹⁸⁵ Data for energy input come from price and quantity of fuels used for heat or power. For the productivity calculation of

¹⁸² Dale W. Jorgenson, F. M. Gollop, and B. M. Fraumeni, *Productivity and U.S. Economic Growth*, *Op. cit.*

¹⁸³ Bureau of Labor Statistics, *Technical Information About the BLS Multifactor Productivity*, *Op. cit.*

¹⁸⁴ *Ibid.*

¹⁸⁵ *Ibid.*

manufacturing industries, materials are non-energy inputs but include fuel-type materials that are used as raw materials in manufacturing. Purchased business services are purchased services from service industries by manufacturing industries. Costs associated with intermediate input purchases should include commodity taxes because they are paid by the producer.¹⁸⁶

The present input-output framework is designed to track material flows. Data is limited regarding contracted labor services and leased capital equipments.¹⁸⁷ As regards to materials flows, data tend to be incomplete outside of the manufacturing industries.¹⁸⁸

¹⁸⁶ W. Erwin Diewert, "Which (Old) Ideas on Productivity Measurement Are Ready to Use?" in *New Developments in Productivity Analysis. Op. cit.*

¹⁸⁷ *Ibid.*

¹⁸⁸ *Ibid.*

A.2 Aggregation Methods

When choosing an aggregation method, it is best to choose one that uses weights that change over time. When fixed weights are used, the quality of estimates that are closer to the base year is generally high, whereas estimates that are further from the base year are likely to be error prone.¹⁸⁹ For instance, when the price of capital increases, the quantity of capital decreases relative to labor. The value share of capital may increase, decrease, or stay constant, but the fixed-weight approach dictates the value share to be constant.¹⁹⁰ Fisher Ideal¹⁹¹ and the Tornqvist indices are aggregated using weights that can change over time. Diewert shows that “certain index number formulas, which he coined “superlative,” such as the Tornqvist and the Fisher Ideal, are consistent with flexible production functions.”¹⁹² “Aggregation methods that use fixed weights are consistent with a more restrictive production function.” Diewert shows that “chained time series of superlative index numbers are approximately consistent.”¹⁹³ On theoretical grounds, Fisher Ideal and Tornqvist indices are both good choices. There is also little difference between these indices in practical applications.¹⁹⁴ As Dumagan shows, the Tornqvist index and the Fisher ideal index numerically approximate each other.¹⁹⁵ As the Tornqvist index requires less data to calculate, it may be more practical to use.¹⁹⁶ Although more researchers prefer the Tornqvist index, there is no strong reason to prefer the Tornqvist index or the Fisher ideal index.¹⁹⁷ BLS uses the Tornqvist index in labor productivity calculations, while the BEA uses the Fisher ideal index for chain-type indices.

¹⁸⁹ William Gullickson, “Measurement of Productivity Growth in U.S. Manufacturing,” *Op. cit.*

¹⁹⁰ Edwin R. Dean, Michael J. Harper, and Mark S. Sherwood, “Productivity Measurement with Changing-Weight Indices of Outputs and Inputs” In *Industry Productivity: International Comparison and Measurement Issues*, OECD (Paris: OECD, 1996).

¹⁹¹ For more information on Fisher Ideal Indices, please see R.D. Rossiter, “Fisher Ideal Indices in the National Income and Product Accounts,” *Journal of Economic Education* Fall (2000): 363-373.

¹⁹² W. Erwin Diewert, “Exact and Superlative Index Numbers,” *Journal of Econometrics* vol 4, no. 4 (1976): 115-45.

¹⁹³ W. Erwin Diewert, “Superlative Index Numbers and Consistency in Aggregation,” *Econometrica* July (1978): 883-900.

¹⁹⁴ William Gullickson, “Measurement of Productivity Growth in U.S. Manufacturing,” *Op. cit.*

¹⁹⁵ Jesus C. Dumagan, “Comparing the superlative Tornqvist and Fisher ideal indices,” *Economic Letters* 76(2002): 251-258.

¹⁹⁶ *Ibid.*

¹⁹⁷ Edwin R. Dean, Michael J. Harper, and Mark S. Sherwood, “Productivity Measurement with Changing-Weight Indices of Outputs and Inputs,” *Op. cit.*

The Tornqvist index can be used to aggregate different types of outputs. It is also the approach used to individually aggregate different types of labor inputs, capital inputs, or intermediate inputs. When different types of labor are aggregated, the weights used in the Tornqvist index formula are relative shares of labor compensation. For aggregation of different capital assets, relative shares of capital income the assets generate are used as weights. Again when finally calculating the productivity, all inputs are aggregated using the Tornqvist index, with weights being each input's share of total costs. Generally it is desirable to use the most disaggregated data and then aggregate different components to a more aggregate level.

A.3 Overview of BLS Productivity Program

BLS does not publish labor productivity or multifactor productivity measures for the construction industry. BLS produces labor productivity for business, private nonfarm business, manufacturing (total, durable, and nondurable sectors), and nonfinancial corporations. Labor productivity is available also for over 400 selected industries in manufacturing, mining, utilities, wholesale and retail trade, and services.

BLS has two multifactor productivity programs—the Major Sector Multifactor Productivity program and the Industry Multifactor Productivity program. In the Major Sector Multifactor Productivity program, the BLS publishes multifactor productivity the private business sector, the private nonfarm business sector, the aggregate manufacturing sector, and 18 3-digit NAICS manufacturing industries and the utility and gas industry. The productivity measures for the private business sector and the private nonfarm business sector are based on value added output, and labor and capital inputs. For the aggregate manufacturing sector and the 18 3-digit NAICS manufacturing industries, gross output, KLEMS inputs (capital, labor, energy, materials, and purchased business services) are used.

In its Industry Multifactor Productivity program, the BLS publishes multifactor productivity for 86 4-digit NAICS manufacturing industries, air transportation, and railroad transportation. For these industries, inputs include employee hours, capital services, and intermediate purchases.

The BLS also produces multifactor productivity for manufacturing industries of U.S., France, and Germany for comparison of productivity trends. These measures are based on value-added and labor and capital inputs.

A.4 Classification Issues

There are two types of classification issues. One issue concerns the different classification systems used in different datasets. For instance, the Economic Census and the Current Employment Statistics are both establishment-based surveys, and they are organized using SIC and NAICS systems. Census Bureau's Value of Construction Put in Place survey uses projects as units of data collection. To create labor productivity measures using output data from the

Value of Construction Put in Place is difficult because labor hours data are collected using NAICS/SIC, but not organized by project types.

The other classification issue concerns the change of classification systems within a dataset. The most prominent example of this issue is the change of industry classification system from SIC (Standard Industrial Classification) system to NAICS (North American Industry Classification System) in 1997. Under the SIC system, establishments were mainly classified by product or activity types, but in some instances, end use, raw materials, or market structure was as the basis for classification. The classification system of SIC was not consistent. NAICS was devised to incorporate new industries that were not covered under SIC and also to provide a consistent framework for classification. With the rapid changes in the composition of the economy, a new classification was needed to accommodate the new and evolving economy. The NAICS system classifies industries by their production processes, as opposed to final products.

Some of the categories of construction under the two classification systems appear to be similar. One such example is SIC 152 General Building Contractors-Residential and NAICS 2361 Residential Building Construction. However, these two categories are not completely comparable due to rule changes, such as the treatment of auxiliary units.¹⁹⁸ Efforts have been made to concord or bridge the two classification systems and reclassify older data using the new classification system.

For instance, the microdata of the Current Employment Statistics survey from March 2001 were coded in both SIC and NAICS.¹⁹⁹ For the data from March 2001, 97.2 % of the employment of the construction industry under NAICS can be classified under the construction industry in the SIC system. A small percent (0.3 %) of the employment of the construction industry falls under mining in the SIC system.²⁰⁰ Some of the employment (1.3 % and 1.2 %, respectively) fall under Finance, insurance, and real estate, and services in the SIC system. These ratios were used to reconstruct historic data.

In the case of the Economic Census, establishments surveyed in 1997 were coded both with SIC and NAICS, and bridge tables were developed based on the 1997 data. The bridge tables list the NAICS codes and the corresponding SIC codes. Since the matches are not always exact, the tables also list the proportions of total SIC sales, receipts, or value of shipment under particular SIC codes that are under specific NAICS codes. Using these proportions to “translate” NAICS codes into SIC codes, or vice versa, could cause problems.

¹⁹⁸ Teresa L. Morisi, “Recent Changes in the National Current Employment Statistics Survey,” *Monthly Labor Review* June (2003): 3-13.

¹⁹⁹ *Ibid.*

²⁰⁰ *Ibid.*

For cases where the SIC code does not translate into a NAICS code, one approach is to directly look at micro-level data. If an establishment is surveyed in 1997, then its 5-digit identification number and its SIC and NAICS assignments can be used to assign NAICS codes to the older data associated with this particular establishment. This is possible because establishments surveyed in 1997 were assigned both a SIC code and a NAICS code.

If establishments cannot be classified with a NAICS code using the above approaches, then they could be assigned NAICS codes by following the procedure outlined below using data from the 1997 Economic Census. Information on the characteristics, such as shipments per worker or hourly wages (in the case of the manufacturing industry), can be used to derive probabilities of specific NAICS code assignment. A NAICS code can then be drawn from the distribution.²⁰¹ Another example is Klimek and Merrell,²⁰² who used 1997 Economic Census data on retail and wholesale industries and established proportions of establishments originally assigned a SIC code that are assigned to a NAICS code. Using these proportions, the authors constructed a distribution from which NAICS codes are randomly drawn and assigned to individual establishments. The newly NAICS coded data were then used to produce aggregate data. This latter approach was shown to be reasonable as two thirds of the establishments that required random assignment were cases where over 90 % of the establishments in 1997 were coded into a single NAICS industry. The authors also suggested that multiple random assignments can be done and used to generate standard errors. Another possibility is to use a firm's NAICS code to assign to its associated establishments.

²⁰¹ Kimberly N. Bayard and Shawn D. Klimek, "Creating a Historical Bridge for Manufacturing Between the Standard Industrial Classification System and the North American Industry Classification System," Proceedings of the American Statistical Association, Business and Economic Statistics Section [CD-ROM] (2004), pp. 478-84.

²⁰² Shawn D. Klimek and David R. Merrell, "On Reclassifying Industries from the Standard Industrial Classification System to the North American Industry Classification System," presented at the Second International Conference on Establishment Surveys, Buffalo, New York, 2000.

Appendix B Productivity and Competitiveness: An Annotated Bibliography

This annotated bibliography summarizes key documents in productivity measurement with an emphasis on the construction industry. Challenges in construction productivity measurement have been recognized for many decades. While some aspects of construction productivity measurement have received attention, and notable improvements have been made, many fundamental challenges exist. This annotated bibliography provides a focused view of the state of knowledge and, for researchers and practitioners, an updated and centralized source of documents in this area.

This annotated bibliography contains three sections. Section B.1 focuses on documents with a specific focus on the productivity, its measurement, and other related issues in the construction industry. Some of the measurement issues are deflators, quality adjustments of output, the definition of what constitutes the construction industry, and the underground economy. Most studies on construction productivity focus either on task-level productivity or industry-level productivity. Metrics used include labor productivity, multifactor productivity, and direct work rate. There are, however, different definitions for each of these metrics in the literature. Some of the causes for differences of changes in construction productivity are economies of scale, labor quality, capital-labor ratio, changes in output mix, and institutional issues (prevailing wage laws, unions, and collective bargaining). This section also includes studies from other countries, such as the United Kingdom and Canada.

Section B.2 is titled “construction data.” It includes a document from the Construction Statistics Data Users’ Conference in 1997, published by the Census Bureau. This document discusses governmental statistics on construction and on how the data are collected and reported.

Section B.3 contains studies that are on the general topic of productivity methods and measurement. It includes many documents published by the Bureau of Labor Statistics. It also includes documents from the OECD and academic sources. Methods and challenges on measurement of output, deflators, capital, labor, and data quality are presented.

B.1 Construction Productivity and Related Issues

Allen, Steve G. “Unionized Construction Workers are More Productive.” *Quarterly Journal of Economics* 99, no. 2 (1984): 251-274.

This is an empirical paper that shows that unionized workers are more productive, controlling for capital-labor ratio, capital recentness, measurable labor quality, scale of production, industry, region, and interstate price differences. Productivity is defined as value added per worker. The factors that contribute to higher productivity levels among unionized workers may include better training, reduced use of unskilled labor, lower foreman to journeyman ratios, reduced recruiting and screening costs, and greater managerial ability.

Allen, Steve G. “Why Construction Industry Productivity is Declining.” *Review of Economics and Statistics* 67, no. 4 (1985): 661-669.

This paper studies the sources of construction productivity change. Value added per employee deflated by the Dodge cost index was regressed on capital per employee, labor hours per establishment (economy of scale), labor quality, union, region variables, and building types. Data used was at the state level for 1972 and 1977. The coefficients from the regression model were then combined with data from these two years to yield percent productivity change due to the various factors. The total predicted productivity change from the regression was -8.8 %. The shift of output mix from commercial, industrial, and institutional projects to residential projects resulted in a reduction in skilled labor, and this was the most important factor that contributed to the decline in productivity.

Alternative deflators were also devised. For instance, a deflator for nonresidential building construction was calculated using the difference between the rate of change of value put in place and the rate of change of square footage put in place. Adjusting for bias in the deflators accounts for -10.5 percentage points in reported productivity, which was -21.4 %. The predicted change from the regression, together with the adjustment of the deflators, therefore can explain 92 % of the productivity change.

Allen, Steven G. “Why Construction Industry Productivity is Declining: Reply.” *Review of Economics and Statistics* 71, no. 3 (1989): 547-549.

Allen responds to Pieper’s (1989) comments. The capital-labor ratio was shown to be decreasing in the original paper, but Pieper showed that it was increasing. The sources of data and the assumptions contribute to this discrepancy. Allen made several adjustments and reported that 56.5 % of the observed decline in productivity can be explained, instead of 92 % as in the original paper.

Allmon, E., C. T. Hass, J. D. Borcharding, and P. M. Goodrum. “U.S. Construction Labor Productivity Trends, 1970-1998.” *Journal of Construction Engineering and Management* 126, no. 2 (2000): 97-104.

This study focuses on task-level productivity. Unit labor costs in constant dollars and daily output factors were obtained from Means cost manuals, for tasks such as hand trenching, welded steel pipe installation, ceiling tile installation, and compaction with a sheepsfoot roller, over three decades. The tasks were chosen such that tasks that are impacted by varying degrees of technology improvement are included in the study. The daily output increased for most of the tasks, and the unit labor costs decreased in real terms for all tasks. The two main reasons for the increase in productivity are low wages and technology improvement. Time use data from 72 projects in Austin, Texas over 25 years were also studied. It was found that direct work rate is positively correlated with construction productivity.

Azari-Rad, Hamid, Peter Philips, and Mark J. Prus, eds. *The Economics of Prevailing Wage Laws*. Hampshire, England: Ashgate Publishing Limited, 2005.

This book is a collection of chapters that examine the prevailing wage laws and how they affect various aspects of the American construction industry. The book presents a history of prevailing wage laws and an overview of the construction industry. The underlying vision of the prevailing wage laws is a society where labor is highly skilled, highly paid, and the industries are capital intensive and utilize advanced technologies. The main thesis of the book is that prevailing wage laws solve a free-rider problem and they allow long-term costs to be paid, such as costs on training, safety, insurance, and pensions, despite the short-term nature of projects. These arguments are supported by empirical evidence based on heterogeneity in prevailing wage laws across states and time. The discussion on productivity is limited.

Baily, Martin Neil and Robert J. Gordon. “The Productivity Slowdown, Measurement Issues, and the Explosion of Computer Power.” *Brookings Papers on Economic Activity*, no. 2 (1988): 347-431.

This paper examines the source of U.S. productivity slowdown after 1973. There is a section that is devoted to construction productivity. The average annual growth rate in terms of GDP per hour in construction was estimated to be between -1.67 % to -1.99 % in the period of 1973-1987. Trends of output and inputs are examined, and the paper documents the “implausibility” of the data. This paper indicates data problems and the need for better data collection, particularly on output. Included at the end of the paper is a discussion by William Nordhaus and David Romer.

Bosch, Gerhard and Peter Philips, eds. *Building Chaos—An International Comparison of Deregulation in the Construction Industry*. London: Routledge, 2003.

This book contains case studies of 9 countries—the Netherlands, Germany, Denmark, Canada (Province of Quebec), Australia, Spain, the United States, the United Kingdom, and Republic of Korea, in descending order of the level of regulation of their construction labor markets. The construction industry is intrinsically volatile. Because construction projects are neither storable nor transportable, the industry is particularly vulnerable to economic downturns. It was shown that construction industries in these countries follow two paths of development. On one path of development, long term costs are paid for. These long term costs include training of workers, health insurance, retirement, compensation for instability of the industry, and development and use of advanced technologies. This model of development is capital intensive, human capital intensive, and “technically dynamic.” Productivity tends to be high in construction industries that are on the technically dynamic path. The other model of development is characterized by a free-rider problem. The long term costs are not paid for because there is no legal requirement or because there are no arrangements made between contractors and organized labor. Labor quality tends to be low and not well equipped, and labor intensity tends to be high.

In the case of the United States, in some regions organized unions and organized employer associations engage in collective bargaining and develop agreements for apprenticeship programs, health insurance, and retirement. In other regions, no such agreements exist.

Canadian Construction Innovation Council. *Measuring the Performance of the Canadian Construction Industry: Metrics*. Ontario: Canadian Construction Innovation Council, 2006.

This is an initial document of the Canadian Construction Innovation Council's effort to assess the performance of the Canadian construction industry. This document includes a literature review of benchmarking efforts and a preliminary set of metrics.

Canadian Construction Innovation Council. *Measuring the Performance of the Canadian Construction Industry: Pilot Project Final Report*. Ontario, 2007.

This report is the follow-up report of the Metrics report dated 2006 (see above). This document reports the findings of the pilot study where metrics of performance were applied to 37 projects, including buildings and water and wastewater piping systems. It was pointed out that only using productivity measurements may not capture a complete picture of the performance. In addition, the industry prefers descriptive measures. The metrics system used is based on benchmarking programs of the Construction Industry Institute in the U.S. and the Movement for Innovation (M4I) in the United Kingdom. Project and organizational performance metrics are the focus, and aggregation from the project or organizational level to industry level is feasible. A project timeline with 6 phases is defined, along with metrics in costs, time, quality, safety, scope, innovation, and sustainability. The results are presented using radar charts, box-and-whisker plots, and cumulative distribution curves.

Centre for the Study of Living Standards. *Productivity Trends in the Construction Sector in Canada: A Case of Lagging Technical Progress*. CSLS Research Report, no. 2001-3. Ontario, 2001.

This report documents the trends in productivity growth in Canada and uses regression models to explain output per hour in the total construction and residential construction sectors. The independent variables include capital intensity, educational attainment, capital utilization, and the unemployment rate. None of these variables can explain the productivity decline in the sector. Comparing the late 1970s with the late 1990s, capital-labor ratio and educational attainment have increased. While increases in both of these factors are expected to increase output per hour, the observed output per hour declined. After examination of other variables, the report concludes that measurement error and lack of technical progress are the main factors for the observed productivity decline. The section on taxation examines how taxation policy affects companies' decisions to make investments in equipment and workforce. Allocation of these resources could potentially have a great impact on productivity growth.

Crawford, Paul and Bernard Vogl. “Measuring Productivity in the Construction Industry.” *Building Research and Information* 34, no. 3 (2006): 208-219.

This paper provides an overview of methods of productivity measurement and presents data on construction productivity in the UK. It points out that labor productivity in the UK is relatively low compared to the rest of Europe, and it is likely a result of low capital intensity adopted in the UK. This observation is consistent with Bosch and Philips (2003), in which construction industries of nine countries are ranked by levels of regulation and two paths of development are identified. The construction industry of the UK is characterized by low levels of regulation, low labor wages, labor intensive production processes, and limited use of technology. Crawford and Vogl point out that high levels of labor productivity can be achieved at the expense of overall productivity due to suboptimal capital-labor allocations. Therefore, measures of labor productivity do not tell the whole story. There is a need to improve existing data and creating new data for productivity measurement. The paper also points out the need to have measures for the quality of inputs.

Eastman, Charles M. and Rafael Sacks. “Relative Productivity in the AEC Industries in the United States for On-site and Off-site Activities.” *Journal of Construction Engineering and Management* 134, no. 7 (2008): 517-526.

This paper examines on-site and off-site sectors of the construction industry. The authors found that productivity is higher for off-site sectors compared to on-site sectors. Furthermore, off-site sector productivity growth is also higher. Some of the off-site sectors are classified as manufacturing under Census. When construction industry productivity measurement does not properly account for the role of intermediate inputs, such as pre-fabricated construction products, productivity estimates could be biased downward. The empirical evidence presented in this paper highlights the importance of treating intermediate inputs properly in the growth accounting framework. More specifically, the evidence is consistent with the productivity measurement approach where output measure is gross output and all inputs, including intermediate inputs, are treated symmetrically.

Goodrum, Paul and Carl T. Haas. “Partial Factor Productivity and Equipment Technology Change at Activity Level in U.S. Construction Industry.” *Journal of Construction Engineering and Management* 128, no. 6 (2002): 463-472.

This paper examines task-level productivity for 200 activities between 1976 and 1998. More specifically, this paper looks at whether equipment technology enhances labor productivity. The data sources were cost estimating guides, including Means, Richardson, and Dodge. Partial factor productivity is defined to be units of physical output divided by the sum of labor costs and fixed capital costs. The authors concluded that activities that experienced a significant change in equipment technology also experienced a greater improvement in partial factor productivity.

Most of the activities examined experienced improvement in partial factor productivity during the study period.

Goodrum, Paul M., Carl T. Haas, and Robert W. Glover. “The Divergence in Aggregate and Activity Estimates of US Construction Productivity.” *Construction Management and Economics* 20, no. 5 (2002): 415-423.

This paper compares productivity estimates at the industry level and at the task level. While industry-level productivity estimates tend to show declining trends, task-level activity productivity estimates tend to suggest productivity increases. This paper discusses productivity output measures, particularly the construction of output deflators. It lists the types of indices used to deflate different types of construction outputs. For task-level productivity, the measure of labor productivity is defined to be units of physical output divided by work hours, and the measure of multifactor productivity is defined to be units of physical output divided by the sum of deflated labor cost and equipment cost. The data used for task-level productivity calculations are from estimating manuals. Labor productivity and multifactor productivity at the task level were shown to be increasing from 1976 to 1998.

Goodrum, Paul M. and Carl T. Haas. “Long Term Impact of Equipment Technology on Labor Productivity in the U.S. Construction Industry at the Activity Level.” *Journal of Construction Engineering and Management* 130, no. 1 (2004):124-133.

Using cost estimating guides, the authors identified 200 construction activities during 1976 and 1998. Five technology factors were identified: amplification of human energy, level of control, functional range, ergonomics, and information processing. 43 types of hand tools and 31 types of machinery associated with the 200 activities were studied and a technology index was developed. The technology index captures changes in the equipment in terms of technology factors. The technology index was constructed for each activity. Using this approach, the authors found that 107 of the 200 activities increased labor productivity, while 30 activities showed a decline and 63 activities showed no change in labor productivity. Equipment technological advances can therefore explain some of the labor productivity increase during the 22-year period.

Centre for the Study of Living Standards. *Can Measurement Error Explain the Weakness of Productivity in the Canadian Construction Industry?* By Peter Harrison. Centre for the Study of Living Standards Research Report, no. 2007-01. Ontario, 2007.

This report presents productivity trends, reviews productivity literature, presents views of industry practitioners, describes methodology used by Statistics Canada, and discusses possible sources of mismeasurement. This report is an accessible, thorough, and comprehensive resource for construction productivity, with a focus on Canada. Statistics Canada’s productivity measurement methodology is described, with an emphasis on the construction of price deflators. This document also contains discussions on prefabrication and demonstrates that incorporating

more productive prefabrication results in higher productivity if the output measure is gross output. If the output measure is value added, then more productive prefabrication does not result in higher construction productivity. The document also indicates lack of empirical evidence regarding the increasingly important role of prefabrication perceived by industry practitioners.

Haas, Carl T., James T. O'Connor, Richard L. Tucker, Jason A. Eickmann, and Walter R. Fagerlund. *Prefabrication and Preassembly Trends and Effects on the Construction Workforce*. Center for Construction Industry Studies Report No. 14. Austin, Texas, 2000.

Prefabrication and preassembly use in the U.S. is studied in this report using a survey instrument. The survey respondents were 29 managers. The survey results indicate increased usage of prefabrication and preassembly, from 14 % in 1984 to 27 % in 1999, as a fraction of overall project work. The main reasons for using prefabrication and preassembly were schedule, workforce issues, and economic factors. The areas where prefabrication and preassembly are most often used are piping, mechanical, equipment, and structural assembly. In addition to the survey design and the results, this report also includes historic accounts and a literature review.

Hendrickson, Chris. "Discussion of 'Is Construction Labor Productivity Really Declining?' by Eddy M. Rojas and Peerapong Aramvareekul" *Journal of Construction Engineering and Management* 131, no. 2 (2005): 269-270.

Hendrickson discusses the paper by Rojas and Peerapong (2003), who argue that the data quality is so low that it is not possible to conclude whether construction productivity is declining or increasing. Hendrickson uses Census of Construction Industries data from 1982 through 1997 and calculated a productivity measure defined as construction contribution to GDP divided by hours. Construction contribution to the GDP was deflated using the GDP price deflator index. This productivity measure increased from 1982 to 1987 and stayed constant through 1997. It is noted that real wage has declined during this period. The author also notes the lack of difference in trends for input costs, output prices, and general price indices. If there is productivity decline, then we would expect output prices to be increasing at a higher rate than input costs and general price inflation.

This discussion is followed by a closure by Rojas. He argues that the conclusion of moderate improvement in productivity reached by Hendrickson was based on the assumption that the output measure is reliable, but he argues that the output measure is not reliable. Both Hendrickson and Rojas agree there is a need for meaningful measures of productivity.

National Bureau of Standards. *Productivity Measurement for the Construction Industry*. NBS Technical Note no. 1172. Washington, D.C., 1983.

This report describes the measurement of single and total factor productivity. It summarizes the approaches and indicates how they apply to the construction industry. This document concludes with a discussion on data availability and challenges. The lack of appropriate price and cost

indices to convert output values into quantity indices is one obstacle. Another challenge is establishing a quantity index for capital input. It also points out that the definitions of some four-digit construction industries have been changed between Economic Censuses. It would not be possible to construct five-year TFP growth estimates based on Census data for these industries. This report contains an appendix with an annotated bibliography on productivity measurement methods, productivity in the construction industry, and productivity measurement case studies.

National Research Council. *Advancing the Competitiveness and Efficiency of the U.S. Construction Industry*. Washington D.C.: National Academies Press, 2009.

In 2008, the National Institute of Standards and Technology requested the National Research Council to appoint an ad hoc committee of experts to provide advice for advancing the competitiveness and productivity of the construction industry. Over 50 industry experts were gathered at a two-day workshop in November of 2008 to brainstorm and identify activities that could lead to breakthrough improvements in construction productivity. The committee incorporated its expert opinions and the discussions from the workshop and produced this report. Five opportunities for breakthrough improvements were identified. They are: (1) Widespread deployment and use of interoperable technology applications, also called Building Information Modeling (BIM); (2) Improved job-site efficiency through more effective interfacing of people, processes, materials, equipment, and information; (3) Greater use of prefabrication, preassembly, modularization, and off-site fabrication techniques and processes; (4) Innovative, widespread use of demonstration installations; and (5) Effective performance measurement to drive efficiency and support innovation. The report also provides three recommendations: (1) greater collaboration among construction industry stakeholders to implement interoperable technology applications, job-site efficiencies, off-site fabrication processes, demonstration installations, and effective performance measures; (2) the development of a technology readiness index to evaluate and mitigate risks of new technologies; and (3) collaborative efforts among governmental agencies to develop industry-level productivity measures.

O’Grady, John, and Prism Economic Analysis. *Estimates of Revenue Losses to Government as a result of Underground Practices in Ontario’s Construction Industry*. Document prepared for the Ontario Construction Secretariat. Ontario, 2001.

This document provides an update to the 1998 report titled “The Underground Economy in Ontario’s Construction Industry.” Estimates in the 1998 report were amended using new input/output data, and new estimates are developed for the 1998-2000 period. The rate of self employment and share of cash in total transactions are two indicators of underground activity. These two indicators have moderately increased. The share of underground income in total construction income has declined from 22 % to 19 %, as a result of changes in composition of the construction activities, increased enforcement, and increase in the share of housing starts by large developers. However, the amount of the underground income has increased, and the underground economy in construction remains a serious problem.

O’Grady, John, Greg Lampert, and Bill Empey. *The Underground Economy in Ontario’s Construction Industry: Estimates of Its Size and the Revenue Losses to Government and the WISB*. Document prepared for the Ontario Construction Secretariat. Ontario, 1998.

This document presents estimates of the size of the underground economy in the construction industry in Ontario, in addition to estimates of the fiscal impact for the governments and the Ontario Workplace Safety and Insurance Board. It is estimated that between 1995 and 1997, the underground employment in the construction industry in Ontario was about 25 % to 35 % of total employment. For residential construction, the estimates are between 35 % and 48 %. For nonresidential construction, it is between 11 % and 17 %. Annual total fiscal cost to the governments is between 1.1 and 1.7 billion dollars. Factors contributing to the growth of the construction underground economy include the introduction of the GST (Goods and Services Tax) in 1991, increase in unemployment due to economic conditions, and increased competition in obtaining contracts.

Oppedahl, David B. “Understanding the (Relative) Fall and Rise of Construction Wages.” *Chicago Fed Letter* July, no. 155 (2000).

Construction wages relative to all private production worker wages peaked in the 1970s and has generally declined, with a slight increase from 1996 to 2000. The decline is a result of a number of factors. Increases in the wage premium associated with higher levels of education reduce the relative wage in construction because construction workers tend to have lower educational attainment. Another factor is immigration of low-skilled workers. Technology improvement is also cited as a reason for deskilling. As advanced technologies become available, the author argues that the skills required by the labor are lessened. An example given in the article is on prefabrication. However, it should be pointed out that when more advanced technologies (such as Building Information Modeling) are used in construction, perhaps more (and different) skills are needed, which may increase labor wages. Other reasons for wage declines include the existence of an informal economy, increased safety, decline in union representation and bargaining power, and wage laws that allow hiring of less skilled workers.

Park, Hee-Sung, Stephen R. Thomas, and Richard L. Tucker. “Benchmarking of Construction Productivity.” *Journal of Construction Engineering and Management* 131, no. 7 (2005): 772-778.

This paper describes the Construction Industry Institute’s (CII) effort on construction productivity benchmarking. The Construction Productivity Metrics System (CPMS) was developed through a consensus of industry experts. CPMS is a framework for data collection and productivity analysis. With a focus on heavy industrial projects, seven activity areas were identified: concrete, structural steel, electrical, piping, instrumentation, equipment, and insulation. Specific tasks to be included in productivity measurement are identified, and units of

measurement are specified. This framework was applied to a sample of 16 industrial projects, and the results indicate this approach can yield meaningful productivity measures.

Pieper, Paul. “Why Construction Industry Productivity is Declining: Comment.” *Review of Economics and Statistics* 71, no. 3 (1989): 543-546.

In Allen (1985), construction productivity decline was explained largely by shifts from high to low productivity sectors, declines in average establishment size, labor quality, capital-labor ratio and percentage union members. Pieper challenges Allen’s conclusion. Pieper argues that the capital-labor ratio has increased, rather than decreased, as claimed by Allen (1985). Pieper agrees with Allen that there is a shift in the sector output, but he believes the effect on productivity is overestimated by Allen. He believes the problem lies in Allen’s use of a cost index to deflate value added. Allen claimed that the BEA deflator was largely based on cost indices. Pieper points out that only about one quarter of construction expenditures are deflated using cost indices. The majority of total construction expenditures are deflated using the Census Single Family Homes and the Federal Highway Administration indices, which are price indices.

Pieper, Paul. “The Measurement of Construction Prices: Retrospect and Prospect.” In *Fifty Years of Economic Measurement: The Jubilee of the Conference on Research in Income and Wealth*, Volume 54. Edited by Ernst R. Berndt and Jack E. Triplett. Chicago: University of Chicago Press, 1990.

This chapter discusses thoroughly the history of the development of construction price indices during the previous 40 years. This work has been highly cited in construction productivity measurement literature. Pieper argues that when the cost index is used to deflate output, productivity is assumed to be constant, and it is considered the least desirable type of price index. The cost index, however, has been commonly used to deflate output due to lack of appropriate output price deflators. Pieper also criticizes Dacy’s (1964, 1965) method of deflation by pointing out that an embedded assumption is that factors of production are not substitutable. He discusses alternative price indices, including bid prices, hedonic price indices, estimation indices, and cost indices. This chapter is followed by comments by Robert Parker from the BEA. Parker summarizes Pieper’s chapter and points out that the construction industry appears to lack a lobbying group that advocates the Congress to appropriate resources for improved construction statistics. Parker also describes BEA’s efforts in this area.

Rojas, Eddy M. and Peerapong Aramvareekul. “Is Construction Labor Productivity Really Declining?” *Journal of Construction Engineering and Management* 129, no. 1 (2003): 41-46.

This paper discusses factors that affect the quality of labor productivity measures at the industry level. It argues that the uncertainty in the data prevents researchers to reach a conclusion as to whether the construction labor productivity is indeed declining during 1979 to 1998. Value Put in Place is collected at the project level. Compared to manufacturing, for which data is collected at the establishment level, there are more data collecting units for the construction industry. This

is one reason for less accurate data for the construction industry. Another reason is the lack of annual data for benchmarking the value put in place data. Problems with lack of price indices for nonresidential construction are also discussed. There is also a disconnect between Construction Put in Place and Census of Construction Industries. For instance, architectural and engineering work, force-account construction, and secondary construction are included in the Construction Put in Place, but not the Census of Construction Industries. This paper also discusses the impact of changes in output mix on labor productivity and suggests that labor productivity should be calculated for different sectors within the construction industry.

Schrivver, William R. and Roger L. Bowlby. “Changes in Productivity and Composition of Output in Building Construction, 1972-1982.” *Review of Economics and Statistics* 67, no. 2 (1985): 318-322.

Cost of building per square foot in 1972 dollars was regressed on characteristics of buildings, such as number of stories, location, end-use, and framing type. The data used was from Dodge contract construction, which is based on ex ante cost estimates. The deflators used were the Census Bureau price index for a new one-family house, Turner Construction Company cost index, and American Appraisal Company cost indices. A shift to more office buildings and less residential construction was observed. Cost increase per square foot of output is interpreted as a decline in total factor productivity. This study concludes that there is a significant decline in total factor productivity in construction from 1980 through 1982, after composition of output is accounted for. There is no decline in productivity from 1972 through 1979.

Stokes, H. Kemble, Jr. An Examination of the Productivity Decline in the Construction Industry. *Review of Economics and Statistics* 63 no. 4 (1981): 495-502.

Labor productivity in the construction industry rose at an annual rate of 2.4 % between 1950 and 1968, and it declined at an annual rate of 2.8 % between 1968 and 1978. Labor productivity is measured using real value added as the output measure. The decline in productivity appears to be robust, regardless of which labor input measure is used (employees, hours paid, or hours worked). This paper examines possible causes of the decline. Only 25 % of the decline could be explained by the factors examined, which include measurement of output, shifts in the output mix, changes in capital-labor ratio, demographic changes of the labor force, economies of scale, regional shifts, and shifts in work practices. BLS has conducted studies on construction labor productivity defined as gross output per employee hour. Labor productivity defined as such increased during the 1970s. This observation is consistent with the increased use of prefabricated materials during the same period. It is also consistent with the decline of value added and an increase in manufactured construction supplies during the same period.

Teicholz, Paul, Paul M. Goodrum and Carl T. Haas. “U.S. Construction Labor Productivity Trends, 1970-1998.” *Journal of Construction Engineering and Management* 127, no. 5 (2001): 427-429.

Teicholz discusses the paper titled “U.S. Construction Labor Productivity Trends” by Allmon, Haas, Borcharding, and Goodrum (2000). Allmon *et al.* (2000) report that decreasing real labor costs and more productive equipment are major factors which made labor costs lower for accomplishing tasks in the 1990s compared to the 1970s. Task-level productivity has increased because the same tasks cost less in terms of labor. Teicholz presents labor productivity trends from 1964 to 1999, for the construction industry and all non-farm industries. Labor productivity is calculated by dividing deflated value of construction put in place with field labor hours. Labor productivity trends downward for the construction industry while it trends upward for all non-farm industries. This graph of labor productivity trends is frequently cited in reports on construction productivity.

Goodrum and Haas respond to the discussion by Teicholz. They provide a concise and informative review of the problem of productivity measurement in construction. They point out the problem with output deflators, which is embedded in the deflated value of construction put in place data. Output deflators are often based on input cost indices, and as a result, tend to overestimate inflation and underestimate output and productivity. The Census Single-Family House Under Construction Index was constructed using a hedonic approach. The authors argue that this index, while preferred, may not capture changes in quality of output. As a result, output could be biased downward. This index is used for over half of the value of construction put in place. Another possible source of underestimation of output comes from decreases in real wage in construction, which is one component of value of construction put in place. Teicholz, Goodrum, and Haas agree there is a significant need for further studies to resolve the productivity puzzle in construction industry.

B.2 Construction Data

Census Bureau. Construction Statistics Data Users’ Conference. October 28, 1997. Washington, D.C. Document issued March, 1999.

This report summarizes the discussions and presentations at the Construction Statistics Data Users’ Conference. The programs that survey and compile construction statistics are described. These programs include the Building Permits Program, Survey of Construction, Value Put in Place Program (VIP), Manufactured (Mobile) Home Program, and Economic Census: Construction Sector. The Building Permits Program, the Survey of Construction, and the Manufactured (Mobile) Home Program all focus on the residential sector. The Survey of Construction is used to develop price indices using hedonic regression models. The nonresidential sector is covered by the Value Put in Place Program and the Economic Census. The Value Put in Place Program is based on construction projects. The Economic Census is an establishment survey. The report points out that the VIP data are not comparable to the Economic Census data. The VIP construction costs include architectural and engineering design, construction management, force-account construction, and secondary construction, in addition to construction performed by the construction industry, as defined by the Economic Census. It also

points out that the Economic Census does not survey establishments with no payrolls. Sole proprietorships and partnerships are therefore not included. Also not included in the Economic Census is construction work performed in the underground economy.

Census Bureau. Statistical Abstracts of the United States

The Census Bureau compiles Statistical Abstracts of the United States using data from a variety of governmental agencies and private sources. Examples of construction related data include number of establishments, number of paid employees, and annual payroll from County Business Patterns, characteristics of commercial buildings from U.S. Energy Information Administration, producer price indices of construction materials from Bureau of Labor Statistics, and Value of Construction Put in Place from the Census Bureau.

R.S. Means Building Construction Cost Data

The R.S. Means Building Construction Cost Data provide, for individual construction tasks, estimates of daily output, crew requirement, labor hours, material cost, labor cost, equipment cost, and overhead and profits. It is a reference guide for budgeting and estimating. The data are based on surveys of contractors and suppliers. This book is published yearly. The latest version is the 67th edition (2009).

B.3 Productivity Data and Measurement

Bureau of Labor Statistics. *Trends in Multifactor Productivity, 1948-81*. Bureau of Labor Statistics Bulletin no. 2178. Washington, D.C., 1983.

This is the first publication on multifactor productivity measurement in BLS. This document describes the methodology and data sources used in BLS's multifactor productivity measurement program, and presents results. It describes the incorporation of recommendations from the Rees Report regarding aggregation methods (the adoption of Tornqvist index) and construction of capital inputs in the BLS productivity program. This document provides a brief and accessible derivation of the multifactor productivity growth equation starting from a production function. It also compares BLS estimates with productivity measures made by Denison, Jorgenson, and Kendrick.

Bureau of Labor Statistics. "Chapter 10. Productivity Measures: Business Sectors and Major Subsectors." In *BLS Handbook of Methods*. Washington, D.C., 1997.

This chapter describes the construction of labor productivity and multifactor productivity indices by the BLS. This chapter, like other chapters in the BLS Handbook of Methods, is succinct and informative. The labor productivity is an output per hour measure. Labor quality is not taken into consideration in this construct. Multifactor productivity is constructed in two ways. First, it is constructed using labor and capital inputs for major sectors. It is also constructed using capital, labor, energy, materials, and purchased business services inputs (KLEMS inputs) for

more detailed manufacturing industries. For multifactor productivity, labor is categorized into 1008 types by education, experience, and gender. Note, however, that BLS recently dropped the experience categorization. It discusses the data sources and procedures used. Inputs are aggregated using a Tornqvist chain index. Properties of this index are nicely described. This document points out that the output data for the construction industry are not satisfactory, and the productivity measures for this industry need to be used with caution. An annotated bibliography is also included.

Bureau of Labor Statistics. “Chapter 11. Industry Productivity Measures.” In *BLS Handbook of Methods*. Washington, D.C., 1997.

This chapter describes labor productivity and multifactor productivity measures. It begins with a history of BLS’s involvement in productivity measurement. It then describes the methodology the BLS uses to calculate the productivity measures, the sources of data, and the assumptions associated with the calculations. Specific details are included for industries for which BLS calculates productivity. This document also discusses the use and limitation of these measures. A list of technical references with brief annotation is included.

Bureau of Labor Statistics. *Labor Composition and U.S. Productivity Growth, 1948-90*. Bureau of Labor Statistics Bulletin no. 2426. Washington, D.C., 1993.

This document is a study that examines labor productivity, taking into account the heterogeneity of labor. Labor is categorized into groups by educational attainment, work experience, and gender. Labor productivity growth is decomposed into two components—changes in labor hours and changes in labor composition. The estimation of earnings is done in two steps. First, experience is econometrically estimated using historic data on actual work experience. Then the estimated experience is used, along with education data, to estimate labor earnings. The hourly earnings for each type of workers were estimated using econometric models instead of using averages from the survey. One reason why this approach was used was that precision is increased due to small sample sizes of some of the worker types. An appendix describes how the productivity growth equation is derived starting from a production function and the assumptions used in the process. Note that BLS recently dropped experience from its labor classification and it uses actual wages rather than estimated wages.

Bureau of Labor Statistics. *Productivity: A Selected Annotated Bibliography, 1983-87*. Bureau of Labor Statistics Bulletin no. 2360. Washington, D.C., 1990.

This annotated bibliography contains over 1000 publications published between 1983-87 on the concepts, methods, measurement, sources of productivity change, the relation between productivity to economic variables such as wages, prices, and employment, and economic growth. An author index and a subject index are included. This bibliography is the 7th in the series. Previous BLS bibliographies include Bulletin 1226 (1958), Bulletin 1514 (1966), Bulletin 1776 (1971), Bulletin 1933 (1977), Bulletin 2051 (1980), and Bulletin 2212 (1984).

Bureau of Labor Statistics. *Construction of Average Weekly Hours for Supervisory and Nonproduction Wage and Salary Workers in Detailed Industries. Technical Note.* Washington D.C., 2005.

This document describes the construction of annual ratios of supervisory worker average weekly hours to nonsupervisory worker average weekly hours using occupation data from the Current Population Survey. It also discusses the issues with classification systems associated with CES and CPS.

Bureau of Labor Statistics. *Construction of Employment and Hours for Self-employed and other Nonfarm workers and for all Farm workers, using Current Population Survey data for primary and secondary jobs.* Washington, D.C., 2006.

This document describes the calculation of number of employed and hours worked for self-employed using the Current Population Survey. Prior to 1994, CPS collected hours worked at all jobs, but only collected industry and occupation information on the primary job. Beginning in 1994, CPS collected hours worked and industry and occupation data on all primary and secondary jobs. This document describes methodologies for estimating hours worked by the self-employed and unpaid family workers using historic and more recent CPS data.

Bureau of Labor Statistics. *Technical Information About the BLS Multifactor Productivity Measures.* Washington, D.C., 2007.

This document describes the BLS multifactor productivity program, including data sources and calculation procedures. It also discusses using a simplified methodology and preliminary data to generate estimates of productivity. The list of references is annotated and is very extensive.

Dean, Edwin R. Michael J. Harper, and Mark S. Sherwood. “Productivity Measurement with Changing-Weight Indices of Outputs and Inputs.” In *Industry Productivity: International Comparison and Measurement Issues.* Paris: OECD, 1996.

This paper details the history of BLS’s productivity program and its improvements over time. Changing-weight indices are more preferable than fixed-weight indices, and the properties of Tornqvist index are discussed. The paper discusses value-added being a more appropriate output measure than gross output in multifactor productivity measurement. However, it might be better to use value-added output measures for international productivity comparisons, since value-added output measures tend to be more readily available in the international arena.

Dean, Edwin R. and Michael J. Harper. “The BLS Productivity Measurement Program.” In *New Developments in Productivity Analysis.* Edited by Charles R. Hulten, Edwin R. Dean, and Michael J. Harper. Chicago: University of Chicago Press, 2001.

This document is a historic account of BLS’s productivity program, including its expansion and improvements due to changes in data availability, developments in the literature, and needs of

data users. It provides an accessible summary of the literature on labor inputs and capital inputs and how they are operationalized in the BLS program. Topics of discussion include production theory, aggregation methods, labor composition, hours at work, the perpetual inventory method, capital deterioration and depreciation, and choices of output measures. In addition to providing links to the economic literature and data sources, the document points out best practices, limitations, and potential improvements.

Diewert, W. Erwin. “Which (Old) Ideas on Productivity Measurement Are Ready to Use?” In *New Developments in Productivity Analysis*. Edited by Charles R. Hulten, Edwin R. Dean, and Michael J. Harper. Chicago: University of Chicago Press, 2001.

This paper describes areas of improvement for productivity measurement. This paragraph summarizes some of the comments. The input-output framework tracks materials flows, but there is limited information on contracted labor services or rented capital equipment. Even for material flows, the data outside of manufacturing tend to be incomplete. The current System of National Accounts does not collect enough information on the self-employed. How the operating surplus of the self-employed is allocated between labor and capital incomes needs to be imputed. This problem could be more significant as the self-employed population grows. Issues associated with capital inputs include the limited data on efficiency declines of assets and service life of assets. Currently the opportunity cost associated with capital purchases is not included in the user cost. Another issue is the interest rate that should be used. There are also comments on a unified national statistical system in which surveys and resultant data are designed and organized in a coordinated way.

Gullickson, William. “Measurement of Productivity Growth in U.S. Manufacturing.” *Monthly Labor Review* July (1995): 13-37.

This paper focuses on the multifactor productivity of the manufacturing industry and describes basic principles in productivity measurement. First, inputs should be as comprehensive as possible. Second, double-counting should be avoided in input and output measures. The third principle is on aggregation. Changing weights, rather than fixed-weights, are preferred in aggregation. It contains a discussion on value-added output concepts and gross output concepts, indicating a distinct drawback associated with the value-added approach, particularly for disaggregated industry analysis. This paper also contains an informative discussion on productivity and prices. Productivity is a residual between output and inputs. It is also a residual between output and input prices. For instance, productivity is the means by which output price can stay constant while input prices increase.

Harper, Michael. “Estimating Capital Inputs for Productivity Measurement: An Overview of U.S. Concepts and Methods.” *International Statistical Review* 67, no. 3 (1999): 327-337.

This paper describes the concepts and methods of capital input calculation adopted by BLS. It is very readable and informative. It begins with a conceptual framework, in which there are two

agents: investors and production managers. The rental price of a capital stock is later tied back to these two decision-makers. This paper lays out the construction of capital input measures in a step-by-step fashion. It includes a discussion on what constitutes capital and why some types of capital are not included in productivity measurement. It discusses the perpetual inventory method for vintage aggregation, the assumption associated with the age/efficiency profile, procedures used to determine the rental prices of the assets, taking account of tax treatment of different assets, and finally, aggregation methods.

Jorgenson, Dale W., Frank M. Gollop, and Barbara M. Fraumeni. *Productivity and U.S. Economic Growth*. Cambridge, Massachusetts: Harvard University Press, 1987.

This book is a study of U.S. productivity from 1948 to 1979. The growth accounting framework has been adopted by the BLS in its productivity programs and has become the international standard (Jorgenson *et al.* 2005). The authors describe in detail their methodology in deriving the components in the productivity growth equation—output, labor, capital, and intermediate inputs. All the inputs were treated symmetrically. That is, labor, capital, and intermediate inputs can all contribute to growth in output, in contrast to a more restricted approach where a value-added function is assumed with an implication that intermediate inputs are not involved in productivity growth. All the inputs are also treated as being heterogeneous in their quality in this framework. Therefore, growth in an input can be due to both growths in the quantity and the quality of this input. Sectoral productivity is calculated and then aggregated to the economy level. Sources of growth were identified. Assumptions, such as Hicks neutrality and the existence of a value-added function, were tested empirically.

Jorgenson, Dale W., Mun S. Ho, and Kevin J. Stiroh. *Productivity Volume 3: Information Technology and the American Growth Resurgence*. Cambridge, Massachusetts and London, England: MIT Press, 2005.

This book presents the Jorgenson's productivity framework, incorporating information technology as an intermediate input. Intermediate inputs are a significant input in more than 70 % of the industries studied. Explicitly accounting for them by using gross output as the output measure is the proper approach. This work also demonstrates that it is fruitful to separate heterogeneous components of inputs. For instance, IT-related capital was separated from other types of capital, and sources of growth can be properly allocated. Similarly, labor was separated into college educated and non-college educated types. More detailed categorization of inputs enables explanation of industry productivity trends.

Lally, Paul R. "How BEA Accounts for Investment in Private Structures." *Survey on Current Business* February (2009): 9-15.

This paper describes the use of recently available BLS PPIs in BEA's estimates on private nonresidential structure investments. These include PPIs for office buildings, warehouses, industrial buildings, and schools. For building types for which PPIs are not currently available,

alternative deflation methods are discussed. This paper also mentions the possibility of developing PPIs for highways, hospitals, retail, communication, power, and lodging structures.

Lawson, Ann M. Brian C. Moyer, Sumiye Okubo, and Mark A. Planting. “Integrating Industry and National Economic Accounts, First Steps and Future Improvements.” In *A New Architecture for the U.S. National Accounts*. Edited by Dale W. Jorgenson, J. Steven Landefeld, and William D. Nordhaus. Chicago and London: University of Chicago Press, 2006.

This chapter discusses quality of value-added estimates in the I-O accounts and in the GDP-by-industry accounts. Value-added estimates from both the I-O accounts and the GDP-by-industry accounts for the construction industry were pointed out to be poor in quality. The poor data quality is due to incomplete coverage in the Economic Census and the large number of low-quality enterprise-establishment adjustments.

Mark, Jerome A. “Problems Encountered in Measuring Single- and Multifactor Productivity.” *Monthly Labor Review* (1986): 3-11.

This paper contains data sources and methods BLS uses for productivity calculation. It also discusses lack of good price deflators, particularly for the construction industry, among other challenges. Construction industry output for nonresidential structures is deflated using cost indices. This results in a productivity index that is biased towards no change. It was pointed out that the lack of appropriate price deflators is the determinant for whether a productivity measure can be derived in many cases. This paper recognizes that productivity measurement is not an easy task. BLS has made many improvements in its productivity program throughout the years, and more improvements will need to be made in the future. While labor productivity is often a less preferred measure of productivity compared to multifactor productivity, it is calculated with much more precision and with fewer assumptions.

National Research Council. *Measurement and Interpretation of Productivity*. Washington D.C.: National Academy of Sciences, 1979.

This document is also known as the Rees report, produced by the Panel to Review Productivity Statistics set up by the National Academy of Sciences. This book consists of two parts—a report that gives an overview of the productivity measurement issues and recommendations, and a collection of papers on productivity measurement. Much of BLS’s improvements on its productivity program can be traced to recommendations in this document.

Notably, a paper by Gollop shows that the assumption of the existence of a value-added sub-function in the production function is too restrictive. This assumption implies that the marginal rates of substitution between the arguments of the value-added sub-function—capital, labor, and time—are independent of intermediate inputs. It also implies that the intermediate inputs are not involved in technological change and that technological change can only occur through capital

and labor. This work shows the importance of explicitly treating capital, labor, and intermediate inputs symmetrically in productivity measurement.

OECD (Organization for Economic Co-operation and Development). *Measuring Productivity—Measurement of Aggregate and Industry-Level Productivity Growth*. Paris: OECD, 2001.

“The OECD Productivity Manual is the first comprehensive guide to the various productivity measures aimed at statisticians, researchers and analysts involved in constructing industry-level productivity indicators.” This users’ guide focuses on productivity growth, rather than productivity levels, at the industry level, using non-parametric methods. A variety of productivity measurements are described. How to choose among the different options depends on the purpose and data availability. This manual is very accessible and practical. It points out desirable qualities associated with different approaches, but it also indicates practical challenges. It focuses on the index number approach in a production theoretic framework, but a section of the manual is devoted to the growth accounting approach. The growth accounting approach integrates the theory of the firm, index number theory and national accounts. The growth accounting technique looks at the rates of changes in output and the rates of changes in inputs. The multifactor productivity growth is determined as the “unexplained” residual.

This manual points out that availability of data poses a significant challenge in the construction of productivity measures. Examples include price indices for output measures by industry, hours worked by industry (in particular, statistics for self-employed individuals, and cross-classification by productivity related characteristics), service life of assets, age-efficiency and age-price profiles of assets, and updated input-output tables integrated with national accounts.

OECD. *Measuring Capital—Measurement of Capital Stocks, Consumption of Fixed Capital and Capital Services*. Paris: OECD, 2001.

This OECD manual describes concepts related to capital measurement and provides guidelines for estimation of capital stocks, consumption of fixed capital, and capital services. In addition to established methods, it discusses alternative methods, what is commonly implemented in practice, alternative data sources, and some unresolved issues. It is a very detailed reference on measuring capital.

Schreyer, Paul, and Dirk Pilat. *Measuring Productivity*. Economic Studies no. 33, OECD, Second Quarter. Paris: OECD, 2001.

This document provides an overview of the growth accounting approach of productivity measurement, with discussions on the comparison issues of productivity growth and levels between countries and across time. Gross output and value added output measures are compared. Using value added output, the relationship between multifactor productivity and labor productivity is derived. Changes in value-added-based labor productivity are shown to be

the sum of labor productivity changes due to changes in capital-labor ratios and effects of multifactor productivity growth. In addition to capital-labor ratio and multifactor productivity growth, changes in gross-output-based labor productivity are also a function of the ratio between intermediate input and labor input. Multifactor productivity measures based on gross output are not comparable across different levels of aggregation due to interindustry transactions. Multifactor productivity measures based on value added are comparable across different levels of aggregation because interindustry flows are subtracted from the output measure. The authors point out that gross output and value added are useful complements. This document also contains an informative discussion on the interpretation of productivity measures.

Zoghi, Cindy. *Measuring Labor Composition: A Comparison of Alternate Methodologies*. Bureau of Labor Statistics. Washington, D.C., 2007.

This paper discusses the issues associated with calculating a labor composition index. Labor quality is often taken into account by sorting labor into types by education, experience, age, gender, occupation, and geographic region. It is not clear exactly which of these variables are the best to use to capture the difference in effectiveness of labor. What determines wage? Wage may not always reflect marginal productivity of labor. A number of theories are discussed. A second issue is regarding whether to use the actual wages as weights in labor input aggregation or to use the estimated wages from Mincer-type human capital wage regressions as weights. It is not possible to determine which approach is best regarding the choice of labor type categorizations and the use of estimated or actual wages.

Appendix C Sources of Construction Data Related to Productivity and Their Availability

This appendix provides a description of data sources that may be relevant to construction productivity measurement. It then describes classification systems, variables, and availability. These materials are tabulated for key sources of data.

C.1 Sources of Construction Data Related to Productivity

County Business Patterns (Census Bureau)

County Business Patterns contains annual data. The variables include number of employees, payroll, and number of establishments by NAICS codes. Number of establishments by employment-size class for NAICS categories is also reported. The online data is available for 1998 through 2006. County Business Patterns does not cover the self employed.

Dodge Reports (McGraw-Hill Construction)

Dodge Reports are lists of construction projects and are available since 1967. Variables include value, month started, square footage, dwelling units (for residential only), state, county, project type, number of stories, and ownership (private and four public categorization). Note that the value associated with a project is an ex ante estimate. Whether a project is new construction, addition, or alteration is also indicated. Framing type is also reported. Examples of building types in the nonresidential sector include stores and restaurants, warehouses, office and bank buildings, parking garages and automotive services, manufacturing plants, warehouses, labs, schools, libraries, hospitals, government service buildings, religious buildings, amusement, social, and recreational buildings, hotels and motels, and dormitories. Nonbuilding categories include streets and highways, bridges, dams, reservoirs, river development, sewage and waste disposal systems, water supply systems, power plants, gas, and communication systems. While the Value Put in Place data have undergone a classification system change, the classification system of the Dodge Reports has remained the same since 1967.

Current Employment Statistics Survey

The BLS Current Employment Statistics (CES) program is used as the primary source of industry employment and hours data. The data are collected monthly and the employment levels are benchmarked yearly using data from State unemployment insurance programs, which covers about 98 % of all nonfarm employees.²⁰³ The classification of industries in this survey is the NAICS system since 2003, and some of the historic data since 1990 were updated using the same

²⁰³ Bureau of Labor Statistics, *Handbook of Methods* (Washington DC: Bureau of Labor Statistics, 1997).

classification system.²⁰⁴ For industries that have the same or similar titles in the SIC and NAICS systems, there could still be some discrepancy in the data classified under these two systems. For example, under the SIC, auxiliary establishments were classified under the same code as the primary activity of the parent enterprise. Under NAICS, auxiliary establishments are classified based on their own primary activity.²⁰⁵ In 2003, in addition to the new classification system, the CES also switched from a quota-based sampling method to a probability based sampling method.²⁰⁶ The historic CES data that remain in the SIC classification system extend back to 1988, 1972, or 1958 for different segments of the construction industry. The longest time series is for the total number of all employees, the annual data for which extend back to 1919. The CES reports the number of all employees, the number of production workers, the number of women workers, the average weekly hours of production workers, the average weekly earnings of production workers, and the average weekly hours of overtime of production workers. In the case of the construction industry, the production workers include “workers, up through the level of working supervisors, who are engaged directly in a construction project, either at the site or in shops or yards, at jobs ordinarily performed by members of construction trades.”²⁰⁷ For non-production workers, which are executive and managerial personnel, professional and technical employees, and workers in routine office jobs, only employment data is available from this survey. One disadvantage of the CES data set is that hours paid, instead of hours worked are reported. Hours paid include vacation, paid sick leave, and holidays, in addition to hours worked. Hours at work includes paid time for traveling between jobs sites, coffee breaks, and machine downtime.²⁰⁸ While some of the hours at work do not increase productivity, others do. One such example is activities that motivate workers and reduce shirking. Paid leave is best considered a benefit.²⁰⁹ Work hours of non-production workers are not collected, and therefore would need to be estimated. The Current Employment Statistics Survey is also an establishment-level survey, and it does not contain information on the self-employed.

²⁰⁴ Teresa L. Morisi, “Recent Changes in the National Current Employment Statistics Survey,” *Monthly Labor Review* June (2003): 3-13. For national series that had classifications with 90% or greater degree of interchangeability under SIC and NAICS, the historic data were reconstructed.

²⁰⁵ *Ibid.*

²⁰⁶ *Ibid.*

²⁰⁷ Bureau of Labor Statistics, *Handbook of Methods*. *Op. cit.*

²⁰⁸ *Ibid.*

²⁰⁹ Lucy P. Eldridge, Marilyn E. Manser, and Phyllis Flohr Otto, “Alternative Measures of Supervisory Employee Hours and Productivity Growth,” *Monthly Labor Review* April (2004): 9-28.

Current Population Survey (Census Bureau and Bureau of Labor Statistics)

The Current Population Survey (CPS) is a household survey. The CPS is used by the BLS to supplement the CES, for data on proprietors and unpaid family workers since self-employed individuals are not included in the CES. This is particularly a concern for the construction industry where a large proportion of the workers are self-employed. The CPS collects hours worked and weeks worked, in addition to industry and occupation information. The construction industry is included in the CPS as one category and is not further divided into subcategories. Number of workers, and average work hours are reported for wage and salary workers, self-employed workers and unpaid family workers. These data are available from 1994 to present.

One limitation of using the Current Population Survey to obtain information on the self employed is the sample size. Coding of industries and reporting are more accurate in establishment level surveys compared to household surveys. For this reason, data from the Current Employment Statistics program is used as a primary source of data, and data from the Current Population Survey is used as a supplemental source of data in BLS's productivity program.²¹⁰

The Current Population Survey is also used by the BLS to derive annual ratios of supervisory (or nonproduction) worker average weekly hours to nonsupervisory (or production) worker average weekly hours, and subsequently nonproduction worker hours. The CPS asks respondents for their occupation and employment status. The information on occupation and employment status is used to sort the data into supervisory (nonproduction) and nonsupervisory (production) categories. CPS collects data on hours worked. Ratios of supervisory (nonproduction) worker average weekly hours to nonsupervisory (production) worker average weekly hours are calculated. These ratios are multiplied by nonsupervisory (production) worker average weekly hours from the CES. Note the hours data from the CES are for hours paid, and therefore some discrepancy is introduced. The ratios between hours worked and hours paid are available at major sector level, but not available at detailed industry level. The resultant number, supervisory (nonproduction) worker average weekly hours, is then multiplied by the number of supervisory workers to yield total supervisory worker weekly hours. Total supervisory worker hours are obtained by multiplying total supervisory worker weekly hours by 52. Total supervisory worker hours are then combined with total nonsupervisory worker hours from CES and total self-employed hours and unpaid family worker hours to yield total hours for an industry.

Economic Census (Census Bureau)

The Economic Census is an establishment survey and it covers establishments with payrolls. The focus of the Economic Census of the Construction Industries is establishments whose primary activity is construction. It is conducted every five years, in years ending with 2 or 7. It

²¹⁰ *Ibid.*

is classified under NAICS in recent years and under SIC in historic years. Many categories of the construction industry under SIC and NAICS are not comparable.²¹¹ This change in classification system presents itself as a break in the time series of Economic Census data. The Economic Census defines value added for construction industries to be “the dollar value of business done less costs for construction work subcontracted to others and payments for materials, components, supplies, and fuels.”²¹² All the components needed to calculate value added are collected in Economic Census surveys. In terms of labor input, the Economic Census contains data on number of construction workers employed and number of other employees.²¹³ There is, however, no information on the work hours or full-time vs. part-time status of the workers. The Economic Census does, however, collect data on labor costs. Percentage of construction work done in various project types, such as office building construction or tunnel construction, is also collected. Percentage of construction work done in different specialty trade activities, such as concrete work or structural steel erection, is also collected for contractors.

While the C30 reports or the Dodge reports contain data on gross output of the construction industry, the Economic Census is the only data source that enables the calculation of a value added measure of output. However, because the Economic Census covers both general contractors and subcontractors, there is a significant amount of double counting in terms of output. This is the reason why output data for the construction industry are not used in the input-output tables produced by the Bureau of Economic Analysis.²¹⁴ As the construction industry is one that wages often are paid in cash, labor wage data may be biased downward.²¹⁵ The Economic Census also does not cover the self-employed. Note also that there has been a change in the definition of value of construction work in the Economic Census surveys. The Economic Censuses for 1987-1997 collected value of construction work. In 2002 Economic Census, receipts, billings, or sales for construction work were collected to enhance the accuracy of estimates by respondents.²¹⁶

²¹¹ <http://www.census.gov/epcd/naics02/SICN02C.HTM#symbols>

²¹² 2002 Economic Census.

²¹³ In 1997 Economic Census, number of construction workers is not available, however, annual payroll costs for construction workers is available.

²¹⁴ Bureau of Economic Analysis, *Concepts and Methods of the U.S. Input-Output Accounts* (Washington DC: Bureau of Economic Analysis, 2009).

²¹⁵ *Ibid.*

²¹⁶ Census Bureau, 2002 Economic Census, Construction, Industry Series (Washington DC: Census Bureau, 2005).

Hours-at-Work Survey and National Compensation Survey (BLS)

BLS has been collecting annual establishment level data on actual hours worked for production and nonsupervisory workers (Hours-at-Work Survey) since 1981. Data from the Hours at Work Survey are used to derive ratios of hours at work to hours paid. This is done for 1-digit Standard Industrial Classification (SIC) industry groups on an annual basis. These ratios are then used to convert hours paid data from the establishment survey. Hours-at-Work Survey by itself, however, is not detailed enough to be used in industry-level productivity measurement. The BLS terminated the Hours at Work Survey in 2000, and replaced the HWS with the National Compensation Survey. The Employment Cost Index (ECI) from the National Compensation Survey is used to convert hours paid to hours worked. The Hours at Work Survey had a few limitations. Eventually, because of stringent data reporting requirements, the response rate decreased to the point where not enough data were usable. The ECI was designed to capture the hourly cost of wages and benefits, including paid leave. Ratios of hours at work to hours paid can also be constructed using the ECI data. These ratios are calculated using the ECI data since 2001. For the years before 2001, the ratios are based on Hours at Work Survey. The HWS survey included production and nonsupervisory workers in nonagricultural establishments. The National Compensation Survey, on the other hand, covers all workers. Another advantage of the National Compensation Survey is that it contains a bigger sample. The sample size is 37000 occupations within 8500 private establishments where as the HWS sampled fewer than 6000 establishments. The response rate associated with the NCS is also higher than that of the HWS.

Input-Output tables (Bureau of Economic Analysis)

The main data source of the input-output tables associated with the construction industry is the Value of Construction Put in Place data. Data from the Economic Census are generally not used to estimate output of the construction industry because of substantial double counting due to the inclusion of both general contractors and subcontractors.²¹⁷ Economic Census data, however, are used for inputs to construction industries. Generally, the I-O tables are organized using NAICS, but for the construction industry, activities are used due to data limitation.²¹⁸

National Income and Product Accounts Fixed Investment and Fixed Assets Data (Bureau of Economic Analysis)

The BEA compiles data on private fixed investment by structure types and investment in government fixed assets by structure types. One data source the BEA uses is the C30 data from the Census Bureau. In addition to construction spending as defined by the Census Value of Construction Put in Place, BEA includes mining exploration, shafts, and wells, brokers'

²¹⁷ Bureau of Economic Analysis, *Concepts and Methods of the U.S. Input-Output Accounts* (Washington DC: Bureau of Economic Analysis, 2009).

²¹⁸ *Ibid.*

commissions on the sale of new and used structures, mobile structures, manufactured homes, and net purchases of used structures. The data series start from 1929 and are available annually through the present. In 1997, the classification system changed. The pre-1997 classification system is based on building types. The new classification system is based on function (or end use). This change in classification system was preceded by the change in the classification in the C30 reports. In the 2009 Comprehensive Revision of the NIPAs, historic data were updated to conform to the new classification system.²¹⁹ For the private sector, real fixed investment is reported along with quantity index and price index. For the public sector, investment in fixed assets is reported in current dollars along with a quantity index. Data of investment by structure types could be used as a gross output measure in productivity analyses.

In addition to fixed investment by structure types, the BEA also compiles data on assets by industry. For the construction industry as a whole, net stocks, depreciation, and investment data are reported both in terms of current dollars and chain-type quantity indices. These data are reported annually by 32 equipment types and 15 structure types. Net stocks, depreciation, and investment data by industry are elements of capital in productivity analysis.

Nonemployer Statistics (Census Bureau)

The Nonemployer Statistics contains annual data and is based on administrative records. The universe of this survey is businesses with no paid employees. The variables include number of establishments and receipts by NAICS codes and by type of establishment (corporations, individual proprietorships, and partnerships). More aggregated NAICS classification is used to report the number of establishments and receipts by states.

Price Deflator (Fisher) Index of New One-Family Houses Under Construction (Census Bureau)

Price deflators for new one-family houses under construction are developed using a hedonic regression approach and are available monthly since 1964.

Price Deflator for New Multifamily Housing (Tabulated by the Census Bureau for the Bureau of Economic Analysis)

Price deflators for new multifamily housing are developed using a hedonic regression approach. This index was first developed in 1993 and extends back to 1978.²²⁰

Producer Price Indices (Bureau of Labor Statistics)

²¹⁹ Eugene P. Seskin and Shelly Smith, "Improved Estimates of the National Income and Product Accounts: Results of the 2009 Comprehensive Revision" *Survey of Current Business* September (2009): 15-35.

²²⁰ Frank de Leeuw, "A Price Index for New Multifamily Housing," *Survey of Current Business* February (1993):33-42.

The Bureau of Labor Statistics has recently developed producer price indices for the nonresidential sector of the construction industry. These producer price indices are a Laspeyres index, which holds quality constant. These price indices are based on prototypical buildings. Producer price indices are produced for the following new building construction categories: (1) new industrial building construction (NAICS 236211); (2) new warehouse building construction (NAICS 236221); (3) new school building construction (NAICS 236222); and (4) new office building construction (NAICS 236223). For each of the four building types, model buildings are developed. The buildings are comprised of a collection of assemblies, or production elements. The estimation of pricing for each assembly includes materials and labor. Sometimes machinery is also required. These costs are estimated by a cost-estimating firm. BLS surveys contractors regarding their margin (overhead and profits). BLS tracks both the costs and the margin. Producer price indices have also been developed for four types of special trades in the nonresidential setting (commercial and industrial): 1) concrete contractors (NAICS 23811), 2) roofing contractors (NAICS 23816), 3) electrical contractors (NAICS 23821), and 4) plumbing/HVAC contractors (NAICS 23822). The producer price indices for the specialty trades are for both new nonresidential building construction and nonresidential building maintenance and repair. Excluded from these producer price indices are residential work, additions, renovations, and non-building construction.

Producer Price Indices for Materials and Supply Inputs to Construction Industries (Bureau of Labor Statistics)

The BLS produces PPIs for materials and supply inputs to construction industries for different types of new construction (single-unit residential, multi-unit residential, non-residential buildings, highway and street construction, and other heavy construction) and for residential and non-residential maintenance and repair construction. These indices only include costs of materials and supplies and do not take into account labor costs, contractor overhead, and profits.

R.S. Means Square Foot Costs

Square Foot Costs data from R.S. Means are available since early 1980s. The BEA has used data from the more recent years along with the hedonic regression approach to develop price deflators for several building types. In the BEA's hedonic model, the natural logarithm of the cost per square foot is regressed on total square feet, dummy variables for combinations of exterior wall and interior supporting-frame type, and dummy variable for the year.²²¹

Survey of Construction (Census Bureau)

The Survey of Construction focuses on new residential buildings. Data are monthly and include start date, completion date, and physical characteristics of each housing unit, such as square

²²¹ Carol E. Moylan and Brooks B. Robinson, "Preview of the 2003 Comprehensive Revision of the National Income and Product Accounts—Statistical Changes," *Survey of Current Business* September (2003): 17-32.

footage and number of bedrooms. These variables are collected for both new single-family and multifamily housing units. In addition, sales date and sales price are collected for one-family houses. Housing starts data have been collected since 1959, housing completions data have been collected since 1963, and housing completions data have been collected since 1968.

Value of Construction Put in Place (Census Bureau)

Value of Construction Put in Place is collected by Census Bureau's Manufacturing, Mining, and Construction Statistics (i.e., Current Construction Report, Series C30: Value of New Construction). Samples for the Value of Construction Put in Place Survey are drawn from the list of construction projects produced by McGraw-Hill Construction (Dodge Reports). Dodge Reports do not usually contain projects in nonpermit areas, and therefore projects in nonpermit areas are identified separately.²²² Value of Construction Put in Place includes architectural and engineering design, construction management, force-account construction, and secondary construction, in addition to construction services performed by the construction industry, as defined by the Economic Census. Since the Value of Construction Put in Place data are collected from owners, this data capture some construction activities not captured by the establishment-based Economic Census. Examples are construction by the self-employed, homeowner construction, and construction done as a secondary source of revenue by nonconstruction establishments.²²³ The definitions of construction in the Value of Construction Put in Place and the Economic Census are also different. For instance, maintenance and repair is part of value of construction work in the Economic Census, but it is not included in the Value of Construction Put in Place. The value of land is excluded. For all sampled projects, a questionnaire is mailed to the owner of the project prior to the start of the project. Estimates of total construction cost, architectural, engineering, and miscellaneous costs are requested in the survey. Value of Construction Put in Place is collected monthly until the project is completed. The Census Bureau reports that about two thirds of Value Put in Place corresponds with work performed by the construction industry as defined by the Economic Census.²²⁴ Data of Value of Construction Put in Place are reported monthly by building types. These data are not deflated, but they are reported with and without seasonal adjustment. Total construction cost, which is the sum of contract construction cost and owner supplied materials and labor, can be obtained from the microdata, but it is not published. Square footage information is reported on survey forms, but it is not published.

The classification system of the Value of Construction Put in Place data changed in 1993. The new system is based on project types by end usage while the older system is based on building

²²² Census Bureau, Construction Statistics Data Users' Conference. October 28, 1997. Washington, DC. Document issued March, 1999.

²²³ Bureau of Economic Analysis, *Concepts and Methods of the U.S. Input-Output Accounts*. *Op. cit.*

²²⁴ C30 Construction Spending Latest News, July 1, 2003.

and nonbuilding types. Data collected under the two classification systems are generally not comparable, particularly at a finer level. Some categories appear to be similar in both classification systems, but there are within-category changes that made the data incomparable. For instance, private medical offices were classified as office buildings in the old classification system, but they are classified under health care in the new classification system. Direct comparisons can only be made at the more aggregate levels, specifically for total, total private, total state and local, total federal, and total public levels for annual and not seasonally adjusted monthly data.²²⁵

C.2 Classification Systems, Variables, and Coverage

The Standard Industry Classification (SIC) of the construction industry is tabulated, in addition to the North American Industry Classification System (NAICS) 1997 and 2007. Data sources that are tabulated are: Current Employment Statistics Survey (Bureau of Labor Statistics) with SIC classification and with NAICS classification, GDP by Industry (Bureau of Economic Analysis), Producer Price Indices (Bureau of Labor Statistics), producer price indices for materials and supply inputs to construction industries (Bureau of Labor Statistics), Investment in government fixed assets and private fixed investment by structure type from National Income and Product Accounts (Bureau of Economic Analysis), Value of Construction Put in Place with old and new classification systems (Census Bureau), Economic Census of the Construction Industries (Census Bureau). For Economic Census of the Construction Industries, sub-industries under SIC or NAICS classifications are tabulated for 1992, 1997, and 2002. Variables and their availability are tabulated separately. This compilation of variables and their availability is based mainly on Industry Series of 1992, 1997, and 2002 and is therefore incomplete. Source data and price indices for BEA's annual estimates of private fixed investment in structures by type are also tabulated.

²²⁵ *Ibid.*

Table C.1 Standard Industry Classification (SIC)

SIC code	Definition
15	Building Construction General Contractors And Operative Builders
152	General Building Contractors-Residential
1521	General Contractors-Single-Family Houses
1522	General Contractors-Residential Buildings, Other Than Single-Family
153	Operative Builders
1531	Operative Builders
154	General Building Contractors-nonresidential
1541	General Contractors-Industrial Building and Warehouses
1542	General Contractors-Nonresidential Buildings, Other Than Industrial Buildings and Warehouses
16	Heavy Construction Other Than Building Construction Contractors
161	Highway and Street Construction, Except Elevated Highways
1611	Highway and Street Construction, Except Elevated Highways
162	Heavy Construction, Except Highway and Street
1622	Bridge, Tunnel, and Elevated Highway Construction
1623	Water, Sewer, Pipeline, and Communications and Power Line Construction
1629	Heavy Construction, Not Elsewhere Classified
17	Construction Special Trade Contractors
171	Plumbing, Heating And Air-conditioning
1711	Plumbing, Heating and Air-Conditioning
172	Painting And Paper Hanging
1721	Painting and Paper Hanging
173	Electrical Work
1731	Electrical Work
174	Masonry, Stonework, Tile Setting, And Plastering
1741	Masonry, Stone Setting, and Other Stone Work
1742	Plastering, Drywall, Acoustical, and Insulation Work
1743	Terrazzo, Tile, Marble, and Mosaic Work
175	Carpentry And Floor Work
1751	Carpentry Work
1752	Floor Laying and Other Floor Work, Not Elsewhere Classified
176	Roofing, Siding, And Sheet Metal Work
1761	Roofing, Siding, and Sheet Metal Work
177	Concrete Work
1771	Concrete Work
178	Water Well Drilling
1781	Water Well Drilling
179	Miscellaneous Special Trade Contractors
1791	Structural Steel Erection
1793	Glass and Glazing Work
1794	Excavation Work
1795	Wrecking and Demolition Work
1796	Installation or Erection of Building Equipment, Not Elsewhere
1799	Special Trade Contractors, Not Elsewhere Classified

Table C.2 North American Industry Classification System (NAICS) 1997

NAICS code	Definition
23	Construction
233	Building, Developing, and General Contracting
2331	Land Subdivision and Land Development
23311	Land Subdivision and Land Development
2332	Residential Building Construction
23321	Single Family Housing Construction
23322	Multifamily Housing Construction
2333	Nonresidential Building Construction
23331	Manufacturing and Industrial Building Construction
23332	Commercial and Institutional Building Construction
234	Heavy Construction
2341	Highway, Street, Bridge, and Tunnel Construction
23411	Highway and Street Construction
23412	Bridge and Tunnel Construction
2349	Other Heavy Construction
23491	Water, Sewer, and Pipeline Construction
23492	Power and Communication Transmission Line Construction
23493	Industrial Nonbuilding Structure Construction
23499	All Other Heavy Construction
235	Special Trade Contractors
2351	Plumbing, Heating, and Air-Conditioning Contractors
23511	Plumbing, Heating, and Air-Conditioning Contractors
2352	Painting and Wall Covering Contractors
23521	Painting and Wall Covering Contractors
2353	Electrical Contractors
23531	Electrical Contractors
2354	Masonry, Drywall, Insulation, and Tile Contractors
23541	Masonry and Stone Contractors
23542	Drywall, Plastering, Acoustical, and Insulation Contractors
23543	Tile, Marble, Terrazzo, and Mosaic Contractors
2355	Carpentry and Floor Contractors
23551	Carpentry Contractors
23552	Floor Laying and Other Floor Contractors
2356	Roofing, Siding, and Sheet Metal Contractors
23561	Roofing, Siding, and Sheet Metal Contractors
2357	Concrete Contractors
23571	Concrete Contractors
2358	Water Well Drilling Contractors
23581	Water Well Drilling Contractors
2359	Other Special Trade Contractors
23591	Structural Steel Erection Contractors
23592	Glass and Glazing Contractors
23593	Excavation Contractors
23594	Wrecking and Demolition Contractors
23595	Building Equipment and Other Machinery Installation Contractors
23599	All Other Special Trade Contractors

Source: U.S. Census Bureau

Table C.3 North American Industry Classification System (NAICS) 2007

NAICS code	Definition
23	Construction
236	Construction of Buildings
2361	Residential Building Construction
23611	Residential Building Construction
236115	New Single-Family Housing Construction (except Operative Builders)
236116	New Multifamily Housing Construction (except Operative Builders)
236117	New Housing Operative Builders
236118	Residential Remodelers
2362	Nonresidential Building Construction
23621	Industrial Building Construction
236210	Industrial Building Construction
23622	Commercial and Institutional Building Construction
236220	Commercial and Institutional Building Construction
237	Heavy and Civil Engineering Construction
2371	Utility System Construction
23711	Water and Sewer Line and Related Structures Construction
237110	Water and Sewer Line and Related Structures Construction
23712	Oil and Gas Pipeline and Related Structures Construction
237120	Oil and Gas Pipeline and Related Structures Construction
23713	Power and Communication Line and Related Structures Construction
237130	Power and Communication Line and Related Structures Construction
2372	Land Subdivision
23721	Land Subdivision
237210	Land Subdivision
2373	Highway, Street, and Bridge Construction
23731	Highway, Street, and Bridge Construction
237310	Highway, Street, and Bridge Construction
2379	Other Heavy and Civil Engineering Construction
23799	Other Heavy and Civil Engineering Construction
237990	Other Heavy and Civil Engineering Construction
238	Specialty Trade Contractors
2381	Foundation, Structure, and Building Exterior Contractors
23811	Poured Concrete Foundation and Structure Contractors
238110	Poured Concrete Foundation and Structure Contractors
23812	Structural Steel and Precast Concrete Contractors
238120	Structural Steel and Precast Concrete Contractors
23813	Framing Contractors
238130	Framing Contractors
23814	Masonry Contractors
238140	Masonry Contractors
23815	Glass and Glazing Contractors
238150	Glass and Glazing Contractors
23816	Roofing Contractors
238160	Roofing Contractors
23817	Siding Contractors
238170	Siding Contractors
23819	Other Foundation, Structure, and Building Exterior Contractors
238190	Other Foundation, Structure, and Building Exterior Contractors

Table C.3 North American Industry Classification System (NAICS) 2007

NAICS code	Definition
2382	Building Equipment Contractors
23821	Electrical Contractors and Other Wiring Installation Contractors
238210	Electrical Contractors and Other Wiring Installation Contractors
23822	Plumbing, Heating, and Air-Conditioning Contractors
238220	Plumbing, Heating, and Air-Conditioning Contractors
23829	Other Building Equipment Contractors
238290	Other Building Equipment Contractors
2383	Building Finishing Contractors
23831	Drywall and Insulation Contractors
238310	Drywall and Insulation Contractors
23832	Painting and Wall Covering Contractors
238320	Painting and Wall Covering Contractors
23833	Flooring Contractors
238330	Flooring Contractors
23834	Tile and Terrazzo Contractors
238340	Tile and Terrazzo Contractors
23835	Finish Carpentry Contractors
238350	Finish Carpentry Contractors
23839	Other Building Finishing Contractors
238390	Other Building Finishing Contractors
2389	Other Specialty Trade Contractors
23891	Site Preparation Contractors
238910	Site Preparation Contractors
23899	All Other Specialty Trade Contractors
238990	All Other Specialty Trade Contractors

Source: Census Bureau

NAICS 2002 is very similar to NAICS 2007. The only difference is that under NAICS 2007, NAICS 23821 and NAICS 238210 are "Electrical Contractors and Other Wiring Installation Contractors," whereas, it is "Electrical Contractors" under NAICS 2002.

Table C.4 Current Employment Statistics Survey with SIC Classification (Bureau of Labor Statistics)

SIC code	Category	Variable	Seasonal adjustment	Coverage
15-17	Construction	number of all employees	both	monthly from January 1939 to April 2003
15-17	Construction	number of all employees	no	annual from 1919
15-17	Construction	number of production workers	both	monthly from January 1947 to April 2003
15-17	Construction	average weekly hours of production workers	both	monthly from January 1947 to April 2003
15	General building contractors	number of all employees	both	monthly from January 1958 to April 2003
15	General building contractors	number of production workers	no	monthly from January 1964 to March 2003
15	General building contractors	average weekly hours of production workers	no	monthly from January 1958 to March 2003
15	Residential building construction	number of all employees	no	monthly from January 1972 to March 2003
152	Residential building construction	number of production workers	no	monthly from January 1972 to March 2003
152	Residential building construction	average weekly hours of production workers	no	monthly from January 1972 to March 2003
153	Operative builders	number of all employees	no	monthly from January 1958 to March 2003
153	Operative builders	number of production workers	no	monthly from January 1972 to March 2003
153	Operative builders	average weekly hours of production workers	no	monthly from January 1972 to March 2003
154	Nonresidential building construction	number of all employees	no	monthly from January 1972 to March 2003
154	Nonresidential building construction	number of production workers	no	monthly from January 1972 to March 2003
154	Nonresidential building construction	average weekly hours of production workers	no	monthly from January 1972 to March 2003
16	Heavy construction, except building	number of all employees	both	monthly from January 1972 to April 2003
16	Heavy construction, except building	number of production workers	no	monthly from January 1972 to March 2003
16	Heavy construction, except building	average weekly hours of production workers	no	monthly from January 1972 to March 2003
161	Highway and street construction	number of all employees	no	monthly from January 1988 to March 2003
161	Highway and street construction	number of production workers	no	monthly from January 1988 to March 2003
161	Highway and street construction	average weekly hours of production workers	no	monthly from January 1988 to March 2003
162	Heavy construction, except highway	number of all employees	no	monthly from January 1972 to March 2003
162	Heavy construction, except highway	number of production workers	no	monthly from January 1972 to March 2003
162	Heavy construction, except highway	average weekly hours of production workers	no	monthly from January 1972 to March 2003
17	Special trade contractors	number of all employees	both	monthly from January 1972 to April 2003
17	Special trade contractors	number of production workers	no	monthly from January 1972 to March 2003
17	Special trade contractors	average weekly hours of production workers	no	monthly from January 1972 to March 2003
171	Plumbing, heating, and air-conditioning	number of all employees	no	monthly from January 1958 to March 2003
171	Plumbing, heating, and air-conditioning	number of production workers	no	monthly from January 1958 to March 2003
171	Plumbing, heating, and air-conditioning	average weekly hours of production workers	no	monthly from January 1958 to March 2003
172	Painting and paper hanging	number of all employees	no	monthly from January 1958 to March 2003
172	Painting and paper hanging	number of production workers	no	monthly from January 1958 to March 2003
172	Painting and paper hanging	average weekly hours of production workers	no	monthly from January 1958 to March 2003

Table C.4 Current Employment Statistics Survey with SIC Classification (Bureau of Labor Statistics)

SIC code	Category	Variable	Seasonal adjustment	Coverage
173	Electrical work	number of all employees	no	monthly from January 1958 to March 2003
173	Electrical work	number of production workers	no	monthly from January 1958 to March 2003
173	Electrical work	average weekly hours of production workers	no	monthly from January 1958 to March 2003
174	Masonry, stonework, and plastering	number of all employees	no	monthly from January 1972 to March 2003
174	Masonry, stonework, and plastering	number of production workers	no	monthly from January 1972 to March 2003
174	Masonry, stonework, and plastering	average weekly hours of production workers	no	monthly from January 1972 to March 2003
175	Carpentry and floor work	number of all employees	no	monthly from January 1972 to March 2003
175	Carpentry and floor work	number of production workers	no	monthly from January 1972 to March 2003
175	Carpentry and floor work	average weekly hours of production workers	no	monthly from January 1972 to March 2003
176	Roofing, siding, and sheet metal work	number of all employees	no	monthly from January 1958 to March 2003
176	Roofing, siding, and sheet metal work	number of production workers	no	monthly from January 1958 to March 2003
176	Roofing, siding, and sheet metal work	average weekly hours of production workers	no	monthly from January 1958 to March 2003

Table C.5 Current Employment Statistics Survey with NAICS Classification (Bureau of Labor Statistics)

NAICS code	Category	Variable	Seasonal adjustment	Coverage
	Construction	number of all employees	both	monthly from January 1939
	Construction	number of production workers	both	monthly from January 1947
	Construction	average weekly hours of production workers	both	monthly from January 1948
236	Construction of buildings	number of all employees	both	monthly from January 1990
236	Construction of buildings	number of production workers	no	monthly from January 1990
237	Construction of buildings	average weekly hours of production workers	no	monthly from January 1990
2361	Construction of residential buildings	number of all employees	both	monthly from January 1985
2361	Construction of residential buildings	number of production workers	no	monthly from January 1990
2362	Construction of residential buildings	average weekly hours of production workers	no	monthly from January 1990
236115	New single-family general contractors	number of all employees	no	monthly from January 1990
236115	New single-family general contractors	number of production workers	no	monthly from January 1990
236115	New single-family general contractors	average weekly hours of production workers	no	monthly from January 1990
236116	New multifamily general contractors	number of all employees	no	monthly from January 1990
236118	Residential remodelers	number of all employees	no	monthly from January 1990
236118	Residential remodelers	number of production workers	no	monthly from January 1990
236118	Residential remodelers	average weekly hours of production workers	no	monthly from January 1990
2362	Construction of nonresidential buildings	number of all employees	both	monthly from January 1990
2362	Construction of nonresidential buildings	number of production workers	no	monthly from January 1990
2362	Construction of nonresidential buildings	average weekly hours of production workers	no	monthly from January 1990
23621	Construction of industrial buildings	number of all employees	no	monthly from January 1990
23621	Construction of industrial buildings	number of production workers	no	monthly from January 1990
23621	Construction of industrial buildings	average weekly hours of production workers	no	monthly from January 1990
23622	Construction of commercial buildings	number of all employees	no	monthly from January 1990
23622	Construction of commercial buildings	number of production workers	no	monthly from January 1990
23622	Construction of commercial buildings	average weekly hours of production workers	no	monthly from January 1990
237	Heavy and civil engineering construction	number of all employees	both	monthly from January 1990
237	Heavy and civil engineering construction	number of production workers	no	monthly from January 1990
237	Heavy and civil engineering construction	average weekly hours of production workers	no	monthly from January 1990
2371	Utility system construction	number of all employees	no	monthly from January 1990
2371	Utility system construction	number of production workers	no	monthly from January 1990
2371	Utility system construction	average weekly hours of production workers	no	monthly from January 1990
23711	Water and sewer system construction	number of all employees	no	monthly from January 1990
23711	Water and sewer system construction	number of production workers	no	monthly from January 1990
23711	Water and sewer system construction	average weekly hours of production workers	no	monthly from January 1990

Table C.5 Current Employment Statistics Survey with NAICS Classification (Bureau of Labor Statistics)

NAICS code	Category	Variable	Seasonal adjustment	Coverage
23712	Oil and gas pipeline construction	number of all employees	no	monthly from January 1990
23712	Oil and gas pipeline construction	number of production workers	no	monthly from January 1990
23712	Oil and gas pipeline construction	average weekly hours of production workers	no	monthly from January 1990
23713	Power and communication system construction	number of all employees	no	monthly from January 1990
23713	Power and communication system construction	number of production workers	no	monthly from January 1990
23713	Power and communication system construction	average weekly hours of production workers	no	monthly from January 1990
2372	Land subdivision	number of all employees	no	monthly from January 1990
2372	Land subdivision	number of production workers	no	monthly from January 1990
2372	Land subdivision	average weekly hours of production workers	no	monthly from January 1990
2373	Highway, street, and bridge construction	number of all employees	no	monthly from January 1990
2373	Highway, street, and bridge construction	number of production workers	no	monthly from January 1990
2373	Highway, street, and bridge construction	average weekly hours of production workers	no	monthly from January 1990
2379	Other heavy construction	number of all employees	no	monthly from January 1990
2379	Other heavy construction	number of production workers	no	monthly from January 1990
2379	Other heavy construction	average weekly hours of production workers	no	monthly from January 1990
238	Specialty trade contractors	number of all employees	both	monthly from January 1976
238	Specialty trade contractors	number of production workers	no	monthly from January 1976
238	Specialty trade contractors	average weekly hours of production workers	no	monthly from January 1976
part of 238	Residential specialty trade contractors	number of all employees	both	monthly from January 2001
part of 238	Nonresidential specialty trade contractors	number of all employees	both	monthly from January 2001
2381	Building foundation and exterior contractors	number of all employees	no	monthly from January 1990
2381	Building foundation and exterior contractors	number of production workers	no	monthly from January 1990
2381	Building foundation and exterior contractors	average weekly hours of production workers	no	monthly from January 1990
part of 2381	Residential building foundation and exterior contractors	number of all employees	no	monthly from January 2001
part of 2381	Nonresidential specialty trade contractors	number of all employees	no	monthly from January 2001
23811	Poured concrete structure contractors	number of all employees	no	monthly from January 1990
23811	Poured concrete structure contractors	number of production workers	no	monthly from January 1990
23811	Poured concrete structure contractors	average weekly hours of production workers	no	monthly from January 1990
23812	Steel and precast concrete contractors	number of all employees	no	monthly from January 1990
23812	Steel and precast concrete contractors	number of production workers	no	monthly from January 1990
23812	Steel and precast concrete contractors	average weekly hours of production workers	no	monthly from January 1990
23813	Framing contractors	number of all employees	no	monthly from January 1990
23813	Framing contractors	number of production workers	no	monthly from January 1990
23813	Framing contractors	average weekly hours of production workers	no	monthly from January 1990
23814	Masonry construction	number of all employees	no	monthly from January 1990
23814	Masonry construction	number of production workers	no	monthly from January 1990
23814	Masonry construction	average weekly hours of production workers	no	monthly from January 1990

Table C.5 Current Employment Statistics Survey with NAICS Classification (Bureau of Labor Statistics)

NAICS code	Category	Variable	Seasonal adjustment	Coverage
23815	Glass and glazing contractors	number of all employees	no	monthly from January 1990
23815	Glass and glazing contractors	number of production workers	no	monthly from January 1990
23815	Glass and glazing contractors	average weekly hours of production workers	no	monthly from January 1990
23816	Roofing contractors	number of all employees	no	monthly from January 1990
23816	Roofing contractors	number of production workers	no	monthly from January 1990
23816	Roofing contractors	average weekly hours of production workers	no	monthly from January 1990
23817	Siding contractors	number of all employees	no	monthly from January 1990
23819	Other building exterior contractors	number of all employees	no	monthly from January 1990
2382	Building equipment contractors	number of all employees	no	monthly from January 1990
2382	Building equipment contractors	number of production workers	no	monthly from January 1990
2382	Building equipment contractors	average weekly hours of production workers	no	monthly from January 1990
2382	Residential building equipment contractors	number of all employees	no	monthly from January 2001
part of 2382	Nonresidential building equipment contractors	number of all employees	no	monthly from January 2001
23821	Electrical contractors	number of all employees	no	monthly from January 1990
23821	Electrical contractors	number of production workers	no	monthly from January 1990
23821	Electrical contractors	average weekly hours of production workers	no	monthly from January 1990
23822	Plumbing and HVAC contractors	number of all employees	no	monthly from January 1990
23822	Plumbing and HVAC contractors	number of production workers	no	monthly from January 1990
23822	Plumbing and HVAC contractors	average weekly hours of production workers	no	monthly from January 1990
23829	Other building equipment contractors	number of all employees	no	monthly from January 1990
23829	Other building equipment contractors	number of production workers	no	monthly from January 1990
23829	Other building equipment contractors	average weekly hours of production workers	no	monthly from January 1990
2383	Building finishing contractors	number of all employees	no	monthly from January 1990
2383	Building finishing contractors	number of production workers	no	monthly from January 1990
2383	Building finishing contractors	average weekly hours of production workers	no	monthly from January 1990
part of 2383	Residential building finishing contractors	number of all employees	no	monthly from January 2001
part of 2383	Non residential building finishing contractors	number of all employees	no	monthly from January 2001
23831	Drywall and insulation contractors	number of all employees	no	monthly from January 1990
23831	Drywall and insulation contractors	number of production workers	no	monthly from January 1990
23831	Drywall and insulation contractors	average weekly hours of production workers	no	monthly from January 1990
23832	Painting and wall covering contractors	number of all employees	no	monthly from January 1990
23832	Painting and wall covering contractors	number of production workers	no	monthly from January 1990
23832	Painting and wall covering contractors	average weekly hours of production workers	no	monthly from January 1990

Table C.5 Current Employment Statistics Survey with NAICS Classification (Bureau of Labor Statistics)

NAICS code	Category	Variable	Seasonal adjustment	Coverage
23833	Flooring contractors	number of all employees	no	monthly from January 1990
23833	Flooring contractors	number of production workers	no	monthly from January 1990
23833	Flooring contractors	average weekly hours of production workers	no	monthly from January 1990
23834	Tile and terrazzo contractors	number of all employees	no	monthly from January 1990
23834	Tile and terrazzo contractors	number of production workers	no	monthly from January 1990
23834	Tile and terrazzo contractors	average weekly hours of production workers	no	monthly from January 1990
23835	Finish carpentry contractors	number of all employees	no	monthly from January 1990
23835	Finish carpentry contractors	number of production workers	no	monthly from January 1990
23835	Finish carpentry contractors	average weekly hours of production workers	no	monthly from January 1990
23839	Other building finishing contractors	number of all employees	no	monthly from January 1990
23839	Other building finishing contractors	number of production workers	no	monthly from January 1990
23839	Other building finishing contractors	average weekly hours of production workers	no	monthly from January 1990
2389	Other specialty trade contractors	number of all employees	no	monthly from January 1990
2389	Other specialty trade contractors	number of production workers	no	monthly from January 1990
2389	Other specialty trade contractors	average weekly hours of production workers	no	monthly from January 1990
part of 2389	Other residential trade contractors	number of all employees	no	monthly from January 2001
part of 2389	Other nonresidential trade contractors	number of all employees	no	monthly from January 2001
23891	Site preparation contractors	number of all employees	no	monthly from January 1990
23891	Site preparation contractors	number of production workers	no	monthly from January 1990
23891	Site preparation contractors	average weekly hours of production workers	no	monthly from January 1990
23899	All other specialty trade contractors	number of all employees	no	monthly from January 1990
23899	All other specialty trade contractors	number of production workers	no	monthly from January 1990
23899	All other specialty trade contractors	average weekly hours of production workers	no	monthly from January 1990

Table C.6 GDP by Industry (Bureau of Economic Analysis)

IO code	Description	Variables	Data availability
230110	New residential 1-unit structures, nonfarm	gross output, quantity indexes, price indexes	annually from 1998-2007
230120	New multifamily housing structures, nonfarm	gross output, quantity indexes, price indexes	annually from 1998-2007
230130	New residential additions and alterations, nonfarm	gross output, quantity indexes, price indexes	annually from 1998-2007
230140	New farm housing units and additions and alterations	gross output, quantity indexes, price indexes	annually from 1998-2007
230210	Manufacturing and industrial buildings	gross output, quantity indexes, price indexes	annually from 1998-2007
230220	Commercial and institutional buildings	gross output, quantity indexes, price indexes	annually from 1998-2007
230230	Highway, street, bridge, and tunnel construction	gross output, quantity indexes, price indexes	annually from 1998-2007
230240	Water, sewer, and pipeline construction	gross output, quantity indexes, price indexes	annually from 1998-2007
230250	Other new construction	gross output, quantity indexes, price indexes	annually from 1998-2007
230310	Maintenance and repair of farm and nonfarm residential structures	gross output, quantity indexes, price indexes	annually from 1998-2007
230320	Maintenance and repair of nonresidential buildings	gross output, quantity indexes, price indexes	annually from 1998-2007
230330	Maintenance and repair of highways, streets, bridges, and tunnels	gross output, quantity indexes, price indexes	annually from 1998-2007
230340	Other maintenance and repair construction	gross output, quantity indexes, price indexes	annually from 1998-2007
	Description	Variables	Data availability
	construction	value added	annually from 1998-2008
	construction	compensation to employees, wages and salaries, supplements to wages and salaries, gross operating surplus	annually from 1998-2007
	construction	chain-type quantity indexes for energy inputs, chain-type quantity indexes for material inputs, chain-type quantity indexes for purchased service inputs	annually from 1997-2007
	construction	full-time and part-time employees	annually from 1948-1997

Table C.7 Producer Price Indices (Bureau of Labor Statistics)

Code	Product	Coverage
236211	New industrial building construction	monthly since June 2007
236221	New warehouse building construction	monthly since December 2004
236222	New school building construction	monthly since December 2005
236223	New office building construction	monthly since June 2006
23816	Roofing contractors, nonresidential work	monthly since December 2007
23811	Concrete contractors, nonresidential work	monthly since December 2007
23821	Electrical contractors, nonresidential work	monthly since December 2007
23822	Plumbing/HVAC contractors, nonresidential work	monthly since December 2007

Table C.8 Producer Price Indices for Materials and Supply Inputs to Construction Industries
(Bureau of Labor Statistics)

Code	Grouping	Coverage
BCON	Inputs to construction industries	monthly from June 1986
BNEW	New construction	monthly from June 1986
BRS1	Single-unit residential	monthly from June 1986
BRSM	Multi-unit residential	monthly from June 1986
BBLD	Non-residential buildings	monthly from June 1986
BHWY	Highway and street construction	monthly from June 1986
BHVY	Other heavy construction	monthly from June 1986
BMRP	Maintenance and repair construction	monthly from June 1986
BMRS	Residential	monthly from June 1986
BMNR	Non-residential	monthly from June 1986

Table C.9 National Income and Product Accounts Private Fixed Investment by Structure Type (Bureau of Economic Analysis)

Category	Variables	Coverage
Private fixed investment in structures	real private fixed investment, quantity index and price index	annual since 1929
Nonresidential	real private fixed investment, quantity index and price index	annual since 1929
Commercial and healthcare	real private fixed investment, quantity index and price index	annual since 1929
Office	real private fixed investment, quantity index and price index	annual since 1929
Health care	real private fixed investment, quantity index and price index	annual since 1929
Hospitals and special care	real private fixed investment, quantity index and price index	annual since 1929
Hospitals	real private fixed investment, quantity index and price index	annual since 1929
Special care	real private fixed investment, quantity index and price index	annual since 1929
Medical buildings	real private fixed investment, quantity index and price index	annual since 1929
Multimerchandise shopping	real private fixed investment, quantity index and price index	annual since 1929
Food and beverage establishments	real private fixed investment, quantity index and price index	annual since 1929
Warehouses	real private fixed investment, quantity index and price index	annual since 1929
Other commercial	real private fixed investment, quantity index and price index	annual since 1929
Manufacturing	real private fixed investment, quantity index and price index	annual since 1929
Power and communication	real private fixed investment, quantity index and price index	annual since 1929
Power	real private fixed investment, quantity index and price index	annual since 1929
Electric	real private fixed investment, quantity index and price index	annual since 1929
Other Power	real private fixed investment, quantity index and price index	annual since 1929
Communication	real private fixed investment, quantity index and price index	annual since 1929
Mining exploration, shafts, and wells	real private fixed investment, quantity index and price index	annual since 1929
Petroleum and natural gas	real private fixed investment, quantity index and price index	annual since 1929
Mining	real private fixed investment, quantity index and price index	annual since 1929
Other structures	real private fixed investment, quantity index and price index	annual since 1929
Religious	real private fixed investment, quantity index and price index	annual since 1929
Educational and vocational lodging	real private fixed investment, quantity index and price index	annual since 1929
Amusement and recreation	real private fixed investment, quantity index and price index	annual since 1929
Transportation	real private fixed investment, quantity index and price index	annual since 1929
Air	real private fixed investment, quantity index and price index	annual since 1929
Land	real private fixed investment, quantity index and price index	annual since 1929
Farm	real private fixed investment, quantity index and price index	annual since 1929
Other	real private fixed investment, quantity index and price index	annual since 1929
Brokers' commission on sale of structures	real private fixed investment, quantity index and price index	annual since 1929
Net Purchases of used structures		

Table C.9 National Income and Product Accounts Private Fixed Investment by Structure Type (Bureau of Economic Analysis)

Category	Variables	Coverage
Residential	real private fixed investment, quantity index and price index	annual since 1929
Permanent site	real private fixed investment, quantity index and price index	annual since 1929
Single-family structures	real private fixed investment, quantity index and price index	annual since 1929
Multifamily structures	real private fixed investment, quantity index and price index	annual since 1929
Other structures	real private fixed investment, quantity index and price index	annual since 1929
Manufactured homes	real private fixed investment, quantity index and price index	annual since 1929
Dormitories	real private fixed investment, quantity index and price index	annual since 1929
Improvements	real private fixed investment, quantity index and price index	annual since 1929
Brokers' commission on sale of structures	real private fixed investment, quantity index and price index	annual since 1929
Net Purchases of used structures		
Addenda	real private fixed investment, quantity index and price index	annual since 1929
Private fixed investment in new structures	real private fixed investment, quantity index and price index	annual since 1929
Nonresidential structures	real private fixed investment, quantity index and price index	annual since 1929
Residential structures	real private fixed investment, quantity index and price index	annual since 1929

Table C.10 National Income and Product Accounts Investment in Government Fixed Assets (Bureau of Economic Analysis)

Category	Variables	Coverage
Government fixed assets	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Equipment and software	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Structures	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Residential	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Industrial	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Office	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Commercial	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Health care	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Educational	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Public safety	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Amusement and recreation	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Transportation	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Power	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Highways and streets	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Military facilities	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Conservation and development	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Other structures	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Federal	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
National defense	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Equipment and software	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Aircraft	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Missiles	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Ships	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Vehicles	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Electronics and software	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Other equipment	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Structures	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Buildings	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Residential	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Industrial	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Military facilities	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929

Table C.10 National Income and Product Accounts Investment in Government Fixed Assets (Bureau of Economic Analysis)

Category	Variables	Coverage
Nondefense	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Equipment and software	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Structures	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Office	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Commercial	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Health care	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Educational	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Public safety	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Amusement and recreation	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Transportation	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Power	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Highways and streets	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Conservation and development	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Other structures	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
State and local	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Equipment and software	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Structures	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Residential	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Office	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Commercial	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Health care	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Educational	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Public safety	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Amusement and recreation	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Transportation	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Power	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Highways and streets	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Sewer systems	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Water systems	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Conservation and development	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Other structures	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929

Table C.10 National Income and Product Accounts Investment in Government Fixed Assets (Bureau of Economic Analysis)

Category	Variables	Coverage
Addenda:	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
General government fixed assets	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Equipment and software	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Structures	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Government enterprise fixed assets	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Equipment and software	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Structures	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Government nonresidential fixed assets	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Equipment and software	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Structures	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Federal	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Defense	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
Nondefense	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929
State and local	investment in government fixed assets (current dollars and chain-type quantity indexes)	annual from 1929

Table C.11 Value of Construction Put in Place (C30 Data), Old Classification System (Census Bureau)

Category	Variable	Seasonal adjustment	Constant or current dollars	Coverage	Comparable with new classification system
Total construction	value of construction put in place	both	both	monthly from 1964 to 2002	yes for annual and not seasonally adjusted monthly data
Private construction	value of construction put in place	both	both	monthly from 1964 to 2002	yes for annual and not seasonally adjusted monthly data
Residential buildings	value of construction put in place	both	both	monthly from 1964 to 2002	no
New housing units	value of construction put in place	both	both	monthly from 1964 to 2002	no
1 unit	value of construction put in place	both	both	monthly from 1964 to 2002	no
2 or more units	value of construction put in place	both	both	monthly from 1964 to 2002	no
Improvements	value of construction put in place	both	both	annual from 1964 to 2002	no
Nonresidential buildings	value of construction put in place	both	both	monthly from 1964 to 2002	no
Industrial	value of construction put in place	both	both	monthly from 1964 to 2002	no
Office	value of construction put in place	both	both	monthly from 1964 to 2002	no
Hotels, motels	value of construction put in place	both	both	monthly from 1964 to 2002	no
Other commercial	value of construction put in place	both	both	monthly from 1964 to 2002	no
Religious	value of construction put in place	both	both	monthly from 1964 to 2002	no
Educational	value of construction put in place	both	both	monthly from 1964 to 2002	no
Hospital and institutional	value of construction put in place	both	both	monthly from 1964 to 2002	no
Miscellaneous	value of construction put in place	both	both	monthly from 1964 to 2002	no
Farm nonresidential	value of construction put in place	both	both	annual from 1964 to 2002	no
Public utilities	value of construction put in place	both	both	annual from 1964 to 2002	no
Telecommunications	value of construction put in place	both	both	monthly from 1964 to 2002	no
Railroad	value of construction put in place	both	both	annual from 1964 to 2002	no
Electric light and power	value of construction put in place	both	both	annual from 1964 to 2002	no
Gas	value of construction put in place	both	both	annual from 1964 to 2002	no
Petroleum pipelines	value of construction put in place	both	both	annual from 1964 to 2002	no
All other private	value of construction put in place	both	both	monthly from 1964 to 2002	no
Public construction	value of construction put in place	both	both	monthly from 1964 to 2002	yes for annual and not seasonally adjusted monthly data
Buildings	value of construction put in place	both	both	monthly from 1964 to 2002	no
Housing and redevelopment	value of construction put in place	both	both	monthly from 1964 to 2002	no
Industrial	value of construction put in place	both	both	monthly from 1964 to 2002	no
Educational	value of construction put in place	both	both	monthly from 1964 to 2002	no
Hospital	value of construction put in place	both	both	monthly from 1964 to 2002	no
Other	value of construction put in place	both	both	monthly from 1964 to 2002	no
Highways and streets	value of construction put in place	both	both	monthly from 1964 to 2002	no
Military facilities	value of construction put in place	both	both	monthly from 1964 to 2002	no
Conservation and development	value of construction put in place	both	both	monthly from 1964 to 2002	no

Table C.11 Value of Construction Put in Place (C30 Data), Old Classification System (Census Bureau)

Category	Variable	Seasonal adjustment	Constant or current dollars	Coverage	Comparable with new classification system
Other public construction	value of construction put in place	both	both	monthly from 1964 to 2002	no
Sewer systems	value of construction put in place	both	both	monthly from 1964 to 2002	no
water supply facilities	value of construction put in place	both	both	monthly from 1964 to 2002	no
Miscellaneous nonbuilding	value of construction put in place	both	both	monthly from 1964 to 2002	no
Total public construction	value of construction put in place	both	both	monthly from 1964 to 2002	yes for annual and not seasonally adjusted monthly data
State and local construction	value of construction put in place	both	both	monthly from 1964 to 2002	yes for annual and not seasonally adjusted monthly data
Total building	value of construction put in place	both	both	monthly from 1964 to 2002	no
Housing and redevelopment	value of construction put in place	both	both	monthly from 1964 to 2002	no
Educational	value of construction put in place	both	both	monthly from 1964 to 2002	no
Hospital	value of construction put in place	both	both	monthly from 1964 to 2002	no
Other	value of construction put in place	both	both	monthly from 1964 to 2002	no
Highways and streets	value of construction put in place	both	both	monthly from 1964 to 2002	no
Conservation and development	value of construction put in place	both	both	monthly from 1964 to 2002	no
Other state and local construction	value of construction put in place	both	both	monthly from 1964 to 2002	no
Sewer systems	value of construction put in place	both	both	monthly from 1964 to 2002	no
Water supply facilities	value of construction put in place	both	both	monthly from 1964 to 2002	no
Miscellaneous nonbuilding	value of construction put in place	both	both	monthly from 1964 to 2002	no
Federal construction	value of construction put in place	both	both	monthly from 1964 to 2002	yes for annual and not seasonally adjusted monthly data
Total building	value of construction put in place	both	both	monthly from 1964 to 2002	no
Housing	value of construction put in place	both	both	monthly from 1964 to 2002	no
Industrial	value of construction put in place	both	both	monthly from 1964 to 2002	no
Educational	value of construction put in place	both	both	monthly from 1964 to 2002	no
Hospital	value of construction put in place	both	both	monthly from 1964 to 2002	no
Other federal buildings	value of construction put in place	both	both	monthly from 1964 to 2002	no
Highways and streets	value of construction put in place	both	both	annual from 1964 to 2002	no
Military Facilities	value of construction put in place	both	both	monthly from 1964 to 2002	no
Conservation and development	value of construction put in place	both	both	monthly from 1964 to 2002	no
Miscellaneous nonbuilding	value of construction put in place	both	both	monthly from 1964 to 2002	no

Total private construction includes public safety, highway and street, sewage and waste disposal, water supply, and conservation and development, which are not reported separately.

Table C.12 Value of Construction Put in Place (C30 Data), New Classification System (Census Bureau)

Category	Variable	Seasonal adjustment	Constant or current dollars	Coverage	Comparable with old classification system
Total construction	value of construction put in place	both	current dollars	monthly since 1993	yes for annual and not seasonally adjusted monthly data
Residential	value of construction put in place	both	current dollars	monthly since 2002	no
Nonresidential	value of construction put in place	both	current dollars	monthly since 2002	no
Lodging	value of construction put in place	both	current dollars	monthly since 2002	no
Office	value of construction put in place	both	current dollars	monthly since 2002	no
Commercial	value of construction put in place	both	current dollars	monthly since 2002	no
Health care	value of construction put in place	both	current dollars	monthly since 2002	no
Educational	value of construction put in place	both	current dollars	monthly since 2002	no
Religious	value of construction put in place	both	current dollars	monthly since 2002	no
Public safety	value of construction put in place	both	current dollars	monthly since 2002	no
Amusement and recreation	value of construction put in place	both	current dollars	monthly since 2002	no
Transportation	value of construction put in place	both	current dollars	monthly since 2002	no
Communication	value of construction put in place	both	current dollars	monthly since 2002	no
Power	value of construction put in place	both	current dollars	monthly since 2002	no
Highway and street	value of construction put in place	both	current dollars	monthly since 2002	no
Sewage and waste disposal	value of construction put in place	both	current dollars	monthly since 2002	no
Water supply	value of construction put in place	both	current dollars	monthly since 2002	no
Conservation and development	value of construction put in place	both	current dollars	monthly since 2002	no
Manufacturing	value of construction put in place	both	current dollars	monthly since 2002	no
Total private construction	value of construction put in place	both	current dollars	monthly since 1993	yes for annual and not seasonally adjusted monthly data
Residential	value of construction put in place	both	current dollars	monthly since 1993	no
Nonresidential	value of construction put in place	both	current dollars	monthly since 1993	no
Lodging	value of construction put in place	both	current dollars	monthly since 1993	no
Office	value of construction put in place	both	current dollars	monthly since 1993	no
Commercial	value of construction put in place	both	current dollars	monthly since 1993	no
Health care	value of construction put in place	both	current dollars	monthly since 1993	no
Educational	value of construction put in place	both	current dollars	monthly since 1993	no
Religious	value of construction put in place	both	current dollars	monthly since 1993	no
Amusement and recreation	value of construction put in place	both	current dollars	monthly since 1993	no
Transportation	value of construction put in place	both	current dollars	monthly since 1993	no
Communication	value of construction put in place	both	current dollars	monthly since 1993	no
Power	value of construction put in place	both	current dollars	monthly since 1993	no
Manufacturing	value of construction put in place	both	current dollars	monthly since 1993	no

Table C.12 Value of Construction Put in Place (C30 Data), New Classification System (Census Bureau)

Category	Variable	Seasonal adjustment	Constant or current dollars	Coverage	Comparable with old classification system
Total public construction	value of construction put in place	both	current dollars	monthly since 1993	yes for annual and not seasonally adjusted monthly data
Residential	value of construction put in place	both	current dollars	monthly since 2002	no
Nonresidential	value of construction put in place	both	current dollars	monthly since 2002	no
Office	value of construction put in place	both	current dollars	monthly since 2002	no
Commercial	value of construction put in place	both	current dollars	monthly since 2002	no
Health care	value of construction put in place	both	current dollars	monthly since 2002	no
Educational	value of construction put in place	both	current dollars	monthly since 2002	no
Public safety	value of construction put in place	both	current dollars	monthly since 2002	no
Amusement and recreation	value of construction put in place	both	current dollars	monthly since 2002	no
Transportation	value of construction put in place	both	current dollars	monthly since 2002	no
Power	value of construction put in place	both	current dollars	monthly since 2002	no
Highway and street	value of construction put in place	both	current dollars	monthly since 2002	no
Sewage and waste disposal	value of construction put in place	both	current dollars	monthly since 2002	no
Water supply	value of construction put in place	both	current dollars	monthly since 2002	no
Conservation and development	value of construction put in place	both	current dollars	monthly since 2002	no

Total private construction includes public safety, highway and street, sewage and waste disposal, water supply, and conservation and development, which are not reported separately.

Total public construction includes lodging, religious, communication and manufacturing, which are not reported separately.

Table C.13 Industry Series of the 1992 Economic Census

SIC code	Description
1521	General Contractors— Single-Family Houses
1522	General Contractors— Residential Buildings, Other Than Single-Family Houses
1531	Operative Builders
1541	General Contractors— Industrial Buildings and Warehouses
1542	General Contractors— Nonresidential Buildings, Other Than Industrial Buildings and Warehouses
1611	Highway and Street Construction Contractors, Except Elevated Highways
1622	Bridge, Tunnel, and Elevated Highway Construction Contractors
1623	Water, Sewer, Pipeline, and Communications and Power Line Construction
1629	Heavy Construction Contractors, Not Elsewhere Classified
1711	Plumbing, Heating, and Air-Conditioning Special Trade Contractors
1721	Painting and Paper Hanging Special Trade Contractors
1731	Electrical Work Special Trade Contractors
1741	Masonry, Stone Setting, and Other Stone Work Special Trade Contractors
1742	Plastering, Drywall, Acoustical and Insulation Work Special Trade Contractors
1743	Terrazzo, Tile, Marble, and Mosaic Work Special Trade Contractors
1751	Carpentry Work Special Trade Contractors
1752	Floor Laying and Other Floor Work Special Trade Contractors, Not Elsewhere Classified
1761	Roofing, Siding, and Sheet Metal Work Special Trade Contractors
1771	Concrete Work Special Trade Contractors
1781	Water Well Drilling Special Trade Contractors
1791	Structural Steel Erection Special Trade Contractors
1793	Glass and Glazing Work Special Trade Contractors
1794	Excavation Work Special Trade Contractors
1795	Wrecking and Demolition Work Special Trade Contractors
1796	Installation or Erection of Building Equipment Special Trade Contractors, Not Elsewhere Classified
1799	Special Trade Contractors, Not Elsewhere Classified

Table C.14 Industry Series of the 1997 Economic Census

NAICS code	SIC code	Description
233110		Land Subdivision and Land Development
	655200	Land subdividers and developers, except cemeteries
233210		Single-Family Housing Construction
	152100	General contractors--single-family houses
	153110	Operative builders (pt)
	874121	Management services (pt)
233220		Multifamily Housing Construction
	152220	General contractors--residential buildings, other than single-family
	153120	Operative builders (pt)
	874122	Management services (pt)
233310		Manufacturing and Industrial Building Construction
	153130	Operative builders (pt)
	154120	General contractors--industrial buildings and warehouses
	874123	Management services (pt)
233320		Commercial and Institutional Building Construction
	152210	General contractors--residential buildings, other than single-family (pt)
	153140	Operative builders (pt)
	154110	General contractors--industrial buildings and warehouses (pt)
	154200	General contractors--nonresidential buildings, except industrial buildings and warehouses
	874124	Management services (pt)
234110		Highway and Street Construction
	161100	Highway and street construction contractors, except elevated highways
	874131	Management services (pt)
234120		Bridge and Tunnel Construction
	162200	Bridge, tunnel, and elevated highway construction contractors
	874132	Management services (pt)
234910		Water, Sewer, Pipeline Construction
	162310	Water, sewer, pipeline, and communication and power line construction (pt)
	874133	Management services (pt)
234920		Power and Communication Transmission Line Construction
	162320	Water, sewer, pipeline, and communication and power line construction (pt)
	874134	Management services (pt)
234930		Industrial Nonbuilding Structure Construction
	162910	Heavy construction, not elsewhere classified (pt)
	874135	Management services (pt)
234990		All Other Heavy Construction
	162920	Heavy construction, not elsewhere classified (pt)
	735320	Heavy construction, equipment rental (pt)
	874136	Management services (pt)
235110		Plumbing, Heating, and Air-Conditioning Contractors
	171100	Plumbing, heating, and air-conditioning special trade contractors
235210		Painting and Wall Covering Contractors
	172100	Painting and paper hanging special trade contractors
	179910	Special trade contractors, not elsewhere classified (pt)
235310		Electrical Contractors
	173100	Electrical work special trade contractors

Table C.14 Industry Series of the 1997 Economic Census

NAICS code	SIC code	Description
235410		Masonry and Stone Contractors
	174100	Masonry, stone setting, and other stone work special trade contractors
235420		Drywall, Plastering, Acoustical, and Insulation Contractors
	174200	Plastering, drywall, acoustical, and insulation work special trade contractors
	174310	Terrazzo, tile, marble, and mosaic work special trade contractors
	177110	Concrete work special trade contractors (pt)
235430		Tile, Marble, Terrazzo, and Mosaic Contractors
	174320	Terrazzo, tile, marble, and mosaic work special trade contractors (pt)
235510		Carpentry Contractors
	175100	Carpentry work special trade contractors
235520		Floor Laying and Other Floor Contractors
	175200	Floor laying and other floor work special trade contractors, not elsewhere classified
235610		Roofing, Siding, and Sheet Metal Contractors
	176100	Roofing, siding, and sheet metal work special trade contractors
235710		Concrete Contractors
	177120	Concrete work special trade contractors (pt)
235810		Water Well Drilling Contractors
	178100	Water well drilling special trade contractors
235910		Structural Steel Erection Contractors
	179100	Structural steel erection special trade contractors
235920		Glass and Glazing Contractors
	179300	Glass and glazing work special trade contractors
	179920	Special trade contractors, not elsewhere classified (pt)
235930		Excavation Contractors
	179400	Excavation work special trade contractors
235940		Wrecking and Demolition Contractors
	179500	Wrecking and demolition work special trade contractors
235950		Building Equipment and Other Machinery Installation Contractors
	179600	Install or erection of building equipment, special trade contractors, not elsewhere classified
235990		All Other Special Trade Contractors
	179940	Special trade contractors, not elsewhere classified (pt)

Table C.15 Industry Series of the 2002 Economic Census

NAICS code	Description
236	Construction of Buildings
236115	New Single-Family Housing Construction (except Operative Builders)
236116	New Multifamily Housing Construction (except Operative Builders)
236117	New Housing Operative Builders
236118	Residential Remodelers
236210	Industrial Building Construction
236220	Commercial and Institutional Building Construction
237	Heavy and Civil Engineering Construction
237110	Water and Sewer Line and Related Structures Construction
237120	Oil and Gas Pipeline and Related Structures Construction
237130	Power and Communication Line and Related Structures Construction
237210	Land Subdivision
237310	Highway, Street, and Bridge Construction
237990	Other Heavy and Civil Engineering Construction
238	Specialty Trade Contractors
238110	Poured Concrete Foundation and Structure Contractors
238120	Structural Steel and Precast Concrete Contractors
238130	Framing Contractors
238140	Masonry Contractors
238150	Glass and Glazing Contractors
238160	Roofing Contractors
238170	Siding Contractors
238190	Other Foundation, Structure, and Building Exterior Contractors
238210	Electrical Contractors
238220	Plumbing, Heating, and Air-Conditioning Contractors
238290	Other Building Equipment Contractors
238310	Drywall and Insulation Contractors
238320	Painting and Wall Covering Contractors
238330	Flooring Contractors
238340	Tile and Terrazzo Contractors
238350	Finish Carpentry Contractors
238390	Other Building Finishing Contractors
238910	Site Preparation Contractors
238990	All Other Specialty Trade Contractors

Table C.16 Variables Reported by Economic Census

Variables	How values are reported	Availability	Notes	
Number of establishments	U.S. total	1977, 1982, 1987, 1992, 1997, 2002		
	by state (location of establishment)	1992, 1997, 2002		
	by size of establishment (number of employees)	1992, 1997, 2002		
	by size of establishment (value of business done)	1992, 1997, 2002		
	by project-type specialization	1992, 1997, 2002		
	by kind-of-business specialization	1997, 2002		
Proprietors and working partners	U.S. total	1977, 1982, 1987, 1992		
Total number of employees	U.S. total	1977, 1982, 1987, 1992, 1997, 2002		
	by state (location of establishment)	1987, 1992, 1997, 2002		
	by size of establishment (number of employees)	1987, 1992, 1997, 2002		
	by size of establishment (value of business done)	1987, 1992, 1997, 2002		
	by project-type specialization	1992, 1997, 2002		
	by kind-of-business specialization	1997, 2002		
Number of construction workers	U.S. total	1977, 1982, 1987, 1992, 1997, 2002		
	by state (location of establishment)	1992, 1997, 2002		
March	U.S. total	1977, 1982, 1987, 1992, 1997, 2002		
May	by state (location of establishment)	1992, 1997, 2002		
August	U.S. total	1977, 1982, 1987, 1992, 1997, 2002		
November	by state (location of establishment)	1992, 1997, 2002		
Number of other employees	U.S. total	1977, 1982, 1987, 1992, 1997, 2002		
	by state (location of establishment)	1997, 2002		
	March	U.S. total	1977, 1982, 1987, 1997, 2002	
	May	by state (location of establishment)	1997, 2002	
	August	U.S. total	1987, 1992, 1997, 2002	
	November	by state (location of establishment)	1997, 2002	
	U.S. total	1987, 1992, 1997, 2002		
	by state (location of establishment)	1997, 2002		
	March	U.S. total	1977, 1982, 1987, 1992, 1997, 2002	
	May	by state (location of establishment)	1997, 2002	
Total payroll	U.S. total	1977, 1982, 1987, 1992, 1997, 2002		
	by state (location of establishment)	1992, 1997, 2002		
	by size of establishment (number of employees)	1992, 1997, 2002		
	by size of establishment (value of business done)	1992, 1997, 2002		
	by project-type specialization	1992, 1997, 2002		
	by kind-of-business specialization	1997, 2002		
Construction workers	U.S. total	1977, 1982, 1987, 1992, 1997, 2002		
Other employees	by state (location of establishment)	1992, 1997, 2002		
U.S. total	1977, 1982, 1987, 1992, 1997, 2002			
by state (location of establishment)	1997, 2002			
First-quarter payroll, all employees	U.S. total	1977, 1982, 1987, 1992, 1997, 2002		
Fringe benefits, all employees	U.S. total	1977, 1982, 1987, 1992, 1997, 2002		
Legally required expenditures	U.S. total	1977, 1982, 1987, 1992, 1997, 2002		
Voluntary expenditures	U.S. total	1977, 1982, 1987, 1992, 1997, 2002		

Table C.16 Variables Reported by Economic Census

Variables	How values are reported	Availability	Notes
Value of business done	U.S. total	1977, 1982, 1987, 1992, 1997, 2002	
	by size of establishment (number of employees)	1992, 1997, 2002	
	by size of establishment (value of business done)	1992, 1997, 2002	
	by type of business	1987, 1992, 1997, 2002	
Value of construction work	U.S. total	1977, 1982, 1987, 1992, 1997, 2002	
	by state (location of establishment)	1987, 1992, 1997, 2002	
	by state (location of construction work)	1987, 1992, 1997, 2002	
	by size of establishment (number of employees)	1987, 1992, 1997, 2002	
	by size of establishment (value of business done)	1987, 1992, 1997, 2002	
	by project-type specialization	1992, 1997, 2002	
	by kind-of-business specialization	1997	
Value of construction work on government owned projects	U.S. total	1977, 1982, 1987, 1992, 1997, 2002	
Value of construction work on federally owned projects	U.S. total	1987, 1992, 1997, 2002	
Value of construction work on state and locally owned projects	U.S. total	1987, 1992, 1997, 2002	
Value of construction work on privately owned projects	U.S. total	1977, 1982, 1987, 1992, 1997, 2002	
Other business receipts	U.S. total	1997, 2002	
Cost of construction work subcontracted out to others	U.S. total	1977, 1982, 1987, 1992, 1997, 2002	
	by state (location of establishment)	1992, 1997, 2002	
	by size of establishment (number of employees)	1992, 1997, 2002	
	by size of establishment (value of business done)	1992, 1997, 2002	
	by project-type specialization	1992, 1997, 2002	
	by kind-of-business specialization	1997, 2002	
Value of construction work subcontracted in from others	U.S. total	1977, 1982, 1987, 1992	
	by state (location of establishment)	1992	
Net value of construction work	U.S. total	1977, 1982, 1987, 1992, 1997, 2002	
	by state (location of establishment)	1992, 1997, 2002	
	by size of establishment (number of employees)	1992, 1997, 2002	
	by size of establishment (value of business done)	1992, 1997, 2002	
	by project-type specialization	1992, 1997, 2002	
	by kind-of-business specialization	1997, 2002	
Value added	U.S. total	1977, 1982, 1987, 1992, 1997, 2002	
	by state (location of establishment)	1987, 1992, 1997, 2002	
	by size of establishment (number of employees)	1987, 1992, 1997, 2002	
	by size of establishment (value of business done)	1987, 1992, 1997, 2002	
	by project-type specialization	1992, 1997, 2002	
	by kind-of-business specialization	1997, 2002	
Value of construction work			
New construction	U.S. total and by project type	1987, 1992, 1997, 2002	
Additions, alterations, or reconstruction	U.S. total and by project type	1987, 1992, 1997, 2002	
Maintenance and repair	U.S. total and by project type	1987, 1992, 1997, 2002	
Selected costs	U.S. total	1977, 1982, 1987, 1992, 1997, 2002	
Materials, components, supplies, and fuel	U.S. total	1992, 1997, 2002	
	by state (location of establishment)	1992, 1997, 2002	
	by size of establishment (number of employees)	1992, 1997, 2002	
	by size of establishment (value of business done)	1992, 1997, 2002	
Materials, parts, and supplies	U.S. total	1977, 1982, 1987, 1992, 1997, 2002	
Power, fuels, and lubricants	U.S. total	1977, 1982, 1987, 1992, 1997, 2002	
Purchased electricity	U.S. total	1977, 1982, 1987, 1992, 1997, 2002	
Natural gas and manufactured gas	U.S. total	1977, 1982, 1987, 1992, 1997, 2002	
Gasoline and diesel fuel	U.S. total	1977, 1982, 1987, 1992, 1997, 2002	
On-highway use of gasoline and diesel fuel	U.S. total	1987, 1992, 1997, 2002	
Off-highway use of gasoline and diesel fuel	U.S. total	1987, 1992, 1997, 2002	
All other fuels and lubricants	U.S. total	1977, 1982, 1987, 1992, 1997, 2002	

Table C.16 Variables Reported by Economic Census

Variables	How values are reported	Availability	Notes
End-of-year gross book value of depreciable assets	U.S. total	1987, 1992, 1997, 2002	Buildings and structures vs. machinery and equipments are also reported separately for 1987 and 1998
	by state (location of establishment)	1992, 1997, 2002	
	by size of establishment (number of employees) by size of establishment (value of business done)	1992 1992	Buildings and structures vs. machinery and equipments are also reported separately for 1987 and 1999
Depreciation charges during year	U.S. total	1987, 1992, 1997, 2002	Buildings and structures vs. machinery and equipments are also reported separately for 1987 and 2000
Number of establishments with inventories	U.S. total	1992, 1997, 2002	
Value of construction work for establishments with inventories	U.S. total	1992, 1997, 2002	
End-of-year (Economic Census year), inventories of materials and supplies	U.S. total	1992, 1997, 2002	
End-of-year (one year before Economic Census year), inventories of materials and supplies	U.S. total	1992, 1997, 2002	
Number of establishments with no inventories	U.S. total	1992, 1997, 2002	
Value of construction work for establishments with no inventories	U.S. total	1992, 1997, 2002	
Establishments not reporting inventories	U.S. total	1992, 1997, 2002	
Value of construction work for establishments not reporting inventories	U.S. total	1992, 1997, 2002	

The above variables are reported by the Economic Census for SIC/NAICS categories. This compilation is an incomplete list.

Table C.17 Source Data and Price Indices for BEA's Annual Estimates of Private Fixed Investment in Structures by Type

Component	Major source data	Price index used to deflate the estimates	Description of the price index
Private fixed investment in structures			
Nonresidential			
Commerical and health care			
Office	Census Bureau monthly construction survey (C30)	BEA price index for office buildings	This quality-adjusted index measures changes in costs and is derived using ordinary least squares hedonic regressions based on square foot costs data from the R.S. Means Company.
Healthcare			
Hospitals and special care			
Hospitals	Census Bureau monthly construction survey (C30)	An unweighted average of Census Bureau's single-family houses under construction index and a Turner Construction Company building cost index	The Census Bureau index measures quality-adjusted changes in the price of new single-family homes under construction. The building cost index is a price index for national building construction costs based on current cost.
Special care	Census Bureau monthly construction survey (C30)	Same as those for hospitals	
Medical buildings	Census Bureau monthly construction survey (C30)	Same as those for hospitals	
Multimerchandise shopping	Census Bureau monthly construction survey (C30)	BLS PPI for warehouses	This PPI measures the quality-adjusted cost for new warehouse construction.
Food and beverage establishments	Census Bureau monthly construction survey (C30)	Same as those used for multimerchandise shopping	
Warehouses	Census Bureau monthly construction survey (C30)	Same as those used for multimerchandise shopping	
Other Commercial	Census Bureau monthly construction survey (C30) and judgemental trend	Same as that used for warehouses and BLS price index for mobile structures	This PPI measures changes in the prices of new residential mobile homes.

Table C.17 Source Data and Price Indices for BEA's Annual Estimates of Private Fixed Investment in Structures by Type

Component	Major source data	Price index used to deflate the estimates	Description of the price index
Manufacturing	Census Bureau monthly construction survey (C30)	BEA price index for factories	This quality-adjusted index measures changes in costs. It is derived using ordinary least squares hedonic regressions based on square foot costs data from the R.S. Means Company.
Power and communication			
Power			These indexes are based on prices for materials, labor costs, and prices of mechanical and electrical equipment for steam operated electric plants in six regions and for reinforced concrete buildings and brick buildings in six regions.
Electric	Census Bureau monthly construction survey (C30)	Weighted average of Handy-Whitman price indexes for electric light and power plants and utility buildings	This index is based on prices for materials, labor costs, and prices of mechanical and electrical equipment for gas plants in six regions.
Other power	Census Bureau monthly construction survey (C30)	Handy-Whitman price index for gas plants	This index is derived from data from operating companies and suppliers on construction methods, plant investment, and component costs.
Communication	Census Bureau monthly construction survey (C30)	AUS Telephone Plant index	
Mining exploration, shafts, and wells			
Petroleum and natural gas	Footage drilled and cost per foot from trade sources extrapolated by BLS producer price index for oil and gas wells.	Weighted average of BLS PPIs for drilling oil and gas wells and for oil and gas field services	These indexes measure changes in prices received by domestic producers.
Mining	Census Bureau annual capital expenditures survey	Same as those used for hospitals	

Table C.17 Source Data and Price Indices for BEA's Annual Estimates of Private Fixed Investment in Structures by type

Component	Major source data	Price index used to deflate the estimates	Description of the price index
Other structures			
Religious	Census Bureau monthly construction survey (C30)	Same as those used for hospitals	This PPI measures the quality-adjusted cost for new school construction.
Educational and vocational	Census Bureau monthly construction survey (C30)	BLS PPI for new school construction	
Lodging	Census Bureau monthly construction survey (C30)	Same as those used for hospitals	
Amusement and recreation	Census Bureau monthly construction survey (C30)	Same as those used for hospitals	
Transportation			
Air	Census Bureau monthly construction survey (C30)	Same as those used for hospitals	The BLS employment cost index measures labor costs. The Bureau of Reclamation construction cost trends index tracks costs such as contractor labor and equipment costs for the Bureau's construction projects. The PPI for material and supply inputs measures prices of input commodities, and the PPI for other communication equipment measures prices of signal equipment.
Land	Census Bureau monthly construction survey (C30)	Weighted average of BLS employment cost index (ECI) for construction industry, Bureau of Reclamation construction cost trends for bridges and for power plants, the BLS PPIs for material and supply inputs into construction industries, BLS PPI for other communication equipment, and the price indexes used for hospitals.	
Farm	Census Bureau monthly construction survey (C30)	Same as those used for hospitals	The Handy-Whitman water utility plant index is based on prices for materials, labor costs, and prices of mechanical and electrical equipment for water utilities in six regions. This PPI measures changes in real estate brokerage fees received from nonresidential property sales and rental.
Other	Census Bureau monthly construction survey (C30)	An unweighted average of the Handy-Whitman water utility plant index, Federal Highway Administration Composite index for highways, and those used for hospitals	
Brokers' commissions on sale of structures	Trend-based estimates	BLS PPI for real estate brokerage, nonresidential property sales and rental including land sales and rental	These implicit price deflators reflect the types of buildings bought and sold by the private sector.
Net purchases of used structures	BEA government fixed asset accounts	An unweighted average of the implicit price deflators for nonresidential buildings, for utilities, for farm buildings, and for other private structures	

Table C.17 Source Data and Price Indices for BEA's Annual Estimates of Private Fixed Investment in Structures by Type

Component	Major source data	Price index used to deflate the estimates	Description of the price index
Residential			
Permanent site			
Single-family structures	Census Bureau monthly construction survey	Census Bureau price index for single-family houses under construction index	This index measures changes in the price of new single-family homes under construction.
Multifamily structures	Census Bureau monthly construction survey	Census Bureau price index for multifamily houses under construction	This index measures changes in the price of new multi-family homes under construction.
Other structures			
Manufactured homes	Shipments from trade source and average retail price from Census Bureau monthly survey	BLS PPI for mobile structures	This PPI measures changes in the prices of new mobile homes.
Dormitories	Census Bureau monthly construction survey	Same as that used for single family structures	
Improvements	Census Bureau survey of residential alterations and repair and survey of consumer expenditures	Average of the Census Bureau index for single-family houses under construction, BLS PPI for home maintenance and repair, and BLS employment cost index for construction industry	See single-family structures for a description of the Census Bureau index. The BLS employment cost index measures labor costs in the construction industry. The PPI measures the cost of residential home maintenance and repair.
Brokers' commissions on sale of structures	Number of single-family houses sold and mean sales price from Census Bureau monthly construction survey and trade source	BLS PPI for real estate brokerage, residential property sales and rental	This PPI measures changes in real estate brokerage fees received from residential property sales and rental.
Net purchases of used structures	BEA government fixed asset accounts	Same as that used for single family structures	

Source: Paul R. Lally, "How BEA Accounts for Investment in Private Structures," Survey of Current Business February 2009: 9-15.

Table C.18 Web Links to Key Sources

Table Number	Web Link (URL)
C.1	http://www.osha.gov/pls/imis/sic_manual.html
C.2	http://www.census.gov/eos/www/naics/
C.3	http://www.census.gov/eos/www/naics/
C.4	http://www.bls.gov/ces/cesbtabs.htm
C.5	http://www.bls.gov/ces/
C.6	http://www.bea.gov/industry/gdpbvind_data.htm
C.7	http://www.bls.gov/ppi/
C.8	http://www.bls.gov/ppi/
C.9	http://www.bea.gov/national/nipaweb/SelectTable.asp?Selected=N
C.10	http://www.bea.gov/national/nipaweb/SelectTable.asp?Selected=N
C.11	http://www.census.gov/const/www/c30index.html
C.12	http://www.census.gov/const/www/c30index.html
C.13	http://www.census.gov/epcd/www/92result.html
C.14	http://www.census.gov/epcd/www/ec97stat.htm
C.15	http://www.census.gov/econ/census02/guide/INDRPT23.HTM
C.16	See Economic Censuses for specific years.
C.17	http://www.bea.gov/scb/pdf/2009/02%20February/0209_briefing_structures.pdf

Appendix D Glossary of Selected Terms

Automation and integration technologies:

Automation technologies: Automation technologies focus on the degree to which individual work functions are automated (e.g., supply management and project management).

Integration technologies: Integration technologies focus on the ability to exchange information between work functions and their associated databases (e.g., exchanges of information among supply management and project management functions).

Deflation: The meaning of deflation is the division of the value of some aggregate by a price index - described as a “deflator” - in order to revalue its quantities at the prices of the price reference period or to revalue the aggregate at the general price level of the price reference period.²²⁶

Establishment: An establishment is a business or industrial unit at a single physical location that produces or distributes goods or performs services.²²⁷

Free-rider: A person or organization who benefits from a public good, but neither provides it nor contributes to the cost of collective provision. They thus free ride on the efforts of others. The free-rider problem means that many public goods are under-provided, or have to be provided by governments which can collect taxes or pay for them.²²⁸

Intermediate inputs: Goods and services, other than fixed assets, used as inputs into the production process of an establishment that are produced elsewhere in the economy or are imported. They may be either transformed or used up by the production process. Land, labor, and capital are primary inputs and are not included among intermediate inputs.²²⁹

Nominal prices: Prices charged by providers of general government services such as health and education and prices that are heavily subsidized through government funding or regulated by

²²⁶ OECD Glossary of Statistical Terms. <http://stats.oecd.org/glossary/detail.asp?ID=3019>. Accessed September 3, 2009.

²²⁷ Census Bureau web page on Economic Census. http://factfinder.census.gov/jsp/saff/SAFFInfo.jsp?_pageId=sp2_economic

²²⁸ John Black, *Oxford Dictionary of Economics*. (Oxford and New York: Oxford University Press, 2002).

²²⁹ OECD Glossary of Statistical Terms. <http://stats.oecd.org/glossary/detail.asp?ID=1395>. Accessed September 3, 2009.

government policy. Such prices are not economically significant and therefore do not provide signals of market driven inflation.²³⁰

Price index: A price index reflects an average of the proportionate changes in the prices of a specified set of goods and services between two periods of time. Usually a price index is assigned a value of 100 in some selected base period and the values of the index for other periods are intended to indicate the average percentage change in prices compared with the base period.²³¹

Productivity: The basic concept underlying construction productivity measures is a comparison of the output of a task, project, or industry with the corresponding factors of production (inputs) required to generate that output.

Three dimensions of productivity:

Task: Tasks refer to specific construction activities such as concrete placement or structural steel erection.

Project: Projects are the collection of tasks required for the construction of a new facility or renovation of an existing constructed facility.

Industry: Industry measures are based on the North American Industrial Classification (NAICS) codes for the construction sector and represent the total portfolio of projects.

Quantity index: A measure reflecting the average of the proportionate changes in the quantities of a specified set of goods and services between two periods of time. Usually a quantity index is assigned a value of 100 in some selected base period and the values of the index for other periods are intended to indicate the average percentage change in quantities compared with the base period. A quantity index is built up from information on quantities such as the number or total weight of goods or the number of services; the quantity index has no meaning from an economic point of view if it involves adding quantities that are not commensurate, although it is often used as a proxy for a volume index.²³²

²³⁰ OECD Glossary of Statistical Terms. <http://stats.oecd.org/glossary/detail.asp?ID=5660>. Accessed September 3, 2009.

²³¹ OECD Glossary of Statistical Terms. <http://stats.oecd.org/glossary/detail.asp?ID=2110>. Accessed September 3, 2009.

²³² OECD Glossary of Statistical Terms. <http://stats.oecd.org/glossary/detail.asp?ID=2221>. Accessed September 3, 2009.

Real terms: Attempts to reduce changes in economic variables to changes in quantities. Real GDP, for example, is the value of gross national product, measured at current prices, deflated by a GDP deflator, or price index.²³³

Sustainability: Sustainability is defined as “meeting the needs of the present generation without compromising the ability of future generations to meet their own needs.”²³⁴

Value added for the construction industry: Value added for the construction industry is defined as the dollar value of business done less costs for construction work subcontracted to others and payments for materials, components, supplies, and fuels.²³⁵

Value of construction put in place: The value of construction put in place is a measure of the value of construction installed or erected at the site during a given period. For an individual project, this includes (1) cost of materials installed or erected, (2) cost of labor (both by contractors and force account) and a proportionate share of the cost of construction equipment rental, (3) contractor’s profit, (4) cost of architectural and engineering work, (5) miscellaneous overhead and office costs chargeable to the project on the owner’s books, and (6) interest and taxes paid during construction (except for state and locally owned projects).²³⁶

Workface Planning: Workface Planning is the process of organizing and delivering all the elements necessary, before work is started, to enable craft persons to perform quality work in a safe, effective and efficient manner.²³⁷

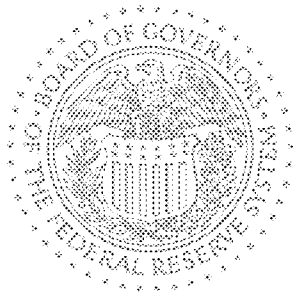
²³³ John Black, *Oxford Dictionary of Economics*. *Op. cit.*

²³⁴ Brundtland GH, editor. 1987. Our common future: Report of the UN Commission on Environment and Development. <http://www.un-documents.net/wced-ocf.htm>. Accessed September 3, 2009.

²³⁵ Economic Census 2002.

²³⁶ Census Bureau Construction Spending Methodology. <http://www.census.gov/const/C30/definitions.pdf>

²³⁷ Construction Owners Association of Alberta (COAA) Best Practices XVI Conference – Workface Planning (WFP) Plenary Presentation. 2003. <http://www.workfaceplan.com/archive.htm>. Accessed September 4, 2009.

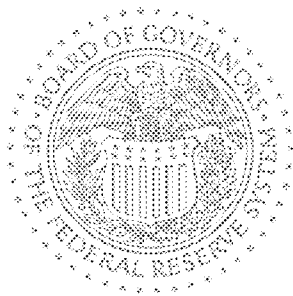


Incentive Compensation Practices:
A Report on the Horizontal Review
of Practices at Large Banking
Organizations

October 2011

BOARD OF GOVERNORS OF THE FEDERAL RESERVE SYSTEM

FTC_AR_0000259



Incentive Compensation Practices:
A Report on the Horizontal Review
of Practices at Large Banking
Organizations

October 2011

BOARD OF GOVERNORS OF THE FEDERAL RESERVE SYSTEM

FTC_AR_0000261

To order additional copies of this or other Federal Reserve Board publications, contact:

Publications Fulfillment
Mail Stop N-127
Board of Governors of the Federal Reserve System
Washington, DC 20551
(ph) 202-452-3245
(fax) 202-728-5886
(e-mail) Publications-BOG@frb.gov

This and other Federal Reserve Board reports are also available online at
www.federalreserve.gov/boarddocs/rptcongress/default.htm.

Contents

Executive Summary	1
Steps Taken by Firms	1
Scope and Status of Reform Effort	3
Introduction	5
Pre-Crisis Conditions and Response	5
Risk-Based Adjustments to Compensation	5
Principles of the Interagency Guidance and Supervisory Expectations	9
Affected Bank Personnel: Executive and Non-Executive Employees	9
Four Methods for Linking Compensation and Risk	9
Avoiding “One-Size-Fits-All” Limits or Formulas	10
Well-Designed Management and Control Functions	10
Timelines for Adoption	10
Incentive Compensation Horizontal Review	11
Scope of the Horizontal Review and Feedback Provided	11
Balancing Incentives at Large Banking Organizations	13
Topic 1: Risk Adjustment and Performance Measures	13
Topic 2: Deferred Incentive Compensation	15
Topic 3: Other Methods that Promote Balanced Risk-Taking Incentives	17
Topic 4: Covered Employees	18
Risk Management, Controls, and Corporate Governance	21
Topic 5: Risk-Management and Control Personnel and the Design of Incentive Arrangements	21
Topic 6: Incentive Compensation Arrangements for Staff in Risk-Management and Control Roles	22
Topic 7: Practices Promoting Reliability	23
Topic 8: Strong Corporate Governance	23
International Context	25
Conformance with Interagency Guidance	25
European Union Approach to Deferred Incentive Compensation	25
Conclusion	27

Executive Summary

Risk-taking incentives provided by incentive compensation arrangements in the financial services industry were a contributing factor to the financial crisis that began in 2007. To address such practices, the Federal Reserve first proposed guidance on incentive compensation in 2009 that was adopted by all of the federal banking agencies in June 2010.

To foster implementation of improved practices, in late 2009 the Federal Reserve initiated a multi-disciplinary, horizontal review of incentive compensation practices at 25 large, complex banking organizations.¹ One goal of this horizontal review was to help fill out our understanding of the range of incentive compensation practices across firms and categories of employees within firms. The second, more important goal was to guide each firm in implementing the interagency guidance.

Given the variety of activities at these complex firms, and the number and range of employees who are in a position to assume significant risk, our approach has been to require each firm to develop, under our supervision, its own practices and governance mechanisms to ensure risk-appropriate incentive compensation that accords with the interagency guidance throughout the organization. Supervisors assessed areas of weakness at the firms, in response to which the firms have developed comprehensive plans outlining how those weaknesses will be addressed. These plans, as modified based on comments from supervi-

sors, will be the basis for further progress and evaluation.

As explained in more detail in this report, every firm in the review has made progress during the review in developing practices and procedures that will internalize the principles in the interagency guidance into the management systems in each firm. Many of these changes are already evident in the actual compensation arrangements of firms. For example, senior executives now have more than 60 percent of their incentive compensation deferred on average, higher than illustrative international guidelines agreed by the Financial Stability Board, and some of the most senior executives have more than 80 percent deferred with additional stock retention requirements after deferred stock vests. Moreover, firms are now attentive to risk-taking incentives for large numbers of employees below the executive level—at many firms thousands or tens of thousands of employees—which was not the case before the beginning of the horizontal review, when most firms paid little attention to risk-taking incentives, or were attentive only for the top employees.

Yet every firm also needs to do more. As oversight of incentive compensation moves into the regular supervisory process, the Federal Reserve will continue to work to ensure progress continues both in the implementation of the firms' plans and in the risk-appropriate character of actual compensation practices.

Steps Taken by Firms

With the oversight of the Federal Reserve and other banking agencies, the firms in the horizontal review have implemented new practices to make employees' incentive compensation sensitive to risk. The following is a brief progress report on four key areas of the review. More details can be found in the report:

¹ The financial institutions in the Incentive Compensation Horizontal Review are Ally Financial Inc.; American Express Company; Bank of America Corporation; The Bank of New York Mellon Corporation; Capital One Financial Corporation; Citigroup Inc.; Discover Financial Services; The Goldman Sachs Group, Inc.; JPMorgan Chase & Co.; Morgan Stanley; Northern Trust Corporation; The PNC Financial Services Group, Inc.; State Street Corporation; SunTrust Banks, Inc.; U.S. Bancorp; and Wells Fargo & Company; and the U.S. operations of Barclays plc, BNP Paribas, Credit Suisse Group AG, Deutsche Bank AG, HSBC Holdings plc, Royal Bank of Canada, The Royal Bank of Scotland Group plc, Societe Generale, and UBS AG.

- **Effective Incentive Compensation Design.** All firms in the horizontal review have implemented new practices to balance risk and financial results in a manner that does not encourage employees to expose their organizations to imprudent risks. The most widely used methods for doing so are risk adjustment of awards and deferral of payments.

—*Risk adjustments* make the amount of an incentive compensation award for an employee take into account the risk the employee's activities may pose to the organization. At the beginning of the horizontal review, no firm had a well-developed strategy to use risk adjustments and many had no effective risk adjustments. Every firm has made progress in developing appropriate risk adjustments, but most have more work to do to ensure the full range of risks are appropriately balanced. An example of a leading-edge practice that is now used by a few firms is including in internal profit measures used in incentive compensation awards a charge for liquidity risk that takes into account stressed conditions. This reduces incentives to take imprudent liquidity risk. An example of a challenge for many firms is development of policies and procedures to guide judgmental adjustments of incentive compensation awards. Such internal guidelines help promote consistency and effectiveness in incentive compensation decisionmaking.

—*Deferring payout* of a portion of incentive compensation awards can help promote prudent incentives if done in a way that takes into account risk taking, especially bad outcomes. Deferring payouts was fairly common before the crisis, especially for senior executives and highly paid employees. However, pre-crisis deferral arrangements typically were not structured to fully take account of risk or actual outcomes. Almost all firms now use vehicles for some employees that adjust downward the amount of deferred incentive compensation that is paid if losses are large. However, most firms still have work to do to implement such arrangements for a larger set of employees and to more closely link such reductions to individual employees' actions, particularly for employees below the senior executive level.

- **Progress in Identifying Key Employees.** At most large banking organizations, thousands or tens of thousands of employees have a hand in risk taking. Yet, before the crisis, the conventional wisdom at most firms was that risk-based incentives were

important only for a small number of senior or highly paid employees and no firm systematically identified the relevant employees who could, either individually or as a group, influence risk. All firms in the horizontal review have made progress in identifying the employees for whom incentive compensation arrangements may, if not properly structured, pose a threat to the organization's safety and soundness. All firms in the horizontal review now recognize the importance of establishing sound incentive compensation programs that do not encourage imprudent risk taking for those who can individually affect the risk profile of the firm. In addition, slightly more than half of the firms have identified groups of similarly compensated employees whose combined actions may expose the organization to material amounts of risk. However, some firms are still working to identify a complete set of mid- and lower-level employees and to fully assess the risks associated with their activities.

- **Changing Risk-Management Processes and Controls.** Because firms did not consider risk in the design of incentive compensation arrangements before the crisis, firms rarely involved risk-management and control personnel when considering and carrying out incentive compensation arrangements. All firms in the horizontal review have changed risk-management processes and internal controls to reinforce and support the development and maintenance of balanced incentive compensation arrangements. Risk-management and control personnel are engaged in the design and operation of incentive compensation arrangements of other employees to ensure that risk is properly considered. Some firms have further work to do to provide sufficiently active and robust engagement by risk management and control staff.
- **Progress in Altering Corporate Governance Frameworks.** At the outset of the horizontal review, the boards of directors of most firms had begun to consider the relationship between incentive compensation and risk, though many were focused exclusively on the incentive compensation of their firm's most senior executives. Since then, all firms in the horizontal review have made progress in altering their corporate governance frameworks to be attentive to risk-taking incentives created by the incentive compensation process for employees throughout the firm. The role of boards of directors in incentive compensation has expanded, as has the amount of risk information provided to boards related to incentive compensation. The

appropriateness of the degree of engagement of the boards will be evaluated after a few years of experience.

Scope and Status of Reform Effort

Supervisors in the horizontal review gathered confidential supervisory information from all firms and found important differences in practices across business lines and banking organizations. Additionally, practices are changing rapidly in response to the Federal Reserve's efforts and industry developments. Therefore, a moment-in-time, comparative analysis of individual firms from the horizontal review is not possible and could be misleading. That said, the Federal Reserve is working to foster market discipline in the area of incentive compensation. On this front, the

Federal Reserve intends to implement the Basel Committee's recent "Pillar 3 disclosure requirements for remuneration," issued in July 2011,² which will provide more complete information about risk-related elements of incentive compensation practices of individual institutions.

In part spurred by the horizontal review, incentive compensation practices at banking organizations are continuing to evolve and develop. We expect this evolution to continue. The Federal Reserve will continue to work with these firms through the supervisory process to ensure improvement and progress are sustained.

² See "Pillar 3 disclosure requirements on remuneration issued by the Basel Committee," *Bank for International Settlements*, (www.bis.org/press/p110701.htm).

Introduction

Risk-taking incentives provided by incentive compensation arrangements in the financial services industry were a contributing factor to the financial crisis that began in 2007. To address such practices, the Federal Reserve first proposed guidance on incentive compensation in 2009 that was adopted by all of the federal banking agencies in June 2010. In 2009, the Federal Reserve announced a horizontal review of incentive compensation practices at a group of large, complex banking organizations. (See “Principles of the Interagency Guidance and Supervisory Expectations” on page 9 and “Incentive Compensation Horizontal Review” on page 11.)

Pre-Crisis Conditions and Response

As discussed in the interagency guidance, the activities of employees may create a wide range of risks for a banking organization, such as credit, market, liquidity, operational, legal, compliance, and reputational risks, as well as other risks to the viability or operation of the organization. Some of these risks may be realized in the short term, while others may become apparent only over the long term. For example, future revenues that are booked as current income may not materialize, and short-term profit-and-loss measures may not appropriately reflect differences in the risks associated with the revenue derived from different activities. In addition, some risks—or combinations of risky strategies and positions—may have a low probability of being realized but would have highly adverse effects on the organization if they were to be realized (“bad tail risks”). While shareholders may have less incentive to guard against bad tail risks because of the infrequency of their realization and the existence of the federal safety net, these risks warrant special attention for safety-and-soundness reasons given the threat they pose to the organization’s solvency and the federal safety net.

Before the crisis, large banking organizations did not pay adequate attention to risk when designing and

operating their incentive compensation systems, and some employees were provided incentives to take imprudent risks. For example, an employee who made a high-risk loan may have generated more revenue in the short run than one who made a low-risk loan. Incentive compensation arrangements based solely on the level of short-term revenue paid more to the employee taking more risk, thereby incentivizing employees to take more, sometimes imprudent, risk. Led by supervisors in the horizontal review, over the past two years banking organizations have improved their incentive compensation arrangements to take appropriate account of risk. The two most common ways to do so—risk adjustments and deferral—make use of risk information that becomes available at different points in time.

Risk-Based Adjustments to Compensation

Information about risks taken that is known before incentive compensation is awarded can be used to make risk adjustments to those awards. For example, if an employee in a lending unit makes many high-risk loans during a year, the estimated profit from the loans can be adjusted when designing the employee’s incentive compensation package, using either quantitative or qualitative information. In all cases, risk adjustments should consider likely losses under stressed conditions, and not merely business-as-usual, so that larger, but lower-probability, loss outcomes can be taken into account.

Both quantitative and qualitative risk information can be used in making such adjustments. They can be applied either through use of a formula or through the exercise of judgment and may play a role in setting amounts of incentive compensation pools (bonus pools), in allocating pools to individuals’ incentive compensation, or both. The effectiveness of the different types of adjustments varies with the situation of the employee and the banking organization, as well as the thoroughness of their implemen-

tation. Banking organizations in the horizontal review have made significant progress in improving their risk adjustments, but most still have work to do. The first topic in “Balancing Incentives at Large Banking Organizations” on page 13 describes the main types of risk adjustments and some areas in which further work is needed.³

Deferred incentive compensation can contribute to prudent incentives because risk taking and risk outcomes often become clearer over time. If payout of a portion of incentive compensation awards is deferred for a period of time after the award date, late-arriving information about risk taking and outcomes of such risk taking can be used to alter the payouts in ways that will improve the balance of risk-taking incentives. Banking organizations in the horizontal review have made progress in improving deferral practices, but many still have work to do on performance conditions for vesting. Deferral practices are described in the second topic in “Balancing Incentives at Large Banking Organizations” on page 15.

Risk adjustments and deferral are not the only ways of improving the balance of risk-taking incentives. Some alternatives, such as the use of longer performance periods when evaluating employees’ performance and awards and reducing the sensitivity of awards to measures of short-term performance are briefly described in the third topic in “Balancing Incentives at Large Banking Organizations” on page 17.

At the beginning of the horizontal review, the conventional wisdom at most firms was that risk-taking incentives were important only for a small number of senior or highly paid employees. Though the decisions and incentives of senior executives are indeed very important, the combined risk taking by a group of similarly compensated employees can also be material to the firm’s risk profile. Thus, identifying the set of employees, who may individually or collectively expose the firm to material amounts of risk, is a key element of practice. The interagency guidance notes that such “covered employees” should include not only those who can individually affect the risk profile of the firm, but also groups of similarly compensated employees whose actions when taken together can affect the risk profile. Examples of such groups may include many types of traders and loan originators. Most firms in the horizontal review have

made progress in identifying covered employees, but some still have work to do. The fourth topic in “Balancing Incentives at Large Banking Organizations” on page 18 discusses covered employees and progress in identifying them.

As described in the interagency guidance, establishment of prudent risk-taking incentives should be critically supported by risk-management and control personnel. In addition, practices to promote improvements in the reliability and effectiveness of incentive compensation systems over time can usefully support development of prudent risk-taking incentives on a sustained basis. These elements are described in “Risk Management, Controls, and Corporate Governance” on page 21, which notes progress in most areas.

Some observers have been particularly interested in comparing progress of incentive compensation practices of firms headquartered in different jurisdictions. Approximately one-third of the large banking organizations included in the horizontal review are headquartered outside the United States (foreign banking organizations, or FBOs). In general, progress in conforming to the interagency guidance is similar at the U.S. banking organizations and at the FBOs in the horizontal review, and progress in conforming to the Financial Stability Board’s (FSB) *Principles for Sound Compensation Practices* (Principles) and the related *Implementation Standards*,⁴ which are somewhat less demanding than the interagency guidance, is also similar, as described in “International Context” on page 25.

As the horizontal review of incentive compensation practices draws to a close, further work on incentive compensation will continue through the normal supervisory process. Much supervisory work is already focused on risk management and control systems. Risk-taking incentives are a complementary focus for supervisors. However, incentive compensation practices are likely to evolve rapidly over the next several years, so both firms and supervisors must continue to adapt and improve. The Federal Reserve also intends to implement the Basel Committee’s recent “Pillar 3 disclosure requirements for remuneration,” issued in July 2011. Increased public disclosure about risk-related incentive compensation practices at major firms may improve market disci-

³ Employees sometimes take risk in pursuit of goals other than short-term financial performance. In such cases, risk adjustments may also contribute to balanced risk-taking incentives.

⁴ The FSB issued the *Principles* in April 2009 and the *Implementation Standards* in September 2009. These FSB documents are available at www.financialstabilityboard.org/list/fsb_publications/#id_123/index.htm.

pline of such practices. Finally, the Federal Reserve is working with other banking and financial regulatory agencies to develop an interagency rule on incentive

compensation practices, as mandated by the Dodd-Frank Wall Street Reform and Consumer Protection Act (Dodd-Frank Act).

Principles of the Interagency Guidance and Supervisory Expectations

The interagency guidance is anchored by three principles:

1. **Balance between risks and results.** Incentive compensation arrangements should balance risk and financial results in a manner that does not encourage employees to expose their organizations to imprudent risks;
2. **Processes and controls that reinforce balance.** A banking organization's risk-management processes and internal controls should reinforce and support the development and maintenance of balanced incentive compensation arrangements; and
3. **Effective corporate governance.** Banking organizations should have strong and effective corporate governance to help ensure sound incentive compensation practices, including active and effective oversight by the board of directors.

The interagency guidance is consistent with both the FSB *Principles* and *Implementation Standards* adopted in 2009.⁵

Affected Bank Personnel: Executive and Non-Executive Employees

Incentive compensation arrangements for executive and non-executive employees able to control or influence risk taking at a banking organization may pose safety-and-soundness risks if not properly struc-

ured. Accordingly, the interagency guidance applies to senior executives as well as other employees who, either individually or as part of a group of similarly compensated employees, have the ability to expose the banking organization to material amounts of risk. In identifying employees covered by the interagency guidance, banking organizations are directed to consider the full range of inherent risks associated with an employee's work activities, rather than just the level or type of risk that may remain after application of the organization's internal controls for managing risk ("residual risk").

Four Methods for Linking Compensation and Risk

The interagency guidance discusses four methods that banking organizations often use to make incentive compensation more sensitive to risk: (1) risk-adjusting incentive compensation awards based on measurements of risk; (2) deferring payment of awards using mechanisms that allow for actual award payouts to be adjusted as risks are realized or become better known; (3) using longer performance periods (for example, more than one year) when evaluating employees' performance and granting awards; and (4) reducing the sensitivity of awards to measures of short-term performance.⁶ Each method has advantages and disadvantages.

A key premise of the interagency guidance is that the methods used to achieve appropriately risk-sensitive incentive compensation arrangements likely will differ across and within firms. Employees' activities and the risks associated with those activities vary significantly across banking organizations and potentially across employees within a particular banking organization. Differences across firms may be based on their principal chosen lines of business and the char-

⁵ On April 14, 2011, as mandated by the Dodd-Frank Act, the Federal Reserve, along with the Office of the Comptroller of the Currency, the Federal Deposit Insurance Corporation, the former Office of Thrift Supervision, the National Credit Union Administration, the Securities and Exchange Commission, and the Federal Housing Finance Agency, issued for comment a proposed rule on incentive compensation practices. The proposed rule builds off the interagency guidance. This report focuses on the observations from the horizontal review, which was conducted in the context of the interagency guidance and does not discuss the proposed rule. The proposed rule is available at www.gpo.gov/efile/fr/pk/g/FR-2011-04-14/pdt/2011-7237.pdf.

⁶ As noted in the interagency guidance, this list of methods is not intended to be exhaustive—other methods may exist or be developed.

acteristics of the markets in which they operate, among other factors, affecting both the types of risk faced by the firm and the time horizon of those risks. Even within firms, employees' activities and the attendant risks can depend on many different variables, including the specific sales targets or business strategies and the nature and degree of control or influence that different employees may have over risk taking. These differences naturally create different opportunities and different potential incentives, broadly speaking, for employees to take or influence risk. Thus, the use of any single, formulaic approach to incentive compensation by banking organizations or supervisors is unlikely to be effective at addressing all incentives to take imprudent risks.

Avoiding "One-Size-Fits-All" Limits or Formulas

The interagency guidance helps to avoid the potential hazards or unintended consequences that would be associated with rigid, one-size-fits-all supervisory limits or formulas. Subject to supervisory oversight, each organization is responsible for ensuring that its incentive compensation arrangements are consistent with its safety and soundness. Methods for achieving balanced incentive compensation arrangements at one organization may not be effective at another organization, in part because of the importance of integrating incentive compensation arrangements with the firm's own risk-management systems and business model. Similarly, the effectiveness of methods is likely to differ across business lines and units within a large banking organization. In general, large banking organizations are likely to need multiple methods to ensure that incentive compensation arrangements do not encourage imprudent risk taking.

Well-Designed Management and Control Functions

The interagency guidance also places great emphasis on the role of risk-management and internal control functions in providing for balanced risk-taking incentives. Poorly designed or implemented incentive compensation arrangements can themselves be a source of risk to banking organizations and undermine

existing controls. For example, unbalanced incentive compensation arrangements can place substantial strain on the risk-management and internal control functions of even well-managed organizations.

Therefore, risk-management and internal control functions should be involved in designing, implementing, and evaluating incentive compensation arrangements to ensure that the arrangements properly take risk into account.

The interagency guidance recognizes that large banking organizations tend to be significant users of incentive compensation arrangements, and that flawed approaches to incentive compensation at these institutions are more likely to have adverse effects on the broader financial system. Accordingly, the interagency guidance elaborates with greater specificity certain supervisory expectations for large banking organizations.⁷

Timelines for Adoption

In adopting the interagency guidance, the banking agencies recognized that achieving conformance with its terms and principles would likely require significant changes and enhancements to firm practices and that fully implementing such changes would require some time. For the large banking organizations in the horizontal review, we communicated our expectation that each firm should demonstrate significant progress toward consistency with the interagency guidance in 2010, should achieve substantial conformance with the interagency guidance by the end of 2011 (affecting the award of incentive compensation awards for the 2011 performance year), and should fully conform thereafter.

⁷ For example, the interagency guidance states that large banking organizations should have a systematic approach to incentive compensation supported by formalized and well-developed policies, procedures, and systems to ensure that incentive compensation arrangements are appropriately balanced and consistent with safety and soundness. Such institutions should also have robust procedures for collecting information about the effects of their incentive compensation programs on employee risk taking, as well as systems and processes for using this information to adjust compensation arrangements to eliminate or reduce unintended incentives for risk taking. Similarly, the interagency guidance urges large banking organizations to actively monitor industry, academic, and regulatory developments in incentive compensation practices and theory and be prepared to incorporate into their incentive compensation systems new or emerging methods that are likely to improve the organization's long-term financial well-being and safety and soundness.

Incentive Compensation Horizontal Review

In late 2009, in conjunction with its initial proposal of principles-based guidance on incentive compensation, the Federal Reserve launched a special simultaneous, horizontal review of incentive compensation practices and related risk management, internal controls, and corporate governance practices at a group of large complex banking organizations. These firms were chosen because flawed approaches to incentive compensation at these institutions are more likely to have adverse effects on the broader financial system and because of their extensive use of incentive compensation practices. The special work associated with the horizontal review is now nearing completion, but supervisory work on incentive compensation will continue through the ongoing supervisory process.

The Federal Reserve has communicated to the firms our assessment of their practices and our expectations for remediation in areas where improvements are needed. The firms, with the oversight and input of the Federal Reserve, have each developed remediation plans. These remediation plans, along with updates and discussion around them, have been a key mechanism for bringing clarity about needed changes.

Scope of the Horizontal Review and Feedback Provided

To carry out this major supervisory initiative, the Federal Reserve made a substantial commitment of staff resources and senior management attention. More than 150 individuals from the Federal Reserve and the other banking agencies have been involved in the horizontal review. In addition to senior supervisory staff, these included a multidisciplinary group of professionals, including supervisors, economists and lawyers, several specially constituted incentive compensation on-site review teams, and the permanent supervisory teams assigned to each of the involved banking organizations. Federal Reserve staff has coordinated with other banking regulators in con-

ducting the horizontal review and communicating with the firms.

To perform the supervisory assessments of conformance with the interagency guidance, we gathered extensive information from the firms on their incentive compensation arrangements and associated processes, policies, and procedures. We reviewed internal documents governing existing incentive compensation practices as well as self-assessments of incentive compensation practices relative to the interagency guidance. We conducted many face-to-face meetings with senior executive officers and members of boards of directors' compensation committees. To supplement this information and to evaluate specifically how incentive compensation programs were implemented at the line-of-business level, the Federal Reserve conducted focused examinations of incentive compensation practices in trading and mortgage-origination business lines at a number of the organizations involved in the horizontal review.

The Federal Reserve has continued to provide individualized feedback to each of the firms as additional information and updates of remediation plans have been received. All of the firms have made progress toward achieving consistency with the interagency guidance. The nature and extent of remaining work varies across organizations and sometimes within organizations. Achieving conformance with the interagency guidance depends on the successful build-out of systems and processes, achievement of intermediate implementation milestones, and successful completion of remediation plans. Even then, in many cases, it will be important for the firms to keep in mind that new systems and practices have not been fully tested by experience, so ongoing monitoring of these new systems and practices will be important.

With regard to FBOs with activities in the United States, we have acknowledged the particular challenges that arise as they seek to conform their U.S. operations with the details of their home-country

consolidated regulator's expectations and those of the interagency guidance. As noted, the interagency guidance is consistent with international regulatory efforts on incentive compensation practices, including the FSB *Principles and Implementation Standards*. We have indicated our intent to follow the comple-

mentary principles of effective consolidated supervision and national treatment of banking organizations operating in the United States.⁸

⁸ For observations regarding incentive compensation practices at FBOs, see "International Context" on page 25.

Balancing Incentives at Large Banking Organizations

This section describes methods firms use to provide employees with prudent risk-taking incentives, as well as identifies the relevant set of employees. It is mostly related to the first of the three principles in the interagency guidance.

Incentive compensation arrangements achieve balance between risk and financial reward when the amount of money ultimately received by an employee depends not only on the employee's performance, but also on the risks taken in achieving this performance. Firms often determine the dollar amount of incentive compensation awards for a performance year immediately after the end of the year. Part of the award may be paid immediately and part may be deferred. Risk adjustments (see *Topic 1* below) are features of incentive compensation arrangements that incorporate information about risks taken into decisions about the total amount of awards. Deferred payouts can also be adjusted for risk using information that becomes available during the deferral period, as described under *Topic 2*. *Topic 3* focuses on other balancing methods, and *Topic 4* on identification of covered employees (those employees for whom prudent risk-taking incentives are particularly important).

Topic 1: Risk Adjustment and Performance Measures

At the beginning of the horizontal review, no firm had a well-developed strategy to use risk adjustments and many had no effective risk adjustments. Currently, all firms in the horizontal review employ some sort of risk adjustment for at least some subset of employees, but the role of risk adjustments in the overall mix of balancing strategies varies across firms and across businesses within firms. Some adjustments rely on quantitative measures of risk, while others are based on perceptions of risks taken by employees or business units. Quantitative measures of risk may be applied mechanically (although this is relatively unusual) or as an element in judgment-

based decisions. Risk adjustments may play a role in setting amounts of bonus pools, in allocating pools to individuals' incentive compensation, or both. In all cases, risk adjustments should consider likely losses under stressed conditions, and not merely business-as-usual, so that larger, but lower-probability loss outcomes can influence incentives to take risk.

Every firm has made progress in developing and implementing appropriate risk adjustments, but the progress is uneven, not only across firms, but within firms. Substantial work remains to be done to achieve consistency and effectiveness of such adjustments in providing balanced risk-taking incentives. Because most incentive compensation decisions involve some judgment, a key element of that work is improved written policies and procedures and improved monitoring practices.

Disciplined, Judgment-Based Decisionmaking

Judgment is an element of decisionmaking at every firm and at nearly every step in the design and operation of incentive compensation arrangements.⁹ This poses two challenges: (1) ensuring that decisions based on judgment are made consistently can be difficult and (2) risk adjustments may be only one of many inputs into decisionmaking about incentive compensation awards. Without appropriate restraint, judgments about other aspects of an employee's performance, such as achieving a certain level of market share, could be made in a way that would undermine the desired incentive effects of the risk adjustments. To promote consistency and effectiveness of the impact of judgment on balanced risk-taking incentives, the interagency guidance notes that firms are expected to have robust policies and procedures to guide the consistent use of judgment, and that decisions should be documented so that firms can review

⁹ An exception is formulaic compensation plans, such as commission sales plans, which sometimes specify amounts of incentive compensation according to a specific formula set at the beginning of the year.

whether policies and procedures are being followed and can assess the effectiveness of the policies and procedures over time.¹⁰

At the beginning of the horizontal review, most firms lacked written policies and procedures to guide managers in making risk adjustments, and policies and procedures for incentive compensation decisionmaking often did not clearly identify the weight to be given to risks taken during the performance year. Such policies and procedures, along with training for managers and *ex post* review of decisions, are important to achieving consistent application of *risk* adjustments. Some firms have made progress in developing written policies and procedures and related processes, but others are still in the process of completing this work.¹¹

Quantitative and Qualitative Risk Measures

In cases where risk adjustments are applied based on a formula, incentive compensation decisions are made using measures of financial performance that are net of a risk charge based on a quantitative measure of risk. Such adjustments balance incentives to take risk to the extent that such charges offset increases in financial performance (or reductions in costs) that are associated with increased risk taking. The use of mechanical risk adjustments is possible when suitable quantitative risk measures are available, and the effectiveness of this type of risk adjustment depends on the quality of the risk measure. One leading edge practice, observed at some firms, is to assess a charge against internal profit measures for

liquidity risk that takes into account stressed conditions and to use this adjusted profit measure in determining incentive compensation awards.

Most firms in the horizontal review also used quantitative risk measures as an input to judgment-based incentive compensation decisionmaking. For example, boards of directors usually take into account available risk measures when making decisions about bonus pools for the firm or about awards for senior executives. Some risk measures can be difficult to convert into quantitative risk charges, but nevertheless convey useful information. However, as noted previously, achieving a consistent balancing impact through judgmental decisionmaking is a challenge. Firms with more well-developed policies and procedures to guide decisionmakers in judgmentally using quantitative risk information seemed more likely to achieve a consistent balancing impact. This is an area in which many firms are working to improve effectiveness.

Almost all firms in the horizontal review use non-quantitative perceptions of risk taking as a basis for some risk adjustments. Such adjustments have the potential to address hard-to-measure risks and limitations of existing data and risk-measurement methods. For example, the manager of a lending business might be aware that some employees of the business make riskier loans and others safer loans, even though the quantitative risk measures available to the manager do not show it. Based on this information, the manager could risk adjust by giving lower incentive compensation awards per unit of revenue to the employees making the riskier loans. As in other cases where incentive compensation awards are based on judgment-based decisionmaking, they are more likely to be consistently effective where firms have clear policies and procedures to guide application. Developing such policies and procedures is particularly challenging because the information about risk is qualitative and the nature of the information tends to change over time.

Risk Adjustment and Bonus Pools

Incentive compensation practices of firms differ in the process of determining the total bonus pools and the allocation of incentive compensation to individuals. In a top-down process, senior management and the board of directors determine the size of an overall amount of funding for the firm as a whole near the end of the performance year, and this bonus pool is then split into sub-pools for each business. Pools

¹⁰ For example, an organization should have policies and procedures that describe how managers are expected to exercise judgment to achieve balance, including a description, as warranted, of the appropriate available information about the employee's risk-taking activities to be considered in making informed judgments. Such policies and procedures need not involve a precise analysis to be followed in developing discretionary risk adjustments, but should provide enough structure and instruction that decisions can be justified and documented on a clear and consistent basis and thereby allow for *ex post* monitoring.

¹¹ Some firms have identified in their policies and procedures specific factors appropriate to the line of business and employee role, including reference points, to be considered by management when making discretionary risk adjustments. Some firms have introduced new management processes aimed at governing discretion-based risk adjustments and aimed at providing documentation sufficient to support review of such decisions by Internal Audit. Some firms also have assigned control-function employees to focus on compliance with enhanced policies and procedures, and on documentation processes. They have improved communication to managers and employees about how risk adjustments work, which is crucial to full impact on risk-taking decisions.

are allocated to individual employees in a manner related to their individual performance. In a bottom-up process, the firm assesses performance of each employee and assigns him or her an incentive compensation award, with the total amount of incentive compensation for the year for the firm as a whole simply being the sum of individual incentive compensation awards. Most firms' processes are a mixture of top-down and bottom-up, but the emphasis can differ markedly.¹²

Risk adjustments balance incentive compensation arrangements to the extent they affect the incentives provided to individuals. The impact on incentives may be limited in cases where a firm makes risk adjustments only when deciding amounts of pools because the award to each employee under the pool will receive the same adjustment. This is appropriate when the nature and extent of risk taking of all employees under the pool is the same, such as cases where a pool applies to a business unit in which all risk decisions are influenced in the same way by all employees. Where individual employees in a single pool can have varied levels of impact on the amount of risk, the differences will not be fully addressed by risk adjustments to the pool alone. In such cases, additional adjustments incorporated into decisions about individual incentive compensation awards would be needed to make the risk adjustment fully effective.

Next Steps

Most of the firms in the horizontal review have made significant changes to their risk adjustment practices for awards for the 2011 performance year. Still, most continue to have work to do, including development of appropriate policies and procedures to guide judgmental adjustments of incentive compensation awards. Most firms should continue to evaluate the effectiveness of the quantitative and qualitative risk adjustments they are using and whether risks are appropriately balanced. Additionally, in 2012 firms should evaluate how effective the risk adjustments used for the 2011 awards were, and make improvements as necessary. The Federal Reserve will continue to work with the firms to make sure progress contin-

ues and to evaluate best practices in this area as they evolve.

Topic 2: Deferred Incentive Compensation

Another method for balancing incentive compensation arrangements is to defer the actual payout of a portion of an award to an employee significantly beyond the end of the performance period, adjusting the payout for actual losses or other aspects of the employee's performance that are realized or become better known only during the deferral period. Such deferral arrangements make it possible for the amount ultimately paid to the employee to reflect information about risks taken that arrives during the deferral period.

The interagency guidance does not require that deferral be used for all employees; does not suggest any specific formula for deferral arrangements; and does not mandate the use of any specific vehicle for payment, such as stock. However, the interagency guidance does have some specific suggestions relating to deferral arrangements for senior executives. A substantial fraction of incentive compensation awards should be deferred for senior executives of the firm because other methods of balancing risk-taking incentives are less likely to be effective by themselves for such individuals.

Elements of Deferral Practices

The proportion of incentive compensation awards to be deferred was substantial at the firms in the horizontal review. For example, senior executives now have more than 60 percent of their incentive compensation deferred on average, higher than illustrative international guidelines agreed by the FSB, and some of the most senior executives have more than 80 percent deferred with additional stock retention requirements after deferred stock vests. Most firms assign deferral rates to employees using a fixed schedule or "cash/stock table" under which employees receiving higher incentive compensation awards generally are subject to higher deferral rates, though deferral rates for the most senior executives are often set separately and are higher than those for other employees.

Deferral periods generally range from three to five years, with three years the most common. Most organizations in the horizontal review use the same deferral period for all employees in a given incentive com-

¹² Even at firms with a bottom-up emphasis, budget constraints place a practical limit on the size of the aggregate bonus for the firm as a whole, so some top-down element is present. Similarly, top-down firms take some account of perceived performance of key individuals in setting pools.

pensation plan and often for all employees. Some firms transfer ownership of the entire deferred award to the employee at the end of the vesting period (“cliff vesting”), while others adopted a schedule under which a portion of the award vests at given intervals.

The most common vehicles for conveying deferred incentive compensation to employees are shares of the firm’s stock, stock options, and performance units (an instrument with a payout value that depends on a measure of performance during the deferral period, often an accounting measure like earnings or return-on-equity). Some firms use deferred cash or debt-like instruments.

Performance-Based Deferral

At the beginning of the horizontal review, few firms adjusted payouts of deferred awards for risk outcomes or other information about risks taken that became available during the deferral period. Without such performance conditions, deferral arrangements are unlikely to contribute to balancing risk-taking incentives (for ease of reference, deferral with performance conditions is referred to as “performance-based deferral”).¹³

¹³ Two common issues with performance-based deferral became clear during the horizontal review. The first is related to payment of deferred incentive compensation in share-based instruments. Where vehicles are share-based, at the time shares are awarded, risk-taking actions during the performance year might have either upside or downside effects on the stock price in the future, so the net effect on incentives is not clear. Moreover, most employees below the senior executive level are not likely to believe that their own risk-taking decisions will have a material impact on the firm’s stock price. For example, if the leader of a business unit knows that a particular strategy may lead to losses that are large from the standpoint of the unit, the leader may believe any such losses would be more than offset by profits from other business units. Thus, the leader would not expect the losses to affect the ultimate value of deferred pay received, and deferral would have little impact on his or her risk-taking incentives. In order for a deferral arrangement to meaningfully contribute to balance, vesting triggers should be based on measures of performance that are linked to the employee’s risk-taking activities, especially those taken before the incentive compensation award.

The second common issue that became clear during the horizontal review related to the particular performance conditions (triggers) chosen by firms. Some firms have performance-based deferral arrangements that allow for a large or outsized payout when the values of triggers reflect positive performance. However, these arrangements may encourage employees to take more risk during the deferral period, in order to maximize the value of such triggers and thus may not balance risk-taking incentives. One example of a trigger that may be appropriate is one that reduces the amount of deferred compensation that is vested if the firm (or business line or unit, depending on the level of the employee) experiences negative net income in any fiscal year during the deferral period. The relevant triggers for any

Firms in the horizontal review have made progress in implementing performance-based deferral arrangements that promote balanced risk-taking incentives. Each firm’s setup is somewhat different, but three broad styles of arrangement were observed—formulaic, judgment-based, and a hybrid of the two. In a formulaic approach, the percentage of the award that vests is directly related to a measure of performance during the deferral period. In a judgment-based arrangement, the circumstances under which less than full vesting will occur are decided judgmentally rather than being linked to fixed values of performance metrics, and the amount of incentive compensation paid out under those circumstances is also decided through a judgment-based process. In a hybrid setup, a specific trigger value of performance is set at the beginning of the deferral period, and if performance falls below that trigger value, a judgment-based process determines how much of the deferred incentive compensation will not vest.¹⁴ To the extent that judgment plays a role in the vesting decision, firms are expected to have robust policies and procedures to guide the consistent use of judgment, and decisions should be appropriately documented so that firms can monitor whether their policies and procedures are being followed.¹⁵ Policies and procedures need to be clear to employees, or they will not have a clear understanding when risk-taking decisions are made of which outcomes will lead to forfeiture, in which case deferral arrangements are not likely to have a significant impact on risk-taking behavior. Many firms still have work to do on their policies and procedures in this area.

Most firms in the horizontal review have clawback arrangements for at least some employees that are triggered by malfeasance, violations of the firm’s policies, and material restatement of financial results.¹⁶ Such clawback provisions can contribute to

performance-based deferral arrangement also should be clearly explained to employees covered by those arrangements.

¹⁴ In a common variant of the hybrid process, once the trigger is met for a particular group (e.g., a business unit), the discretionary process determines not only the percentage of incentive compensation that vests, but also which employees are subject to less than full vesting, usually based on which employees were responsible for losses or for imprudent risk taking.

¹⁵ Concerns about the use of discretion in deferral arrangements are similar to concerns about the use of discretion in *ex ante* risk adjustment, as discussed under Topic 1 of this report.

¹⁶ The word “clawback” is sometimes used to refer to any deferral-of-payment method. The term “clawback” also may refer specifically to an arrangement under which an employee must return incentive compensation payments previously received by the employee if certain risk outcomes occur. Section 304 of the Sarbanes-Oxley Act of 2002 (15 U.S.C. 7243), which applies to

balanced risk-taking incentives by discouraging specific types of behavior. While potentially effective, they do not affect most risk-related decisions and are not triggered by most risk outcomes—the narrow focus of these arrangements mean that they are unlikely to contribute meaningfully to balance.

Progress on performance-based deferral for the 2010 performance year was most common for senior executives. Many firms are now in the process of revising arrangements to be used for the 2011 performance year and are extending performance-based deferral coverage to more employees as a mechanism to provide prudent risk-taking incentives. Some firms have implemented, or are implementing, performance-based deferral for all employees receiving deferred incentive compensation, while others are doing so mainly for employees whose authorities and influence over risk taking are such that risk adjustments might have only limited effectiveness in balancing risk-taking incentives, such as senior managers within business lines and other employees engaged in activities that involve risks over a long duration.

Next Steps

Most of the firms in the horizontal review have made significant changes to their deferral arrangements. Many firms in the horizontal review have increased the fraction of incentive compensation that is deferred for both senior executives and other employees. All firms have more work to do to improve their performance-based deferral arrangements. Firms may also fine-tune the role of deferral relative to risk adjustments as they gain experience with how the two work together. As firms develop and fine-tune deferral arrangements, firms should evaluate how well these deferral arrangements have worked and make improvements as necessary. The Federal Reserve will monitor and encourage progress and work to ensure that practices are effective.

Topic 3: Other Methods that Promote Balanced Risk-Taking Incentives

Risk adjustments and deferral with performance-sensitive features represent important mechanisms

chief executive officers and chief financial officers of public banking organizations, is an example of this more specific type of “clawback” requirement. Nearly all U.S.-based firms in the horizontal review are publicly traded, and therefore subject to this provision.

for achieving balanced incentives for taking risk. The interagency guidance also identifies the use of longer performance periods (for example, more than one year) and reduced sensitivity of awards to short-term performance as methods for achieving balance. During the horizontal review, we observed the use of both methods, though neither was universally used.

Evaluating Performance: Emphasis on Long-Term over Short-Term

Firms used longer performance periods (that is, a backward-looking multiyear assessment horizon), for example, for senior executives in some cases, and in others for non-executive employees. Measuring and evaluating performance or awards on a multiyear basis allows for a greater portion of risks and risk outcomes to be observed within the performance assessment horizon, thus garnering many of the benefits of a deferral arrangement with performance-sensitive features. One simple variation involves using risk outcomes from prior-year actions as a consideration in reducing current-year incentive compensation award decisions. To be effective, multiyear assessments should be based on policies and procedures that give appropriate weight to poor outcomes due to past decisions. Otherwise, adverse outcomes may be effectively ignored due to an emphasis on current-year performance.

Damping the sensitivity of incentives to measures of short-term performance was a choice made by some institutions to rein in incentives when, for example, concerns arose about the significance of the incentives or risks involved. For example, increasing bonus pools or individual award amounts at a lower rate when financial performance is well above target levels can limit incentives to take large risks to achieve extreme levels of performance. A cap on incentive compensation awards beyond a certain level of performance is another example. However, in the horizontal review, there were few instances where such caps and reduced sensitivity were sufficient by themselves to balance risk-taking incentives.

Next Steps

The interagency guidance urges large banking organizations to actively monitor industry, academic, and regulatory developments in incentive compensation practices and theory to identify new or emerging methods that are likely to improve the organization’s long-term financial well-being and safety and sound-

ness. The Federal Reserve will do the same and will encourage firms to use methods that are most appropriate for their circumstances.

Topic 4: Covered Employees

Identifying the full set of employees who may individually or collectively expose the firm to material amounts of risk is a crucial step toward managing risks associated with incentive compensation. Without identifying the relevant employees, a firm cannot be sure it has properly designed its incentive compensation arrangements to provide appropriate risk-taking incentives.

Three Categories of Covered Employees

The interagency guidance describes three categories of such employees, which together are referred to as “covered employees”:

- senior executives;
- other individual employees able to take or influence material risks; and
- groups of similarly compensated individuals who, in aggregate, can take or influence material risks.

Incentive compensation arrangements for all covered employees should be appropriately balanced, regardless of whether the covered employee is a senior executive, an individual, or part of a group of similarly compensated individuals. Though the Federal Reserve has no target number or quota of covered employees for any firm, many of the largest firms have determined they have thousands or tens of thousands of covered employees.

Standard Approaches to Covered Employee Identification

Firms follow one of two general approaches to identify covered employees. One approach involves developing and following a systematic process that identifies types of risk that each employee (or group of employees) takes or influences and that assesses the materiality of the risks. Such a process should “cast a wide net” and should consider the full range of types and severities of risk. Some firms have invested in enhanced information systems to facilitate this process. Many firms in the horizontal review follow this approach.

The second approach designates a very large set of employees as covered, such as all employees receiving any incentive compensation, or all employees subject to a subset of the firm’s incentive compensation plans. Although this reduces the effort required to identify covered employees, firms still need to identify the relevant types and severities of risks that are incentivized through incentive compensation arrangements to be sure incentives to take such risks are balanced.

Many firms appropriately identify at least some groups of similarly compensated employees who may collectively expose the firm to material risk. Examples include originators of mortgages, commercial lending officers, or groups of traders subject to similar incentive compensation arrangements.

Establishing Robust Processes Going Forward

Several firms have yet to establish robust processes for identifying covered employees that are consistent with the interagency guidance, especially for identifying groups of covered employees. Some firms rely heavily on mechanical materiality thresholds in their identification process. For example, only employees able to make decisions that commit at least \$1 billion of the firm’s economic capital might be eligible for consideration as covered employees, or only employees above a given level of total compensation. Such materiality thresholds as applied by most firms to exclude employees from being considered covered employees have three common weaknesses: (1) they often fail to capture the full extent to which an employee may expose the firm to risk, (2) they tend to exclude potential covered employees who may significantly influence risk taking but do not make final risk decisions, and (3) they often ignore groups of similarly compensated employees. In reviewing the firms’ use of thresholds, we found that under some circumstances, a suitably chosen materiality threshold could appropriately play a complementary role in identifying covered employees if used to include employees as covered employees.

FBOs with U.S. operations that were part of the horizontal review face special challenges in developing procedures for identifying covered employees for purposes of the interagency guidance. Generally, home-country supervisors expect their standards to be met by the consolidated organization, and so in its

U.S. operations, an FBO must meet both home-country and U.S. regulatory expectations. Many of these firms have home-country supervisors whose regulations focus on a more limited set of employees than described in the interagency guidance.¹⁷ As a result, these firms need to develop processes to identify both covered employees in their U.S. operations for application of the interagency guidance and those employees subject to home-country regulation. The number of covered employees for purposes of the interagency guidance in U.S. operations of an FBO may exceed the number of employees subject to home-country regulation.

Next Steps

All firms in the horizontal review now recognize the importance of establishing sound incentive compen-

sation programs that do not encourage imprudent risk taking for those employees who can individually affect the risk profile of the firm. In addition, many firms have identified groups of similarly compensated employees whose combined actions may expose the organization to material amounts of risk. Some firms have put in place a robust process for identifying relevant individuals and groups of employees, with the flexibility to adapt to the changing business environment over time. However, some firms are still working to identify a complete set of mid- and lower-level employees, and others are working to ensure their process is sufficiently robust. The Federal Reserve will work with the firms to ensure that progress continues.

¹⁷ Supervisors in many other jurisdictions require their firms to identify only their equivalent of individual covered employees, often using materiality standards that restrict attention to a relatively small number of individuals.

Risk Management, Controls, and Corporate Governance

Establishment of balanced risk-taking incentives should be supported by the engagement of risk-management and control personnel in the design and implementation of incentive compensation arrangements, incentive compensation for such personnel that is independent of the financial performance of the businesses they oversee (in order to limit conflicts of interest), practices to promote improvements in the reliability and effectiveness of incentive compensation systems over time, and improvements in corporate governance. These features are discussed in topics 5 through 8 below.

Topic 5: Risk-Management and Control Personnel and the Design of Incentive Arrangements

Properly identifying risks attendant to employees' activities and setting suitable balancing mechanisms are critical elements of providing balanced risk-taking incentives. The interagency guidance notes that risk-management processes and internal controls should reinforce and support the development and maintenance of balanced incentive compensation arrangements. Risk-management and control personnel (including Internal Audit) should be involved in the design, operation, and monitoring of incentive compensation arrangements because their skills and expertise provide essential perspective and support. Risk-management staff, in particular, should participate in the firm's analysis and decisionmaking regarding the identification of covered employees, the selection of any risk-sensitive performance metrics, the development of risk-adjustment methodologies and vesting triggers, and the overall effectiveness of the firm's balancing efforts.

At all firms in the horizontal review, certain functions, such as human resources and finance, traditionally were involved in incentive compensation decisions and in the design and implementation of incentive compensation arrangements. However, this

role traditionally involved little or no focus on incentives to take risk or the risk associated with the employee's activities. Risk-management personnel traditionally had relatively little involvement in incentive compensation design, and their involvement in decisionmaking was often limited, for example, to only supplying information about breaches of internal policy and procedure by individual employees or units. However, a few firms did incorporate risk measures produced by risk-management personnel into financial performance measures used in incentive compensation decisionmaking before the crisis.

Increased Involvement of Risk-Management Personnel in Design and Decisionmaking

Risk-management personnel are now involved in incentive compensation system design and decisionmaking at virtually all firms in the horizontal review. However, the intensity and nature of involvement varies. For example, risk-management functions now provide significant risk-related input to the board-level decisionmaking process for individual senior executive incentive compensation at all firms and for bonus pool size decisions at firms at which pools play a role. Most firms consider some quantitative risk measures in making at least some incentive compensation decisions; and these are usually provided by the risk and finance functions. Nonetheless, at some firms, risk experts primarily play a peripheral or informal role.

Control, finance, and risk-management staff members provide some input to individual employee performance reviews at many firms. For example, they report breaches of policy and procedure or rate the "risk awareness" or adherence to the firm's risk appetite of individual employees or business units. At firms that use committee structures in their incentive compensation decisionmaking process, control, finance, or risk-management personnel usually are among the members of committees. At most firms in

the horizontal review, risk-management and control functions are also involved in identification of covered employees.

At firms where risk-management personnel are intensely involved in basic design decisions for the incentive compensation system, as well as in determining details of the risk-related elements of the incentive compensation process overall, progress on risk-taking incentives has tended to be faster. At firms where risk experts play a peripheral, informal role, progress has tended to be slower, primarily because other personnel tend to have less experience and expertise in designing risk identification and measurement features. Several firms remain in the latter category.

Next Steps

The main challenge going forward is to ensure that risk-management and control personnel are actively engaged with incentive compensation and that improvements in risk management and in recognition of risks the firm takes are incorporated into incentive compensation decisionmaking. The Federal Reserve will continue to work with firms to ensure that such personnel have an appropriate role.

Topic 6: Incentive Compensation Arrangements for Staff in Risk-Management and Control Roles

Improper incentive compensation arrangements can compromise the independence of staff in risk-management and control roles. For example, a conflict of interest is created if the performance measures applied to them, or the bonus pool from which their awards are drawn, depend substantially on the financial results of the lines of business or business activities that such staff oversee. Such dependence can give staff an incentive to allow or foster risk taking that is inconsistent with the firm's risk-management policies and control framework or the safety and soundness of the firm. Thus, risk-management and control personnel should be compensated in a way that makes their incentives independent of the lines of business whose risk taking and incentive compensation they monitor and control. Such staff includes not only employees assigned to firmwide risk-management or control functions, but also employees who perform similar roles while

embedded within individual lines of business within the firm.

Maintaining the Independence of Risk-Management and Control Personnel

The firms in the horizontal review have completed much of the necessary work in this area. Performance measures applied to staff in risk-management and control roles are usually oriented to the performance of their oversight duties and not the performance of the line of business they oversee. Their incentive compensation may be indirectly related to financial performance, if, for example, the bonus pool is drawn from the firmwide pool, which is related to firmwide performance. In most cases, linkage to firmwide performance is likely to be too weakly linked to control and risk-management decisions to pose a significant conflict of interest.

Where more direct or substantial potential conflicts of interest have arisen, some firms achieved independence by moving risk-management and control function personnel out of line-of-business incentive compensation plans or line-of-business bonus pools, establishing separate plans or pools for them. Other firms established separate bonus pools for staff in risk-management and control roles, the sizes of which do not depend directly on the financial performance of a particular line of business or business activity.

At some firms, lower-level risk-management or control staff members who are embedded in business lines receive their incentive compensation awards from the business line bonus pool. Such practices can be acceptable if the relevant staff members perform functions that are unrelated to risk-taking decisions and if the product of their work is unrelated to incentive compensation decisionmaking.

Some firms include comments from cross-function reviews (such as 360 degree reviews) in incentive compensation decisionmaking for all staff members. This raises the possibility that business line reviews could influence incentive compensation decisions for risk-management and control staff members even if no formal link to financial performance exists. In addition, some firms have incentive compensation arrangements for staff in risk-management and control functions that are subject to adjustments based on management judgment. Clear guidance from policies and procedures, clear documentation of indi-

vidual judgment-based adjustments (and decisions made under such policies and procedures), and review by internal audit help to ensure the incentive compensation awards are not swayed by business line results.

Next Steps

As part of its normal supervision of the independence of risk and control functions, the Federal Reserve will continue to be attentive to the risk-related incentives provided by the incentive compensation arrangements for their personnel.

Topic 7: Practices Promoting Reliability

Firms should regularly review whether the design and implementation of their incentive compensation systems deliver appropriate risk-taking incentives and should correct deficiencies and make improvements that are suggested by the findings. The interagency guidance mentions several practices that can contribute to the effectiveness of such activity, including internal reviews and audits of compliance with policies and procedures, monitoring of results relative to expectations, and simulation of the operation of incentive compensation arrangements before implementation.

Importance of Internal Reviews and Audits

Internal reviews and audits of compliance with policies and procedures are important to ensure that the incentive compensation system is implemented as intended by those employees involved in incentive compensation decisionmaking. For example, if procedures require that specific quantitative measures of risk are to be included in financial performance measures used in decisionmaking, but they are not, the sensitivity of decisions to risk taking probably would not be as intended. Though the internal audit function should play a key role in this activity, other functions such as risk management, finance, and human resources also should be involved.

An incentive compensation system may be implemented as intended, but it may still fail to achieve the desired relationship between risk and reward because features of its design and operation do not work out as expected. Detecting such problems requires that a firm monitor relationships among measures of short- and long-run financial performance, amounts of

incentive compensation awards, measures of risk and risk outcomes, amounts of ultimate payments of deferred incentive compensation, and other factors relevant to incentive compensation decisions. Such monitoring bears some resemblance to the “backtesting” that is often done for risk-management models and systems. To be effective, such monitoring should include some quantitative analysis, but because all incentive compensation systems involve some exercise of human judgment in decisionmaking, effective monitoring is not likely to be purely quantitative or mechanical. Large banking organizations are more likely to require some use of automated systems to adequately monitor the effectiveness of incentive compensation arrangements in balancing risk-taking incentives, especially systems that support capture of relevant data in databases that support monitoring and analysis.

Next Steps

All organizations in the horizontal review have considerable work remaining to fully implement practices promoting balanced risk incentives in their incentive compensation arrangements. Few organizations performed extensive reviews and analyses related to risk-taking incentives before the crisis. In some cases internal audit reviewed other aspects of incentive compensation activities, such as incentive compensation award disbursement practices or adherence to vesting policies related to time-of-service.

Over time, as incentive compensation is awarded and paid out and risk outcomes become better known, firms and their supervisors will learn more about the reliability of methods for balancing risk-taking incentives and the effectiveness of different methods of assessing reliability. In the meantime, the Federal Reserve will work with firms as they develop the necessary systems and capabilities and will promote experimentation and innovation.

Topic 8: Strong Corporate Governance

Active and effective oversight of incentive compensation practices by the board of directors is a key element of the interagency guidance. The board of directors of a large banking organization, or its delegated committee, should actively oversee the development and operation of the organization’s incentive compensation policies, systems, and related control

processes. The board of directors or the delegated committees of such organizations should also monitor the effectiveness of incentive compensation arrangements in balancing the risk-taking incentives of covered employees.

Most of the firms in the horizontal review already had in place a board-level compensation committee composed of independent directors. While historically these committees have been actively engaged in decisions relating to the incentive compensation arrangements for certain senior executives, their involvement in overseeing the incentive compensation practices and arrangements relating to other covered employees (including non-executives) has increased considerably during the horizontal review. All firms in the horizontal review have enhanced the role of the board in overseeing the incentive compensation system for all covered employees and are now paying increased attention to risk-related aspects of incentive compensation. Some firms have established management committees that include representatives of risk-management and control functions to support their efforts. Notwithstanding progress made to date, firms indicated that they will continue to implement enhanced corporate governance practices and that these practices will continue to evolve.

Progress in Facilitating Effective Internal Communications

Most firms have established mechanisms to facilitate communication between the compensation committee and the risk and audit committees. Many firms have members of the compensation committee that are also members of the risk and audit committees. Other firms rely on regular meetings between the compensation and risk committees, while others have not yet enhanced their communications systems and rely on communications that are more ad hoc in nature.

The board of directors or its delegated committee should review and approve policies and procedures that appropriately address corporate standards and processes governing the design, approval, administration, and monitoring of incentive compensation arrangements for covered employees. At some firms in the horizontal review, the relevant body is not yet consistently reviewing and approving these standards.

The board of directors should regularly review the results of monitoring of incentive compensation arrangements described in the previous section and results of other activities undertaken to promote reliability of the incentive compensation system. For example, boards should receive periodic reports that review incentive compensation awards and payments relative to risk outcomes on a backward-looking basis to determine whether the organization's incentive compensation arrangements may be promoting imprudent risk taking. As noted previously, at most firms such reports are at a relatively early stage of development. While some boards undertake an annual review of the effectiveness of incentive compensation in avoiding inappropriate incentives to incur risk, many currently rely on periodic presentations by the chief risk officer or other risk-management staff to the board of directors or its compensation committee, the content of which varies considerably from firm to firm.

Next Steps

Though firms have implemented improved corporate governance practices, the effectiveness of such practices will not be known until some years of experience have been accumulated. Effectiveness will depend on the attentiveness of members of compensation committees to risk-taking incentives. The Federal Reserve will continue to work to promote effective governance of incentive compensation practices at banking organizations.

International Context

Some observers have been interested in comparing progress of firms headquartered in different jurisdictions in improving their incentive compensation practices, for example, in progress relative to the FSB *Principles and Implementation Standards*.

About one-third of the large banking organizations included in the horizontal review are headquartered outside the United States. Almost all of the FBOs in the horizontal review are headquartered in Europe (including the United Kingdom). We observed progress in implementing the interagency guidance, which is consistent with the FSB documents, at both U.S. banking organizations and FBOs. However, the interagency guidance, while consistent with the FSB *Principles and Implementation Standards*, is more detailed and demanding in many respects. Thus, satisfying the expectations implied by the FSB documents is not necessarily enough to satisfy the expectations in the interagency guidance.

Conformance with Interagency Guidance

In general, progress on conforming to the interagency guidance is similar at the U.S. banking organizations and at the FBOs in the horizontal review. Firms that are more and less far along can be found in both sets of firms. With respect to particular aspects of the guidance, the FBOs have had more difficulty in identifying covered employees in their U.S. operations (as noted previously, few foreign supervisors employ the concept of groups of covered employees, instead focusing their attention on relatively small numbers of senior and highly paid employees). Progress on conforming to the elements of the interagency guidance that focus on corporate governance and the role of risk-management and control personnel is similar at FBOs and U.S. banking organizations.

Progress on achieving balanced incentive compensation arrangements is similar on the whole across the two groups, but the balancing methods employed and

the rate of innovation are different between the groups. For risk adjustments, some foreign supervisors have emphasized risk adjustments mainly at the level of firmwide or business line bonus pools. Thus, some FBOs have made progress risk adjusting such pools but have made less progress implementing risk adjustments down to the level of the individual employee.

Some observers have been particularly interested in the details of deferral practices, focusing on the share of incentive compensation awards that is deferred and the use of equity as a vehicle for deferred incentive compensation. Numerical examples of deferral fractions set out in the FSB *Principles and Implementation Standards* are sometimes used as a benchmark (60 percent or more for senior executives, 40 percent or more for other individual “material risk takers,” which are not the same as covered employees). Deferral fractions are at or above these benchmarks at both the U.S. banking organizations and the FBOs in the horizontal review.

In some cases, substantial deferral fractions are achieved in different ways. As noted previously, most U.S. firms and some FBOs use a cash-stock table that increases the deferral rate as the amount of incentive compensation increases. As a practical matter, this results in substantial deferral rates for senior executives and for some employees. In contrast, as noted previously, some European Union (EU) supervisors prescribe some elements of pay structure for some employees at EU banking organizations. This also results in substantial deferral rates for those employees.

European Union Approach to Deferred Incentive Compensation

In many cases the pay structure under the EU regulation is somewhat different than that seen at U.S. banking organizations. Under some national implementations within the EU, the deferred portion of an

incentive compensation award is required to be granted half in an equity-linked instrument and half in cash or a cash-like vehicle. The upfront portion of the incentive compensation award is required to be paid half in cash and half in stock subject to a retention requirement of six months to one year. Though the overall fraction of the incentive compensation award granted in stock is substantial in such implementations, the upfront stock subject to a retention requirement is likely to have a limited balancing impact on risk-taking incentives due to the short retention period. The impact of the deferred portion depends on performance conditions; in the absence of performance conditions, deferred cash will have only a modest balancing impact since the amount ultimately received by the employee is reduced only in the event of the firm's failure.

Overall, the net exposure of an employee to a firm's performance over time is not necessarily larger under

the EU regulation than under the simpler structures often seen at U.S. firms. For example, if 60 percent of an incentive compensation award is deferred for three years, half in stock and half in cash that vests unless the firm fails, then only 30 percent of the incentive compensation award is exposed to poor performance short of failure. In contrast, suppose all deferred awards are in stock deferred for three years, as is common in the United States. If the same 60 percent of the incentive compensation award is deferred, the whole 60 percent is exposed to the variation in the value of the stock. If the stock is also subject to effective performance conditions, the whole 60 percent is exposed to the conditions. The details of vesting and other performance conditions are particularly important to the overall balancing impact.

Conclusion

Reinforced by the supervisory activities undertaken through the horizontal review, the large banking organizations in the review have made significant progress toward enhancing their incentive compensation arrangements in ways that provide appropriately balanced incentives to take risks (as outlined in the interagency guidance) and promote safety and soundness. As described in this report, however, most firms still have significant work to do to achieve full conformance with the interagency guidance.

The Federal Reserve remains committed to helping move the industry forward in developing and imple-

menting incentive compensation practices that are consistent with prudent risk management and safety and soundness. Continued supervisory attention will be focused on further refinement and implementation and on making appropriate changes as business conditions change and business strategies evolve.



Talent Wants to Be Free

Orly Lobel

2013

The F.T.C. considered Talent Wants to Be Free as part of the rulemaking process. A copy of Talent Wants to Be Free is on file with the F.T.C.

Governing Innovation in Practice: The Role of the Top Management Team

Companies that compete through new products or services have, of necessity, a new product development system, organization, and process. These have generally been in place for a decade or more and are regularly updated as conditions change and new practices and processes are developed and adopted. As part of these updates, senior functional and business managers may change the allocation of responsibilities for the planning, design, production, and introduction of new offerings. They also invest in new tools and work approaches, and they regularly introduce new targets in terms of new product quality, cost, and lead times. In many companies, this process now works reasonably well and smoothly. Relying on the second line of command to supervise all these activities, the top management team of large companies may not be involved directly, except when there are significant changes, which can include both new opportunities and failures.

But despite having a good product development process in the company, CEOs often complain about the relative lack of market and financial impact of innovation efforts, at least given the investments. New products are developed, for sure, but the results are often disappointing when compared to the predictions and promises of product managers and others responsible for the introduction of innovations. New products may provide a benefit to customers and help to maintain the company's profitability, but too few are real "game changers." How many of these CEOs, reading constantly about Apple's series of market hits, ask themselves how they can emulate the success of the Silicon Valley giant? And reflecting on the fate of Nokia, that troubled innovation star – at least in the eyes of the business media – how many wonder about what suddenly happened to the mobile phone pioneer's top management team? How can one explain why these brilliant Finnish leaders, who could launch new phones with amazing frequency, somehow took their eye off the ball and missed a deep turn in the smartphone market?

Well, these questions highlight the fact that even if a company has a competent new product development process, this does not mean that it will be able to develop a range of market-winning innovations and sustain a high level of creativity and productivity over time. Neither does it guarantee that the company will be able to detect and react adequately to all opportunities and threats. As we stressed earlier, although the new product development process was designed to enable product developers to work across the company's functions and activities, the scope of innovation is both richer in results – for example, when it leads to the creation of new business models – and more complex because it involves a combination of "hard" and "soft" elements. Because of this complexity, because innovation affects the entire company, no process or set of processes can be sufficient to meet all the demands. However, the existence of satisfactory new product development processes makes it possible to implement a comprehensive innovation management system – steered by the C-suite – which is conducive to generating streams of market-leading innovations and avoiding competitive pitfalls.

Setting up a formal innovation management system requires proactive, personal engagement by the top team. Unfortunately, the C-suite is often simply too busy with strategic, financial, and operational issues to devote time to steering innovation on a day-to-day basis and creating that unique environment and culture. The system in place generally reflects past legacies that are seldom challenged by management. Occasionally, a new CEO or CTO will launch an “innovation revival” campaign, but it is often limited in scope and duration. Old habits tend to survive!

It is therefore healthy practice for the top management team to regularly engage in a comprehensive reassessment of the company's innovation system – how it is organized, its processes, environment, and culture – and to introduce new innovation governance guidelines. The role of the top team in this effort is critical. It goes beyond making minor structural changes and appointing new people in charge of existing departments. Governing innovation effectively involves at least six priorities:

Setting an overall frame for innovation by clarifying a vision and mission for innovation, proposing a set of values to guide innovation activities and auditing current performance.

Defining how the company will identify sources of value from innovation, how it will create value, and how it intends to capture value.

Choosing organizational models for the allocation of primary and supporting governance responsibilities for innovation, and setting up dedicated process management mechanisms.

Establishing priorities and allocating resources for innovation as part of an explicit innovation strategy and plan in support of the company's objectives.

Identifying and overcoming current obstacles in the company's organizational system and sources of resistance in order to build a lasting innovation environment.

Monitoring and evaluating results on an ongoing basis, and setting up a process to address conflicts of interest within the top management team in order to make innovation sustainable.

We shall now explore each of these six innovation governance areas in more detail.

Setting an Overall Frame for Innovation

In some companies, the innovation tradition and culture seems almost like a magic potion that is part of their DNA and ensures that all activities focus on innovation – think of Apple, Google, P&G, or 3M. But even in such companies, it is useful for top management to reflect at regular intervals on how innovation can contribute to the realization of the company's overall mission and vision. This requires a willingness to align business and innovation visions, to propose and enforce a set of values that are conducive to innovation, and to conduct comprehensive innovation audits.

At P&G, the Consumer is Boss

“Procter & Gamble is known for its highly capable and motivated workforce. But in the early 2000s, our people were not oriented to any common strategic purpose. We had a corporate mission to meaningfully improve the everyday lives of the customers we served. [...] But we hadn't explicitly or inspirationally enrolled enough of our 100,000-plus people around the world in our mission; it was neither fully embraced by employees nor fully leveraged by the company's leadership. Our innovation efforts suffered accordingly. So we expanded our mission to include the idea that ‘the consumer is boss.’”

“The people who buy and use P&G products are valued not just for their money, but as a rich source of information and direction. If we can develop better ways of learning from them – by listening to them, observing them in their daily lives, and even living with them – then our mission is more likely to succeed. ‘The consumer is boss’ became far more than a slogan to us. It was a clear, simple, and inclusive cultural priority for both our employees and our external stakeholders, such as suppliers and retail partners. We also linked the concept directly to innovation. From the ideation stage through the purchase of a product, the consumer should be ‘the heart of all we do’ at P&G.”

A.G. Lafley, former CEO

P&G convincingly illustrates the link its management sees between its overall vision of innovation and its culture. Its motto “the consumer is boss”¹ (see box) shows that visions and missions are not something ethereal. They can lead to very concrete actions in favor of innovation and shape the values and culture of the company. In framing innovation in this way, P&G's top management team, under the inspired leadership of A.G. Lafley, demonstrated an innate sense of innovation governance.

Aligning Business and Innovation Visions

Aligning visions means discussing and agreeing on what management wants to achieve business-wise and how innovation can help achieve it. This is vital to ensure that innovation is closely tied to the company's overall mission. The company's vision – how it wants to see its future – can generally be expressed in the form of three basic questions:

Who do we want to be? What kinds of activities do we want to pursue and what do we want to stand for as a company vis-à-vis our stakeholders? (This defines the company's desired identity.)

What business do we want to be in? Which segments and customers do we want to serve as a priority? (This delineates the company's desired business boundaries and focus.)

What do we want our offerings to mean to our customers? How do we intend to become the preferred supplier for our customers? (This provides a set of competitive values for the company.)

Similarly, the company's innovation vision – hence the scope of management's innovation governance mission – can be expressed in the form of the three questions on the content of innovation that we proposed in Chapter 1:

Why innovate? What concrete benefits are we trying to achieve given our current market and competitive position?

Where to innovate? In what areas should we concentrate our efforts beyond our traditional product renewal activities?

How much to innovate? How ambitious and open to risk should we be, and indeed can we afford to be, and for what objective?

These are all questions worth asking regularly, even if nothing special is happening in the company and its markets. This can be done, for example, as part of an annual top management off-site strategy retreat. Formally reviewing the mission and purpose of innovation and its desired focus may generate interesting new perspectives. But even if it only confirms current management views, it will at least ensure that all members of the C-suite are aligned behind common beliefs and a shared innovation vision and can therefore speak with one voice to the rest of the organization.

Expressing Innovation-enhancing Values

These innovation-specific management discussions may also be useful for reaffirming a set of specific values concerning innovation. It is therefore the role of the CEO, and his/her direct reports, to regularly review and specify the values they want to promote, values that can then be broadcast through management publications, speeches, and individual performance reviews. Of course, values should not be changed too often. However, they deserve to be clarified if they are too simplistic.

Google's 10 Core Values

- 1) We want to work with great people
- 2) Technology innovation is our lifeblood
- 3) Working at Google is fun
- 4) Be actively involved; you are Google
- 5) Don't take success for granted
- 6) Do the right thing; don't be evil
- 7) Earn customer and user loyalty and respect every day
- 8) Sustainable long-term growth and profitability are key to our success
- 9) Google cares about and supports the communities where we work and live
- 10) We aspire to improve and change the world.

For example, including “innovation” or “innovativeness” in the company's core values – as found frequently in annual reports and other company publications – does not really say much. Management

needs to express in a concrete and explicit fashion what this means practically in terms of personal attitudes and interactions.

Some of these values can be expressed as short, punchy sentences that can do a lot to promote the kind of culture management aspires to create. P&G's "the consumer is boss" motto, noted earlier, indeed conveys a clear and simple message about the company's main focus. The same can be said about Steve Jobs' early slogan at Apple – "Let's Be Pirates" – which called for a rebellion against the dominance of the WinTel PC.² Similarly, Andy Grove's famous book title – Only the Paranoid Survive – was effective in conveying to all at Intel the importance of humility and the conviction that no innovation battle is won forever.

Apple's Values (as viewed by Gary Hamel)

Be passionate

vs. Be rational

Lead, don't follow

vs. Be cautious

Aim to surprise

vs. Aim to satisfy

Be unreasonable

vs. Be practical

Innovate incessantly

vs. Innovate here and there

Sweat the details

vs. Get it mostly right

Think like an engineer, feel like an artist

vs. Think like an engineer, feel like an accountant.

Google provides a good example of a number of innovation-oriented values because they sustain its unique environment and culture – that is, unique in the type of people the company hires, as well as in their attitudes and ambitions. And each of these values, including its most famous one – "don't be evil" – is broken down into concrete elements.³

When a company has developed a strong innovation culture and supporting values, keen external observers of that company are generally able to highlight the main elements of the culture, even though

management may not have specifically broadcast them as such. This is the case with Apple's culture as viewed by management expert Gary Hamel, who adds: "I can't even be sure whether the values I've outlined [see box] are the ones that really drive Apple – but if they aren't, they should be! For me, the case of Apple is just a convenient and plausible vehicle for driving home a fundamental truth: You can't improve a company's performance without improving its values."⁴ This last statement is so crucial, particularly in terms of innovation performance that it should be posted in gold letters in the CEO's corner office, in the C-suite meeting room and in the boardroom.

Auditing and Improving Innovation Performance

Finally, setting the frame for innovation includes conducting a thorough innovation audit to establish the starting base before launching improvement programs. This allows management to understand how the process currently works in reality, what its deficiencies are, and what general obstacles – whether organizational or cultural – are hindering the company's innovation effectiveness.

A thorough audit generally includes some benchmarking of the company's current innovation practices against those of companies with a great innovation track record. The results of this benchmarking may be instrumental in convincing management, and the wider organization, that the company needs to change and in indicating major areas where such change is warranted. It is also a good way to silence the skeptics and proponents of the status quo. Innovation audits can be outsourced – a number of specialized consultants offer their benchmarking services. But it can also be carried out internally using an established framework,⁵ ideally focusing on the whole value creation process – business design, value identification, and value realization. Maximizing value creation is indeed one of the most important management priorities in innovation governance, as we will see below.

Many companies participate in peer-to-peer benchmarking through membership in organizations such as the Product Development and Management Association (PDMA) and the International Association for Product Development (IAPD). These organizations have helped innovators to learn from one another and in many cases have provided a venue for the adoption of new and emerging practices. They have also enabled members to create a network within which more formal benchmarking visits have taken place as workshop participants identify peers with whom they can explore specific practices.

Tetra Pak's Innovation Benchmark

"In '96, I was given the responsibility to take a hard look at how first-class companies went about the innovation process. And that took me around to companies like Dupont, 3M, Canon, Ericsson, BMW, just to mention some of them, to see what we could learn from them. I brought that back to Tetra Pak and we had several very good discussions in our group management about what we needed to change and what change in focus we would have to bring about in order to have an innovation process that works better than it does today. From that, we discussed the implications for top management in Tetra Pak.

One of them was the need to get ourselves much more involved than we had been in the past, particularly with the most important projects, which we came to call 'Pace Plus' projects."

Bo Wirsén

Tetra Pak, the world's leader of packaging systems for liquid food, provides a good illustration of the power of benchmarking. The company, whose innovation governance system will be presented in Chapter 11, was founded on a radical innovation, Tetra-Brik®, an effective carton packaging system using aseptic technology for long-life milk and juices. The company was managed for many years by its charismatic owner, Ruben Rausing, and later on successively by his two sons, Hans and Gad. Each of them promoted a creative environment, particularly in R&D. But for years Tetra Pak was unable to translate its superb R&D capabilities into successful new products because it lacked adequate processes to sense the market, select the best ideas, and manage new product development projects time- and cost-effectively.

So, in 1996 management set up a small steering group of four senior managers whose mission was to recommend steps to improve the company's innovation performance. This small group was directed by the very senior vice president in charge of European operations, Bo Wirsén. As a group, they knew, from having experienced them, many of the deficiencies of their innovation process, but they lacked references about best practices. This prompted Wirsén to visit a number of companies that had impressed him.⁶

What was unique in Tetra Pak's initial audit was that such a senior member of the top management team took the initiative to conduct these benchmarking visits himself. This gave him strong personal credibility when improvement targets were decided. It also provided him with new insights into critical innovation success factors that an outsourced benchmarking exercise would not necessarily have provided. For example, through his benchmarking visits Wirsén realized that the company might benefit from creating two new functions with a strong role to play in innovation – chief technology officer and strategic marketing officer. At Tetra Pak, this initial benchmarking exercise was used to kick off an innovation improvement program. But it was not referred to later or used as a formal auditing system.

DSM, a global life sciences and materials sciences company, whose governance system will be described in Chapter 10, provides another good example of the importance of starting an innovation improvement program with a thorough audit. When top management decided to change the company's innovation governance system in 2006 and set up a corporate innovation center, it entrusted responsibility for the center to a high-level chief innovation officer (CIO), Rob van Leen, a former group vice president for food and nutrition. Starting from scratch – the company had thus far managed innovation in traditional ways, through R&D – Van Leen felt the need to build a common language and set a base through a company-wide auditing exercise. Some of DSM's groups had a good innovation track record, others less good. All had to go through a thorough benchmarking exercise structured around a number of critical processes and capabilities, an initiative that some of them resented as being too administrative. The outcome of this exercise was a mind opener to all, as Van Leen noted (see box).⁷

DSM's Innovation Audit

“Basically, we had to start from scratch. Nobody knew how to do it. We had to find targets but we really didn't have good definitions behind it, so that is why we started definitions, reporting and so on. Then we started to put in place a diagnostic, which was built with the help of a consultant, to compare ourselves on many innovation practices with the rest of the industry. And when we saw that we actually scored below the average in the industry, we started to put in place best practices. We got them from everywhere, sometimes from within the company because there were business groups doing very well, and sometimes we brought them in from outside.”

Rob van Leen, chief innovation officer

Interestingly, this audit was turned into a real management tool. First, Van Leen distributed the results widely, including to the company's board of management, which forced some of the skeptics to take it seriously. He also decided to redo the assessment at regular intervals to measure progress. But the most powerful use of the tool was to initiate regular review meetings around this audit between the controller of the innovation center and members of the management teams of each business group. During these meetings, progress and remaining obstacles were discussed, together with some of the business group's most meaningful innovation projects. The review meetings were then documented in a detailed and widely distributed report. These practices have created a propensity for emulation among business group managers – the better performing groups want to stay on top and the poorer ones feel the need to show progress.

Defining How to Generate Value from Innovation

It is a truism that innovation is about turning market opportunities into value. In established management theories, this means identifying, evaluating, creating, and – arguably the most difficult step – capturing value.

Without a clear mandate from top management, most companies will naturally search for value within their current industries and markets. In this way, value is most usually generated by developing and introducing new products or services that replace or complement existing product lines. Some of these products or services will be incrementally better or cheaper; others will be more radically new. But their common denominator is that they remain, for the most part, within the company's existing industry value chain and keep converging toward the same competitive arena, the same “red ocean” as Kim and Mauborgne put it.⁸ This is why the potential value created by most new products is seldom fully captured.

In fact, it is not rare to hear CEOs complain that the new products or services generated by their organization are often less profitable than the original ones on which the company built its growth. These new products or services may revitalize current market segments, but they do not lead to a

sustainable competitive advantage since they are quickly imitated or superseded by competitors' entries. An important element of top management's innovation governance mission is therefore to stop this "new product merry-go-round" and initiate new ways to redefine value.

Redefining value requires broadening the scope of the search for opportunities, as we proposed in Chapter 1. This can be done by introducing a totally new basis of competition, as well as by creating new market space using previously neglected yet critical attributes – Kim and Mauborgne's "blue ocean strategy."⁹ It can also result from a systematic exploitation of opportunities to redesign the industry value chain – some authors call it the "value constellation" or "value network" – to one's advantage, or in some cases to create a totally new value chain. Such a move requires a thorough understanding of industry value chain dynamics, alternative business models and competitors' blind spots.

Charlie Fine of MIT, author of the best-seller *Clockspeed*,¹⁰ emphasizes the need to understand the dynamic relationships between suppliers, partners, and other industry value chain players to identify opportunities to take over parts of the value chain and therefore increase total profits. In his seminar *Driving Strategic Innovation*, conducted jointly with IMD, he encourages senior managers to identify strategic opportunities in their industry value chain through a systematic three-step approach:

Step 1: Assess your value chain dynamics, i.e. what factors will affect the dynamics of:

Your industry's technologies (S-curves) and its innovation pattern?

Your customers?

Your competitors?

Your industry structure?

Your governmental and regulatory agencies?

Your environment?

Step 2: Analyze your industry value chain:

What are the key elements in your industry value chain?

Who has power in the chain?

Who makes the profits in the chain?

What are the sources of power and profit in the chain (technology, brand, etc.)?

What are the key dynamic processes influencing the power structure in the chain?

Where is the locus of innovations in the chain?

What is the clock speed of each element in the chain and what are the drivers?

Step 3: Design/modify your value chain strategy

Review your insourcing/outsourcing options and decisions (make/buy choices and/or vertical integration).

Analyze your partner selection options and decisions (e.g. choice of suppliers and partners for the chain).

Evaluate your contractual relationship options and decisions (arm's length, joint venture, long-term contract, strategic alliance, equity participation, etc.).

Apple provides a striking example of this value creation strategy. Its financial success is in large part the result of having recognized – before any of its hardware competitors – the importance of content for sustainable value creation and of having cornered this value through its novel and proprietary iTunes system and its focus on smartphone applications. Apple's winning value identification strategy consisted of controlling the marketing, sales, and distribution of other companies' content by making its customers and suppliers captive, thus capturing a large part of the value of the content. This strategy is largely attributed to Steve Jobs and his top management team. They fully exercised their innovation governance role, which was to steer the company toward greener pastures – integrating hardware, software, and content – rather than leaving it to compete against the conventional pure hardware business model of its early competitors.

Choosing an Innovation Governance Model

As we stated earlier, steering, promoting, and sustaining innovation in the broadest sense of the term – not just the new product development process – is a major task that spans all company functions and organizational units. As such, it needs to be handled directly by the CEO or entrusted explicitly by the top team either to a very senior leader or to a group of managers fully empowered to exercise that responsibility. That assignment must be public, i.e. everyone in the organization should know who is in overall charge of innovation and how that overall responsibility is redistributed across the organization. Any change in the allocation of responsibilities – because changes are bound to happen over time – must also be explained and broadcast.

In our research, we have identified nine models for the primary allocation of overall responsibility for innovation. Some companies also use one or another of the same nine models to support the primary model. As we shall describe in Chapter 4, in some of these models overall responsibility for innovation is assigned to a single individual. The CEO may hold this responsibility, which is most likely to be the case if he/she is the company founder. Other individuals who may hold this role are the chief technology or research officer (CTO or CRO), a dedicated chief innovation officer (CIO) – whose actual title can be quite fancy like 'Chief Yahoo' – or a high-level innovation manager. In the financial industry the chief information officer can play this role; in other non-manufacturing sectors another CXO or a business unit manager can assume this responsibility. There are also models in which a group of leaders takes on responsibility for innovation collectively, whether they represent a subset of the top management team or constitute a high-level cross-functional steering group or a network of "champions."

There are therefore a number of models to choose from, each with its own advantages and shortcomings. It is top management's responsibility to weigh up the pros and cons of each model and how it suits the company's position and leadership resources. The choice will indeed depend on the personal preferences of the top team – do they want to remain involved personally or do they prefer to delegate responsibility to the level below? It will also reflect the type of innovation that is pursued – for example, if technology is the main driver, this would justify allocating overall responsibility to the chief technology or research officer – and of course the availability of suitable candidates for the job. Given that choice is available, top management would be well advised to refrain from sticking to the model they adopted years ago, or choosing the one most frequently found in their industry, for example the CTO model in the engineering industry.

Choosing a suitable organizational model is essential, but it is equally important to realize that conditions change. It is therefore good practice to review regularly the adequacy of the model in use given the company's changing market situation, leadership structure, and strategy.

In Chapter 4 we will explain the nature and purpose of these organizational models and in Chapters 5 and 6 we will describe these models individually and discuss how effective they seem to be, at least in the perception of companies that have adopted them.

Establishing Innovation Priorities and Allocating Resources

Steering innovation, i.e. deciding on the company's priorities concerning where, how much, and in what domain to invest in innovation, is one of the key governance missions of top management. It is generally done, at least indirectly, through project portfolio decisions. Managers understand the value of seeing the portfolio as a way of distributing resources across incremental, platform, and “radical” projects.

Going Beyond Traditional Portfolio Management Approaches

Business units typically identify their most attractive projects and management can then consolidate the various portfolios to check whether, once combined, they provide the right balance of growth, margin, and risk. Such a bottom-up approach is sometimes complemented by the addition of a few corporate projects resulting from a proactive and ambition-led, top-down innovation push. The sum of business projects included in the consolidated portfolio reflects the company's implicit innovation strategy.

The limitation of this approach is that it allocates corporate resources on the basis of the perceived attractiveness of projects as seen by business units, since business portfolios tend to weigh heavily in the company's total resource allocation. This business project attractiveness often reflects the perceived level of competitive urgency of the projects and their impact on short-term business performance in

terms of sales and profitability. In other words, unless the portfolio includes proper guidelines for investment, “game-changing” projects with a long-term impact on the company are at risk of being short-changed, thus weakening the implementation of the company's vision.

To offset that risk and ensure that the strategy will meet the company's innovation priorities, and to provide investment guidelines, it is useful for top management, as a first step, to decide on how much the company should and can afford to spend on innovation in general and on innovative projects in particular. This will determine an overall “envelope” of resources for innovation, which can then be compared with other investment funding needs. This envelope should cover not only R&D – as a total amount and as a percentage of sales – and other product development expenses (the upstream investments) but also investments in manufacturing capacity and commercialization (downstream investments).

Allocating Resources between Different Innovation Thrusts

This broad “innovation envelope” should then be allocated among the different types of innovations being pursued, and this is the second step in the resource allocation process. To do so, it is necessary to characterize the main innovation thrusts being pursued. The book *Innovation Leaders*¹¹ proposed to do so by combining broad options derived from the questions listed in Chapter 1:

Why innovate? (Innovation objective)

Innovations can be pursued for two broad objectives: to energize and expand a current business in its existing markets or to create a totally new business. These objectives can be combined.

Where to innovate? (Innovation scope or focus)

Innovations can focus on products or services – introducing a new “black box” or stand-alone service – or, alternatively, on developing a new business model or business system.

How much to innovate? (Innovation intensity level)

Innovations can be incremental in the changes brought to current products, services, or processes, or they can be more radical, leading to completely new product and service concepts.

With whom to innovate? (Innovation boundaries)

Innovations can be developed and implemented internally, using the company's capabilities and resources, or externally through deliberate collaboration with partners.

Note that both innovation intensity and boundaries – if restricted to an either/or option (incremental or radical; internal or external) – are always subject to debate. An innovation that is radical in one company may be characterized as incremental by a competitor. The level of innovation is relative to the reference models of the beholder. Also, innovations are rarely conducted only internally – external factors like

suppliers are usually involved – which means most innovation projects fall somewhere between the two extremes.

These four dimensions can be combined, as shown in Figure 3.1, into four entirely different innovation thrusts. Top management should recognize them explicitly, choose the ones that it will pursue as a priority, and use them to characterize and communicate its innovation strategy and investments, which will usually come from a combination of the chosen thrusts.

Figure 3.1: Typology of Innovation by Strategic Focus

These four thrusts propose a simple typology of innovation choices:

The internal development and launch of a new and/or improved product, process, or service offering, typically to grow and reinforce the current business in an incremental innovation mode.

The internal development of a totally new product category or service offering, typically to grow and create a totally new business, next to the existing ones, in a radical innovation mode.

The development and launch, together with selected partners, of a totally new business model or integrated system, typically to grow and create a new business in a radical innovation mode.

The development and launch, together with partners or complementors, of a new and/or improved customer solution or customer system, typically to grow and reinforce the current business in an incremental innovation mode.

This classification reflects the fact that, from a management point of view, developing a “black-box” product or service is very different – and carries a different type of risk – than introducing a new business model or business system, or even a complex product solution. Indeed, whereas the development of a new product or service is often the result of an internal process, even though it may involve the use of outsourced technology and suppliers, the development of a radically new business model or system often requires the cooperation of several external partners, outsourcing suppliers, or complementors.

As mentioned earlier, these four thrusts are not mutually exclusive – innovations can be pursued simultaneously across several of these areas. Once displayed on a two-by-two matrix (refer to Figure 3.1), they provide a useful framework and lens for examining the complex reality of innovation thrusts.

Ideally, once the overall innovation envelope has been established, management should propose how the envelope should be split between the four quadrants:

How much the company should spend internally on incremental projects to reinforce the current businesses.

How much it should invest, again internally, in radical projects to create a totally new business next to the existing ones.

How much it should commit to attempts to introduce a radical new system or business model with partners.

How much it should devote to the creation of incrementally innovative customer solutions, once again with partners.

Management can now go back to the original project portfolio and position specific innovation projects in the four quadrants to see whether the investments that they represent add up to the predetermined envelopes (refer to Figure 3.2). If they do not – if some quadrants lack projects – then management can indicate where additional efforts are expected (see arrows on Figure 3.2) and how much they represent in terms of new resources to be committed, thus starting a search for new opportunities.

Figure 3.2: Portfolio of Innovation Projects by Type

This approach allows management to introduce an innovation dimension in the traditional project portfolio approach by: (1) indicating how the various objectives of the company's innovation strategy will be funded; (2) specifying how much management plans to spend on innovation in general, and on activities that will reinforce current businesses vs. innovative efforts to create entirely new activities; (3) providing guidelines on how much should be spent on each main type of innovation; and (4) suggesting new market domains that need to be explored as a priority for these new activities.

Three general remarks can be made on this definition of priorities and management allocation of resources.

First, in their effort to reinforce their current market position, most business management teams tend to focus on only a few areas where innovation can make a difference, i.e. new better and cheaper products, new technologies, and new production processes. It is therefore useful for the C-suite to stress the importance of other reinforcing innovations, for example in new business models, in the supply chain and/or value chain, in service, in marketing and channel distribution, and the like. These could stimulate business managers to look more broadly at innovation, as recommended in Chapter 1.

Second, deciding how much to invest by type of innovation, i.e. incremental vs. radical, determines how much risk the company is willing to take (or avoid). By addressing this issue directly – for example, by setting up specific envelopes by quadrant – management can establish a general company policy that may be helpful when the company is leaning too far to one side or the other. For example, some managers always seem to look for breakthroughs. They behave as if incremental innovations such as product derivatives are not worth their efforts, with the result that they miss major market and profit opportunities. Other managers, in contrast, stay permanently within their comfort zone and shy away from risky developments. In each case, it will help if management specifies, for each business, what it considers as the right balance between incremental and radical innovation.

Third, and finally, defining a policy on open innovation is an important element of an innovation strategy, particularly in the new social network environment and with the growing importance of crowdsourcing. It goes beyond a simple exhortation to build upon external ideas and competencies. A policy on open innovation ought to specify:

the domains where external cooperation is desirable;

the boundaries of cooperative deals and the types of partners to be considered off-limits;

considerations on the protection of intellectual property; and

indicators to measure the level of achievement of the policy.

By specifying this type of broad resource allocation – covering not only R&D but also other upstream and downstream investments – management can achieve three important benefits of good governance:

Send a clear message regarding the company's priorities.

Set the frame for the development of its new business activities.

Ensure that these activities will be adequately funded.

Overcoming Obstacles and Building a Favorable Innovation Environment

As Gary Hamel suggested, there is generally a strong correlation between innovation culture and innovation performance. The success of Google, for example, cannot be separated from the emphasis the company puts on its “can-do” entrepreneurial culture or from the concrete steps management takes to sustain it. Google's famous rule – modeled on 3M's “15% rule” – that allows people to pursue their own ideas for up to 20% of their time, is just one example of the company's innovation-enhancing environment. By contrast, some excellent companies with huge technological resources never seem to reach the status of top performers in their industry, largely because of an internal culture that stifles innovation.

Innovation calls for openness, experimentation and risk taking, and, above all, cooperation and constructive challenges across functions and organizational units, and all of these aspects need to be explicitly encouraged by management. In a seminal article, two Harvard Business School professors¹² summarized the lessons learned from a two-day colloquium held at their school on “Creativity, Entrepreneurship, and Organizations of the Future.” The colloquium gathered over a hundred people who were “deeply concerned with the workings of creativity in organizations,” including research scholars and business leaders from companies whose success depends on creativity – such as design consultancy IDEO, technology innovator E-Ink, internet giant Google, software specialist Intuit, and pharmaceutical leader Novartis among others. Even though the colloquium focused on creativity, it provided a number of lessons that apply more generally to innovation. The lessons can be summarized in a number of concrete exhortations to senior management to:

Draw on the right minds:

Tap ideas from all ranks

Encourage and enable collaboration

Open the organization to diverse perspectives.

Bring process to bear carefully:

Map the phases of creative work

Manage the commercialization hand-off

Provide paths through the bureaucracy

Create a filtering mechanism.

Fan the flames of motivation:

Provide intellectual challenge

Allow people to pursue their passions

Be an appreciative audience

Embrace the certainty of failure

Provide the setting for “good work.”

But creating this type of open and creative environment may not be sufficient. Management must also address several organizational and cultural obstacles that hinder innovation. We have observed them – the seven vicious innovation killers – in a wide range of companies and propose a number of antidotes to each below.

Killer # 1: Excessive Operational Pressure

The first innovation killer, present in most companies, is the excessive pressure put on managers as a result of their operational and organizational responsibilities and of a constant fire-fighting atmosphere within the business. These pressures tend to be reinforced by a management performance evaluation system that encourages short-term results. Managers may be willing to spend time on innovation, but there may simply not be enough time for these innovative undertakings in many organizations.

Management can counteract this pressure in two ways, which have proved to be effective antidotes.

First, the pressure can be alleviated if management identifies, appoints, and guides dedicated and passionate innovation coaches to motivate, challenge, and support local innovation teams. These champions are generally found among younger high-potential managers. To be effective, these coaches or champions should be highly energetic as well as socially skilled so that they are not viewed as interfering in the business or “bossing” the local managers. They should also be practical and resourceful in identifying bottlenecks, suggesting solutions, proposing best practices and tools, and generally helping business managers move forward with their innovation agenda and projects.

Second, management can create a counter-pressure in favor of innovation, for example by introducing specific innovation performance measures in every manager's balanced scorecard. This assumes, of course, that management is true to the very principle of balanced scorecards – in other words that it judges managers on all dimensions of the scorecard and not just on financial or budget performance. Once managers are penalized in their personal performance review for letting their innovation activities lag behind, even if they make their budget, it is probable that they will revive their interest in these undertakings and find a better balance of their time and efforts.

Killer # 2: Fear of Experimentation and Taking Risks

This second innovation killer usually results from unrealistic financial benchmarks or from a culture that does not tolerate failure. Financial benchmarks – for example, assigning unreasonably high hurdle rates of return on totally new and innovative projects – are an innovation killer because they may discourage people from undertaking uncertain projects. Note that full-blooded entre-preneurs will often pay lip service to these financial goals – be it in terms of net present value created or payback time – and they will provide whatever numbers management expects to see, knowing full well that such kinds of number games are irrelevant at the early stage of risky innovation projects. But circumventing the existing benchmarks can only work sometimes. After a failure or two the real true-blue entrepreneurs will soon find it more attractive to find a company that really values innovation.

Risk averse innovation cultures exist in many companies, particularly those with a strong focus on operational excellence and performance predictability. Even though management may encourage risk taking in their speeches, managers are quick to sense what the top team really means. Most companies carry a whole cemetery of failed projects and ventures, and managers are quick to find out what fate befell their promoters. This often kills early desires to take risks.

There are two powerful antidotes to the fear, which can convince managers that top management values and actively seeks risk taking.

Sponsoring Post-it Notes©

The most cited innovation of all time was developed under the auspices of 3M vice president Geoffrey Nicholson. Even though it was in most ways a “skunk works” project, a senior manager took the time and the risk of working with the team and providing “coverage.” If it had failed it would have been on his watch. It succeeded beyond anyone's wildest dreams, but Nicholson's role is rarely mentioned.

First, management can set the example at the top by asking senior leaders to personally coach high-risk/high-reward projects. Often, this means making themselves regularly available to the project team – for example, after normal hours – for informal reviews and problem solving.

When this is done – and this is the first benefit – the whole organization quickly learns that (1) risky projects are perceived as acceptable and management is ready to back them, even if they end up failing, and (2) a high mortality rate for such projects is considered normal and nobody should be penalized for trying and failing.

A second benefit of this approach is that within the top management team innovation discussions become more concrete as policy decisions can be tested on real projects.

Bringing Silicon Valley Inside

“Venture capitalists are risk takers, but they are not big risk takers. [...] Out of 5000 ideas, a five-partner VC firm may invest in ten which it views as a portfolio of options. Out of that ten, five will be total write-offs, three will be modest successes, one will double the initial investment, and one will return the investment 50- to 100-fold. The goal is to make sure you have a big winner, not to make sure there are no losers.”

Gary Hamel

If every member of the top team personally sponsors a project, the third benefit is that the decision to pull the plug on a given project is taken collectively without undue pressure on the sponsoring leader who does not risk losing face.

The second antidote is for management to adopt the philosophy of venture capitalists (VCs) regarding investments, resource allocation, coaching and return expectations, as recommended by Gary Hamel in his famous article on Silicon Valley (see box).¹³ Hamel advocates creating an internal market for ideas, talent, and capital, and making projects compete for resources. Knowing that only a small proportion of projects will be successful, VCs look for the upside and not the downside of projects, and they will do their utmost to support the projects that have the highest chances of winning. In cases of failure, they move fast to start new ones.

Killer # 3: Insufficient Customer and User Orientation

Relying on superficial market knowledge or outdated knowledge or, worse, believing that the company “knows better” than customers – the typical arrogance of established market leaders – are also innovation dampers. Note that the reverse – taking customers' expressed wishes at face value – can also be misleading, since customers are often unable to talk about their latent or future needs – the needs that, if well addressed, can build competition-crushing, or disruptive, innovation. Who would ever have thought that we “needed” and would pay for a telephone that holds an entire address book and calendar! Insufficient customer and user orientation can also lead companies to neglect to define and target specific customer groups. Companies that launch a new product concept without clearly identifying a specific target group beforehand – at least initially – and without understanding how that customer group will benefit from it are likely to waste resources. The approach of “raising-the-flag-and-seeing-if-anyone-salutes” is a costly way to bring innovations to market. It is only valid if the company is very agile, learns fast from the initial launch, and quickly reorients and relaunches to target a specific customer group.

Market Immersion at Hilti

Hilti, one of the world's global leaders of professional hand tools for the construction industry, sells directly and through retailers to thousands of building contractors around the world. The company has built its innovations and market success on an in-depth understanding of its customers and users. Visiting customers, not to sell to them but to observe them, is a habit that came from the very top. Hilti's former CEO, Dr Pius Baschera, was known for regularly accompanying salesmen on their customer visits ... without telling them in advance, thus avoiding being presented only to friendly customers. In so doing, he passed a strong message to his organization, i.e. “To succeed, know thy customers deeply!”

Once again, there are at least two types of antidotes to this lack of customer intimacy.

First, management can overcome this deficiency, not by multiplying ad nauseam the amount of traditional market research done by the company, but by making staff temporarily share the life of various customers to understand their total experience.¹⁴ The point is not so much to search for what customers say they want – they may often trail behind the times – but to become immersed in their environment in order to understand what they do, how they feel, what frustrates and delights them,

thus being able to anticipate what they might need and want in the future. Customer-oriented companies encourage a significant proportion of their staff – and not just marketing specialists – to conduct such customer immersions at regular intervals.

Second, management can also encourage staff to engage selected customers to join them in their idea searches and innovation projects. Some industries, like aerospace, do this as a matter of routine. No new aircraft could be developed and commercialized without the active involvement of lead customers, for example airlines, from the very early stages onward. But this habit, which some companies refer to as organizing “customer clinics,” is not always encouraged, for fear of losing confidentiality on new products or for lack of trust in the wisdom of customers, or often because selecting customers for such tasks is not easy. Whatever their actual contribution to the company's projects, this habit of involving customers in idea searches and projects creates strong customer intimacy.

Killer # 4: Uncertainty on Innovation Priorities

Not knowing what management expects from innovation is often perceived by many in the organization as a major innovation obstacle, particularly if it is combined with a risk averse culture. It leads to ad hoc idea generation (where should we search for ideas?); difficult concept evaluations (against what objectives should we evaluate our ideas?); fuzzy screening and selection (on what basis should we favor one project over another?); and poor project justifications.

As we recommended earlier, this uncertainty can be overcome by clarifying the company's innovation strategy, which means defining and broadcasting why, where, how, and with whom to innovate.

Management can also beef up the project briefing process by requesting that projects be linked explicitly to the company's announced innovation objectives and strategy.

Killer # 5: Lack of Management Patience Regarding Results

Leaders who press their teams unduly for faster results – not so much for shorter lead times, which is understandable, but for quick returns on investment – can be strong innovation inhibitors. Indeed, if short payback is introduced as an important criterion for the selection of innovative project ideas, then staff will, of necessity, screen out all ideas with a long-term high-risk/high-impact profile, to focus exclusively on predictable, incremental innovations. The same leaders may also be tempted to pull the plug too soon on very attractive projects with a long incubation phase and payback outlook.

Nestlé's Patience with Nespresso

With \$3.8 billion in highly profitable sales in 2012, Nespresso is a jewel in Nestlé's product and brand portfolio. Yet few outsiders know that it took 16 years for the project to reach breakeven, as all kinds of market applications and channels were tried one after the other before achieving success. How did Nespresso manage to escape the hatchet of Nestlé's corporate financial controllers for so long? The answer is that there was strong advocacy for the project from several members of Nestlé's top management team, including Camillo Pagano, former executive vice president in charge of strategic business divisions and marketing worldwide.

The top management team can fight this temptation in two ways.

First, it can earmark specific resources for long-term projects, alongside the company's "normal" R&D budget, and personally become involved in selecting these high-impact projects.

Second, it can ensure that hurdle rates of return, whatever the type, are not introduced as criteria in the initial screening process for innovative project ideas. Financial payback considerations should appear only at a much later stage, when big capital investments are being considered. Instead of payback criteria, management should emphasize the project's potential to create value, i.e. the superiority of the future product or service provided to the company's customers, as perceived by the market, and ideally the price these customers will be ready to pay for the product or service, which will make the project attractive. This potential to create a quantum leap in value should obviously be validated through customer contacts and early feedback. If positive, this should convince management to be patient!

Killer # 6: Functional and Regional Silos

Large, complex organizations are often characterized by the coexistence of communities of specialists, each with its own identity, values, and professional norms. These communities exist at headquarters at the functional level. They are also present in decentralized operations, manufacturing plants, or regional and national commercial organizations. Unless strongly unified under the same corporate culture banner, these various groups tend to develop an "us vs. them" mentality, which can be detrimental to a cross-functional and cross-disciplinary process like innovation.

There are multiple dangers:

Organizational isolationism, which slows the process down by making functional project handovers complex since each function wants to keep full control over its own field of expertise.

The inability to build on one another's ideas because of the lack of opportunities to work together on the project from the start. This often happens when regional organizations feel left out of the initial project specifications, which are decided upon at headquarters level "for the world."

Domineering attitudes of some departments, which claim to have the final say in all project matters. This can be the case with marketing dominating R&D in fast-moving consumer goods, or with engineering over-powering everyone in technology-intensive companies.

Fights over ideas and budgets may also prevent people from cooperating across organizational boundaries. This can happen when a project team requires additional resources from functional departments.

Working across Disciplines at Eli Lilly

Eli Lilly was developing a drug intended to prevent breast cancer in women. Trials showed that, although it worked, it did not work well enough to be worth marketing. Under usual circumstances, the product would have been shelved, along with several years of the complex activities that go into drug development. Instead, several people who had worked on the project began to realize, after attending a PDMA Frontier Dialogue, that the essential hormone makeup of the drug was able to build bones. They crossed therapeutic areas, a move almost unheard of at the time, and several years later the successful osteoporosis medicine Evista hit the market.

The best and most classic antidote to this danger is the systematic adoption of A to Z cross-functional and/or cross-regional innovation project teams – from idea and concept to market launch – combined with a high degree of empowerment of the project leaders in relation to the functional organization. By working together on the same intense projects, people start building bridges across functions and geographical areas, and they are more likely to adopt a “we” attitude as opposed to an “us vs. them” mindset. In addition, because they share the same performance measures – it is the team that succeeds or fails, not individuals or functions – this helps create a strong sense of solidarity within the team.

Another way to fight silos and develop a “one company spirit” is to multiply opportunities for joint innovation training programs. When people from different functions and regions spend time together discussing current management issues, visions and perceptions do change and collaboration becomes easier.

Killer # 7: Rigid and Over-regimented Environment

Last but not least among the most widespread innovation killers is an overregulated environment. This situation exists in many large and traditional companies. Company policy, management rules, and standard operating procedures are definitely necessary to run operations but, by limiting the freedom of would-be entrepreneurs and slowing down teams with unnecessary paperwork and controls, they can discourage people and ultimately stifle innovation.

Management can overcome this risk in two steps.

First, it can make an explicit effort to review all the management rules that were designed primarily for conducting normal operations and generally controlling non-project expenses, and free the project teams from most of these rules. Among the rules to be eliminated are all those that (1) impose standard work processes; (2) limit or organize horizontal and vertical communications; (3) restrict the project team's free access to customers, suppliers, and partners; and (4) require considerable justifications for an authorization to travel or to spend small amounts of money, for example for information or tools.

But because a certain number of rules are necessary, management should ask the project team to define the process they intend to follow and the specific rules that they are willing to accept and apply in their work. Management could then check on how well team members abide by the process and rules that they themselves have chosen.

Monitoring and Evaluating Results

Finally, management needs to set up and monitor a range of performance indicators to track progress and identify new improvement targets as some of the initial goals are reached. At the very least, indicators ought to cover both input factors – how many resources the company pumps into innovation – and output measures – how much the company is getting out of its innovation investments. But advanced innovators will typically go beyond these two broad categories and introduce a pyramid of metrics with four types of carefully selected indicators:

Lagging indicators measure process results, typically on the basis of market or financial performance. The percentage of sales that comes from products introduced in the past few years, depending on the industry life cycle, is a typical lagging indicator. So is “time to profit,” which measures the time it takes for cumulated profits to pass cumulated investments.

Leading indicators measure process input quality and/or quantity or factors conditioning innovation. The number of patents issued and granted is an example of a leading indicator – and not an overall innovation performance indicator as some companies believe! Another example is the percentage of R&D spent on long-term, high-risk/high-impact projects.

In-process indicators measure process quality in terms of deliverables and time or cost compliance. Classic indicators in this category include the number of non-value-adding changes in projects past a certain point, or the percentage of project review gates passed according to schedule.

Learning indicators which measure the improvement rate on critical performance targets for the business. Examples include the product stabilization period (from launch until quality and performance meet expectations), or more generally the “half-life” of a specific improvement (the time it takes to improve a given performance by 50%).

Innovation Scorecard at Solvay

“The Innovation Scorecard seeks to provide an overall picture of the Group's performance in terms of innovation. It takes account of the extent of employee participation in innovation projects and the

proportion of projects carried out in collaboration with external partners, which is a key component of Solvay's sustainability strategy.”

From Solvay's website

The range of innovation performance indicators companies use varies from very few (typically reflecting sales and profit growth from new products) to too many! It is indeed difficult to find the right balance and mix.

Solvay, the global chemicals and polymer group, provides a good illustration of a balanced innovation scorecard. It is impressive because it focuses on a manageable number of ratios, eight in total, but in the main categories of performance which reflect the company's innovation priorities: results, the growth pipeline, partnerships, ideas generated, people involvement, and R&D (refer to Figure 3.3 shown in Solvay's 2012 annual report).

Figure 3.3: Solvay's Innovation Scorecard

Reproduced with permission of Solvay SA. All rights reserved.

The merit of Solvay's approach is that it has set specific corporate targets in three key areas that have been selected as main innovation challenges within the Group, as indicated on its website:

Growth objective: 30% of Group income should come from new products or technologies.

Partnerships objective: 50% of projects should be developed in partnership with external partners (customers, universities, public authorities, start-ups, etc.) in the framework of structured agreements.

People objective: 100% of executives should define their personal innovation objective every year and have the occasion to evaluate it at least once with their managers. All employees should produce at least one innovative idea every year.¹⁵

In Conclusion: A Call for Action

The six innovation governance areas described in this chapter highlight a number of responsibilities that will typically not be carried out by the second or third line of a company's hierarchy. These employees can be expected to manage processes and projects within a set of overall guidelines, not to come up with an overall framework for innovation.

The six domains essential for organizing and mobilizing for innovation are:

setting an overall frame for innovation;
defining value;
choosing an innovation governance model;
establishing innovation priorities and allocating resources;
overcoming obstacles and building an innovation culture; and
monitoring and evaluating results.

They condition the way innovation is carried out and sustained by the organization. They therefore belong to the prime innovation governance duties of the top management team. It is vital for the C-suite to address them collectively, broadcast their outcomes and include them as a regular topic on the top management agenda.

We conclude with one caveat: the mission of innovation leaders is to steer and support innovators. Governing innovation means making sure that innovators have as smooth a path as possible, that their commitment and hard work pay off as much and as often as possible. We have seen many cases where people work hard on projects that should have been successful, only to see their work side-lined, defeated, or disrupted by the kinds of “killers” outlined in this chapter. This problem is often caused by the leaders who should be in charge of smoothing the path to success. They fail to follow the kinds of practices we have discussed above. Problems lie in the way the system is designed and the way the work is organized. Now that companies have discovered increasingly better ways of designing and organizing the work of innovation, it is time for top management to take full responsibility for making sure that the design and organization are optimized so that the innovators have a chance to produce the value they are capable of delivering.

Notes

1 Lafley, A.G. and Charan, R. (introduction) (2008). P&G's Innovation Culture: How We Built a World-class Organic Growth Engine by Investing in People. Strategy+Business, August 26, <http://www.strategy-business.com/article/08304?pg=all>

2 Refer to Young, J.S. and Simon, W.L. (2005). iCon Steve Jobs, The Greatest Second Act in the History of Business. Wiley, p. 88.

3 For the full list of these values, refer to <http://www.askstudent.com/google/list-of-google-core-values/>

4 Hamel, G. (2010). Deconstructing Apple – Part 2. Wall Street Journal, March 8.

5 Refer to the framework proposed by Beebe Nelson and Valerie Kijewski. <http://www.nkaudits.com/>

6 Excerpt from a 2005 videotaped interview “Steering and Coaching an Innovation Project” with Nils Björkman and Bo Wirsén, members of the Group Leadership Team of Tetra Pak International by Prof. J.P. Deschamps, IMD, Lausanne, Switzerland.

7 Excerpt from a 2009 videotaped interview “DSM: Mobilizing the Organization to Grow through Innovation” with Rob van Leen, chief innovation officer at DSM, by Prof. J.P. Deschamps, IMD, Lausanne, Switzerland.

8 Chan Kim, W. and Mauborgne, R. (2005). Blue Ocean Strategy – How to Create Uncontested Market Space and Make Competition Irrelevant. Harvard Business School Press.

9 Chan Kim, W. and Mauborgne, R. (1997). Value Innovation: The Strategic Logic of High Growth. Harvard Business Review, January–February 1997.

10 Fine, C.H. (1998). Clockspeed: Winning Industry Control in the Age of Temporary Advantage. Perseus Books. Used with permission.

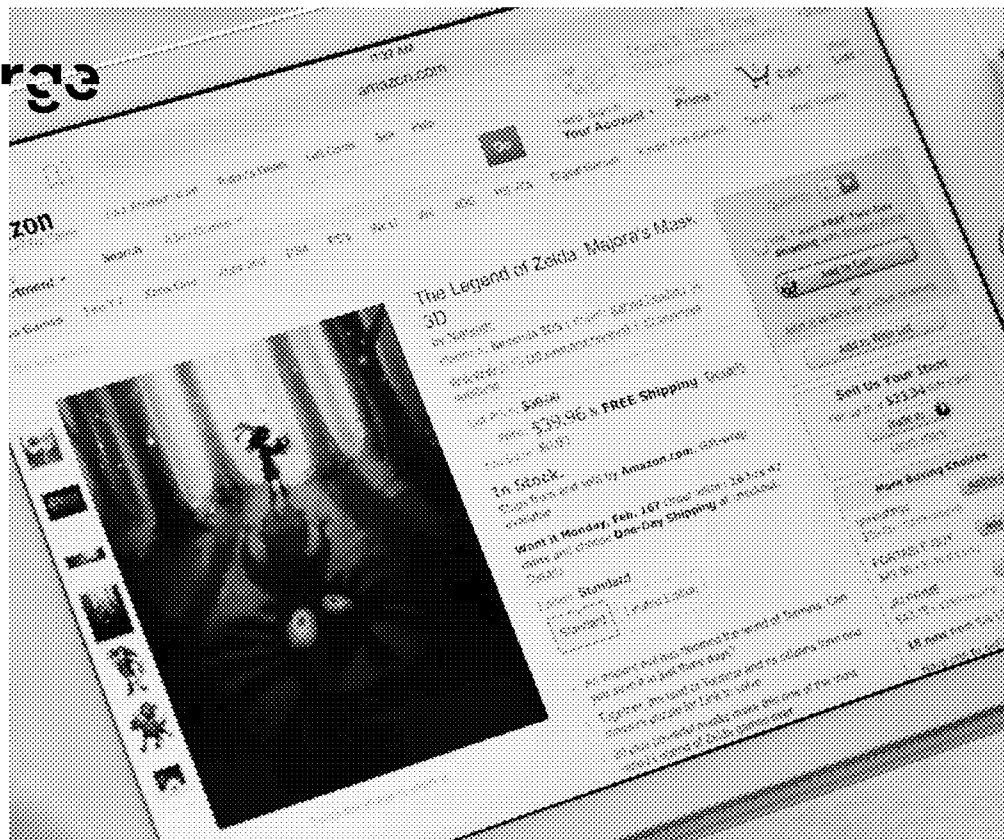
11 Deschamps, J.-P. (2008). Innovation Leaders: How Senior Executives Stimulate, Steer and Sustain Innovation. San Francisco: Wiley/Jossey-Bass, Chapter 6.

12 Amabile, T.M. and Khairé, M. (2008). Creativity and the Role of the Leader. Harvard Business Review, October.

13 Hamel, G. (1999). Bringing Silicon Valley Inside. Harvard Business Review, September/October.

14 Gouillart, F.J. and Sturdivant, F.D. (2009). Spend a Day in the Life of Your Customers. Harvard Business Review, March.

15 Innovation Scorecard, copyright Solvay 2013, source:
<http://www.solvay.com/EN/Innovation/innovationscorecard.aspx>



REPORT

EXCLUSIVE: AMAZON MAKES EVEN TEMPORARY WAREHOUSE WORKERS SIGN 18-MONTH NON-COMPETES

Contract says it can limit jobs across the globe

By Spencer Woodman | @spencerwoodman | Mar 26, 2015, 11:44am EDT | 0 Comments / 0 New

Amazon is the country's largest and most sophisticated online retailer, but it still runs largely on manual labor. Scattered around the country are massive warehouses staffed by workers who spend their days picking objects off shelves and putting them in boxes. During the holiday season, the company calls on a huge reserve army of temporary laborers.

The work is repetitive and physically demanding and can pay several dollars above minimum wage, yet Amazon is requiring these workers — even seasonal ones — to sign strict and far-reaching noncompete agreements. The Amazon contract, obtained by *The Verge*, requires employees to promise that they will not work at any company where they

"directly or indirectly" support any good or service that competes with those they helped support at Amazon, for a year and a half *after* their brief stints at Amazon end. Of course, the company's warehouses are the beating heart of Amazon's online shopping empire, the extraordinary breadth of which has earned it the title of "the Everything Store," so Amazon appears to be requiring temp workers to forswear a sizable portion of the global economy in exchange for a several-months-long hourly warehouse gig.

The company has even required its permanent warehouse workers who get laid off to reaffirm their non-compete contracts as a condition of receiving severance pay.

"IT IS QUITE BROAD IN ITS SCOPE."

When Amazon shut down a massive warehouse in Coffeyville, Kansas, earlier this year, hundreds of employees lost work. One laid-off warehouse worker, who earned just over \$12 an hour unloading inbound freight at the Coffeyville facility, showed *The Verge* a clause in her severance agreement that admonished her to "fully comply" with the noncompetition agreement. This worker wished to remain anonymous because of a non-disclosure agreement she signed with Amazon.

"It is quite broad in its scope," says Orly Lobel a professor of labor and employment law at University of San Diego, who has studied noncompetes extensively and reviewed the Amazon agreement.

"During employment and for 18 months after the Separation Date, Employee will not, directly or indirectly, whether on Employee's own behalf or on behalf of any other entity (for example, as an employee, agent, partner, or consultant), engage in or support the development, manufacture, marketing, or sale of any product or service that competes or is intended to compete with any product or service sold, offered, or otherwise provided by Amazon (or intended to be sold, offered, or otherwise provided by

the methods of fulfilling these needs: smaller stores, addresses, and pricing; and Amazon selling products, marketing strategies, research projects or developments, products, legal affairs, and future plans relating to any aspect of Amazon's present or anticipated business. "Confidential Information" does not include the terms and conditions of Employee's own employment.

3.2 Prevention of Unauthorized Release of Confidential Information. Employee will take reasonable measures to prevent unauthorized persons or entities from obtaining, receiving, or getting access to any Confidential Information in Employee's possession or control.

3.3 Confidential Information of Third Parties. Employee will preserve as confidential any information that Employee learns or obtains from a third party or conveys to a third party (such as a client, customer, affiliate, partner, or vendor) that is not readily available to the public or that Amazon is obligated to treat as confidential, and Employee will treat such information as Confidential Information.

3.4 Return of Confidential Documents. On the Separation Date, or at any time thereafter requested by Amazon, Employee will immediately return all Confidential Information and other things belonging to Amazon, including tools, equipment, devices, or other property, and all documents, records, notes, e-mails, and tangible notes containing or embodying any Confidential Information, including any copies (whether stored in paper, electronic, magnetic, or other form) used in Employee's possession or control, whether prepared by Employee or others.

4. RESTRICTIVE COVENANTS

4.1 Non-Competition. During employment and for 12 months after the Separation Date, Employee will not, directly or indirectly, whether on Employee's own behalf or on behalf of any other entity (for example, as an employee, agent, partner, or consultant), engage in or support the development, modification, marketing, or sale of any product or service that competes or is intended to compete with the product or service sold, offered, or otherwise provided by Amazon or intended to be sold, offered, or otherwise provided by Amazon at the location that Employee worked at or supported or about which Employee obtained or received Confidential Information.

4.2 Non-Solicitation. During employment and for 12 months after the Separation Date, Employee will not, directly or indirectly, whether on Employee's own behalf or on behalf of any other entity (for example, as an employee, agent, partner, or consultant), by action or inaction, recruit any Customer of any product or service that Employee worked at or supported or about which Employee obtained or received Confidential Information, or for any other entity, or encourage any Customer to cease doing business with Amazon or to terminate or limit an existing relationship or arrangement with Amazon. For purposes of this Agreement, "Customer" means any individual or entity that was a customer or client of Amazon during Employee's employment, or with which Amazon engaged in discussions before the Separation Date regarding the possibility that such party might become a customer or client of Amazon, and "Business Partner" means any individual or entity with which, before the Separation Date, Amazon was involved in any business arrangement or engaged in discussions regarding the possibility of entering into such an arrangement.

4.3 Non-Interference. During employment and for 12 months after the Separation Date, Employee will not, directly or indirectly, whether on Employee's own behalf or on behalf of any other entity (for example, as an employee, agent, partner, or consultant), (a) solicit or otherwise encourage any employee, contractor, or consultant of Amazon ("Amazon Personnel") to terminate the employment or contractual relationship with Amazon; (b) disclose information to any other individual or entity about Amazon Personnel that could be used to solicit or otherwise encourage Amazon Personnel to form new business relationships with that or another individual or entity; or (c) otherwise interfere with the performance or employment of Amazon Personnel. Amazon reserves all rights and responsibilities as Amazon, including in this Section 4.3, restricts Employee from exercising rights provided under the National Labor Relations Act.

4.4 Reasonableness of Restrictions. Employee recognizes that the restrictions in this Section 4 may significantly limit Employee's future flexibility in many ways. For example, the restriction in Section 4.1

Amazon in the future) that Employee worked on or supported, or about which Employee obtained or received Confidential Information." Noncompete agreements have traditionally been associated with highly skilled, white collar jobs where, in exchange for signing a restrictive contract, employees might gain specialized training and learn trade secrets that enable professional advancement. More recently, such contracts have been seeping into low-skilled and low-wage occupations that

require little on-the-job training.

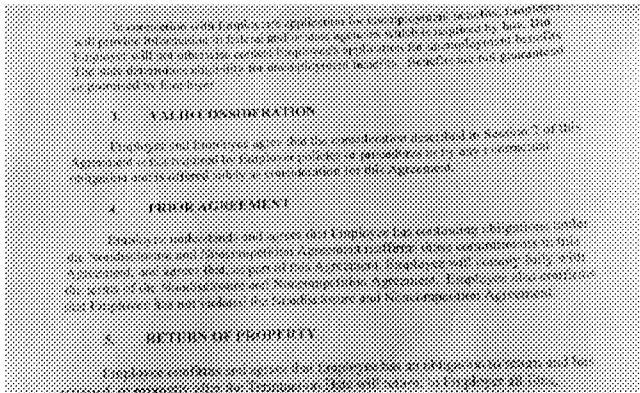
This trend is likely occurring at least in part because employers know they can get away with it in today's economy, where jobs are scarce, says Charlotte Garden, a law professor at Seattle University School of Law. "When you have a more vulnerable workforce applying for jobs," Garden says, "they're not going to attempt to negotiate the terms of the contract they're handed."

The expansion of noncompetes into low-wage work came to national attention last year, when the *Huffington Post* reported that Jimmy John's had some of its permanent workers sign noncompete agreements that covered sandwich sellers within three miles of Jimmy John's locations. US Congress members called for a federal investigation into the sandwich chain's use of the agreements. The Amazon contract appears more extreme: it is not only being pushed on temporary workers, who will have their opportunities inevitably constrained upon their planned dismissal, but it is also explicit in its potentially limitless geographic reach.

"Employee recognizes that the restrictions in this section 4 may significantly limit Employee's future flexibility in many ways," the agreement asserts, referencing the section containing the noncompete agreement and three other clauses. "Employee

further recognizes that the geographic areas for many of Amazon's products and services — and, by extension, the geographic areas applicable to certain restrictions in this Section 4 — are extremely broad and in many cases worldwide."

The contract — which was obtained through applying and being accepted to a seasonal Amazon warehouse position — even includes a provision that requires employees who sign it to "disclose and provide a true and correct copy of this Agreement to any prospective new employer [...] BEFORE accepting employment[...]"



Laid-off employees were asked to reaffirm the noncompete contract as a condition of receiving severance: "Employee understands and agrees that Employee has continuing obligations under the Nondisclosure and Noncompetition Agreement reaffirms

those commitments in this Agreement, and agrees that, as part of this Agreement, Employee will comply fully with the terms of the Nondisclosure and Noncompetition Agreement."

It's unclear whether Amazon has attempted to enforce its noncompete contracts with hourly warehouse workers, and Amazon did not respond when asked about this by *The Verge*. But the company does have a history of aggressively pursuing such cases against white collar workers. Last year, after a former Amazon marketing manager took a job at Google, Amazon leveled a suit against him that was said to test the limits of noncompete law. The willingness of courts to validate such agreements can vary dramatically across states. But regardless of whether courts are willing to enforce them, noncompetes can still affect workers' behavior.

Regina Lee, a seasonal Amazon worker who signed a noncompete, takes the agreement seriously. In recent years, as Lee has struggled to find decent-paying work, she has become a both loyal and grateful member of Amazon's army of seasonal warehouse workers. "Before I worked at Amazon, I applied to Walmart, and I didn't get anywhere, so I'm just happy to have a job," Lee says. "Especially a job I can go back to every winter." Lee says that, during last year's pre-Christmas rush, Amazon's human resources team in

Coffeyville, Kansas, helped accommodate her when she suffered an allergic reaction that caused her to need to switch to a different job within the warehouse. "It was above and beyond anything my previous employers would have done," Lee said.

"IT WAS ABOVE AND BEYOND ANYTHING MY PREVIOUS EMPLOYERS WOULD HAVE DONE."

Lee wants to continue her seasonal work at Amazon, and because of the noncompete that she's signed, she would be careful if she were to apply for a second job at an Amazon competitor like Sam's Club, the wholesale subsidiary of Walmart. Lee says, in this hypothetical scenario, she would be clear with the hiring agents at Sam's Club about the noncompete she'd signed at Amazon and would also contact Amazon to ask for permission for working at Sam's Club.

"I'd send Amazon a thing and say: 'I applied at Sam's Club. How do you feel?'" Lee said. "Then it would be up to Sam's Club to hire me and up to Amazon to say yes or no."

Lee's husband, Ray, is also a seasonal Amazon worker and says that he believes the noncompete contract only applies to trade secrets: "How technical is it to go and box stuff everyday and send it off?" Apart from their Amazon work, the couple has sought work at theme parks and campgrounds and have not had their behavior affected by noncompete agreements.

Several former Amazon workers in Kansas and Tennessee said that they had vague recollections of signing a noncompete agreement but did not give it much consideration. Two workers who had left Amazon warehouse jobs in 2012 and 2013 said they had no recollection of signing a non-compete agreement. It is unclear when Amazon began having warehouse workers sign this agreement, and the company did not respond to questions sent by *The Verge* about this. Two other Amazon workers approached by *The Verge* cited the nondisclosure agreement they had signed with the company in refusing to share their experiences for this story.

Amazon did not respond to a question asking for examples of jobs the agreement would bar its former warehouse workers from taking. But it should be noted that some of

America's largest employers, retailers like Walmart and Target, have established across-the-board programs to match prices of goods sold on Amazon.com, bringing the products sold in thousands of retail stores across the country into even more direct competition with those sorted through Amazon warehouses across the country.

"THEN IT WOULD BE UP TO SAM'S CLUB TO HIRE ME AND UP TO AMAZON TO SAY YES OR NO."

Although companies may push noncompetes on low-wage workers to keep trade secrets from leaking, there's also a more cynical explanation: to simply deprive competitors of employees to hire, according to Lobel. Noncompetes can also depress workers' wages. Traditionally, a key strategy to keep employees from defecting to a competitor has been simply to offer competitive wages, but a company that uses non-compete agreements can feel less pressure to pay well.

In this way, noncompetes can exacerbate structural inequalities in the current job market, inequalities which themselves make noncompetes easier for companies to demand. In America's post-recession economy, job seekers continue to vastly outnumber openings for good jobs. In this setting, workers don't have much leverage when haggling with employers over terms and conditions of work. One effect of this has been the expansion of the so-called "gig economy", where apps like Uber and TaskRabbit draw on a pool of freelancers ready to perform quick jobs that become available with no attendant promise of benefits or job security. Large numbers of unemployed and underemployed have also fueled the boom in temp-agency staffing that has accounted for significant portions of the country's post-recession job gains.

A lack of negotiating power can lead workers to sign noncompete contracts, Lobel says, and those contracts further erode their negotiating power. Because noncompetes make job loss more perilous by limiting post-employment opportunities, the agreements can tether workers to their current job, making them less likely to address grievances with management or attempt to look for better or more fitting work.

"The decisions of whether to leave or not become distorted by the fact that there's this stick over their heads in the post-employment phase," says Lobel, who examined the economic effects of noncompetes in her recent book titled *Talent Wants to Be Free*.

NONCOMPETES MAKE JOB LOSS MORE PERILOUS

"People very well might decide that, despite their unhappiness with their job, despite the fact that they think they can do better with another employer, they might decide that it's just not worth the risk and that they should just lay low."

Courts are often reluctant to enforce noncompete agreements that cover the entire United States, let alone the whole world, according to Garden, who notes that the standard of "reasonableness" is the main legal test of the agreements. Yet different states have far different ideas of what counts as reasonable. (In an apparent nod to this, the Amazon contract stipulates that the signer consents that "each and every covenant and restraint in this Agreement is reasonable.") California law bans the enforcement of noncompetes. Oregon, North Dakota, and Colorado have also enacted strict limits on noncompetes. "Then there are states like Texas and Florida and a bunch of others that are on the other end of the spectrum," says Lobel, "that think of it as a simple contract issue, and if you sign the contract and you breach it then, well, you've breached the contract, and they'll enforce it, and they'll give injunctions quite easily."

"THEY MIGHT DECIDE THAT IT'S JUST NOT WORTH THE RISK AND THAT THEY SHOULD JUST LAY LOW."

Such threats can have very concrete effects on workers. In the case of Jimmy John's, a number of former workers have sued the company in part for what they call the "oppressive" effects of the sandwich chain's noncompete clause. One former Jimmy John's employee in the suit alleges that, after leaving the sandwich chain, she had taken a telemarketing job that paid less than she could have earned bartending — all because of the sandwich company's noncompete. The AP recently found that Jimmy John's workers are not alone in suffering real-life consequences of signing the agreements, which, according to court records, have also ensnared nail stylists, maids, and agricultural workers. In Massachusetts, a barbershop forced one of its former hairdressers who had signed a noncompete into unemployment after he took a job at a competitor.

The signing of noncompete agreements are more prevalent in states that are more willing to enforce the contracts, according to preliminary findings of a study conducted by scholars at University of Illinois at Urbana Champaign and the University of Michigan. The paper, which polled more than 10,000 workers across the country with an online survey, also indicates that noncompetes are prevalent in jobs that often require little training. The survey found that at least 12 percent of US workers, or at least 19 million Americans, are working under the agreements. The paper found that roughly 9 percent of transportation and warehousing workers who answered the survey were working under noncompete contracts. (This statistic is an underestimation, says Evan Starr, a co-author of the study, because many people who sign noncompetes are unaware that they've done so.) The paper goes on to assert that while the millions of low-skilled workers signing the contracts "are far less likely to bargain over their noncompetes, they receive little in return for signing, but may bear serious costs."

"THEY RECEIVE LITTLE IN RETURN FOR SIGNING, BUT MAY BEAR SERIOUS COSTS."

Starr, who reviewed the Amazon agreement, said that while attorneys may differ in their interpretations on which services count as having been "supported" by a warehouse employee, the 18-month duration seems "incredibly long," especially for a temporary job. In the case of a stint lasting three months, the restrictions would stretch six times longer than the actual length of employment, Starr noted in an email. "A restriction like this could only be credible if the type of information the individual learned in a short time could be very damaging to the firms."

Yet Garden, the Seattle University law professor, notes that such a contract being legally enforceable may in fact be entirely beside the point in a low-wage workplace. "One way to look at this is as a kind of invidious approach to having workers sign a contract that is very likely to be unenforceable," Garden says. "Knowing that people who have been working for 10 and 11 dollars an hour are not going to be able to hire a lawyer to fight for them later on."

BUSINESS / TECH / AMAZON

Amazon does an about-face on controversial warehouse worker non-compete contracts / Company says it's removing a clause that could keep hourly employees from working elsewhere for 18 months

By Josh Lowensohn
Source The Guardian

Mar 27, 2015, 6:36 PM EDT



0 Comments (0 New)



Stephen Brasheax/Getty Images


Amazon is rolling back a controversial non-compete clause in its contracts for warehouse workers (including temporary ones) that could have kept them from working at competing companies for a year and a half. "That clause hasn't been

applied to hourly associates, and we're removing it," a company spokesperson told *The Guardian*.

"A very wide-ranging clause"

The controversial contract details were first reported by *The Verge* yesterday, and quickly drew public ire given their broad scope. Amazon sells just about everything, and the clause in question prohibited workers from going to a company that "directly or indirectly" competed with Amazon for a year and a half after their tenure. Even if Amazon never enforced the contract, it could still discourage workers from seeking employment elsewhere. Amazon also required workers to reaffirm their contracts, which included the non-compete clause, in exchange for severance after being laid off.

Amazon did not respond to repeated requests for comments before the publishing of *The Verge's* initial report on the contracts, and did not immediately respond to a request for comment on this latest change.

 0 COMMENTS (0 NEW)

FEATURED VIDEOS FROM THE VERGE

What's next for Microsoft's giant Activision Blizzard \$68.7 billion deal?

ENTREPRENEURSHIP POLICY DIGEST

SEPTEMBER 25, 2014
 UPDATED SEPTEMBER 14, 2015

THE IMPORTANCE OF YOUNG FIRMS FOR ECONOMIC GROWTH

Nearly eight years since the beginning of the Great Recession, the American economy finally gained back all of the jobs lost during the economic downturn. While this is positive news, underlying structural concerns remain, resulting in historically low labor force participation, high rates of unemployment and underemployment, and a "missing generation" of firms. Together, these factors are a drag on the economy, sapping dynamism.

Policymakers often think of small business as the employment engine of the economy. But when it comes to job-creating power, it is not the size of the business that matters as much as it is the age. New and young companies are the primary source of job creation in the American economy. Not only that, but these firms also contribute to economic dynamism by injecting competition into markets and spurring innovation.

Representing 95 percent of all U.S. companies, businesses with fewer than fifty employees are undoubtedly important to overall economic strength. So too are the relatively few large companies that employ millions of Americans. Yet, neither group contributes to new job creation in the way young, entrepreneurial firms do. In fact, between 1988 and 2012, companies more than five years old destroyed more jobs than they created in all but eight of those years.

Yet, the startup news is not all good. The rate at which new businesses are opening has been steadily declining until 2014. Because of their out-sized contributions, this decline has troubling implications for economic dynamism and growth if it is not reversed.

YOUNG FIRMS DRIVE JOB GROWTH AND ECONOMIC DYNAMISM

- New businesses account for nearly all net new job creation and almost 20 percent of gross job creation, whereas small businesses do not have a significant impact on job growth when age is accounted for.
- Companies less than one year old have created an average of 1.5 million jobs per year over the past three decades.

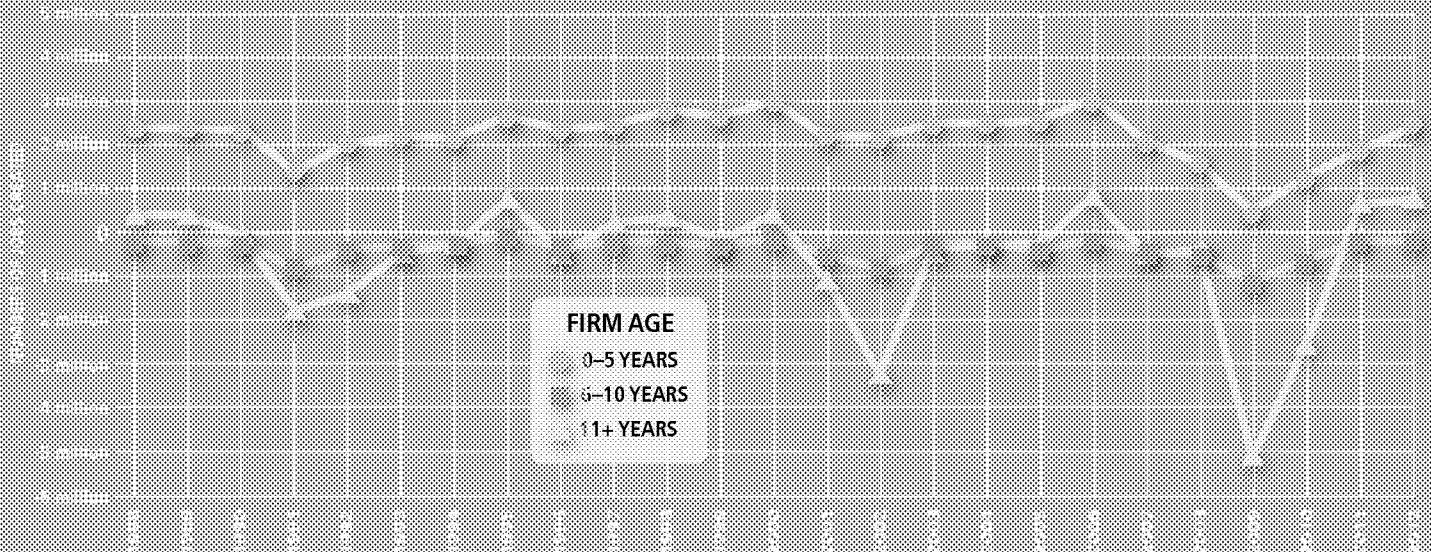
continued

SOURCES OF QUALITY DATA ON YOUNG FIRMS

Understanding the economic impact of young firms requires dependable data. Sources from the U.S. Census Bureau include:

- Business Dynamics Statistics
- Survey of Business Owners
- Quarterly Workforce Indicators
- Annual Survey of Entrepreneurs

NET JOB CREATION



SOURCE: U.S. CENSUS BUREAU BUSINESS DYNAMICS STATISTICS

- Many young firms exhibit an “up or out” dynamic, in which innovative and successful firms grow rapidly and become a wellspring of job and economic growth, or quickly fail and exit the market, allowing capital to be put to more productive uses.
- Young firms were hit hard during the Great Recession. Even still, from 2006 to 2009, young and small firms (fewer than five years old and twenty employees) remained a positive source of net employment growth (8.6 percent), whereas older and larger firms shed more jobs than they created.

DECLINING STARTUP RATES THREATEN GROWTH

- New businesses represent a declining share of the business community. According to Census data, new firms represented as much as 16 percent of all firms in the late 1970s. By 2011, that share had declined to 8 percent.
- Not only are there fewer new firms, but those startups that do exist are creating fewer jobs. The gross number of jobs created by new firms fell by more than two million between 2005 and 2010.
- Startup activity has been subdued across the country. Firm entry rates were lower between 2009 and 2011 than they were between 1978 and 1980 in every state and Metropolitan Statistical Area except one.

PAVE THE WAY FOR AN ENTREPRENEURIAL RENAISSANCE

Policies at the federal, state, and local levels influence an individual’s ability to start a business and impact firm growth and survival. Policymakers at all levels can help create an environment more conducive to business formation.

FEDERAL

Welcome Immigrants

- Immigrants were nearly twice as likely as native-born Americans to start businesses in 2014. The creation of a visa for immigrant entrepreneurs would allow these job creators to start companies in the United States.

Remove Regulatory Barriers to Growth

- As regulations build up over time, they represent an increasing and disproportionate cost to entrepreneurial firms. Ideas to counter regulatory accumulation include the establishment of a commission to review and recommend regulatory changes to Congress and implementing sunset dates on major regulations.

STATE

Simplify Tax Codes and Payment Systems

- Taxes matter, but what entrepreneurs are most concerned about is tax complexity. Simplifying tax codes and payment systems so they are easier to understand will relieve what many entrepreneurs feel is a burden on them and their businesses.

Encourage Competition and Labor Mobility

- Occupational licensing and non-compete agreements can depress entrepreneurship by artificially inflating the cost to enter a new market and restricting the free movement of individuals. Reconsider licensing requirements and adjust non-competes to spur entrepreneurial growth.

LOCAL

Cultivate Human Capital

- Higher levels of education are associated with increased entrepreneurial activity. An analysis of 356 U.S. metropolitan areas found that high school and college completion is important to startup rates.

FOR MORE INFORMATION

Click on the links for access to the following resources, or contact Jason Wiens at jwiens@kauffman.org:

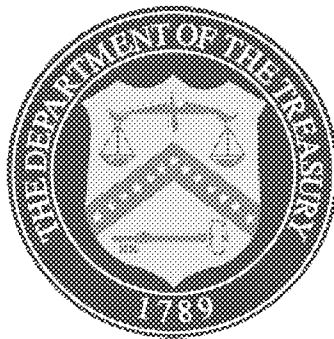
- Watch the Kauffman Foundation’s *Three Things* sketchbook
- Read Kauffman Foundation testimony before the U.S. House Committee on Small Business
- Read the Kauffman Foundation’s *Startup Act* and *Startup Act for the States*
- Read *Young Firms are the Job Creators*
- Read *Can Millennials Reverse America’s Declining Rates of Entrepreneurship?*



To sign up to receive subsequent Policy Digests, go to www.kauffman.org/policydigest.

1001 PENNSYLVANIA

4801 Rockhill Road • Kansas City, Missouri 64110 • www.kauffman.org



Office of Economic Policy
U.S. Department of the Treasury

**Non-compete Contracts: Economic Effects and Policy
Implications**

March 2016

Table of Contents

Executive Summary	3
Section I. Non-competes and Their Justifications	6
Section II. What Can We Say About the Justifications?	11
Section III. The Details of Non-compete Enforcement	14
Section IV. Effects of Non-compete Enforcement	18
Section V. Directions for Reform	24
Appendix A	27
Appendix B	32
Works Cited	34

Executive Summary

Non-compete agreements are contracts between workers and firms that delay employees' ability to work for competing firms. Employers use these agreements for a variety of reasons: they can protect trade secrets, reduce labor turnover, impose costs on competing firms, and improve employer leverage in future negotiations with workers. However, many of these benefits come at the expense of workers and the broader economy. Recent research suggests that a considerable number of American workers (18 percent of all workers, or nearly 30 million people) are covered by non-compete agreements.¹ The prevalence of such agreements raises important questions about how they affect worker welfare, job mobility, business dynamics, and economic growth more generally. This report presents insights from economic theory and evidence on the economic effects of non-compete agreements. It goes on to discuss policy implications, starting a discussion about how such agreements could be used in a way that balances the interests of firms with those of workers and society as a whole.

Non-compete agreements have social benefits in some situations.

- Non-competes are sometimes used to protect trade secrets, which can promote innovation.
- By reducing the probability of worker exit, non-competes may increase employers' incentives to provide costly training.
- Employers with especially high turnover costs could use non-competes to match with workers who have a low desire to switch jobs in the future.

But non-compete agreements can also impose large costs on workers.

- Worker bargaining power is reduced after a non-compete is signed, possibly leading to lower wages.
- Non-competes sometimes induce workers to leave their occupations entirely, foregoing accumulated training and experience in their fields.

¹ These and other similar numbers throughout the executive summary and report are from Starr, Bishara, and Prescott (2015) and private correspondence with the authors. Note that all figures are preliminary and may change slightly.

- Reduced job churn caused by non-competes is itself a concern for the U.S. economy. Job churn helps to raise labor productivity by achieving a better matching of workers and firms, and may facilitate the development of industrial clusters like Silicon Valley.

Moreover, there is reason to believe that many specific instances of non-compete agreements are less likely to produce social benefits.

- Non-competes are often used by employers in non-transparent ways:
 - Many workers do not realize when they accept a job that they have signed a non-compete, or they do not understand its implications.
 - Many workers are asked to sign a non-compete only after accepting a job offer. One lower-bound estimate is that 37 percent of workers are in this position.
 - Many firms ask workers to sign non-competes that are entirely or partly unenforceable in certain jurisdictions, suggesting that firms may be relying on a lack of worker knowledge. For instance, California workers are bound by non-competes at a rate slightly higher than the national average (19 percent), despite the fact that, with limited exceptions, non-competes are not enforced in that state.²
- Only 24 percent of workers report that they possess trade secrets. Moreover, less than half of workers who have non-competes also report possessing trade secrets, suggesting that trade secrets cannot explain the majority of non-compete activity.
- Non-competes are common among workers who report lower rates of trade secret possession: 15 percent of workers without a four-year college degree are subject to non-competes, and 14 percent of workers earning less than \$40,000 have non-competes. This is true even though workers without four-year degrees are half as likely to possess trade secrets as those with four-year degrees, and workers earning less than \$40,000 possess trade secrets at less than half the rate of their higher-earning counterparts.
- Available evidence suggests that workers with a low initial desire to switch jobs are *not* more likely to match with employers who require non-competes.
- In some cases, non-competes prevent workers from finding new employment even after being fired without cause; in such cases, it is difficult to believe that non-competes yield social benefits.

² Depending on the facts of the individual case, such non-competes may be enforced in other states.

States vary greatly in the manner and degree to which they will enforce non-competes.

- In some states, non-compete enforcement is determined by statute, while in others it is determined exclusively by case law.
- Some states refuse to enforce non-competes, or refuse to enforce non-competes that contain any unenforceable provisions (“red-pencil” doctrine), although a majority of states will modify overbroad non-compete contracts to render them enforceable (“blue-pencil” and “equitable reform” doctrines).

The analysis in this report suggests several broad recommendations that would minimize the harms associated with non-compete agreements.

- Increase transparency in the offering of non-competes.
- Encourage employers to use enforceable non-compete contracts.
- Require that firms provide “consideration” to workers bound by non-compete contracts in exchange for both signing and abiding by non-competes.

I. Non-competes and Their Justifications³

Non-compete contracts – agreements between workers and firms that restrict workers’ ability to take new employment – have a long history, but their scope, prevalence, and enforcement have varied widely across time and place. With the recent development of more comprehensive data on their usage, it has become more apparent that non-competes are an important labor market institution meriting careful study. Recent research shows that as many as 30 million workers are currently covered by non-compete agreements. While in some cases non-compete agreements can promote innovation, their misuse can benefit firms at the expense of workers and the broader economy. Details of non-competes and their enforcement have implications for worker bargaining power, job mobility, and economic growth. This report draws on insights from economic theory, as well as a rapidly growing body of empirical evidence, to help clarify thinking about non-competes and non-compete reform.

What are non-competes and who is bound by them?

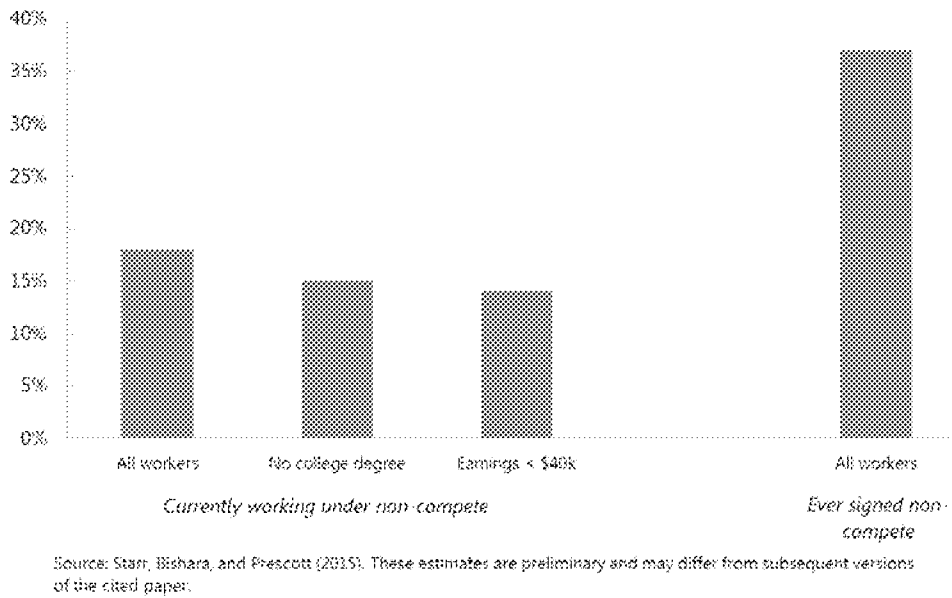
Many employers ask their employees to sign non-compete agreements. The details of these contracts vary greatly across firms and states, but they share a common purpose: restricting the ability of a worker to compete with his or her current employer for some specified period of time, often in a specified geographic area. Typically, this takes the form of a prohibition on taking employment at a rival firm, where “rival” may be interpreted quite broadly to include all firms within a given industry.

Non-compete agreements have become quite common among a variety of types of workers. As shown in the chart below, roughly 18 percent of workers currently report working under a non-compete agreement and about 37 percent of workers report having worked under one at some point during their career. Although such agreements are less common among less-educated workers and lower-income workers, the fractions of these workers operating under one are still substantial.⁴

³ This report benefited greatly from discussions with Professor Evan Starr, and we are grateful for his time and expertise. We also make extensive use of Starr, Bishara, and Prescott (2015). However, the views expressed here are not necessarily those of Starr and his coauthors, nor are they implicated in any errors.

⁴ See Starr, Bishara, and Prescott (2015).

Percentage of Workers with Non-compete Agreements, by Group



How are non-competes typically justified?

The conventional picture of a workplace characterized by non-compete agreements is one that features trade secrets, including sophisticated technical information and business practices that firms have a strong interest in protecting. By preventing a worker from taking such secrets to a firm’s competitors, the non-compete essentially solves a “hold-up” problem: *ex ante*, both worker and firm have an interest in sharing vital information, as this raises the worker’s productivity. But *ex post*, the worker has an incentive to threaten the firm with divulgence of the information, raising his or her compensation by some amount equal to or less than the firm’s valuation of the information. Predicting this state of affairs, the firm is unwilling to share the information in the first place unless it has some legal recourse like a non-compete contract.

Occasionally, client relationships are included along with trade secrets in this explanation (and are sometimes treated similarly as a matter of state law). However, it is not clear that relationships with clients constitute a socially valuable investment analogous to trade secrets.⁵ For this reason, trade secrets will be the focus of discussion in this report.

⁵ For instance, a trade secret involving intellectual property may be the product of expensive investments. If the investment had not been made, none of the benefits of the property would have been realized. By contrast, the

While non-competes help solve the trade secrets “hold-up” problem, they are not the only tool at employers’ disposal. States generally have laws prohibiting theft or disclosure of trade secrets. In addition, employers can use compensation schemes that discourage turnover for workers with trade secret access (e.g., employers may provide additional compensation contingent on the worker remaining at the firm).⁶ We provide further evidence regarding trade secrets later in the report.

What are other possible explanations?

What might explain the existence of non-competes among workers who are not plausibly affected by the sort of trade secrets discussed previously? A number of explanations have been suggested. One possibility (*training*) – which may coexist with either of the next two explanations – is that firms and workers use non-competes to encourage more investment in workers. In general, firms are reluctant to pay for training that improves a worker’s “general” skills and makes her more valuable to it and other firms alike. Economists usually think of general training as occurring when workers accept wage cuts to compensate their employer for its expenses in providing the training.⁷ For various practical reasons, however, workers may be unwilling to pay for training.⁸ Non-competes offer an alternative: firms get an assurance that workers are unlikely to leave for some period of time, allowing the firm to capture more of the increased productivity from costly training it provides, and workers receive more training than they otherwise would.

Another possibility (*screening*) is that non-competes are an attempt by firms to preferentially hire workers with a low likelihood of departure. Underlying this alternative is the assumption that firms face substantial costs for hiring and separating with workers.⁹ Moreover, it is not obvious to firms which workers are most likely to exit, and workers cannot credibly assert their probability of leaving (i.e., all workers will pretend to have a very low probability, as this raises their perceived value to the firm). By making non-competes a condition of employment, firms

client, and their need for a good or service, presumably exist independently of any investment made by the employer.

⁶ See Salop and Salop (1976) for one discussion of such a mechanism.

⁷ See Becker (1962).

⁸ For instance, workers may be credit-constrained and unable to finance the training, or workers may have difficulty observing the quality of the training, rendering them less willing to pay for it.

⁹ See Hamermesh (1995).

reduce the value of the job to those workers who know they are likely to depart. For those workers who do not expect to leave imminently, the non-compete is less of an imposition. Note that in order for this explanation to be correct, prospective workers must understand the non-compete and its implications.

A final explanation (henceforth referred to as *lack of salience*) is that workers do not pay attention to non-compete contracts and do not realize how much bargaining power and future employment flexibility they are foregoing. Only later, when workers consider exiting a firm, do they become aware of the existence and/or implications of the non-compete agreement.¹⁰ Other workers may be aware of the non-compete, but only after it is presented to them once they have accepted a position or started working, and not at the time the job offer was originally extended. According to this explanation, only employers benefit from the non-compete, as they obtain increased bargaining power in future wage negotiations, reduced turnover costs, and possible impairment of rivals' ability to hire.

How do the different non-compete explanations affect the optimal policy response?

The explanations for non-compete agreements described above have different implications for the desirability of such agreements. Thinking through these implications helps to shed light on the appropriate policy response. The first three explanations – trade secrets, training, and screening – suggest that non-competes can be socially desirable. The last explanation, lack of salience, suggests that non-competes are socially harmful.

The conventional explanation for non-compete agreements involving protection of trade secrets is a potentially strong justification for such agreements where it genuinely applies, and where other devices for protection of employers (like trade secrets law) are not effective. As previously discussed, non-competes can encourage additional economic activity and broader information sharing when trade secrets are significant.

The training and screening explanations for non-compete agreements also suggest social benefits. If worker training is sufficiently enhanced by the availability of non-competes, or if

¹⁰ Research in other contexts has found a large role for salience considerations. See Kahneman (2003) for a discussion of salience as it relates to behavioral economics, and Feldman, Katuscak, and Kawano (2016) for an example from the tax literature.

firms with unusually high separation costs are able to match more appropriately with workers, both worker and firm are better off. Balanced against these benefits are the social costs associated with diminished mobility.

The final explanation for non-compete agreements – lack of salience – implies that non-competes are merely a costly transfer from workers to firms, made possible by workers' lack of awareness. According to this explanation, non-competes lead to diminished worker mobility and a loss of human capital, with no corresponding benefit to society. When workers are legally prevented from accepting competitors' offers, those workers have less leverage in wage negotiations and fewer opportunities to develop their careers outside of their current firm. By contrast, the firms using non-competes benefit through reduced turnover costs, increased bargaining power, and denial of valuable employees to competitors.

Constructing ideal policy for non-competes requires determining which explanation is most relevant for a particular type of worker (i.e., for low-skill service workers vs. high-skill IT workers), and balancing the trade-offs between non-competes' benefits and their undesirable consequences. For instance, low-wage workers may be particularly poorly served by non-competes due to the lower likelihood that trade secrets are relevant.

However, it is not always easy to distinguish among the different explanations for non-competes, and several possible reforms are beneficial regardless of the underlying explanation. For example, measures to improve the salience and transparency of non-competes and non-compete enforceability are broadly useful and will help to minimize the worst effects of non-competes.

In Section V, some directions for policy reform are described and their reasoning briefly explained.

II. What Can We Say About the Justifications?

Research on non-competes is still at an early stage. However, a recent paper provides comprehensive data on workers with non-competes, answering many of the most important questions about these workers.¹¹ In addition to collecting information on the characteristics of workers who sign non-competes, this research also examines the extent to which workers with non-competes actually interact with clients, have access to client-specific information, and work with trade secrets.¹² This section summarizes the literature examining the different rationales for non-compete agreements.

Protecting trade secrets. If protection of trade secrets were the main explanation for non-compete agreements, then one would expect such agreements to be highly concentrated among workers with advanced education and occupations likely to feature trade secrets.^{13,14} However, the fraction of workers without a four-year college degree reporting a current non-compete agreement is about 15 percent, only slightly below the 18 percent share for all workers.¹⁵ While engineering and computer/mathematical occupations have the highest non-compete prevalence at slightly more than one-third, occupations like personal services and installation and repair also include many workers with non-competes, at about 18 percent. When entry-level workers at fast food restaurants are asked to sign two-year non-competes, it becomes less plausible that trade secrets are always the primary motivation for such agreements.¹⁶

Unsurprisingly, workers who reported access to trade secrets were much more likely to be bound by a non-compete, with about a 25 percentage point higher probability than those who report no interaction with clients, no access to client-specific information, and no possession of trade secrets. The link between client access and non-competes is not as strong: those who report such

¹¹ See Starr, Bishara, and Prescott (2015).

¹² As the authors' data is collected through an online survey, achieving a representative sample may be challenging. The authors note, however, that more traditional survey designs face similar difficulties.

¹³ Note that not all trade secrets are equivalent from an economic perspective. Though the legal definition of trade secrets embraces a wide variety of private information (e.g., fast-food recipes), some of these examples may not involve a substantial "hold-up" problem of the kind described above.

¹⁴ See Starr, Bishara, and Prescott (2015) for evidence that occupations and income groups differ substantially in the degree to which they involve trade secrets.

¹⁵ See Starr, Bishara, and Prescott (2015).

¹⁶ See <http://www.forbes.com/sites/clareoconnor/2014/10/15/does-jimmy-johns-non-compete-clause-for-sandwich-makers-have-legal-legs/>.

access (but no trade secrets) have about a 7 percentage point higher probability of a non-compete. However, less than half of all workers with non-competes report possessing trade secrets. Together, these findings suggest that the trade secrets explanation is likely part, but not all, of the story of non-competes.

Encouraging training. Non-compete enforcement is associated with more worker training. Evan Starr finds that a “one standard deviation increase in a state’s overall enforceability level increases the probability that the average high litigation occupation receives firm-sponsored training by 2.4% relative to low litigation occupations.”¹⁷ Interestingly, this work finds that when states require firms to offer substantial “consideration” along with a non-compete (e.g., promotions, training, and higher wages), both training and wage outcomes for workers are improved.

Facilitating screening. Starr, Bishara, and Prescott have developed data that are directly relevant to the question of screening by asking their survey respondents how long they expected to work for their current employer, then comparing the responses of workers who have and have not signed non-competes. Interestingly, after controlling for various demographic and economic variables, there is no relationship between expected tenure and likelihood of having signed a non-compete. This result suggests that screening is not an important part of the non-compete story.

Exploiting lack of salience. Several pieces of evidence suggest that employers are relying on workers’ incomplete understanding of non-compete agreements. First, employers often require that workers sign non-compete agreements even in states that refuse to enforce them. For example, in California, which (with limited exceptions) does not enforce non-compete agreements, the fraction of workers currently under a non-compete is 19 percent, which is slightly higher than the national average.

Second, a separate survey, exclusively focused on members of the Institute of Electrical and Electronics Engineers, reports that “...barely 3 in 10 workers reported that they were told about the non-compete in their job offer. In nearly 70% of cases, the worker was asked to sign the

¹⁷ See Starr (2015), page 3. “Enforceability level” is defined by Starr to capture all the dimensions of non-compete enforcement, and “high-litigation” refers to occupations characterized by more legal action related to non-compete contracts.

non-compete after accepting the offer – and, consequently, after having turned down (all) other offers. Nearly half the time, the non-compete was not presented to employees until or after the first day at work.”¹⁸ This evidence is especially powerful insofar as it applies to highly-educated, high-wage workers who might be considered more likely to understand the process surrounding non-competes. Even in cases where the conventional explanation of trade secrets has a surface plausibility, firms often delay the presentation of non-competes. This behavior would not be necessary if non-competes were a mutually-beneficial arrangement.

Finally, Starr, Bishara, and Prescott (2015) find that only 10 percent of workers with non-competes report bargaining over their non-compete, with 38 percent of the non-bargainers not realizing that they could even negotiate.¹⁹ Moreover, workers appear confused as to whether non-competes are even enforceable in their states. In preliminary work by Starr and coauthors, workers are shown to be frequently incorrect or unsure as to whether their non-competes are actually enforceable. Again, this is not consistent with a “perfect information” setting in which workers knowingly accepted the limitations imposed by non-competes.

¹⁸ See Marx and Fleming (2012), page 49.

¹⁹ See Starr, Bishara, and Prescott (2015).

III. The Details of Non-compete Enforcement

Non-compete enforcement differs significantly across states. Some relevant terms of art are defined below.

Non-compete contract: A contract that delays or in some other way restricts a worker's ability to compete with a previous employer. Typically this entails restrictions on future employment.

Consideration: A benefit received by a signatory to a contract. Generally, both parties must receive consideration in order for a contract to be valid. Consideration commonly includes property or promises of specific actions. In the case of a non-compete, consideration may sometimes refer to wage increases, promotions, or continued employment (sometimes including hiring).

Protectable interests: These are the aspects of an employer-employee relationship that provide the legal motivation for a non-compete agreement. They vary state to state, but frequently include trade secrets, confidential information, goodwill, and/or client relationships. Some states additionally provide protection for special training.

Red-pencil doctrine: Doctrine prevailing in some states requiring that courts must declare an entire non-compete contract void if one or more of its provisions are found to be defective under state law or precedent.

Blue-pencil doctrine: Doctrine prevailing in some states requiring that courts delete provisions of a non-compete contract that render it overbroad or otherwise defective, retaining the enforceable subset of the contract.

Equitable reform, aka Reformation: Doctrine prevailing in some states requiring that courts may rewrite a non-compete contract so as to render it non-defective. Unlike blue-pencil doctrine, this may entail insertions of new text.

Currently, nearly all states will enforce non-compete agreements to some extent. Within those states, non-compete enforcement may be restricted in a variety of ways that vary from state to state. See Beck Reed Riden LLP for a summary of state rules.²⁰

Judicial modification of non-competes. Rather than declaring specific contracts completely enforceable or unenforceable, courts in certain states may alter the contracts themselves. In those states, judges may declare portions of a contract void but other parts to be valid under what is called “blue pencil doctrine.”

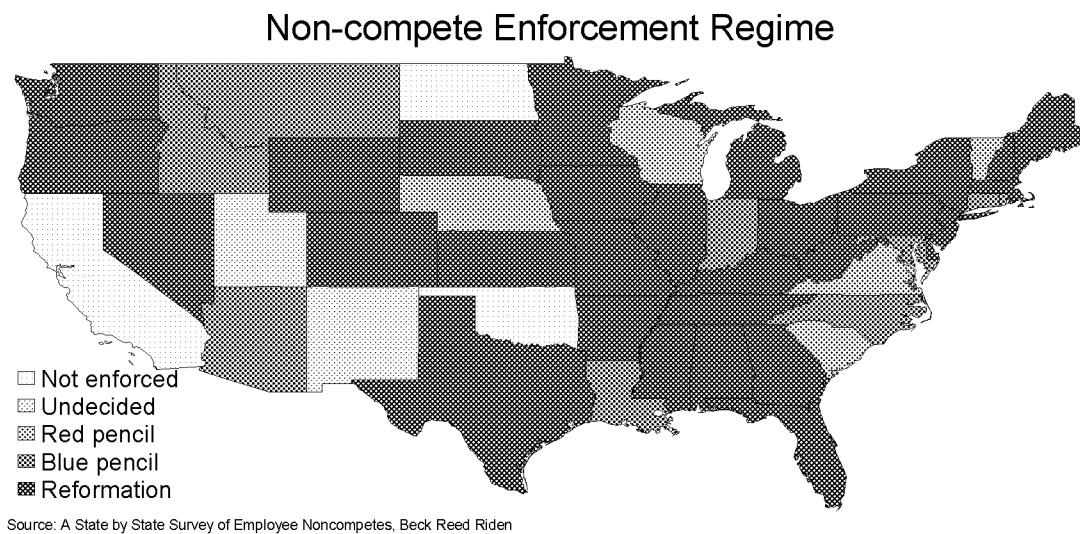
The following stylized example may help to explain how this doctrine might work. Suppose that a contract states that “The employee agrees not to work for any business competitive with the employer for one year in the following counties: Leelanau, Benzie, and Manistee.” Purely hypothetically, a judge might find the inclusion of Benzie to be overbroad, and could determine that the non-compete is valid once Benzie County is removed. However, as blue-pencil doctrine does not allow a court to add terms to a contract, the contract could not be revised to add “agrees not to work *in an administrative capacity*”, were the court to hold that this qualifier was necessary to prevent the contract from being overbroad.

In other states, an “equitable reform” or “reformation” doctrine allows judges to amend the language in question to generate an enforceable contract consistent with the original intent of the existing contract.²¹ This allows more flexibility than the blue pencil rule and increases the likelihood of a non-compete being upheld in some form, all else equal. It may also encourage firms to take risks in the writing of contracts, including provisions likely to be struck down. If workers do not have a good sense of which parts of a contract are enforceable, then these untenable provisions may still affect their behavior. On the opposite end of the spectrum, some states simply do not allow any judicial modification of contracts, but instead hold that any unenforceable provisions render the entire contract unenforceable. This is sometimes known as “red-pencil” doctrine.

²⁰ Other summaries of non-compete law exist and are in some cases slightly inconsistent with the Beck Reed Riden table we use; see “Summary of Covenants Not to Compete: A Global Perspective” by Fenwick and West LLP, for one alternative.

²¹ In some cases, this doctrine is (confusingly) also referred to as “blue-pencil.”

The figure below illustrates findings by one survey of the use of each rule by state.²² See Appendix B for additional figures illustrating the survey’s findings regarding other important dimensions of non-compete enforcement, including treatment of trade secrets, enforceability in case of firing without cause, and whether “continued employment” counts as worker consideration in exchange for a non-compete.



Quits vs. Layoffs. The paradigmatic case of non-compete enforcement is one in which an employee quits and is prevented from working for a competitor. However, even fired workers are often bound by non-compete contracts. One survey reports that, as of 2015, non-competes were enforceable against employees discharged without cause in about half of states.²³

Recent changes in non-compete enforcement. Several states have recently altered their approaches to non-compete enforcement. Notably, Georgia amended its constitution in 2011 to allow for increased enforcement of non-compete agreements.²⁴ Other states have altered their statutes to extend or limit the reaches of non-competes, as with a recent statute in Alabama that more explicitly explains what is and is not a valid protectable interest. Like Alabama, Oregon passed a statute that more clearly defines the bounds of a non-compete. As of 2016, new non-competes in Oregon will be limited to a maximum of 18-month duration. New Mexico also

²² Alaska and Hawaii, not shown, are both “reformation” states.

²³ See Beck Reed Riden LLP (2015).

²⁴ See <http://www.lexology.com/library/detail.aspx?g=aadbea62-9a31-4ac3-92dd-8198906c37f6>.

more clearly defined the bounds of non-competes, restricting their enforceability for certain health care practitioners. In Hawaii, non-competes have been prohibited for tech workers.²⁵

In other states, legislators have recently proposed significant changes. A bill similar to that passed in Hawaii was introduced, but not enacted, in Missouri. In New Jersey and Maryland, bills were proposed that would render non-competes unenforceable for any workers eligible to receive unemployment compensation. State legislators in Massachusetts, Michigan, and Washington have proposed that non-competes be made largely unenforceable in their states. Finally, Senators Franken and Murphy have proposed that firms be prohibited from entering into non-compete agreements with workers making less than \$15 per hour.²⁶

Appendix A provides a brief summary of the development of non-compete law over the long run.

²⁵ See <https://legiscan.com/AL/bill/HB352/2015>, <http://gov.oregonlive.com/bill/2015/HB3236/>, <http://www.nmlegis.gov/lcs/legislation.aspx?chamber=S&legtype=B&legno=325&year=15>, and http://www.capitol.hawaii.gov/session2015/bills/HB1090_CD1_.HTM, for Alabama, Oregon, New Mexico, and Hawaii law, respectively.

²⁶ Private correspondence with Evan Starr. See <http://house.mo.gov/billtracking/bills151/billpdf/intro/HB05971.PDF>, http://www.njleg.state.nj.us/2012/Bills/A4000/3970_11.HTM, http://mgaleg.maryland.gov/2013RS/fnotes/bil_0001/sb0051.pdf, <https://malegislature.gov/Bills/189/House/H1701>, [http://www.legislature.mi.gov/\(S\(eemdorjddevzki1vwou5lszu\)\)/mileg.aspx?page=GetObject&objectName=2015-HB-4198](http://www.legislature.mi.gov/(S(eemdorjddevzki1vwou5lszu))/mileg.aspx?page=GetObject&objectName=2015-HB-4198), <http://app.leg.wa.gov/billinfo/summary.aspx?year=2015&bill=2931>, for Missouri, New Jersey, Maryland, Massachusetts, Michigan, and Washington, respectively. See <http://www Franken.senate.gov/files/documents/150604MOVEsummary.pdf> for the proposed federal bill.

IV. Effects of Non-compete Enforcement

The effects of non-compete enforcement on mobility. According to authors of a recent study, the state of Michigan inadvertently “legalized” non-competes in 1985.²⁷ This presented a rarely available opportunity to study the effect of non-compete enforcement. Typically, it is difficult to rule out the possibility that changes in law reflect changes in current or expected economic circumstances. Thus, a simple comparison of economic outcomes before and after a state legalizes non-competes will include the effects of both these changes in circumstances and non-compete enforcement itself, making it difficult to separately estimate the latter effect. But in the case of Michigan, with its allegedly accidental and unanticipated change in the enforceability of non-competes, researchers can more reliably interpret changes in outcomes (e.g., labor mobility) as being caused by non-compete enforcement.

Marx, Strumsky, and Fleming exploit this natural experiment, showing that worker job mobility fell by 8 percent when non-competes were made enforceable, with the effect even larger for workers with more narrowly-focused human capital. However, other authors dispute these findings, arguing that the inadvertent legalization was not retroactive and that some states were inappropriately labeled as “non-enforcing.”²⁸ In separate work, Marx finds that workers who do switch jobs are more likely to leave their industry if they are covered by a non-compete, with the attendant “reduced compensation, atrophy of their skills, and estrangement from their professional networks” that would be expected to occur.²⁹

The effects of non-compete enforcement on wages. The literature on the effect of non-competes on wages is small, consisting largely of case studies, surveys of specific professions (e.g., electrical engineers), theoretical papers, and a recent analysis based on a broad online survey.³⁰ We therefore combine information from previous literature on enforceability and non-compete prevalence with standard labor market data, generating suggestive evidence on the wage impacts

²⁷ See Marx and Fleming (2012) for details.

²⁸ See Sichelman and Barnett (2015).

²⁹ See Marx, Strumsky, and Fleming (2009) and Marx (2011).

³⁰ See various papers by Marx, Marx and Fleming (2012), Meccheri (2009), and Starr, Bishara, and Prescott (2015), respectively. We are not aware of any panel data with individual responses to questions about non-competes, and existing work typically does not present population-wide inferences about the wage effects of non-compete enforcement.

of non-compete enforcement.³¹ Interestingly, we find stricter non-compete enforcement to be associated with both lower wage growth and lower initial wages.³²

The first column of Table 1 shows the percentage change in wages from a one-unit increase in a non-compete enforceability index, holding constant a number of worker characteristics.³³ It suggests that a standard deviation in non-compete enforcement reduces wages by about 1.4 percent. Recent work by Starr and coauthors finds broadly similar results to those presented here.³⁴

It is possible to refine this approach by focusing more narrowly on populations likely to be affected by non-competes. Workers with bachelor's degrees are more than 50 percent more likely to be bound by non-competes than those without, suggesting that one might better approximate the “eligible” subgroup by restricting the sample to workers with bachelor's degrees. This is shown in Table 1, column 2. Note that the magnitude of the wage effect of non-compete enforcement increases for this subgroup, as expected. A slightly more nuanced approach makes use of the occupational breakdown provided in recent work. Rather than omitting non-college workers, we instead reweight the sample to be more representative of workers with non-competes. For example, this will imply placing a higher weight on workers in the architecture and engineering occupations than in the personal services occupations. Table 1, column 3 shows results from this reweighted approach. The magnitude of the wage impact is again above that of column 1, but not dramatically so.³⁵

³¹ We use the 2014 merged outgoing rotation groups of the Current Population Survey (CPS), which provide a cross section of population-representative workers. Merged with this data is the Starr-Bishara index of non-compete enforceability by state (generously provided by Evan Starr), as well as the fraction of workers with non-competes by major occupation from Starr, Bishara, and Prescott (2015).

³² Here again, the particular proposed explanation for non-competes is important. For instance, if *screening* is the dominant explanation, and workers are fully informed about non-competes, we would expect stricter enforcement to cause an initial wage premium but slower subsequent wage growth. Workers would only be willing to sign the non-compete if they were compensated at the time of signing. If, on the other hand, *salience* is the dominant explanation, we would expect no initial premium and slower wage growth, as workers are prevented from taking advantage of outside opportunities or using outside opportunities as leverage for wage growth at the current firm.

³³ These controls consist of education, age, gender, marital status, occupation, industry, public sector status, and union status.

³⁴ See forthcoming work by Balasubramanian, Chang, Sakakibara, Sivadasan, and Starr, as well as Starr, Ganco, and Campbell.

³⁵ This is perhaps to be expected given the fact that non-competes are used quite broadly. While non-competes are more common in particular occupations (e.g., management, computer and mathematical, and architectural and engineering occupations), they are also found in a wide variety of unexpected occupations and education levels.

Table 1. Wage effect of one standard deviation of noncompete enforcement

	Original sample	College-only	Occupation-reweighted
Enforcement	-1.38% (-13.55)	-1.86% (9.73)	-1.52% (12.49)
N	163,252	57,156	163,252
R ²	0.43	0.25	0.43

Source: 2014 Current Population Survey, Starr et al. (2015), private correspondence with Starr, and Treasury calculations. All estimates are conditional on education, marital status, union status, sex, major occupation and industry, public sector status, and a quadratic in age. T-statistics are in parentheses.

Much of the research on non-competes has focused on their relationship with on-the-job training. Non-competes and non-compete enforceability may affect the rate at which wages grow with employee tenure and experience. We therefore examine the association of non-compete enforceability with age-wage profiles, i.e., the rate at which wages increase with age.

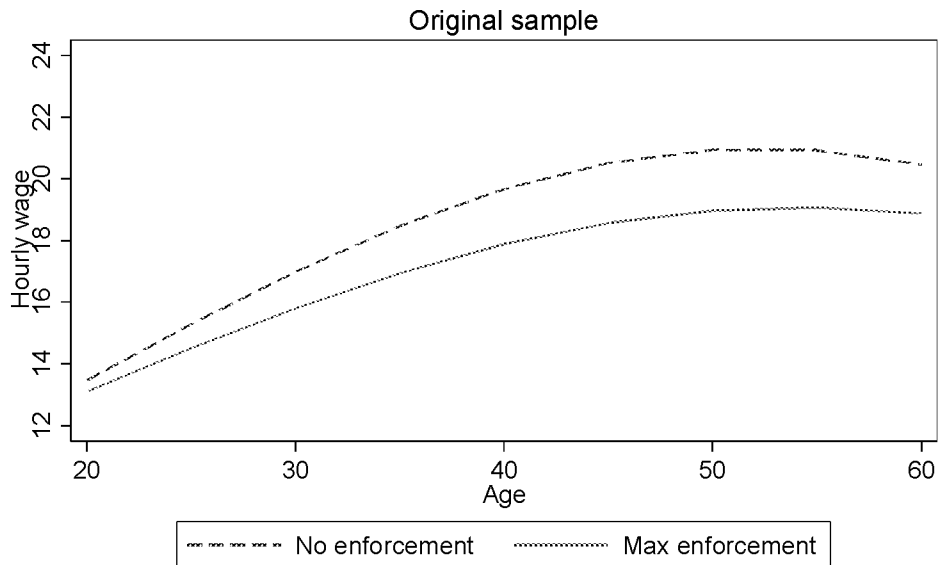
Figures 1 and 2 below are plots of age-wage profiles in a minimally-enforcing state and a maximally-enforcing state, for original and occupation-reweighted samples, respectively. As workers age, the effect of tightened non-compete enforcement appears to rise: using the original sample, the effect of maximal enforcement, relative to minimal enforcement, is 5 percent at age 25 and 10 percent at age 50. As with the previous results, the occupation-reweighted projections show a somewhat larger difference between wages in minimally- and maximally-enforcing states.

Are these results surprising? If non-competes existed exclusively to promote training, one would expect states with stronger enforcement to see faster wage growth over the life cycle. If, on the other hand, non-competes are the product of a lack of salience for workers, one would instead expect to see the pattern shown in Figures 1 and 2.³⁶ As workers progress through their careers, switching jobs is more difficult in states that stringently enforce non-competes. Given that job

³⁶ When interpreting any of the results just described, it should be remembered that we are not exploiting variation over time in non-compete enforcement; rather, the wage estimates are derived from variation across states. Even after controlling for available worker-level variables, states may differ in ways that are both relevant to wage growth and non-compete enforcement. As such, the results shown here should be seen as merely suggestive.

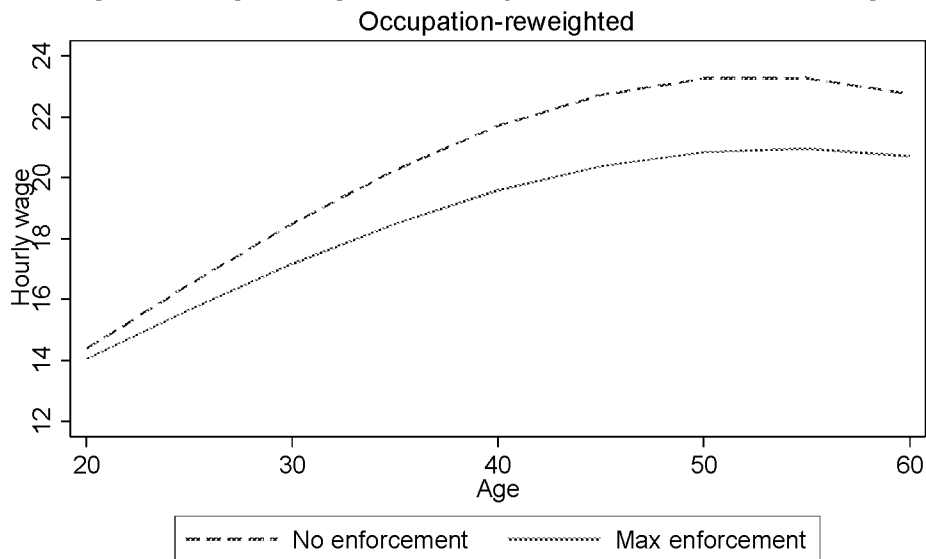
switching is generally associated with substantial wage increases, this increased difficulty of switching would reduce wage growth over time.³⁷

Figure 1. Age-Wage Profile by State Enforcement Regime



Source: 2014 Current Population Survey, Starr et al. (2015), and Treasury calculations. All variables, with the exception of enforcement index and age, are held constant at their means.

Figure 2. Age-Wage Profile by State Enforcement Regime



Source: 2014 Current Population Survey, Starr et al. (2015), and Treasury calculations. All variables, with the exception of enforcement index and age, are held constant at their means.

³⁷ See Topel and Ward (1992).

Non-compete enforcement and aggregate impacts. Thus far, we have discussed the effects of non-competes on individual workers. But non-compete enforcement may matter for cities, states, and regions in ways that cannot be fully understood at the individual level. Whether non-competes are beneficial or harmful for a single worker and a single firm, there are potential spillovers across workers and firms, particularly related to information.

In urban economics, regions are subject to so-called “agglomeration effects.” For instance, high-tech firms do not locate randomly, but tend to cluster in places like Silicon Valley. This clustering is due to a number of factors that include the availability of a large, deep pool of workers with relevant skills, a more competitive market of suppliers, and information spillovers across workers and firms. This last factor is important in connection with non-competes. When firms in a given industry are clustered, it becomes easier for their workers to share expertise and discoveries. While not always in the interest of a particular firm, this sharing can redound to the advantage of the larger economy, making the cluster an attractive destination for firms.

One important facilitator of this sharing is, unsurprisingly, the movement of workers across firms within industry. Employee departures impose costs on their firms, but yield benefits for destination firms and act to broadly disseminate improvements in technologies and best practices. Non-compete enforcement can stifle this mobility, thereby limiting the process that leads to agglomeration economies.

Many observers have suggested that Silicon Valley is a prime example of this phenomenon.³⁸ California, along with some other states, generally does not enforce non-compete agreements. It would be difficult to reach definitive conclusions about one instance of an industrial cluster, of course. One fact contradicting the hypothesis of free mobility is that high-tech firms in Silicon Valley have been alleged to collude to suppress wages and reduce “poaching.”³⁹ We do not know precisely how this behavior interacts with use of non-competes; in other words, the California firms may have been colluding as a substitute for using non-competes. Nevertheless, the Silicon Valley example highlights the importance of information sharing facilitated by worker mobility in some industrial clusters.

³⁸ For example, see Gilson (1999).

³⁹ See <http://www.wsj.com/articles/judge-rejects-settlement-in-silicon-valley-wage-case-1407528633>.

Singh and Marx look more broadly at informational spillovers and find that non-compete enforcement reduces their scope. Furthermore, using the Michigan natural experiment and cross-sectional data, Marx, Singh, and Fleming find that highly skilled workers tend to move from enforcing to non-enforcing states. This suggests that non-competes play a role in “brain drain,” potentially harming states that enforce non-competes more stringently.⁴⁰

Samila and Sorenson (2011) also examine the relationship between non-compete enforcement and regional employment and entrepreneurship. They find that more stringent enforcement is negatively related to both employment growth and entrepreneurship, consistent with results from Marx and coauthors.

⁴⁰ See Singh and Marx (2011) and Marx, Singh, and Fleming (2011). Note that this particular finding does not speak to whether strict non-compete enforcement is harmful to the nation as a whole.

V. Directions for Reform

Until recently, the lack of comprehensive data and analysis of non-competes made it difficult to evaluate the institution from a public policy perspective. However, recent research has underlined some important stylized facts that help to inform ideal policy and distinguish between various possible explanations for non-competes. First, non-competes are common in the labor market across educational, occupational, and income groups. Many workers who do not report possessing trade secrets are nonetheless covered by non-competes.⁴¹ Second, workers are often poorly informed about the existence and details of their non-competes, as well the relevant legal implications. Some employers appear to be exploiting this lack of understanding in ways that harm workers without producing corresponding benefits to society. Finally, while non-compete enforcement is associated with increased training for some workers, the details of this enforcement are important: strong “consideration” requirements can support training and wage growth while diminishing the likelihood that non-compete contracts result purely from inadequate worker knowledge.

The following are general reform recommendations related to the enforcement and use of non-compete contracts. They are not intended to be detailed or exhaustive. Nonetheless, these are promising avenues for state and/or federal policymakers to explore.

Increase transparency in the offering of non-competes.

Policymakers should act to inject transparency into the world of non-competes. To the extent that firms are simply misleading their prospective workers, non-competes are straightforwardly negative for employees. It is important to be precise about the forms that worker confusion can take. Some workers may simply not realize that they have signed a non-compete or fail to understand its ramifications. This sort of confusion could be addressed by a requirement that employers make the contracts, as well as their implications for future mobility, more salient for workers at the outset of an employment relationship. Relatedly, some workers who are aware of their non-compete contract may nonetheless be confused about its legal enforceability.

⁴¹ It is worth noting, however, that this is based on worker self-reports; employers may disagree.

Encourage employers to use enforceable non-compete contracts.

Many firms write non-compete contracts that contain unenforceable, overbroad provisions. Given the well-documented worker confusion about these contracts and the very low cost of writing an unenforceable contract, employers can exert a chilling effect on worker behavior even when their contracts are unenforceable. Conversely, states should explicitly specify the constraints on enforceability of non-compete contracts, where possible.

Require that firms provide “consideration” to workers bound by non-compete contracts in exchange for both signing and abiding by non-competes.

Some firms already provide severance payments to workers with non-competes.⁴² For instance, a worker who quits may receive 50 percent of her previous salary in exchange for abiding by the terms of the non-compete. This limits the harm to workers while ensuring that firms retain the ability to protect their interests with non-competes. Importantly, by requiring that firms incur a cost when requesting a non-compete, this policy preserves the most socially valuable non-compete agreements and discourages the least valuable, for which firms would not be willing to pay.

Conclusion

Non-competes are a central labor market institution, with nearly one fifth of all American workers currently bound by such a contract. Surprisingly, non-competes are widely distributed across education, occupation, and income groups. Understanding the consequences of this institution for workers and the broader economy is therefore of great importance, especially in light of its central role in determining workers’ prospects for wage growth and job mobility.

⁴² See <http://www.sec.gov/Archives/edgar/data/320187/000119312510161874/dex1023.htm>.

Though non-compete contracts can have important social benefits, principally related to the protection of trade secrets, a growing body of evidence suggests that they are frequently used in ways that are inimical to the interests of workers and the broader economy. Enhancing the transparency of non-competes, better aligning them with legitimate social purposes like protection of trade secrets, and instituting minimal worker protections can all help to ensure that non-compete contracts contribute to economic growth without unduly burdening workers.

Ryan Nunn in the Office of Economic Policy was the principal drafter of this report. Inquiries should be directed to the Office of Economic Policy at (202) 622-2200.

Appendix A

Modern interpretations of non-compete agreements are often said to have their origin in 15th and 16th century English common law and are best understood in the context of that period's economic structure. The guild economy largely comprised three types of workers: the apprentice, the journeyman, and the master craftsman. Custom required apprentices to train under master craftsmen for an extended period until graduating to the status of journeyman. Once a journeyman, the individual was free to work wherever he wished while he sought entrance into the inner circle of master craftsmen. Non-compete agreements likely originated in this context as journeymen replaced retiring master craftsmen by purchasing their businesses.⁴³ However, available case law suggests English courts tended to disfavor restraints on trade – especially restraints initiated by an employer.

The most cited example from this period comes from *The Dyer's Case* of 1414.⁴⁴ This case is perhaps the first known example of a contractual restraint of trade. A London practitioner prohibited his apprentice from pursuing his trade in the same city for six months following his apprenticeship. The court ruled against the covenant.⁴⁵ According to some commentators, the result produced two fundamental pillars of employment law.⁴⁶ The first was a policy in favor of retaining skilled labor in the public domain. The second pillar promoted the right of all individuals to seek a livelihood. These principles guided legal precedent for the next century.

Over time, some master craftsmen began to take on more apprentices than customary so as to employ a larger staff at low cost.⁴⁷ The consequence of this strategy was an influx of journeymen looking for ways to unseat master craftsmen. Some craftsmen addressed the increased levels of competition by requiring apprentices and journeymen to sign non-compete agreements.⁴⁸ The English Parliament brought attention to some of these practices in 1536 by authoring the Act for Avoiding of Extracting Taken upon Apprentices.⁴⁹ The law attempted to

⁴³ Harlan M. Blake, *Employee Agreements Not to Compete*, 73 *Harvard Law Review* 638 (1960).

⁴⁴ *The Dyer's Case*, Y.B. Mich. 2 Hen. 5, fol. 5, pl. 26 (1414).

⁴⁵ *Ibid.*

⁴⁶ Dan Messeloff, *Giving the Green Light to Silicon Alley Employees: No-Compete Agreements between Internet Companies and Employees under New York Law*, *Fordham Intellectual Property, Media and Entertainment Law Journal*, (vol. 11, issue 3, 2001), at 710-711. Much of this appendix benefits from this article.

⁴⁷ Blake, *supra* note 39, at 633.

⁴⁸ *Ibid.*

⁴⁹ Bland, Brown & Tawney, *English Economic History – Select Documents*, (1919), at 284-286.

restrain some of the practices of guild masters – including non-compete contracts. In 1563, the Statute of Artificers restricted the privileges of workers while also shifting power from guild masters to the evolving English state.⁵⁰ The law established national constraints on maximum wages and the length of apprenticeships.⁵¹

By the beginning of the 17th century, courts continued to disfavor employment restraints, whether in the form of time or place. An excerpt from *Colgate v. Bachelier* (1602) notes, “For as well as [employers] may restrain [employees] for one time, or one place, [they] may restrain [them] for longer times, and more places, which is against the benefit of the Common-wealth. . . . For he ought not be abridged of his Trade, and Living.”⁵² Others worried that non-compete covenants forced young men into “idleness”.⁵³ However, as a new economic system emerged, English courts began to rethink their position on non-compete covenants.

Mitchel v. Reynolds (1711) marked a distinct shift away from the practice of completely banning non-competes.⁵⁴ Reynolds, a baker, agreed to rent his bakery for five years. In return, Mitchel pledged Reynolds a bond worth 50 pounds on the condition that Reynolds would not resume his trade within St. Andrew Holborn Parish for 5 years. The latter failed to keep the agreement and Mitchel sued. Chief Justice Parker ruled in favor of the agreement.⁵⁵ He reasoned that while general restraints on trade were unlawful, as they benefited neither party, some partial restraints were reasonable.⁵⁶ Effectively, the ruling permitted individuals to enter agreements even if they restricted one’s ability to work in a particular location or for a certain period, as long as both parties and the affected communities benefited from the arrangement. However, employers were required to demonstrate the economic necessity of any such agreement.

⁵⁰ Donald Woodward, *The Background to the Statute of Artificers: The Genesis of Labour Policy, 1558-63*, *The Economic History Review* (vol. 33, no.1) 1980, at 32-44.

⁵¹ *Ibid.*

⁵² *Cro. Eliz.* 872, 78 English Report 1097, (Queen’s Bench 1602).

⁵³ *Case of Tailors of Ipswich*, 77 English Report 1218, 1219 (King’s Bench 1614).

⁵⁴ *Mitchel v. Reynolds*, 24 English Report 347 (Queen’s Bench 1711).

⁵⁵ Dan Messeloff, *Giving the Green Light to Silicon Alley Employees: No-Compete Agreements between Internet Companies and Employees under New York Law*, *Fordham Intellectual Property, Media and Entertainment Law Journal*, (vol. 11, issue 3, 2001), at 710-711.

⁵⁶ “General” restraints were defined as those with unlimited scope in either time or space, while “partial” restraints were those limited in both dimensions.

The economic significance of non-competes evolved as new technology accompanied the Industrial Revolution.⁵⁷ Once limited to local markets, companies began expanding into national and international markets, exposing themselves to new rivals.⁵⁸ Moreover, corporations were increasingly concerned with worker mobility. Leaving one's town no longer carried the same economic and physical risks. *Homer v. Ashford* (1825) describes the logic applied by English courts on matters of non-compete covenants:

A merchant or manufacture would soon find a rival in every one of his servants if he could not prevent them from using to his prejudice the knowledge they acquired in his employ. Engagements of this sort between masters and servants are not injurious restraints of trade, but securities necessary for those who engage in it. The effect of such contracts is to encourage rather than cramp the employment of capital in trade and the promotion of industry.⁵⁹

Some took the argument of the court to suggest that non-compete clauses were permissible in most circumstances. Six years later, the court clarified that while employers should have access to protection, *Mitchel's* test-of-reason still applied. In *Horner v. Graves* (1831), a dentist's assistant contracted to not practice independently within 100 miles of the original employer.⁶⁰ Soon after parting with his employer, the assistant broke the agreement, prompting the dentist to sue. In response, the court sided with the defendant, explaining that a reasonable restraint must also account for the interests of the public. From the public's perspective, the dentist had sought to withhold a valuable service within the 100 mile radius of his practice in order to protect himself. The court determined that the burden placed on the public was greater than the need to protect the interests of the previous employer and that the requirement was unreasonably broad.⁶¹

The intermittent reweighting of employer, worker, and public interests continued as the 19th century wore on. By 1841, although most English courts still rejected general restraints, some began to enforce them as businesses globalized.⁶² A trend toward pro-employer policy

⁵⁷ See Messeloff, *supra* note 42, at 712-713.

⁵⁸ Blake, *supra* note 39, at 638.

⁵⁹ *Homer v. Ashford*, 3 Bing. 322, 327 (1825).

⁶⁰ 7 Bing. 735, 131 English Report 284 (C.P. 1831).

⁶¹ *Ibid.* at 743.

⁶² Blake, *supra* note 39, at 624.

continued in 1853 when the Queen’s Bench ruled that the burden of showing unreasonableness rested on the employee rather than employer.⁶³ In 1875, the court ruled that while contracts must remain reasonable, a central value of the liberal economic philosophy permitted men of sound mind to enter arrangements as they saw fit.⁶⁴ Increasing emphasis on freedom of contract was evident in *Rousillon v. Rousillon* (1880), where the court allowed covenantal protection to extend beyond national borders. The court reasoned that if the contract was reasonable in scope at the negotiation, changing economic circumstances should not bar enforcement.^{65,66}

As English courts were moving toward pro-employer policies, American courts started developing their own body of common law. In 1851, *Lawrence v. Kidder*, a case before the New York Supreme Court, established a precedent that the state’s priority was to deter monopolies.^{67,68} The court reasoned that as far as possible, the state must ensure that all citizens be permitted to work.⁶⁹ As such, the court viewed agreements which barred individuals from practicing their occupations based on state or territory boundaries as unlawful.

A Pennsylvania court made an important distinction in 1866 between the sale of “handicraft” and the sale of “property”.⁷⁰ The Pennsylvania court deemed restrictions on property much more reasonable than restrictions on the use of an employee’s skills. This distinction laid the foundation for the landmark Supreme Court decision in *Oregon Steam Navigation Co. v. Winsor* (1874).⁷¹ The California Steam Navigation Company sold the Oregon Steam Navigation Company a boat under the condition that they would not operate the vessel within California for a period of ten years. The Oregon Steam Navigation Company subsequently sold it to Winsor, who at the time of sale was engaged in the navigation of water in Washington. The sale was subject to a condition (among others) that Winsor would not operate the boat in California for a period of ten years. The court upheld the condition, noting that there was no injury to the public.⁷²

⁶³ *Tallis v. Tallis*, 1 El. & B. 391, 118 English Report 482 (Queen’s Bench 1853).

⁶⁴ *Printing & Numerical Registering Co. v. Sampson*, L.R. 19 Eq. 462 (1875).

⁶⁵ 14 Ch. D. 351 (1880).

⁶⁶ *Blake*, *supra* note 39, at 641.

⁶⁷ 10 Barb. 641 (N.Y. Supreme Court 1851)

⁶⁸ *Blake*, *supra* note 39, at 644.

⁶⁹ *Ibid.*

⁷⁰ *Keeler v. Taylor*, 53 Pa. 467, 470 (1866).

⁷¹ *Messeloff*, *supra* note 42, at 720-721.

⁷² *Ibid.*

The New York Court of Appeals echoed the opinion of the Supreme Court in 1887 when it ruled in favor of a non-compete clause which restricted selling matches in the states of Nevada and Montana.⁷³ The court found that the condition was a “partial” restraint even though it covered the entire state of New York, while noting that the distinction between “general” and “partial” restraints, while still good law, was weakening.⁷⁴

Non-compete policies began diverging across states by the end of the 19th century. Notably, the California legislature rendered non-competes generally unenforceable.⁷⁵ Outside of legal opinions, the most influential American documents on contract law are the “Restatement of Contracts” of 1932 and its revision in 1979.⁷⁶ Though non-binding, these writings, published by the American Law Institute, codify case law. Both versions of the Restatement of Contracts state that restraints are unlawful if they unjustly benefit employers and impose undue hardship on the employee or public – reflecting the opinion in *Horner v. Graves*.⁷⁷ The second Restatement of Contracts protects the employee further by increasing the standard by which an employer must demonstrate legitimate need for non-compete protection.⁷⁸

⁷³ *Diamond Match v. Roeber* 106 N.Y. 473 (1887).

⁷⁴ Messeloff, *supra* note 42, at 722.

⁷⁵ Messeloff, *supra* note 42, at 714.

⁷⁶ *Ibid.* at 723-724.

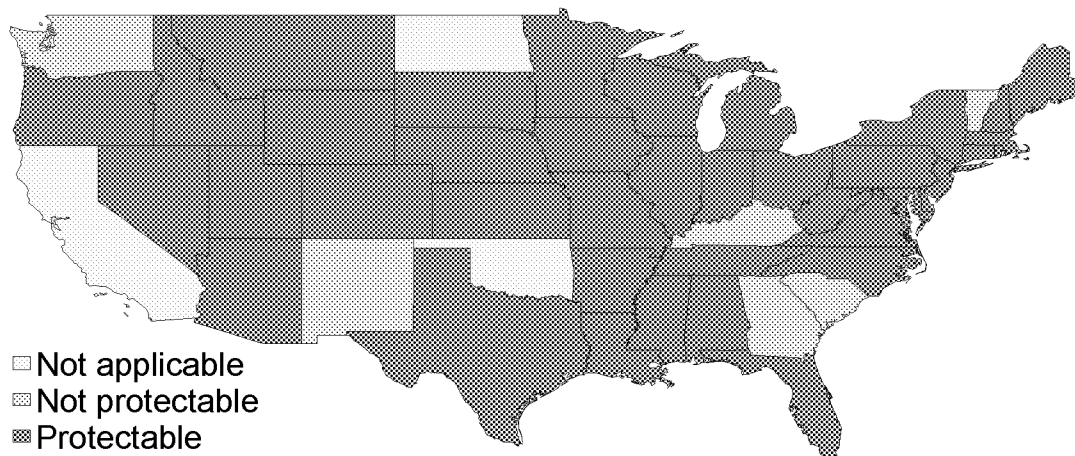
⁷⁷ *Ibid.*

⁷⁸ *Ibid.*

Appendix B

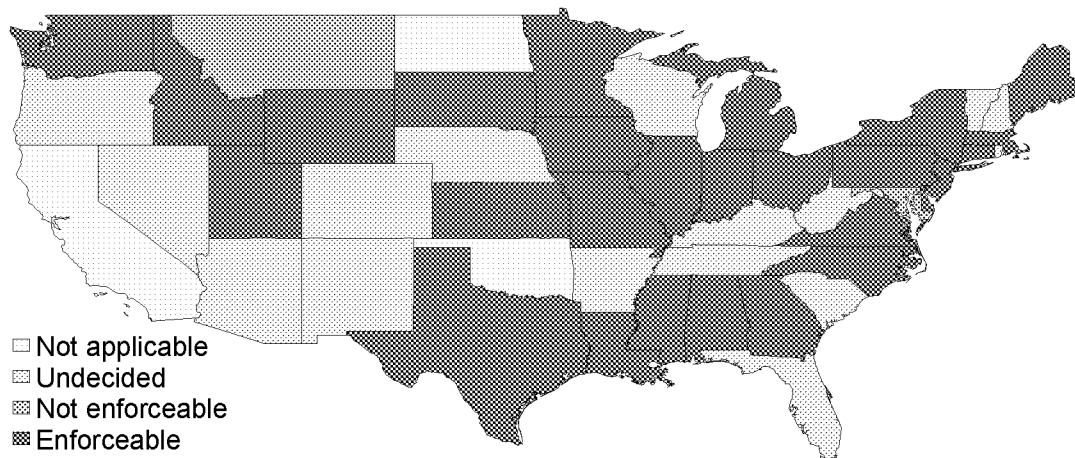
The following figures show some of the state heterogeneity in non-compete enforcement as of 2015. Note, however, that they reflect one particular expert's view of state law, and may elide distinctions relevant to some specific cases.⁷⁹

Trade Secrets as Legitimate Interests



Source: A State by State Survey of Employee Noncompetes, Beck Reed Riden

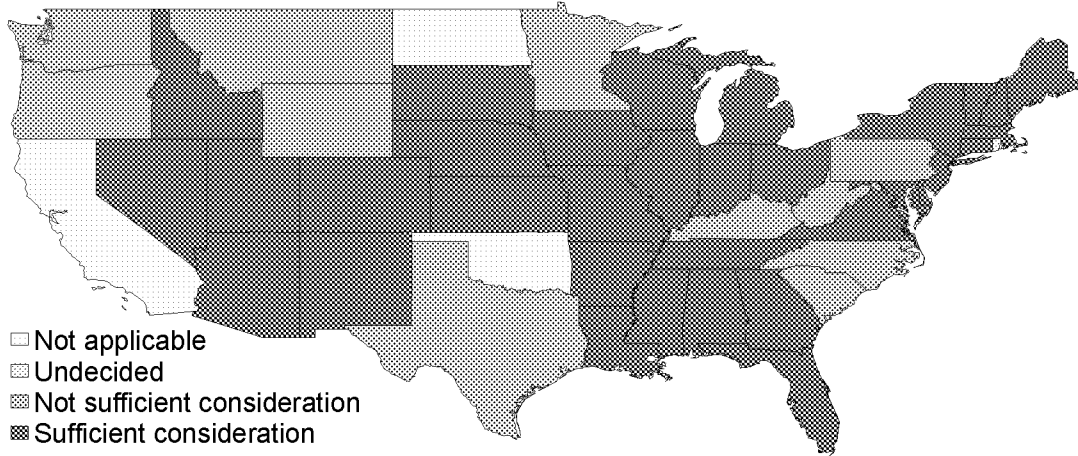
Enforceable Against Workers Fired Without Cause



Source: A State by State Survey of Employee Non-competes, Beck Reed Riden

⁷⁹ Hawaii considers trade secrets to be a protectable interest, is undecided on the question of enforcement against workers fired without cause, and regards continued employment as sufficient consideration. Alaska is identical, with the exception that it is undecided as to whether continued employment constitutes sufficient consideration.

Continued Employment as Sufficient Consideration



Source: A State by State Survey of Employee Noncompetes, Beck Reed Riden

Works Cited

Balasubramanian, Natarajan, Jin Woo Chang, Mariko Sakakibara, Jadadeesh Sivadasan, and Evan Starr. 2016. “Locked In? Noncompete Enforceability and the Mobility and Earnings of High Tech Workers.” Working paper.

Beck Reed Riden LLP. 2015. “Employee Noncompetes: A State by State Survey.”
<http://www.beckreedriden.com/50-state-noncompete-survey/>.

Becker, Gary S. 1962. *Investment in Human Capital: A Theoretical Analysis*. *Journal of Political Economy* 70 (5): 9–49.

Blake, Harlan M. 1960. “Employee Agreements Not to Compete.” *Harvard Law Review* 73: 625–691.

Bland, Brown, and Tawney. 1915. “English Economic History—Select Documents.” *The American Historical Review* 20 (3): 621–623.

Case of Tailors of Ipswich, (1614) 77 K.B. 1218, 1219 (E.R.).

Cro. Eliz. 872, (1602) 78 Q.B. 1097, (E.R.).

Diamond Match v. Roeber (1887) 106 N.Y. 473 (N.Y. Court of Appeals).

Feldman, Katuscak, and Kawano. 2016. “Taxpayer Confusion: Evidence from the Child Tax Credit.” Forthcoming in the *American Economic Review*.

Forbes. “Does Jimmy Johns Non-Compete Clause for Sandwich Makers have Legal Legs?”
<http://www.forbes.com/sites/clareoconnor/2014/10/15/does-jimmy-johns-non-compete-clause-for-sandwich-makers-have-legal-legs/>.

Gilson, Ronald J. 1999. “The Legal Infrastructure of High Technology Industrial Districts: Silicon Valley, Route 128, and Covenants Not to Compete.” *New York University Law Review* 74 (3): 575–629.

Hamermesh, D. S. 1995. “Labour Demand and the Source of Adjustment Costs.” *Economic Journal* 105 (430): 620–634.

Homer v. Ashford, (1825) 3 Bing. 322, 327 (E.R.).

Horner v. Graves (1831) 7 Bing. 735, 131 C.P. 284 (E.R.).

Kahneman, Daniel. 2003. “Maps of Bounded Rationality: Psychology for Behavioral Economics.” *The American Economic Review* 93 (5): 1449–1475.

Keeler v. Taylor, (1866) 53 Pa. 467, 470 (Penn. Supreme Court).

Lawrence v. Kidder (1851) 10 Barb. 641 (N.Y. Supreme Court).

Malsburger, Brian. 1998. "Covenants Not to Compete: A State by State Survey." Arlington: The Bureau of National Affairs.

Marx, Matt, and Lee Fleming. 2012. "Non-compete Agreements: Barriers to Entry... and Exit?" In *Innovation Policy and the Economy*, Volume 12, 39-64. Chicago: University of Chicago Press.

Marx, Matt, Deborah Strumsky and Lee Flemming. 2009. "Mobility, Skills, and the Michigan Non-Compete Experiment." *Management Science* 55 (6): 875–889.

Marx, Matt, Jasjit Singh and Lee Flemming. 2011. "Regional Disadvantage? Non-Compete Agreements and Brain Drain." *Research Policy* 44 (2): 394–404.

Marx, Matt. 2011. "The Firm Strikes Back: Non-compete Agreements and the Mobility of the Technical Professionals." *American Sociological Review* 76 (5): 695–712.

Mason v. Provident Clothing & Supply Co. (1913) A.C. 724 (H.R.) (Eng.) and Herbert Morris, Ltd. v. Saxelby, (1916) I A.C. 688 (H.L.) (Eng.).

Messeloff, Daniel. 2001. "Giving the Green Light to Silicon Alley Employees: No-Compete Agreements between Internet Companies and Employees under New York Law." *Fordham Intellectual Property, Media and Entertainment Law Journal* 11 (3): 710–746.

Mitchel v. Reynolds, (1711) 24 Q.B. 347 (E.R.).

Nordenfelt v. Maxim Nordenfelt Guns & Ammunition Co., (1984) A.C. 535, 536 (H.L.) (Eng.).

Oregon Steam Navigation Co. v. Winsor 87 S. Ct. (20 Wall.) 64 (1873).

Printing & Numerical Registering Co. v. Sampson, (1875) 19 Eq. 462 (L.R.).

Rousillon v. Rousillon (1880) 14 Ch. D. 351.

Salop, Joanne and Steven Salop. 1976. "Self-Selection and Turnover in the Labor Market." *The Quarterly Journal of Economics* 90 (4): 619-27.

Samila, Sampsa and Olav Sorenson. 2011. "Noncompete Covenants: Incentives to Innovate or Impediments to Growth." *Management Science* 57: 425:38.

Sichelman, Ted and Jonathan Barnett. 2015. "Revisiting Labor Mobility in Innovation Markets." Working paper.

Starr, Evan, Martin Ganco and Ben Campbell. 2016. "Redirect and Retain: Why and How Firms Capitalize on Noncompete Enforceability in Technical and Business Occupations." Working paper.

Starr, Evan, Norman Bishara and JJ Present. 2015. "Noncompetes in the U.S. Labor Force." Working paper.

Starr, Evan. 2015. "Consider This: Firm-Sponsored Training and the Enforceability of Covenants Not to Compete." Working paper.

Tallis v. Tallis, 1 El. & B. 391, (1853) 118 Q.B. 482 (E.R.).

The Dyer's Case, (1414) Y.B. Mich. 2 Hen. 5, fol. 5, pl. 26.

Thompson Hine LLP. "Georgia's Restrictive Covenants Act: Putting the "Non" in Noncompete." Lexology, <http://www.lexology.com/library/detail.aspx?g=aadbea62-9a31-4ae3-92dd-8198906c37f6>.

Topel, Robert H. and Michael P. Ward. 1992. "Job Mobility and the Careers of Young Men." *The Quarterly Journal of Economics* 107 (2): 439–179.

Wall Street Journal. "Judge Rejects Settlement in Silicon Valley Wage Case." <http://www.wsj.com/articles/judge-rejects-settlement-in-silicon-valley-wage-case-1407528633>.

Woodward, Donald. 1980. "The Background to the Statute of Artificers: The Genesis of Labour Policy, 1558-63." *The Economic History Review* 33 (1): 32–44.

Calendar No. 355

114TH CONGRESS } SENATE { REPORT
2d Session } { 114-220

DEFEND TRADE SECRETS ACT OF 2016

MARCH 7, 2016.—Ordered to be printed

Mr. GRASSLEY, from the Committee on the Judiciary,
submitted the following

R E P O R T

[To accompany S. 1890]

[Including cost estimate of the Congressional Budget Office]

The Committee on the Judiciary, to which was referred the bill (S. 1890), to amend chapter 90 of title 18, United States Code, to provide Federal jurisdiction for the theft of trade secrets, and for other purposes, having considered the same, reports favorably thereon, with an amendment, and recommends that the bill, as amended, do pass.

CONTENTS

	Page
I. Background and Purpose of the Defend Trade Secrets Act of 2016	1
II. History of the Bill and Committee Consideration	4
III. Section-by-Section Summary of the Bill	5
IV. Congressional Budget Office Cost Estimate	13
V. Regulatory Impact Evaluation	14
VI. Conclusion	14
VII. Changes to Existing Law Made by the Bill, as Reported	15

I. BACKGROUND AND PURPOSE OF THE DEFEND TRADE SECRETS ACT

Trade secrets are a form of intellectual property that allow for the legal protection of commercially valuable, proprietary information and make up an increasingly important part of American companies' intellectual property portfolios. Comprising all types of financial, scientific, technical, engineering, or other forms of information, trade secrets are an integral part of the operation, competitive advantage, and financial success of many U.S.-based companies.

The growing importance of trade secrets as a form of intellectual property makes their theft a particularly economically damaging

crime. In a recent report, the Commission on the Theft of American Intellectual Property estimated that annual losses to the American economy caused by trade secret theft are over \$300 billion, comparable to the current annual level of U.S. exports to Asia.¹ This same report found that trade secret theft has led to the loss of 2.1 million American jobs each year and that the illegal theft of intellectual property is undermining the means and incentive for entrepreneurs to innovate. This in turn is slowing the development of new inventions and industries that could raise the prosperity and quality of life for everyone.² In another study, PricewaterhouseCoopers LLP and the Center for Responsible Enterprise and Trade found that the annual cost of trade secret theft may be as high as \$480 billion.³

Protecting trade secrets has become increasingly difficult given ever-evolving technological advancements. Thieves are using increasingly sophisticated methods to steal trade secrets and the growing use of technology and cyberspace has made trade secret theft detection particularly difficult.⁴ The growing problem of trade secret theft has been acknowledged by industry, Congress,⁵ and the administration—with Attorney General Eric Holder stating during a White House conference in 2013, “There are only two categories of companies affected by trade-secret theft: those that know they’ve been compromised and those that don’t know yet.”⁶

Unlike other types of intellectual property, which are primarily protected under Federal law, trade secrets are primarily governed by State law. The Uniform Trade Secrets Act (UTSA) has been adopted (in its entirety or with some modifications) in 47 States and the District of Columbia.⁷ State laws that follow the UTSA provide trade secret owners with the ability to file civil lawsuits against a party who misappropriates trade secrets. Although the differences between State laws and the UTSA are generally relatively minor, they can prove case-dispositive: they may affect which party has the burden of establishing that a trade secret is not readily ascertainable, whether the owner has any rights against a party that innocently acquires a trade secret, the scope of information protectable as trade secret, and what measures are necessary to satisfy the requirement that the owner employ “rea-

¹The IP Commission, *The Report of the Commission on the Theft of American Intellectual Property* (May 2013), available at <http://www.ipcommission.org/report/IPCommissionReport052213.pdf>.

²Report of the Commission of the Theft of American Intellectual Property, at 1, 10 (May 2013), available at <http://www.ipcommission.org/report/IPCommissionReport052213.pdf>.

³Richard A. Hertling & Aaron Cooper, *Trade Secret Theft: The Need for a Federal Civil Remedy*, *The National Law Review* (June 25, 2014), available at <http://www.natlawreview.com/article/trade-secret-theft-need-federal-civil-remedy>.

⁴Brian T. Yeh, *Protection of Trade Secrets: Overview of Current Law and Legislation*, CRS Report No. R43714 (2014), available at <http://www.crs.gov/pages/Reports.aspx?PRODCODE=R43714&Source=search#fn12>.

⁵*Economic Espionage and Trade Secret Theft: Are Our Laws Adequate for Today's Threats? Hearing Before the Senate Judiciary Comm., Subcomm. on Crime and Terrorism*, 113th Cong. (2014); *Trade Secrets: Promoting American Innovation, Competitiveness and Market Access in Foreign Markets: Hearing Before the House Judiciary Comm.*, 113th Cong. (2014); *Protecting Trade Secrets: the Impact of Trade Secret Theft on American Competitiveness and Potential Solutions to Remedy this Harm: Hearing Before the Senate Judiciary Comm.*, 114th Cong. (2015).

⁶Sibohan Gorman and Jared A. Favole, *U.S. Ups Ante for Spying on Firms*, *Wall Street Journal* (Feb. 21, 2013) (reproducing a statement made by Attorney General Holder at a White House conference), available at <http://www.wsj.com/articles/SB10001424127887323549204578316413319639782>.

⁷Uniform Law Commission: *The National Conference of Commissioners on Uniform State Laws, Uniform Trade Secrets Act*, available at <http://www.uniformlaws.org/Act.aspx?title=Trade+Secrets+Act>.

sonable measures” to maintain secrecy of the information. At the Federal level, the Economic Espionage Act of 1996 (EEA), codified at 18 U.S.C. §§ 1831 *et seq.*, makes it a Federal criminal offense to misappropriate a trade secret that has an interstate or foreign nexus. The EEA, however, does not give trade secret owners a private right of action in Federal court. The Committee learned that, while fighting economic espionage and the theft of trade secrets is a top priority for Federal law enforcement,⁸ criminal enforcement remains a limited solution to stopping trade secret theft as the Federal Bureau of Investigation and Department of Justice are limited in the resources they can bring to bear.⁹

S. 1890 amends the Economic Espionage Act of 1996 to provide a Federal civil remedy for the misappropriation of trade secrets. A Federal cause of action will allow trade secret owners to protect their innovations by seeking redress in Federal court, bringing their rights into alignment with those long enjoyed by owners of other forms of intellectual property, including copyrights, patents, and trademarks. Modelling its definition of misappropriation on the UTSA, the bill provides for equitable remedies and the award of damages for the misappropriation of a trade secret. It also provides for expedited relief on an *ex parte* basis in the form of a seizure of property from the party accused of misappropriation, a remedy available under extraordinary circumstances where necessary to preserve evidence or prevent dissemination of a trade secret. The *ex parte* seizure provision is an important remedy for trade secret owners because it “enable[s] a trade secret owner under limited, controlled conditions, to proactively contain a theft before it progresses and the trade secret is lost.”¹⁰ For example, the damage caused by the large-scale 2006 theft of know-how related to DuPont’s innovative Kevlar product, in which there was significant destruction of evidence, would likely have been mitigated by the existence of a seizure remedy.

The bill balances the need for efficient recovery of a stolen trade secret with the rights of defendants and third-parties. Seizure orders must therefore minimize interruption to the business operations of third parties, protect the seized property from disclosure, and set a hearing date at the earliest possible time.

By improving trade secret protection, the Defend Trade Secrets Act of 2016 will incentivize future innovation while protecting and encouraging the creation of American jobs.

⁸*Economic Espionage and Trade Secret Theft: Are Our Laws Adequate for Today’s Threats?: Hearing Before the Senate Judiciary Comm., Subcomm. on Crime and Terrorism*, 113th Cong. (2014) (statement of Randall C. Coleman, Assistant Director, Counterintelligence Division, FBI), available at <https://www.fbi.gov/news/testimony/combating-economic-espionage-and-trade-secret-theft>.

⁹*Trade Secrets: Promoting and Protecting American Innovation, Competitiveness, and Market Access in Foreign Markets: Hearing Before the House Judiciary Comm., Subcomm. on Courts, Intellectual Property, and the Internet*, 113th Cong. (2014) (statement of Richard A. Hertling, Of Counsel, Covington & Burling, LLP, Protect Trade Secrets Coalition), available at <http://judiciary.house.gov/cache/files/5311b6c1-9a4f-49e5-a477-451a3ee228bf/113-97-88436.pdf>.

¹⁰*Protecting Trade Secrets: The Impact of Trade Secret Theft on American Competitiveness and Potential Solutions to Remedy this Harm: Hearing Before the S. Comm. on the Judiciary*, 114th Cong (2015), Statement of Karen Cochran, Associate General Counsel and Chief Intellectual Property Counsel, E.I. DuPont de Nemours & Co., at *4-5.

II. HISTORY OF THE BILL AND COMMITTEE CONSIDERATION

A. INTRODUCTION OF THE BILL

On July 29, 2015, Senators Hatch and Coons introduced the Defend Trade Secrets Act 2015. Senators Baldwin, Durbin, Flake, and Tillis were original cosponsors. The bill was referred to the Committee on the Judiciary. The bill built on previous legislation introduced in the Senate in two prior Congresses: S. 3389, the Protecting American Trade Secrets and Innovation Act of 2012, which was introduced by Senators Kohl, Coons, and Whitehouse in the 112th Congress and S. 2267, the Defend Trade Secrets Act of 2014, which was introduced by Senators Coons and Hatch in the 113th Congress.

B. COMMITTEE CONSIDERATION

On December 2, 2015, Senator Grassley chaired a Committee hearing on the subject of trade secret theft, entitled “Protecting Trade Secrets: the Impact of Trade Secret Theft on American Competitiveness and Potential Solutions to Remedy This Harm.” The hearing examined the importance of trade secrets to American companies, the adequacy of existing civil remedies, and the potential impact of a uniform Federal civil remedy for trade secret misappropriation. Testimony was received from Ms. Karen Cochran, Chief Intellectual Property Counsel, E.I. DuPont de Nemours and Company, Wilmington, DE;¹¹ Mr. Tom Beall, Vice President and Chief Intellectual Property Counsel, Corning Incorporated, Corning, NY;¹² Mr. James Pooley, Principal, James Pooley, PLC, Menlo Park, CA;¹³ Ms. Sharon Sandeen, Professor of Law, Hamline University School of Law, St. Paul, MN.¹⁴

The Committee’s Subcommittee on Crime and Terrorism previously held a hearing on the subject of trade secret theft during the 113th Congress on May 13, 2014, entitled, “Economic Espionage and Trade Secret Theft: Are Our Laws Adequate for Today’s Threats?” Testimony was received from Randall C. Coleman, Assistant Director, Counterintelligence Division, Federal Bureau of Investigation; Peter L. Hoffman, Vice President, Intellectual Property Management, The Boeing Company, Chicago, IL; Ms. Pamela Passman, President and Chief Executive Officer, Center for Responsible Enterprise and Trade, Washington, DC; Mr. Drew Greenblatt, President, Marlin Steel Wire Products, Baltimore, MD; and Mr. Douglas K. Norman, Vice President and General Patent Counsel, Eli Lilly and Company, Indianapolis, IN.

The Committee considered S. 1890 on January 28, 2016, in open session. Senators Hatch and Coons offered a substitute amendment

¹¹ *Protecting Trade Secrets: the Impact of Trade Secret Theft on American Competitiveness and Potential Solutions to Remedy This Harm, Before the Senate Comm. On the Judiciary*, 114th Cong. (2015) (statement of Ms. Karen Cochran, Chief Intellectual Property Counsel, E.I. DuPont de Nemours and Company), available at <http://www.judiciary.senate.gov/imo/media/doc/12-02-15%20Cochran%20Testimony.pdf>.

¹² *Id.* statement of Mr. Tom Beall, Vice President and Chief Intellectual Property Counsel, Corning Incorporated, available at <http://www.judiciary.senate.gov/imo/media/doc/12-02-15%20Beall%20Testimony.pdf>.

¹³ *Id.* statement of Mr. James Pooley, Principal, James Pooley, available at <http://www.judiciary.senate.gov/imo/media/doc/12-02-15%20Pooley%20Testimony.pdf>.

¹⁴ *Id.* statement of Ms. Sharon Sandeen, Professor of Law, Hamline University School of Law, available at <http://www.judiciary.senate.gov/imo/media/doc/12-02-15%20Sandeen%20Testimony.pdf>.

reflecting the input of several members of the Committee. The amendment provides that only the owner of a trade secret may bring a civil action for the secret's misappropriation, reduces the period of limitations from 5 to 3 years to align with the UTSA, and amends the definitions of "trade secret" and "improper means." The amendment also makes clear that ex parte seizures are only available in "extraordinary circumstances," and places other limitations on the breadth of seizures. The amendment further clarifies the appropriate scope of injunctions relating to employment to ensure that court orders are not contrary to applicable State laws. Finally, the amendment adds language expressing the sense of Congress that it is important to balance the interests of all parties when issuing an ex parte seizure, and instructing the Federal Judicial Center to develop best practices for the execution of seizures and the storage of seized information. The amendment was accepted by a voice vote without objection. Senators Leahy and Grassley offered an amendment to provide protection to whistleblowers who disclose trade secrets to law enforcement in confidence for the purpose of reporting or investigating a suspected violation of law. The amendment also immunizes the confidential disclosure of a trade secret in a lawsuit, including an anti-retaliation proceeding. The amendment was accepted by a voice vote without objection.

The Committee unanimously adopted both amendments by voice vote. The Committee then voted to report the Defend Trade Secrets Act of 2016, as amended, favorably to the Senate by voice vote.

III. SECTION-BY-SECTION SUMMARY OF THE BILL

Section 1. Short title

Section 1 provides that the short title of S. 1890 is the "Defend Trade Secrets Act of 2016."

Sec. 2. Federal jurisdiction for theft of trade secrets

Section 2(a) amends § 1836 of title 18 by striking subsection (b), which provides that the Federal district courts have exclusive jurisdiction over civil actions brought by the Attorney General for trade secret misappropriation. In its place, the new provision creates a Federal civil remedy for private parties for trade secret misappropriation.

In general

The new § 1836(b) in paragraph (1) authorizes the owner of a trade secret that is misappropriated to bring a civil action in Federal court if the trade secret that is related to a product or service used in, or intended for use in, interstate or foreign commerce. This jurisdictional nexus to interstate or foreign commerce is identical to the existing language required for Federal jurisdiction over the criminal theft of a trade secret under § 1832(a).

Civil seizure

The new § 1836(b) authorizes a Federal court to issue an order, in extraordinary circumstances and upon an ex parte application based on an affidavit or verified complaint, to provide for seizure of property necessary to preserve evidence or to prevent the propagation or dissemination of the trade secret. Ex parte seizures will

issue only when the prerequisites for the issuance of a seizure order are present. The issuance of a seizure order is limited to “extraordinary circumstances.” Subparagraph (A)(ii) lists requirements for issuing a seizure order. For example, this authority is not available if an injunction under existing rules of civil procedure would be sufficient. The ex parte seizure provision is expected to be used in instances in which a defendant is seeking to flee the country or planning to disclose the trade secret to a third party immediately or is otherwise not amenable to the enforcement of the court’s orders.

Subparagraph (A)(ii) contains numerous limitations, described below, and is not intended to affect the authority of the Federal courts to provide equitable relief and issue appropriate orders pursuant to Rule 65 of the Federal Rules of Civil Procedure, the All Writs Act (28 U.S.C. 1651), or any other authority, including the court’s inherent authority.

Subparagraph (A)(ii) of section 1836(b) specifies that that a court may not grant a seizure order unless it finds that it clearly appears from specific facts that (1) a temporary restraining order issued pursuant to Federal Rule of Civil Procedure 65(b) would be inadequate because the party to which the order would be issued would evade, avoid, or otherwise not comply with it; (2) immediate and irreparable injury will occur if the seizure is not ordered; (3) the harm to the applicant of denying the application outweighs the harm to the legitimate interests of the person against whom the seizure is ordered and substantially outweighs the harm to any third parties; (4) the applicant is likely to succeed in showing that the person against whom the seizure is ordered misappropriated the trade secret by improper means, or conspired to misappropriate the trade secret by improper means, and is in actual possession of it and any property to be seized; (5) the applicant describes with reasonable particularity the matter to be seized and, to the extent reasonable, identifies the location where the matter is to be seized; (6) the person against whom the seizure would be ordered, or those working in concert with that person, would destroy, move, hide, or otherwise make such matter inaccessible if the applicant were to provide that person notice; and (7) the applicant has not publicized the requested seizure.

Before granting an ex parte seizure order, it is the Committee’s expectation that courts will require applicants to describe the trade secret that would be the subject of the order with sufficient particularity so that the court may evaluate the request. The requirement of actual possession contained in clause (V) serves to protect third-parties from seizure. For instance, the operator of a server on which another party has stored a misappropriated trade secret, or an online intermediary such as an Internet service provider, would not be subject to seizure because their servers, and the data stored upon them, would not be in the actual possession of the defendant against whom seizure was ordered. While the court may not order a seizure against the third party under this provision, the court may decide to issue a third-party injunction preventing disclosure of the trade secret using its existing authority to provide equitable relief. The requirement relating to improper means is intended to prevent the seizure provision from being used against a party who may know it is in possession of a trade secret that was misappropriated.

priated, but did not use, or conspire to use, improper means to acquire such trade secret.¹⁵ Seizure of a trade secret that was stolen by one party and handed off to an accomplice is allowed under the clause.

Subparagraph (B) of new § 1836(b)(2) provides that a seizure order shall (i) set forth findings of fact and conclusions of law required for the order; (ii) provide for the narrowest seizure of property necessary to protect the trade secret, in a manner that minimizes any interruption of the business operations of third parties and, to the extent possible, does not interrupt the legitimate business operations of the person accused of misappropriating the trade secret; (iii) be accompanied by an order protecting the seized property from disclosure by prohibiting access by the applicant or the person against whom the order is directed, and prohibiting any copies of the seized property, until such parties have an opportunity to be heard in court (iv) provide guidance to law enforcement officials executing the seizure that clearly delineates the scope of their authority, including the hours during which the seizure may be executed and whether force may be used to access locked areas; (v) set a date for a hearing at the earliest possible time, and no later than seven days after the order has issued, unless parties involved consent to another date; and (vi) require the person obtaining the order to provide the security determined adequate by the court for payment of damages that person may be entitled to recover as a result of a wrongful or excessive seizure, or attempted seizure.

Subparagraph (C) of new § 1836(b)(2) requires a court, in issuing a seizure order, to take appropriate action to protect the target of the order from publicity, by or at the behest of the person obtaining the order, about such order and any seizure under such order.

Subparagraph (D) states that any materials seized pursuant to an order shall be taken into the custody of the court, which shall secure the material from physical and electronic access. In implementing this subparagraph, unless there is consent from the parties, the court should be careful to keep any electronic data or storage media secure and disconnected from any network or the Internet, thereby increasing security of the materials. The court shall take appropriate measures to protect the confidentiality of seized materials that are unrelated to the trade secret, unless the person against whom the order is entered consents to the disclosure of the material. The court may appoint a special master, bound by a non-disclosure agreement approved by the court, to locate and isolate all misappropriated trade secret information and facilitate the return of unrelated property and data to the person from whom the property was seized.

¹⁵The Act's protections against the misappropriation of trade secrets—and the remedies it provides against such misappropriation—are not intended to displace or restrict protections for members of the press recognized under the First Amendment. The Act should be applied consistently with the First Amendment and with the Supreme Court's decision in *Bartnicki v. Vopper*, 532 U.S. 514 (2001). That case held that the First Amendment protects members of the press against liability (including in civil actions) for disclosing information, even if the information was improperly or illegally obtained by another party in the first instance, particularly if the information relates to a matter of public concern. Indeed, *Bartnicki* recognized that the Supreme Court "has repeatedly held that 'if a newspaper lawfully obtains truthful information about a matter of public significance then state officials may not constitutionally punish publication of the information, absent a need . . . of the highest order.'" See *Bartnicki*, 532 U.S. at 528 (quoting *Smith v. Daily Mail Publ'g Co.*, 443 U.S. 97, 102 (1979)).

Subparagraph (E) requires service of the court's order and the submissions of the applicant on the party against whom the order is directed. The order must be carried out by a Federal law enforcement officer. The court may allow State and local law enforcement officials to participate but may not allow the applicant or its agents to participate. At the request of law enforcement, the court may appoint a neutral technical expert, bound by a nondisclosure agreement, to assist in the seizure if the court determines that the expert's participation would minimize the burden of the seizure.

Subparagraph (F) provides that the court shall hold a hearing at which the party who obtained the order shall have the burden to prove the facts supporting the findings of fact and conclusions of law necessary to prove the order. If a party fails to meet the burden for its proposed seizure, the seizure order shall be dissolved or modified appropriately. A party against whom the order has been issued, or any person harmed by the order, may move the court at any time to dissolve or modify the order.

Subparagraph (G) provides that a person who suffers damage by reason of a wrongful or excessive seizure has a cause of action against the applicant for the order under which the seizure was made, to recover damages, including punitive damages, and reasonable attorney's fees.

Subparagraph (H) provides that a party or other person who claims to have an interest in the subject matter seized may move to encrypt any seized materials.

Remedies

Paragraph (3) of new § 1836(b) provides the remedies for the misappropriation of a trade secret.

Subparagraph (A) specifies the equitable relief available and is drawn directly from § 2 of the Uniform Trade Secrets Act ("UTSA"), which forms the basis of trade secrets law in almost every State. Provided an order does not prevent a person from entering into an employment relationship or otherwise conflict with applicable State laws prohibiting restraints on trade, a court may grant an injunction to prevent any actual or threatened misappropriation. Any conditions placed by a court on employment must be based on evidence of threatened misappropriation, and not merely on information a person knows.¹⁶ These limitations on injunctive relief were included to protect employee mobility, as some members, including Senator Feinstein, voiced concern that the injunctive relief authorized under the bill could override state-law limitations that safeguard employee mobility and thus could be a substantial departure from existing law in those states. If determined appropriate, a court may require affirmative actions to be taken to protect the trade secret, and, in exceptional circumstances that render an injunction inequitable, may condition future use of the trade secret

¹⁶The Committee notes that courts interpreting State trade secret laws have reached different conclusions on the applicability of the inevitable disclosure doctrine. *Compare PepsiCo, Inc. v. Redmond*, 54 F.3d 1262, 1269 (7th Cir. 1995) ("[A] plaintiff may prove a claim of trade secret misappropriation by demonstrating that [the] defendant's new employment will inevitably lead him to rely on the plaintiff's trade secrets"), with *Whyte v. Schlage Lock Co.*, 125 Cal. Rptr. 2d 277, 281 (Ct. App. 2002) (rejecting explicitly the inevitable disclosure doctrine under California law).

upon payment of a reasonable royalty for no longer than the period of time for which such use would have been prohibited.

Section (3)(A)(i)(1)(I) reinforces the importance of employment mobility and contains some limitations on injunctive relief that may be ordered. However, as Senator Feinstein explained when the Committee considered this bill at its executive business meeting, if a State's trade secrets law authorizes additional remedies, those State-law remedies will still be available. Some courts have found, based on the information possessed by the employee alone, that an injunction may issue to enjoin a former employee from working in a job that would inevitably result in the improper use of trade secrets. Consistent with the overall intent of the Defense Trade Secret Act and, in particular, Section (2)(f), which provides that the bill does not "preempt any other provision of law," the remedies provided in Section (3)(A)(i)(1)(I) are intended to coexist with, and not to preempt, influence, or modify applicable State law governing when an injunction should issue in a trade secret misappropriation matter.

Subparagraph (B), drawn directly from § 3 of the UTSA, specifies the damage award that a court may issue. Specifically, it authorizes an award of damages for the actual loss and any unjust enrichment caused by the misappropriation of the trade secret, or, in lieu of damages measured by any other method, an award of a reasonable royalty. It is not the Committee's intent to encourage the use of reasonable royalties to resolve trade secret misappropriation. Rather, the Committee prefers other remedies that, first, halt the misappropriator's use and dissemination of the misappropriated trade secret and, second, make available appropriate damages.¹⁷

Subparagraph (C) authorizes an award of exemplary damages, not exceeding twice the compensatory damages awarded, if the trade secret is willfully and maliciously misappropriated. This provision is similar to § 3(b) of the UTSA.

Subparagraph (D) allows that attorney's fees may be awarded to the prevailing party if a claim of misappropriation is made in bad faith, there is willful and malicious misappropriation, or a motion to terminate an injunction is made or opposed in bad faith. This provision is modeled on § 4 of the UTSA.

Jurisdiction

Subsection (c) of new § 1836 provides that district courts of the United States shall have original jurisdiction of civil actions brought under the section. This is identical to current subsection (b).

Period of limitations

Subsection (d) of new § 1836 provides a three-year period of limitations in which to bring a claim under the section. This limitations period, which was reduced from five years during the Committee's markup, is now identical to the limitations period of the

¹⁷The Committee notes that courts interpreting the UTSA's analogous provision have held that the award of reasonable royalties is a remedy of last resort. See e.g., *Progressive Prod., Inc. v. Swartz*, 258 P.2d 969, 979–80 (Kan. 2011) (citing the comment to § 2 of the UTSA and explaining that an award of royalties is reserved for "special situation[s]," including "exceptional circumstances" in which an overriding public interest makes an injunction untenable).

UTSA, although a number of States have modified the limitations period in enacting the UTSA.

Definitions; Rule of construction; Conforming amendments

Section 2(b) of the Act amends § 1839 of title 18 to add three new definitions.

The intent of Section 2(b)(1)(A)—striking “the public” and inserting “another person who can obtain economic value from the disclosure or use of the information”—is to bring the Federal definition of a trade secret in conformity with the definition used in the Uniform Trade Secrets Act (“UTSA”). Both the Court of Appeals for the Seventh Circuit, in *United States v. Lange*, 312 F.3d 263, 267 (7th Cir. 2002), and the Court of Appeals for the Third Circuit, in *United States v. Hsu*, 155 F.3d 189, 196 (3d Cir. 1998), have identified this difference between the UTSA and the Federal definition of a trade secret as potentially meaningful. While other minor differences between the UTSA and Federal definition of a trade secret remain, the Committee does not intend for the definition of a trade secret to be meaningfully different from the scope of that definition as understood by courts in States that have adopted the UTSA.

First, “misappropriation” is defined identically in all relevant respects to the definition of misappropriation in § 1(2) of the UTSA. The Committee intentionally used this established definition to make clear that this Act is not intended to alter the balance of current trade secret law or alter specific court decisions.

Second, the subsection defines “improper means.” The definition contained in subparagraph (A) is identical to the definition in § 1(1) of the UTSA and includes theft, bribery, misrepresentation, breach, or inducement of a breach of a duty to maintain secrecy, or espionage through electronic or other means. Subparagraph (B) serves to clarify that reverse engineering and independent derivation of the trade secret do not constitute improper means.

Third, the subsection defines “Trademark Act of 1946,” commonly called the Lanham Act, which provides the basis for recovery by a party harmed by a wrongful or excessive seizure.

Subsection 2(c) of the Act ensures that nothing in the legislation is read to create a private right of action for conduct of a governmental entity or (following the amendment of 18 U.S.C. 1833 by section 7 of this Act) for disclosing trade secret information to the Government or in a court filing in accordance with new 18 U.S.C. 1833(b).

Subsection 2(d) of the Act is a conforming amendment that updates the title of section 1836 in the section heading and table of sections based on the changes made by this Act.

Subsection 2(e) provides that amendments made by section 2 of the Act shall apply to any misappropriation for which any act occurs on or after the date of enactment of the Act.

Subsection 2(f) of the Act clarifies that nothing in this Act modifies the rule of construction in § 1838 of title 18, and, as a result State trade secret laws are not preempted or affected by this Act. Further, nothing in this Act affects an otherwise lawful disclosure under the Freedom of Information Act.

Subsection 2(g) of the Act also specifies that the new civil remedy created by this Act is not to be construed as a law pertaining to intellectual property for purposes of any other Act of Congress.

Sec. 3. Trade secret theft enforcement

Subsection 3(a) of the Act amends § 1832(b) of title 18 by revising the maximum penalty for a violation under § 1832(a) to be the greater of \$5,000,000 or three times the value of the stolen trade secret to the organization, including expenses for research and design and other costs that the organization has thereby avoided.

Subsection 3(a) also amends § 1835 of title 18 by adding a new subsection (b), which provides that the court may not direct the disclosure of any material the owner asserts to be a trade secret unless the court allows the owner to file a submission under seal describing the interest of the owner in keeping the information confidential. The provision or disclosure of information relating to a trade secret to the United States or to the court in connection with a prosecution does not constitute waiver of trade secret protection unless the owner expressly consents to such waiver. The provision is also intended to ensure that in a prosecution for conspiracy related to the alleged theft of a trade secret, the actual trade secret itself is not subject to disclosure to the defense, because the actual secrecy of the information that is the object of the conspiracy is not relevant to the prosecution of a conspiracy charge.

Subsection 3(b) of the Act amends section 1961(1) of title 18 to include sections 1831 and 1832 relating to economic espionage and theft of trade secrets as predicate offenses for the Racketeer Influenced and Corrupt Organizations (RICO) Act.

Sec. 4. Report on theft of trade secrets occurring abroad

Section 4 of the Act requires, not later than one year after the date of enactment of this act and biannually thereafter, a report by the Attorney General, in consultation with the Intellectual Property Enforcement Coordinator, the Director of the United States Patent and Trademark Office, and the heads of other appropriate agencies, to the Committees on the Judiciary of the Senate and the House of Representatives, on:

- (1) the scope and breadth of trade secret theft from United States companies occurring outside the United States;
- (2) the extent to which trade secret theft occurring outside of the United States is sponsored by foreign governments, agents, or instrumentalities;
- (3) the threat posed by trade secret theft occurring outside of the United States;
- (4) the ability and limitations of trade secret owners to prevent the trade secret misappropriation of trade secrets outside of the United States, to enforce judgment against foreign entities for such theft, and to prevent imports based on theft of trade secrets overseas;
- (5) the trade secret protections afforded United States companies by each country that is a trading partner of the United States and specific information about enforcement efforts available and undertaken in each such country, including a list of specific countries where trade secret theft is a significant problem for United States companies;
- (6) instances of the Federal Government working with foreign countries to investigate, arrest, and prosecute entities and individuals involved in the theft of trade secrets outside of the United States;

(7) specific progress made under trade agreements and treaties, including any new remedies enacted by foreign countries, to protect United States companies from trade secret theft outside the United States; and

(8) recommendations for legislative and executive branch actions that may be undertaken to (A) reduce the threat of and economic impact caused by the theft of the trade secrets of United States companies occurring outside of the United States; (B) educate United States companies regarding threats to their trade secrets when taken outside of the United States; (C) provide assistance to United States companies to reduce the risk of loss of their trade secrets when taken outside of the United States; and (D) provide a mechanism for United States companies to confidentially or anonymously report the theft of trade secrets occurring outside the United States.

Sec. 5. Sense of Congress

Section 5 of the Act provides that it is the sense of Congress that trade secret theft occurs domestically and around the world, and that it is harmful to United States companies that own and depend on trade secrets. The Economic Espionage Act of 1996 protects trade secrets from theft under the criminal law. In enacting a civil remedy, it is important when seizing information to balance the need to prevent or remedy misappropriation with the need to avoid interrupting the legitimate interests of the party against whom a seizure is issued, and the business of third parties.

Sec. 6. Best practices

Section 6 directs the Federal Judicial Center to develop recommended best practices for seizure, storage, and security of information under this Act, within two years of the enactment. A copy of the recommendations and any updates made shall be provided to the Committees on the Judiciary of the Senate and the House of Representatives.

Sec. 7. Immunity from liability for confidential disclosure of a trade secret to the Government or in a court filing

Section 7 of the Act amends § 1833 of title 18 by adding a new subsection (b). The new § 1833(b)(1) provides for criminal and civil immunity for anyone who discloses a trade secret under two circumstances. Subparagraph (A) addresses disclosures in confidence to a Federal, State, or local government official, or to an attorney, for the purpose of reporting or investigating a suspected violation of the law. Subparagraph (B) applies to disclosure in a complaint or other document filed under seal in a judicial proceeding. The Committee stresses that this provision immunizes the act of disclosure in the limited circumstances set forth in the provision itself; it does not immunize acts that are otherwise prohibited by law, such as the unlawful access of material by unauthorized means.

Section 1833(b)(2) created by this Act provides that an individual who files a lawsuit against an employer for retaliation for reporting a suspected violation of the law may disclose a trade secret to an attorney for use in the proceeding, provided the individual files any document containing the trade secret under seal and does not disclose the trade secret other than pursuant to a court order.

Section 1833(b)(3) requires notice of the immunity in this subsection to be set forth in any employment contract that governs the use of trade secrets, although an employer may choose to provide such notice by reference to a policy document setting forth the employer's reporting policy for a suspected violation of the law that provides notice of the immunity. An employer may not be awarded exemplary damages or attorney's fees under this Act against an employee to whom such notice was not provided. The notice requirements apply to contracts entered into or updated after the date of enactment of this subsection.

Section 1833(b)(4) defines the term "employee" to include any individual performing work as a contractor or consultant.

Section 1833(b)(5) is a conforming amendment to update section 1838 of title 18 in the section heading and table of sections based on the changes made by this Act.

IV. CONGRESSIONAL BUDGET OFFICE COST ESTIMATE

The Committee sets forth, with respect to the bill, S. 1890, the following estimate and comparison prepared by the Director of the Congressional Budget Office under section 402 of the Congressional Budget Act of 1974:

FEBRUARY 25, 2016.

Hon. CHUCK GRASSLEY,
Chairman, Committee on the Judiciary,
U.S. Senate, Washington, DC.

DEAR MR. CHAIRMAN: The Congressional Budget Office has prepared the enclosed cost estimate for S. 1890, Defend Trade Secrets Act of 2016.

If you wish further details on this estimate, we will be pleased to provide them. The CBO staff contact is Marin Burnett.

Sincerely,

KEITH HALL.

Enclosure.

S. 1890—Defend Trade Secrets Act of 2016

S. 1890 would establish a federal remedy for individuals seeking relief from the misappropriation of trade secrets. Under the bill, an owner of a trade secret could file a civil action in a district court and the court could issue an order to seize any property necessary to preserve evidence for the civil action. The legislation would require information gathered or stored during a legal proceeding related to trade secrets to be secured to protect its confidentiality. The bill also would increase the fines that may be collected in the event of the theft of a trade secret. Finally, the legislation would require the Department of Justice (DOJ) and the Federal Judicial Center to submit periodic reports concerning the theft of trade secrets in the United States.

Based on information from DOJ and the Administrative Office of the U.S. Courts, CBO estimates that implementing S. 1890 would have no significant effect on the federal budget. Because enacting S. 1890 would affect direct spending and revenues, pay-as-you-go procedures apply. Specifically, the bill would affect civil court filing fees and potentially increase certain fines, which are recorded in

the budget as revenues. A portion of those revenues would be spent without further appropriation. On net, CBO estimates that the budgetary effect of those provisions would be negligible for each year and over the 2016–2026 period.

CBO estimates that enacting S. 1890 would not increase net direct spending or on-budget deficits in any of the four consecutive 10-year period beginning in 2027.

S. 1890 would preempt state laws that govern matters of individual liability when trade secrets are disclosed to governmental officials during the course of an investigation or legal proceeding. That preemption would be a mandate as defined in the Unfunded Mandates Reform Act (UMRA) because it would limit the authority of states to apply their own laws. However, CBO estimates that the preemption would not affect the budgets of state, local, or tribal governments because it would impose no duty on states that would result in additional spending or loss of revenue.

S. 1890 also would impose a private-sector mandate as defined in UMRA by extending civil and criminal liability protection to individuals who disclose trade secrets to government authorities during the course of an investigation or as a part of certain legal proceedings. By providing such liability protection, the bill would prevent entities from seeking compensation for damages from those individuals under trade secret laws. The cost of the mandate would be the forgone value of judgements and compensation for damages for such disclosures that entities would be awarded under a trade secrets claim. The bill would strengthen existing whistleblower protections to protect individuals from potential trade secret claims. The available literature suggests that few of those types of lawsuits have been brought against individuals under current law. Consequently, CBO estimates the cost of the mandate would probably fall below the annual threshold established in UMRA for private-sector mandates (\$154 million in 2016, adjusted annually for inflation).

The CBO staff contacts for this estimate are Marin Burnett (for federal costs), Rachel Austin (for intergovernmental mandates), and Logan Smith (for private-sector mandates). The estimate was approved by H. Samuel Papenfuss, Deputy Assistant Director for Budget Analysis.

V. REGULATORY IMPACT EVALUATION

In compliance with rule XXVI of the Standing Rules of the Senate, the Committee finds that no significant regulatory impact will result from the enactment of S. 1890.

VI. CONCLUSION

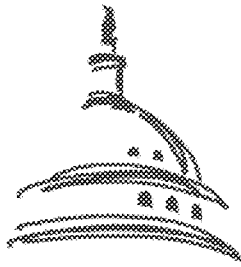
The Defend Trade Secrets Act, S. 1890, as amended, offers a needed update to Federal law to provide a Federal civil remedy for trade secret misappropriation. Carefully balanced to ensure an effective and efficient remedy for trade secret owners whose intellectual property has been stolen, the legislation is designed to avoid disruption of legitimate business, without preempting State law. This narrowly drawn legislation will provide a single, national standard for trade secret misappropriation with clear rules and predictability for everyone involved. Victims will be able to move

quickly to Federal court, with certainty of the rules, standards, and practices to stop trade secrets from winding up being disseminated and losing their value. As trade secret owners increasingly face threats from both at home and abroad, the bill equips them with the tools they need to effectively protect their intellectual property and ensures continued growth and innovation in the American economy.

VII. CHANGES TO EXISTING LAW MADE BY THE BILL, AS REPORTED

In compliance with paragraph 12 of rule XXVI of the Standing Rules of the Senate, the Committee finds that it is necessary to dispense with the requirement of paragraph 12 to expedite the business of the Senate.





**Congressional
Research Service**

Informing the legislative debate since 1914

Protection of Trade Secrets: Overview of Current Law and Legislation

Updated April 22, 2016

Congressional Research Service

<https://crsreports.congress.gov>

R43714

Summary

A trade secret is confidential, commercially valuable information that provides a company with a competitive advantage, such as customer lists, methods of production, marketing strategies, pricing information, and chemical formulae. (Well-known examples of trade secrets include the formula for Coca-Cola, the recipe for Kentucky Fried Chicken, and the algorithm used by Google's search engine.) To succeed in the global marketplace, U.S. firms depend upon their trade secrets, which increasingly are becoming their most valuable intangible assets.

However, U.S. companies annually suffer billions of dollars in losses due to the theft of their trade secrets by employees, corporate competitors, and even foreign governments. Stealing trade secrets has increasingly involved the use of cyberspace, advanced computer technologies, and mobile communication devices, thus making the theft relatively anonymous and difficult to detect. The Chinese and Russian governments have been particularly active and persistent perpetrators of economic espionage with respect to U.S. trade secrets and proprietary information.

In contrast to other types of intellectual property (trademarks, patents, and copyrights) that are governed primarily by federal law, trade secret protection is primarily a matter of state law. Thus, trade secret owners have more limited legal recourse when their rights are violated. State law provides trade secret owners with the power to file civil lawsuits against misappropriators. A federal criminal statute, the Economic Espionage Act (EEA), allows U.S. Attorneys to prosecute anyone who engages in "economic espionage" or the "theft of trade secrets." The EEA's "economic espionage" provision punishes those who misappropriate trade secrets with the intent or knowledge that the offense will benefit a foreign government, instrumentality, or agent. The EEA's "theft of trade secrets" prohibition is of more general application, involving the intentional theft of a trade secret related to a product or service used in or intended for use in interstate or foreign commerce, with the intent or knowledge that such action will injure the trade secret owner. In addition to criminal enforcement of the statute, the EEA authorizes the Attorney General to bring a civil action to obtain injunctive relief against any violation of the EEA.

However, because the U.S. Department of Justice and its Federal Bureau of Investigation have limited investigative and prosecutorial resources, as well as competing enforcement priorities, some observers assert that the federal government cannot adequately protect U.S. trade secrets from domestic and foreign threats. They have urged Congress to adopt a comprehensive, federal trade secret law in order to promote uniformity in trade secret law throughout the United States and to more effectively deal with trade secret theft that crosses state and international borders (a challenging problem for state courts to address). Among other things, they support the establishment of a federal civil cause of action for trade secret misappropriation, to allow U.S. companies to obtain monetary and injunctive relief when their trade secret assets are stolen.

In the 114th Congress, the Defend Trade Secrets Act (DTSA) (H.R. 3326 and S. 1890) has been introduced that would create a federal private right of action for trade secret misappropriation. S. 1890 was reported out of the Senate Judiciary Committee in late January 2016 with an amendment in the nature of a substitute. On April 4, 2016, the Senate passed S. 1890 by a vote of 87-0. On April 20, the House Judiciary Committee unanimously approved S. 1890.

Contents

Introduction	1
Background	2
Definition of a Trade Secret	2
Eligible Subject Matter and Acquisition of Rights	2
Duration of Protection	3
Misappropriation	3
Trade Secrets As a Form of Intellectual Property	4
Purpose of Trade Secret Law and Comparison to Patent Law	4
Historical Development of Trade Secret Law	5
Current Legal Landscape for Trade Secret Protection.....	6
State Law	6
Federal Law	7
Trade Secrets Act	7
Economic Espionage Act	7
International Law	11
The Growing Problem of Trade Secret Theft and Economic Espionage.....	13
Measuring Economic Loss	13
Types of Offenders	14
Domestic	14
Foreign	14
Enforcement of Trade Secret Rights.....	15
Litigation and Prosecution	15
Executive Branch Actions	16
Administration Strategy	16
Special 301	17
Free Trade Agreements (TPP and TTIP).....	17
Limitations of Current Law and Proposed Changes.....	18
In Support of a Federal Civil Cause of Action for Trade Secret Theft.....	18
In Opposition to a Federal Civil Trade Secret Remedy.....	20
Legislation in the 114 th Congress: The Defend Trade Secrets Act	22

Contacts

Author Information.....	24
-------------------------	----

Introduction¹

U.S. corporations face a “growing and persistent threat” by individuals, rival companies, and foreign governments that seek to steal some of their most valuable intangible assets—their trade secrets.² The tools, tactics, and methods used by such perpetrators vary widely but increasingly have involved the use of cyberspace and sophisticated technologies that “mak[e] it possible for malicious actors, whether they are corrupted insiders or foreign intelligence services (FIS), to quickly steal and transfer massive quantities of data while remaining anonymous and hard to detect.”³ As former Attorney General Eric Holder once opined,

There are only two categories of companies affected by trade-secret theft: those that know they’ve been compromised and those that don’t know yet. ... A hacker in China can acquire source code from a software company in Virginia without leaving his or her desk.⁴

Globalization has been cited as a major contributor to the increased incidents of trade secret theft:

In many ways, trade-secret theft is a foreseeable outgrowth of expanding international markets. When large multinational companies expand their overseas operations, they almost inevitably face challenges related to supply accountability and protection against such theft. Their foreign manufacturing operations and joint-venture partners require customer lists, internal standards, manufacturing processes, information on sources of goods, recipes, and production and sales strategies in order to carry out their operational responsibilities. Each new piece of information that is sent overseas opens a company’s supply chain and puts its valuable [intellectual property] at risk.⁵

There is significant congressional interest in reducing the problems of trade secret theft and economic espionage that U.S. businesses currently face, as demonstrated by significant legislative activity in the 114th Congress and hearings held in the 114th and 113th Congresses.⁷ This report provides an overview of existing federal, state, and international laws governing trade secret

¹ Portions of this report have been borrowed and adapted from CRS Report RL34109, *Intellectual Property Rights Violations: Federal Civil Remedies and Criminal Penalties Related to Copyrights, Trademarks, and Patents*, by Brian T. Yeh; CRS Report R41391, *The Role of Trade Secrets in Innovation Policy*, by John R. Thomas; and CRS Report R42681, *Stealing Trade Secrets and Economic Espionage: An Overview of 18 U.S.C. 1831 and 1832*, by Charles Doyle.

² Office of the National Counterintelligence Executive, *Foreign Spies Stealing US Economic Secrets in Cyberspace*, October 2011, at i, available at http://www.ncix.gov/publications/reports/fecie_all/Foreign_Economic_Collection_2011.pdf.

³ *Id.*

⁴ Siobhan Gorman and Jared A. Favole, *U.S. Ups Ante for Spying on Firms*, WALL ST. JOURNAL, February 21, 2013 (reproducing a statement made by Attorney General Holder at a White House conference).

⁵ *The Report of the Commission on the Theft of American Intellectual Property*, at 41 (May 2013), available at http://www.ipcommission.org/report/IP_Commission_Report_052213.pdf. This commission is a private, bipartisan initiative led by former U.S. Director of National Intelligence Dennis Blair and former U.S. Ambassador to China Jon Huntsman.

⁶ *Protecting Trade Secrets: the Impact of Trade Secret Theft on American Competitiveness and Potential Solutions to Remedy This Harm: Hearings Before the Senate Judiciary Comm.*, 114th Cong. 1st Sess. (2015).

⁷ *Cyber Espionage and the Theft of U.S. Intellectual Property and Technology: Hearings Before the House Energy & Commerce Comm., Subcomm. on Oversight and Investigations*, 113th Cong. 1st Sess. (2013); *Economic Espionage and Trade Secret Theft: Are Our Laws Adequate for Today’s Threats?: Hearings Before the Senate Judiciary Comm., Subcomm. on Crime and Terrorism*, 113th Cong. 2d Sess. (2014); *Trade Secrets: Promoting and Protecting American Innovation, Competitiveness and Market Access in Foreign Markets: Hearings Before the House Judiciary Comm., Subcomm. on Courts, Intellectual Property and Internet*, 113th Cong. 2d Sess. (2014).

protection, describes the limitations of these legal regimes, and reviews pending legislation, the Defend Trade Secrets Act (S. 1890), that is intended to address such deficiencies.

Background

Definition of a Trade Secret

U.S. trade secret law protects secret, valuable business information from theft and espionage. While it has been said that an “exact definition of a trade secret is not possible,”⁸ a trade secret generally consists of confidential, commercially valuable information.⁹ One U.S. federal court has described trade secrets as follows:

A trade secret is really just a piece of information (such as a customer list, or a method of production, or a secret formula for a soft drink) that the holder tries to keep secret by executing confidentiality agreements with employees and others and by hiding the information from outsiders by means of fences, safes, encryption, and other means of concealment, so that the only way the secret can be unmasked is by a breach of contract or a tort.¹⁰

Whether information qualifies as a “trade secret” under federal or state law is a question of fact that may be determined by a jury.¹¹ A jury may consider several factors in assessing whether certain material is a trade secret, including the following:

- the extent to which the information is known outside of the company;
- the extent to which it is known by employees and others involved in the company;
- the extent of measures taken by the company to guard the secrecy of the information;
- the value of the information to the company and to its competitors;
- the amount of effort or money expended by the company in developing the information; and
- the ease or difficulty with which the information could be properly acquired or duplicated by others.¹²

Eligible Subject Matter and Acquisition of Rights

The U.S. Supreme Court has explained that for subject matter to be protected as a trade secret, the material must meet minimal standards of novelty and inventiveness to avoid extending trade secret protection to matters of general or common knowledge in the industry in which it is used.¹³

⁸ Restatement (First) of Torts §757, comment b.

⁹ Uniform Trade Secrets Act §1(4).

¹⁰ *ConFold Pac. v. Polaris Indus.*, 433 F.3d 952, 959 (7th Cir. 2006) (citations omitted).

¹¹ 4-15 ROGER M. MILGRIM, *MILGRIM ON TRADE SECRETS* §15.01.

¹² Restatement (First) of Torts §757, comment b.

¹³ *Kewanee Oil Co. v. Bicron Corp.*, 416 U.S. 470, 476 (1974) (“[S]ome novelty will be required, if merely because that which does not possess novelty is usually known; secrecy, in the context of trade secrets, thus implies at least minimal novelty.”); *see also* *Ruckelshaus v. Monsanto Co.*, 467 U.S. 986, 1002 (1984) (“Information that is public knowledge or that is generally known in an industry cannot be a trade secret.”).

In addition, the Supreme Court has held that a person can have a property interest in a trade secret (protected by the Taking Clause of the Fifth Amendment), although “[b]ecause of the intangible nature of a trade secret, the extent of the property right therein is defined by the extent to which the owner of the secret protects his interest from disclosure to others.”¹⁴ Therefore, companies may acquire a protectable trade secret property right by putting into place reasonable measures to maintain the confidentiality of certain business information “that is sufficiently valuable ... to afford an actual or potential economic advantage over others.”¹⁵ This expansive standard means that trade secret protection could be available to a wide range of proprietary information and technologies that companies rely on to give them an economic advantage over their competitors, including customer lists, methods of production, marketing strategies, pricing information, and chemical formulae.

Duration of Protection

Trade secret protection may extend indefinitely, lasting as long as the subject matter of the trade secret is commercially valuable and is kept confidential.¹⁶ However, the trade secret status of information may be lost if the information is accidentally or intentionally disclosed by anyone.¹⁷ Once a trade secret has been exposed to the public, its protected character is lost and cannot later be retrieved.¹⁸ However, disclosures of trade secrets to third parties for certain limited reasons do *not* waive trade secret protections, so long as the trade secret owner took reasonable measures to maintain its secrecy before and during disclosure, such as requiring non-disclosure or confidentiality agreements from each recipient of confidential information.¹⁹

Misappropriation

Misappropriation of a trade secret is a tort that may occur in several ways. One is when an individual acquires the trade secret through improper means, such as theft, bribery, misrepresentation, or espionage.²⁰ Another is when the individual uses or discloses the trade secret through a breach of confidence. For example, an employee might switch jobs and then disclose his previous employer’s trade secrets in violation of a confidentiality agreement.²¹ Finally, a trade secret may be misappropriated if it is used or disclosed with knowledge that the trade secret had been acquired improperly or through mistake. A person who uses information that he knows to have been stolen by another is therefore also guilty of misappropriation.²²

It is not a violation of trade secret law for another party to independently develop the subject matter of a trade secret, or for a party to analyze publicly available products or information in order to discover the secret information.²³ In addition, “reverse engineering,” which involves

¹⁴ *Ruckelshaus*, 467 U.S. at 1002.

¹⁵ Restatement (Third) of Unfair Competition §39.

¹⁶ *United States v. Dubilier Condenser Corp.*, 289 U.S. 178, 186 (1933) (explaining that rather than seek patent protection, an inventor “may keep his invention secret and reap its fruits indefinitely.”).

¹⁷ See *Religious Tech. Ctr. v. Netcom On-Line Communication Servs.*, 923 F. Supp. 1231, 1256 (N.D. Cal. 1995).

¹⁸ *In re Remington Arms Co.*, 952 F.2d 1029, 1033 (8th Cir. 1991).

¹⁹ 1-1 ROGER MILGRIM, MILGRIM ON TRADE SECRETS §1.04.

²⁰ Restatement (Third) of Unfair Competition §40 (1994).

²¹ See Jennifer Brockett, *Protecting Intellectual Property During Layoffs*, 32 LOS ANGELES LAWYER (April 2009).

²² Restatement (Third) of Unfair Competition §40 (1994).

²³ *Id.* at §43.

“starting with the known product and working backward to divine the process which aided in its development or manufacture,” is not considered an improper means of acquiring the subject matter of another’s trade secret.²⁴

Misappropriation of a trade secret may be enjoined by a court and the defendant may also be liable for compensatory and punitive damages.²⁵

Trade Secrets As a Form of Intellectual Property

Intellectual property encompasses a broad range of intangible property, including the following four categories of subject matter: (1) original artistic and literary works of authorship, such as motion pictures, books, art, photographs, music, and sound recordings (protected by copyright law); (2) symbols, names, colors, sounds, and words that distinguish commercially offered goods and services (protected by trademark law); (3) inventions of processes, machines, manufactures, and compositions of matter that are useful, new, and nonobvious (protected by patent law); and (4) confidential and proprietary business information (protected by trade secrets law). Federal law grants certain exclusive rights to the owners of patents, trademarks, and copyrights and provides remedies in the event that those rights are violated (an act referred to as an infringement).²⁶

Owners of these three types of intellectual property may enforce their rights by bringing a lawsuit against an alleged infringer in federal court. The U.S. Department of Justice may also criminally prosecute particularly egregious violators of the copyright and trademark laws²⁷ in order to impose greater punishment and possibly deter other would-be violators. (The Patent Act only provides civil remedies in the event of patent infringement.²⁸)

In contrast to the other three types of intellectual property that are governed primarily by federal law, trade secrets are primarily governed under state law,²⁹ and thus owners of trade secrets have more limited legal recourse when their rights are violated by others. State law provides trade secret owners with the power to file civil lawsuits against those who misappropriate trade secrets. Federal law allows U.S. Attorneys to prosecute such offenders but does not currently give trade secret owners a private right of action in federal court against parties that have engaged in trade secret theft.

Purpose of Trade Secret Law and Comparison to Patent Law

Trade secret law serves as the primary alternative to the patent system,³⁰ granting inventors proprietary rights to particular technologies, processes, designs, or formula that may not be able to satisfy the rigorous statutory standards for patentability. Companies may choose to maintain an invention as a trade secret rather than obtain a patent because their trade secret rights are not

²⁴ *Kewanee Oil Co.*, 416 U.S. at 476.

²⁵ Restatement (Third) of Unfair Competition §§44, 45.

²⁶ For a comprehensive description, see CRS Report RL34109, *Intellectual Property Rights Violations: Federal Civil Remedies and Criminal Penalties Related to Copyrights, Trademarks, and Patents*, by Brian T. Yeh.

²⁷ For copyright, 17 U.S.C. §506, 18 U.S.C. §2319; for trademark, 18 U.S.C. §2320.

²⁸ 35 U.S.C. §281.

²⁹ The U.S. Supreme Court in *Kewanee Oil Co. v. Bicron Corp.*, 416 U.S. 470 (1974), held that state trade secret laws are not preempted by either the Patent Clause of the U.S. Constitution (Article I, §8, cl. 8) or the federal patent statute (35 U.S.C. §§101 et seq.) Although both trade secret law and patent law protect certain kinds of information, the two fields of law are distinct. For a detailed comparison of patent law and trade secret law, see CRS Report R41391, *The Role of Trade Secrets in Innovation Policy*, by John R. Thomas.

³⁰ ROGER E. SCHECHTER & JOHN R. THOMAS, *INTELLECTUAL PROPERTY: THE LAW OF COPYRIGHTS, PATENTS AND TRADEMARKS*, §24.

restricted to a limited number of years—unlike patent protection, which lasts less than 20 years and upon expiration, thrusts the invention into the public domain. In addition, trade secret protection is far easier, quicker, and cheaper to obtain (immediately receiving legal protection upon a company taking reasonable efforts to maintain the secrecy of valuable business information), compared to the complicated, lengthy, and expensive process of acquiring a patent, which can take several years and requires the involvement of a federal government agency, the U.S. Patent & Trademark Office. However, obtaining patent protection may be more appropriate in certain instances, such as when a technology is difficult to maintain as a secret because competitors could easily reverse-engineer or independently discover it.

The U.S. Supreme Court has explained that the purpose of trade secret law is to provide companies with incentives to innovate and develop valuable information that may not be patentable:

Trade secret law will encourage invention in areas where patent law does not reach, and will prompt the independent innovator to proceed with the discovery and exploitation of his invention. Competition is fostered and the public is not deprived of the use of valuable, if not quite patentable, invention.³¹

In addition, by establishing legal remedies for trade secret misappropriation, trade secret law deters individuals who “have as their sole purpose and effect the redistribution of wealth from one firm to another.”³²

Historical Development of Trade Secret Law

Unlike other forms of intellectual property that can trace their origins back several hundreds of years, trade secret law is a creation of state court opinions from the middle of the 19th century. As noted by one legal scholar, the principles of trade secret law

evolved out of a series of related common law torts: breach of confidence, breach of confidential relationship, common law misappropriation, unfair competition, unjust enrichment, and torts related to trespass or unauthorized access to a plaintiff’s property. It also evolved out of a series of legal rules—contract and common law—governing the employment relationship.³³

In 1939, the American Law Institute (ALI), a group of lawyers, judges, and legal scholars, published a treatise titled the “Restatement of Torts,” which was an effort to provide a “clear formulation[]of common law and its statutory elements or variations and reflect the law as it presently stands or might plausibly be stated by a court.”³⁴ The Restatement of Torts included two sections dealing with the law of trade secrets. Section 757 explained the subject matter of trade secrets, while Section 758 spelled out the elements of a trade secret misappropriation cause of action. The ALI later addressed trade secrets in sections 39-45 of its 1993 “Restatement (Third) of Unfair Competition.”

In addition, the National Conference of Commissioners on Uniform State Law (NCCUSL) issued the Uniform Trade Secrets Act (UTSA) in 1979, which represents “the first comprehensive effort

³¹ *Kewanee Oil Co.*, 416 U.S. at 484-85.

³² *Rockwell Graphic Systems, Inc. v. DEV Industries, Inc.*, 925 F.2d 174, 178 (7th Cir. 1991).

³³ Mark A. Lemley, *The Surprising Virtues of Treating Trade Secrets as IP Rights*, 61 *STANFORD L. REV.* 311, 316 (2008).

³⁴ ALI, *Publications Catalog FAQ*, at <http://www.ali.org/index.cfm?fuseaction=publications.faq>.

to codify the law of trade secrets protection, incorporating the major common law principles while filling gaps left by the courts.³⁵ The NCCUSL consists of a group of academics, attorneys, and judges who draft statutes addressing a variety of issues, and then propose that each state enact them.³⁶ However, the NCCUSL lacks direct legislative authority itself. Its uniform acts become law only to the extent that state legislatures choose to adopt them.

The federal government did not take steps to provide national trade secret protection until the mid-1990s, when Congress enacted the Economic Espionage Act of 1996. This federal criminal law is described in detail in the following section.

Current Legal Landscape for Trade Secret Protection

State Law

As noted in the section above, trade secrets primarily receive protection from misappropriation under state law. Individuals or corporations may seek civil damages in state courts by pursuing a common law tort action for misappropriation or through a specific state statute. The Uniform Trade Secrets Act (UTSA) codifies the basic principles of common law trade secret protection and has been adopted by 47 states and the District of Columbia,³⁷ although many state legislatures made some changes to the original model text before enacting it. These state laws provide definitions for the key terms “trade secret,” “misappropriation,” and “improper means,”³⁸ and specify various forms of injunctive and monetary relief (including compensatory damages, punitive damages, and attorney’s fees) in a civil action for misappropriation of a trade secret.³⁹ A few states even recognize the theft of trade secrets as a prosecutable crime.⁴⁰

However, according to a March 2016 Senate Judiciary Committee report, state law variations from the UTSA have led to different procedural and substantive standards being applied by state courts in trade secret cases:

Although the differences between State laws and the UTSA are generally relatively minor, they can prove case-dispositive: they may affect which party has the burden of establishing that a trade secret is not readily ascertainable, whether the owner has any rights against a party that innocently acquires a trade secret, the scope of information protectable as a trade secret, and what measures are necessary to satisfy the requirement that the owner employ “reasonable measures” to maintain secrecy of the information.⁴¹

³⁵ NCCUSL, *Why States Should Adopt UTSA*, at <http://www.uniformlaws.org/Narrative.aspx?title=Why%20States%20Should%20Adopt%20UTSA>.

³⁶ For more information about the NCCUSL, see <http://www.uniformlaws.org/>.

³⁷ Only New York, Massachusetts and North Carolina have not enacted the UTSA, though they offer protection through a distinct statute or the common law.

³⁸ Uniform Trade Secrets Act §1.

³⁹ Restatement (Third) of Unfair Competition §§44, 45 (1994).

⁴⁰ For example, California provides that anyone who acquires, discloses, or uses trade secrets without authorization shall be punished by imprisonment of up to one year in a county jail, by a fine of up to \$5,000, or by both penalties. CAL. PENAL CODE §499c. In Texas, the knowing theft of a trade secret carries a criminal sentence of at least two years imprisonment (up to a maximum of 10 years) and a fine of up to \$10,000. TEX. PENAL CODE §31.05. See also N.J. STAT. ANN. §2C:20-1; N.Y. PENAL LAW §165.07.

⁴¹ S.Rept. 114-220, at 2-3.

Federal Law

Trade Secrets Act

Before 1996, arguably the most significant federal legislation regarding trade secrets was the Trade Secrets Act.⁴² This statute, enacted in 1948, is actually of narrow applicability. It forbids federal government employees and government contractors from making an unauthorized disclosure of confidential government information, including trade secrets. The sanctions for violating this criminal offense are removal from office or employment, and a fine and/or imprisonment of not more than one year. The law does not apply to state or local government actors or to private sector employees.

Economic Espionage Act

In 1996, Congress enacted a far broader piece of legislation pertaining to trade secrets, the Economic Espionage Act of 1996 (EEA).⁴³ The legislative history of the EEA reveals the congressional concerns over growing international and domestic economic espionage against U.S. businesses that prompted the establishment of a more comprehensive, federal scheme protecting trade secrets:

American companies and the U.S. Government spend billions on research and development. The benefits reaped from these expenditures can easily come to nothing, however, if a competitor can simply steal the trade secrets without expending the development costs. ... For years now, there has been mounting evidence that many foreign nations and their corporations have been seeking to gain competitive advantage by stealing the trade secrets, the intangible intellectual property of inventors in this country. ... [S]ince the end of the cold war, foreign nations have increasingly put their espionage resources to work trying to steal American economic secrets.⁴⁴

The EEA defines two separate criminal offenses: (1) theft of a trade secret for the benefit of a foreign entity (economic espionage, 18 U.S.C. Section 1831), and (2) trade secret theft intended to confer an economic benefit to another party (theft of trade secrets, 18 U.S.C. Section 1832).⁴⁵ As a threshold matter, to trigger an action under either provision of the EEA, the information must qualify as a trade secret. The EEA expansively defines a “trade secret” to encompass

[A]ll forms and types of financial, business, scientific, technical, economic, or engineering information, including patterns, plans, compilations, program devices, formulas, designs, prototypes, methods, techniques, processes, procedures, programs, or codes, whether tangible or intangible, and whether or how stored, compiled, or memorialized physically, electronically, graphically, photographically, or in writing if—

- a) the owner thereof has taken reasonable measures to keep such information secret; and
- b) the information derives independent economic value, actual or potential, from not being generally known to, and not being readily ascertainable through proper means by, the public.⁴⁶

⁴² 18 U.S.C. §1905.

⁴³ P.L. 104-294, 110 Stat. 3488 (1996).

⁴⁴ 142 CONG. REC. S12207, S12208 (daily ed. October 2, 1996) (statement of Sen. Specter).

⁴⁵ For a comprehensive description and analysis of all the statutory elements of the EEA, see CRS Report R42681, *Stealing Trade Secrets and Economic Espionage: An Overview of 18 U.S.C. 1831 and 1832*, by Charles Doyle.

⁴⁶ 18 U.S.C. §1839(3). This definition is substantially similar to that used by the UTSA, although it is broader in

Economic Espionage

The EEA's "economic espionage" provision, 18 U.S.C. Section 1831, punishes those who misappropriate, or attempt or conspire to misappropriate, trade secrets with the intent or knowledge that the offense will benefit a foreign government, instrumentality, or agent.⁴⁷ Such misappropriation must have been committed "knowingly"; in other words, the individual must have known that the information taken was valuable to its owner and that its owner had taken steps to keep it confidential.⁴⁸

According to the legislative history of the EEA, the "benefit" derived from a foreign espionage effort includes not only an economic benefit, but also "reputational, strategic, or tactical benefit."⁴⁹ A "foreign instrumentality" includes any "entity that is substantially owned, controlled, sponsored, commanded, managed, or dominated by a foreign government."⁵⁰ Therefore, a foreign corporation that engages in espionage without any evidence of sponsorship or control from a foreign government may not be subjected to a Section 1831 prosecution. However, an individual or organization that engages in theft of trade secrets, although not intending to benefit a foreign entity, could be liable for violating the more general criminal trade secrets provision contained in Section 1832, described in the section below.

Theft of Trade Secrets

The EEA's "theft of trade secrets" prohibition, 18 U.S.C. Section 1832, is of more general application. The principal elements of an EEA claim for theft of trade secrets are (1) the intentional and/or knowing theft, appropriation, destruction, alteration, or duplication of (2) a trade secret related to a product or service used in or intended for use in interstate or foreign commerce (3) with intent to convert the trade secret and (4) intent or knowledge that such action will injure the owner.⁵¹

Scrutiny of these additional elements reveals several fundamental differences between Sections 1832 and 1831. First, Section 1832 does not require that the offense benefit or intend to benefit a foreign entity; it is a law of general applicability. Section 1832 also requires that the theft *economically* benefit someone other than the trade secret owner, whereas Section 1831, the foreign economic espionage provision, more broadly encompasses misappropriation for any purpose, including non-economic benefits such as "reputational, strategic, or tactical benefit[s]."⁵² Establishing that the offender intended to cause injury to the trade secret owner "does not require

coverage. For a comparison of the language of the EEA and UTSA, see James H.A. Pooley et al., *Understanding the Economic Espionage Act of 1996*, 5 TEX. INTELL. PROP. L.J. 177, 188-197 (1997).

⁴⁷ 18 U.S.C. §1831.

⁴⁸ The legislative history of the EEA opined that this mens rea element of the offense would not be too difficult for government prosecutors to establish: "Most companies go to considerable pains to protect their trade secrets. Documents are marked proprietary; security measures put in place; and employees often sign confidentiality agreements to ensure that the theft of intangible information is prohibited in the same way that the theft of physical items are protected." 142 CONG. REC. S12213 (daily ed. October 2, 1996) (Managers' Statement for H.R. 3723, The Economic Espionage Bill).

⁴⁹ H.R. Rep. No. 104-788, at 11 (1996).

⁵⁰ 18 U.S.C. §1839(1).

⁵¹ 18 U.S.C. §1832.

⁵² H.R. Rep. No. 104-788, at 11 (1996).

the government to prove malice or evil intent, but merely that the actor knew or was aware to a practical certainty that his conduct would cause some disadvantage to the rightful owner.”⁵³

In 2014, an FBI assistant director testified before Congress about the logistical difficulties of bringing a prosecution under Section 1831 compared to Section 1832:

Often, the greatest challenge in prosecuting economic espionage, as opposed to trade secret theft, is being able to prove that the theft was intended to benefit a foreign government or foreign instrumentality. The beneficiary of the stolen trade secrets may be traced to an overseas entity, but obtaining evidence that proves the entity’s relationship with a foreign government can be difficult. The decision to pursue these cases under Section 1832 (theft of trade secrets) instead of Section 1831 (economic espionage) may depend upon the availability of foreign evidence and witnesses, diplomatic concerns, and the presence of classified or sensitive information required to prove the foreign nexus element.⁵⁴

Authorized Penalties Under the EEA

The EEA authorizes substantial criminal fines and imprisonment penalties for economic espionage and theft of trade secrets. For economic espionage, the maximum penalties increase to \$5 million for individuals and imprisonment of 15 years;⁵⁵ in the case of corporations that are found guilty of this offense, the applicable maximum fine is the greater of (a) \$10 million or (b) three times the value of the stolen trade secret.⁵⁶ Theft of trade secrets for commercial advantage is punishable by a fine of up to \$250,000 for individuals as well as imprisonment of up to 10 years, whereas organizations can be fined up to \$5 million.⁵⁷ The EEA also authorizes the criminal or civil forfeiture of “any property used, or intended to be used ... to commit or facilitate” an EEA violation as well as “any property constituting, or derived from, any proceeds obtained directly or indirectly as a result of” an EEA offense.⁵⁸ Offenders must also pay victims of trade secret theft restitution.⁵⁹

In addition, during any prosecution or proceeding under the EEA, federal district courts are required to enter protective orders, or to take other measures, “as may be necessary and appropriate to preserve the confidentiality of trade secrets, consistent with the requirements of the Federal Rules of Criminal and Civil Procedure, the Federal Rules of Evidence, and all other applicable laws.”⁶⁰ The legislative history of the EEA reveals the congressional interest in ensuring that courts use protective orders to guard against trade secret disclosures:

We have been deeply concerned about the efforts taken by courts to protect the confidentiality of a trade secret. It is important that in the early stages of a prosecution the issue whether material is a trade secret not be litigated. Rather, courts should, when entering these orders, always assume that the material at issue is in fact a trade secret.⁶¹

⁵³ *Id.* at 11-12.

⁵⁴ *Economic Espionage and Trade Secret Theft: Are Our Laws Adequate for Today’s Threats?: Hearings Before the Senate Judiciary Comm., Subcomm. on Crime and Terrorism*, 113th Cong. 2d Sess. (2014) (statement of Randall C. Coleman, Assistant Director, Counterintelligence Division, FBI).

⁵⁵ 18 U.S.C. §1831.

⁵⁶ 18 U.S.C. §1831.

⁵⁷ 18 U.S.C. §1832.

⁵⁸ 18 U.S.C. §§1834; 2323.

⁵⁹ *Id.*

⁶⁰ 18 U.S.C. §1835.

⁶¹ 142 CONG. REC. S12213 (daily ed. October 2, 1996) (Managers’ Statement for H.R. 3723, The Economic Espionage

The EEA also allows the Attorney General to bring a civil action to obtain “appropriate injunctive relief” against any violation of the EEA provisions regarding the protection of trade secrets.⁶² However, the EEA does not provide victims of trade secret theft with a private civil cause of action.⁶³

Extraterritorial Application of the EEA

Trade secret violations that occur both domestically and outside the United States may be subject to criminal prosecution by the federal government under the EEA. The U.S. Supreme Court has said on a number of occasions that “[i]t is a longstanding principle of American law ‘that legislation of Congress, unless a contrary intent appears, is meant to apply only within the territorial jurisdiction of the United States’”⁶⁴ With this in mind, Congress specifically identified the circumstances under which it intended the economic espionage and theft of trade secrets provisions of the EEA to apply overseas.⁶⁵ Either offense may be prosecuted if (1) the offender is a U.S. citizen or permanent resident alien or an organization organized under U.S. law, or (2) an act in furtherance of the offense is committed within the United States.⁶⁶

Statutory Exceptions to EEA Prohibitions

The EEA provides two express exceptions to the conduct that it prohibits (1) any otherwise lawful activity conducted by a governmental entity of the United States, a state, or a political subdivision of a state; or (2) the reporting of a suspected violation of law to any governmental entity of the United States, a state, or a political subdivision of a state, if such entity has lawful authority with respect to that violation.⁶⁷ The first exception permits the government to conduct an otherwise lawful “investigative, protective, or intelligence activity” with respect to the trade secret.⁶⁸ The second exception allows for the reporting of suspected criminal activity to law enforcement.⁶⁹

Bill).

⁶² 18 U.S.C. §1836.

⁶³ See *Barnes v. J.C. Penney Co.*, 2004 U.S. Dist. LEXIS 17557, *10 (N.D. Tex. 2004) (explaining that “[t]his criminal law provision [18 U.S.C. §1832] does not create a private cause of action. Any decision regarding prosecution under this provision is vested in the sole discretion of the United States Department of Justice and Plaintiff has no standing to seek relief under its terms.”).

⁶⁴ *Morrison v. National Australia Bank Ltd.*, 130 S.Ct. 2869, 2877 (2010), quoting *EEOC v. Arabian American Oil Co.*, 499 U.S. 244, 248 (1991) and *Foley Bros., Inc. v. Filardo*, 336 U.S. 281 (1949). See generally, CRS Report 94-166, *Extraterritorial Application of American Criminal Law*, by Charles Doyle.

⁶⁵ H.Rept. 104-788, at 14 (1996).

⁶⁶ 18 U.S.C. §1837. This broad grant of extraterritorial authority may raise enforcement problems if an act of economic espionage does not have any connection with the United States. For example, it has been suggested that “if a United States citizen residing abroad steals a Russian trade secret on behalf of the Chinese government, that act is a violation of the EEA ...” James H.A. Pooley et al., *Understanding the Economic Espionage Act of 1996*, 5 TEX. INTELL. PROP. L.J. 177, 204 (1997). Yet the Department of Justice would likely not bring an action under the EEA for this violation, “both to conserve its resources and to avoid the danger of intervening in what is essentially an internal dispute in a foreign country.” *Id.*

⁶⁷ 18 U.S.C. §1833.

⁶⁸ H.R. Rep. No. 104-788, at 14 (1996).

⁶⁹ *Id.*

Non-Preemption of Other Federal and State Laws

While the EEA was enacted in part due to the apparent shortcomings of other federal laws concerning the protection of trade secrets, the EEA expressly states that the act does not preempt or displace any other civil or criminal remedies provided by other federal or state laws for the misappropriation of a trade secret.⁷⁰ Federal prosecutors thus may bring criminal charges under the following laws in addition to, or instead of, the EEA, assuming that the conduct involved in the EEA violation also violates these federal criminal statutes: (1) the Computer Fraud and Abuse Act,⁷¹ which penalizes anyone who accesses certain computers without authorization or in excess of authorization, with the intent to defraud; (2) the National Stolen Property Act (NSPA),⁷² which prohibits the interstate transportation of tangible stolen “goods, wares, or merchandise,” or the knowing receipt of such property; and (3) the federal wire fraud statute,⁷³ which makes it illegal to use wire, radio, or television communications for purposes of executing a scheme to defraud.

International Law

The United States offers a more sophisticated and robust legal regime protecting trade secrets than most other countries. It has been noted that,

Much of the rest of the world has very weak laws or enforcement practices, with the issue particularly acute in many of the largest emerging economies, such as China, Brazil, Russia, and India. Thus, as supply chains and operations expand globally, a company’s ability to protect its trade secrets may be significantly diminished by weak rule of law and ineffective or non-existent enforcement in a number of countries.⁷⁴

There is no international treaty specifically pertaining to the protection of trade secrets. However, one of the agreements reached during the Uruguay Round of Multilateral Trade Negotiations (that concluded with the signing of the Marrakesh Agreement Establishing the World Trade Organization (WTO))⁷⁵ was the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS). TRIPS establishes minimum standards of protection for patents, copyrights, trademarks, and trade secrets that each WTO signatory state must give to the intellectual property of fellow WTO members.⁷⁶ Compliance with TRIPS is a prerequisite for WTO membership.

⁷⁰ 18 U.S.C. §1838.

⁷¹ 18 U.S.C. §1030(a)(4), (e)(2). For more information about this statute, see CRS Report 97-1025, *Cybercrime: An Overview of the Federal Computer Fraud and Abuse Statute and Related Federal Criminal Laws*, by Charles Doyle.

⁷² 18 U.S.C. §§2314, 2315. The NSPA has been interpreted by the federal courts to *exclude* the theft of *intangible* intellectual property. See *United States v. Aleynikov*, 676 F.3d 71, 77-78 (2d Cir. 2012) (“Some tangible property must be taken from the owner for there to be deemed a ‘good’ that is ‘stolen’ for purposes of the NSPA. ... [T]he theft and subsequent interstate transmission of purely intangible property is beyond the scope of the NSPA.”); *United States v. Agrawal*, 726 F.3d 235, 252 (2d Cir. 2013) (“[A] defendant such as Agrawal, who steals papers on which intangible intellectual property is reproduced, does assume physical control over something tangible as is necessary for the item to be a ‘good’ ... for purposes of the NSPA.”) (internal quotations and citations omitted).

⁷³ 18 U.S.C. §1343. For more information about this statute, see CRS Report R41930, *Mail and Wire Fraud: A Brief Overview of Federal Criminal Law*, by Charles Doyle.

⁷⁴ George Washington University Homeland Security Policy Institute, *Economic Espionage and Trade Secret Theft: An Overview of the Legal Landscape and Policy Response*, at 5 (September 2013), available at http://homelandsecurity.gwu.edu/sites/homelandsecurity.gwu.edu/files/downloads/Covington_SpecialIssueBrief.pdf.

⁷⁵ For more information about the WTO, see CRS Report RS22154, *World Trade Organization (WTO) Decisions and Their Effect in U.S. Law*, by Jane M. Smith, Brandon J. Murrill, and Daniel T. Shedd.

⁷⁶ World Trade Organization, *Understanding the WTO - Intellectual Property: Protection and Enforcement*, at http://www.wto.org/english/thewto_e/whatis_e/tif_e/agrm7_e.htm.

TRIPS does not explicitly refer to “trade secrets.” However, in order to “ensur[e] effective protection against unfair competition,”⁷⁷ TRIPS does refer to “protection of undisclosed information” and uses a definition that is similar to that of the traditional trade secret definition described above. Article 39 of TRIPS obliges WTO members to protect individuals and corporations⁷⁸ who own or control “undisclosed information” from unauthorized disclosure, acquisition, or use “without their consent in a manner contrary to honest commercial practices.”⁷⁹ A footnote defines “a manner contrary to honest commercial practices” to mean “practices such as breach of contract, breach of confidence and inducement to breach, and includes the acquisition of undisclosed information by third parties who knew, or were grossly negligent in failing to know, that such practices were involved in the acquisition.”⁸⁰

Article 39 also defines “undisclosed information” as information that

1. “is secret in the sense that it is not, as a body or in the precise configuration and assembly of its components, generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question;
2. has commercial value because it is secret; and
3. has been subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep it secret.”⁸¹

Note that unlike the federal Economic Espionage Act that provides an extensive list of the various types of information that may be considered a trade secret, Article 39 lacks such specificity and thus the term “information” could be subject to broad or narrow interpretation by WTO members. In addition, recent testimony before Congress criticized the vagueness of the protection mandated by Article 39:

The heart of the relevant clause in TRIPS is vague; it asks whether the trade secret has been acquired or used “in a manner contrary to honest commercial practices.” As a result, in Europe alone, trade secret law, which to date is not yet controlled by a European Union Directive, is a patchwork of different forms of protection. What is contrary to honest commercial practices in one country may be considered acceptable in other countries.⁸²

Nevertheless, Article 39 of TRIPS is the first time that protection of trade secrets has appeared in a multilateral treaty.⁸³ According to a legal commentator, the “TRIPS Agreement includes a requirement that member nations enact trade secret law that is very similar to U.S. trade secret law. ... This is significant in light of the fact that trade secret law either did not exist or was undeveloped in many countries prior to the TRIPS Agreement.”⁸⁴

⁷⁷ TRIPS Agreement, art. 39, para. 1, available at http://www.wto.org/english/docs_e/legal_e/27-trips_04d_e.htm#7.

⁷⁸ The TRIPS Agreement refers to “individuals and corporations” as “natural and legal persons.”

⁷⁹ TRIPS Agreement, art. 39, para. 2.

⁸⁰ *Id.* n.10.

⁸¹ *Id.*, art. 39, para. 2.

⁸² *Trade Secrets: Promoting and Protecting American Innovation, Competitiveness and Market Access in Foreign Markets: Hearings Before the House Judiciary Comm., Subcomm. on Courts, Intellectual Property and Internet*, 113th Cong. 2d Sess. (2014) (statement of David M. Simon, Senior Vice President, salesforce.com, Inc.).

⁸³ Francois Dessemontet, *Arbitration and Confidentiality*, 7 AM. REV. INT'L ARB. 299, 307 (1996).

⁸⁴ Andrew Beckerman-Rodau, *Patent Law - Balancing Profit Maximization and Public Access to Technology*, 4 COLUM. SCI. & TECH. L. REV. 1, 20 n.108. (2002).

The WTO has the power to resolve disputes between member states for alleged violations of the TRIPS Agreement, including its provisions governing “undisclosed information.” However, such cases appear to be very rare; a search of the WTO’s dispute cases revealed that a complaint involving Article 39 has occurred only once, and that case was eventually withdrawn after the parties (China and the European Communities) reached an agreement in the form of a Memorandum of Understanding.⁸⁵ In May 2014, Senator Schumer sent a letter to the U.S. Trade Representative (USTR) Michael Froman, urging him to “initiate a case at the World Trade Organization (WTO) against China for state-backed cyber espionage against American businesses and workers.”⁸⁶ The letter argues that Chinese policies that sanction cyber espionage are in clear violation of the TRIPS agreement that obliges WTO members to protect trade secrets.⁸⁷ As of the date of this report, the USTR has not filed a WTO complaint against China over this matter.⁸⁸

The United States has entered into numerous bilateral and multilateral free trade agreements (FTAs) that require their signatories to provide higher levels of intellectual property protection than are required under the TRIPS Agreement. These intellectual property obligations exceed those of the TRIPS Agreement and are commonly referred to as “TRIPS-plus agreements.” The United States has for many years pursued a policy of encouraging its trading partners to adopt TRIPS-plus provisions, which include more robust protections for trade secrets. Negotiating the inclusion of trade secret protection as part of these FTAs is discussed later in this report.

The Growing Problem of Trade Secret Theft and Economic Espionage

Measuring Economic Loss

It is difficult to determine the total value of trade secrets to U.S. businesses, although a report issued by the U.S. Chamber of Commerce stated that “[p]ublicly traded U.S. companies own an estimated \$5 trillion worth of trade secrets.”⁸⁹ A recent study by PricewaterhouseCoopers (PwC) and the Center for Responsible Enterprise and Trade (CREATe.org) suggested that the economic loss attributable to trade secret theft is between 1% to 3% of U.S. Gross Domestic Product.⁹⁰ A more precise calculation of the economic impact of trade secret theft is impeded by several factors identified by the Office of the National Counterintelligence Executive (ONCIX):

1. A company may not realize that its sensitive information has been stolen until years after the crime.

⁸⁵ WTO, Dispute Settlement DS372, available at http://www.wto.org/english/tratop_e/dispu_e/cases_e/ds372_e.htm.

⁸⁶ Senator Schumer, *Press Release: Schumer Calls on U.S. Trade Rep to File WTO Suit in Response to Chinese Cyber-Attacks*, May 22, 2014, available at <http://www.schumer.senate.gov/Newsroom/record.cfm?id=351779>.

⁸⁷ *Id.*

⁸⁸ For more information on this topic, see CRS Report IN10079, *Alleged Chinese Government Cyber Theft of U.S. Commercial Trade Secrets*, by Wayne M. Morrison, Susan V. Lawrence, and John W. Rollins.

⁸⁹ U.S. Chamber of Commerce, *The Case for Enhanced Protection of Trade Secrets in the Trans-Pacific Partnership Agreement*, at 10, available at https://www.uschamber.com/sites/default/files/legacy/international/files/Final%20TPP%20Trade%20Secrets%208_0.pdf.

⁹⁰ PwC & CREATe.org, *Economic Impact of Trade Secret Theft: A Framework for Companies to Safeguard Trade Secrets and Mitigate Potential Threats*, at 3 (February 2014), available at http://www.pwc.com/en_US/us/forensic-services/publications/assets/economic-impact.pdf.

2. Reporting security breaches to the FBI or other law enforcement entity could harm the company's reputation and stock prices, or damage its corporate relationships.
3. Publicly accusing a foreign government or business competitor of trade secret theft carries the risk of offending the company's potential customers or business partners.
4. It may be very difficult, if not impossible, to measure the monetary value of some forms of sensitive information.⁹¹

ONCIX further opined that the “[e]stimates from academic literature on the losses from economic espionage range so widely as to be meaningless—from \$2 billion to \$400 billion or more a year—reflecting the scarcity of data and the variety of methods used to calculate losses.”⁹²

Types of Offenders

Domestic

In the vast majority (over 90%) of trade secret cases that are litigated in state court, the alleged misappropriator is someone the trade secret owner knows, either a current or former employee or a business partner.⁹³ Given this statistic, it has been suggested that “a prudent trade secret owner should focus its efforts in large part on protecting trade secrets from unscrupulous employees and, to a somewhat lesser extent, business partners.”⁹⁴

Foreign

In its October 2011 report to Congress, ONCIX warned that “[b]ecause the United States is a leader in the development of new technologies and a central player in global finance and trade networks, foreign attempts to collect US technological and economic information will continue at a high level and will represent a growing and persistent threat to US economic security.”⁹⁵ ONCIX raised particular concerns about the use of the Internet, computer technologies, and mobile communication devices to steal the trade secrets of U.S. businesses:

[N]early all business records, research results, and other sensitive economic or technology-related information now exist primarily in digital form. Cyberspace makes it possible for foreign collectors to gather enormous quantities of information quickly and with little risk, whether via remote exploitation of victims' computer networks, downloads of data to external media devices, or e-mail messages transmitting sensitive information.⁹⁶

⁹¹ Office of the National Counterintelligence Executive, *Foreign Spies Stealing US Economic Secrets in Cyberspace*, October 2011, at 3.

⁹² *Id.* at 4.

⁹³ David S. Almeling et al., *A Statistical Analysis of Trade Secret Litigation in State Courts*, 46 GONZAGA L. REV. 57, 68 (2010).

⁹⁴ *Id.*

⁹⁵ Office of the National Counterintelligence Executive, *Foreign Spies Stealing US Economic Secrets in Cyberspace*, October 2011, at i, available at http://www.ncix.gov/publications/reports/fecie_all/Foreign_Economic_Collection_2011.pdf.

⁹⁶ *Id.* at iii.

While cyber-enabled methods of trade secret theft are getting increased attention from the federal government,⁹⁷ it is important to realize that many actors (foreign intelligence services, corporate competitors, transnational criminal organizations) “still rely on physical means such as recruitment of insiders and placement of agents within companies for purposes of stealing critical data.”⁹⁸ The motivation for trade secret theft varies, with some perpetrators “seek[ing] personal financial gain, while others hope to advance national interests or political and social causes.”⁹⁹

According to ONCIX, the governments of China and Russia are particularly “aggressive and capable collectors of sensitive U.S. economic information and technologies,” and “Chinese actors are the world’s most active and persistent perpetrators of economic espionage.”¹⁰⁰ The U.S. International Trade Commission (USITC) released a report indicating that U.S. firms lost approximately \$1.1 billion in the year 2009 due to Chinese trade secret misappropriation.¹⁰¹ Between January 2009 and January 2013, China was involved in 17 criminal prosecutions (out of a total of 20) that the U.S. Department of Justice brought pursuant to the EEA.¹⁰²

Enforcement of Trade Secret Rights

Litigation and Prosecution

At the state level, enforcement of trade secret laws is generally the responsibility of the trade secret owner (by filing a civil suit in state court against an individual or organization alleged to have misappropriated the trade secret in order to obtain remedies such as injunctive relief and compensatory and punitive damages).¹⁰³ In addition, as discussed above, a few states have enacted criminal laws against trade secret theft under which state prosecutors may bring criminal charges against defendants in trade secret cases.

At the federal level, the Economic Espionage Unit located within the Federal Bureau of Investigation’s (FBI’s) Counterintelligence Division has primary responsibility for investigating offenses under the EEA.¹⁰⁴ The U.S. Department of Justice (DOJ) and its U.S. Attorneys have the power to prosecute cases involving corporate and state-sponsored trade secret theft.¹⁰⁵ The

⁹⁷ See, e.g., CRS Report IN10079, *Alleged Chinese Government Cyber Theft of U.S. Commercial Trade Secrets*, by Wayne M. Morrison, Susan V. Lawrence, and John W. Rollins.

⁹⁸ PwC & CREATE.org, *Economic Impact of Trade Secret Theft: A Framework for Companies to Safeguard Trade Secrets and Mitigate Potential Threats*, at 4.

⁹⁹ *Id.* at 10.

¹⁰⁰ Office of the National Counterintelligence Executive, *Foreign Spies Stealing US Economic Secrets in Cyberspace*, October 2011, at i-ii.

¹⁰¹ USITC, *China: Effects of Intellectual Property Infringement and Indigenous Innovation Policies on the U.S. Economy*, Investigation no. 332-519, USITC Publication 4226, May 2011, 3-42, available at <http://www.usitc.gov/publications/332/pub4226.pdf>.

¹⁰² Executive Office of the President, *Administration Strategy on Mitigating the Theft of U.S. Trade Secrets*, February 2013, at 23-31, available at http://www.whitehouse.gov/sites/default/files/omb/IPEC/admin_strategy_on_mitigating_the_theft_of_u.s._trade_secrets.pdf.

¹⁰³ ROGER E. SCHECHTER & JOHN R. THOMAS, *INTELLECTUAL PROPERTY: THE LAW OF COPYRIGHTS, PATENTS AND TRADEMARKS*, §24.4.

¹⁰⁴ *Economic Espionage and Trade Secret Theft: Are Our Laws Adequate for Today’s Threats?: Hearings Before the Senate Judiciary Comm., Subcomm. on Crime and Terrorism*, 113th Cong. 2d Sess. (2014) (statement of Randall C. Coleman, Assistant Director, Counterintelligence Division, FBI).

¹⁰⁵ The 93 U.S. Attorneys’ Offices located across the United States and its territories have primary responsibility for prosecution of intellectual property offenses. Every office has at least one Computer Hacking and Intellectual Property

Attorney General is also authorized by the EEA to bring a civil action in federal court to obtain “appropriate injunctive relief” against any violation of the EEA.¹⁰⁶ However, as discussed in detail later in this report, federal law does not currently provide a private, federal cause of action for trade secret misappropriation.

Executive Branch Actions

Administration Strategy

In February 2013, the White House issued a report, *The Administration Strategy on Mitigating the Theft of U.S. Trade Secrets*, which describes its plan for “vigorously ... combat[ing] the theft of U.S. trade secrets that could be used by foreign companies or foreign governments to gain an unfair economic edge.”¹⁰⁷ The report noted that the theft of valuable U.S. trade secrets has several negative consequences, including the loss of U.S. companies’ intellectual property, the harm to American business innovation and global competitiveness, damage to national and economic security, possible reduction of U.S. exports, and the increased risk of American job losses.¹⁰⁸

The report contains five “strategy action items” that are intended to provide a “means for improved coordination within the U.S. government” to protect the integrity of trade secrets.¹⁰⁹

1. Focusing diplomatic efforts and pressure on other countries to protect trade secrets and discourage their theft, including (through the U.S. Trade Representative, or USTR) seeking provisions in bilateral, regional, and multilateral trade agreements¹¹⁰ that require parties to establish remedies for trade secret theft similar to those provided for in U.S. law;
2. Promoting the development and adoption of voluntary best practices by private industry to protect trade secrets;
3. Enhancing domestic law enforcement operations by having the FBI and DOJ prioritize trade secret theft investigations and prosecutions, as well as having the Office of the Director of National Intelligence share information with the private sector about potential foreign espionage threats;
4. Improving domestic legislation to ensure that federal laws are effective in protecting trade secrets; and
5. Conducting education and outreach efforts to raise public awareness of the detrimental effects of trade secret theft.

(CHIP) Coordinator, who are Assistant U.S. Attorneys with expertise in prosecuting IP and computer crimes. U.S. Dep’t of Justice, Computer Crime & Intellectual Property Section, Prosecuting Intellectual Property Crimes (4th ed. 2013), available at http://www.justice.gov/criminal/cybercrime/docs/prosecuting_ip_crimes_manual_2013.pdf.

¹⁰⁶ 18 U.S.C. §1836.

¹⁰⁷ Executive Office of the President, *Administration Strategy on Mitigating the Theft of U.S. Trade Secrets*, February 2013, at 1-2.

¹⁰⁸ *Id.* at 1.

¹⁰⁹ *Id.* at 2.

¹¹⁰ For a comprehensive explanation of how the federal government may promote the protection of U.S. intellectual property through its international trade policy, see CRS Report RL34292, *Intellectual Property Rights and International Trade*, by Shayerah Ilias Akhtar and Ian F. Fergusson.

Special 301

The USTR is required¹¹¹ to conduct an annual review of foreign countries' intellectual property policies and practices and to publish a "Special 301" Report that identifies countries that lack adequate and effective intellectual property protection and enforcement regimes. The 2013 Special 301 Report was the first time that the USTR included a section dedicated to "the growing problem of misappropriation of trade secrets in China and elsewhere."¹¹² The Report "urge[d] its trading partners to ensure that they have robust systems for protecting trade secrets, including deterrent penalties for criminal trade secret theft" and promised that the "USTR will monitor developments in this area."¹¹³

In a 2014 congressional hearing, a witness described the negative consequences of overseas trade secret theft as follows: "Inadequate protection of trade secrets abroad harms not only companies whose property is stolen, but also the country where the theft occurs, because companies are then less likely to form joint ventures and make high-value global supply chain investments in those countries."¹¹⁴

Free Trade Agreements (TPP and TTIP)

Currently, the USTR is seeking to improve trade secret protection in countries with which it has been negotiating two free trade agreements: (1) the Trans-Pacific Partnership (TPP),¹¹⁵ which involves 11 countries in the Asia-Pacific region, and (2) the Transatlantic Trade and Investment Partnership (TTIP),¹¹⁶ with the European Union. The U.S. Chamber of Commerce has argued that the legal regimes of TPP countries need significant improvement in the area of trade secret protection:

Some TPP countries, such as Canada, Australia, Malaysia, and Singapore, have no laws criminalizing traditional trade secret disclosure or misappropriation. ... Among those countries that do criminalize trade secret misappropriation or disclosure, the penalties often vary from those that would not provide sufficient deterrent effect to those that would but only if applied consistently. ... The low criminal penalties or lack thereof in some TPP jurisdictions are particularly troublesome, as criminal penalties are believed to provide a greater deterrent to the would-be trade secret thief than the prospect of a civil penalty alone.¹¹⁷

Such variation in trade secret protection is also present in the TTIP negotiations, as the European Union currently lacks a consistent, harmonized legal system governing trade secret protection;

¹¹¹ P.L. 93-618, as amended by P.L. 100-418.

¹¹² USTR, *2013 Special 301 Report*, at 4 (May 2013), available at <http://www.ustr.gov/sites/default/files/05012013%202013%20Special%20301%20Report.pdf>.

¹¹³ *Id.* at 13.

¹¹⁴ *Trade Secrets: Promoting and Protecting American Innovation, Competitiveness and Market Access in Foreign Markets: Hearings Before the House Judiciary Comm., Subcomm. on Courts, Intellectual Property and Internet*, 113th Cong. 2d Sess. (2014) (statement of Thaddeus Burns, Senior Counsel, General Electric, on behalf of the Intellectual Property Owners Association).

¹¹⁵ For more information on the TPP and intellectual property rights, see CRS Report R42694, *The Trans-Pacific Partnership (TPP) Negotiations and Issues for Congress*, coordinated by Ian F. Fergusson.

¹¹⁶ For more information on the TTIP and intellectual property rights, see CRS Report R43387, *Transatlantic Trade and Investment Partnership (T-TIP) Negotiations*, by Shayerah Ilias Akhtar, Vivian C. Jones, and Renée Johnson.

¹¹⁷ U.S. Chamber of Commerce, *The Case for Enhanced Protection of Trade Secrets in the Trans-Pacific Partnership Agreement*, at 23.

instead, there are disparities across the 27 EU Member States in “what [trade secrets] can be protected, in what circumstances, and what the courts can or will do.”¹¹⁸

Limitations of Current Law and Proposed Changes

It has been argued that “federal law has not kept pace with the technological innovation that has enabled increased trade secret theft.”¹¹⁹ The lack of a federal civil cause of action for trade secret misappropriation is perhaps the most widely cited deficiency in U.S. trade secret law. As one legal practitioner has argued,

Unfortunately the EEA has not deterred trade secret theft and foreign economic espionage. The Computer Crime and Intellectual Property Section of the United States Department of Justice has done an excellent job, but the burden on the government is too great. Without a federal civil cause of action, U.S. companies cannot adequately protect U.S. trade secret assets in a worldwide economy that now crosses international boundaries.¹²⁰

Another problem companies have encountered in having only federal criminal statutes protecting trade secrets is that “criminal law punishes the defendant, but the process for compensating the victim is unwieldy, particularly when compared to relief available under civil law.”¹²¹ Others have highlighted the limitations of the EEA’s extraterritorial application, noting that “prosecutors lack enforcement and proper service mechanisms against individuals and entities located outside the United States ... Prosecutors cannot charge alleged violators of the EEA until they cross U.S. borders.”¹²² Reportedly, since the enactment of the EEA in 1996, there have been relatively few cases prosecuted under the law: approximately 125 indictments¹²³ and 10 convictions.¹²⁴

In Support of a Federal Civil Cause of Action for Trade Secret Theft

Some observers have urged Congress to adopt a comprehensive, federal trade secret law in order to promote uniformity in trade secret law throughout the United States.¹²⁵ Supporters of such

¹¹⁸ Robert Anderson & Sarah Turner, *Report on Trade Secrets for the European Commission* (January 2012), at 44, available at http://ec.europa.eu/internal_market/ipenforcement/docs/trade-secrets/120113_study_en.pdf.

¹¹⁹ *Trade Secrets: Promoting and Protecting American Innovation, Competitiveness and Market Access in Foreign Markets: Hearings Before the House Judiciary Comm., Subcomm. on Courts, Intellectual Property and Internet*, 113th Cong. 2d Sess. (2014) (statement of Thaddeus Burns, Senior Counsel, General Electric, on behalf of the Intellectual Property Owners Association).

¹²⁰ R. Mark Halligan, *Protecting U.S. Trade Secret Assets in the 21st Century*, 6:1 LANDSLIDE (September/October 2013), available at http://www.americanbar.org/publications/landslide/2013-14/september-october-2013/protecting_us_trade_secret_assets_the_21st_century.html.

¹²¹ *Economic Espionage and Trade Secret Theft: Are Our Laws Adequate for Today’s Threats?: Hearings Before the Senate Judiciary Comm., Subcomm. on Crime and Terrorism*, 113th Cong. 2d Sess. (2014) (statement of Douglas K. Norman, Vice President & General Patent Counsel, Eli Lilly and Company).

¹²² *The Report of the Commission on the Theft of American Intellectual Property*, at 42 (May 2013).

¹²³ *Can You Keep a Secret?*, THE ECONOMIST, March 16, 2013.

¹²⁴ *Economic Espionage and Trade Secret Theft: Are Our Laws Adequate for Today’s Threats?: Hearings Before the Senate Judiciary Comm., Subcomm. on Crime and Terrorism*, 113th Cong. 2d Sess. (2014) (statement of Randall C. Coleman, Assistant Director, Counterintelligence Division, FBI); see also News Release, *Senator Coons, Hatch Introduce Bill to Combat Theft of Trade Secrets and Protect Jobs*, April 29, 2014, at <http://www.coons.senate.gov/newsroom/releases/release/senators-coons-hatch-introduce-bill-to-combat-theft-of-trade-secrets-and-protect-jobs> (“Current federal criminal law is insufficient. Although the Economic Espionage Act of 1996 made trade secret theft a crime, the Department of Justice brought only 25 trade secret theft cases last year.”).

¹²⁵ See, e.g., Marina Lao, *Federalizing Trade Secrets Law in an Information Economy*, 59 OHIO STATE L. J. 1633

legislation have argued that a federal trade secrets law would create procedural and substantive standards for the trade secret misappropriation offense on a uniform nationwide basis, in response to the current situation of state trade secret laws in which there are “fundamental differences about what constitutes a trade secret, what is required to misappropriate it, and what remedies are available” due to state-by-state variations in statutory text and state court interpretations.¹²⁶ In addition, Senator Coons has observed that, in contrast to state courts, “[f]ederal courts are better suited to working across state and national boundaries to facilitate discovery, serve defendants or witnesses, or prevent a party from leaving the country.”¹²⁷ Representative Nadler has also asserted that the limitations of state trade secret law are impediments to the effective protection of U.S. corporate trade secrets in a global economy:

While this system [of state law remedies] appears to have worked relatively well for local and intrastate disputes, it has not proven efficient or effective for [trade secret theft] incidents that cross state, and sometimes international, borders. ...

[A] fifty-state system does not work well in our increasingly mobile and globally interconnected world. Former employees and industrial spies are likely to carry or transfer secret information across state borders or overseas. The limited jurisdiction of the state court system makes it more difficult to obtain discovery or to act quickly enough to enforce an order that might stop the immediate loss of company secrets.¹²⁸

Some commentators argue that trade secrets deserve to receive the same robust legal protections available to the three other types of intellectual property.¹²⁹ For example, owners of patents, copyright, and trademarks have the right to file a lawsuit against infringers in federal court to recover damages and possibly to enjoin further infringement, and yet there is no similar right afforded to trade secret owners,¹³⁰ despite the fact that trade secrets are often considered by many companies as their most valuable and important intellectual property asset.¹³¹ Instead, at the federal level, companies must rely on the federal government (and its limited resources) to enforce their trade secret rights.

(1998); Rebel J. Pace, *The Case for a Federal Trade Secrets Act*, 8 HARVARD J. OF LAW & TECHNOLOGY (1995).

¹²⁶ David S. Almeling, *Four Reasons to Enact a Federal Trade Secrets Act*, FORDHAM INTELLECTUAL PROPERTY, MEDIA & ENTERTAINMENT LAW JOURNAL XIX.3 (2009), at 774.

¹²⁷ News Release, *Senators Coons, Hatch Introduce Bill to Combat Theft of Trade Secrets and Protect Jobs*, April 29, 2014.

¹²⁸ Press Release, *Rep. Nadler on Protecting Trade Secrets of American Companies*, June 24, 2014, available at <http://nadler.house.gov/press-release/rep-nadler-protecting-trade-secrets-american-companies>.

¹²⁹ *Id.* (noting that U.S. law “already protect[s] trademarks, copyrights, and patents through federal civil remedies. It is time to do the same for trade secrets.”); *Economic Espionage and Trade Secret Theft: Are Our Laws Adequate for Today’s Threats?: Hearings Before the Senate Judiciary Comm., Subcomm. on Crime and Terrorism*, 113th Cong. 2d Sess. (2014) (statement of Drew Greenblatt, President and Owner, Marlin Steel Wire Products) (“Despite their strategic economic importance, trade secrets misappropriation is the only form of U.S. intellectual property violation for which the owner lacks access to federal court. This leaves U.S. firms without a key tool to prevent trade secret theft and recover any losses.”).

¹³⁰ *Trade Secrets: Promoting and Protecting American Innovation, Competitiveness and Market Access in Foreign Markets: Hearings Before the House Judiciary Comm., Subcomm. on Courts, Intellectual Property and Internet*, 113th Cong. 2d Sess. (2014) (statement of Thaddeus Burns, Senior Counsel, General Electric, on behalf of the Intellectual Property Owners Association).

¹³¹ U.S. Chamber of Commerce, *The Case for Enhanced Protection of Trade Secrets in the Trans-Pacific Partnership Agreement*, at 10; see also David Kappos, *Trade Secrets: Promise of Federal Protection Brings New Hope for Critical IP Law*, TheHill.com, June 30, 2014, at <http://thehill.com/blogs/congress-blog/technology/210848-trade-secrets-promise-of-federal-protection-brings-new-hope> (“Despite accounting for an average of two-thirds of U.S. companies’ information value, trade secrets suffer from extremely limited recognition under federal law.”).

Supporters of a federal civil remedy for trade secret misappropriation believe that Congress should empower federal courts to issue ex parte orders to seize stolen trade secrets in certain limited circumstances, such as “to prevent an imminent misappropriation, the dissemination of a stolen trade secret, and to preserve evidence.”¹³² However, they note that any legislation should contain proper safeguards to prevent abuse of the ex parte process, “including damages in the event of wrongful seizure and protection of the information seized to protect against inappropriate access to the information.”¹³³

Finally, it has been asserted that “the United States has not consistently received cooperation from international jurisdictions in protecting trade secrets in part because it does not have its own federal civil statute to reference in encouraging the adoption and enforcement of similar legislation by its treaty partners.”¹³⁴

In Opposition to a Federal Civil Trade Secret Remedy

The establishment of a federal civil trade secret remedy has many proponents, yet there have been some opposing views. In 2007, the Trade Secrets Committee of the American Intellectual Property Law Association (AIPLA) issued a report that advised against federalizing trade secret law, in part out of a concern that such action may create additional burdens and costs upon the federal judiciary:

The Committee believes that the problem of disparate state trade secret laws may have been overstated, because the various state statutes share much in common, especially those based upon the Uniform Trade Secrets Act (UTSA). Furthermore, many trade secret cases are already heard in federal court through diversity or supplemental jurisdiction, providing at least federal procedure, if not substantive law, benefits to private litigants. Others have argued, and the Committee agrees, that the current state regulation of trade secrets, although far from perfect, is functioning adequately and that federalizing state trade secret law would, therefore, needlessly burden the already overworked federal judiciary.¹³⁵

However, AIPLA has since changed its position on this matter, as revealed in an April 2013 letter to the U.S. Intellectual Property Enforcement Coordinator (IPEC). In response to the IPEC’s request for public comments for an administration legislative review related to economic espionage and trade secret theft, the President of AIPLA wrote that because of the increase in foreign trade secret theft in recent years, “AIPLA believes that the time has come to consider a federal civil remedy for international trade secret misappropriation.”¹³⁶ Furthermore, the AIPLA letter argued that “[a]ny federal legislation should not preempt state trade secret laws, but should

¹³² *Trade Secrets: Promoting and Protecting American Innovation, Competitiveness and Market Access in Foreign Markets: Hearings Before the House Judiciary Comm., Subcomm. on Courts, Intellectual Property and Internet*, 113th Cong. 2d Sess. (2014) (statement of Thaddeus Burns, Senior Counsel, General Electric, on behalf of the Intellectual Property Owners Association).

¹³³ *Id.*

¹³⁴ *Id.*

¹³⁵ American Intellectual Property Law Association, *Report of the AIPLA Trade Secrets Committee* (2007), at 2, available at http://www2.aipla.org/MSTemplate.cfm?Section=Proposal_to_Federalize_Trade_Secret_Law&Site=Trade_Secret_Law&Template=/ContentManagement/ContentDisplay.cfm&ContentID=7041.

¹³⁶ AIPLA Comments on Trade Secret Theft Strategy Legislative Review, April 22, 2013, at 2, available at <http://www.aipla.org/advocacy/executive/Documents/AIPLA%20Letter%20to%20IPEC%20on%20Trade%20Secrets%20-%2004.22.13.pdf>.

instead complement them and should provide jurisdiction for civil actions involving claims involving the international theft of trade secrets.”¹³⁷

A group of law school professors has urged Congress to reject the Defend Trade Secrets Act of 2015 (DTSA) (discussed in detail in the following section of this report) because they believe that the legislation, which would establish a new private cause of action under the EEA, “is likely to create new problems that could adversely impact domestic innovation, increase the duration and cost of trade secret litigation, and ultimately negatively affect economic growth.”¹³⁸ The letter was authored or signed by professors who teach intellectual property law, trade secret law, innovation policy, and information law throughout the United States. In the view of these law professors, the DTSA is not necessary and could even cause unintentional harm. They argue that: “(1) it will not address the cyberespionage problem that is most often used to justify the adoption of a federal trade secret law; (2) a federal trade secret law is not needed to protect U.S. trade secrets because there is already a robust set of state laws for the protection of such secrets; and (3) there are significant costs to creating a federal civil cause of action for trade secret misappropriation.”¹³⁹

An attorney who specializes in patent and trade secret litigation has identified two potential problems with the DTSA’s lack of a provision expressly preempting state trade secret laws:¹⁴⁰

First, the need for the DTSA stems in part from state-by-state variations in trade secret laws and the transactional and substantive problems that such variations impose. The DTSA leaves those variations in place. Worse, the DTSA adds another law to the already cluttered landscape of 48 UTSA states (with their variations), two non-UTSA states, the federal Economic Espionage Act, and a federal common trade secret law.

Second, the DTSA opens a backdoor to common-law and other causes of action that are precluded in most states. The UTSA “displaces tort, restitutionary, and other laws...providing civil remedies for misappropriation of a trade secret.” The DTSA doesn’t displace anything.

Under the DTSA, trade secret plaintiffs would have the option of pursuing their claim in state or federal court and, if they choose federal court, the additional option of asserting duplicative causes of actions that aren’t available in state courts.¹⁴¹

A legal commentator testified before Congress in December 2015 that legislation creating a federal civil trade secret remedy may increase the length and cost of trade secret litigation:

[A]s there is no federal civil trade secret jurisprudence, numerous issues that have long been resolved at the state level are sure to be highly litigated at the federal level. As the federal courts develop their jurisprudence, they will have multiple sources of existing state law to borrow from, but with no direction from Congress on which to choose. ... thereby leading to less uniformity in trade secret doctrine, not more.¹⁴²

¹³⁷ *Id.* at 3.

¹³⁸ Professors’ Letter in Opposition to the Defend Trade Secrets Act of 2015 (S. 1890, H.R. 3326), November 17, 2015, available at <https://cyberlaw.stanford.edu/files/blogs/2015%20Professors%20Letter%20in%20Opposition%20to%20DTSA%20FINAL.pdf>.

¹³⁹ *Id.* at 2.

¹⁴⁰ The DTSA has a “rule of construction” provision that expresses that Congress does not intend for the DTSA “to preempt any other provision of law.” S. 1890, §2(f).

¹⁴¹ David S. Almeling, *Guest Post: Defend Trade Secrets Act – A Primer, an Endorsement, and a Criticism*, Patently-O, May 30, 2014, at <http://patentlyo.com/patent/2014/05/secrets-endorsement-criticism.html>.

¹⁴² *Protecting Trade Secrets: the Impact of Trade Secret Theft on American Competitiveness and Potential Solutions to*

Legislation in the 114th Congress: The Defend Trade Secrets Act

Two bills have been introduced in the 114th Congress related to trade secret misappropriation, S. 1890 and H.R. 3326 (the Defend Trade Secrets Act (DTSA) of 2015). As introduced on July 29, 2015 by Senator Hatch and Representative Doug Collins, respectively, the bills are substantively identical. S. 1890 has seen all the legislative activity to date. On January 28, 2016, the Senate Judiciary Committee, by a unanimous voice vote, reported S. 1890 with an amendment in the nature of a substitute. Senator Grassley filed a written report on March 7, 2016.¹⁴³ The Senate passed S. 1890 by a vote of 87-0 on April 4, 2016. On April 20, the House Judiciary Committee unanimously approved S. 1890 by voice vote.¹⁴⁴ The following summarizes the key provisions of the DTSA (S. 1890), as passed by the Senate and the House Judiciary Committee.

The DTSA would create a private cause of action in federal courts for trade secret owners to sue misappropriators. The DTSA would establish this new private right by adding a subsection entitled “private civil actions” to the provision of the EEA that currently authorizes the Attorney General to bring a civil action to obtain “appropriate injunctive relief” against any violation of the EEA, codified at 18 U.S.C. Section 1836.

The DTSA would allow an owner of a trade secret that is misappropriated to bring a civil action if the trade secret is related to a product or service used in, or intended to be used in, interstate or foreign commerce.¹⁴⁵ The legislation would amend the EEA’s definition section (18 U.S.C. Section 1839) to include definitions of the terms “misappropriation” and “improper means” that largely mirror the definitions in the Uniform Trade Secrets Act.¹⁴⁶

The DTSA would provide a court with the power to issue civil ex parte orders, “only in extraordinary circumstances,”¹⁴⁷ for the “seizure of property necessary to prevent the propagation or dissemination of the trade secret that is the subject of the action.”¹⁴⁸ According to the Senate Judiciary Committee report, “[t]he ex parte seizure provision is expected to be used in instances in which a defendant is seeking to flee the country or planning to disclose the trade secret to a third party immediately or is otherwise not amenable to the enforcement of the court’s orders.”¹⁴⁹ To avoid potential abuse of these seizure orders, the DTSA includes several detailed requirements that the court must follow before granting a seizure order, including (1) a finding that immediate and irreparable injury will occur if the seizure is not ordered; (2) the applicant for the seizure order is likely to succeed in showing that the information is a trade secret and that the person to whom the seizure is ordered misappropriated it by improper means; (3) the harm to the applicant

Remedy This Harm: Hearings Before the Senate Judiciary Comm., 114th Cong. 1st Sess. (2015) (statement of Professor Sharon K. Sandeen, at 4-5) (citation omitted).

¹⁴³ S.Rept. 114-220.

¹⁴⁴ Press Release: Judiciary Committee Approves Bipartisan Trade Secrets Legislation, April 20, 2016, at <https://judiciary.house.gov/press-release/judiciary-committee-approves-bipartisan-trade-secrets-legislation/>.

¹⁴⁵ S. 1890, §2(a), adding new 18 U.S.C. §1836(b)(1).

¹⁴⁶ *Id.* §2(b), amending 18 U.S.C. §1839. Note that the EEA’s definition section already includes an expansive definition of “trade secret” as well as “owner” (includes a person or entity).

¹⁴⁷ As introduced, S. 1890 did not include this qualification. The Senate Judiciary Committee adopted an amendment that added this language to the bill.

¹⁴⁸ S. 1890, §2(a), adding new 18 U.S.C. §1836(b)(2)(A)(i).

¹⁴⁹ S.Rept. 114-220, at 6.

in denying the order outweighs the harm to the legitimate interests of the party against whom the seizure is ordered and substantially outweighs the harm to any third parties; and (4) the matter to be seized would be destroyed, moved, hidden, or otherwise rendered inaccessible if the party in possession of such property were given advance notice.¹⁵⁰ In addition, the DTSA requires that any seizure order must contain several elements, including¹⁵¹

1. findings of fact and conclusions of law required for the order;
2. the narrowest seizure of property necessary to protect the trade secret;
3. a direction that the seizure be conducted in a manner that minimizes any interruption of the business operations of third parties and, to the extent possible, does not interrupt the legitimate business operations of the person accused of misappropriating the trade secret;
4. a date for a hearing regarding the seizure order at the earliest possible time, but not later than seven days after the order has issued (unless the parties involved consent to another date);
5. a requirement that the applicant provide a security that the court finds is adequate for the payment of damages that any person may be entitled to recover as a result of a wrongful or excessive seizure.

The DTSA would require the court to take custody of any seized materials and secure it from physical and electronic access.¹⁵² The DTSA provides a cause of action for any person who suffers damage by reason of a wrongful or excessive seizure; the person is entitled to the same relief that is provided by the Trademark Act of 1946 (15 U.S.C. Section 1116(d)(11)) concerning the wrongful seizure of goods and counterfeit trademarks (including damages for lost profits, cost of materials, loss of good will, punitive damages, and reasonable attorney's fees).¹⁵³

The DTSA would empower a court to offer civil remedies for trade secret misappropriation,¹⁵⁴ including injunctive relief, damages (for actual loss and any unjust enrichment caused by the misappropriation of the trade secret, or in lieu of damages measured by any other method, an award of a reasonable royalty), punitive damages of up to two times the amount of damages (if the trade secret is willfully and maliciously misappropriated),¹⁵⁵ and reasonable attorney's fees (if the claim of misappropriation is made in bad faith, there is willful and malicious misappropriation, or a motion to terminate an injunction is made or opposed in bad faith). The DTSA would require evidence of actual or threatened misappropriation before a court may issue an injunction to prevent it.¹⁵⁶ However, such an order for injunctive relief may not "prevent a person from entering into an employment relationship" or "otherwise conflict with an applicable State law prohibiting restraints on the practice of a lawful profession, trade, or business."¹⁵⁷ In

¹⁵⁰ *Id.* §2(a), adding new 18 U.S.C. §1836(b)(2)(A)(ii).

¹⁵¹ *Id.* §2(a), adding new 18 U.S.C. §1836(b)(2)(B).

¹⁵² *Id.* §2(a), adding new 18 U.S.C. §1836(b)(2)(D).

¹⁵³ *Id.* §2(a), adding new 18 U.S.C. §1836(b)(2)(G).

¹⁵⁴ *Id.* §2(a), adding new 18 U.S.C. §1836(b)(3).

¹⁵⁵ As introduced, S. 1890 would have authorized an exemplary damage award of up to three times the amount of compensatory damages. The Senate Judiciary Committee approved an amendment to the bill that limited punitive damages to two times compensatory damages.

¹⁵⁶ S. 1890, §2(a), adding new 18 U.S.C. §1836(b)(3)(A)(i).

¹⁵⁷ *Id.* §2(a), adding new 18 U.S.C. §1836(b)(3)(A)(i)(I), (II).

addition, “conditions placed on such employment shall be based on evidence of threatened misappropriation and not merely on the information the person knows.”¹⁵⁸

The DTSA would establish a three-year statute of limitations period for the misappropriation of trade secret civil action, which is similar to that provided by the Uniform Trade Secrets Act and under most state laws.¹⁵⁹ Finally, the DTSA includes a “rule of construction” provision¹⁶⁰ that declares that nothing in the DTSA shall be construed (1) to preempt any other provision of law or (2) to modify the EEA’s existing rule of construction (codified at 18 U.S.C. Section 1838) stating that the EEA does not preempt or displace any civil or criminal remedies provided by federal or state law for the misappropriation of a trade secret.

Author Information

Brian T. Yeh
Legislative Attorney

Disclaimer

This document was prepared by the Congressional Research Service (CRS). CRS serves as nonpartisan shared staff to congressional committees and Members of Congress. It operates solely at the behest of and under the direction of Congress. Information in a CRS Report should not be relied upon for purposes other than public understanding of information that has been provided by CRS to Members of Congress in connection with CRS’s institutional role. CRS Reports, as a work of the United States Government, are not subject to copyright protection in the United States. Any CRS Report may be reproduced and distributed in its entirety without permission from CRS. However, as a CRS Report may include copyrighted images or material from a third party, you may need to obtain the permission of the copyright holder if you wish to copy or otherwise use copyrighted material.

¹⁵⁸ *Id.* §2(a), adding new 18 U.S.C. §1836(b)(3)(A)(i)(I). These limitations were added to the bill by the Senate Judiciary Committee, in response to concerns that the original language of the bill would have negatively impacted the ability of an employee to take new jobs with other companies. *See* S.Rept. 114-220, at 8.

¹⁵⁹ As introduced, S. 1890 would have created a five-year limitations period. The Senate Judiciary Committee approved an amendment to the bill that reduced the time period to three years.

¹⁶⁰ S. 1890, §2(f).

ANNEX 1C

AGREEMENT ON TRADE-RELATED ASPECTS OF INTELLECTUAL PROPERTY RIGHTS (as amended on 23 January 2017)

PART I GENERAL PROVISIONS AND BASIC PRINCIPLES

PART II STANDARDS CONCERNING THE AVAILABILITY, SCOPE AND USE OF INTELLECTUAL PROPERTY RIGHTS

1. Copyright and Related Rights
2. Trademarks
3. Geographical Indications
4. Industrial Designs
5. Patents
6. Layout-Designs (Topographies) of Integrated Circuits
7. Protection of Undisclosed Information
8. Control of Anti-Competitive Practices in Contractual Licences

PART III ENFORCEMENT OF INTELLECTUAL PROPERTY RIGHTS

1. General Obligations
2. Civil and Administrative Procedures and Remedies
3. Provisional Measures
4. Special Requirements Related to Border Measures
5. Criminal Procedures

PART IV ACQUISITION AND MAINTENANCE OF INTELLECTUAL PROPERTY RIGHTS AND RELATED *INTER-PARTES* PROCEDURES

PART V DISPUTE PREVENTION AND SETTLEMENT

PART VI TRANSITIONAL ARRANGEMENTS

PART VII INSTITUTIONAL ARRANGEMENTS; FINAL PROVISIONS

ANNEX AND APPENDIX TO THE TRIPS AGREEMENT

AGREEMENT ON TRADE-RELATED ASPECTS OF INTELLECTUAL PROPERTY RIGHTS

Members,

Desiring to reduce distortions and impediments to international trade, and taking into account the need to promote effective and adequate protection of intellectual property rights, and to ensure that measures and procedures to enforce intellectual property rights do not themselves become barriers to legitimate trade;

Recognizing, to this end, the need for new rules and disciplines concerning:

- (a) the applicability of the basic principles of GATT 1994 and of relevant international intellectual property agreements or conventions;
- (b) the provision of adequate standards and principles concerning the availability, scope and use of trade-related intellectual property rights;
- (c) the provision of effective and appropriate means for the enforcement of trade-related intellectual property rights, taking into account differences in national legal systems;
- (d) the provision of effective and expeditious procedures for the multilateral prevention and settlement of disputes between governments; and
- (e) transitional arrangements aiming at the fullest participation in the results of the negotiations;

Recognizing the need for a multilateral framework of principles, rules and disciplines dealing with international trade in counterfeit goods;

Recognizing that intellectual property rights are private rights;

Recognizing the underlying public policy objectives of national systems for the protection of intellectual property, including developmental and technological objectives;

Recognizing also the special needs of the least-developed country Members in respect of maximum flexibility in the domestic implementation of laws and regulations in order to enable them to create a sound and viable technological base;

Emphasizing the importance of reducing tensions by reaching strengthened commitments to resolve disputes on trade-related intellectual property issues through multilateral procedures;

Desiring to establish a mutually supportive relationship between the WTO and the World Intellectual Property Organization (referred to in this Agreement as "WIPO") as well as other relevant international organizations;

Hereby agree as follows:

PART I

GENERAL PROVISIONS AND BASIC PRINCIPLES

Article 1

Nature and Scope of Obligations

1. Members shall give effect to the provisions of this Agreement. Members may, but shall not be obliged to, implement in their law more extensive protection than is required by this Agreement, provided that such protection does not contravene the provisions of this Agreement. Members shall be free to determine the appropriate method of implementing the provisions of this Agreement within their own legal system and practice.

2. For the purposes of this Agreement, the term "intellectual property" refers to all categories of intellectual property that are the subject of Sections 1 through 7 of Part II.

3. Members shall accord the treatment provided for in this Agreement to the nationals of other Members.¹ In respect of the relevant intellectual property right, the nationals of other Members shall be understood as those natural or legal persons that would meet the criteria for eligibility for protection provided for in the Paris Convention (1967), the Berne Convention (1971), the Rome Convention and the Treaty on Intellectual Property in Respect of Integrated Circuits, were all Members of the WTO members of those conventions.² Any Member availing itself of the possibilities provided in paragraph 3 of Article 5 or paragraph 2 of Article 6 of the Rome Convention shall make a notification as foreseen in those provisions to the Council for Trade-Related Aspects of Intellectual Property Rights (the "Council for TRIPS").

Article 2

Intellectual Property Conventions

1. In respect of Parts II, III and IV of this Agreement, Members shall comply with Articles 1 through 12, and Article 19, of the Paris Convention (1967).

2. Nothing in Parts I to IV of this Agreement shall derogate from existing obligations that Members may have to each other under the Paris Convention, the Berne Convention, the Rome Convention and the Treaty on Intellectual Property in Respect of Integrated Circuits.

Article 3

National Treatment

1. Each Member shall accord to the nationals of other Members treatment no less favourable than that it accords to its own nationals with regard to the protection³ of intellectual property, subject to the exceptions already provided in, respectively, the Paris Convention (1967), the Berne Convention (1971), the Rome Convention or the Treaty on Intellectual Property in Respect of Integrated Circuits. In respect of performers, producers of phonograms and broadcasting organizations, this obligation only applies in respect of the rights provided under this Agreement. Any Member availing itself of the possibilities provided in Article 6 of the Berne Convention (1971) or paragraph 1(b) of Article 16 of the Rome Convention shall make a notification as foreseen in those provisions to the Council for TRIPS.

¹ When "nationals" are referred to in this Agreement, they shall be deemed, in the case of a separate customs territory Member of the WTO, to mean persons, natural or legal, who are domiciled or who have a real and effective industrial or commercial establishment in that customs territory.

² In this Agreement, "Paris Convention" refers to the Paris Convention for the Protection of Industrial Property; "Paris Convention (1967)" refers to the Stockholm Act of this Convention of 14 July 1967. "Berne Convention" refers to the Berne Convention for the Protection of Literary and Artistic Works; "Berne Convention (1971)" refers to the Paris Act of this Convention of 24 July 1971. "Rome Convention" refers to the International Convention for the Protection of Performers, Producers of Phonograms and Broadcasting Organizations, adopted at Rome on 26 October 1961. "Treaty on Intellectual Property in Respect of Integrated Circuits" (IPIC Treaty) refers to the Treaty on Intellectual Property in Respect of Integrated Circuits, adopted at Washington on 26 May 1989. "WTO Agreement" refers to the Agreement Establishing the WTO.

³ For the purposes of Articles 3 and 4, "protection" shall include matters affecting the availability, acquisition, scope, maintenance and enforcement of intellectual property rights as well as those matters affecting the use of intellectual property rights specifically addressed in this Agreement.

2. Members may avail themselves of the exceptions permitted under paragraph 1 in relation to judicial and administrative procedures, including the designation of an address for service or the appointment of an agent within the jurisdiction of a Member, only where such exceptions are necessary to secure compliance with laws and regulations which are not inconsistent with the provisions of this Agreement and where such practices are not applied in a manner which would constitute a disguised restriction on trade.

Article 4

Most-Favoured-Nation Treatment

With regard to the protection of intellectual property, any advantage, favour, privilege or immunity granted by a Member to the nationals of any other country shall be accorded immediately and unconditionally to the nationals of all other Members. Exempted from this obligation are any advantage, favour, privilege or immunity accorded by a Member:

- (a) deriving from international agreements on judicial assistance or law enforcement of a general nature and not particularly confined to the protection of intellectual property;
- (b) granted in accordance with the provisions of the Berne Convention (1971) or the Rome Convention authorizing that the treatment accorded be a function not of national treatment but of the treatment accorded in another country;
- (c) in respect of the rights of performers, producers of phonograms and broadcasting organizations not provided under this Agreement;
- (d) deriving from international agreements related to the protection of intellectual property which entered into force prior to the entry into force of the WTO Agreement, provided that such agreements are notified to the Council for TRIPS and do not constitute an arbitrary or unjustifiable discrimination against nationals of other Members.

Article 5

Multilateral Agreements on Acquisition or Maintenance of Protection

The obligations under Articles 3 and 4 do not apply to procedures provided in multilateral agreements concluded under the auspices of WIPO relating to the acquisition or maintenance of intellectual property rights.

Article 6

Exhaustion

For the purposes of dispute settlement under this Agreement, subject to the provisions of Articles 3 and 4 nothing in this Agreement shall be used to address the issue of the exhaustion of intellectual property rights.

Article 7

Objectives

The protection and enforcement of intellectual property rights should contribute to the promotion of technological innovation and to the transfer and dissemination of technology, to the mutual advantage of producers and users of technological knowledge and in a manner conducive to social and economic welfare, and to a balance of rights and obligations.

Article 8

Principles

1. Members may, in formulating or amending their laws and regulations, adopt measures necessary to protect public health and nutrition, and to promote the public interest in sectors of vital importance to their socio-economic and technological development, provided that such measures are consistent with the provisions of this Agreement.
2. Appropriate measures, provided that they are consistent with the provisions of this Agreement, may be needed to prevent the abuse of intellectual property rights by right holders or the resort to practices which unreasonably restrain trade or adversely affect the international transfer of technology.

PART II

STANDARDS CONCERNING THE AVAILABILITY, SCOPE AND USE OF INTELLECTUAL PROPERTY RIGHTS

SECTION 1: COPYRIGHT AND RELATED RIGHTS

Article 9

Relation to the Berne Convention

1. Members shall comply with Articles 1 through 21 of the Berne Convention (1971) and the Appendix thereto. However, Members shall not have rights or obligations under this Agreement in respect of the rights conferred under Article 6bis of that Convention or of the rights derived therefrom.
2. Copyright protection shall extend to expressions and not to ideas, procedures, methods of operation or mathematical concepts as such.

Article 10

Computer Programs and Compilations of Data

1. Computer programs, whether in source or object code, shall be protected as literary works under the Berne Convention (1971).

2. Compilations of data or other material, whether in machine readable or other form, which by reason of the selection or arrangement of their contents constitute intellectual creations shall be protected as such. Such protection, which shall not extend to the data or material itself, shall be without prejudice to any copyright subsisting in the data or material itself.

Article 11

Rental Rights

In respect of at least computer programs and cinematographic works, a Member shall provide authors and their successors in title the right to authorize or to prohibit the commercial rental to the public of originals or copies of their copyright works. A Member shall be excepted from this obligation in respect of cinematographic works unless such rental has led to widespread copying of such works which is materially impairing the exclusive right of reproduction conferred in that Member on authors and their successors in title. In respect of computer programs, this obligation does not apply to rentals where the program itself is not the essential object of the rental.

Article 12

Term of Protection

Whenever the term of protection of a work, other than a photographic work or a work of applied art, is calculated on a basis other than the life of a natural person, such term shall be no less than 50 years from the end of the calendar year of authorized publication, or, failing such authorized publication within 50 years from the making of the work, 50 years from the end of the calendar year of making.

Article 13

Limitations and Exceptions

Members shall confine limitations or exceptions to exclusive rights to certain special cases which do not conflict with a normal exploitation of the work and do not unreasonably prejudice the legitimate interests of the right holder.

Article 14

Protection of Performers, Producers of Phonograms (Sound Recordings) and Broadcasting Organizations

1. In respect of a fixation of their performance on a phonogram, performers shall have the possibility of preventing the following acts when undertaken without their authorization: the fixation of their unfixed performance and the reproduction of such fixation. Performers shall also have the possibility of preventing the following acts when undertaken without their authorization: the broadcasting by wireless means and the communication to the public of their live performance.

2. Producers of phonograms shall enjoy the right to authorize or prohibit the direct or indirect reproduction of their phonograms.

3. Broadcasting organizations shall have the right to prohibit the following acts when undertaken without their authorization: the fixation, the reproduction of fixations, and the rebroadcasting by wireless means of broadcasts, as well as the communication to the public of television broadcasts of the same. Where Members do not grant such rights to broadcasting organizations, they shall provide owners of copyright in the subject matter of broadcasts with the possibility of preventing the above acts, subject to the provisions of the Berne Convention (1971).

4. The provisions of Article 11 in respect of computer programs shall apply *mutatis mutandis* to producers of phonograms and any other right holders in phonograms as determined in a Member's law. If on 15 April 1994 a Member has in force a system of equitable remuneration of right holders in respect of the rental of phonograms, it may maintain such system provided that the commercial rental of phonograms is not giving rise to the material impairment of the exclusive rights of reproduction of right holders.

5. The term of the protection available under this Agreement to performers and producers of phonograms shall last at least until the end of a period of 50 years computed from the end of the calendar year in which the fixation was made or the performance took place. The term of protection granted pursuant to paragraph 3 shall last for at least 20 years from the end of the calendar year in which the broadcast took place.

6. Any Member may, in relation to the rights conferred under paragraphs 1, 2 and 3, provide for conditions, limitations, exceptions and reservations to the extent permitted by the Rome Convention. However, the provisions of Article 18 of the Berne Convention (1971) shall also apply, *mutatis mutandis*, to the rights of performers and producers of phonograms in phonograms.

SECTION 2: TRADEMARKS

Article 15

Protectable Subject Matter

1. Any sign, or any combination of signs, capable of distinguishing the goods or services of one undertaking from those of other undertakings, shall be capable of constituting a trademark. Such signs, in particular words including personal names, letters, numerals, figurative elements and combinations of colours as well as any combination of such signs, shall be eligible for registration as trademarks. Where signs are not inherently capable of distinguishing the relevant goods or services, Members may make registrability depend on distinctiveness acquired through use. Members may require, as a condition of registration, that signs be visually perceptible.

2. Paragraph 1 shall not be understood to prevent a Member from denying registration of a trademark on other grounds, provided that they do not derogate from the provisions of the Paris Convention (1967).

3. Members may make registrability depend on use. However, actual use of a trademark shall not be a condition for filing an application for registration. An application shall not be refused solely on the ground that intended use has not taken place before the expiry of a period of three years from the date of application.

4. The nature of the goods or services to which a trademark is to be applied shall in no case form an obstacle to registration of the trademark.

5. Members shall publish each trademark either before it is registered or promptly after it is registered and shall afford a reasonable opportunity for petitions to cancel the registration. In addition, Members may afford an opportunity for the registration of a trademark to be opposed.

Article 16

Rights Conferred

1. The owner of a registered trademark shall have the exclusive right to prevent all third parties not having the owner's consent from using in the course of trade identical or similar signs for goods or services which are identical or similar to those in respect of which the trademark is registered where such use would result in a likelihood of confusion. In case of the use of an identical sign for identical goods or services, a likelihood of confusion shall be presumed. The rights described above shall not prejudice any existing prior rights, nor shall they affect the possibility of Members making rights available on the basis of use.

2. Article 6bis of the Paris Convention (1967) shall apply, *mutatis mutandis*, to services. In determining whether a trademark is well-known, Members shall take account of the knowledge of the trademark in the relevant sector of the public, including knowledge in the Member concerned which has been obtained as a result of the promotion of the trademark.

3. Article 6bis of the Paris Convention (1967) shall apply, *mutatis mutandis*, to goods or services which are not similar to those in respect of which a trademark is registered, provided that use of that trademark in relation to those goods or services would indicate a connection between those goods or services and the owner of the registered trademark and provided that the interests of the owner of the registered trademark are likely to be damaged by such use.

Article 17

Exceptions

Members may provide limited exceptions to the rights conferred by a trademark, such as fair use of descriptive terms, provided that such exceptions take account of the legitimate interests of the owner of the trademark and of third parties.

Article 18

Term of Protection

Initial registration, and each renewal of registration, of a trademark shall be for a term of no less than seven years. The registration of a trademark shall be renewable indefinitely.

Article 19

Requirement of Use

1. If use is required to maintain a registration, the registration may be cancelled only after an uninterrupted period of at least three years of non-use, unless valid reasons based on the existence of obstacles to such use are shown by the trademark owner. Circumstances arising independently of the

will of the owner of the trademark which constitute an obstacle to the use of the trademark, such as import restrictions on or other government requirements for goods or services protected by the trademark, shall be recognized as valid reasons for non-use.

2. When subject to the control of its owner, use of a trademark by another person shall be recognized as use of the trademark for the purpose of maintaining the registration.

Article 20

Other Requirements

The use of a trademark in the course of trade shall not be unjustifiably encumbered by special requirements, such as use with another trademark, use in a special form or use in a manner detrimental to its capability to distinguish the goods or services of one undertaking from those of other undertakings. This will not preclude a requirement prescribing the use of the trademark identifying the undertaking producing the goods or services along with, but without linking it to, the trademark distinguishing the specific goods or services in question of that undertaking.

Article 21

Licensing and Assignment

Members may determine conditions on the licensing and assignment of trademarks, it being understood that the compulsory licensing of trademarks shall not be permitted and that the owner of a registered trademark shall have the right to assign the trademark with or without the transfer of the business to which the trademark belongs.

SECTION 3: GEOGRAPHICAL INDICATIONS

Article 22

Protection of Geographical Indications

1. Geographical indications are, for the purposes of this Agreement, indications which identify a good as originating in the territory of a Member, or a region or locality in that territory, where a given quality, reputation or other characteristic of the good is essentially attributable to its geographical origin.

2. In respect of geographical indications, Members shall provide the legal means for interested parties to prevent:

- (a) the use of any means in the designation or presentation of a good that indicates or suggests that the good in question originates in a geographical area other than the true place of origin in a manner which misleads the public as to the geographical origin of the good;
- (b) any use which constitutes an act of unfair competition within the meaning of Article 10*bis* of the Paris Convention (1967).

3. A Member shall, *ex officio* if its legislation so permits or at the request of an interested party, refuse or invalidate the registration of a trademark which contains or consists of a geographical indication with respect to goods not originating in the territory indicated, if use of the indication in the trademark for such goods in that Member is of such a nature as to mislead the public as to the true place of origin.

4. The protection under paragraphs 1, 2 and 3 shall be applicable against a geographical indication which, although literally true as to the territory, region or locality in which the goods originate, falsely represents to the public that the goods originate in another territory.

Article 23

Additional Protection for Geographical Indications for Wines and Spirits

1. Each Member shall provide the legal means for interested parties to prevent use of a geographical indication identifying wines for wines not originating in the place indicated by the geographical indication in question or identifying spirits for spirits not originating in the place indicated by the geographical indication in question, even where the true origin of the goods is indicated or the geographical indication is used in translation or accompanied by expressions such as "kind", "type", "style", "imitation" or the like.⁴

2. The registration of a trademark for wines which contains or consists of a geographical indication identifying wines or for spirits which contains or consists of a geographical indication identifying spirits shall be refused or invalidated, *ex officio* if a Member's legislation so permits or at the request of an interested party, with respect to such wines or spirits not having this origin.

3. In the case of homonymous geographical indications for wines, protection shall be accorded to each indication, subject to the provisions of paragraph 4 of Article 22. Each Member shall determine the practical conditions under which the homonymous indications in question will be differentiated from each other, taking into account the need to ensure equitable treatment of the producers concerned and that consumers are not misled.

4. In order to facilitate the protection of geographical indications for wines, negotiations shall be undertaken in the Council for TRIPS concerning the establishment of a multilateral system of notification and registration of geographical indications for wines eligible for protection in those Members participating in the system.

Article 24

International Negotiations; Exceptions

1. Members agree to enter into negotiations aimed at increasing the protection of individual geographical indications under Article 23. The provisions of paragraphs 4 through 8 below shall not be used by a Member to refuse to conduct negotiations or to conclude bilateral or multilateral agreements. In the context of such negotiations, Members shall be willing to consider the continued applicability of these provisions to individual geographical indications whose use was the subject of such negotiations.

⁴ Notwithstanding the first sentence of Article 42, Members may, with respect to these obligations, instead provide for enforcement by administrative action.

2. The Council for TRIPS shall keep under review the application of the provisions of this Section; the first such review shall take place within two years of the entry into force of the WTO Agreement. Any matter affecting the compliance with the obligations under these provisions may be drawn to the attention of the Council, which, at the request of a Member, shall consult with any Member or Members in respect of such matter in respect of which it has not been possible to find a satisfactory solution through bilateral or plurilateral consultations between the Members concerned. The Council shall take such action as may be agreed to facilitate the operation and further the objectives of this Section.

3. In implementing this Section, a Member shall not diminish the protection of geographical indications that existed in that Member immediately prior to the date of entry into force of the WTO Agreement.

4. Nothing in this Section shall require a Member to prevent continued and similar use of a particular geographical indication of another Member identifying wines or spirits in connection with goods or services by any of its nationals or domiciliaries who have used that geographical indication in a continuous manner with regard to the same or related goods or services in the territory of that Member either (a) for at least 10 years preceding 15 April 1994 or (b) in good faith preceding that date.

5. Where a trademark has been applied for or registered in good faith, or where rights to a trademark have been acquired through use in good faith either:

- (a) before the date of application of these provisions in that Member as defined in Part VI; or
- (b) before the geographical indication is protected in its country of origin;

measures adopted to implement this Section shall not prejudice eligibility for or the validity of the registration of a trademark, or the right to use a trademark, on the basis that such a trademark is identical with, or similar to, a geographical indication.

6. Nothing in this Section shall require a Member to apply its provisions in respect of a geographical indication of any other Member with respect to goods or services for which the relevant indication is identical with the term customary in common language as the common name for such goods or services in the territory of that Member. Nothing in this Section shall require a Member to apply its provisions in respect of a geographical indication of any other Member with respect to products of the vine for which the relevant indication is identical with the customary name of a grape variety existing in the territory of that Member as of the date of entry into force of the WTO Agreement.

7. A Member may provide that any request made under this Section in connection with the use or registration of a trademark must be presented within five years after the adverse use of the protected indication has become generally known in that Member or after the date of registration of the trademark in that Member provided that the trademark has been published by that date, if such date is earlier than the date on which the adverse use became generally known in that Member, provided that the geographical indication is not used or registered in bad faith.

8. The provisions of this Section shall in no way prejudice the right of any person to use, in the course of trade, that person's name or the name of that person's predecessor in business, except where such name is used in such a manner as to mislead the public.

9. There shall be no obligation under this Agreement to protect geographical indications which are not or cease to be protected in their country of origin, or which have fallen into disuse in that country.

SECTION 4: INDUSTRIAL DESIGNS

Article 25

Requirements for Protection

1. Members shall provide for the protection of independently created industrial designs that are new or original. Members may provide that designs are not new or original if they do not significantly differ from known designs or combinations of known design features. Members may provide that such protection shall not extend to designs dictated essentially by technical or functional considerations.

2. Each Member shall ensure that requirements for securing protection for textile designs, in particular in regard to any cost, examination or publication, do not unreasonably impair the opportunity to seek and obtain such protection. Members shall be free to meet this obligation through industrial design law or through copyright law.

Article 26

Protection

1. The owner of a protected industrial design shall have the right to prevent third parties not having the owner's consent from making, selling or importing articles bearing or embodying a design which is a copy, or substantially a copy, of the protected design, when such acts are undertaken for commercial purposes.

2. Members may provide limited exceptions to the protection of industrial designs, provided that such exceptions do not unreasonably conflict with the normal exploitation of protected industrial designs and do not unreasonably prejudice the legitimate interests of the owner of the protected design, taking account of the legitimate interests of third parties.

3. The duration of protection available shall amount to at least 10 years.

SECTION 5: PATENTS

Article 27

Patentable Subject Matter

1. Subject to the provisions of paragraphs 2 and 3, patents shall be available for any inventions, whether products or processes, in all fields of technology, provided that they are new, involve an inventive step and are capable of industrial application.⁵ Subject to paragraph 4 of Article 65,

⁵ For the purposes of this Article, the terms "inventive step" and "capable of industrial application" may be deemed by a Member to be synonymous with the terms "non-obvious" and "useful" respectively.

paragraph 8 of Article 70 and paragraph 3 of this Article, patents shall be available and patent rights enjoyable without discrimination as to the place of invention, the field of technology and whether products are imported or locally produced.

2. Members may exclude from patentability inventions, the prevention within their territory of the commercial exploitation of which is necessary to protect *ordre public* or morality, including to protect human, animal or plant life or health or to avoid serious prejudice to the environment, provided that such exclusion is not made merely because the exploitation is prohibited by their law.

3. Members may also exclude from patentability:

- (a) diagnostic, therapeutic and surgical methods for the treatment of humans or animals;
- (b) plants and animals other than micro-organisms, and essentially biological processes for the production of plants or animals other than non-biological and microbiological processes. However, Members shall provide for the protection of plant varieties either by patents or by an effective *sui generis* system or by any combination thereof. The provisions of this subparagraph shall be reviewed four years after the date of entry into force of the WTO Agreement.

Article 28

Rights Conferred

1. A patent shall confer on its owner the following exclusive rights:

- (a) where the subject matter of a patent is a product, to prevent third parties not having the owner's consent from the acts of: making, using, offering for sale, selling, or importing⁶ for these purposes that product;
- (b) where the subject matter of a patent is a process, to prevent third parties not having the owner's consent from the act of using the process, and from the acts of: using, offering for sale, selling, or importing for these purposes at least the product obtained directly by that process.

2. Patent owners shall also have the right to assign, or transfer by succession, the patent and to conclude licensing contracts.

Article 29

Conditions on Patent Applicants

1. Members shall require that an applicant for a patent shall disclose the invention in a manner sufficiently clear and complete for the invention to be carried out by a person skilled in the art and may require the applicant to indicate the best mode for carrying out the invention known to the inventor at the filing date or, where priority is claimed, at the priority date of the application.

2. Members may require an applicant for a patent to provide information concerning the applicant's corresponding foreign applications and grants.

⁶ This right, like all other rights conferred under this Agreement in respect of the use, sale, importation or other distribution of goods, is subject to the provisions of Article 6.

Article 30

Exceptions to Rights Conferred

Members may provide limited exceptions to the exclusive rights conferred by a patent, provided that such exceptions do not unreasonably conflict with a normal exploitation of the patent and do not unreasonably prejudice the legitimate interests of the patent owner, taking account of the legitimate interests of third parties.

Article 31

Other Use Without Authorization of the Right Holder

Where the law of a Member allows for other use⁷ of the subject matter of a patent without the authorization of the right holder, including use by the government or third parties authorized by the government, the following provisions shall be respected:

- (a) authorization of such use shall be considered on its individual merits;
- (b) such use may only be permitted if, prior to such use, the proposed user has made efforts to obtain authorization from the right holder on reasonable commercial terms and conditions and that such efforts have not been successful within a reasonable period of time. This requirement may be waived by a Member in the case of a national emergency or other circumstances of extreme urgency or in cases of public non-commercial use. In situations of national emergency or other circumstances of extreme urgency, the right holder shall, nevertheless, be notified as soon as reasonably practicable. In the case of public non-commercial use, where the government or contractor, without making a patent search, knows or has demonstrable grounds to know that a valid patent is or will be used by or for the government, the right holder shall be informed promptly;
- (c) the scope and duration of such use shall be limited to the purpose for which it was authorized, and in the case of semi-conductor technology shall only be for public non-commercial use or to remedy a practice determined after judicial or administrative process to be anti-competitive;
- (d) such use shall be non-exclusive;
- (e) such use shall be non-assignable, except with that part of the enterprise or goodwill which enjoys such use;
- (f) any such use shall be authorized predominantly for the supply of the domestic market of the Member authorizing such use;
- (g) authorization for such use shall be liable, subject to adequate protection of the legitimate interests of the persons so authorized, to be terminated if and when the circumstances which led to it cease to exist and are unlikely to recur. The competent authority shall have the authority to review, upon motivated request, the continued existence of these circumstances;

⁷ "Other use" refers to use other than that allowed under Article 30.

- (h) the right holder shall be paid adequate remuneration in the circumstances of each case, taking into account the economic value of the authorization;
- (i) the legal validity of any decision relating to the authorization of such use shall be subject to judicial review or other independent review by a distinct higher authority in that Member;
- (j) any decision relating to the remuneration provided in respect of such use shall be subject to judicial review or other independent review by a distinct higher authority in that Member;
- (k) Members are not obliged to apply the conditions set forth in subparagraphs (b) and (f) where such use is permitted to remedy a practice determined after judicial or administrative process to be anti-competitive. The need to correct anti-competitive practices may be taken into account in determining the amount of remuneration in such cases. Competent authorities shall have the authority to refuse termination of authorization if and when the conditions which led to such authorization are likely to recur;
- (l) where such use is authorized to permit the exploitation of a patent ("the second patent") which cannot be exploited without infringing another patent ("the first patent"), the following additional conditions shall apply:
 - (i) the invention claimed in the second patent shall involve an important technical advance of considerable economic significance in relation to the invention claimed in the first patent;
 - (ii) the owner of the first patent shall be entitled to a cross-licence on reasonable terms to use the invention claimed in the second patent; and
 - (iii) the use authorized in respect of the first patent shall be non-assignable except with the assignment of the second patent.

Article 31bis

1. The obligations of an exporting Member under Article 31(f) shall not apply with respect to the grant by it of a compulsory licence to the extent necessary for the purposes of production of a pharmaceutical product(s) and its export to an eligible importing Member(s) in accordance with the terms set out in paragraph 2 of the Annex to this Agreement.

2. Where a compulsory licence is granted by an exporting Member under the system set out in this Article and the Annex to this Agreement, adequate remuneration pursuant to Article 31(h) shall be paid in that Member taking into account the economic value to the importing Member of the use that has been authorized in the exporting Member. Where a compulsory licence is granted for the same products in the eligible importing Member, the obligation of that Member under Article 31(h) shall not apply in respect of those products for which remuneration in accordance with the first sentence of this paragraph is paid in the exporting Member.

3. With a view to harnessing economies of scale for the purposes of enhancing purchasing power for, and facilitating the local production of, pharmaceutical products: where a developing or

least-developed country WTO Member is a party to a regional trade agreement within the meaning of Article XXIV of the GATT 1994 and the Decision of 28 November 1979 on Differential and More Favourable Treatment Reciprocity and Fuller Participation of Developing Countries (L/4903), at least half of the current membership of which is made up of countries presently on the United Nations list of least-developed countries, the obligation of that Member under Article 31(f) shall not apply to the extent necessary to enable a pharmaceutical product produced or imported under a compulsory licence in that Member to be exported to the markets of those other developing or least-developed country parties to the regional trade agreement that share the health problem in question. It is understood that this will not prejudice the territorial nature of the patent rights in question.

4. Members shall not challenge any measures taken in conformity with the provisions of this Article and the Annex to this Agreement under subparagraphs 1(b) and 1(c) of Article XXIII of GATT 1994.

5. This Article and the Annex to this Agreement are without prejudice to the rights, obligations and flexibilities that Members have under the provisions of this Agreement other than paragraphs (f) and (h) of Article 31, including those reaffirmed by the Declaration on the TRIPS Agreement and Public Health (WT/MIN(01)/DEC/2), and to their interpretation. They are also without prejudice to the extent to which pharmaceutical products produced under a compulsory licence can be exported under the provisions of Article 31(f).

Article 32

Revocation/Forfeiture

An opportunity for judicial review of any decision to revoke or forfeit a patent shall be available.

Article 33

Term of Protection

The term of protection available shall not end before the expiration of a period of twenty years counted from the filing date.⁸

Article 34

Process Patents: Burden of Proof

1. For the purposes of civil proceedings in respect of the infringement of the rights of the owner referred to in paragraph 1(b) of Article 28, if the subject matter of a patent is a process for obtaining a product, the judicial authorities shall have the authority to order the defendant to prove that the process to obtain an identical product is different from the patented process. Therefore, Members shall provide, in at least one of the following circumstances, that any identical product when produced without the consent of the patent owner shall, in the absence of proof to the contrary, be deemed to have been obtained by the patented process:

- (a) if the product obtained by the patented process is new;

⁸ It is understood that those Members which do not have a system of original grant may provide that the term of protection shall be computed from the filing date in the system of original grant.

- (b) if there is a substantial likelihood that the identical product was made by the process and the owner of the patent has been unable through reasonable efforts to determine the process actually used.

2. Any Member shall be free to provide that the burden of proof indicated in paragraph 1 shall be on the alleged infringer only if the condition referred to in subparagraph (a) is fulfilled or only if the condition referred to in subparagraph (b) is fulfilled.

3. In the adduction of proof to the contrary, the legitimate interests of defendants in protecting their manufacturing and business secrets shall be taken into account.

SECTION 6: LAYOUT-DESIGNS (TOPOGRAPHIES) OF INTEGRATED CIRCUITS

Article 35

Relation to the IPIC Treaty

Members agree to provide protection to the layout-designs (topographies) of integrated circuits (referred to in this Agreement as "layout-designs") in accordance with Articles 2 through 7 (other than paragraph 3 of Article 6), Article 12 and paragraph 3 of Article 16 of the Treaty on Intellectual Property in Respect of Integrated Circuits and, in addition, to comply with the following provisions.

Article 36

Scope of the Protection

Subject to the provisions of paragraph 1 of Article 37, Members shall consider unlawful the following acts if performed without the authorization of the right holder:⁹ importing, selling, or otherwise distributing for commercial purposes a protected layout-design, an integrated circuit in which a protected layout-design is incorporated, or an article incorporating such an integrated circuit only in so far as it continues to contain an unlawfully reproduced layout-design.

Article 37

Acts Not Requiring the Authorization of the Right Holder

1. Notwithstanding Article 36, no Member shall consider unlawful the performance of any of the acts referred to in that Article in respect of an integrated circuit incorporating an unlawfully reproduced layout-design or any article incorporating such an integrated circuit where the person performing or ordering such acts did not know and had no reasonable ground to know, when acquiring the integrated circuit or article incorporating such an integrated circuit, that it incorporated an unlawfully reproduced layout-design. Members shall provide that, after the time that such person has received sufficient notice that the layout-design was unlawfully reproduced, that person may perform any of the acts with respect to the stock on hand or ordered before such time, but shall be

⁹ The term "right holder" in this Section shall be understood as having the same meaning as the term "holder of the right" in the IPIC Treaty.

liable to pay to the right holder a sum equivalent to a reasonable royalty such as would be payable under a freely negotiated licence in respect of such a layout-design.

2. The conditions set out in subparagraphs (a) through (k) of Article 31 shall apply *mutatis mutandis* in the event of any non-voluntary licensing of a layout-design or of its use by or for the government without the authorization of the right holder.

Article 38

Term of Protection

1. In Members requiring registration as a condition of protection, the term of protection of layout-designs shall not end before the expiration of a period of 10 years counted from the date of filing an application for registration or from the first commercial exploitation wherever in the world it occurs.

2. In Members not requiring registration as a condition for protection, layout-designs shall be protected for a term of no less than 10 years from the date of the first commercial exploitation wherever in the world it occurs.

3. Notwithstanding paragraphs 1 and 2, a Member may provide that protection shall lapse 15 years after the creation of the layout-design.

SECTION 7: PROTECTION OF UNDISCLOSED INFORMATION

Article 39

1. In the course of ensuring effective protection against unfair competition as provided in Article 10*bis* of the Paris Convention (1967), Members shall protect undisclosed information in accordance with paragraph 2 and data submitted to governments or governmental agencies in accordance with paragraph 3.

2. Natural and legal persons shall have the possibility of preventing information lawfully within their control from being disclosed to, acquired by, or used by others without their consent in a manner contrary to honest commercial practices¹⁰ so long as such information:

- (a) is secret in the sense that it is not, as a body or in the precise configuration and assembly of its components, generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question;
- (b) has commercial value because it is secret; and
- (c) has been subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep it secret.

¹⁰ For the purpose of this provision, "a manner contrary to honest commercial practices" shall mean at least practices such as breach of contract, breach of confidence and inducement to breach, and includes the acquisition of undisclosed information by third parties who knew, or were grossly negligent in failing to know, that such practices were involved in the acquisition.

3. Members, when requiring, as a condition of approving the marketing of pharmaceutical or of agricultural chemical products which utilize new chemical entities, the submission of undisclosed test or other data, the origination of which involves a considerable effort, shall protect such data against unfair commercial use. In addition, Members shall protect such data against disclosure, except where necessary to protect the public, or unless steps are taken to ensure that the data are protected against unfair commercial use.

SECTION 8: CONTROL OF ANTI-COMPETITIVE PRACTICES IN CONTRACTUAL LICENCES

Article 40

1. Members agree that some licensing practices or conditions pertaining to intellectual property rights which restrain competition may have adverse effects on trade and may impede the transfer and dissemination of technology.

2. Nothing in this Agreement shall prevent Members from specifying in their legislation licensing practices or conditions that may in particular cases constitute an abuse of intellectual property rights having an adverse effect on competition in the relevant market. As provided above, a Member may adopt, consistently with the other provisions of this Agreement, appropriate measures to prevent or control such practices, which may include for example exclusive grantback conditions, conditions preventing challenges to validity and coercive package licensing, in the light of the relevant laws and regulations of that Member.

3. Each Member shall enter, upon request, into consultations with any other Member which has cause to believe that an intellectual property right owner that is a national or domiciliary of the Member to which the request for consultations has been addressed is undertaking practices in violation of the requesting Member's laws and regulations on the subject matter of this Section, and which wishes to secure compliance with such legislation, without prejudice to any action under the law and to the full freedom of an ultimate decision of either Member. The Member addressed shall accord full and sympathetic consideration to, and shall afford adequate opportunity for, consultations with the requesting Member, and shall cooperate through supply of publicly available non-confidential information of relevance to the matter in question and of other information available to the Member, subject to domestic law and to the conclusion of mutually satisfactory agreements concerning the safeguarding of its confidentiality by the requesting Member.

4. A Member whose nationals or domiciliaries are subject to proceedings in another Member concerning alleged violation of that other Member's laws and regulations on the subject matter of this Section shall, upon request, be granted an opportunity for consultations by the other Member under the same conditions as those foreseen in paragraph 3.

PART III

ENFORCEMENT OF INTELLECTUAL PROPERTY RIGHTS

SECTION 1: GENERAL OBLIGATIONS

Article 41

1. Members shall ensure that enforcement procedures as specified in this Part are available under their law so as to permit effective action against any act of infringement of intellectual property rights covered by this Agreement, including expeditious remedies to prevent infringements and remedies which constitute a deterrent to further infringements. These procedures shall be applied in such a manner as to avoid the creation of barriers to legitimate trade and to provide for safeguards against their abuse.
2. Procedures concerning the enforcement of intellectual property rights shall be fair and equitable. They shall not be unnecessarily complicated or costly, or entail unreasonable time-limits or unwarranted delays.
3. Decisions on the merits of a case shall preferably be in writing and reasoned. They shall be made available at least to the parties to the proceeding without undue delay. Decisions on the merits of a case shall be based only on evidence in respect of which parties were offered the opportunity to be heard.
4. Parties to a proceeding shall have an opportunity for review by a judicial authority of final administrative decisions and, subject to jurisdictional provisions in a Member's law concerning the importance of a case, of at least the legal aspects of initial judicial decisions on the merits of a case. However, there shall be no obligation to provide an opportunity for review of acquittals in criminal cases.
5. It is understood that this Part does not create any obligation to put in place a judicial system for the enforcement of intellectual property rights distinct from that for the enforcement of law in general, nor does it affect the capacity of Members to enforce their law in general. Nothing in this Part creates any obligation with respect to the distribution of resources as between enforcement of intellectual property rights and the enforcement of law in general.

SECTION 2: CIVIL AND ADMINISTRATIVE PROCEDURES AND REMEDIES

Article 42

Fair and Equitable Procedures

Members shall make available to right holders¹¹ civil judicial procedures concerning the enforcement of any intellectual property right covered by this Agreement. Defendants shall have the right to written notice which is timely and contains sufficient detail, including the basis of the claims. Parties shall be allowed to be represented by independent legal counsel, and procedures shall not impose overly burdensome requirements concerning mandatory personal appearances. All parties to such procedures shall be duly entitled to substantiate their claims and to present all relevant evidence. The procedure shall provide a means to identify and protect confidential information, unless this would be contrary to existing constitutional requirements.

Article 43

Evidence

¹¹ For the purpose of this Part, the term "right holder" includes federations and associations having legal standing to assert such rights.

1. The judicial authorities shall have the authority, where a party has presented reasonably available evidence sufficient to support its claims and has specified evidence relevant to substantiation of its claims which lies in the control of the opposing party, to order that this evidence be produced by the opposing party, subject in appropriate cases to conditions which ensure the protection of confidential information.

2. In cases in which a party to a proceeding voluntarily and without good reason refuses access to, or otherwise does not provide necessary information within a reasonable period, or significantly impedes a procedure relating to an enforcement action, a Member may accord judicial authorities the authority to make preliminary and final determinations, affirmative or negative, on the basis of the information presented to them, including the complaint or the allegation presented by the party adversely affected by the denial of access to information, subject to providing the parties an opportunity to be heard on the allegations or evidence.

Article 44

Injunctions

1. The judicial authorities shall have the authority to order a party to desist from an infringement, *inter alia* to prevent the entry into the channels of commerce in their jurisdiction of imported goods that involve the infringement of an intellectual property right, immediately after customs clearance of such goods. Members are not obliged to accord such authority in respect of protected subject matter acquired or ordered by a person prior to knowing or having reasonable grounds to know that dealing in such subject matter would entail the infringement of an intellectual property right.

2. Notwithstanding the other provisions of this Part and provided that the provisions of Part II specifically addressing use by governments, or by third parties authorized by a government, without the authorization of the right holder are complied with, Members may limit the remedies available against such use to payment of remuneration in accordance with subparagraph (h) of Article 31. In other cases, the remedies under this Part shall apply or, where these remedies are inconsistent with a Member's law, declaratory judgments and adequate compensation shall be available.

Article 45

Damages

1. The judicial authorities shall have the authority to order the infringer to pay the right holder damages adequate to compensate for the injury the right holder has suffered because of an infringement of that person's intellectual property right by an infringer who knowingly, or with reasonable grounds to know, engaged in infringing activity.

2. The judicial authorities shall also have the authority to order the infringer to pay the right holder expenses, which may include appropriate attorney's fees. In appropriate cases, Members may authorize the judicial authorities to order recovery of profits and/or payment of pre-established damages even where the infringer did not knowingly, or with reasonable grounds to know, engage in infringing activity.

Article 46

Other Remedies

In order to create an effective deterrent to infringement, the judicial authorities shall have the authority to order that goods that they have found to be infringing be, without compensation of any sort, disposed of outside the channels of commerce in such a manner as to avoid any harm caused to the right holder, or, unless this would be contrary to existing constitutional requirements, destroyed. The judicial authorities shall also have the authority to order that materials and implements the predominant use of which has been in the creation of the infringing goods be, without compensation of any sort, disposed of outside the channels of commerce in such a manner as to minimize the risks of further infringements. In considering such requests, the need for proportionality between the seriousness of the infringement and the remedies ordered as well as the interests of third parties shall be taken into account. In regard to counterfeit trademark goods, the simple removal of the trademark unlawfully affixed shall not be sufficient, other than in exceptional cases, to permit release of the goods into the channels of commerce.

Article 47

Right of Information

Members may provide that the judicial authorities shall have the authority, unless this would be out of proportion to the seriousness of the infringement, to order the infringer to inform the right holder of the identity of third persons involved in the production and distribution of the infringing goods or services and of their channels of distribution.

Article 48

Indemnification of the Defendant

1. The judicial authorities shall have the authority to order a party at whose request measures were taken and who has abused enforcement procedures to provide to a party wrongfully enjoined or restrained adequate compensation for the injury suffered because of such abuse. The judicial authorities shall also have the authority to order the applicant to pay the defendant expenses, which may include appropriate attorney's fees.

2. In respect of the administration of any law pertaining to the protection or enforcement of intellectual property rights, Members shall only exempt both public authorities and officials from liability to appropriate remedial measures where actions are taken or intended in good faith in the course of the administration of that law.

Article 49

Administrative Procedures

To the extent that any civil remedy can be ordered as a result of administrative procedures on the merits of a case, such procedures shall conform to principles equivalent in substance to those set forth in this Section.

SECTION 3: PROVISIONAL MEASURES

Article 50

1. The judicial authorities shall have the authority to order prompt and effective provisional measures:

- (a) to prevent an infringement of any intellectual property right from occurring, and in particular to prevent the entry into the channels of commerce in their jurisdiction of goods, including imported goods immediately after customs clearance;
- (b) to preserve relevant evidence in regard to the alleged infringement.

2. The judicial authorities shall have the authority to adopt provisional measures *inaudita altera parte* where appropriate, in particular where any delay is likely to cause irreparable harm to the right holder, or where there is a demonstrable risk of evidence being destroyed.

3. The judicial authorities shall have the authority to require the applicant to provide any reasonably available evidence in order to satisfy themselves with a sufficient degree of certainty that the applicant is the right holder and that the applicant's right is being infringed or that such infringement is imminent, and to order the applicant to provide a security or equivalent assurance sufficient to protect the defendant and to prevent abuse.

4. Where provisional measures have been adopted *inaudita altera parte*, the parties affected shall be given notice, without delay after the execution of the measures at the latest. A review, including a right to be heard, shall take place upon request of the defendant with a view to deciding, within a reasonable period after the notification of the measures, whether these measures shall be modified, revoked or confirmed.

5. The applicant may be required to supply other information necessary for the identification of the goods concerned by the authority that will execute the provisional measures.

6. Without prejudice to paragraph 4, provisional measures taken on the basis of paragraphs 1 and 2 shall, upon request by the defendant, be revoked or otherwise cease to have effect, if proceedings leading to a decision on the merits of the case are not initiated within a reasonable period, to be determined by the judicial authority ordering the measures where a Member's law so permits or, in the absence of such a determination, not to exceed 20 working days or 31 calendar days, whichever is the longer.

7. Where the provisional measures are revoked or where they lapse due to any act or omission by the applicant, or where it is subsequently found that there has been no infringement or threat of infringement of an intellectual property right, the judicial authorities shall have the authority to order the applicant, upon request of the defendant, to provide the defendant appropriate compensation for any injury caused by these measures.

8. To the extent that any provisional measure can be ordered as a result of administrative procedures, such procedures shall conform to principles equivalent in substance to those set forth in this Section.

SECTION 4: SPECIAL REQUIREMENTS RELATED TO BORDER MEASURES¹²

¹² Where a Member has dismantled substantially all controls over movement of goods across its border with another Member with which it forms part of a customs union, it shall not be required to apply the provisions of this Section at that border.

Article 51

Suspension of Release by Customs Authorities

Members shall, in conformity with the provisions set out below, adopt procedures¹³ to enable a right holder, who has valid grounds for suspecting that the importation of counterfeit trademark or pirated copyright goods¹⁴ may take place, to lodge an application in writing with competent authorities, administrative or judicial, for the suspension by the customs authorities of the release into free circulation of such goods. Members may enable such an application to be made in respect of goods which involve other infringements of intellectual property rights, provided that the requirements of this Section are met. Members may also provide for corresponding procedures concerning the suspension by the customs authorities of the release of infringing goods destined for exportation from their territories.

Article 52

Application

Any right holder initiating the procedures under Article 51 shall be required to provide adequate evidence to satisfy the competent authorities that, under the laws of the country of importation, there is *prima facie* an infringement of the right holder's intellectual property right and to supply a sufficiently detailed description of the goods to make them readily recognizable by the customs authorities. The competent authorities shall inform the applicant within a reasonable period whether they have accepted the application and, where determined by the competent authorities, the period for which the customs authorities will take action.

Article 53

Security or Equivalent Assurance

1. The competent authorities shall have the authority to require an applicant to provide a security or equivalent assurance sufficient to protect the defendant and the competent authorities and to prevent abuse. Such security or equivalent assurance shall not unreasonably deter recourse to these procedures.

2. Where pursuant to an application under this Section the release of goods involving industrial designs, patents, layout-designs or undisclosed information into free circulation has been suspended by customs authorities on the basis of a decision other than by a judicial or other independent authority, and the period provided for in Article 55 has expired without the granting of provisional relief by the duly empowered authority, and provided that all other conditions for importation have been complied with, the owner, importer, or consignee of such goods shall be entitled to their release

¹³ It is understood that there shall be no obligation to apply such procedures to imports of goods put on the market in another country by or with the consent of the right holder, or to goods in transit.

¹⁴ For the purposes of this Agreement:

(a) "counterfeit trademark goods" shall mean any goods, including packaging, bearing without authorization a trademark which is identical to the trademark validly registered in respect of such goods, or which cannot be distinguished in its essential aspects from such a trademark, and which thereby infringes the rights of the owner of the trademark in question under the law of the country of importation;

(b) "pirated copyright goods" shall mean any goods which are copies made without the consent of the right holder or person duly authorized by the right holder in the country of production and which are made directly or indirectly from an article where the making of that copy would have constituted an infringement of a copyright or a related right under the law of the country of importation.

on the posting of a security in an amount sufficient to protect the right holder for any infringement. Payment of such security shall not prejudice any other remedy available to the right holder, it being understood that the security shall be released if the right holder fails to pursue the right of action within a reasonable period of time.

Article 54

Notice of Suspension

The importer and the applicant shall be promptly notified of the suspension of the release of goods according to Article 51.

Article 55

Duration of Suspension

If, within a period not exceeding 10 working days after the applicant has been served notice of the suspension, the customs authorities have not been informed that proceedings leading to a decision on the merits of the case have been initiated by a party other than the defendant, or that the duly empowered authority has taken provisional measures prolonging the suspension of the release of the goods, the goods shall be released, provided that all other conditions for importation or exportation have been complied with; in appropriate cases, this time-limit may be extended by another 10 working days. If proceedings leading to a decision on the merits of the case have been initiated, a review, including a right to be heard, shall take place upon request of the defendant with a view to deciding, within a reasonable period, whether these measures shall be modified, revoked or confirmed. Notwithstanding the above, where the suspension of the release of goods is carried out or continued in accordance with a provisional judicial measure, the provisions of paragraph 6 of Article 50 shall apply.

Article 56

Indemnification of the Importer and of the Owner of the Goods

Relevant authorities shall have the authority to order the applicant to pay the importer, the consignee and the owner of the goods appropriate compensation for any injury caused to them through the wrongful detention of goods or through the detention of goods released pursuant to Article 55.

Article 57

Right of Inspection and Information

Without prejudice to the protection of confidential information, Members shall provide the competent authorities the authority to give the right holder sufficient opportunity to have any goods detained by the customs authorities inspected in order to substantiate the right holder's claims. The competent authorities shall also have authority to give the importer an equivalent opportunity to have any such goods inspected. Where a positive determination has been made on the merits of a case, Members may provide the competent authorities the authority to inform the right holder of the names

and addresses of the consignor, the importer and the consignee and of the quantity of the goods in question.

Article 58

Ex Officio Action

Where Members require competent authorities to act upon their own initiative and to suspend the release of goods in respect of which they have acquired *prima facie* evidence that an intellectual property right is being infringed:

- (a) the competent authorities may at any time seek from the right holder any information that may assist them to exercise these powers;
- (b) the importer and the right holder shall be promptly notified of the suspension. Where the importer has lodged an appeal against the suspension with the competent authorities, the suspension shall be subject to the conditions, *mutatis mutandis*, set out at Article 55;
- (c) Members shall only exempt both public authorities and officials from liability to appropriate remedial measures where actions are taken or intended in good faith.

Article 59

Remedies

Without prejudice to other rights of action open to the right holder and subject to the right of the defendant to seek review by a judicial authority, competent authorities shall have the authority to order the destruction or disposal of infringing goods in accordance with the principles set out in Article 46. In regard to counterfeit trademark goods, the authorities shall not allow the re-exportation of the infringing goods in an unaltered state or subject them to a different customs procedure, other than in exceptional circumstances.

Article 60

De Minimis Imports

Members may exclude from the application of the above provisions small quantities of goods of a non-commercial nature contained in travellers' personal luggage or sent in small consignments.

SECTION 5: CRIMINAL PROCEDURES

Article 61

Members shall provide for criminal procedures and penalties to be applied at least in cases of wilful trademark counterfeiting or copyright piracy on a commercial scale. Remedies available shall include imprisonment and/or monetary fines sufficient to provide a deterrent, consistently with the level of penalties applied for crimes of a corresponding gravity. In appropriate cases, remedies available shall also include the seizure, forfeiture and destruction of the infringing goods and of any

materials and implements the predominant use of which has been in the commission of the offence. Members may provide for criminal procedures and penalties to be applied in other cases of infringement of intellectual property rights, in particular where they are committed wilfully and on a commercial scale.

PART IV

ACQUISITION AND MAINTENANCE OF INTELLECTUAL PROPERTY RIGHTS AND RELATED *INTER-PARTES* PROCEDURES

Article 62

1. Members may require, as a condition of the acquisition or maintenance of the intellectual property rights provided for under Sections 2 through 6 of Part II, compliance with reasonable procedures and formalities. Such procedures and formalities shall be consistent with the provisions of this Agreement.

2. Where the acquisition of an intellectual property right is subject to the right being granted or registered, Members shall ensure that the procedures for grant or registration, subject to compliance with the substantive conditions for acquisition of the right, permit the granting or registration of the right within a reasonable period of time so as to avoid unwarranted curtailment of the period of protection.

3. Article 4 of the Paris Convention (1967) shall apply *mutatis mutandis* to service marks.

4. Procedures concerning the acquisition or maintenance of intellectual property rights and, where a Member's law provides for such procedures, administrative revocation and *inter partes* procedures such as opposition, revocation and cancellation, shall be governed by the general principles set out in paragraphs 2 and 3 of Article 41.

5. Final administrative decisions in any of the procedures referred to under paragraph 4 shall be subject to review by a judicial or quasi-judicial authority. However, there shall be no obligation to provide an opportunity for such review of decisions in cases of unsuccessful opposition or administrative revocation, provided that the grounds for such procedures can be the subject of invalidation procedures.

PART V

DISPUTE PREVENTION AND SETTLEMENT

Article 63

Transparency

1. Laws and regulations, and final judicial decisions and administrative rulings of general application, made effective by a Member pertaining to the subject matter of this Agreement (the availability, scope, acquisition, enforcement and prevention of the abuse of intellectual property rights) shall be published, or where such publication is not practicable made publicly available, in a national language, in such a manner as to enable governments and right holders to become acquainted with them. Agreements concerning the subject matter of this Agreement which are in force between

the government or a governmental agency of a Member and the government or a governmental agency of another Member shall also be published.

2. Members shall notify the laws and regulations referred to in paragraph 1 to the Council for TRIPS in order to assist that Council in its review of the operation of this Agreement. The Council shall attempt to minimize the burden on Members in carrying out this obligation and may decide to waive the obligation to notify such laws and regulations directly to the Council if consultations with WIPO on the establishment of a common register containing these laws and regulations are successful. The Council shall also consider in this connection any action required regarding notifications pursuant to the obligations under this Agreement stemming from the provisions of Article 6^{ter} of the Paris Convention (1967).

3. Each Member shall be prepared to supply, in response to a written request from another Member, information of the sort referred to in paragraph 1. A Member, having reason to believe that a specific judicial decision or administrative ruling or bilateral agreement in the area of intellectual property rights affects its rights under this Agreement, may also request in writing to be given access to or be informed in sufficient detail of such specific judicial decisions or administrative rulings or bilateral agreements.

4. Nothing in paragraphs 1, 2 and 3 shall require Members to disclose confidential information which would impede law enforcement or otherwise be contrary to the public interest or would prejudice the legitimate commercial interests of particular enterprises, public or private.

Article 64

Dispute Settlement

1. The provisions of Articles XXII and XXIII of GATT 1994 as elaborated and applied by the Dispute Settlement Understanding shall apply to consultations and the settlement of disputes under this Agreement except as otherwise specifically provided herein.

2. Subparagraphs 1(b) and 1(c) of Article XXIII of GATT 1994 shall not apply to the settlement of disputes under this Agreement for a period of five years from the date of entry into force of the WTO Agreement.

3. During the time period referred to in paragraph 2, the Council for TRIPS shall examine the scope and modalities for complaints of the type provided for under subparagraphs 1(b) and 1(c) of Article XXIII of GATT 1994 made pursuant to this Agreement, and submit its recommendations to the Ministerial Conference for approval. Any decision of the Ministerial Conference to approve such recommendations or to extend the period in paragraph 2 shall be made only by consensus, and approved recommendations shall be effective for all Members without further formal acceptance process.

PART VI

TRANSITIONAL ARRANGEMENTS

Article 65

Transitional Arrangements

1. Subject to the provisions of paragraphs 2, 3 and 4, no Member shall be obliged to apply the provisions of this Agreement before the expiry of a general period of one year following the date of entry into force of the WTO Agreement.
2. A developing country Member is entitled to delay for a further period of four years the date of application, as defined in paragraph 1, of the provisions of this Agreement other than Articles 3, 4 and 5.
3. Any other Member which is in the process of transformation from a centrally-planned into a market, free-enterprise economy and which is undertaking structural reform of its intellectual property system and facing special problems in the preparation and implementation of intellectual property laws and regulations, may also benefit from a period of delay as foreseen in paragraph 2.
4. To the extent that a developing country Member is obliged by this Agreement to extend product patent protection to areas of technology not so protectable in its territory on the general date of application of this Agreement for that Member, as defined in paragraph 2, it may delay the application of the provisions on product patents of Section 5 of Part II to such areas of technology for an additional period of five years.
5. A Member availing itself of a transitional period under paragraphs 1, 2, 3 or 4 shall ensure that any changes in its laws, regulations and practice made during that period do not result in a lesser degree of consistency with the provisions of this Agreement.

Article 66

Least-Developed Country Members

1. In view of the special needs and requirements of least-developed country Members, their economic, financial and administrative constraints, and their need for flexibility to create a viable technological base, such Members shall not be required to apply the provisions of this Agreement, other than Articles 3, 4 and 5, for a period of 10 years from the date of application as defined under paragraph 1 of Article 65. The Council for TRIPS shall, upon duly motivated request by a least-developed country Member, accord extensions of this period.
2. Developed country Members shall provide incentives to enterprises and institutions in their territories for the purpose of promoting and encouraging technology transfer to least-developed country Members in order to enable them to create a sound and viable technological base.

Article 67

Technical Cooperation

In order to facilitate the implementation of this Agreement, developed country Members shall provide, on request and on mutually agreed terms and conditions, technical and financial cooperation in favour of developing and least-developed country Members. Such cooperation shall include assistance in the preparation of laws and regulations on the protection and enforcement of intellectual property rights as well as on the prevention of their abuse, and shall include support regarding the establishment or reinforcement of domestic offices and agencies relevant to these matters, including the training of personnel.

PART VII

INSTITUTIONAL ARRANGEMENTS; FINAL PROVISIONS

Article 68

Council for Trade-Related Aspects of Intellectual Property Rights

The Council for TRIPS shall monitor the operation of this Agreement and, in particular, Members' compliance with their obligations hereunder, and shall afford Members the opportunity of consulting on matters relating to the trade-related aspects of intellectual property rights. It shall carry out such other responsibilities as assigned to it by the Members, and it shall, in particular, provide any assistance requested by them in the context of dispute settlement procedures. In carrying out its functions, the Council for TRIPS may consult with and seek information from any source it deems appropriate. In consultation with WIPO, the Council shall seek to establish, within one year of its first meeting, appropriate arrangements for cooperation with bodies of that Organization.

Article 69

International Cooperation

Members agree to cooperate with each other with a view to eliminating international trade in goods infringing intellectual property rights. For this purpose, they shall establish and notify contact points in their administrations and be ready to exchange information on trade in infringing goods. They shall, in particular, promote the exchange of information and cooperation between customs authorities with regard to trade in counterfeit trademark goods and pirated copyright goods.

Article 70

Protection of Existing Subject Matter

1. This Agreement does not give rise to obligations in respect of acts which occurred before the date of application of the Agreement for the Member in question.
2. Except as otherwise provided for in this Agreement, this Agreement gives rise to obligations in respect of all subject matter existing at the date of application of this Agreement for the Member in question, and which is protected in that Member on the said date, or which meets or comes subsequently to meet the criteria for protection under the terms of this Agreement. In respect of this paragraph and paragraphs 3 and 4, copyright obligations with respect to existing works shall be solely determined under Article 18 of the Berne Convention (1971), and obligations with respect to the rights of producers of phonograms and performers in existing phonograms shall be determined solely under Article 18 of the Berne Convention (1971) as made applicable under paragraph 6 of Article 14 of this Agreement.
3. There shall be no obligation to restore protection to subject matter which on the date of application of this Agreement for the Member in question has fallen into the public domain.
4. In respect of any acts in respect of specific objects embodying protected subject matter which become infringing under the terms of legislation in conformity with this Agreement, and which were commenced, or in respect of which a significant investment was made, before the date of acceptance of the WTO Agreement by that Member, any Member may provide for a limitation of the remedies

available to the right holder as to the continued performance of such acts after the date of application of this Agreement for that Member. In such cases the Member shall, however, at least provide for the payment of equitable remuneration.

5. A Member is not obliged to apply the provisions of Article 11 and of paragraph 4 of Article 14 with respect to originals or copies purchased prior to the date of application of this Agreement for that Member.

6. Members shall not be required to apply Article 31, or the requirement in paragraph 1 of Article 27 that patent rights shall be enjoyable without discrimination as to the field of technology, to use without the authorization of the right holder where authorization for such use was granted by the government before the date this Agreement became known.

7. In the case of intellectual property rights for which protection is conditional upon registration, applications for protection which are pending on the date of application of this Agreement for the Member in question shall be permitted to be amended to claim any enhanced protection provided under the provisions of this Agreement. Such amendments shall not include new matter.

8. Where a Member does not make available as of the date of entry into force of the WTO Agreement patent protection for pharmaceutical and agricultural chemical products commensurate with its obligations under Article 27, that Member shall:

- (a) notwithstanding the provisions of Part VI, provide as from the date of entry into force of the WTO Agreement a means by which applications for patents for such inventions can be filed;
- (b) apply to these applications, as of the date of application of this Agreement, the criteria for patentability as laid down in this Agreement as if those criteria were being applied on the date of filing in that Member or, where priority is available and claimed, the priority date of the application; and
- (c) provide patent protection in accordance with this Agreement as from the grant of the patent and for the remainder of the patent term, counted from the filing date in accordance with Article 33 of this Agreement, for those of these applications that meet the criteria for protection referred to in subparagraph (b).

9. Where a product is the subject of a patent application in a Member in accordance with paragraph 8(a), exclusive marketing rights shall be granted, notwithstanding the provisions of Part VI, for a period of five years after obtaining marketing approval in that Member or until a product patent is granted or rejected in that Member, whichever period is shorter, provided that, subsequent to the entry into force of the WTO Agreement, a patent application has been filed and a patent granted for that product in another Member and marketing approval obtained in such other Member.

Article 71

Review and Amendment

1. The Council for TRIPS shall review the implementation of this Agreement after the expiration of the transitional period referred to in paragraph 2 of Article 65. The Council shall, having regard to the experience gained in its implementation, review it two years after that date, and at identical intervals thereafter. The Council may also undertake reviews in the light of any relevant new developments which might warrant modification or amendment of this Agreement.

2. Amendments merely serving the purpose of adjusting to higher levels of protection of intellectual property rights achieved, and in force, in other multilateral agreements and accepted under those agreements by all Members of the WTO may be referred to the Ministerial Conference for action in accordance with paragraph 6 of Article X of the WTO Agreement on the basis of a consensus proposal from the Council for TRIPS.

Article 72

Reservations

Reservations may not be entered in respect of any of the provisions of this Agreement without the consent of the other Members.

Article 73

Security Exceptions

Nothing in this Agreement shall be construed:

- (a) to require a Member to furnish any information the disclosure of which it considers contrary to its essential security interests; or
- (b) to prevent a Member from taking any action which it considers necessary for the protection of its essential security interests;
 - (i) relating to fissionable materials or the materials from which they are derived;
 - (ii) relating to the traffic in arms, ammunition and implements of war and to such traffic in other goods and materials as is carried on directly or indirectly for the purpose of supplying a military establishment;
 - (iii) taken in time of war or other emergency in international relations; or
- (c) to prevent a Member from taking any action in pursuance of its obligations under the United Nations Charter for the maintenance of international peace and security.

ANNEX TO THE TRIPS AGREEMENT

1. For the purposes of Article 31*bis* and this Annex:
 - (a) "pharmaceutical product" means any patented product, or product manufactured through a patented process, of the pharmaceutical sector needed to address the public health problems as recognized in paragraph 1 of the Declaration on the TRIPS Agreement and Public Health (WT/MIN(01)/DEC/2). It is understood that active ingredients necessary for its manufacture and diagnostic kits needed for its use would be included¹;
 - (b) "eligible importing Member" means any least-developed country Member, and any other Member that has made a notification² to the Council for TRIPS of its intention to use the system set out in Article 31*bis* and this Annex ("system") as an importer, it being understood that a Member may notify at any time that it will use the system in whole or in a limited way, for example only in the case of a national emergency or other circumstances of extreme urgency or in cases of public non-commercial use. It is noted that some Members will not use the system as importing Members³ and that some other Members have stated that, if they use the system, it would be in no more than situations of national emergency or other circumstances of extreme urgency;
 - (c) "exporting Member" means a Member using the system to produce pharmaceutical products for, and export them to, an eligible importing Member.
2. The terms referred to in paragraph 1 of Article 31*bis* are that:
 - (a) the eligible importing Member(s)⁴ has made a notification² to the Council for TRIPS, that:
 - (i) specifies the names and expected quantities of the product(s) needed⁵;
 - (ii) confirms that the eligible importing Member in question, other than a least-developed country Member, has established that it has insufficient or no manufacturing capacities in the pharmaceutical sector for the product(s) in question in one of the ways set out in the Appendix to this Annex; and
 - (iii) confirms that, where a pharmaceutical product is patented in its territory, it has granted or intends to grant a compulsory licence in accordance with Articles 31 and 31*bis* of this Agreement and the provisions of this Annex⁶;

¹ This subparagraph is without prejudice to subparagraph 1(b).

² It is understood that this notification does not need to be approved by a WTO body in order to use the system.

³ Australia, Canada, the European Communities with, for the purposes of Article 31*bis* and this Annex, its member States, Iceland, Japan, New Zealand, Norway, Switzerland, and the United States.

⁴ Joint notifications providing the information required under this subparagraph may be made by the regional organizations referred to in paragraph 3 of Article 31*bis* on behalf of eligible importing Members using the system that are parties to them, with the agreement of those parties.

⁵ The notification will be made available publicly by the WTO Secretariat through a page on the WTO website dedicated to the system.

⁶ This subparagraph is without prejudice to Article 66.1 of this Agreement.

- (b) the compulsory licence issued by the exporting Member under the system shall contain the following conditions:
- (i) only the amount necessary to meet the needs of the eligible importing Member(s) may be manufactured under the licence and the entirety of this production shall be exported to the Member(s) which has notified its needs to the Council for TRIPS;
 - (ii) products produced under the licence shall be clearly identified as being produced under the system through specific labelling or marking. Suppliers should distinguish such products through special packaging and/or special colouring/shaping of the products themselves, provided that such distinction is feasible and does not have a significant impact on price; and
 - (iii) before shipment begins, the licensee shall post on a website⁷ the following information:
 - the quantities being supplied to each destination as referred to in indent (i) above; and
 - the distinguishing features of the product(s) referred to in indent (ii) above;
- (c) the exporting Member shall notify⁸ the Council for TRIPS of the grant of the licence, including the conditions attached to it.⁹ The information provided shall include the name and address of the licensee, the product(s) for which the licence has been granted, the quantity(ies) for which it has been granted, the country(ies) to which the product(s) is (are) to be supplied and the duration of the licence. The notification shall also indicate the address of the website referred to in subparagraph (b)(iii) above.

3. In order to ensure that the products imported under the system are used for the public health purposes underlying their importation, eligible importing Members shall take reasonable measures within their means, proportionate to their administrative capacities and to the risk of trade diversion to prevent re-exportation of the products that have actually been imported into their territories under the system. In the event that an eligible importing Member that is a developing country Member or a least-developed country Member experiences difficulty in implementing this provision, developed country Members shall provide, on request and on mutually agreed terms and conditions, technical and financial cooperation in order to facilitate its implementation.

4. Members shall ensure the availability of effective legal means to prevent the importation into, and sale in, their territories of products produced under the system and diverted to their markets inconsistently with its provisions, using the means already required to be available under this Agreement. If any Member considers that such measures are proving insufficient for this purpose, the matter may be reviewed in the Council for TRIPS at the request of that Member.

5. With a view to harnessing economies of scale for the purposes of enhancing purchasing power for, and facilitating the local production of, pharmaceutical products, it is recognized that the

⁷ The licensee may use for this purpose its own website or, with the assistance of the WTO Secretariat, the page on the WTO website dedicated to the system.

⁸ It is understood that this notification does not need to be approved by a WTO body in order to use the system.

⁹ The notification will be made available publicly by the WTO Secretariat through a page on the WTO website dedicated to the system.

development of systems providing for the grant of regional patents to be applicable in the Members described in paragraph 3 of Article 31*bis* should be promoted. To this end, developed country Members undertake to provide technical cooperation in accordance with Article 67 of this Agreement, including in conjunction with other relevant intergovernmental organizations.

6. Members recognize the desirability of promoting the transfer of technology and capacity building in the pharmaceutical sector in order to overcome the problem faced by Members with insufficient or no manufacturing capacities in the pharmaceutical sector. To this end, eligible importing Members and exporting Members are encouraged to use the system in a way which would promote this objective. Members undertake to cooperate in paying special attention to the transfer of technology and capacity building in the pharmaceutical sector in the work to be undertaken pursuant to Article 66.2 of this Agreement, paragraph 7 of the Declaration on the TRIPS Agreement and Public Health and any other relevant work of the Council for TRIPS.

7. The Council for TRIPS shall review annually the functioning of the system with a view to ensuring its effective operation and shall annually report on its operation to the General Council.

APPENDIX TO THE ANNEX TO THE TRIPS AGREEMENT

Assessment of Manufacturing Capacities in the Pharmaceutical Sector

Least-developed country Members are deemed to have insufficient or no manufacturing capacities in the pharmaceutical sector.

For other eligible importing Members insufficient or no manufacturing capacities for the product(s) in question may be established in either of the following ways:

- (i) the Member in question has established that it has no manufacturing capacity in the pharmaceutical sector;

or

 - (ii) where the Member has some manufacturing capacity in this sector, it has examined this capacity and found that, excluding any capacity owned or controlled by the patent owner, it is currently insufficient for the purposes of meeting its needs. When it is established that such capacity has become sufficient to meet the Member's needs, the system shall no longer apply.
-

CHAPTER 11: OPINIONS ON FINANCING & DELIVERY OF HEALTH CARE

The Opinions in this chapter are offered as ethics guidance for physicians and are not intended to establish standards of clinical practice or rules of law.

11.1 Access to Health Care

- 11.1.1 Defining Basic Health Care
- 11.1.2 Physician Stewardship of Health Care Resources
- 11.1.3 Allocating Limited Health Care Resources
- 11.1.4 Financial Barriers to Health Care Access

11.2 Health Care Organizations & Physician Practice

- 11.2.1 Professionalism in Health Care Systems
- 11.2.2 Conflicts of Interest in Patient Care
- 11.2.3 Contracts to Deliver Health Care Services
 - 11.2.3.1 Restrictive Covenants
- 11.2.4 Transparency in Health Care
- 11.2.5 Retainer Practices

11.3 Fees & Charges

- 11.3.1 Fees for Medical Services
- 11.3.2 Fees for Nonclinical & Administrative Services
- 11.3.3 Interest & Finance Charges
- 11.3.4 Fee Splitting



11.1.1 Defining Basic Health Care

Health care is a fundamental human good because it affects our opportunity to pursue life goals, reduces our pain and suffering, helps prevent premature loss of life, and provides information needed to plan for our lives. Society has an obligation to make access to an adequate level of care available to all its members, regardless of ability to pay.

Physicians regularly confront the effects of lack of access to adequate care and have a corresponding responsibility to contribute their expertise to societal decisions about what health care services should be included in a minimum package of care for all.

Individually and collectively as a profession, physicians should advocate for fair, informed decision making about basic health care that:

- (a) Is transparent.
- (b) Strives to include input from all stakeholders, including the public, throughout the process.
- (c) Protects the most vulnerable patients and populations, with special attention to historically disadvantaged groups.
- (d) Considers best available scientific data about the efficacy and safety of health care services.
- (e) Seeks to improve health outcomes to the greatest extent possible, in keeping with principles of wise stewardship.

- (f) Monitors for variations in care that cannot be explained on medical grounds to ensure that the defined threshold of basic care does not have discriminatory impact.
- (g) Provides for ongoing review and adjustment in consideration of innovation in medical science and practice to ensure continued, broad public support for the defined threshold of basic care.

AMA Principles of Medical Ethics: VII

11.1.2 Physician Stewardship of Health Care Resources

Physicians' primary ethical obligation is to promote the well-being of individual patients. Physicians also have a long-recognized obligation to patients in general to promote public health and access to care. This obligation requires physicians to be prudent stewards of the shared societal resources with which they are entrusted. Managing health care resources responsibly for the benefit of all patients is compatible with physicians' primary obligation to serve the interests of individual patients.

To fulfill their obligation to be prudent stewards of health care resources, physicians should:

- (a) Base recommendations and decisions on patients' medical needs.
- (b) Use scientifically grounded evidence to inform professional decisions when available.
- (c) Help patients articulate their health care goals and help patients and their families form realistic expectations about whether a particular intervention is likely to achieve those goals.
- (d) Endorse recommendations that offer reasonable likelihood of achieving the patient's health care goals.
- (e) Choose the course of action that requires fewer resources when alternative courses of action offer similar likelihood and degree of anticipated benefit compared to anticipated harm for the individual patient but require different levels of resources.
- (f) Be transparent about alternatives, including disclosing when resource constraints play a role in decision making.
- (g) Participate in efforts to resolve persistent disagreement about whether a costly intervention is worthwhile, which may include consulting other physicians, an ethics committee, or other appropriate resource.

Physicians are in a unique position to affect health care spending. But individual physicians alone cannot and should not be expected to address the systemic challenges of wisely managing health care resources. Medicine as a profession must create conditions for practice that make it feasible for individual physicians to be prudent stewards by:

- (h) Encouraging health care administrators and organizations to make cost data transparent (including cost accounting methodologies) so that physicians can exercise well-informed stewardship.
- (i) Ensuring that physicians have the training they need to be informed about health care costs and how their decisions affect overall health care spending.

- (j) Advocating for policy changes, such as medical liability reform, that promote professional judgment and address systemic barriers that impede responsible stewardship.

AMA Principles of Medical Ethics: I,V,VII,VIII,IX

11.1.3 Allocating Limited Health Care Resources

Physicians' primary ethical obligation is to promote the well-being of their patients. Policies for allocating scarce health care resources can impede their ability to fulfill that obligation, whether those policies address situations of chronically limited resources, such as ICU (intensive care unit) beds, medications, or solid organs for transplantation, or "triage" situations in times of scarcity, such as access to ventilators during an influenza pandemic.

As professionals dedicated to protecting the interests of their patients, physicians thus have a responsibility to contribute their expertise to developing allocation policies that are fair and safeguard the welfare of patients.

Individually and collectively through the profession, physicians should advocate for policies and procedures that allocate scarce health care resources fairly among patients, in keeping with the following criteria:

- (a) Base allocation policies on criteria relating to medical need, including urgency of need, likelihood and anticipated duration of benefit, and change in quality of life. In limited circumstances, it may be appropriate to take into consideration the amount of resources required for successful treatment. It is not appropriate to base allocation policies on social worth, perceived obstacles to treatment, patient contribution to illness, past use of resources, or other non-medical characteristics.
- (b) Give first priority to those patients for whom treatment will avoid premature death or extremely poor outcomes, then to patients who will experience the greatest change in quality of life, when there are very substantial differences among patients who need access to the scarce resource(s).
- (c) Use an objective, flexible, transparent mechanism to determine which patients will receive the resource(s) when there are not substantial differences among patients who need access to the scarce resource(s).
- (d) Explain the applicable allocation policies or procedures to patients who are denied access to the scarce resource(s) and to the public.

AMA Principles of Medical Ethics: I,VII

11.1.4 Financial Barriers to Health Care Access

Health care is a fundamental human good because it affects our opportunity to pursue life goals, reduces our pain and suffering, helps prevent premature loss of life, and provides information needed to plan for our lives. As professionals, physicians individually and collectively have an ethical responsibility to ensure that all persons have access to needed care regardless of their economic means.

In view of this obligation,

- (a) Individual physicians should:
- (i) take steps to promote access to care for individual patients, such as providing pro bono care in their office or through freestanding facilities or government programs that provide health care for the poor, or, when permissible, waiving insurance copayments in individual cases of hardship. Physicians in the poorest communities should be able to turn for assistance to colleagues in more prosperous communities.
 - (ii) help patients obtain needed care through public or charitable programs when patients cannot do so themselves.
- (b) Physicians, individually and collectively through their professional organizations and institutions, should participate in the political process as advocates for patients (or support those who do) so as to diminish financial obstacles to access health care.
- (c) The medical profession must work to ensure that societal decisions about the distribution of health resources safeguard the interests of all patients and promote access to health services.
- (d) All stakeholders in health care, including physicians, health facilities, health insurers, professional medical societies, and public policymakers must work together to ensure sufficient access to appropriate health care for all people.

AMA Principles of Medical Ethics: I,II,VI,VII,IX

11.2.1 Professionalism in Health Care Systems

Containing costs, promoting high-quality care for all patients, and sustaining physician professionalism are important goals. Models for financing and organizing the delivery of health care services often aim to promote patient safety and to improve quality and efficiency. However, they can also pose ethical challenges for physicians that could undermine the trust essential to patient-physician relationships.

Payment models and financial incentives can create conflicts of interest among patients, health care organizations, and physicians. They can encourage undertreatment and overtreatment, as well as dictate goals that are not individualized for the particular patient.

Structures that influence where and by whom care is delivered—such as accountable care organizations, group practices, health maintenance organizations, and other entities that may emerge in the future—can affect patients’ choices, the patient-physician relationship, and physicians’ relationships with fellow health care professionals.

Formularies, clinical practice guidelines, and other tools intended to influence decision making, may impinge on physicians’ exercise of professional judgment and ability to advocate effectively for their patients, depending on how they are designed and implemented.

Physicians in leadership positions within health care organizations should ensure that practices for financing and organizing the delivery of care:

- (a) Are transparent.
- (b) Reflect input from key stakeholders, including physicians and patients.

- (c) Recognize that over reliance on financial incentives may undermine physician professionalism.
- (d) Ensure ethically acceptable incentives that:
 - (i) are designed in keeping with sound principles and solid scientific evidence. Financial incentives should be based on appropriate comparison groups and cost data and adjusted to reflect complexity, case mix, and other factors that affect physician practice profiles. Practice guidelines, formularies, and other tools should be based on best available evidence and developed in keeping with ethics guidance;
 - (ii) are implemented fairly and do not disadvantage identifiable populations of patients or physicians or exacerbate health care disparities;
 - (iii) are implemented in conjunction with the infrastructure and resources needed to support high-value care and physician professionalism;
 - (iv) mitigate possible conflicts between physicians' financial interests and patient interests by minimizing the financial impact of patient care decisions and the overall financial risk for individual physicians.
- (e) Encourage, rather than discourage, physicians (and others) to:
 - (i) provide care for patients with difficult to manage medical conditions;
 - (ii) practice at their full capacity, but not beyond.
- (f) Recognize physicians' primary obligation to their patients by enabling physicians to respond to the unique needs of individual patients and providing avenues for meaningful appeal and advocacy on behalf of patients.
- (g) Are routinely monitored to:
 - (i) identify and address adverse consequences;
 - (ii) identify and encourage dissemination of positive outcomes.

All physicians should:

- (h) Hold physician-leaders accountable to meeting conditions for professionalism in health care systems.
- (i) Advocate for changes in health care payment and delivery models to promote access to high-quality care for all patients.

AMA Principles of Medical Ethics: I,II III,V

11.2.2 Conflicts of Interest in Patient Care

The primary objective of the medical profession is to render service to humanity; reward or financial gain is a subordinate consideration. Under no circumstances may physicians place their own financial interests above the welfare of their patients.

Treatment or hospitalization that is willfully excessive or inadequate constitutes unethical practice. Physicians should not provide wasteful and unnecessary treatment that may cause needless expense solely for the physician's financial benefit or for the benefit of a hospital or other health care organization with which the physician is affiliated.

Where the economic interests of the hospital, health care organization, or other entity are in conflict with patient welfare, patient welfare takes priority.

AMA Principles of Medical Ethics: II

11.2.3 Contracts to Deliver Health Care Services

Physicians have a fundamental ethical obligation to put the welfare of patients ahead of other considerations, including personal financial interests. This obligation requires them to consider carefully the terms and conditions of contracts to deliver health care services before entering into such contracts to ensure that those contracts do not create untenable conflicts of interests.

Ongoing evolution in the health care system continues to bring changes to medicine, including changes in reimbursement mechanisms, models for health care delivery, restrictions on referral and use of services, clinical practice guidelines, and limitations on benefits packages. While these changes are intended to enhance quality, efficiency, and safety in health care, they can also put at risk physicians' ability to uphold professional ethical standards of informed consent and fidelity to patients and can impede physicians' freedom to exercise independent professional judgment and tailor care to meet the needs of individual patients.

As physicians enter into various differently structured contracts to deliver health care services—with group practices, hospitals, health plans, or other entities—they should be mindful that while many arrangements have the potential to promote desired improvements in care, some arrangements also have the potential to impede patients' interests.

When contracting to provide health care services, physicians should:

- (a) Carefully review the terms of proposed contracts or have a representative do so on their behalf to assure themselves that the arrangement:
 - (i) minimizes conflict of interest with respect to proposed reimbursement mechanisms, financial or performance incentives, restrictions on care, or other mechanisms intended to influence physicians' treatment recommendations or direct what care patients receive, in keeping with ethics guidance;
 - (ii) does not compromise the physician's own financial well-being or ability to provide high-quality care through unrealistic expectations regarding utilization of services or terms that expose the physician to excessive financial risk;
 - (iii) allows the physician to appropriately exercise professional judgment;
 - (iv) includes a mechanism to address grievances and supports advocacy on behalf of individual patients;
 - (v) permits disclosure to patients.

- (b) Negotiate modification or removal of any terms that unduly compromise physicians' ability to uphold ethical standards.

AMA Principles of Medical Ethics: I,II,III,V,VI,VIII,IX

11.2.3.1 Restrictive Covenants

Competition among physicians is ethically justifiable when it is based on such factors as quality of services, skill, experience, conveniences offered to patients, fees, or credit terms.

Covenants-not-to-compete restrict competition, can disrupt continuity of care, and may limit access to care.

Physicians should not enter into covenants that:

- (a) Unreasonably restrict the right of a physician to practice medicine for a specified period of time or in a specified geographic area on termination of a contractual relationship; and
- (b) Do not make reasonable accommodation for patients' choice of physician.

Physicians in training should not be asked to sign covenants not to compete as a condition of entry into any residency or fellowship program.

AMA Principles of Medical Ethics: III,IV,VI,VII

11.2.4 Transparency in Health Care

Respect for patients' autonomy is a cornerstone of medical ethics. Patients must rely on their physicians to provide information that patients would reasonably want to know to make informed, well-considered decisions about their health care. Thus, physicians have an obligation to inform patients about all appropriate treatment options, the risks and benefits of alternatives, and other information that may be pertinent, including the existence of payment models, financial incentives; and formularies, guidelines or other tools that influence treatment recommendations and care. Restrictions on disclosure can impede communication between patient and physician and undermine trust, patient choice, and quality of care.

Although health plans and other entities may have primary responsibility to inform patient-members about plan provisions that will affect the availability of care, physicians share in this responsibility.

Individually, physicians should:

- (a) Disclose any financial and other factors that could affect the patient's care.
- (b) Disclose relevant treatment alternatives, including those that may not be covered under the patient's health plan.
- (c) Encourage patients to be aware of the provisions of their health plan.

Collectively, physicians should advocate that health plans with which they contract disclose to patient-members:

- (d) Plan provisions that limit care, such as formularies or constraints on referrals.
- (e) Plan provisions for obtaining desired care that would otherwise not be provided, such as provision for off-formulary prescribing.
- (f) Plan relationships with pharmacy benefit management organizations and other commercial entities that have an interest in physicians' treatment recommendations.

AMA Principles of Medical Ethics: I,II,III,V,VI

11.2.5 Retainer Practices

Physicians are free to enter into contracts to provide special non-medical services and amenities with individual patients who are willing and able to pay additional costs out of pocket for such services. While such retainer contracts are one among many diverse models for delivering and paying for health care, they can also raise ethical concerns about access, quality, and continuity of care.

Regardless of the model within which they practice, physicians must uphold their primary professional obligation of fidelity and their responsibility to treat all patients with courtesy and respect for patients' rights and dignity, and ensure that all patients in the physician's practice receive the same quality of medical care, regardless of contractual arrangements for special, non-medical services and amenities.

Physicians who enter into retainer contracts with patients must:

- (a) Present the terms of the retainer arrangement clearly to patients, including implications for the patient's current health care insurance, if known, and take care not to imply that more or better medical services will be provided under a retainer contract.
- (b) Ensure that patient decisions to accept retainer contracts are voluntary and that patients are free to opt-out of entering into a retainer agreement.
- (c) Facilitate transfer of care for any patient who chooses not to participate in a retainer practice. If it is not feasible to transfer a patient's care to another local physician, the physician should continue to provide care under the terms of the patient's existing health care insurance until other appropriate arrangements for ongoing care can be made.
- (d) Ensure that treatment recommendations for all patients are based on scientific evidence, relevant professional guidelines, sound professional judgment, and prudent stewardship.
- (e) Uphold standards of honesty and transparency in billing and clearly distinguish charges for special services or amenities provided under a retainer contract from medical services reimbursable by the patient's health care insurance or third-party payer.
- (f) Uphold professional obligations to promote access to health care and to provide care to those in need regardless of ability to pay, in keeping with ethics guidance.

AMA Principles of Medical Ethics: I,II,VI,VIII,IX

11.3.1 Fees for Medical Services

Physicians are expected to conduct themselves as honest, responsible professionals. They should be knowledgeable about and conform to relevant laws and should adhere to professional ethical standards and sound business practice. Physicians should not recommend, provide, or charge for unnecessary medical services. Nor should they make intentional misrepresentations to increase the level of payment they receive or to secure noncovered health benefits for their patients.

With regard to fees for medical services, physicians should:

- (a) Charge reasonable fees based on the:
 - (i) kind of service(s);
 - (ii) difficulty or uniqueness of the service(s) performed;
 - (iii) time required to perform the service(s);
 - (iv) skill required to perform the service(s);
 - (v) experience of the physician;
 - (vi) quality of the physician's performance.
- (b) Charge only for the service(s) that are personally rendered or for services performed under the physician's direct personal observation, direction, or supervision. If possible, when services are provided by more than one physician, each physician should submit his or her own bill to the patient and be compensated separately. When physicians have professional colleagues assist in the performance of a service, the physician may pay a reasonable amount for such assistance and recoup that amount through fees charged to the patient, provided the patient is notified in advance of the financial arrangement.
- (c) Itemize separately charges for diagnostic, laboratory, or clinical services provided by other health care professionals and indicate who provided the service when fees for others' services cannot be billed directly to the patient, in addition to charges for the physician's own professional services.
- (d) Not charge excessive fees, contingent fees, or fees solely to facilitate hospital admission. Physicians must not charge a markup or commission, or profit on services rendered by other health care professionals.
- (e) Extend professional courtesy at their discretion, recognizing that it is not an ethical requirement and is prohibited in many jurisdictions.

AMA Principles of Medical Ethics: II, VI

11.3.2 Fees for Nonclinical & Administrative Services

Physicians individually and collectively should promote access to care for individual patients, in part through being prudent stewards of resources. Thus physicians have a responsibility to balance patients' needs and expectations with responsible business practices.

With respect to fees for nonclinical or administrative services provided in conjunction with patient care, physicians should:

- (a) Clearly notify patients in advance of fees charged by the practice (if any) for nonclinical or administrative services.
- (b) Base fees (if any) on reasonable costs to the practice for:
 - (i) providing special documentation on patient request for such purposes as insurance reimbursement to the patient, certification of immunization or fitness, or similar nonclinical services;
 - (ii) missed appointments or appointments not cancelled in advance in keeping with the published policy of the practice;
 - (iii) acquisition or processing charges in relation to diagnostic, laboratory, or clinical services, copies of medical records, or similar nonclinical services.

AMA Principles of Medical Ethics: II, VI

11.3.3 Interest & Finance Charges

Financial obstacles to medical care can directly affect patients' well-being and may diminish physicians' ability to use their knowledge and skills on patients' behalf. Physicians should not be expected to risk the viability of their practices or compromise quality of care by routinely providing care without compensation. Patients should make reasonable efforts to meet their financial responsibilities or to discuss financial hardships with their physicians.

To preserve patients' dignity and help sustain the patient-physician relationship, physicians should be candid about financial matters and:

- (a) Clearly notify patients in advance about policy and practice with respect to delinquent accounts, including under what circumstances:
 - (i) payment will be requested at the time of service;
 - (ii) interest or finance charges may be levied;
 - (iii) a past due account will be sent to a collection agency.
- (b) Ensure that no bills are sent to collection without the physician's knowledge.

- (c) Use discretion and compassion in hardship cases, in keeping with ethics guidance regarding financial barriers to health care access.

AMA Principles of Medical Ethics: II,VI,IX

11.3.4 Fee Splitting

Patients must be able to trust that their physicians will be honest with them and will make treatment recommendations, including referrals, based on medical need, the skill of other health care professionals or facilities to whom the patient is referred, and the quality of products or services provided.

Payment by or to a physician or health care institution solely for referral of a patient is fee splitting and is unethical.

Physicians may not accept:

- (a) Any payment of any kind, from any source for referring a patient other than distributions of a health care organization's revenues as permitted by law.
- (b) Any payment of any kind, from any source for prescribing a specific drug, product, or service.
- (c) Payment for services relating to the care of a patient from any health care facility/organization to which the physician has referred the patient.
- (d) Payment referring a patient to a research study.

Physicians in a capitated primary care practice may not refer patients based on whether the referring physician has negotiated a discount for specialty services.

AMA Principles of Medical Ethics: II



WILEY

The accuracy of hospital merger screening methods

Author(s): Christopher Garmon

Source: *The RAND Journal of Economics*, Winter 2017, Vol. 48, No. 4 (Winter 2017), pp. 1068-1102

Published by: Wiley on behalf of RAND Corporation

Stable URL: <https://www.jstor.org/stable/26305434>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

RAND Corporation and Wiley are collaborating with JSTOR to digitize, preserve and extend access to *The RAND Journal of Economics*

The accuracy of hospital merger screening methods

Christopher Garmon*

This article analyzes the accuracy of various prospective hospital merger screening methods used by antitrust agencies and the courts. The predictions of the screening methods calculated with pre-merger data are compared with the actual post-merger price changes of 28 hospital mergers measured relative to controls. The evaluated screening methods include traditional structural measures (e.g., Herfindahl-Hirschman Index), measures derived from hospital competition models (e.g., diversion ratios, willingness-to-pay, and upward pricing pressure), and hospital merger simulation. Willingness-to-pay and upward pricing pressure are found to be more accurate at flagging potentially anticompetitive mergers for further investigation than traditional methods.

1. Introduction

■ The hospital industry is one of the largest and most dynamic sectors in the United States economy. In 2015, hospital services accounted for 5.7% of US Gross Domestic Product (GDP), more than any other category of health expenditure.¹ A large fraction of US hospital expenditures (40%) are financed with private health insurance or patient out-of-pocket payments. In recent years, the growth of privately financed hospital expenditures has been driven almost entirely by hospital price increases.² In most states, hospital prices charged to private health insurance companies are unregulated and determined by negotiations between hospitals and health insurance companies. The negotiated prices are determined in large part by local competitive conditions and the ability of health insurance companies to substitute with competing hospitals in their managed care networks. Hospital antitrust enforcement plays a significant role in US healthcare

*University of Missouri—Kansas City; garmonc@umkc.edu.

This research was conducted while I was employed by the Bureau of Economics, Federal Trade Commission. The views expressed in this article are the author's and not necessarily those of the Commission or any individual Commissioner. I thank anonymous reviewers, Keith Brand, Leemore Dafny, Dan Hosken, Devesh Raval, Sean May, Seth Sacher, David Schmidt, Loren Smith, Michael Vita, Nathan Wilson, Zenon Zabinski, conference and seminar participants at the International Industrial Organization Conference, FTC Microeconomics Conference, DC Industrial Organization Conference, Kenyon College, and RAND Health for their suggestions, and Michael Bohne, Chris Carman, Laura Kmitch, and Jordan Rhodes for their research assistance.

¹Centers for Medicare and Medicaid Services, Historical National Health Expenditure Data, www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsHistorical.html (accessed on 6/6/2017).

²2015 Health Care Cost and Utilization Report, Health Care Cost Institute, www.healthcostinstitute.org (accessed on 6/6/2017).

cost containment by preserving hospital competition and limiting hospital price growth, while also promoting quality and access to healthcare.

Over the past 25 years, significant changes have occurred in US hospital antitrust enforcement. Between 1993 and 2000, during the largest hospital merger wave in US history, federal and state antitrust authorities challenged eight proposed hospital mergers in federal court and failed in each attempt. This string of setbacks led to an explosion of research on hospital competition and the effects of hospital mergers. One branch of the literature retrospectively studied the effects of past hospital mergers and found that the tools and assumptions upon which courts relied during the 1990s often led to incorrect conclusions about the likely effects of hospital mergers. Another branch of the literature attempted to model price formation in hospital markets and developed a set of tools to directly assess the lost competition from hospital mergers and predict their price effects. These tools (e.g., diversion ratios, willingness-to-pay, upward pricing pressure, and merger simulation) were used by the federal antitrust agencies in recent hospital merger challenges and, unlike the 1990s, many of these challenges have been successful.

The use of tools that directly measure lost competition in hospital antitrust enforcement has occurred alongside the general evolution of merger review in differentiated product markets and the increasing reliance on direct measures of lost competition by the federal antitrust agencies in the United States. The Federal Trade Commission (FTC) and the Department of Justice (DOJ) revised their Horizontal Merger Guidelines (HMG) in 2010 to emphasize direct measures of competition (e.g., diversion ratios and the value of diverted sales) and de-emphasize traditional concentration measures (e.g., the Herfindahl-Hirschman Index [HHI]) in differentiated product markets.³

With the recent use of new screening tools in hospital antitrust enforcement and the emphasis on similar direct measures of competition in the review of mergers in other differentiated product markets, it is important to evaluate whether these new tools are accurate in predicting post-merger price changes. The original articles that developed the hospital screening tools did not assess the accuracy of their predictions against actual post-merger outcomes. This article offers the first comprehensive comparison of the predictions of a wide range of screening tools against the actual post-merger price changes of a relatively large sample of hospital mergers. The actual post-merger price changes (measured relative to controls) of 28 hospital mergers are compared to the predictions of various screening methods. The screening methods include direct measures of the competition between the merging parties (i.e., diversion ratios, upward pricing pressure [UPP], and willingness-to-pay [WTP]), merger simulation, and traditional concentration measures (i.e., Herfindahl-Hirschman Index [HHI]) calculated with various market definitions and market share metrics.

The focus of the analysis is on evaluating methods that can be implemented with data that are likely available to regulators during the initial preliminary investigation of a merger. It is at this stage that delineating between possible anticompetitive mergers and beneficial or innocuous mergers is most useful and imposes the least regulatory cost. Although a full-phase investigation can provide the regulator with detailed data and other evidence to increase the precision of its estimates, a full-phase investigation imposes significant costs on the merging parties and the regulator. The ideal screen for an initial investigation avoids “casting a wide net” and instead focuses the regulator on the mergers most likely to be anticompetitive. All of the screening methods evaluated in this article can be calculated with data often available without a full-phase investigation: patient discharge data and other public data sets. It is important to note that this excludes merger simulations calibrated with health insurance claims data, such as Gowrisankaran, Nevo, and Town (2015). This article evaluates only merger simulations calibrated with less-detailed hospital and discharge data.

³ For example, from Section 6.1 of the 2010 HMG: “The Agencies rely much more on the value of diverted sales than on the level of the HHI for diagnosing unilateral price effects in markets with differentiated products.”

Any evaluation of merger screening methods is complicated by active antitrust enforcement. In an era of active and effective hospital antitrust enforcement, most mergers that are likely to be anticompetitive or balance are blocked or never proposed. Thus, a sample of consummated mergers taken from a period of active antitrust enforcement may be truncated and biased toward mergers with limited reductions in competition and significant procompetitive effects (e.g., cost savings) (Carlton, 2009). To address this issue, our sample of consummated hospital mergers includes 12 mergers in North Carolina and Missouri that occurred between 1997 and 2001. This period was at the tail end of the federal and state hospital antitrust losing streak and before the successful hospital merger challenges of recent years. In addition, North Carolina introduced a hospital Certificate of Public Advantage (COPA) regulatory program in 1995 that gave merging hospitals participating in the program antitrust immunity. Only one pair of merging hospitals participated in North Carolina's COPA program,⁴ but the option to participate, coupled with recent court rulings favoring hospital mergers, likely contributed to an environment in which competing hospitals felt safe to merge with less risk of an antitrust challenge. The Missouri hospital mergers in the sample include one merger challenged by the FTC, but allowed by the courts. We also address the bias from antitrust enforcement by directly measuring post-merger variable cost changes, as well as price changes, and focusing on the mergers that were not associated with significant efficiencies.

Analyzing hospital mergers from North Carolina and Missouri in the late 1990s and early 2000s may lessen the truncation problems caused by antitrust enforcement. However, the hospital industry has undergone many changes since the early 2000s, potentially limiting the applicability of findings from that period. Some have argued that methods used in hospital merger review and enforcement should evolve and account for the changes in healthcare delivery and finance that have occurred since the passage of the Affordable Care Act (Guerin-Calvert, Maki, and Vladeck, 2015). To address these concerns and test the accuracy of hospital merger screens in this potentially new regime, our sample of hospital mergers also includes 16 recent transactions from 2007–2012.

The comparison of actual post-merger price changes against the pre-merger predictions of the screening tools reveals that, apart from merger simulation, the new screening tools (in particular, WTP and UPP) are more accurate than traditional concentration measures at flagging potentially anticompetitive hospital mergers for further review. However, the relationship between the new screening tools and the post-merger price changes is not precise or robust to alternate price change measurements, so care should be taken when using the tools to screen mergers for further investigation. Merger simulation performs poorly, but this may be due to the limited data available to calibrate the simulation in the initial investigation. Among the traditional concentration measures, those that employ market shares based on flexible geographic boundaries are more accurate at predicting post-merger price changes than concentration measures based on fixed boundaries.

The article is arranged as follows. Section 2 reviews hospital antitrust enforcement over the past 20 years, and the hospital competition literature that developed alongside it. Section 3 describes the evaluated screening tools in detail. Section 4 describes the data, the criteria for merger selection, price measurement and price change estimation, and the construction of the screening tools. Section 5 compares the screening tools to the post-merger price changes, and Section 6 concludes.

⁴ In December 1995, Memorial Mission Hospital and St. Joseph's Hospital—the only two short-term, general acute care hospitals in Asheville, NC—entered into a joint management agreement to form Mission Health and simultaneously entered into a COPA agreement with the state of North Carolina, granting the merger antitrust immunity in exchange for regulation of Mission Health by the state. This merger is not included in the sample of hospital mergers analyzed in this article.

2. Literature and case review

■ Starting with the FTC's failed attempt in 1994 to block the merger of the only two hospitals in Ukiah, California, the federal and state antitrust authorities unsuccessfully challenged eight hospital mergers between 1994 and 2001 (Ashenfelter et al., 2011). In six of the eight challenges, the courts found that the proposed merger was unlikely to reduce competition significantly because of the presence of numerous remaining competitors. These determinations were based on the courts' acceptance of relatively large geographic antitrust markets established using the Elzinga-Hogarty (EH) test (Elzinga and Hogarty, 1973), and Critical Loss Analysis (CLA).

The EH test posits that the relevant geographic market for antitrust analysis is the area for which inflows (i.e., sales by firms in the area to customers from outside the area) and outflows (i.e., sales by firms outside the area to customers living in the area) are sufficiently small. The two most common EH inflow/outflow thresholds used for market definition are 25% (a "weak" EH market, i.e., if the inflows into and outflows from an area are both less than 25%) and 10% (a "strong" EH market). Operationally, to determine an EH market for a particular merger, one would first find the smallest area from which the merging firms, and other nearby firms, draw 75% (or 90% for the "strong" standard) of their customers. If more than 25% (or 10%) of the customers who live in this area go outside to purchase the good, areas are added to the base draw area until the inflows and outflows are both below 25% (or 10%).

Critical Loss Analysis (CLA) is another related method for defining geographic markets in hospital merger challenges using patient flows. CLA calculates the loss of patients above which a small price increase (e.g., 5%) would be unprofitable for a hypothetical owner of all of the hospitals in an area (i.e., the "critical loss"). If estimates of the actual loss in response to the price increase exceed the critical loss, adjacent areas and hospitals are added to the market until the estimated actual loss no longer exceeds the critical loss.

Although the EH algorithm and CLA do not necessarily produce a unique area, the ubiquity of patient discharge data and the relative ease with which EH/CLA markets can be calculated with patient discharge data made the EH and CLA methods widespread in hospital merger challenges in the 1990s. However, as became apparent in the hospital merger challenges of the 1990s, the EH test and CLA often produce extremely large geographic hospital markets, particularly when the "strong" 10% criterion is applied in the EH test. In urban areas, the weak EH criterion will almost always result in a geographic market encompassing the entire metropolitan area, and the strong criterion will often produce a market larger than the metropolitan area. For example, in overturning a lower court's ruling that the merger of the only two hospitals in Poplar Bluff, Missouri, would be anticompetitive, the Eighth Circuit US Court of Appeals found that the relevant market included competing hospitals in Cape Girardeau (85 miles away from Poplar Bluff) and St. Louis (150 miles away) because significant numbers of patients in the merging parties' service area sought treatment in Cape Girardeau and St. Louis. Echoing the defendants' CLA arguments, the Eighth Circuit US Court of Appeals concluded that "the compelling and essentially unrefuted evidence that the switch to another provider by a small percentage of patients would constrain a price increase, shows that the FTC's proposed market is too narrow."⁵

In two of the eight unsuccessful hospital merger challenges in the 1990s, the merging parties argued—and the courts agreed—that the merging parties would not exercise any additional market power obtained through the merger because they were nonprofit hospitals. Together, the courts' use of large EH/CLA-inspired geographic markets and their limited acceptance of the merging parties' nonprofit defense prevented federal and state antitrust authorities from blocking proposed hospital mergers they felt were anticompetitive. This spurred health economists to study the effects of hospital competition. Starting in the mid-1990s, a large hospital competition literature developed along two tracks. In the first track, economists empirically measured the

⁵ United States Court of Appeals, Eighth Circuit. Federal Trade Commission; State of Missouri, by and through its Attorney General, Jeremiah W. (Jay) Nixon, v. Tenet Health Care Corporation; Poplar Bluff Physicians Group, Inc., doing business as Doctors Regional Medical Center; 186 F.3d 1045.

cross-sectional relationship between hospital competition and outcomes (both price and quality) and retrospectively analyzed past hospital mergers to study their effects. Vogt and Town (2006) summarize this literature and conclude regarding price that “the great weight of the literature shows that hospital consolidation leads to price increases, although a few studies reach the opposite conclusion. Studies that examine consolidation among hospitals that are geographically close to one another consistently find that consolidation leads to price increases of 40% or more.” Summarizing the literature on hospital competition and quality, Vogt and Town (2006) conclude that “on balance, the evidence suggests that increasing hospital concentration lowers quality.” Further, most analyses of hospital competition found a positive correlation between concentration and price, even among nonprofit hospitals. In addition, numerous retrospective studies of mergers of competing nonprofit hospitals found significant post-merger price increases, casting doubt on the argument that nonprofit hospitals do not exercise post-merger market power (Vita and Sacher, 2001; Haas-Wilson and Garmon, 2011; Tenn, 2011).

Although the cross-sectional and retrospective hospital competition literature severely undermined the logic behind the courts’ rulings in the 1990s, it did not provide tools to replace the EH test and CLA. In the second track of the new hospital competition literature, economists modelled hospital markets and developed new screening methods for hospital mergers. Town and Vistnes (2001) and Capps, Dranove, and Satterthwaite (2003) developed a new market power measure for hospitals—commonly referred to as willingness-to-pay (WTP)—from a bargaining model of the negotiation between health insurance companies and hospitals. Gaynor and Vogt (2003) developed a Bertrand model of hospital price competition and, from this model, Antwi, Gaynor, and Vogt (2013) derived a market power measure for hospitals, which they named the Logit Competition Index (LOCI). Both WTP and LOCI are based on the first-order pricing incentives of hospitals. In that regard, they are similar to upward pricing pressure (UPP) (Farrell and Shapiro, 2010; Jaffe and Weyl, 2013), which is often used to measure the potential lost competition from a merger in differentiated product markets. Like UPP, both WTP and LOCI predict significant lost competition after a hospital merger when the merging hospitals are close substitutes, as measured by diversion ratios. Both WTP and LOCI can also be used as the basis for reduced-form hospital merger simulations. Gowrisankaran, Nevo and Town (2015) recently developed a generalized model of hospital price formation in which the WTP and LOCI-based models are special cases.

In recent years, WTP, LOCI, UPP, diversion ratios, and merger simulation have been used in hospital and physician antitrust litigation and regulation in both the United States and abroad. Table 1 lists the screening tools used in recent hospital and physician merger challenges brought before federal district courts in the United States, apart from the traditional concentration measures (e.g., HHI) used in all merger challenges. Diversion ratios and WTP were used by the plaintiff’s economic expert in all of the recent challenges, and UPP was used in the two most recent challenges. In all of these cases, the antitrust agencies were ultimately successful in blocking the mergers. Even though LOCI has not been used in a healthcare provider merger challenge in the United States, the UK Competition Commission recently used LOCI to measure competition as part of its investigation into the private hospital industry in Britain.⁶

To date, there has been limited research on the accuracy of the new screening tools despite their widespread use in antitrust challenges and regulation. Four recent articles have explored the accuracy of the most widely used of the new hospital merger screening tools, willingness-to-pay (WTP). Fournier and Gai (2007) find that post-merger WTP changes estimated using pre-merger data are accurate predictors of actual post-merger WTP changes that occurred after two hospital mergers. For one of these mergers, they also find that the price change implied by a WTP-based merger simulation using pre-merger data produced a conservative estimate of the actual post-merger price change. (Data limitations prevented the measurement of the post-merger price

⁶ www.competition-commission.org.uk/assets/competitioncommission/docs/2012/private-healthcare-market-investigation/ais_app_b_toh_1_annex_2_loci_note_housestyled.pdf (accessed on March 21, 2014).

TABLE 1 Recent Healthcare Provider Merger Challenges in Federal District Court

Plaintiffs ^a	Defendants	Year	New Screening Tools Used ^b	Outcome
FTC/State of Ohio	ProMedica Health/St. Luke's Hospital	2011	Diversion ratios, WTP, merger simulation	Preliminary Injunction (PI) granted ^c
FTC	OSF Healthcare/Rockford Health	2012	Diversion ratios, WTP	PI granted ^d
FTC/State of Idaho	St. Luke's Health/Saltzer Medical Group	2013	Diversion ratios	PI granted ^e
FTC/Commonwealth of Pennsylvania	Pinnacle Health/Penn State Hershey	2016	Diversion ratios, WTP, UPP	PI granted ^f
FTC/State of Illinois	Advocate Health Care/NorthShore University Health	2016	Diversion ratios, WTP, UPP	PI granted ^g

^aThis list excludes the FTC and State of Georgia's challenge of the acquisition of Palmyra Medical Center by the Phoebe Putney Health System because the courts' decisions at all levels (district, appeals, and Supreme) in that case were based on legal state-action issues, not on the competitive effect of the acquisition.

^bIn most cases, the new screening tools were used by the plaintiffs' economic expert. In the Pinnacle Health/Penn State Hershey trial, diversion ratios and UPP were used by both the plaintiffs' and defendants' economic experts.

^cwww.ftc.gov/enforcement/cases-proceedings/101-0167/promedica-health-system-inc-corporation-matter (accessed on September 1, 2016).

^dwww.ftc.gov/enforcement/cases-proceedings/d-9349-111-0102/osf-healthcare-system-rockford-health-system (accessed on September 1, 2016).

^ewww.ftc.gov/enforcement/cases-proceedings/121-0069/st-lukes-health-system-ltd-saltzer-medical-group-pa (accessed on September 1, 2016).

^fwww.ftc.gov/enforcement/cases-proceedings/141-0191-d09368/penn-state-hershey-medical-center-ftc-commonwealth (accessed on September 1, 2016). The federal district court initially denied the plaintiffs' motion for a preliminary injunction, but this decision was subsequently reversed on appeal.

^gwww.ftc.gov/enforcement/cases-proceedings/1410231/ftc-v-advocate-health-care-network (accessed on June 2, 2017). The federal district court initially denied the plaintiffs' motion for a preliminary injunction, but this decision was subsequently reversed after the case was remanded back to the district court by the Seventh Circuit Court of Appeals.

change for the second merger in their study.) May and Noether (2014) compare the predictions of WTP changes estimated using pre-merger data against the actual post-merger price changes of two hospital mergers, and find that the merger predicted to have the largest post-merger price increase had the smallest actual price increase of the two mergers. Dranove and Ody (2016) evaluate the accuracy of WTP, HHI, and market shares in explaining cross-sectional differences in hospital prices due to market power, but they do not evaluate merger predictions based on these measures. Brand and Balan (2013) conduct a Monte Carlo-like exercise in which they compare the predictions of various merger screens (including WTP) against data produced by a bargaining model of the negotiations between hospitals and health insurance companies, and find that diversion ratios, WTP changes, and merger simulation produce accurate predictions of post-merger price changes simulated by the bargaining model. Although this finding implies the new hospital merger screening tools are theoretically sound, the evidence from Fournier and Gai (2007) and May and Noether (2014) comparing the predictions of the WTP screen against actual post-merger price changes is mixed and limited to a meta-sample of only three mergers. Apart from May and Noether (2014) and Fournier and Gai (2007), there has been little research on the accuracy and reliability of WTP, LOCI, UPP, diversion ratios, and hospital merger simulation in predicting the price effects of actual hospital mergers.⁷

Outside the realm of hospital mergers, some research exists on the accuracy of UPP and merger simulation when applied to other differentiated product markets. A limited literature (e.g.,

⁷ Ron Kemp, a senior economic officer for the Dutch Authority for Consumers and Markets, has recently compared the post-merger price changes of 12 hospital mergers in the Netherlands (described in Kemp, Kersten, and Severijnen, 2012) with the pre-merger price increase predictions of LOCI. This comparison is described in the April 22, 2015 presentation "Ex-Post Analysis of Dutch Hospital Mergers," available at www.oecd.org/daf/competition/workshop-expost-evaluation-competition-enforcement-decisions.htm (accessed on June 2, 2015).

Peters, 2006; Weinberg, 2011; Weinberg and Hosken, 2013) tests the pre-merger predictions of merger simulations against actual post-merger price changes across a variety of industries with mixed results. We are not aware of any studies that evaluate the accuracy of UPP using actual post-merger outcomes. However, Miller et al. (2017) use Monte Carlo methods to evaluate the accuracy of UPP and find that it is a good predictor of post-merger price changes, even absent pass-through adjustment.

The current article makes a number of significant contributions to the literature. First, we significantly add to the sample size of hospital mergers considered by Fournier and Gai (2007) and May and Noether (2014), increasing the likelihood of a meaningful evaluation of the accuracy of hospital merger screening tools. Second, this article provides the only evaluation to date of the accuracy of UPP using actual post-merger outcomes for hospital mergers in the United States. Third, the analysis is not confined to the evaluation of a particular screen in isolation (as is common in the literature), but instead compares the predictions of various screening methods. Although it is useful to assess the absolute accuracy of a screening tool, it is more important to evaluate its relative accuracy compared with traditional screening methods. Do any of the new screening tools provide information beyond the traditional market-based screening methods that can help more accurately predict the effects of hospital mergers? If so, are some of the new screening tools better than others at providing this additional information?

Finally, one advantage of the new hospital merger screening tools is that most (e.g., diversion ratios, UPP, and WTP) can be implemented without the traditional exercise of defining product and geographic markets and calculating market shares and concentration measures, such as HHI. However, courts continue to require the definition of a relevant antitrust product and geographic market as part of any merger challenge. Gaynor, Kleiner, and Vogt (2013) use hospital merger simulation methods to define antitrust markets and find that markets implied by these methods are narrower than markets defined using traditional approaches like EH and CLA. However, they do not test merger simulation predictions or the predictions of concentration measures against actual post-merger outcomes. In addition to evaluating the relative accuracy of the new screening tools, this article also evaluates the accuracy of the traditional market power measure, HHI, by comparing predictions of post-merger HHI levels and changes to actual post-merger price changes, under various market definitions and share metrics.

All of the screening tools are evaluated by comparing their predictions for each merger to the merger's actual price change. Other potentially important effects of mergers, such as changes in quality or access to care, are not evaluated.

The following section describes the new hospital merger screening tools in more detail, as well as traditional concentration measures used as a benchmark for comparison.

3. Merger screening methods

■ The merger screening methods most commonly used in healthcare provider antitrust enforcement fall into one of three categories. The first are methods that are based on the first-order necessary conditions describing profit-maximizing price formation in equilibrium models of provider competition (e.g., diversion ratios, LOCI, UPP, and WTP). On their own, these methods typically do not provide a specific prediction of a post-merger price change. Instead, they produce measures that should be proportional to post-merger price changes, all else equal. The second type of screening method is hospital merger simulation, which produces a specific prediction of the post-merger price change. Hospital merger simulations are either reduced-form simulations or fully specified simulations. Reduced-form simulations involve estimating the relationship between a market power measure (e.g., WTP) and price and then using the estimated relationship to predict the post-merger price change (e.g., Capps, Dranove, and Satterthwaite, 2003). Fully specified simulations estimate the parameters of a provider competition model and then use the estimated parameters to predict the post-merger price change (e.g., Gowrisankaran, Nevo, and Town, 2015). The feasibility of hospital merger simulation depends on the quality and detail of the

data available and other factors (Brand and Garmon, 2014). If a merger simulation is conducted, it usually occurs during a full-phase merger investigation. This is because fully specified merger simulations require detailed, payer-specific claims data, which are available only in full-phase investigations. Testing the accuracy of fully specified merger simulations is beyond the scope of this article, due to data availability. Finally, hospital merger screening methods include traditional concentration measures (e.g., HHI) calculated from hospital shares of a defined product and geographic market.

□ **Willingness-to-pay (WTP).** Willingness-to-pay (WTP) was developed by Town and Vistnes (2001) and Capps, Dranove, and Satterthwaite (2003) from a bargaining model of the negotiation between a managed care organization (MCO) and a hospital over the contractual price per admission paid by the MCO for its members seeking care at the hospital. Consider MCO k that negotiates with each hospital in the set Φ of hospitals that would provide positive value to k 's network of hospitals. We assume there are no impediments to reaching an agreement (e.g., asymmetric information, negotiation deadlines with random communication delays, etc.) if a positive surplus is available, so that, in equilibrium, each hospital in Φ will reach an agreement with k . The focus of the model is on the negotiation between k and hospital $j \in \Phi$, assuming that the outcomes of the negotiations between k and the remaining hospitals Φ_{-j} are taken as given (e.g., because of simultaneous negotiations). Assume that MCO k 's profits are given by:

$$\pi_k = \rho_k X_k - C_k(X_k) - \sum_{h \in \Phi} p_h y_{kh}, \tag{1}$$

where ρ_k is k 's health insurance premium, X_k is k 's total membership demand (which is a function of ρ_k , Φ , and the premiums and network configurations of its competitors), C_k are k 's nonhospital costs, $p = \{p_i\}_{i \in \Phi}$ is the vector of hospital prices negotiated by MCO k and the hospitals in its network, and y_{kh} is the number of k 's patients treated at hospital h . Denote as R_{kj} k 's profits apart from payments to hospital j (i.e., $R_{kj} = \pi_k + p_j y_{kj}$). If we assume that the negotiation between MCO k and hospital j follows generalized Nash bargaining and we assume, without loss of generality, that hospital j loses all of k 's members if it is not part of k 's network, then the price that k and j negotiate will solve:

$$\max_{p_j} \{ [R_{kj}(\Phi) - \pi_k(\Phi_{-j}) - p_j y_{kj}]^{(1-\gamma)} [p_j y_{kj} - c_j(y_{kj})]^\gamma \}, \tag{2}$$

where $c_j(y_{kj})$ are hospital j 's costs of serving k 's patients and γ is a split parameter reflecting the relative bargaining abilities of the hospital and MCO. If we assume that hospital j 's price to k does not affect a member's demand for hospital j , as long as j is in k 's network (e.g., k 's health plan design is a PPO that charges the same copay for in-network hospitals), then the price that solves (2) is:

$$p_j = \frac{\gamma(R_{kj}(\Phi) - \pi_k(\Phi_{-j})) + (1 - \gamma)(c_j(y_{kj}))}{y_{kj}}. \tag{3}$$

As seen in (3), hospital j 's market power is proportional to $(R_{kj}(\Phi) - \pi_k(\Phi_{-j}))$, the additional profit k receives from having hospital j in its network. Town and Vistnes (2001) and Capps, Dranove, and Satterthwaite (2003) proxy for $(R_{kj}(\Phi) - \pi_k(\Phi_{-j}))$ with the aggregate consumer surplus hospital j adds to k 's network. The change in consumer surplus associated with j 's inclusion into k 's network is WTP. Assume that, conditional on needing hospitalization, each of k 's members have preferences over the hospitals in Φ of the form $U_{hi} = V_{hi} + \epsilon_i$ where V_{hi} is a linear function of hospital characteristics and the stochastic term ϵ_i is independently and identically distributed according to the extreme value distribution (i.e., the distributional assumption consistent with logit estimation). In this case, WTP for hospital j is defined as the aggregate change in consumer

surplus associated with adding hospital j to k 's network:

$$\begin{aligned} WTP_{kj} &= \sum_{i \in I_k} \frac{1}{\alpha_i} \left[E \left[\max_{h \in \Phi} (V_{hi} + \epsilon_i) \right] - E \left[\max_{h \in \Phi_{-j}} (V_{hi} + \epsilon_i) \right] \right] \\ &= \sum_{i \in I_k} \frac{1}{\alpha_i} \left[\ln \left(\frac{1}{1 - s_i^j} \right) \right], \end{aligned} \quad (4)$$

where α_i is the marginal utility of income for patient i and s_i^j is the probability that patient i chooses hospital j . The predicted probabilities associated with the conditional logit estimation of V_{ji} are used as estimates of s_i^j . We cannot directly observe α_i , so WTP is operationalized by ignoring α_i , although it is subsumed within the estimated WTP coefficient of WTP-based merger simulations (described below). Furthermore, individual MCOs usually cannot be observed in the discharge data most commonly used to estimate WTP, so WTP is typically estimated across all commercial MCOs:

$$WTP_j = \sum_i \left[\ln \left(\frac{1}{1 - s_i^j} \right) \right], \quad (5)$$

WTP can be used to analyze the market power created by the merger of two competing hospitals (or hospital systems) by measuring the net change in WTP associated with the combination of the two hospitals (or systems). When used in this way, it is implicitly assumed that the combined hospitals will negotiate in an all-or-nothing manner (i.e., in order to contract with either hospital, the MCO must contract with both). This collective negotiation by multiple competing hospitals worsens the MCO's threat point in (2). For example, consider the merger of two competing hospitals, L and M. Before the merger, if the MCO fails to reach an agreement with L, the loss in welfare for the MCO's members may be relatively small with M available as an alternative in the payer's network. Post-merger, if L and M negotiate on an all-or-nothing basis, the loss in welfare if the MCO fails to contract with both hospitals will be greater than the sum of the losses associated with each hospital individually. This is because L and M are competing hospitals, in the sense that some of the MCO's members who prefer L see M as an alternative and vice versa. If there were no members for which this were true, the WTP of L and M would equal the sum of the WTP of L and the WTP of M and there would be no net increase in WTP associated with the merger. In this way, a merger of competing hospitals, along with post-merger all-or-nothing negotiation by the merged hospital system, may lead to a disproportionate worsening of the MCO's threat point, resulting in a price increase. The net change in WTP with the merger can be used as a measure of the worsening of the MCO's threat point.

□ **Diversion ratios and upward pricing pressure (UPP).** A merger of competing hospitals can also lead to a price increase, even if the hospitals do not negotiate in an all-or-nothing manner and continue to negotiate with the MCO separately. In this case, the merger improves the threat point of the hospital in (2). For example, before the merger, if L fails to reach an agreement with the MCO, some of the MCO's current patients who prefer L will instead seek care at M. Post-merger, when L is negotiating with the MCO, both parties know that failure to reach an agreement will result in less real diversion from L, as those patients who switch to M remain internal to the combined entity. This improved threat point will lead L to be more aggressive in the negotiation, resulting in a higher price despite the lack of all-or-nothing bargaining from L and M. A similar post-merger dynamic exists in M's negotiation with the MCO. This effect of the merger, due to the change in each hospital's threat point, occurs because of the collective ownership of both hospitals, even if the negotiations remain separate after the merger.

Whether the post-merger negotiations are collective or separate, both of the merger effects are driven by the potential diversion between the merging hospitals, which is a measure of the substitutability of the hospitals in the eyes of the MCO's members. In other words, the effect of the

merger on the negotiated price should be proportional to the number of patients who would switch from a hospital to its merger partner in the event the former is dropped from the MCO's network. Therefore, another measure of the lost competition between two merging hospitals is the diversion ratio: the percentage of the patients treated at a hospital who would go to its merger partner if the former is dropped from the MCO's network. Using the predicted probabilities associated with the conditional logit estimation of V_{ji} and the logit property that diversion is proportional to the probability of selection for each patient type, the diversion ratio from hospital L to hospital M is:

$$d_{LM} = \frac{\sum_i \left(\frac{s_i^L s_i^M}{1-s_i^L} \right)}{\sum_i s_i^L} \tag{6}$$

This diversion ratio calculation implicitly assumes that no patients would continue to seek treatment at L if it is dropped from the MCO's network. Otherwise, the diversion ratio could be calculated by incorporating the probability that each patient type would stay with his or her preferred hospital if it is dropped from the network.

Returning to the previous bargaining model between the hospital and the MCO, assume that the hospital's costs of treating the MCO's patients are proportional to volume (i.e., $c_j(y_{kj}) = c_j y_{kj}$). If hospitals L and M merge, but the combined system negotiates on behalf of L separately from M (i.e., not in an all-or-nothing manner), then the price for L negotiated between the merged firm and MCO k will satisfy:

$$\max_{p_L} \left\{ [R_{kL}(\Phi) - \pi_k(\Phi_{-L}) - p_L y_{kL}]^{(1-\gamma)} [(p_L - c_L) y_{kL} - (p_M - c_M) d_{LM} y_{kL}]^\gamma \right\} \tag{7}$$

The first-order condition for (7) implies that L's post-merger price satisfies:

$$p_L^{post} = \frac{\gamma(R_{kL}(\Phi) - \pi_k(\Phi_{-L})) + (1-\gamma)(c_j y_{kL})}{y_{kL}} + (1-\gamma)(p_M - c_M) d_{LM} \tag{8}$$

Along with (3), this implies that the post-merger increase in price (expressed as a percentage of the pre-merger price p_L) is:

$$\frac{p_L^{post} - p_L}{p_L} = (1-\gamma) \frac{(p_M - c_M)}{p_M} \left(\frac{p_M}{p_L} \right) d_{LM} \tag{9}$$

This is analogous to the standard UPP post-merger price increase approximation, except that diversion in this case is in response to the hypothetical exclusion of L from the MCO's network and pass-through (of UPP to price) is a function of the bargaining split parameter.⁸

Gaynor and Vogt (2003) and Antwi, Gaynor, and Vogt (2013) do not use a bargaining model as the basis for their measure of hospital market power, but instead develop a measure from a Bertrand model of hospital price competition, which they refer to as the Logit Competition Index (LOCI). They also assume that patients have preferences over hospitals of the form $U_{hi} = V_{hi} + \epsilon_i$, where V_{hi} is a linear function of hospital characteristics (including price) and the stochastic term ϵ_i is independently and identically distributed according to the extreme value distribution. Using this model, Antwi, Gaynor, and Vogt (2013) develop a first-order approximation of the price increase associated with the merger of two hospitals, which is closely related to the UPP price increase approximation in (9). Because the LOCI-based price increase approximation is closely related to the UPP approximation, we report only the latter in the merger screens that are evaluated.⁹

□ **Reduced-form merger simulation.** Returning to the bargaining model between the hospital and the MCO, equation (3) can also be used as the basis of a reduced-form merger simulation.

⁸ This derivation is adapted from Haas-Wilson and Garmon (2009).

⁹ However, the LOCI price increase approximation for each merger is reported in the Online Appendix C, available at www.researchgate.net/publication/281494031_The_Accuracy_of_Hospital_Merger_Screening_Methods.

WTP per adjusted discharge WTP_PAD (where the adjustment accounts for variations in acuity across patients) can be used as a proxy for $((R_{kj}(\Phi) - \pi_k^*(\Phi_{-j}))/y_{kj})$ in equation (3) and form the basis of a reduced-form econometric model:

$$p_h = \gamma + \beta_1 WTP_PAD_h + \beta_2 c_h + \beta X_h + \varepsilon_h, \quad (10)$$

where p_h is the case-mix-adjusted price of hospital system h , c_h is the average variable cost of hospital system h , and X_h is a vector of other determinants of price. The coefficient of WTP_PAD is estimated (e.g., via OLS) and then used along with the predicted post-merger change in WTP_PAD to estimate the post-merger price change.

As described in Brand and Garmon (2014), the usefulness of a reduced-form merger simulation may be limited by the data available to estimate (10). In most cases, only cross-sectional data are available to estimate (10), and the estimated WTP_PAD coefficient may suffer from omitted variable bias if there are factors related to the hospital/MCO negotiation that cannot be observed or measured. This bias may be compounded in the context of an initial investigation when payer data are not available and only hospital-level price estimates can be calculated. Furthermore, the use of accounting data to measure average variable cost may introduce endogeneity bias as nonprofit hospitals with market power may classify some profits as costs. (Robinson (2011))

□ **Herfindahl-Hirschman Index.** For each hospital merger, the new merger screening tools (WTP, UPP, and merger simulation) are juxtaposed against the traditional measure of market power used by the antitrust agencies and the courts: the Herfindahl-Hirschman Index (HHI) or the sum of the squared market shares (s_h) for the hospitals in the market (M):

$$HHI_M = \sum_{h \in M} (s_h)^2. \quad (11)$$

The HHI depends on the definition of the product and geographic market M , and the method and metric used to calculate the market shares. One significant conceptual benefit of the new tools over the HHI is that none of the new tools require a product or geographic market for calculation, except for restrictions on products and hospitals necessary for the estimation of the choice model.

Three geographic markets and share calculation methods are used to calculate HHIs. First, as a conservative approximation of the concentration measures used by the courts in the hospital merger challenges of the 1990s, an HHI is calculated based on the Hospital Referral Region (HRR) (defined by the Dartmouth Atlas of Health Care) of the merging hospitals with shares based on the staffed beds of the hospitals located within this HRR. The HRR is used instead of an EH-defined market because the EH procedure will not necessarily produce a unique area for each merger. However, each HRR is roughly similar to a hospital's 90% service area, as it is designed to capture the market for high-acuity services. The retrospective hospital competition literature has found that geographic markets of this size are often too large to correctly predict the effects of a hospital merger with an HHI. In addition, this share calculation method suffers from the "all-in-or-all-out" problem. Hospitals within the market are factored into the HHI calculation with their full capacity, even if they are not located near the merging hospitals, and hospitals located outside of the market are not counted at all, even if they serve many of the patients living near the merging hospitals. To rectify these problems, we calculate two alternative HHIs.

The second HHI uses the Hospital Service Area (HSA) (defined by the Dartmouth Atlas) of the acquired hospital and calculates the market shares based on the patients residing in the area, not the hospitals located within the area. The HSA is typically smaller than the HRR and is meant to capture the market for low- and medium-acuity cases. Shares are calculated based on the admissions of patients residing in the area, even if they are treated at hospitals outside of the area. Defining the geographic market with the HSA is still arbitrary, so the final HHI calculation employs a weighted service area (WSA). For each zip code, the share of each hospital is calculated based on the patients who reside in that zip code (regardless of the location of the hospital). Then, these shares are weighted based on the importance of the zip code to the merging hospitals (i.e., the

weights are the percentage of the combined hospitals' admissions that come from the zip code). The HHI is then calculated as the sum of the squares of the weighted shares. This HHI completely solves the "all-in-or-all-out" problem by including all patients and all hospitals, regardless of location, but doing so in a way that focuses on the area most important to the merging hospitals. This measure is also similar to weighted concentration measures commonly used in the hospital competition literature (e.g., Capps and Dranove, 2004). Appendix A provides an example of each HHI using a simple hypothetical hospital market.

For the latter two HHIs, the product market used to calculate the market shares consists of services to commercially insured patients with general acute care (GAC) conditions who are treated at short-term, GAC hospitals. Excluded from the product market are services not offered at both the acquired hospital and primary acquiring hospital, to isolate the services that form the basis of competition between the merging hospitals.¹⁰ This product market definition is similar to the "cluster" product markets established by the courts in most of the hospital merger challenges over the past 25 years.

4. Data and estimation

■ All of the hospital merger screening methods described above can be implemented with patient-level inpatient discharge data and data on the characteristics of the merging hospitals and potential competitors. The discharge data used for the analysis came from the Arkansas Department of Health (2007–2011), the Connecticut Department of Public Health, Office of Health Care Access (2007–2013), the Georgia Hospital Association (2007–2013), the Oklahoma Department of Health (2007–2011), the Pennsylvania Health Care Cost Containment Council (PHC4) (2007–2013), the New York Department of Health (2007–2012), the Missouri Hospital Association (1996–2003), and the company formerly known as Solucient for North Carolina (1997–2002). The discharge data are restricted to patients treated at nonfederal, short-term, GAC hospitals. In other words, patients treated at federal hospitals (Veterans Affairs or military), long-term acute care hospitals, rehabilitation hospitals, and psychiatric and substance abuse facilities are excluded. In addition, patients treated at nonfederal, short-term GAC hospitals for non-GAC conditions (i.e., rehabilitation, psychiatry, and substance abuse) are excluded. Finally, newborns, patients transferred from other hospitals, and court-ordered admissions are excluded to avoid double-counting patient choices or counting admissions that were mandatory.

Hospital characteristics are taken from the American Hospital Association's (AHA) Annual Survey and the Centers for Medicare and Medicaid Services' (CMS) Healthcare Cost Report Information System (HCRIS). Hospital ownership and changes in ownership are taken from the AHA data and confirmed with background research. The HCRIS and discharge data are used to construct the hospital price estimates as described below. For each merger, the screening tools are constructed using data from the calendar year before the year in which the merger was consummated. If the merging hospitals are located near a state border or are located in different states, we use the discharge data of both states to construct the screening tools. Otherwise, only the discharge data from the merging hospitals' state are used.

To focus on mergers of competing hospitals, we include all of the mergers captured in our discharge data between short-term GAC hospitals in the same metropolitan statistical area (MSA) or adjacent MSAs, as long as we have at least one year of pre-merger discharge and price data and at least one year of post-merger discharge and price data.¹¹ We exclude acquisitions of Critical Access Hospitals (CAH) and acquisitions of failing or failed hospitals. The former are excluded because CAHs are small hospitals serving isolated rural areas and, thus, usually do not compete

¹⁰ The HHI based on market shares calculated using the bed capacity of the hospitals in the HRR implicitly assumes a product market of all hospital services.

¹¹ For rural mergers, we include all mergers involving hospitals in the same county or adjacent counties. Mergers involving hospitals close to a border of a state for which we do not have discharge data are excluded.

with other hospitals.¹² The latter are excluded because merger screens are irrelevant if one of the parties involved in the merger would exit the market absent the merger.¹³ Finally, we also exclude cases in which a hospital system acquires multiple hospitals at the same time. The selection criteria result in a sample of 28 mergers: 16 of which occurred between 2007 and 2012 between hospitals in Arkansas, Connecticut, Georgia, New York, Oklahoma, and Pennsylvania, and 12 of which occurred between hospitals in Missouri and North Carolina between 1997 and 2001. The mergers included in our sample are listed in Appendix B, in alphabetical order of the acquired hospital.

The ideal sample to assess the accuracy of a merger screen would be a random selection of mergers that are as likely to trigger the screen as not. However, mergers occurring in a period of active antitrust enforcement are more likely to be those that have or would have passed through the screen. In an era of antitrust enforcement in which the screen is actively used, mergers that the screen would identify as anticompetitive are less likely to occur because they are blocked or deterred. Thus, sampling mergers from such an era will hamper an analyst's ability to fairly assess the accuracy of a screen in predicting post-merger effects, particularly for mergers that the screen identifies as anticompetitive (Carlton, 2009). The Missouri and North Carolina mergers were added to the sample to ameliorate the bias caused by merger selection during a period of antitrust enforcement. Hospitals merging in North Carolina in 1997–2001 did not possess blanket antitrust immunity. However, North Carolina hospitals merging in this period likely felt relatively safe from antitrust challenges for two reasons. First, this period was at the tail end of the federal and state hospital antitrust losing streak and before the successful hospital merger challenges of recent years. Second, North Carolina introduced a hospital Certificate of Public Advantage (COPA) regulatory program in 1995, that gave merging hospitals participating in the program antitrust immunity conditional on submitting to state regulation. None of the hospitals in our sample participated in North Carolina's COPA program, but the option to participate, if the merger were challenged by federal or state antitrust authorities, may have reduced the likelihood of an antitrust challenge for these mergers. The Missouri mergers in the sample include a merger (Tenet's acquisition of Doctors Regional Medical Center in Poplar Bluff, Missouri) that was challenged by the FTC, but allowed to proceed by the courts. The sample of recent mergers also includes a merger (Phoebe Putney Health System's acquisition of Palmyra Medical Center in Albany, Georgia) that was challenged by the FTC, but allowed to proceed by the courts.¹⁴

□ **Price measurement.** Hospital prices are difficult to measure due to the variety and complexity of services offered. A typical short-term, acute care hospital offers services that support the treatment of patients across a broad range of diagnoses, exhibiting a broad range of severity. The price charged to any particular patient and his or her insurance company can be a function of many factors that affect the cost of treating the patient: the patient's diagnosis, the severity of the diagnosis, the procedures performed in treating the patient, the manner in which the patient was admitted (e.g., through the emergency room), additional conditions present in the patient when admitted to the hospital (i.e., comorbidities), complications that arise during treatment, etc. To accurately measure the overall hospital price paid by patients with private commercial insurance, one not only needs to accurately measure the payments made by the insurer and patient to the hospital, it is also necessary to properly adjust these payments to account for changes

¹² By law, CAHs can have no more than 25 acute care beds and must be at least 35 miles from the next nearest hospital (except in areas of mountainous terrain or other unique circumstances where the hospital is deemed a "necessary provider," despite proximity to another hospital of less than 35 miles).

¹³ An earlier draft of this article included two acquisitions of failing or closed hospitals. Their inclusion would not materially affect the results.

¹⁴ There is some evidence that prices in the years following the first post-merger year may have been constrained by regulation or in response to the legal appeals in this case (Garmon and Kmitch, 2017). Therefore, we exclude the years following the first post-merger year when estimating the price increase for this merger. Also, excluding the Phoebe/Palmyra merger from the sample of mergers does not materially affect the results.

in the diagnoses treated and procedures performed, along with changes in severity, complications, and comorbidities. These latter adjustments are often collectively referred to as “case-mix adjustment.”

The ideal data for hospital price measurement are comprehensive claims data that provide detailed information on each patient and treatment episode and the amounts actually paid by (not just the list price charged to) the patient and insurance company for each treatment and procedure. The discharge data sets described above provide detailed information about each patient stay, but they include only the total list price for all services, not the amount actually paid. Unfortunately, few states collect all-payer claims data in a form available to researchers and have a panel of hospital claims data stretching back far enough to capture pre- and post-merger periods. We are aware of no mergers of competing hospitals in the few states with all-payer claims data during the time period of their claims data collection. Private collections of claims data available for research either fail to provide a comprehensive collection of commercial insurers or do not identify individual hospitals.

Alternatively, some states collect aggregate hospital financial data and make it available to researchers. A few states collect aggregate financial data in sufficient detail to allow estimates of each hospital’s commercial price when the data are combined with discharge data. For instance, an estimate of each hospital’s average inpatient commercial discount can be calculated from the financial data and applied to the hospital’s commercially insured inpatients listed in the discharge data to estimate the hospital’s case-mix-adjusted commercial inpatient price. Numerous researchers studying hospital competition have used this approach, and Levit, Friedman, and Wong (2013) find that commercial prices calculated with state-level financial data are accurate estimates of commercial prices calculated from private claims data. Unfortunately, the states that collect and disseminate hospital financial data with detail sufficient to accurately estimate commercial inpatient prices are too few to allow the study of more than a handful of hospital mergers.

To estimate hospital prices for a relatively large sample of hospital mergers spread across multiple states and across time, we use financial information in the HCRIS data and the commercial price estimation procedure described in Dafny (2009). Dafny (2009) estimates the case-mix-adjusted commercial price for each hospital using estimates of net inpatient commercial revenue and commercial inpatient discharges derived from HCRIS data and each hospital’s case-mix index taken from CMS’s Impact Files. Each hospital’s estimated price is:

$$P_h^d = \frac{(IPSC_h + IPIC_h + IPANC_h) \left(1 - \frac{CONTDISC_h}{GROSSREV_h}\right) - MCPRIM_h - MCAP_h}{(DISCH_h - MDISCH_h)CMI_h}, \quad (12)$$

where $IPSC_h$ is the hospital’s inpatient routine service charges, $IPIC_h$ is intensive care charges, $IPANC_h$ is inpatient ancillary charges, $CONTDISC_h$ is contractual discounts, $GROSSREV_h$ is gross revenues, $MCPRIM_h$ is the hospital’s Medicare primary payer amounts, $MCAP_h$ is the Medicare total amount payable,¹⁵ $DISCH_h$ is the hospital’s total inpatient discharges, $MDISCH_h$ is Medicare inpatient discharges, and CMI_h is the hospital’s case-mix index (i.e., the average Diagnosis Related Group [DRG] weight for its inpatients). The only change we make to the Dafny (2009) formula in (12) is to substitute the hospital’s case-mix index for commercial inpatients calculated from the discharge data for the Impact File case-mix index, which reflects the hospital’s Medicare population.

The price estimate in (12) is not an ideal proxy for each hospital’s commercial price, as it does not deduct Medicaid revenue, and the discount factor applied to inpatient charges reflects inpatient and outpatient discounts. However, under certain circumstances, the price estimate in (12) may provide unbiased estimates when used to measure price changes over time.¹⁶ Suppose hospital h ’s true commercial price at time t is given by P_{ht} and the price measure described in

¹⁵ $MCPRIM+MCAP$ is the total reimbursement to the hospital for Medicare inpatients.

¹⁶ Lewis and Pflum (forthcoming) use a similar approach to study post-merger hospital price changes.

(12) for hospital h at time t is given by P_{ht}^d . We can express the relationship between the two as follows:

$$P_{ht} = \alpha_t P_{ht}^d + \varepsilon_{ht}, \quad (13)$$

where ε_{ht} has mean zero. If α_t is constant over time (i.e., $\alpha_t = \alpha$ for all t), then the distribution of price change estimates calculated using (12) should approximate the distribution of real price changes. However, if α_t varies over time, price changes estimated using (12) will be biased.

As the primary source of potential bias in (12) is the inclusion of Medicaid revenue, and Medicaid provider reimbursement policy is primarily determined at the state level, it is reasonable to assume that α_t is constant within each state at each time t (i.e., $\alpha_t = \alpha_{ts}$ for state s).¹⁷ Therefore, we can aggregate (13) across the hospitals within each metropolitan statistical area (MSA) M (as long as the MSA does not cross a state boundary) to produce:

$$P_{Mt} = \sum_{h \in M} w_{ht}^M P_{ht} = \alpha_{ts} \sum_{h \in M} w_{ht}^M P_{ht}^d + \sum_{h \in M} w_{ht}^M \varepsilon_{ht}, \quad (14)$$

where P_{Mt} is the average commercial hospital price in M at time t and w_{ht}^M are weights reflecting hospital h 's commercial share in M at time t .

To evaluate the potential bias associated with the use of (12) to measure price changes, we estimate (14) for each state and year using an independent data source to construct P_{Mt} . To construct the commercial hospital price in each MSA in each time period, we use data from Truven Health's MarketScan Commercial Claims and Encounters (CCE) database. The CCE database is a large nationwide claims-level data set describing treatment episodes for commercially insured patients with employer-sponsored health plans. It is an almost ideal data source for calculating prices, because it has the actual amount paid to the hospital for each patient along with full diagnosis and procedure information to facilitate case-mix-adjustment, and it includes most major health insurers. Individual hospitals are not identified in the CCE database, so it cannot be used on its own to measure hospital price changes. Using the CCE data, we estimate (14) separately for each state and year (excluding MSAs that cross state boundaries), recover $\hat{\alpha}_{ts}$, and apply these estimated coefficients to the price estimates in (12) to produce an estimate of each hospital's commercial inpatient price, P_{ht} . We then use these "adjusted" prices to estimate the post-merger price change for each merger using the synthetic control estimation method described in the next subsection. These "adjusted" price changes are compared to the unadjusted price changes to evaluate the bias associated with using (12) as a price change measure. This adjustment procedure is possible for only 15 mergers in our sample, as we have CCE data only for 2007 through 2014. As seen in Figure 1, the adjusted and unadjusted price changes are almost identical for most mergers, suggesting that unbiased estimates of post-merger price changes can be obtained with the price measure described in (12).¹⁸

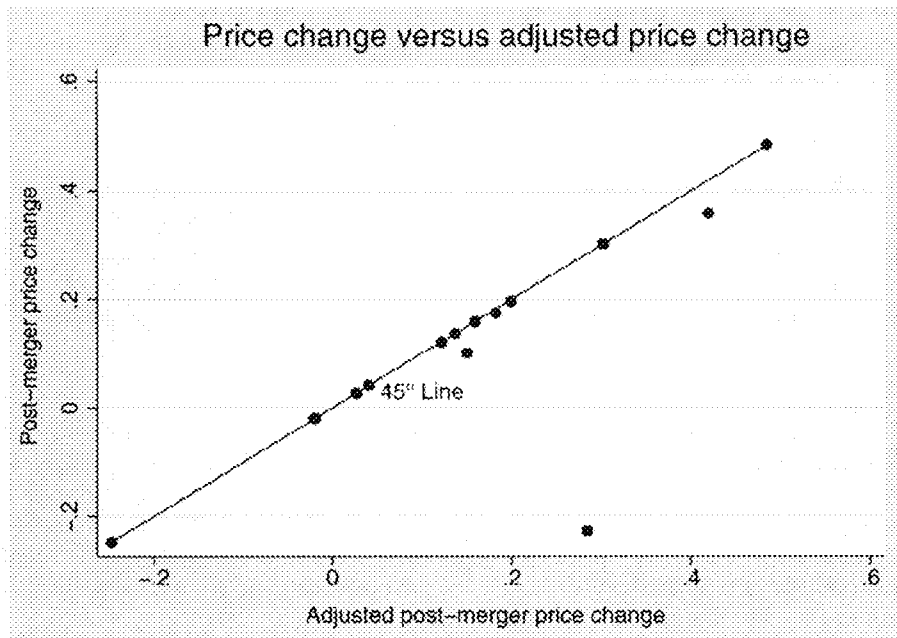
□ **Post-merger price and cost change estimation.** The merger screens described above are meant to capture the loss of competition associated with a hospital merger. To assess the accuracy of their predictions, the screens should be compared to the price change associated with the loss of competition from the merger, apart from other changes that occur coincident with the merger. In other words, the screens should be compared to the difference between the post-merger price

¹⁷ The use of a combined inpatient and outpatient discount factor in (12) is also a potential source of bias. However, hospitals and health insurers usually negotiate inpatient and outpatient prices together, so a change of bargaining power associated with a merger could manifest in either inpatient or outpatient prices.

¹⁸ Using financial data from PHC4 for Pennsylvania short-term, GAC hospitals, we also calculated case-mix-adjusted commercial inpatient prices and regressed these estimates onto commercial inpatient prices calculating using (12), while suppressing the constant and restricting the sample to hospitals with at least 200 commercial discharges per year. The resulting estimated coefficient was 0.99 with an R_2 of 0.90. Furthermore, we cannot reject the null hypothesis that the coefficient is constant across years. This provides additional support to the use of (12) as the basis for unbiased estimates of post-merger price changes.

FIGURE 1

PRICE CHANGE VERSUS ADJUSTED PRICE CHANGE [Color figure can be viewed at wileyonlinelibrary.com]



change and what it would have been absent the merger. Apart from the price measurement issues described above, the former price difference is straightforward to calculate. The latter price difference is impossible to calculate as the merger did, in fact, occur. Thus, following the general difference-in-differences (DID) literature, we select control groups of hospitals to serve as a proxy for the merging hospitals in the counterfactual, absent the merger. These control groups should be made up of nonmerging hospitals that are similar to the merging hospitals.

To select nonmerging hospitals that are similar to the merging hospitals, we use the synthetic control method of Abadie, Diamond, and Hainmueller (2010), in which a “synthetic” control hospital is constructed as a weighted average of the controls so that the synthetic control is similar to the merging hospitals in the pre-merger period with regard to price and predictors of price. The weights are constructed (according to the algorithm described in Abadie, Diamond, and Hainmueller, 2010) by matching potential controls to the merging hospitals based on pre-merger prices and the following predictors of price: operating cost per adjusted admission (a measure of average variable cost), residents and interns per bed (a measure of teaching intensity), and occupancy rate (a measure of excess capacity).¹⁹

Excluded from the potential controls are nonmerging hospitals in the same MSA (or county, for rural mergers) as the merging hospitals. A merger may affect nearby nonmerging hospitals by lowering the overall level of competition in the market (Dafny, 2009). Thus, the use of nearby hospitals in the control group, which otherwise may be optimal because they face cost and demand conditions similar to the merging hospitals, may bias the estimated price change. In addition, the controls exclude hospitals specializing in the treatment of children, hospitals with fewer than 200 commercial admissions in any pre- or post-merger year, and Critical Access Hospitals. Finally,

¹⁹ Following the recommendation of Kaul et al. (2017), we do not use all of the pre-merger prices to match the merging hospitals with the synthetic control. Instead, we use the prices from alternating pre-merger years (e.g., all of the odd-numbered pre-merger years, if the merger occurred in an even-numbered year).

hospitals outside of the merged hospitals' MSA that are owned by the acquiring hospital system are excluded from the control group, and the merging hospitals' price change estimates. As a robustness check, we also estimate post-merger price changes using a standard DID estimator, in which the control group is matched to the merging hospitals based on bed size.²⁰

The post-merger price change for each merger is measured relative to the price change of the synthetic control. The relative price change for each merger is calculated by estimating the following equation using least squares estimation:

$$P_{ht} = \alpha + \beta_1 \text{POST}_{ht} + \beta_2 \text{POSTM}_{ht} + \delta_h + \varepsilon_{ht}, \quad (15)$$

where POST_{ht} is an indicator for the post-merger period, POSTM_{ht} is an indicator for the merged hospitals in the post-merger period, and δ_h is a synthetic control indicator. For the merging hospitals, P_{ht} is the log of the merging hospitals' weighted average commercial price. For the synthetic control, P_{ht} is the log of the weighted average commercial price for control hospitals with positive matching weights. The relative post-merger price change is calculated as:

$$P = e^{\hat{\beta}_2} - 1. \quad (16)$$

The statistical significance of the relative price change is evaluated using an inference procedure similar to that described in Abadie, Diamond, and Hainmueller (2010), in which the relative price change is compared to the distribution of "placebo" effects. The price change relative to a synthetic control is calculated for each hospital in the potential controls as if it were the merging hospitals. The estimated price change for the actual merging hospitals is then compared to the distribution of estimated price changes for the placebos to determine the probability that the relative post-merger price change is due to chance.

Using similar methods, we also measure the post-merger change in average variable costs relative to controls to assess whether the merger resulted in any variable cost efficiencies. From the HCRIS data, we calculate each hospital's annual operating cost per adjusted admission, where the adjustment captures the hospital's combined inpatient and outpatient volume. As with the price changes, the controls are selected using the synthetic weighting procedure of Abadie, Diamond, and Hainmueller (2010), where control hospitals are matched to the merging hospitals based on pre-merger average variable costs, residents and interns per bed, and occupancy rate. Unlike the post-merger price change estimation, local nonmerging hospitals are included as potential controls. Otherwise, exclusions similar to those applied in the price change estimation are applied before the control matching procedure.

□ **Choice model estimation.** The hospital merger screening tools based on first-order approximations (e.g., WTP, UPP) are constructed from the predicted probabilities of a conditional logit choice model. Recent research on the accuracy of conditional logit model predictions in the context of hospital choice indicates that there is no one modelling approach that is superior in all cases²¹ (Raval, Rosenbaum, and Wilson, 2016). Therefore, we use the estimates from three different choice model specifications to construct the new screening tools. The first model is

²⁰ For most of the mergers that involve urban hospitals, the alternate controls are urban, short-term GAC hospitals in the same state with more than 100 staffed beds. For mergers that involve rural hospitals with more than 100 staffed beds, the control group includes all nonmerging urban and rural short-term GAC hospitals in the same state with more than 100 staffed beds. In cases in which the merger involves the acquisition of a hospital with fewer than 100 staffed beds, the control group is selected with a smaller bed-size threshold, and these thresholds are listed in Appendix B. The previously described exclusions also apply to this alternate control group.

²¹ Raval, Rosenbaum, and Wilson (2016) conclude that "... different types of models provide more accurate predictions for different types of patients. These findings suggest that combining the predictions from multiple models may lead to better predictions of behavior than using a single 'preferred model.' "

a parametric choice model in which the patient's choice is modelled as a function of hospital characteristics and patient characteristics. The probability that patient i selects hospital h is:

$$s_i^h = \frac{e^{(\gamma Z_{ih} + \beta X_i Y_h + \theta X_i Z_{ih})}}{\sum_{j \in H} e^{(\gamma Z_{ij} + \beta X_i Y_j + \theta X_i Z_{ij})}}, \quad (17)$$

where Z_{ih} are characteristics specific to patient i and hospital h , X_i is a vector of patient characteristics, and Y_h is a vector of hospital characteristics. In other words, patient i 's hospital choice is assumed to be a function of hospital characteristics unique to the patient and characteristics common to all patients. Further, patient preferences for these characteristics are allowed to vary across patient types. The patient-specific hospital characteristics Z_{ih} consist of the driving time (under normal traffic conditions) between the center of the patient's zip code and the hospital and the driving time squared.²² The patient characteristics consist of the patient's DRG weight, a gender indicator, an indicator for emergency room admissions, an indicator for obstetrics, and an indicator for cardiac surgery. The hospital characteristics consist of the hospital's residents and interns per bed (a measure of teaching intensity), a for-profit indicator, an indicator for hospitals that offer obstetrics services, and an indicator for hospitals that offer cardiac surgery.

The second model is similar to the first, except that the hospital characteristics Y_h in (17) are replaced by hospital fixed effects.²³ The third model is the semiparametric choice model described in Raval, Rosenbaum, and Tenn (2017), in which patient bins are defined iteratively and probabilities are estimated using the observed shares within each bin. As in Raval, Rosenbaum, and Tenn (2017), the bins are defined (with a minimum bin size of 25 patients) using the following patient characteristics, in declining order of importance: patient's county, patient's zip code, major diagnostic category (MDC), whether the patient was admitted through the emergency room, whether the service provided was medical or surgical, the severity of the patient's diagnosis (using DRG-weight quartiles), DRG, age category, and gender. These three specifications cover most of the modelling approaches used in the hospital competition literature and past hospital merger challenges.

All three models are estimated over all of the commercial GAC inpatients in the acquired hospital's HRR in the year prior to the merger. The choice set H is restricted to all hospitals that served at least 0.5% of these patients.²⁴ The choice of a hospital outside of this set is aggregated into an outside option.²⁵ For each model and each merger screen calculated directly from the predicted probabilities (i.e., diversion ratios and WTP change), standard errors are calculated via bootstrap methods. UPP is constructed as the weighted average of the merging hospitals' UPP, where each hospital's UPP is the product of the diversion ratio to its merger partner and the partner's pre-merger absolute margin (calculated as the difference between its pre-merger price and operating cost per adjusted admission) divided by its own pre-merger price.

□ **Merger simulation.** The basis for the merger simulation, condition (10), is estimated via a system-level regression of price on WTP_PAD, average variable cost, and other covariates. The dependent variable is the weighted average commercial price (as calculated in (12)) across all of the hospitals in each system in each metropolitan statistical area (MSA) in the year prior to the merger (where the weights are based on the number of commercial discharges). WTP_PAD is constructed by first estimating the semiparametric choice model across all of the patients in each MSA in the state in the year prior to the merger. Each system's WTP is then calculated

²² Driving times are taken from ArcGIS, version 10.

²³ Unlike the first parametric model, the fixed effects are included on their own and interacted with indicators for the patient's major diagnostic category.

²⁴ For two mergers, the HRR is so large that one of the merging hospitals is not included in the choice set using this inclusion criterion. In these cases, the choice model is estimated over all of the commercial GAC patients in the merging hospitals' combined primary service area instead of the HRR.

²⁵ The coefficient estimates, standard errors, and fit statistics for the estimation of (17) for both parametric choice models for each merger are available upon request.

from the predicted probabilities and divided by the system's aggregate DRG weight to produce the system's WTP_PAD. The system-wide operating cost per adjusted admission (where the adjustment accounts for the hospital's outpatient scale) is used as the proxy for average variable cost. Finally, the other covariates consist of a for-profit indicator (to capture differences in for-profit and nonprofit pricing) and MSA indicators (to capture differences in market conditions across MSAs).

The estimated coefficient of WTP_PAD is then used to predict the post-merger price change by applying it to the predicted change in WTP_PAD associated with the merger.²⁶ Because the estimation of (10) is carried out only for urban hospital systems in each state, a merger simulation is not conducted for the rural hospital mergers.

5. Results

■ *Ex ante* merger screens designed to identify mergers that are likely anticompetitive can be used in one of two ways. First, they can be used to establish a threshold above which the merger is presumed to be anticompetitive, absent extenuating circumstances (e.g., mergers with a change in the screen above X% are likely anticompetitive). We will refer to this as the "threshold" approach. Second, they can be used to make a prediction about the likely effects of a merger (e.g., an X% increase in the screen is associated with a Y% increase in price on average). We will refer to this as the "relationship" approach. The first approach largely differs from the second by making no presumption about mergers below the threshold. We will evaluate the merger screens with both uses in mind. First, are there thresholds of the merger screen above which price increases are likely and price decreases are unlikely? Second, is there a relationship between the merger screen and the post-merger price change?

The distribution of post-merger price changes relative to controls is shown in Figure 2.²⁷ Nine of the 28 mergers resulted in statistically significant price increases relative to controls, whereas 6 resulted in statistically significant relative price decreases (i.e., an absolute price decrease or an increase less than the mean increase across the controls).²⁸ The latter mergers highlight the fact that not all mergers of competing hospitals are anticompetitive, and some may lead to lower prices (or smaller price increases than normal) due to, for instance, cost savings resulting from the merger. The mean price change relative to controls across all 28 mergers is 8.9% and the median is 9.6%.

The primary objective of evaluating merger screens is to determine their accuracy in identifying presumptively anticompetitive mergers, absent extenuating circumstances, such as merger-specific cost savings. Evaluating screens with consummated mergers in a period of active antitrust enforcement introduces bias because mergers that significantly eliminate competition without any ameliorating forces are likely to be blocked or never proposed, whereas those associated with a minimal reduction in competition or those with significant ameliorating factors (e.g., merger-specific cost savings) are more likely to proceed. As mentioned previously, we partially address this bias by including mergers from Missouri and North Carolina from the late 1990s and early 2000s. However, we can also address this bias by isolating the mergers that may have induced merger-specific cost reductions that ameliorate competition reductions. Of the six mergers that resulted in statistically significant relative price decreases, three resulted in statistically significant

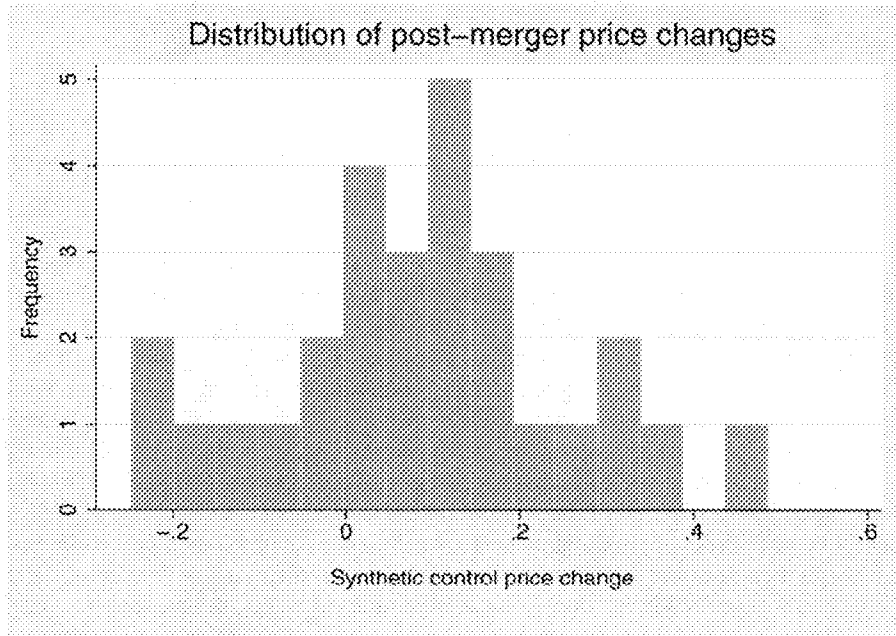
²⁶ The bootstrapped 95% confidence intervals of the price change estimate for each merger are listed in Appendix C, along with each estimate.

²⁷ Appendix C (available at www.researchgate.net/publication/281494031_The_Accuracy_of_Hospital_Merger_Screening_Methods) lists the price change relative to controls for each merger, along with each merger screen calculated using pre-merger data.

²⁸ Statistical significance is measured at the 95% level. It is important to note that a lack of statistical significance does not necessarily imply economic insignificance. For instance, one merger was associated with a 13.7% relative price increase, potentially indicating an anticompetitive merger, but the null hypothesis that this merger's price increase is the same as the mean control price increase cannot be rejected because of the high variance of control price changes.

FIGURE 2

DISTRIBUTION OF POST-MERGER PRICE CHANGES [Color figure can be viewed at wileyonlinelibrary.com]



reductions in average variable cost relative to controls and the mean cost reduction across all six mergers was 4.2%.

Overall, seven mergers resulted in statistically significant reductions in average variable cost, and eight resulted in statistically significant cost increases.²⁹ For the latter, the mean cost increase was 26.4%. For the former, the mean cost decrease was 22.5% and the mean price change was 0.5%. The mergers with significant variable cost reductions do not seem to have any distinguishing characteristics that set them apart from other mergers. Of the seven with significant variable cost reductions, some involve a national for-profit chain acquiring an independent hospital, others involve a nonprofit system acquiring a local independent competitor, and still others involve the combination of proximate independent hospitals.

Whether the variable cost reductions resulted in price reductions or not, the inclusion of mergers with cost efficiencies may bias the evaluation of the screens by introducing mergers with post-merger price changes not fully reflective of the loss of competition. For the analysis that follows, in addition to presenting results for the full sample of mergers, we also present results after excluding mergers with a statistically significant price decrease, mergers with a statistically significant variable cost decrease, and mergers with a cost decrease exceeding 5%, whether statistically significant or not, to minimize the bias associated with antitrust enforcement. For this reason, it is important to emphasize that the screens are being evaluated to determine how well they identify mergers that warrant further investigation, *not* how well they identify anticompetitive mergers in isolation from other evidence.

The screens that are based on first-order approximations of the loss of competition from the merger (i.e., diversion ratios, WTP, and UPP) are all closely related and highly correlated.³⁰ UPP

²⁹ These results are broadly consistent with Schmitt (2017), who finds little evidence of cost savings on average from mergers of geographically proximate hospitals.

³⁰ The correlation between each screen is at least 0.75, regardless of the choice model.

TABLE 2 Correlation between Price Changes and Combinations of Merger Screens

Combining rule	Entire Sample		Excluding Mergers with Variable Cost Savings	
	WTP	UPP	WTP	UPP
Minimum	0.15	0.21	0.47	0.40
Maximum	-0.04	0.04	0.34	0.32
Mean	0.02	0.11	0.40	0.36
Convex combination ^a	0.03	0.12	0.43	0.37

^aUsing weights from Raval, Rosenbaum, and Wilson (2016): Semiparametric (0.5), Parametric w/Hospital Characteristics (0.26), Parametric w/Hospital Fixed Effects (0.24).

is a function of the pre-merger margins and diversion ratios, and the post-merger change in WTP is also determined by the diversion ratios. We will focus on the performance of UPP and WTP for two reasons. First, these screens were the primary tools employed to measure lost competition in the most recent hospital merger challenges. Second, UPP and WTP capture different post-merger negotiating postures. The change in WTP captures the post-merger change in the health insurer's threat point if the merged hospitals bargain in an all-or-nothing manner, whereas the UPP price change approximation captures the post-merger change in the hospitals' threat points if the merged hospitals continue to bargain separately. Although the discussion below focuses exclusively on UPP and WTP, all of the screens are listed in Appendix C.

Each of the first-order approximation screens is based on the predicted probabilities from a model of hospital choice, and Appendix C lists each screen calculated with each of the choice models previously described in Section 4: the semiparametric model, the parametric model with hospital characteristics, and the parametric model with hospital fixed effects. Raval, Rosenbaum, and Wilson (2016) advise against using any one model in isolation as the basis for predictions and instead suggest a convex combination of the three models with the semiparametric receiving the largest weight. Table 2 lists other possible combinations of UPP and WTP—including the minimum value, the maximum value, and the unweighted mean—along with the correlation between the post-merger price change and the combination of screens. For the entire sample and for mergers without variable cost savings, the minimum value (i.e., the most conservative estimate of the lost competition from the merger across the three models) is most correlated with the post-merger price changes. Therefore, the analysis that follows will focus on the minimum UPP and the minimum WTP across the three choice models, along with the traditional concentration measures and merger simulation.

Turning to the threshold approach of screening mergers, the Federal Trade Commission (FTC) and Department of Justice (DOJ) joint Horizontal Merger Guidelines³¹ state that mergers resulting in a post-merger HHI of 2500 or more with an HHI increase of 200 or more “will be presumed to be likely to enhance market power.” As described in the Guidelines, this presumption is not sufficient to conclude that a merger is likely to substantially lessen competition. However, an antitrust regulator using the HHI as a screen is likely to focus on mergers with a post-merger HHI greater than 2500 and a change greater than 200 for further investigation.

For the direct measures of lost competition (UPP and WTP), there are no established thresholds—like the Guidelines' HHI thresholds—above which a merger is presumed problematic. In the FTC's challenge of Promedica's acquisition of St. Luke's, a projected 13.5% change in WTP was cited by the FTC in its public decision.³² In the FTC's challenge of the proposed merger between OSF Healthcare and Rockford Health, a WTP change of 19% was cited by FTC staff in

³¹ www.ftc.gov/sites/default/files/attachments/merger-review/100819hmg.pdf (accessed on 3/27/2014).

³² www.ftc.gov/sites/default/files/documents/cases/2012/06/120625promedicaopinion.pdf, page 49 (accessed on September 30, 2014).

its pre-trial brief to the court.³³ However, a threshold lower than 13.5% may be more appropriate when using WTP to screen for mergers that warrant further investigation. Because the screens are used to select mergers for further investigation, the optimal threshold for a screen should be that which maximizes correct predictions (i.e., flagged mergers that resulted in a price increase or nonflagged mergers that did not result in a price increase) and minimizes false negatives (i.e., nonflagged mergers that resulted in a price increase). Therefore, we selected the threshold value for each screen that maximizes the difference between the number of correct predictions and the number of false negatives. For the minimum change in WTP, the optimal threshold value is 6%. For the minimum UPP, the optimal threshold is 4%.

Tables 3 and 4 compare the threshold performance of WTP and UPP (using the 6% and 4% thresholds, respectively) to the three HHI screens using the Guidelines' thresholds. Table 3 lists the number of correct predictions, false positives, and false negatives for each screen using the entire sample of mergers, whereas Table 4 lists those for mergers without variable cost savings. Of the three HHI measures, the HHI calculated using bed shares in the merging parties' HRR is the most likely to produce a false negative. Of the nine mergers with post-merger price increases, this screen flags only four. Even after excluding mergers with cost efficiencies, the HRR HHI fails to flag five mergers with post-merger price increases. This largely confirms the criticisms of most health economists that the EH-based geographic markets used by the courts in the 1990s (which closely resemble HRR markets) were, in many cases, too large to accurately predict the effect of a hospital merger.

On the other hand, the HHI calculated using discharge shares in the acquired party's HSA is most likely to produce a false positive. This screen flags all but one of the 28 mergers as potentially problematic, including all six of the mergers with post-merger relative price decreases. Even after excluding mergers with cost efficiencies, the HSA HHI flags 16 of the 17 remaining mergers as potentially problematic. When evaluating screens based on their ability to flag mergers for further investigation, false negatives are more problematic than false positives, so the HSA HHI is superior to the HRR HHI in this regard. However, a screen that flags virtually every merger as potentially problematic is not a screen in any practical sense.

The three remaining screens (the HHI calculated using discharge shares in the merging parties' WSA, the minimum change in WTP, and the minimum UPP) perform better than the HRR and HSA HHIs at flagging potentially problematic mergers, but WTP and UPP perform best. Of the five screens, WTP makes the correct prediction (i.e., flags mergers associated with statistically significant price increases and does not flag mergers not associated with statistically significant price increases) most often. In addition, WTP has the lowest incidence of false positives (i.e., flagging a merger as potentially problematic when the merger is not associated with a statistically significant price increase) apart from the HRR HHI. Of the nine mergers with a statistically significant price increase, seven had a minimum WTP change greater than 6%. Of the six mergers with a statistically significant price decrease, three had a minimum WTP change of less than 6%.

UPP also performs well at flagging potentially problematic mergers. For the entire sample, UPP's rate of correct predictions is slightly less than WTP's and its rate of false positives is higher, but, unlike WTP, UPP correctly flagged all nine mergers with a statistically significant price increase. After excluding mergers with variable cost savings, UPP's rate of correct prediction matches that of WTP.

In contrast, the WSA HHI, although it performed better than the other HHIs, did not perform as well as WTP and UPP. For instance, the WSA HHI had the second highest rate of false positives, behind only the HSA HHI, and failed to flag one merger with a statistically significant price increase. One might attribute the poor performance of the HHIs to the Guidelines' thresholds, but altering these thresholds does not improve the HHI's performance. For instance, lowering the WSA HHI threshold (from 2500 to 2000 post-merger with

³³ www.ftc.gov/sites/default/files/documents/cases/2012/04/120404ccpretrialbrief.pdf, page 44 (accessed on June 9, 2015).

TABLE 3 Selection Based on Thresholds: Entire Sample

	Flagged Mergers	Correct Positive ^a	Correct Negative ^b	False Positive ^c	Strong False Positive ^d	False Negative ^e	Mean Relative Price Change for Flagged Mergers	Informedness (DeltaP')	Markedness (DeltaP)	Matthews Correlation Coefficient
HHI (HRR bed shares) Guidelines ^f	9	4	14	5	4	5	6.9%	0.18	0.18	0.18
HHI (HSA discharge shares) Guidelines ^f	27	9	1	18	6	0	8.9%	0.05	0.33	0.13
HHI (WSA discharge shares) Guidelines ^f	21	8	6	13	5	1	10.0%	0.20	0.23	0.22
HHI (WSA discharge shares) <i>Post</i> > 2000, <i>Delta</i> > 100	24	8	3	16	5	1	9.4%	0.05	0.08	0.06
HHI (WSA discharge shares) <i>Post</i> > 3000, <i>Delta</i> > 250	18	7	8	11	5	2	9.9%	0.20	0.19	0.19
Minimum change in <i>WTP</i> > 6%	14	7	12	7	3	2	12.9%	0.41	0.36	0.38
Minimum <i>UPP</i> > 4%	20	9	8	11	4	0	12.2%	0.42	0.45	0.44
Minimum change in <i>WTP</i> > 6% Or WSA HHI Guidelines ^f	22	8	5	14	5	1	9.8%	0.15	0.20	0.17

^aFlagged merger as problematic and merger associated with statistically significant relative price increase.

^bDid not flag merger as problematic and merger not associated with statistically significant relative price increase.

^cFlagged merger as problematic and merger not associated with statistically significant relative price increase.

^dFlagged merger as problematic and merger associated with statistically significant relative price decrease.

^eDid not flag merger as problematic and merger associated with statistically significant relative price increase.

^fHorizontal Merger Guidelines thresholds = Post-Merger *HHI* > 2500 and *HHI Delta* > 200.

TABLE 4 Selection Based on Thresholds: Excluding Mergers with Variable Cost Savings

	Flagged Mergers	Correct Positive ^a	Correct Negative ^b	False Positive ^c	False Negative ^d	Mean Relative Price Change for Flagged Mergers	Informedness (DeltaP')	Markedness (DeltaP)	Matthews Correlation Coefficient
HHI (HRR bed shares) Guidelines ^e	3	3	9	0	5	28.4%	0.38	0.64	0.49
HHI (HSA discharge shares) Guidelines ^e	16	8	1	8	0	16.2%	0.11	0.50	0.24
HHI (WSA discharge shares) Guidelines ^e	11	7	5	4	1	20.3%	0.43	0.47	0.45
HHI (WSA discharge shares) <i>Post > 2000, Delta > 100</i>	14	7	2	7	1	17.0%	0.10	0.17	0.13
HHI (WSA discharge shares) <i>Post > 3000, Delta > 250</i>	9	6	6	3	2	22.7%	0.42	0.42	0.42
Minimum change in <i>WTP > 6%</i>	8	6	7	2	2	23.3%	0.53	0.53	0.53
Minimum <i>UPP > 4%</i>	12	8	5	4	0	20.6%	0.56	0.67	0.61
Minimum change in <i>WTP > 6%</i> Or WSA HHI Guidelines ^e	12	7	4	5	1	18.9%	0.32	0.38	0.35

^aFlagged merger as problematic and merger associated with statistically significant relative price increase.

^bDid not flag merger as problematic and merger not associated with statistically significant relative price increase.

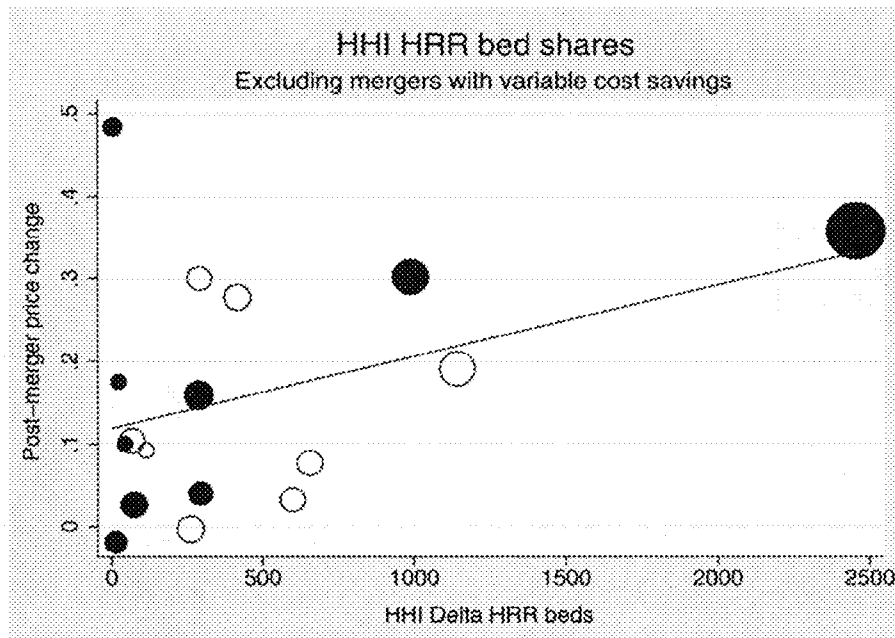
^cFlagged merger as problematic and merger not associated with statistically significant relative price increase.

^dDid not flag merger as problematic and merger associated with statistically significant relative price increase.

^eHorizontal Merger Guidelines thresholds = Post-Merger *HHI > 2500* and *HHI Delta > 200*.

FIGURE 3

POST-MERGER PRICE CHANGE AND HHI DELTA (BED SHARES IN THE HRR) [Color figure can be viewed at wileyonlinelibrary.com]



(Larger dots indicate larger post-merger HHI levels. Hollow dots are Missouri/North Carolina.)

a change of 100) increases the rate of false positives without any improvement in false negatives. Increasing the WSA HHI threshold to 3000 post-merger and a change of 250 improves the rate of correct predictions, but at a cost of one more false negative. In addition, combining the WSA HHI with one of the direct measures of lost competition (e.g., flag mergers in which the minimum WTP change is greater than 6% or the post-merger WSA HHI is greater than 2500 with a change of more than 200) does not improve the performance of either screen.

Finally, Tables 3 and 4 list measures of accuracy commonly used when evaluating binary diagnostic tests and machine learning algorithms (Powers, 2011). Markedness (also referred to as DeltaP) measures how often predictions (for both positive and negative outcomes) are correct.³⁴ Informedness (also referred to as DeltaP') measures how often the outcomes are correctly predicted by the test.³⁵ Matthews Correlation Coefficient is the geometric mean of Markedness and Informedness. All three measures are scaled to fall between negative one and one, with one representing perfect prediction and zero representing random prediction. Overall, using the threshold approach, the direct first-order measures of lost competition (WTP and UPP) more accurately flag hospital mergers for further investigation than any of the evaluated HHI screens using the traditional Guidelines or altered thresholds.

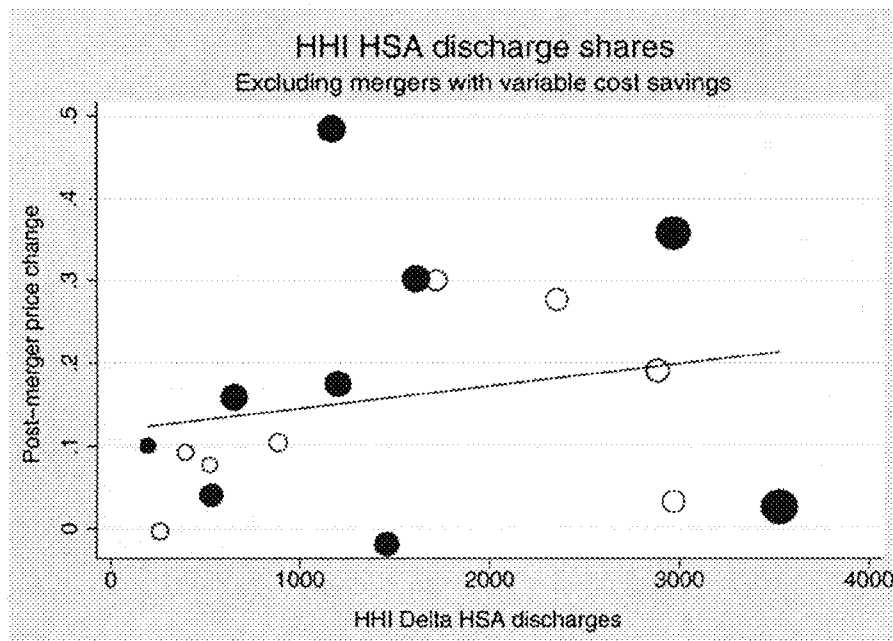
Figures 3 through 8 plot the post-merger price changes against each merger screening tool for the mergers without variable cost savings. For the HHIs, the price change is plotted against the HHI change (i.e., "HHI Delta") with larger dots signifying larger post-merger HHI levels. Hollow dots

³⁴ Specifically, Markedness equals Precision (i.e., the ratio of correct positives to all positive predictions) plus Inverse Precision (i.e., the ratio of correct negatives to all negative predictions) minus one.

³⁵ Specifically, Informedness equals Sensitivity (i.e., the ratio of correct positives to all positive outcomes) plus Inverse Sensitivity (i.e., the ratio of correct negatives to all negative outcomes) minus one.

FIGURE 4

POST-MERGER PRICE CHANGE AND HHI DELTA (DISCHARGE SHARES IN THE HSA) [Color figure can be viewed at wileyonlinelibrary.com]



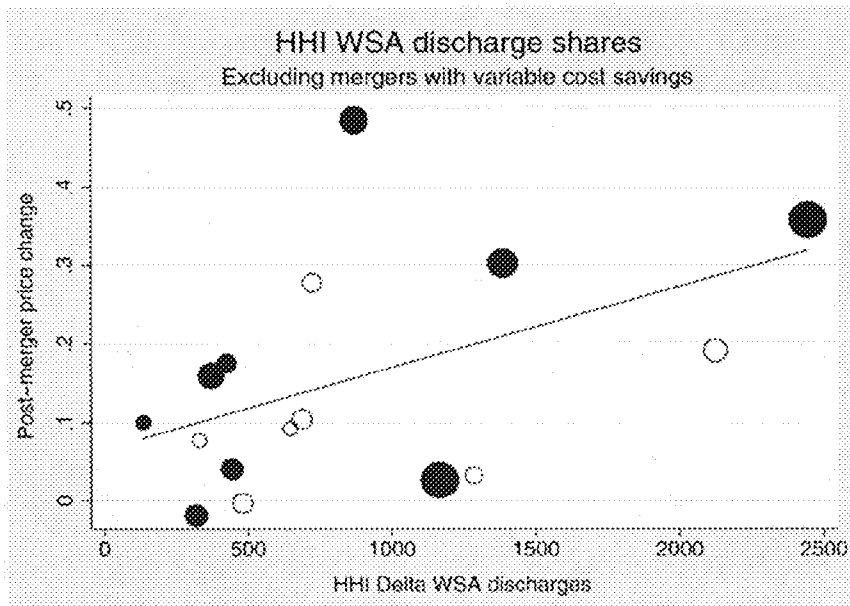
(Larger dots indicate larger post-merger HHI levels. Hollow dots are Missouri/North Carolina.)

represent the mergers in Missouri and North Carolina between 1997 and 2001, whereas solid dots represent the recent mergers. Table 5 lists the coefficient estimates and fit of the OLS regression of the price change on each merger screen. In each case, there is a great deal of unexplained variation in the price changes. This is not surprising, as the screens are meant to capture only the loss of competition resulting from the merger and not other changes coincident with the merger (e.g., management changes). As seen in the first two columns of Table 5, all of the screens perform poorly predicting price changes across the entire sample of mergers and for the earlier mergers in Missouri and North Carolina. When considering just the mergers without variable cost savings (column 3), the merger screens that measure lost competition through first-order approximations (particularly WTP) and the weighted service area HHI do a better job of predicting price changes than the other HHIs. For the minimum WTP change, the relationship between the merger screen and the post-merger price change is positive and statistically significant. The same is true of the relationship between the weighted service area HHI change and the post-merger price change. However, in both cases, the relationship is statistically significant at the lowest level of significance and the relationship between UPP (which is highly correlated with WTP) and the price change is almost, but not quite, statistically significant ($p = 0.11$).

As seen in Figures 6 and 7, WTP and UPP leave a great deal of unexplained variation in the price changes, even after excluding mergers with cost savings. In particular, there are two mergers with large price increases exceeding 20%, but relatively small changes in WTP and UPP. This highlights that methods that approximate the lost competition from a merger like WTP and UPP, although more accurate than most concentration measures, may miss some merger effects that impact price. Recent research (Dafny, Ho, and Lee, 2016; Lewis and Flum, 2017) suggests that certain factors unrelated to the patient overlap measured by WTP and UPP may affect the

FIGURE 5

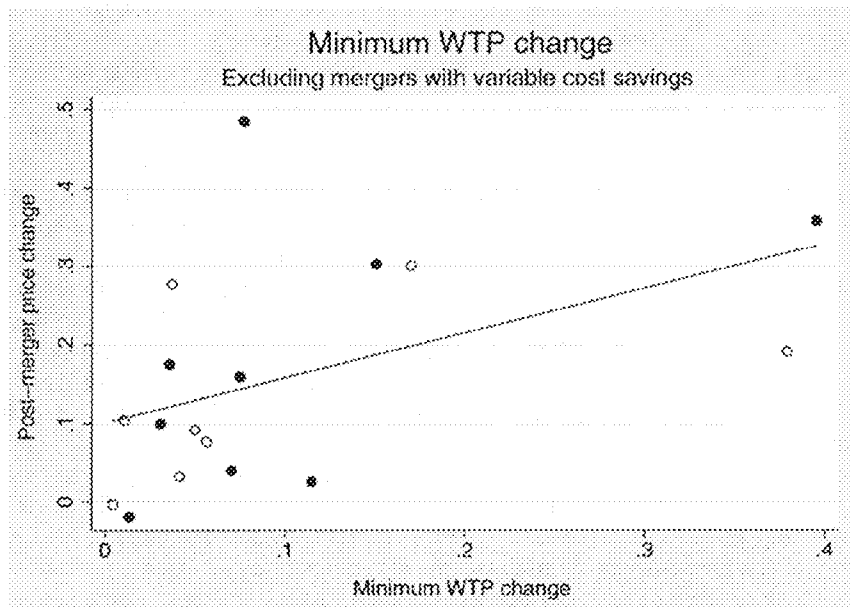
POST-MERGER PRICE CHANGE AND HHI DELTA (DISCHARGE SHARES IN THE WSA) [Color figure can be viewed at wileyonlinelibrary.com]



(Larger dots indicate larger post-merger HHI levels. Hollow dots are Missouri/North Carolina.)

FIGURE 6

POST-MERGER PRICE CHANGE AND MINIMUM WTP CHANGE [Color figure can be viewed at wileyonlinelibrary.com]

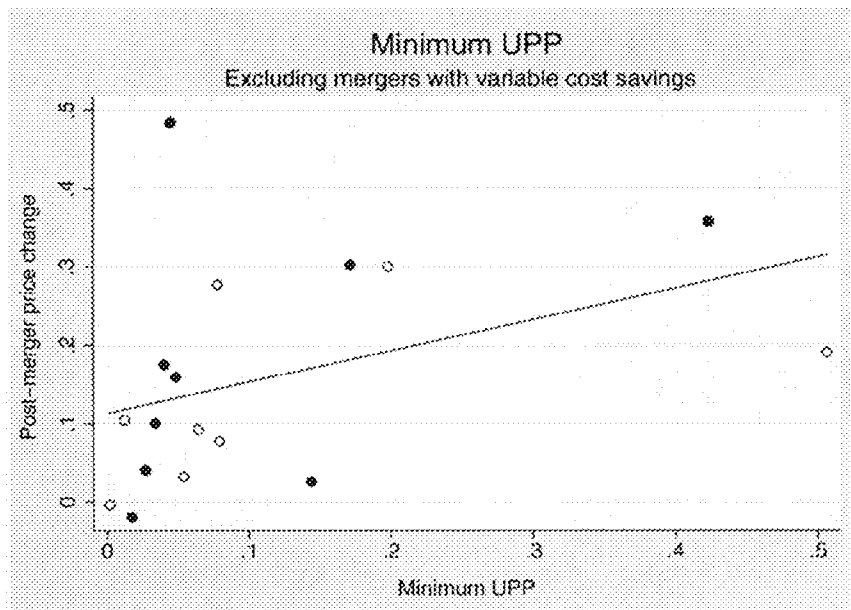


(Hollow dots are Missouri/North Carolina.)

© The RAND Corporation 2017.

FIGURE 7

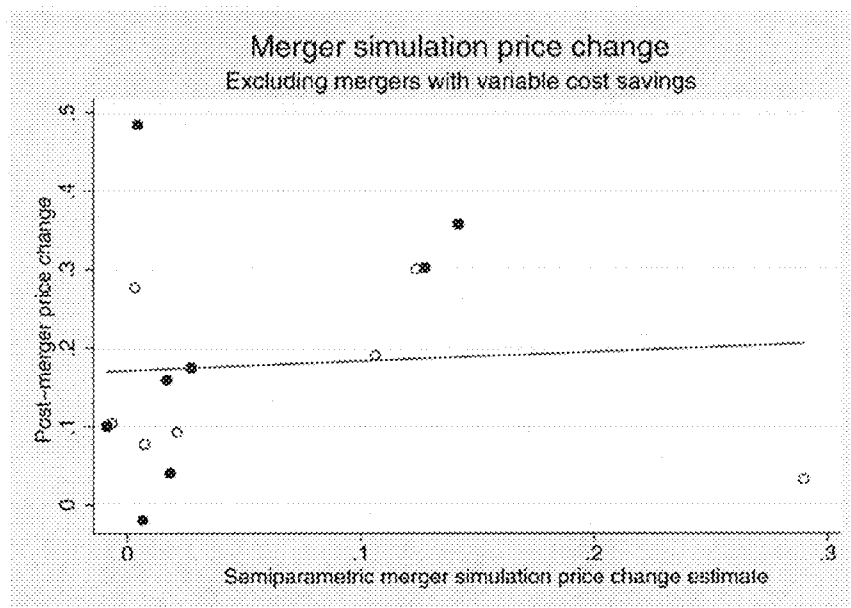
POST-MERGER PRICE CHANGE AND MINIMUM UPP [Color figure can be viewed at wileyonlinelibrary.com]



(Hollow dots are Missouri/North Carolina.)

FIGURE 8

POST-MERGER PRICE CHANGE AND WTP-BASED MERGER SIMULATION [Color figure can be viewed at wileyonlinelibrary.com]



(Hollow dots are Missouri/North Carolina.)

© The RAND Corporation 2017.

TABLE 5 Regressions of Price Change on Screening Tools

Screen		Entire Sample (N=28)		Missouri/North Carolina (N=12)		Excluding Mergers w/Variable Cost Savings (N=17)		Excluding Mergers w/Variable Cost Savings: Missouri/North Carolina (N=8)	
		Coefficient (SE)	R ²	Coefficient (SE)	R ²	Coefficient (SE)	R ²	Coefficient (SE)	R ²
Bed shares in HRR	Post-HHI	-5.2x10 ⁻⁵ (5.2x10 ⁻⁵)	0.06	-4.2x10 ⁻⁵ (5.8x10 ⁻⁵)	0.30	-4.7x10 ⁻⁶ (6.9x10 ⁻⁵)	0.14	3.5x10 ⁻⁵ (8.9x10 ⁻⁵)	0.04
	HHI-Delta	1.9x10 ⁻⁴ (1.6x10 ⁻⁴)		-3.2x10 ⁻⁵ (1.7x10 ⁻⁴)		1.0x10 ⁻⁴ (2.1x10 ⁻⁴)		-3.3x10 ⁻⁵ (2.3x10 ⁻⁴)	
Discharge shares in HSA	Post-HHI	2.0x10 ⁻⁶ (3.4x10 ⁻⁵)	0.01	2.1x10 ⁻⁵ (4.9x10 ⁻⁵)	0.05	3.0x10 ⁻⁵ (4.9x10 ⁻⁵)	0.07	1.3x10 ⁻⁴ (8.7x10 ⁻⁵)	0.44
	HHI Delta	-1.5x10 ⁻⁵ (6.0x10 ⁻⁵)		-5.0x10 ⁻⁵ (7.5x10 ⁻⁵)		-1.7x10 ⁻⁵ (7.9x10 ⁻⁵)		-1.4x10 ⁻⁴ (1.3x10 ⁻⁴)	
Discharge shares in WSA	Post-HHI	-2.7x10 ⁻⁵ (3.6x10 ⁻⁵)	0.02	-5.3x10 ⁻⁵ (4.1x10 ⁻⁵)	0.17	-3.1x10 ⁻⁵ (3.8x10 ⁻⁵)	0.26	-10.0x10 ⁻⁵ (5.6x10 ⁻⁵)	0.18
	HHI Delta	4.2x10 ⁻⁵ (7.6x10 ⁻⁵)		8.6x10 ⁻⁵ (7.9x10 ⁻⁵)		1.6x10 ⁻⁴ * (8.3x10 ⁻⁵)		6.4x10 ⁻⁵ (1.0x10 ⁻⁴)	
Minimum change in Willingness-to-pay (WTP)		0.23(0.29)	0.02	0.24(0.37)	0.04	0.57*(0.28)	0.22	0.40(0.32)	0.21
Minimum Upward Pricing Pressure (UPP)		0.26(0.24)	0.05	0.21(0.25)	0.06	0.40(0.23)	0.16	0.31(0.24)	0.21
WTP-based merger simulation ^a		-0.24(0.40)	0.02	-0.38(0.44)	0.10	0.12(0.46)	0.01	-0.20(0.42)	0.04

Each cell represents a separate regression. Standard error (SE) in parentheses. The coefficients are statistically significant at the 99% (***) , 95% (**), and 90% (*) levels as indicated.

^aFor the WTP-based merger simulation, the sample size is reduced by six to account for rural mergers. The sample size for the Missouri/North Carolina mergers is nine. The sample size for the mergers without variable cost savings is 15. The sample size for the Missouri/North Carolina mergers without variable cost savings is seven.

negotiation between hospitals and health insurers and may lead to merger effects that WTP and UPP fail to measure.

Merger simulation performs poorly, likely due to the limited data available to identify the relationship between WTP and price and the other limitations of merger simulation described above. With the data available in an initial investigation, one can measure only the commercial price of each hospital and not the price charged to each MCO, which would more closely fit the first-order condition in (3). With limited cross-sectional observations, the estimates of the relationship between price and WTP are imprecise. Although merger simulation may be a useful tool for analyzing a hospital merger with detailed payer data, this suggests that it is not a worthwhile exercise in the initial stages of an investigation when data are limited.

The estimated relationships between the price changes and the screens for the 1997–2001 Missouri and North Carolina mergers (the second and fourth columns in Table 5) are similar to the overall sample. The relationships between the weighted service area HHI, WTP, and the price changes have the same sign as the estimates for the overall sample, but are smaller and not statistically significant. When using the alternate difference-in-differences price change instead of the synthetic control price change, the results are similar, but the estimated relationships between the screens and the price changes are no longer statistically significant for the mergers without cost savings.³⁶

6. Conclusion

■ Recent research on hospital competition has produced new screening tools that attempt to capture the post-merger pricing incentives of hospitals better than the traditional techniques of concentration measurement. This article is the first large-scale evaluation of the new screening tools, comparing their predictions to the actual price effects of a relatively large sample of past consummated hospital mergers, using pre-merger data like that readily available in an initial investigation. The results suggest that screening tools based on the first-order pricing incentives of merged hospitals (in particular, WTP and UPP) are more accurate at flagging mergers that are potentially anticompetitive than the traditional tools of market definition and concentration measurement. However, the relationship between the new merger screens and post-merger price changes is not robust to all samples and price change measurements. In particular, the relationship holds only for mergers without variable cost savings, highlighting that the merger screens should not be used in isolation. WTP and UPP are useful for flagging mergers for further investigation, but not for identifying anticompetitive mergers on their own apart from other evidence. However, mergers without merger-specific variable cost savings and with large WTP changes are likely to produce significant price increases due to lost competition.

WTP-based merger simulation performs poorly at predicting post-merger price changes, but this may be due to the limited data available to calibrate the simulation in an initial investigation. Merger simulations may be more accurate when calibrated with detailed health insurance claims data. Finally, for the traditional exercise of market definition and concentration measurement, the weighted service area HHI is more accurate at flagging potentially anticompetitive mergers for further review and predicting post-merger price changes than HHIs calculated using other geographic market definitions. Going forward, better data and more detailed hospital competition models are needed to more accurately predict post-merger effects.

³⁶ These estimates are listed in Online Appendix D, which is available at www.researchgate.net/publication/281494031_The_Accuracy_of_Hospital_Merger_Screening_Methods.

Appendix A

This appendix provides examples of how to calculate the HHI using the three market definitions and share metrics described in the article.

Consider a hypothetical hospital market in which there are five hospitals, each located in one of five contiguous areas (e.g., zip codes). The hospitals, locations, and sizes are listed in Table A1.

In each area live 100 patients. Table A2 lists the patients by their hospital of treatment and their resident location.

Consider a proposed merger between hospitals A and B. Suppose As' and Bs' HRR is the combination of areas 1 through 4. Using the first HHI calculation method (bed shares in the HRR), only hospitals A, B, C, and D are in the market with shares of 20% (100/500) for A, 30% (150/500) for B, 10% (50/500) for C, and 40% (200/500) for D. The pre-merger HHI is $(20)^2 + (30)^2 + (10)^2 + (40)^2 = 3000$ with a delta of $2(Share_A)(Share_B) = 1200$ for a post-merger HHI of 4200.

Suppose Bs' HSA is the combination of areas 1 and 2. Using the second HHI calculation method (discharge shares in the HSA), A has a pre-merger share of $45/200 = 22.5\%$, B has a pre-merger share of $75/200 = 37.5\%$, C has a pre-merger share of $10/200 = 5\%$, D has a pre-merger share of $30/200 = 15\%$, and E has a pre-merger share of $40/200 = 20\%$. The pre-merger HHI is $(22.5)^2 + (37.5)^2 + (5)^2 + (15)^2 + (20)^2 = 2562.5$ with a delta of $2(Share_A)(Share_B) = 1687.5$ for a post-merger HHI of 4250.

For the third HHI calculation method (discharge shares in the weighted service area), we first have to determine the fraction of the merged entities' business that originates from each area. Of the 175 total patients treated by A and B, 65 (37.1%) come from Area 1, 55 (31.4%) come from Area 2, 25 (14.3%) come from Area 3, 20 (11.4%) come from Area 4, and 10 (5.7%) come from Area 5. As' weighted share is $(0.371)(40) + (0.314)(5) + (0.143)(10) + (0.114)(10) + (0.057)(5) = 19.3\%$. Bs' weighted share is $(0.371)(25) + (0.314)(50) + (0.143)(15) + (0.114)(10) + (0.057)(5) = 28.6\%$. Cs' weighted share is $(0.371)(5) + (0.314)(5) + (0.143)(30) + (0.114)(5) + (0.057)(5) = 8.6\%$. Ds' weighted share is $(0.371)(10) + (0.314)(20) + (0.143)(20) + (0.114)(60) + (0.057)(10) = 20.3\%$. Es' weighted share is $(0.371)(20) + (0.314)(20) + (0.143)(25) + (0.114)(15) + (0.057)(75) = 23.3\%$. The pre-merger HHI is $(19.3)^2 + (28.6)^2 + (8.6)^2 + (20.3)^2 + (23.3)^2 = 2215.5$ with a delta of $2(Share_A)(Share_B) = 1102$ for a post-merger HHI 3317.5.

TABLE A1 Hospitals, Areas, and Beds

Hospital	Area	Staffed Beds
A	1	100
B	2	150
C	3	50
D	4	200
E	5	250

TABLE A2 Patient Treatment Distribution

Hospital	Area 1 Patients	Area 2 Patients	Area 3 Patients	Area 4 Patients	Area 5 Patients	Total Patients Treated at Hospital
A	40	5	10	10	5	70
B	25	50	15	10	5	105
C	5	5	30	5	5	50
D	10	20	20	60	10	120
E	20	20	25	15	75	155

Appendix B: Sample of Mergers

State Data Used	Alternate Control Bed Threshold	MSA	Choice Population	Acquiring System	Primary Hospital in Acquiring System	Acquired Hospital
NC	100		HRR	Moses Cone	Moses Cone Memorial Hospital	Annie Penn Hospital
PA/NY	100		Combined PSA	Upper Allegheny Health System	Olean General Hospital	Bradford Regional Medical Center
NC	100	Wilmington	HRR	New Hanover Regional Medical Center	New Hanover Regional Medical Center	Cape Fear Memorial Hospital
MO	100	Columbia	HRR	University of Missouri Health Care	University Hospitals and Clinics	Columbia Regional Hospital
NC	100	Greensboro-Winston-Salem-High Point	HRR	Novant Health	Forsyth Memorial Hospital	Community General Hospital
NY	100	Syracuse	HRR	Upstate University Hospital	Upstate University Hospital	Community-General Hospital of Greater Syracuse
MO	100		HRR	Tenet	Lucy Lee Hospital	Doctors Regional Medical Center
NC	100	Raleigh-Durham-Chapel Hill	HRR	Duke	Duke University Medical Center	Durham Regional Hospital
PA	100		HRR	Schuylkill Health System	Pottsville Hospital	Good Samaritan Hospital
NC	100	Hickory-Morganton-Lenoir	Combined PSA	Carolinas Healthcare	Valdese General Hospital	Grace Hospital
AR	0	Hot Springs	HRR	Mercy	St. Joseph's Mercy Health Center	Healthpark Hospital
GA	100	Atlanta-Sandy Springs-Marietta	HRR	Piedmont	Piedmont Hospital	Henry Medical Center
NC	100	Greenville & Rocky Mount	HRR	UHS East	Pitt County Memorial Hospital	Heritage Hospital
NC	100	Fayetteville	HRR	Cape Fear Valley Health System	Cape Fear Valley Health System	Highsmith-Rainey Memorial Hospital
CT	100	New Haven-Milford	HRR	Yale	Yale-New Haven Hospital	Hospital of Saint Raphael
PA	100	Philadelphia-Camden-Wilmington	HRR	Abington Health	Abington Memorial Hospital	Lansdale Hospital

(Continued)

State Data Used	Alternate Control Bed Threshold	MSA	Choice Population	Acquiring System	Primary Hospital in Acquiring System	Acquired Hospital
PA	100	Pittsburgh	HRR	UPMC	UPMC Presbyterian Shadyside	Mercy Hospital
PA	100	Scranton–Wilkes-Barre	HRR	Community Health Systems	Regional Hospital of Scranton	Moses Taylor Hospital
CT	50		HRR	Western	Danbury Hospital	New Milford Hospital
GA	100	Albany	HRR	Phoebe	Phoebe Putney Memorial Hospital	Palmyra Medical Center
GA	100	Atlanta-Sandy Springs-Marietta	HRR	Piedmont	Piedmont Hospital	Piedmont Newnan Hospital
NC	50	Charlotte-Gastonia-Rock Hill	HRR	Novant Health	Presbyterian Hospital	Presbyterian Orthopaedic Hospital
NC	100	Raleigh-Durham-Chapel Hill	HRR	University of North Carolina Hospitals	University of North Carolina Hospitals	Rex Healthcare
GA	100	Atlanta-Sandy Springs-Marietta	HRR	Emory	Emory University Hospital	Saint Joseph's Hospital of Atlanta
AR/OK	0	Fayetteville-Springdale-Rogers	HRR	Community Health Systems	Northwest Medical Center	Siloam Springs Memorial Hospital
NY/PA	100	Elmira	HRR	Arnot Health	Arnot Ogden Medical Center	St. Joseph's Hospital
CT	100	Hartford-West Hartford-East Hartford	HRR	Hartford	Hartford Hospital	The Hospital of Central Connecticut
NC	100	Greensboro-Winston-Salem-High Point	HRR	Moses Cone	Moses Cone Memorial Hospital	Wesley Long Community Hospital

© The RAND Corporation 2017.

This content downloaded from
136.226.18.202 on Tue, 16 Apr 2024 13:46:20 +00:00
All use subject to <https://about.jstor.org/terms>

References

- ABADIE, A., DIAMOND, A., AND HAINMUELLER, J. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association*, Vol. 105 (2010), pp. 493–505.
- ANTWI, Y.A., GAYNOR, M., AND VOGT, W. "A Competition Index for Differentiated Products Oligopoly with an Application to Hospital Markets." Working Paper, 2013. http://www.cla.auburn.edu/economics/assets/File/AntwiGaynorVogt_Loci.pdf. Accessed June 6, 2016.
- ASHENFELTER, O., HOSKEN, D., VITA, M., AND WEINBERG, M. "Retrospective Analysis of Hospital Mergers." *International Journal of the Economics of Business*, Vol. 18 (2011), pp. 5–16.
- BRAND, K. AND BALAN, D. "Simulating Hospital Merger Simulations." Working Paper, 2013. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=993213. Accessed June 6, 2016.
- BRAND, K. AND GARMON, C. *Hospital Merger Simulation*. Member Briefing, American Health Lawyers Association, 2014. www.healthlawyers.org. Accessed June 6, 2016.
- CAPPS, C. AND DRANOVE, D. "Hospital Negotiation and Negotiated PPO Prices." *Health Affairs*, Vol. 23 (2004), pp. 175–181.
- CAPPS, C., DRANOVE, D., AND SATTERTHWAITE, M. "Competition and Market Power in Option Demand Markets." *RAND Journal of Economics*, Vol. 34 (2003), pp. 737–763.
- CARLTON, D. "Why We Need to Measure the Effect of Merger Policy and How to Do It." *Competition Policy International*, Vol. 5 (2009), pp. 77–90.
- DAFNY, L. "Estimation and Identification of Merger Effects: An Application to Hospital Mergers." *Journal of Law and Economics*, Vol. 52 (2009), pp. 523–550.
- DAFNY, L., HO, K., AND LEE, R.S. "The Price Effects of Cross-Market Mergers." NBER Working Paper no. 22106, 2016.
- DRANOVE, D. AND ODY, C. "Evolving Measures of Provider Market Power." *American Journal of Health Economics*, Vol. 2 (2016), pp. 145–160.
- ELZINGA, K. AND HOGARTY, T. "The Problem of Geographic Market Delineation in Antimerger Suits." *Antitrust Bulletin*, Vol. 18 (1973), pp. 45–81.
- FARRELL, J. AND SHAPIRO, C. "Antitrust Evaluation of Horizontal Mergers: An Economic Alternative to Market Definition." *The BE Journal of Theoretical Economics*, Vol. 10 (2010), article 9. <http://faculty.haas.berkeley.edu/shapiro/alternative.pdf>. Accessed June 6, 2016.
- FOURNIER, G. AND GAI, Y. "What Does Willingness-to-Pay Reveal About Hospital Market Power in Merger Cases?" Working Paper, 2007. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=993213. Accessed June 6, 2016.
- GARMON, C. AND KMITCH, L. "Health Care Competition or Regulation: The Unusual Case of Albany, Georgia." Working Paper, 2017.
- GAYNOR, M.S., KLEINER, S.A., AND VOGT, W.B. "A Structural Approach to Market Definition with an Application to the Hospital Industry." *Journal of Industrial Economics*, Vol. 61 (2013), pp. 243–289.
- GAYNOR, M. AND TOWN, R. "The Impact of Hospital Consolidation—Update." The Synthesis Project, Robert Wood Johnson Foundation, 2012.
- GAYNOR, M. AND VOGT, W. "Competition Among Hospitals." *RAND Journal of Economics*, Vol. 34 (2003), pp. 764–785.
- GOWRISANKARAN, G., NEVO, A., AND TOWN, R. "Mergers When Prices Are Negotiated: Evidence from the Hospital Industry." *American Economic Review*, Vol. 105 (2015), pp. 172–203.
- GUERIN-CALVERT, M., MAKI, J., AND VLADECK, B. "Re-Aligning Prospective Hospital Merger Guidance: Moving Beyond Concentration to More Meaningful Approaches." Working Paper, 2015. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2593165. Accessed June 6, 2016.
- HAAS-WILSON, D. AND GARMON, C. "Two Hospital Mergers on Chicago's North Shore: A Retrospective Study." *FTC Bureau of Economics Working Paper no. 294*, 2009.
- . "Hospital Mergers and Competitive Effects: Two Retrospective Analyses." *International Journal of the Economics of Business*, Vol. 18 (2011), pp. 17–32.
- JAFFE, S. AND WEYL, E.G. "The First-Order Approach to Merger Analysis." *American Economic Journal: Microeconomics*, Vol. 5 (2013), pp. 188–218.
- KAUL, A., KLOBNER, S., PFEIFER, G., AND SCHIELER, M. "Synthetic Control Methods: Never Use All Pre-Intervention Outcomes Together with Covariates." Working Paper, July 29, 2017. http://www.oekonometrie.uni-saarland.de/papers/SCM_Predictors.pdf. Accessed June 6, 2016.
- KEMP, R., KERSTEN, N., AND SEVERIJENEN, A. "Price Effects of Dutch Hospital Mergers: An Ex-Post Assessment of Hip Surgery." *De Economist*, Vol. 160 (2012), pp. 237–255.
- LEVIT, K., FRIEDMAN, B., AND WONG, H. "Estimating Inpatient Hospital Prices from State Administrative Data and Hospital Financial Reports." *Health Services Research*, Vol. 48 (2013), pp. 1779–1797.
- LEWIS, M. AND PFLUM, K. "Hospital Systems and Bargaining Power: Evidence from Out-of-Market Acquisitions." *RAND Journal of Economics*, Vol. 48 (2017), pp. 579–610.
- MAY, S. AND NOETHER, M. "Predicting the Price Effects of Hospital Mergers: An Evaluation of the Willingness-to-Pay Technique." Working Paper, 2014. <http://www.crai.com/sites/default/files/publications/Predicting-the-price-effects-of-hospital-mergers.pdf>. Accessed June 6, 2016.

- MILLER, N.H., REMER, M., RYAN, C., AND SHEU, G. "Upward Pricing Pressure as a Predictor of Merger Price Effects." *International Journal of Industrial Organization*, Vol. 52 (2017), pp. 216–247.
- PETERS, C. "Evaluating the Performance of Merger Simulation: Evidence from the U.S. Airline Industry." *Journal of Law and Economics*, Vol. 49 (2006), pp. 627–649.
- POWERS, D.M.W. "Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness and Correlation." *Journal of Machine Learning Technologies*, Vol. 2 (2011), pp. 37–63.
- RAVAL, D., ROSENBAUM, T., AND TENN, S. "A Semiparametric Discrete Choice Model: An Application to Hospital Mergers." *Economic Inquiry*, Vol. 55 (2017), pp. 1919–1944.
- RAVAL, D., ROSENBAUM, T., AND WILSON, N. "Industrial Reorganization: Learning About Substitution Patterns from Natural Experiments." *FTC Bureau of Economics Working Paper no. 329*, May 2016.
- ROBINSON, J. "Hospitals Respond to Medicare Payment Shortfalls By Both Shifting Costs and Cutting Them, Based on Market Concentration." *Health Affairs*, Vol. 30 (2011), pp. 1265–1271.
- SCHMITT, M. "Do Hospital Mergers Reduce Costs?" *Journal of Health Economics*, Vol. 52 (2017), pp. 74–94.
- TENN, S. "The Price Effects of Hospital Mergers: A Case Study of the Sutter-Summit Transaction." *International Journal of the Economics of Business*, Vol. 18 (2011), pp. 65–82.
- TOWN, R. AND VISTNES, G. "Hospital Competition in HMO Networks." *Journal of Health Economics*, Vol. 20 (2001), pp. 733–753.
- VITA, M. AND SACHER, S. "The Competitive Effects of Not-for-Profit Hospital Mergers: A Case Study." *Journal of Industrial Economics*, Vol. 49 (2001), pp. 63–84.
- VOGT, W.B. AND TOWN, R. "How Has Hospital Consolidation Affected the Price and Quality of Hospital Care?" The Synthesis Project, Robert Wood Johnson Foundation, 2006.
- WEINBERG, M.C. "More Evidence on the Performance of Merger Simulations." *American Economic Review: Papers & Proceedings*, Vol. 101 (2011), pp. 51–55.
- WEINBERG, M.C. AND HOSKEN, D. "Evidence on the Accuracy of Merger Simulations." *The Review of Economics and Statistics*, Vol. 95 (2013), pp. 1584–1600.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's website:

Appendix C

Appendix D: Regressions of DID Price Change on Screening Tools

Living Paycheck to Paycheck is a Way of Life for Majority of U.S. Workers, According to New CareerBuilder Survey

Study Highlights:

- **78 percent of U.S. workers live paycheck to paycheck to make ends meet**
- **Nearly one in 10 workers making \$100,000+ live paycheck to paycheck**
- **More than 1 in 4 workers do not set aside any savings each month**
- **Nearly 3 in 4 workers say they are in debt today - more than half think they will always be**
- **More than half of minimum wage workers say they have to work more than one job to make ends meet**

CHICAGO and ATLANTA, Aug. 24, 2017 /PRNewswire/ -- Do you countdown to payday? You're not alone. More than three-quarters of workers (78 percent) are living paycheck-to-paycheck to make ends meet — up from 75 percent last year and a trait more common in women than men — 81 vs. 75 percent, according to new CareerBuilder research. Thirty-eight percent of employees said they sometimes live paycheck-to-paycheck, 17 percent said they usually do and 23 percent said they always do.

The national survey, which was conducted online by Harris Poll on behalf of CareerBuilder from May 24 to June 16, 2017, included representative samples of 2,369 full-time employers and 3,462 full-time U.S. workers across industries and company sizes in the private sector.

More Money, Less Financial Headaches?

Having a higher salary doesn't necessarily mean money woes are behind you, with nearly one in 10 workers making \$100,000 or more (9 percent) saying they usually or always live paycheck-to-paycheck and 59 percent in that income bracket in debt. Twenty-eight percent of workers making \$50,000-\$99,999 usually or always live paycheck to paycheck, 70 percent are in debt; and 51 percent of those making less than \$50,000 usually or always live paycheck to paycheck to make ends meet, 73 percent are in debt.

"As an employer, your employees' financial problems become your financial problems," said Rosemary Haefner, chief human resources officer for CareerBuilder. "If workers are constantly thinking about their financial struggles, their quality of work can decrease, and it can take a hit on their morale and productivity. If you do what you can to help people keep their finances under control — by doing things such as matching 401(k) contributions or hosting financial planning seminars — you'll ease some of their financial worries and it will be less likely to have a negative impact on your business."

Debt is a Growing Issue for Workers

A quarter of workers (25 percent) have not been able to make ends meet every month in the last year, and 20 percent have missed payment on some smaller bills. Further, 71 percent of all workers say they're in debt — up from 68 percent last year. While 46 percent say their debt is manageable, more than half of those in debt (56 percent) say they feel they will always be in debt. And it should be noted that 18 percent of all workers have reduced their 401k contribution and/or personal savings in the last year, more than a third (38 percent) do not participate in a 401k plan, IRA or comparable retirement plan, and 26 percent have not set aside any savings each month in the last year.

Less than a third of workers (32 percent) stick to a clearly defined budget and a slight majority (56 percent) save \$100 or less a month:

- None: 26 percent
- Less than \$50: 15 percent
- \$51 to \$100: 16 percent
- \$101 to \$250: 14 percent
- \$251 to \$500: 11 percent
- \$501 to \$750: 5 percent
- \$751 to \$1,000: 4 percent
- More than \$1,000: 10 percent

Still, despite financial woes, there are certain things employees aren't willing to give up. When asked what they'd absolutely not give up, regardless of financial concerns, employees cited:

- Internet connection: 54 percent
- Mobile device (smart phone, tablet, etc.): 53 percent
- Driving: 48 percent
- Pets: 37 percent
- Cable: 21 percent

- Going out to eat: 19 percent
- Traveling: 17 percent
- Education: 13 percent
- Buying gifts for people: 13 percent
- Alcohol: 11 percent

Minimum Wage Workers Can't Make Ends Meet

The majority of workers (81 percent) have worked a minimum wage job, and 71 percent of them were not able to make ends meet financially during that time — more than half (54 percent) had to work more than one job.

To alleviate some financial burden, 83 percent of employers that are hiring minimum wage workers this year (45 percent) will be raising the minimum wage at their organization.

Survey Methodology

This survey was conducted online within the U.S. by Harris Poll on behalf of CareerBuilder among 2,369 hiring and human resource managers ages 18 and over (employed full-time, not self-employed, non-government) and 3,462 employees ages 18 and over (employed full-time, not self-employed, non-government) between May 24 and June 16, 2017 (percentages for some questions are based on a subset, based on their responses to certain questions). With pure probability samples of 2,369 and 3,462, one could say with a 95 percent probability that the overall results have sampling errors of +/- 2.01 and +/- 1.67 percentage points, respectively. Sampling error for data from sub-samples is higher and varies.

About CareerBuilder®

CareerBuilder is a global, end-to-end human capital solutions company focused on helping employers find, hire and manage great talent. Combining advertising, software and services, CareerBuilder leads the industry in recruiting solutions, employment screening and human capital management. It also operates top job sites around the world. CareerBuilder and its subsidiaries operate in the United States, Europe, Canada and Asia. For more information, visit www.careerbuilder.com.

Media Contact

Ladan Nikravan Hayes
312.698.0538

ladan.hayes@careerbuilder.com
<http://www.twitter.com/CareerBuilderPR>

SOURCE CareerBuilder

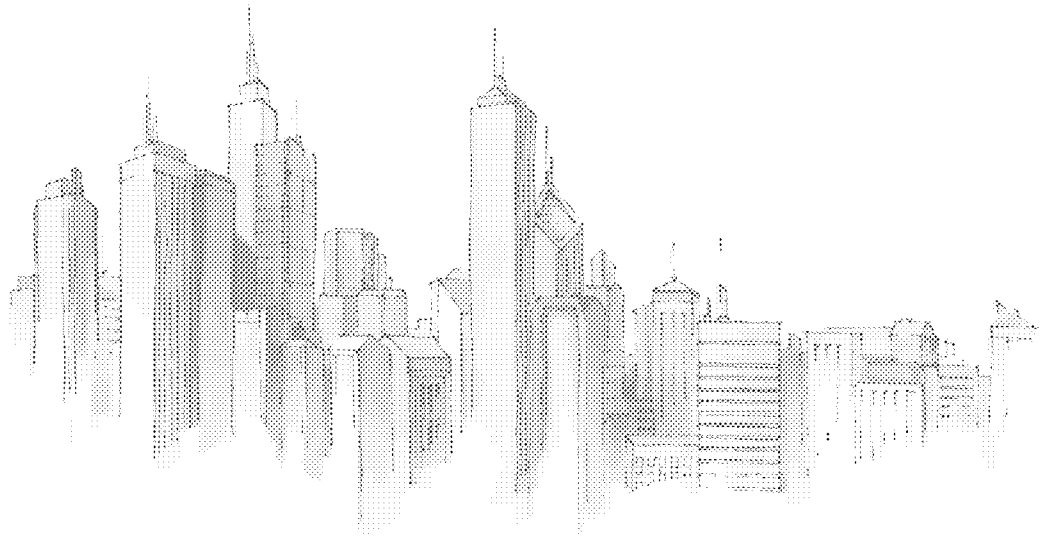
Additional assets available online: [Photos \(1\)](#)

<https://press.careerbuilder.com/2017-08-24-Living-Paycheck-to-Paycheck-is-a-Way-of-Life-for-Majority-of-U.S.-Workers-According-to-New-CareerBuilder-Survey>



A GUIDE FOR GOVERNMENT AGENCIES

How to Comply with the Regulatory Flexibility Act



Published August 2017



Created by Congress in 1976, the Office of Advocacy of the U.S. Small Business Administration is an independent voice for small business within the federal government.

The office is the watchdog of the Regulatory Flexibility Act (RFA) and the source of small business statistics. The office advances the views and concerns of small businesses before Congress, the White House, the federal agencies, the federal courts, and state policymakers.



Website

www.sba.gov/advocacy



Email

advocacy@sba.gov



Blog

advocacysba.sites.usa.gov



Facebook

AdvocacySBA



Twitter

[@AdvocacySBA](https://twitter.com/AdvocacySBA)



Listservs

www.sba.gov/content/connect-us-0



FOREWORD

Economic freedom is the foundation for individual success and prosperity. This freedom is evident in the entrepreneurial small business sector, which creates most of the new jobs and a large share of the innovations in the American economy. When government takes small businesses into consideration in developing regulations, it saves time and money and supports the growth of the nation's most productive sector.

The Regulatory Flexibility Act (RFA) and related laws and executive orders require federal agencies to consider the effects of regulations on small entities. Executive Order 13,272, signed on August 13, 2002, directs the Small Business Administration's Office of Advocacy to provide federal agencies with training and information on how to comply with the RFA. This manual is a sourcebook for agencies to comply with the Act.

The Office of Advocacy continues to provide training to agency personnel in RFA compliance and welcomes additional opportunities to assist in new phases of training. This compliance guide, prepared with input from regulatory agencies, is designed to be used by agency rule writers and policy analysts as a step-by-step manual for complying with the RFA. A careful review of the requirements is recommended before policy analysts begin to draft regulations, and then again at each stage of the process.

Thanks to all who contributed by reviewing and commenting on this guide. Further suggestions for improvements are welcome. For more information about the RFA, E.O. 13,272, and subsequent developments, visit the Advocacy website at www.sba.gov/advocacy, or call us at (202) 205-6533.

To those charged to carry out the nation's regulatory flexibility requirements, the Office of Advocacy offers its strong support and encouragement. You have a crucial role in keeping the nation on track for economic growth by ensuring the strength of the resilient small business sector.

Office of Advocacy
U.S. Small Business Administration
August 2017

ACKNOWLEDGMENTS

As a tool for effective implementation of the Regulatory Flexibility Act, the guide helps create fairer and more effective regulation for all small entities, especially small businesses.

The Office of Advocacy appreciates the efforts of all who reviewed this guide, including representatives of the Department of Labor, the Environmental Protection Agency, the Food and Drug Administration's Center for Food Safety and Applied Nutrition, the Office of Management and Budget's Office of Information and Regulatory Affairs, the House Committee on Small Business, and Advocacy staff.

A NOTE ABOUT THE AUGUST 2017 EDITION

The August 2017 edition is the first update to *A Guide for Government Agencies on How to Comply with the Regulatory Flexibility Act* since May 2012. The information detailing how to comply with the Act remains unchanged. The introduction has been updated to include new executive orders.

There are several updates to the appendixes. The section called "RFA recent developments" which was formerly Appendix B was removed. The two appendixes dealing with the economics of small businesses have been brought up to date (Appendixes B and C in the 2017 edition). Three additional executive orders have been added in three new appendixes:

- E.O. 13,610, Identifying and Reducing Regulatory Burdens (Appendix I);
- E.O. 13,771, Reducing Regulation and Controlling Regulatory Costs, (Appendix J); and
- E.O. 13,777, Enforcing the Regulatory Reform Agenda (Appendix K).

CONTENTS

Foreword	iii
Acknowledgments.....	iv
A Note about the August 2017 Edition	iv
Contents	v
Introduction: The RFA and related law	1
Chapter 1 Where do we begin? First steps of RFA analysis.....	7
The RFA decision process	8
Does the RFA apply?	9
RFA exemptions	9
RFA now applies to certain IRS interpretative rules	10
The RFA threshold analysis: Can we certify?	11
Organizing the threshold report	12
“Factual basis” requirement for certification	13
What is the definition of a small entity?	14
Changing a size standard	15
Assessing the impact on small entities.....	15
Which segment of the economy or industry will be regulated?.....	16
How to categorize small entity sectors	18
Determination of “significant impact”	18
Determination of “substantial number”	21
Direct versus indirect impact	22
Adverse versus beneficial impact	23
What adequate and inadequate certifications look like.....	24
An example of an adequate certification	25
Examples of inadequate certifications	25
Certification checklist.....	29
Chapter 2 Preparing a proposed rule: The initial regulatory flexibility analysis.....	31
The RFA decision process	31
Issues to be addressed in the analysis	32
Elements of an IRFA	34
Reasons action is being considered.....	35
Objectives of the proposed rule	35
Description and estimate of the number of small entities.....	36
Estimating compliance requirements	36
Significant alternatives considered	37
Duplicative, overlapping, and conflicting rules.....	40
Using other analyses to satisfy the IRFA requirements.....	41
When an IRFA may be waived or delayed	41
What an IRFA should look like: A real-life example	42

Chapter 3 Preparing a final rule: The final regulatory flexibility analysis	43
The RFA decision process	43
Issues to be addressed in the analysis	44
Additional questions to be addressed in a FRFA	46
Have all significant issues been assessed?	46
Has the number of small entities been estimated?	46
Has the adverse economic impact on small entities been minimized?	47
Have all significant alternatives been reviewed?	48
Permissible delays in publication; provision for lapse of final rule	48
What a FRFA should look like: A real-life example	49
Chapter 4 SBREFA panels.....	51
Who must hold SBREFA panels?	51
How is the decision to hold a SBREFA panel made?	52
How does a SBREFA panel work?	52
Suggested SBREFA panel timeline	53
Chapter 5 RFA litigation: What the courts have said	55
Where do we begin? First steps of RFA rule analysis	55
Does the RFA apply?	55
What qualifies as a rulemaking under the APA?	56
Exemptions under the APA	56
The certification statement.....	58
The decision process	58
Size standards.....	61
The agency must conduct an adequate analysis before certifying.....	62
Direct versus indirect impact on small entities	64
The initial regulatory flexibility analysis	68
The final regulatory flexibility analysis	69
General content	69
Is a FRFA always required?	70
Considering alternatives to the final rule	71
Analysis of the economic impact	74
Public comments	76
Judicial review	77
Chapter 6 Section 610 review of existing rules	79
RFA section 610 review.....	80
Review of rules that were originally certified.....	80
Scope of the review.....	81
Timing of the review.....	82
Outreach to regulated small entities.....	83
Using other agency reviews to satisfy section 610	84
Examples of successful retrospective rule reviews.....	84
Section 610 assistance from the Office of Advocacy	85

Chapter 7 Additional RFA and SBREFA requirements	87
Semi-annual regulatory agenda.....	87
Small entity compliance guides	88
Informal compliance assistance	89
Regulatory enforcement fairness	90
Penalty reduction policies.....	90
Congressional review.....	91
Conclusion	92
Appendix A The Regulatory Flexibility Act	93
Appendix B Small business by the numbers	102
Appendix C Small business statistics for regulatory analysis	106
Appendix D Executive Order 12,866	109
Appendix E Executive Order 13,272	121
Appendix F Executive Order 13,563 and memoranda	123
Appendix G Memorandum on cumulative impacts.....	130
Appendix H Executive Order 13,579	132
Appendix I Executive Order 13,610	134
Appendix J Executive Order 13,771	136
Appendix K Executive Order 13,777	139
Appendix L Example of a successful IRFA.....	143
Appendix M Example of a successful FRFA.....	175
Appendix N Abbreviations used in this guide.....	204
Appendix O Office of Advocacy staff.....	206
Index	207

Figures and Tables

Figure 1. The RFA decision process: First steps	8
Figure 2. The RFA decision process: IRFA.....	31
Figure 3. The RFA decision process: FRFA.....	43
Figure 4. Suggested SBREFA panel timeline.....	54
Table 1. Certification checklist.....	29

INTRODUCTION: THE RFA AND RELATED LAW

In June 1976, Congress created the Office of Advocacy, headed by a Chief Counsel appointed by the President from the private sector and confirmed by the Senate. Congress concluded that small businesses needed a voice in the councils of government—a voice that was both independent and credible. Congress specifically required the Office of Advocacy to measure the costs and impacts of regulation on small business. The Chief Counsel’s mandate, therefore, is to be an independent voice for small business in policy deliberations—a unique mission in the federal government.

The Regulatory Flexibility Act (RFA),¹ enacted in September 1980, requires agencies to consider the impact of their regulatory proposals on small entities, analyze effective alternatives that minimize small entity impacts, and make their analyses available for public comment. The RFA applies to a wide range of small entities, including small businesses, not-for-profit organizations, and small governmental jurisdictions.

The RFA does not seek preferential treatment for small entities, nor does it require agencies to adopt regulations that impose the least burden on them, or mandate exemptions for them. Rather, it requires agencies to examine public policy issues using an analytical process that identifies barriers to small business competitiveness and seeks a level playing field for small entities, not an unfair advantage.

The size of the business, government unit, or not-for-profit organization being regulated has a bearing on its ability to comply with federal regulations. For example, the costs of complying with a particular regulation—measured in staff time, recordkeeping, outside expertise, and other direct compliance costs—might be roughly the same for a company with sales of \$10 million as for a company with sales of \$1 million. In a larger business, however, the costs of compliance can be spread over a larger volume of production. For small entities, a burdensome regulation could affect the ability to set competitive prices, to devise innovations, or even to make a profit.² In some cases, a small business may be unable to stay in business because of the cost of a regulation. Simply stated, fixed costs have a greater impact on small entities because small entities have fewer options for recovering them. Without the necessary facts, it is possible for an agency to cause serious unintended or unforeseen adverse impacts on small businesses.

In essence, the RFA asks agencies to be aware of the economic structure of the entities they regulate and the effect their regulations may have on small entities. To this end, the RFA requires agencies to analyze the economic impact of proposed regulations when there is likely to be a significant economic impact on a substantial number of small entities, and to consider regulatory alternatives that will achieve the agency’s goal while

¹ Regulatory Flexibility Act, Pub. L. No. 96-354, 94 Stat. 1164 (codified at 5 U.S.C. § 601).

² See Todd A. Morrison, *Economies of Scale in Regulatory Compliance: Evidence of the Differential Impacts of Regulation by Firm Size*, report no. PB85-178861, prepared by Jack Faucett Associates, Inc., for the U.S. Small Business Administration, Office of Advocacy (Springfield, Va.: National Technical Information Service, 1985).

minimizing the burden on small entities. The concept underlying this analytical requirement is that agencies will revise their decision-making processes to take account of small entity concerns in the same manner that agency decision-making processes were modified subsequent to the enactment of the National Environmental Policy Act (NEPA)³ The RFA then acts as a statutorily mandated analytical tool to further assist agencies in meeting the rational rulemaking standard set forth in the Administrative Procedure Act (APA), just as NEPA was intended to rationalize decisions concerning major federal actions that would affect the environment.

The Small Business Regulatory Enforcement Fairness Act (SBREFA), enacted in March 1996,⁴ amended the RFA and provided additional tools to aid small business in the fight for regulatory fairness. The amendments made by SBREFA include:

- Judicial review of agency compliance with some of the RFA's provisions.
- Requirements for more detailed and substantive regulatory flexibility analyses.
- Expanded participation by small entities in the development of rules by the Occupational Safety and Health Administration (OSHA) and the Environmental Protection Agency (EPA).
- Requirements that agencies prepare and publish guides to assist small entities in complying with final rules.

Subsequently, several laws have been passed amending and strengthening SBREFA and the RFA.

The Small Business and Work Opportunity Act of 2007 amended SBREFA to strengthen the requirement that agencies prepare compliance guides for any rule for which they must prepare a final regulatory flexibility analysis. Agencies are required to publish the guides not later than the effective date of the requirements, post them to websites, distribute them to industry contacts, and report annually to Congress.⁵

The Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010 established the Consumer Financial Protection Bureau (CFPB) and required the agency to comply with the RFA Section 609 panel process, making it the third such agency, along with the EPA and OSHA.⁶ In addition to the regular requirements of the initial regulatory flexibility analysis (IRFA) found in 5 USC 603, a CFPB IRFA must include:

a description of (A) any projected increase in the cost of credit for small entities; (B) any significant alternatives to the proposed rule which accomplish the stated objectives of applicable statutes and which minimize any increase in the cost of credit for small entities; and (C)

³ See *Associated Fisheries of Maine v. Daley*, 127 F.3d 104, 114 (1st Cir. 1997) noting parallels between NEPA and the RFA.

⁴ Small Business Regulatory Enforcement Fairness Act of 1996, Pub. L. No. 104-121, 110 Stat. 857 (codified at 5 U.S.C. § 601 et seq.).

⁵ Pub. Law 110-28 (May 27, 2007), Title 6, subtitle B, § 7005.

⁶ Pub. Law 111-203 (July 21, 2010).

advice and recommendations of representatives of small entities relating to issues described in subparagraphs (A) and (B) and subsection (b).⁷

When the CFPB produces a final regulatory flexibility analysis, it must include “a description of the steps the agency has taken to minimize any additional cost of credit for small entities.”

The Small Business Jobs Act of 2010 amended several requirements in the final regulatory flexibility analysis (FRFA) section of the RFA.⁸

- It struck the word “succinct” from the requirement for “a succinct statement of the need for, and objectives of, the rule.”⁹
- It replaced the word “summary” with the word “statement” twice in the requirement for “a summary of the significant issues raised by the public comments . . . , a summary of the assessment of the agency of such issues . . . ,”¹⁰ and
- It codified a requirement of Executive Order 13,272¹¹ by adding a paragraph requiring the FRFA to include “the response of the agency to any comments filed by the Chief Counsel for Advocacy of the Small Business Administration in response to the proposed rule, and a detailed statement of any change made to the proposed rule in the final rule as a result of the comments.”¹²

In addition to this legislation, a number of executive orders have been issued to improve agency compliance with the RFA. The complete text of these executive orders is in Appendixes D through K.

Executive Order 12,866, Regulatory Planning and Review. Issued by President Bill Clinton in 1993, this order lays out an analytical framework for rulemaking agencies and establishes regulatory goals to help agencies¹³ understand the importance of conducting regulatory flexibility analyses. Its regulatory philosophy provides a relevant context for discussions surrounding an agency’s certification decision.

Federal agencies should promulgate only such regulations as are required by law, are necessary to interpret the law, or are made necessary by compelling public need, such as material failures of private markets to protect or improve the health and safety of the public, the environment, or the well-being of the American people. In deciding whether and how to regulate, agencies should assess all costs and benefits of available regulatory alternatives, including the alternative of not regulating. Costs

⁷ Title X, § 1100G(b) of the Dodd-Frank Act amending 5 U.S.C. § 603(d)(1)-(2), §604(a)(6) and § 609(d)(2). See Appendix A.

⁸ Pub. Law 111-240 (September 27, 2010).

⁹ *Id.*, § 1601(1), amending 5 U.S.C. § 604(a)(1).

¹⁰ *Id.*, § 1601(2), amending 5 U.S.C. § 604(a)(2).

¹¹ See the full executive order in Appendix E.

¹² *Id.*, § 1601(3), amending 5 U.S.C. § 604(a)(3).

¹³ Exec. Order No. 12,866 does not apply to independent regulatory commissions such as the Federal Election Commission, the Federal Communications Commission, and the Securities and Exchange Commission. Appendix D contains the complete text.

and benefits should include both quantifiable measures (to the fullest extent possible) and qualitative measures of costs and benefits that are difficult to quantify, but essential to consider.¹⁴

Executive Order 12,866 also specifies 12 principles for agencies to follow in developing regulations. The eleventh principle has particular relevance to the RFA certification decision¹⁵ and the analysis needed to prepare a factual basis for that decision:

Each agency shall tailor its regulations to impose the least burden on society, including individuals, businesses of differing sizes, and other entities (including small communities and governmental entities), consistent with obtaining the regulatory objectives, taking into account, among other things, and to the extent practicable, the costs of cumulative regulations.¹⁶

Executive Order 13,272, Proper Consideration of Small Entities in Agency Rulemaking. Issued by President George W. Bush in August 2002, this order’s purpose is to ensure that federal agencies work closely with Advocacy to address small business issues as early as possible in the regulatory process, particularly as they relate to disproportionate regulatory burden. It also requires agencies to publish how they will comply with the statutory mandates of the RFA.¹⁷ The order sets out a series of responsibilities for both regulating agencies and the Office of Advocacy.

- Agencies will establish policies on how to measure their impact on small entities and will work with Advocacy to establish those procedures.
- The Office of Advocacy is instructed to train agencies on how to properly account for small entity impact when agencies draft regulations and to continue to work with agencies.
- Agencies are to submit proposed rules with significant small entity effects to the Office of Advocacy prior to publication and are required to consider the Office of Advocacy’s comments on the rule.¹⁸
- The Office of Advocacy is required to report annually to the Office of Management and Budget (OMB) on whether agencies are complying with this executive order.

Executive Order 13,563, Improving Regulation and Regulatory Review.¹⁹ Issued in January 2011, this order reaffirms and amplifies the principles embodied in E.O. 12,866

¹⁴ Exec. Order No. 12,866 § 1(a), 58 Fed. Reg. 51,735 (Sept. 30, 1993). For complete text, see Appendix D.

¹⁵ 5 U.S.C. § 605(b). The RFA permits an agency to certify that a proposed rule would not have a significant economic impact on a substantial number of small entities, if the preliminary (threshold) analysis supports such a decision.

¹⁶ Exec. Order No. 12,866 § 1(b). Note that E.O. 12,866 applies to individuals and requires that regulations impose the least burden on society—standards that differ from those of the RFA. However, the fact that application of the order must be “consistent with” maintaining an agency’s regulatory objectives makes the order somewhat parallel to the RFA.

¹⁷ Exec. Order No. 13,272, 67 Fed. Reg. 53,461 (Aug. 13, 2002). For complete text, see Appendix E.

¹⁸ The Small Business Jobs Act, P.L. 111-240, codified this requirement of E.O. 13,272 in 5 U.S.C. § 604(a)(3).

by encouraging agencies to coordinate their regulatory activities, and to consider regulatory approaches that reduce the burden of regulation while maintaining flexibility and freedom of choice for the public. The order also mandates the retrospective review of existing regulations, a process made permanent by Executive Order 13,610, Identifying and Reducing Regulatory Burdens, in 2012.²⁰

A memorandum titled Regulatory Flexibility, Small Business, and Job Creation was issued concurrently with E.O. 13,563. The memorandum reaffirms E.O. 12,866 and reiterates the RFA's provisions for providing regulatory flexibility. The memorandum directs agencies to provide an explicit justification "whenever an executive agency chooses, for reasons other than legal limitations, not to provide such flexibility in a proposed or final rule that is likely to have a significant economic impact on a substantial number of small entities."²¹

Executive Order 13,579, Regulation and Independent Regulatory Agencies, reaffirms the directives of Executive Order 13,563 and extends it, to the extent permitted by law, to independent agencies.²²

In 2017, President Donald Trump issued Executive Orders 13,771 and 13,777 to streamline and eliminate unnecessary regulations.

Executive Order 13,771, Reducing Regulations and Controlling Regulatory Costs.²³ Known as "One In, Two Out," this order mandates that if an agency publishes a proposed rule for notice and comment or promulgates a new rule, it must identify at least two existing regulations for repeal. Further, total incremental costs for new rules must be no greater than zero (with some exceptions). Agencies must also identify in their annual Regulatory Plan offsetting regulations for each regulation that increases incremental costs, and regulations not included in most recent Unified Regulatory Agenda shall not be issued without prior approval.

Executive Order 13,777, Enforcing the Regulatory Reform Agenda.²⁴ This order requires each agency to designate a Regulatory Reform Officer (RRO) to oversee regulatory reform initiatives, such as E.O. 12,866, E.O. 13,563, and E.O. 13,771. Each agency must also establish a Regulatory Reform Task Force to evaluate existing regulations and make recommendations regarding the repeal, replacement, or modification of regulations, especially those that eliminate jobs or inhibit job creation, are outdated, unnecessary, or ineffective, or impose costs that exceed benefits. Each task force must seek input and assistance from affected entities such as state, local, and tribal governments, small businesses, consumers, non-governmental organizations, and trade associations. Agencies must issue a report detailing progress toward improving

¹⁹ Exec. Order No. 13,563, 76 Fed. Reg. 3,821 (Jan. 21, 2011). For complete text, see Appendix F.

²⁰ Exec. Order No. 13,610, 77 Fed. Reg. 28,469 (May 14, 2012). For the complete text, see Appendix I.

²¹ This language mirrors the RFA language at 5 U.S.C. § 604 (a)(6) for final rules.

²² Exec. Order No. 13,579 § 1(c). For complete text, see Appendix H.

²³ Exec. Order No. 13,771, 82 Fed. Reg. 9,339 (Jan. 30, 2017). For complete text, see Appendix J.

²⁴ Exec. Order No. 13,777, 82 Fed. Reg. 12,285 (Feb. 24, 2017). For complete text, see Appendix K.

implementation of regulatory reform initiatives and identifying regulations for repeal, replacement, or modification. Agencies will measure and incorporate progress into their performance indicators required by the Government Performance and Results Act.

These executive orders reinforce executive intent that agencies give serious attention to their rules' impacts on small entities and that they develop a comprehensive set of regulatory alternatives to reduce the regulatory burden on small entities.

Regional Regulatory Reform Roundtables. In order to help facilitate this effort, the Office of Advocacy has been meeting with federal agencies to discuss their regulatory reform agendas. Advocacy has also embarked on a series of regional regulatory reform roundtables to obtain direct input from small businesses on which regulations are most burdensome and in need of retrospective review and reform. Advocacy plans to report to the White House, federal agencies, and Congress on the results of these efforts.

Using this Guide to Comply with the RFA

The Office of Advocacy's compliance guide should be utilized by regulatory agencies as a tool for following the requirements of the Regulatory Flexibility Act and related law and executive orders. This revised guide is the product of Advocacy's decades of experience with the RFA and reflects the spirit of interagency cooperation, as well as the vital importance of small business to the economy. Advocacy hopes the guide will be a useful tool and welcomes comments on ways to improve its usefulness to regulatory agencies.

The guide includes how-to information on determining when the RFA applies to a proposed regulation, performing initial and final regulatory flexibility analyses, and meeting other RFA requirements, including periodic review of existing rules and small business compliance guides. Also included are sections on litigation so that agencies may learn how courts have ruled on RFA compliance and examples of actual agency regulatory analyses. For more assistance, contact the Office of Advocacy at (202) 205-6533, or one of the Advocacy contacts listed in Appendix O.

CHAPTER 1 WHERE DO WE BEGIN? FIRST STEPS OF RFA ANALYSIS

We begin by briefly examining the general purpose of the Regulatory Flexibility Act and its overall requirements. The Regulatory Flexibility Act requires agencies to consider the impact of their rules on small entities.²⁵ When the proposed regulation will impose a significant economic impact on a substantial number of small entities, the agency must evaluate alternatives that would accomplish the objectives of the rule without unduly burdening small entities. Inherent in the RFA is a desire to remove barriers to competition and encourage agencies to consider ways of tailoring regulations to the size of the regulated entities.²⁶

The RFA, like the National Environmental Policy Act, imposes analytical requirements on federal agencies. Both statutes require disclosure of effects and mechanisms to reduce adverse consequences and improve beneficial consequences.²⁷ The RFA does not require that agencies necessarily minimize a rule's impact on small entities if there are significant legal, policy, factual, or other reasons for not minimizing the impact. The RFA requires only that agencies determine, to the extent practicable, the rule's economic impact on small entities and explore regulatory alternatives for reducing any significant economic impact on a substantial number of such entities. Once that process is finished, agencies must explain the reasons for their ultimate regulatory choices.

The goal of Congress in creating the RFA was to change the regulatory culture in agencies and mandate that they consider regulatory alternatives that still achieve statutory purposes, while minimizing the impacts on small entities. Regulatory flexibility analyses built into the regulatory development process at the earliest stages will help agency decision makers achieve regulatory goals with realistic, cost-effective, and less burdensome regulations.

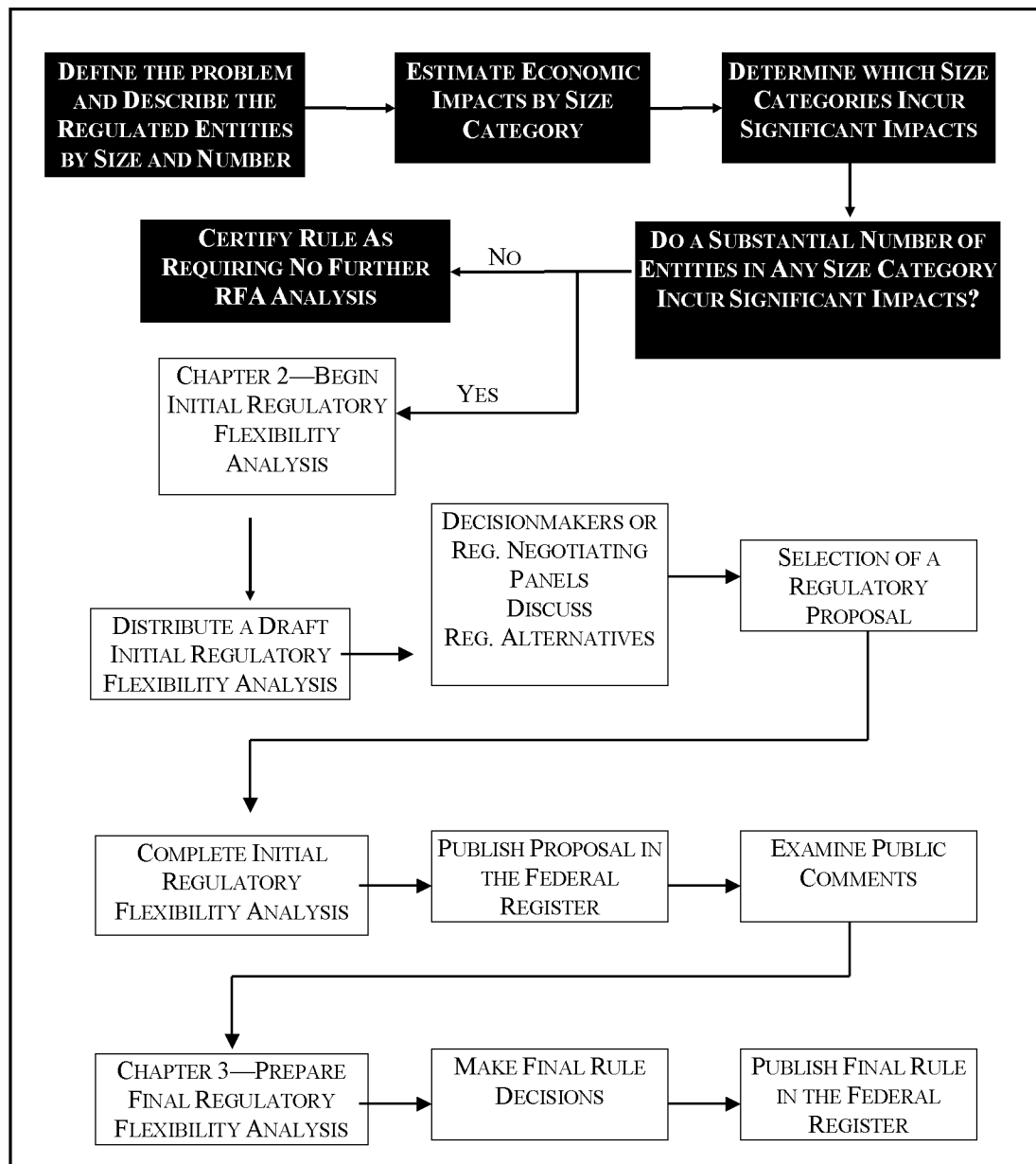
The following chart (Figure 1) shows an overall picture of the RFA decision-making process. This chapter focuses on the first steps, highlighted in the chart.

²⁵ See this chapter's section titled "What is the definition of a small entity?"

²⁶ See generally, Findings and Purposes, Sec. 2(a)–(b).

²⁷ Nothing in the RFA states that an economic impact must be adverse prior to performing an analysis.

Figure 1. The RFA decision process: First steps



Does the RFA apply?

One of the first decisions to make is whether the Regulatory Flexibility Act applies to the particular regulation. The RFA applies to any rule subject to notice and comment rulemaking under section 553(b) of the Administrative Procedure Act (APA)²⁸ or any other law. This includes any rule of general applicability governing federal grants to state and local governments, for which agency procedures provide opportunity for notice and comment. For instance, some agencies, such as the Rural Utilities Service, have their own administrative rules that require notice and comment even though the agency's rules may be exempt from the APA notice and comment requirement.²⁹

RFA exemptions

Rules that are exempt from APA notice and comment requirements are also exempt from the RFA requirements when any of the following is involved: (1) a military or foreign affairs function of the United States, or (2) a matter relating to agency management or personnel or to public property, loans, grants, benefits, or contracts.³⁰ In addition, except where notice or hearing is required by statute, the APA does not apply (1) to interpretative rules,³¹ general statements of policy, or rules of agency organization, procedure or practice; or (2) when the agency for good cause finds (and incorporates the finding and a brief statement of reasons therefor in the rules issued) that notice and public procedure thereon are impracticable, unnecessary, or contrary to the public interest.³² Under any of the circumstances described above, the RFA would not apply.

Interpretative rules generally interpret the intent expressed by Congress. The easiest type of interpretative rule to recognize is one in which an agency does not insert its own judgments in implementing a rule, and simply regurgitates statutory language. One legal treatise on the subject says that interpretative rules are any rules that an agency issues without exercising delegated legislative power to make law through rules.³³ The treatise goes on to state that the difference between legislative and interpretative rulemaking is the weight courts give the agency decisions on review.³⁴

In the case of legislative rules, agencies are given the authority to establish requirements not specifically mentioned in the authorizing statute that may be the basis for a rule. An example of this would be setting an ambient air quality standard or regulating in the public interest as set out in the Communications Act of 1934. See *Whitman v. American*

²⁸ 5 U.S.C § 553(b); see also § 601(2).

²⁹ The "other law" requirement includes situations where the agency binds itself by rule to act through rulemaking rather than by a procedure that does not require notice and comment.

³⁰ *Id.* at § 553(a). There are statutes, such as the Competition in Contracting Act, the Federal Acquisition Streamlining Act, and the Federal Acquisition Reform Act, that mandate that changes to contracting rules be issued pursuant to notice and comment.

³¹ The exception is certain Internal Revenue Service interpretative rules. See the discussion below.

³² *Id.* at § 553(b)(A).

³³ K. Davis, *Administrative Law Treatise*, § 7:8 (1958).

³⁴ Davis at §§ 7:8-7:13.

Trucking Associations for a discussion of what constitutes a standard governing delegation of legislative authority by Congress to the executive branch.³⁵

The RFA presents its own exemptions as well. Section 601(2) states that the RFA does not apply to rules of *particular applicability* relating to rates, wages, corporate or financial structures, or reorganizations thereof, prices, facilities, appliances, services or allowances therefor, or to valuations, costs or accounting, or practices relating to such rates, wages, structures, appliances, services, or allowances.³⁶ The RFA's definition of a rule is less inclusive than the definition of a rule under the Administrative Procedure Act, which defines a "rule" as "an agency statement of *general or particular applicability*." The original draft of the APA limited the definition of rules to "statements of general applicability" or "having a *general* application to members of a broadly identifiable class."³⁷ This is contrasted with statements of "particular applicability" or applying "only to specific individuals or situations" or "named parties."³⁸ Therefore, the RFA applies to rules affecting the general public, as opposed to those that affect specific individuals.

RFA now applies to certain IRS interpretative rules

The Small Business Regulatory Enforcement Fairness Act amended the RFA to bring certain interpretative rulemakings of the Internal Revenue Service (IRS) within the scope of the RFA. The law now applies to those IRS rules published in the *Federal Register* (that would normally be exempt from the RFA as interpretative rules) that impose a "collection of information" requirement on small entities.³⁹ Congress took care to define the term "collection of information" to be identical to the term used in the Paperwork Reduction Act, which means that a collection of information includes any reporting or recordkeeping requirement for more than nine people.⁴⁰

³⁵ American Trucking Ass'ns v. EPA, 175 F.3d 1027, 1044 (D.C. Cir. 1999); Whitman v. American Trucking Ass'ns, 531 I/S/ 457 (2001).

³⁶ 5 U.S.C. § 601(2).

³⁷ Bernard Schwartz, *Administrative Law: A Casebook*, at 255-262 (2d ed. 1984).

³⁸ *Id.*

³⁹ *Id.* at § 603(a).

⁴⁰ *Id.* at § 601(7).

(7) The term "collection of information"

(A) means the obtaining, causing to be obtained, soliciting, or requiring the disclosure to third parties or the public, of facts or opinions by or for an agency, regardless of form or format, calling for either—

(i) answers to identical questions posed to, or identical reporting or recordkeeping requirements imposed on, 10 or more persons, other than agencies, instrumentalities, or employees of the United States; or

(ii) answers to questions posed to agencies, instrumentalities, or employees of the United States which are to be used for general statistical purposes; and

(B) shall not include a collection of information described under section 3518(c)(1) of title 44, United States Code. (8) The term "record-keeping requirement" means a requirement imposed by an agency on persons to maintain specified records.

The RFA threshold analysis: Can we certify?

After an agency begins regulatory development and determines that the RFA applies, it must decide whether to conduct a full regulatory flexibility analysis or to certify that the proposed rule will not “have a significant economic impact on a substantial number of small entities.”⁴¹ The record an agency builds to support a decision to certify is subject to judicial review.⁴²

In order to certify a rule under the RFA, an agency should be able to answer the following types of questions:

- Which small entities will be affected?
- Have adequate economic data been obtained?
- What are the economic implications/impacts of the proposal or do the data reveal a significant economic impact on a substantial number of small entities?

If, after conducting an analysis for a proposed or final rule, an agency determines that a rule will not have a significant economic impact on a substantial number of small entities, section 605(b) provides that the head of the agency may so certify. The certification must include a statement providing the *factual basis* for this determination, and the certification shall be published in the *Federal Register* at the time the proposed or final rule is published for public comment. The agency is also required to provide such certification and statement to the SBA’s Chief Counsel for Advocacy.⁴³ A certification must include, at a minimum, a description of the affected entities and the impacts that clearly justify the “no impact” certification. The agency’s reasoning and assumptions underlying its certification should be explicit in order to obtain meaningful public comment and thus receive information that would be used to re-evaluate the certification.

Clearly, an agency should identify the scope of the problem and the impact of the solution on affected entities before moving forward with a regulatory proposal. At times, despite a good-faith effort on the part of an agency to obtain data, an agency may still be uncertain about whether to certify. In those instances, an advance notice of proposed rulemaking (ANPRM) may be necessary to solicit data. As a final recourse, the agency should err on the side of caution and perform an initial regulatory flexibility analysis (IRFA) with the available data and information, and solicit comments from small entities regarding impact.⁴⁴ Then, if appropriate, the agency can certify the final rule. If an

⁴¹ 5 U.S.C. § 605(b). The decision to certify a rule parallels the finding of no significant impact under NEPA. As with a NEPA determination, the decision to certify, because it is subject to judicial review, should be based on a sound threshold analysis similar to the environmental assessment mandated in Council on Environmental Quality regulations to support a finding of no significant impact or laying the groundwork for a full environmental impact statement.

⁴² *Id.* at § 611(a).

⁴³ There are circumstances where it may be appropriate to publish an IRFA for the proposed rule, and based on comments received, publish a certification for the first time in the final rule. See Chapter 3 of this guide for a detailed discussion of final regulatory flexibility analyses.

⁴⁴ 5 U.S.C. § 605(b). The Office of Advocacy would expect this situation to be rare because agency efforts to develop the rule should include a reasonable effort to explore all the effects of the rule, including the

agency lacks sufficient information to make a certification decision, the agency should engage in reasonable outreach efforts.⁴⁵

Organizing the threshold report

Certification analysis discussed in this chapter does not require the depth of analysis necessary in an initial regulatory flexibility analysis,⁴⁶ as discussed in Chapter 2 of this guide. Nevertheless, this “threshold” analysis can offer important insights into the nature of regulatory impacts. Although a study of alternatives is not required at this stage, it often leads to the skeleton of regulatory alternatives that can reduce or eliminate any disproportionate impacts on small entities. For this reason, Advocacy encourages certification analysis as early in the rule development process as possible.

Agency certifications of final rules are subject to judicial review⁴⁷ and courts evaluate them by determining whether the statement of basis and purpose accompanying the rule identifies a “factual basis” to support the certification.⁴⁸ A helpful threshold report will directly support the elements that must appear in the *Federal Register* Notice of Proposed Rulemaking preamble. The Office of Advocacy believes the threshold analysis should discuss the following items:⁴⁹

- 1) Description of small entities affected
 - ✓ A brief economic and technical statement on the regulated community, describing some of the following types of information:⁵⁰
 - a) The diversity in size of regulated entities
 - b) Revenues in each size grouping
 - c) Profitability in each size grouping
- 2) Economic impacts on small entities
 - ✓ A fair, first estimate of expected cost impacts, or a reasonable basis for assuming costs would be *de minimis* or insignificant within all economic or size groupings of the “small” regulated community
 - ✓ The rationale for the certification decision, based on the analysis presented
- 3) Significant economic impact criteria
 - ✓ The criteria used to examine whether first-estimate costs are significant
- 4) Substantial number criteria

effects on small entities. For more information on preparing an initial regulatory flexibility analysis, see Chapter 2.

⁴⁵ *Id.* at § 609. Outreach is important to obtain information required by the RFA, to obtain relevant input from affected small entities. See Chapters 4 and 7 for a discussion of agency outreach to small entities.

⁴⁶ An initial regulatory flexibility analysis (IRFA) is a document containing the agency’s data and analysis regarding the potential impact of the proposed rule. A detailed description of the requirements of an IRFA can be found in Chapter 2 of this guide.

⁴⁷ 5 U.S.C. § 611.

⁴⁸ *Id.* at § 605(b).

⁴⁹ For additional detail, see the certification checklist at the end of this chapter.

⁵⁰ When an agency does not have quantitative data to support its certification, the agency should explain why such data are not available and request comments.

- ✓ The criteria used to examine whether the entities experiencing significant impacts constitute a substantial number of entities in any of the regulated size groupings
- 5) Description of assumptions and uncertainties
 - ✓ The sources of data used in the economic and technical analysis⁵¹
 - ✓ The degree of uncertainty in the cost estimates, when uncertainty is large
- 6) Certification statement

“Factual basis” requirement for certification

What is a “factual basis”? The Office of Advocacy interprets the “factual basis” requirement to mean that, at a minimum, a certification should contain a description of the number of affected entities and the size of the economic impacts and why either the number of entities or the size of the impacts justifies the certification.

The agency’s reasoning and assumptions underlying its certification should be explicit in order to elicit public comment. Certifications of “no significant economic impact on a substantial number of small entities” have major legal implications for agencies. Consequently, certifications that simply state that the agency has found that the proposed or final rule will not have a significant economic impact on a substantial number of small entities are not sufficient under section 605(b).

The “more than just a few” standard for determining if a rule will have a significant economic impact on a “substantial number of small entities” is a rigorous test for agencies to follow. However, the Office of Advocacy encourages a conservative approach.⁵² In other words, if an agency has set its standard for determining “substantial number” too high, the certification may give rise to court challenges that could have been avoidable.⁵³

Prior to the enactment of the SBREFA amendments in 1996, the RFA required only that a certification be supported by a “succinct statement explaining the reasons for the

⁵¹ Section 607 of the RFA directs agencies to provide a “quantifiable or numerical description of the effects of the proposed rule or alternatives to the proposed rule” and allows a qualitative approach if “quantification is not practical or reliable.” Thus, agencies are expected to make reasonable efforts to acquire quantitative or other information to support analysis of the rules under sections 603 and 604 of the RFA. Such a standard is not required for section 605 certifications, but some agencies use section 607 as a model for preparing certifications. With regard to certification analyses, EPA advises its rulewriters that “where the information necessary to conduct a quantitative analysis is not reasonably available, it may be appropriate to certify the rule based on the qualitative assessment alone.” Regulatory Management Division, EPA Office of Policy, *EPA’s Action Development Process: Final Guidance for EPA Rulewriters: Regulatory Flexibility Act, as amended by the Small Business Regulatory Enforcement Fairness Act* (November 2006), p. 20.

⁵² Five small firms in an industry with more than 1,000 small firms is not likely to be interpreted as a “substantial number”; on the other hand, the same five small firms in an industry with only 20 firms would be a substantial number. See the discussion of the definitions of “significant” and “substantial” later in this chapter.

⁵³ See Chapter 5 of this guide for information on what the courts have held in these types of cases.

certification,”⁵⁴ and since such statements were not subject to judicial review, even as part of the record on review, agencies could avoid substantive explanations by using boilerplate certifications. The amended version of the RFA now requires that certifications be supported by a “statement of factual basis.” In amending the RFA, Congress intended that agencies should do more than provide boilerplate and unsubstantiated statements to support their RFA certifications. Courts will overturn an agency’s final certification if it is not adequate.⁵⁵

What is the definition of a small entity?

The definition of “small entity” is important because it is the starting point for determining the degree of impact a regulation will have on small entities. Three types of small entities are defined in the RFA:⁵⁶

Small business. Section 601(3) of the RFA defines a “small business” as having the same meaning as “small business concern” under section 3 of the Small Business Act. This includes any firm that is “independently owned and operated” and is “not dominant in its field of operation.”⁵⁷ The Small Business Administration (SBA) has developed size standards to carry out the purposes of the Small Business Act and those size standards can be found in 13 C.F.R., section 121.201. The Small Business Act prohibits an agency from adopting a different definition of small business when promulgating regulations to carry out a delegation of authority from Congress unless the agency follows the procedures set forth in SBA’s regulations.⁵⁸ In addition, an agency may feel that the classification used by the Administrator for a particular sector is inappropriate in doing the analysis required by the RFA. The agency is then authorized to use a different definition, solely for purposes of complying with the RFA, after consultation with the Chief Counsel. That consultation does not obviate the need for the agency to comply with section 3 of the Small Business Act should the agency be interested in promulgating a regulation that utilizes a different definition of small business than that developed by the Administrator.⁵⁹

Small organization. Section 601(4) defines a small organization as any not-for-profit enterprise that is independently owned and operated and not dominant in its field (for example, private hospitals and educational institutions). Agencies may develop one or more alternative definitions of “small organization” for purposes of this chapter, provided that they: (1) give an opportunity for public comment and (2) publish the final definition in the *Federal Register*. However, an agency that decides a different definition is appropriate for purposes of complying with the RFA is required to follow the procedures set forth in section 601(4).

⁵⁴ See *Lehigh Valley Farmers, Inc., v. Block*, 640 F. Supp. (E.D. Pa. 1986), *aff’d on other grounds*, 828 F.2d.

⁵⁵ See *North Carolina Fisheries Association v. Daley*, 27 F. Supp. 2d 650 (E.D. Va. 1998).

⁵⁶ Appendix C lists data sources that may be helpful in drawing distinctions between large and small entities.

⁵⁷ 15 U.S.C. § 632.

⁵⁸ 13 C.F.R. § 121.902(b).

⁵⁹ *Northwest Mining Ass’n v. Babbitt*, 5 F. Supp. 2d 9 (D.D.C. 1998).

Small governmental jurisdiction. Section 601(5) defines small governmental jurisdictions as governments of cities, counties, towns, townships, villages, school districts, or special districts with a population of less than 50,000. Agencies may develop one or more alternative definitions for this term provided that they: (1) give opportunity for public comment, (2) base definitions on factors such as low population density and limited revenues, and (3) publish final definitions in the *Federal Register*. The alternative definition developed under this section applies only to the agency's compliance with the RFA. The agency may develop different size standards for small governmental jurisdictions in the development of its regulations. Any agency size standard determination that differs from the SBA's size standard is subject to judicial review.⁶⁰

Changing a size standard

It is important to draw a distinction when it comes to determining appropriate size standards. If an agency chooses to change a size standard after a determination that SBA's size standard is inadequate, the agency must either consult with the Office of Advocacy or seek approval of SBA's Administrator, depending on the circumstances.

For RFA analysis purposes, if an agency wants to use a different size standard, the agency can do so only after consultation with the Office of Advocacy and after an opportunity for public comment. In addition, that new size standard must be published in the *Federal Register*.

On the other hand, if an agency seeks to change the definition of a small business for rulemaking purposes, that is, for purposes of determining how to apply a regulation to a business of a certain size, the agency must seek approval from the SBA's Administrator.⁶¹

Assessing the impact on small entities

Determining a rule's impact on small entities is an important part of the rulemaking process. The RFA requires agencies to conduct sufficient analyses to measure and consider the regulatory impacts of the rule to determine whether there will be a significant economic impact on a substantial number of small entities. No single definition can apply to all rules, given the dynamics of the economy and changes that are constantly occurring in the structure of small-entity sectors.

Every rule is different. The level, scope, and complexity of analysis may vary significantly depending on the characteristics and composition of the industry or small-entity sectors to be regulated. This is why it is important that agencies make every effort

⁶⁰ 5 U.S.C. § 611(a); see also Chapter 5 of this guide for a discussion of how the courts have handled this issue.

⁶¹ Section 3(a)(2)(C)(i)-(ii) of the Small Business Act and SBA's regulations found in 13 CFR 121.902(b) essentially outline the information an agency needs to submit in order for SBA's Administrator to approve a new size standard, as well as when in the rulemaking process an agency needs to obtain that approval.

to conduct a sufficient and meaningful analysis when promulgating rules. The preparation of the required analysis calls for due diligence, knowledge of the regulated small entity community, sound economic and technical analysis, and good professional judgment.⁶² One of the first steps in the analytical process includes understanding the nature and economics of the industry/entities being regulated, and identifying how much each sector is contributing to the problem the agency is trying to address and mitigate. A goal of the entire APA/RFA process is to give the public a complete understanding of what the agency is doing. Small businesses cannot provide informed comments if the agency fails to identify the rule as one that will have a significant impact on a substantial number of small businesses. In turn, informed comments provide useful tools for the agency to construct the least burdensome, most effective regulations.

Because almost every industrial category will have more small than large businesses,⁶³ determining the impact on small businesses plays a key role in compliance with the RFA. In turn, to the extent that the costs of compliance are sufficiently significant that some entities will be unable to comply, the agency's selected regulatory solution probably will not achieve its statutory goal. Thus the analytical requirements, including the decision to certify, play a key role in the agency meeting its overall requirement of rational rulemaking, i.e., that the solution selected by the agency will achieve the agency's objectives.

As discussed in the previous section defining a small entity, it is important that agencies also examine the impact of their proposed regulations on small governmental jurisdictions. There are tens of thousands of these small jurisdictions throughout the United States that fall under the RFA's threshold of a population of less than 50,000. The growing demand for government services has far exceeded the financial capacities of many local governments, particularly the smallest ones, to provide those services while maintaining long-term fiscal viability. Costly federal regulations, both new and existing, often exacerbate an already difficult situation for many small communities. Like small businesses, small communities face economic challenges, lack the economies of scale, and in most cases have fewer technical and financial options available to them. All of these factors increase a small jurisdiction's cost to undertake and complete mandated regulatory initiatives.

Which segment of the economy or industry will be regulated?

To know whether a regulatory proposal affects a substantial number of small entities, the regulator must first know how many regulated entities exist and which are small. In examining this, the analyst best serves the process by identifying each group of regulated entities with similar economic and industrial characteristics. Each group constitutes its own universe of regulated small entities that the proposal may influence significantly. If the regulated community is segmented properly, the members of each group will have

⁶² See OMB's government-wide information guidelines, 67 Fed. Reg. 8452 (Feb. 22, 2002). These guidelines were issued under authority contained in the Information Quality Act, Pub. L. 106-554.

⁶³ This does not mean that small businesses dominate that sector of the market; for example, in telecommunications, although there are many small businesses, a handful of large regional telephone companies still dominate the market.

similar economic characteristics, and an examination of a typical entity or use of the group's mean characteristics will normally allow very rapid economic analysis for the group. This approach allows identification of those groups covered by the RFA.

Congress enacted the Small Business Regulatory Enforcement Fairness Act to achieve “fundamental changes . . . needed in the regulatory and enforcement culture of Federal agencies to make agencies more responsive to small business . . . without compromising the statutory missions of the agencies.”⁶⁴ Thus, to meet the basic SBREFA goal, analysts will routinely want to economically segment industrial sectors into several appropriate size categories smaller than the Small Business Act section 3 definition. Only by so doing will the analyst accurately identify and analyze those entities covered by the RFA where there is a large disparity in economic and industrial characteristics within the single category of small entities.⁶⁵ Consider the following example of how the SBA definition of a small business may not adequately address the nuances that exist within the universe of affected small entities:

SBA established a size standard for the drinking water supply industry at \$5 million in revenues, equating approximately to a city serving 30,000 people. EPA has proposed an alternative definition—a small water supply would serve no more than 10,000 people. Such a system generates somewhat less than a million dollars in annual revenue. However, EPA does not stop by looking only at the supply serving 10,000 people. It also examines sub-populations of the water supply industry serving fewer than 100 people, 101-500 people, 501-3,300 and 3,300-10,000. Water supplies in the smallest size category generate revenues less than one-tenth that of those in the 10,000-25,000 size category. More significantly, 90 percent of regulated water supplies serve fewer than 500 people, and on average, water supplies in those two size categories have net losses, costs being spread to other municipal revenue streams. EPA typically examines each of these small water supply size categories and, in keeping with the Regulatory Flexibility Act, has proposed different “available treatment technologies” for each water supply size, reflecting the wide range in economic viability within the industry. Each of the size categories below the “small water supply” size cut-off stands as its own universe of economically similar regulated entities. EPA recognized the regulatory significance of this and incorporated it into its analysis.⁶⁶

Agencies should identify and examine various economically similar small regulated entities so that they will have a baseline from which to determine whether a significant regulatory cost will have an impact on a substantial number of small entities. An understanding of the differences in economic impacts across the various regulated communities often generates different regulatory alternatives. A sound analysis requires

⁶⁴ SBREFA § 202(3).

⁶⁵ Conversely, if all small entities are equally affected by the proposed regulation, subcategorization is not required.

⁶⁶ For a full discussion of this issue, see EPA's National Primary Drinking Water Regulations: Arsenic Clarifications to Compliance and New Source Contaminants Monitoring, 66 Fed. Reg. 6976, 6987 (Jan. 22, 2001).

that agencies examine the various subsectors of the regulated community, the differences among them, and additional appropriate regulatory alternatives that can achieve the statutory mission while mitigating unnecessary economic impacts on small entities.

How to categorize small entity sectors

The agency's first step in a threshold analysis consists of identifying the industry, governmental and nonprofit sectors they intend to regulate. Using the North American Industry Classification System (NAICS) classifications, SBA defines small businesses in terms of firm revenues or employees.⁶⁷ Different criteria may be helpful to agencies in assessing the composition of a small entity sector. The IRS categorizes firm (corporation and partnership) size by assets. Industry associations apply some or all of these three criteria (revenues, employment, and/or assets) and often add to or replace them with their own technical criteria. In addition to SBA definitions, federal regulators may use any one or multiple criteria to identify their universes of small regulated entities.⁶⁸

Determination of "significant impact"

The agency's second step in a threshold analysis is to determine whether there is a significant economic impact on a substantial number of small entities. The RFA does not define "significant" or "substantial." In the absence of statutory specificity, what is "significant" will vary depending on the economics of the industry or sector to be regulated. The agency is in the best position to gauge the small entity impacts of its regulations.

Significance should not be viewed in absolute terms, but should be seen as relative to the size of the business, the size of the competitor's business, and the impact the regulation has on larger competitors. For example, a regulation may be significant solely because the disparity in impact on small entities may make it more difficult for them to compete in a particular sector of the economy than large businesses. This may relate to their ability to pass costs through to customers or to reduce the marginal cost of such a regulation to an insignificant element of their production functions.

One measure for determining economic impact is the percentage of revenue or percentage of gross revenues affected. For example, if the cost of implementing a particular rule represents 3 percent of the profits in a particular sector of the economy and the profit margin in that industry is 2 percent of gross revenues (an economic structure that occurs in the food marketing industry, where profits are often less than 2 percent), the implementation of the proposal would drive many businesses out of business (all except

⁶⁷ Effective January 1, 1997, the federal government, for statistical purposes, replaced the Standard Industrial Classification (SIC) system with NAICS. For purposes of small business size standards, SBA adopted the NAICS definitions for all industries effective October 1, 2000. NAICS made changes to the descriptions of many industry structures.

⁶⁸ The SBA definitions here are found in § 3(a)(2) of the Small Business Act and are not the RFA definitions referenced above. See <http://www.sba.gov/category/navigation-structure/contracting/contracting-officials/eligibility-size-standards>.

the ones that beat a 3 percent profit margin). That would be a significant economic impact.

However, the economic impact does not have to completely erase profit margins to be significant. For example, the implementation of a rule might reduce the ability of the firm to make future capital investment, thereby severely harming its competitive ability, particularly against larger firms. This scenario may occur in the telecommunications industry, where a regulatory regime that harms the ability of small companies to invest in needed capital will not put them out of business immediately, but over time may make it impossible for them to compete against companies with significantly larger capitalizations. The impact of that rule would then be significant for smaller telecommunications companies.

Other measures may be used; to illustrate, the impact could be significant if the cost of the proposed regulation (a) eliminates more than 10 percent of the businesses' profits; (b) exceeds 1 percent of the gross revenues of the entities in a particular sector or (c) exceeds 5 percent of the labor costs of the entities in the sector.

Some agencies have already developed criteria for determining whether a particular economic impact is significant. Standards must be flexible enough to work for the individual agency. The following examples are meant to be illustrative of different types of criteria that may be used. They are not meant to imply a standard, acceptable formula. Advocacy welcomes input from other agencies on their standards.

- The Department of Health and Human Services (HHS) has determined that a rule is significant if it would reduce revenues or raise costs of any class of affected entities by more than 3 to 5 percent within five years. This approach may work well for an agency, depending upon the circumstances. It becomes complex, however, in the attempt to apply a simple rule fairly to varied industries and regulatory schemes. A 2 percent reduction in revenues in one industrial category would be significant if the industry's profits are only 3 percent of revenues. More than 60 percent of small businesses do not claim a profit and do not pay taxes; therefore, an agency would not be able to apply a profit-based criterion to these firms.
- The EPA has prepared extensive guidance for its rulewriters concerning "significant economic impact" and "substantial number." With respect to small businesses, the agency advises that the offices compare the annualized costs as a percentage of sales ("sales test") to examine significant economic effect. For the same purpose, it also discusses alternative uses of a cash flow test and a profits test.⁶⁹

⁶⁹ U.S. Environmental Protection Agency, Office of Policy, Regulatory Management Division, EPA's Action Development Process: Final Guidance for EPA Rulewriters: Regulatory Flexibility Act as amended by the Small Business Regulatory Enforcement Fairness Act (November 2006), section 2.6.2, available at <http://www.epa.gov/rfa/documents/Guidance-RegFlexAct.pdf>.

Legislative history of “significant economic impact.” The absence of a particularized definition of either “significant” or “substantial” does not mean that Congress left the terms completely ambiguous or open to unreasonable interpretations. Thus, the Office of Advocacy relies on legislative history for general guidance in defining these terms.⁷⁰ With regard to the term “significant economic impact,” Congress said:

The term ‘significant economic impact’ is, of necessity, not an exact standard. Because of the diversity of both the community of small entities and of rules themselves, any more precise definition is virtually impossible and may be counterproductive. Any more specific definition would require preliminary work to determine whether the regulatory analysis must be prepared.⁷¹

Congress also stated that,

Agencies should not give a narrow reading to what constitutes a “significant economic impact” . . . a determination of significant economic effect is not limited to easily quantifiable costs.⁷²

Congress has identified several examples of “significant impact”: a rule that provides a strong disincentive to seek capital;⁷³ 175 staff hours per year for recordkeeping;⁷⁴ impacts greater than the \$500 fine (in 1980 dollars) imposed for noncompliance;⁷⁵ new capital requirements beyond the reach of the entity;⁷⁶ and any impact less cost-efficient than another reasonable regulatory alternative.⁷⁷ Note that even below these thresholds, impacts may be significant. Other, more specific examples are contained in the House of Representatives Report on the RFA.⁷⁸

⁷⁰ Admittedly, throughout this guide, references are made to “adverse” impacts and efforts to “mitigate” impacts. This, after all, is the primary concern of the law. Legislative history, however, makes it clear that Congress intended that regulatory flexibility analyses also address “beneficial” impacts. Therefore, an agency cannot certify a proposed rule if the economic impact will be significant but positive. If an agency so finds, it should conduct a regulatory flexibility analysis to determine if alternatives can enhance the economic benefits flowing to small entities. See discussion in this chapter on adverse versus beneficial impacts.

⁷¹ 126 Cong. Rec. S10,942 (Aug. 6, 1980).

⁷² *Id.* at S10,940.

⁷³ *Id.* at S10,938.

⁷⁴ *Id.*

⁷⁵ 126 Cong. Rec. H24,578 (Sept. 8, 1980).

⁷⁶ *Id.* at H24,593.

⁷⁷ *Id.* at H24,595.

⁷⁸ “A gas station owner spent 600 hours last year filling out just his federal reporting forms. An Idaho businessman paid a \$500 fine [in 1980 dollars] rather than fill out a federal form which was 63 feet long. A New Hampshire radio station paid \$26.23 in postage to mail its license renewal back to Washington. A dairy plant licensed by 250 local governments, 3 states, and 20 agencies had 47 inspections in one month. A butcher had one Federal agency tell him to put a grated floor in his shop one month and then the next month was told by another federal agency he could not have a grated floor. A company was forced out of the toy business because one of its main products was inadvertently placed on a federal ban list. An Oregon company with three small shops received Federal forms weighing 45 pounds.” 126 Cong. Rec. H8,467 (Sept. 8, 1980).

Determination of “substantial number”

The next step is to determine whether it is a substantial number of small entities for whom the rule has a significant economic impact. In this instance, the number may be a ratio or it may be a whole number. In some instances, a very small number of small businesses who would experience a significant economic impact can represent the entire universe of affected small businesses. However, if a very small number of small businesses represents a small fraction of the universe of affected small businesses, the agency can conclude that the number is not substantial.

For example, suppose a rule is expected to affect 20 small entities in a given category. The agency must determine, as best it can, how extensive the economic impact will be on those small entities. Suppose further that the agency can conclude that for five of these small entities, the impact will be significant. Is five a “substantial number” of the small entities affected? When a rule will have a significant economic impact on 25 percent of the small entities affected, this would be considered a substantial number.

Legislative history of “substantial number.” Legislative history also says that the term “substantial” is intended to mean a substantial number of entities within a particular economic or other activity⁷⁹ The intent of the RFA, therefore, was not to require that agencies find that a large number of the entire universe of small entities would be affected by a rule. Quantification of “substantial” may be industry- or rule-specific. However, it is very important that agencies use the broadest category, “more than just a few,” when initially reviewing a regulation before making the decision to certify or do an initial regulatory flexibility analysis. The goal at this stage of the process is to ensure that the broadest possible impacts are fully considered.

“Substantial number” depends on the number of regulated entities and the size of the regulated industry. The interpretation of the term “substantial number” is not likely to be five small firms in an industry with more than 1,000 small firms. On the other hand, it is important to recognize that five small firms in an industry with only 20 small firms would be a substantial number. Depending on the rule, the substantiality of the number of small businesses affected should be determined on an industry-specific basis and/or on the number of small businesses overall. For example, the Internal Revenue Service, when changing the tax deposit rules, would examine the entire universe of small businesses to see how many would be affected. On the other hand, a change by the Food and Drug Administration in the regulation of meat irradiators might affect only 15 firms, but that would be the entire industry.

As EPA explains in its guidance, “analysts should examine both the total number and percentage of regulated small entities experiencing significant economic impacts when determining whether a ‘substantial number’ of small entities may be significantly affected.”⁸⁰ In its guidance, EPA provides a matrix of different combinations of “significant economic impact” in terms of annual costs/sales and “substantial number”

⁷⁹ 126 Cong. Rec. S10,938 (Aug. 8, 1980) (Section-by-Section Analysis of the Regulatory Flexibility Act).

⁸⁰ 2006 EPA Final Guidance, section 2.7.2.

for its certification decision. The larger the economic impacts, the smaller the substantial number that would eliminate the basis for a certification of no impact.⁸¹ For example, for a group of 100-999 affected small entities, EPA presumes no significant effect on a substantial number, where the costs/sales are 1 percent or greater for one or more of the affected small entities and the percent of small entities experiencing a given impact is less than 20 percent of all regulated small entities. However, if the costs/sales were 3 percent or greater, the presumption would no longer apply where the percentage was less than 20 percent of all regulated entities, for the same number of regulated entities.

In calculating the percentage of small entities significantly regulated within a regulated industry for the purpose of making the certification determination, the agency should be careful to count in the denominator only the firms that are regulated by the rule. For example, a regulation of firms with concentrated animal feeding operations (CAFOs) should count only the farms with CAFO operations, and not all farms, when calculating the percentage of CAFOs with a given economic impact. If all farms are included in the denominator rather than all CAFOs, heavy regulation of a segment of the CAFOs would be much less likely to exceed a 1 percent or 3 percent cost/sales ratio. Thus the impact would be underestimated. This is a common mistake by agencies using percentage tests. EPA further explains that analysts should aggregate the impacts of entities of the same type (such as small businesses or small governments) in making this determination. In addition, EPA explains where the rule applies to more than one type of small entity, the impacts should be analyzed separately for each type of entity, using an economic measure appropriate to each type of entity.⁸²

Direct versus indirect impact

The courts have held that the RFA requires an agency to perform a regulatory flexibility analysis of small entity impacts only when a rule directly regulates small entities.

The primary case on the issue of direct versus indirect impacts for RFA purposes is *Mid-Tex Electric Cooperative, Inc., v. FERC (Mid-Tex)*.⁸³ In *Mid-Tex*, the Federal Energy Regulatory Commission (FERC) was proposing regulations affecting how generating utilities included construction work in progress in their rates. Generating utilities were large businesses, but their customers included numerous small entities, such as electric cooperatives. FERC authorized large electric utilities to pass these costs through to their transmitting and retail utility customers. This increased the cost to the transmitting utilities, which may or may not have been able (because of regulation by their rates commissions) to pass the costs on to their residential and business customers. These smaller utilities challenged the rule, asserting that the impact on them should have been considered. The court concluded that an agency may certify the rule pursuant to section 605(b) when it determines that the rule will not have a direct impact on small entities.⁸⁴

⁸¹ 2006 EPA Final Guidance, section 2.7.1, table 2.

⁸² 2006 EPA Final Guidance, section 2.7.3.2.

⁸³ *Mid-Tex Elec. Coop v. FERC*, 773 F.2d 327 (D.C. Cir. 1985).

⁸⁴ *Id.* at 342.

The U.S. Court of Appeals for the District of Columbia applied the holding of the *Mid-Tex* case in *American Trucking Associations, Inc., v. EPA*⁸⁵ (hereafter *ATA*). In the *ATA* case, EPA established a primary national ambient air quality standard (NAAQS) for ozone. The basis of the EPA's certification was that the NAAQS regulated small entities indirectly through state implementation plans. The plans impose requirements on the small entities, whereas states are required to take action to attain compliance with the NAAQS standards. The court found that since the states, not EPA, had the direct authority to impose the burden on small entities, EPA's regulation did not have a direct impact on small entities.

Although it is not required by the RFA, the Office of Advocacy believes that it is good public policy for the agency to perform a regulatory flexibility analysis even when the impacts of its regulation are indirect. An agency should examine the reasonably foreseeable effects on small entities that purchase products or services from, sell products or services to, or otherwise conduct business with entities directly regulated by the rule. In the case of the NAAQS standard at issue in *ATA*, EPA had to estimate the impacts of the proposed rules on small entities in order to comply with the mandate of E.O. 12,866. Therefore, the agency could have examined alternatives that would have been less burdensome on small entities (and is required to under the E.O. 12,866). If an agency can accomplish its statutory mission in a more cost-effective manner, the Office of Advocacy believes that it is good public policy to do so. The only way an agency can determine this is if it does not certify regulations that it knows will have a significant impact on small entities, even if the small entities are regulated by a delegation of authority from the federal agency to some other governing body.⁸⁶

Adverse versus beneficial impact

Congress considered the term “significant” to be neutral with respect to whether the impact is beneficial or harmful to small businesses. Therefore, agencies need to consider both beneficial and adverse impacts in an analysis. The RFA legislative history has explicit insights into congressional intent with respect to beneficial impacts:

Agencies may undertake initiatives which would directly benefit such small entities. Thus, the term ‘significant economic impact’ is neutral with respect to whether such impact is beneficial or adverse. The statute is designed not only to avoid harm to small entities but also to promote the growth and well-being of such entities.⁸⁷

Moreover, early drafts of the RFA used the term “significant adverse” impact, but the final bill used only the term “significant impact.”⁸⁸

⁸⁵ *American Trucking Ass'ns v. EPA*, 175 F.3d 1027, 1044 (D.C. Cir. 1999), *aff'd in part and rev'd in part on other grounds*, *Whitman v. American Trucking Ass'ns*, 531 I/S/ 457 (2001).

⁸⁶ See Chapter 5 of this guide for a more detailed discussion of the direct versus indirect impact issue.

⁸⁷ 126 Cong. Rec. H8,468 (daily ed. Sept. 8, 1980).

⁸⁸ See an early draft of the RFA, S2147, 1st Sess. (1979).

Courts have applied definitions for “significant impact” in cases involving other statutes. For example, in a case involving the National Environmental Policy Act (NEPA), *Friends of Fiery Gizzard v. Farmers Home Administration*,⁸⁹ the court held that a full environmental impact statement (EIS) does not need to be prepared if the only impact of the project will be beneficial. However, the court acknowledged that when both negative and beneficial effects are present, an EIS must be prepared even if the agency feels that the beneficial effects outweigh the negative ones.⁹⁰ (This case does not say that beneficial impacts should not be considered for the preliminary assessment, nor does it say that beneficial impacts are never a factor.) Earlier cases interpreting NEPA held that beneficial impacts should be a consideration in the rulemaking process.⁹¹

Several agencies have taken issue with the Office of Advocacy’s interpretation of significant economic impact. However, the Office of Advocacy believes that its interpretation is consistent with the legislative history and overall purposes of the RFA. The Office of Advocacy does not dispute that the RFA intends for agencies to “minimize the significant economic impact.”⁹² However, the Office of Advocacy’s interpretation does not necessarily mean that agencies should minimize beneficial impacts—that certainly would be contrary to the purposes of the RFA. Instead, Advocacy believes that it is often possible to analyze beneficial impacts with minimal effort and without necessarily triggering the need for an IRFA. Moreover, analyzing beneficial impacts lends credibility to the alternatives selected by the agency.

Once the certification decision is made, the agency must notify the Office of Advocacy and publish its certification in the *Federal Register*. It is good regulatory practice to get the notice to Advocacy as soon as possible. It has been useful to the agency to share a draft certification statement with Advocacy for confidential feedback on the adequacy of the statement. At a minimum, the notification should come at the same time as publication. Publication of a proposal alone can work for most certified regulations, but there will always be those proposals for which solid community comments in advance can be vitally important (e.g., through an advance notice of proposed rulemaking).

What adequate and inadequate certifications look like

Refer to the certification checklist at the end of this chapter (Table 1) for a review of the elements of a certification that meets all requirements.

⁸⁹ *Friends of Fiery Gizzard v. Farmers Home Admin.*, 61 F.3d 501, 505 (6th Cir. 1995).

⁹⁰ *Id.* at 505.

⁹¹ See *Hiram Clarke Civic Club v. Lynn*, 476 F.2d 421, 426-27 (5th Cir. 1973), (Considering only negative impacts “raises serious questions about the adequacy of the investigatory basis underlying the HUD decision not to file an EIS.”); *Environmental Defense Fund v. Marsh*, 651 F.2d 983, 993 (5th Cir. 1981), stating “[A] beneficial impact must nevertheless be discussed in an EIS, so long as it’s significant. NEPA is concerned with all significant environmental effects, not merely adverse ones.”

⁹² 5 U.S.C. Chapter 6, Congressional Findings and Declaration of Purpose.

An example of an adequate certification

The following example of an adequate certification by the U.S. Small Business Administration is from the proposed rule on small business investment companies.

When an agency issues a rulemaking proposal, the Regulatory Flexibility Act (RFA) requires the agency to “prepare and make available for public comment an initial regulatory flexibility analysis” which will “describe the impact of the proposed rule on small entities.” (5 U.S.C. §. 603(a)). Section 605 of the RFA allows an agency to certify a rule, in lieu of preparing an analysis, if the proposed rulemaking is not expected to have a significant economic impact on a substantial number of small entities.

This proposed rule directly affects all SBICs, of which there are currently 432. SBA estimates that approximately 75 percent of these SBICs are small entities. Therefore, SBA has determined that this proposed rule will have an impact on a substantial number of small entities.

However, SBA has determined that the impact on entities affected by the proposed rule will not be significant. The effect of the proposed rule will be to allow SBICs the flexibility to choose the optimal structure for their investments without having to notify or seek approval from SBA. SBA expects the impact of the proposed rule will be a reduction in the paperwork burden for SBICs. SBA asserts that the economic impact of the reduction in paperwork, if any, will be minimal and entirely beneficial to small SBICs. Accordingly, the Administrator of the SBA hereby certifies that this rule will not have a significant economic impact on a substantial number of small entities. SBA invites comment from members of the public who believe there will be a significant impact either on SBICs, or on companies that receive funding from SBICs.⁹³

Examples of inadequate certifications

Following are examples of inadequate certifications that were effectively challenged and refuted through formal comments to the agency or through the courts.⁹⁴

Shark Protection. *Southern Offshore Fishing Association v. Daley*⁹⁵ offers a landmark legal decision recognizing the failure of an agency to adequately examine the market to determine whether there was a significant impact on a substantial number of small entities. On December 20, 1996, the National Marine Fisheries Service (NMFS), National Oceanic and Atmospheric Administration (NOAA) published the proposed rule for the

⁹³ 67 Fed. Reg. 35,055, at 35,056 (May 17, 2002). Note that although this certification addressed beneficial impacts, the agency acknowledged that even those impacts would be minimal and therefore correctly certified the rule.

⁹⁴ For another example of an improper certification, see Chapter 5 under the discussion of *North Carolina Fisheries v. Daley*.

⁹⁵ *Southern Offshore Fishing Ass’n v. Daley*, 995 F. Supp. 1411 (M.S. Fla. 1998).

Atlantic Shark Fisheries: Quotas, Bag Limits, Prohibitions, and Requirements.⁹⁶ The proposed rule, among other things, reduced the commercial quotas for sharks by 50 percent. NMFS prepared a certification in lieu of an IRFA for the proposal. As the basis for the certification NMFS stated, in part:

Reducing the commercial quota is not expected to have a significant impact on a substantial number of small entities primarily because of the large degree of diversification in fishing operations that exist in the fleet and the already short shark fishing season, as outlined in the Regulatory Impact Review.

Advocacy submitted comments asserting that the certification was inappropriate. In its comments, Advocacy pointed out that under NMFS's own criteria for assessing regulatory impact, the proposal would have a significant economic impact on a substantial number of small entities.⁹⁷ NMFS's regulatory impact review stated that the majority of the participants in the fishing industry are small businesses and that there were 326 fisherman, 134 of which qualified for direct permits in the shark fishery. Approximately 41 percent of the shark fishery consisted of fishermen who only fished for sharks. The remaining fishermen were pelagic longline fishermen who also primarily fished for tuna and swordfish. Advocacy, therefore, concluded that the rule would have an impact on a substantial number of small entities.

In terms of significant economic impact, the Office of Advocacy argued that it was logical to infer that a 50 percent reduction in catch would result in a loss in revenue of at least 5 percent. The Office of Advocacy supported its inference with information obtained from fishery associations. For example, the Directed Shark Fishery Association asserted that the majority of the 134 directed shark vessels would lose more than 20 percent of their income. Some were expected to lose as much as 50 percent of their income. Similarly, the North Carolina Fisheries Association contended that more than 20 percent of their full-time shark fishermen would go out of business as a result the proposed rulemaking. Accordingly, Advocacy concluded that by the criteria set forth by NMFS, the impact of the proposed rulemaking would be significant.

Advocacy also presented information that indicated that NMFS's assumption that the affected industries would diversify was not realistic. Advocacy asserted that the cost of converting to another fishery could range from \$3,000 to \$25,000 per boat, depending on the vessel. At that time, Advocacy's statistics indicated that the average gross revenue of a sole fisherman was \$139,000 per year. Obtaining the equipment necessary to diversify

⁹⁶ 61 Fed. Reg. 67,295.

⁹⁷ At that time, NMFS criteria provided that a rule had a significant impact on a substantial number of small entities if 20 percent of those engaged in the fishery had either a reduction in gross revenues by more than 5 percent, an increase in total costs of production by more than 5 percent, or a 10 percent increase in compliance costs; or if 2 percent of small business entities were forced to cease business operations. NMFS no longer uses these criteria. Advocacy was pleased with NMFS's decision to abandon these criteria and institute new guidelines for determining economic impact on the fishing industry.

could amount to approximately 18 percent of the business's gross revenues, which would also be a significant economic impact.⁹⁸

The members of the fishing industry successfully challenged NMFS's RFA compliance in *Southern Offshore Fishing Association v. Daley*.⁹⁹ The court found that the agency certified without making a "reasonable, good-faith effort," prior to issuance of the final rule, to inform the public about the potential adverse effects of its proposals and about less harmful alternatives.

Telecommunications System Construction and Specifications. In another case, the Rural Utilities Service (RUS) certified that the final rule did not have a significant economic impact on a substantial number of small entities because small entities were not subject to any requirements that were not applied equally to large entities. While the rule did subject all entities to the same regulation, this justification ignored the disproportionate impact regulations often have on small businesses. In addition, RUS was depriving itself of the opportunity to learn about the rule's impact on small businesses. The Office of Advocacy filed the following comment with the RUS:

Congress knew about the tendency of agencies to impose "one-size-fits-all" regulations and specifically rejected it. As Congress states, one-size-fits-all regulations are unnecessary and disproportionately burdensome to small businesses. . . . Because of the disparity of the impact of governmental regulations, the agency cannot certify a rule on the basis that all entities have the same regulatory obligations.¹⁰⁰

Offshore Oil and Gas Well Operations. One of the responsibilities of the Minerals Management Service (MMS) of the Department of the Interior is to ensure safety in offshore oil and gas well operations. In February 1998, MMS proposed a rule to update and clarify MMS regulations on postlease operations. MMS prepared a certification in lieu of an IRFA for the proposal. As a basis for the certification, MMS stated:

In general, a company needs large technical and financial resources and experience to safely conduct offshore activities. However, many of the leases and operators have less than 500 employees and are small businesses. It is likely that a State lessee applying for a right-of-use and easement on the OCS may be a small business. The costs associated with obtaining the benefit (right of use and easement) would be minimal. The application fee is estimated to be \$2,350 per application and the rental is estimated to be \$5,000.

⁹⁸ U.S. Small Business Administration, Office of Advocacy comment letter to NMFS and NOAA dated February 6, 1997. See <http://archive.sba.gov/advo/laws/comments/noaa2-6.html>.

⁹⁹ *Southern Offshore Fishing*. This case is also discussed in Chapter 5 of this guide.

¹⁰⁰ See http://archive.sba.gov/advo/laws/comments/rus02_0308.pdf.

Advocacy submitted comments¹⁰¹ asserting that the certification was based on generalizations and unsubstantiated assumptions. In its comments, Advocacy identified databases and a means for a threshold analysis to help determine whether the agency should have certified, finding that the MMS had not provided sufficient information to document a rational basis for its decision to certify the rule. Advocacy stated:

For the purposes of its analysis, the Office of Advocacy referred to SIC 1381, Drilling Oil and Gas Wells. While Advocacy acknowledges that SIC 1381 may include more than drilling on the outer Continental Shelf, Advocacy submits the numbers for the sake of argument in an effort to point out the inherent weaknesses in MMS's certification.

According to this SIC data, there are a total of 1,380 firms that drill oil and gas wells. Of that 1,380 firms, 1,341 or 97% qualify as small firms in that they have fewer than 500 employees; 654 firms have 1-4 employees. The 654 firms constitute 47 percent of all firms large and small. Needless to say, 47 percent of an industry represents a substantial number of firms and suggests that certification of this rulemaking may be improper.

In the 1-4 employee sector, the estimated receipts for a firm are \$46,774, with an annual payroll of \$32,187. The estimated cost of the proposed rule is \$7,350 (\$2,350 per application and \$5,000 for the rental) per year. The \$7,350 amounts to approximately 16 percent of the annual receipts for that sector. Although there are no hard rules for defining significant economic impact on a substantial number of small entities, a proposal that will impose on 47 percent of an industry an additional cost of 16 percent of annual receipts should at least raise a warning sign for a regulatory agency that the proposal could interfere with profits and company survival. It should also indicate to the agency that certification may be improper under the RFA.

¹⁰¹ It should be noted that in the comments, Advocacy also commended MMS for the improvement that it made in its certification process. Instead of an unsupported allegation of no significant economic impact on a substantial number of small entities, MMS did provide a basis for the certification. MMS has continued to work with Advocacy to improve its RFA compliance.

Table 1. Certification checklist

<p>1. Request for comment on proposed rules</p>	<p>Look for:</p> <ul style="list-style-type: none"> √ A request for comment on the certification; and, √ A request for comment on the threshold economic analysis and its underlying assumptions.
<p>2. Description and estimate of number of small entities to which the rule applies</p>	<p>Look for:</p> <ul style="list-style-type: none"> √ The North American Industry Classification System (NAICS codes) categories for those entities subject to the regulation; √ A breakdown of each industry by several entity sizes, which should include the SBA size standard for each industry; √ Any alternative operational size definition used to tier requirements under the rule; √ For each size category in each industry, information on revenues, profit or other measures of economic sustainability.
<p>3. Estimate of economic impacts on small entities</p>	<p>Look for:</p> <ul style="list-style-type: none"> √ A set of tables, charts and discussion for a typical entity in each size category in each industry; √ Estimates of the cost impacts of the proposal; √ Estimates of the beneficial impacts of the proposal.
<p>4. Criteria for “significant economic impacts”</p>	<p>The best analyses will not use a preset criterion, but instead will examine one or more of the following:</p> <ul style="list-style-type: none"> √ <i>Long-term insolvency</i>, measured as regulatory costs significantly reducing typical profits for the size category; √ <i>Short-term insolvency</i>, measured as increased operating expenses or new debt larger than cash reserves and cash flow can support, causing nonmarginal firms to close; √ <i>Disproportionality</i>, based on whether regulations place small entities at a significant competitive disadvantage; √ <i>Inefficiency</i> based on whether the social costs imposed on small entities outweigh the social benefits of regulating them. <p>Look for a cogent explanation underlying any conclusionary statements about preset “criteria.”</p>

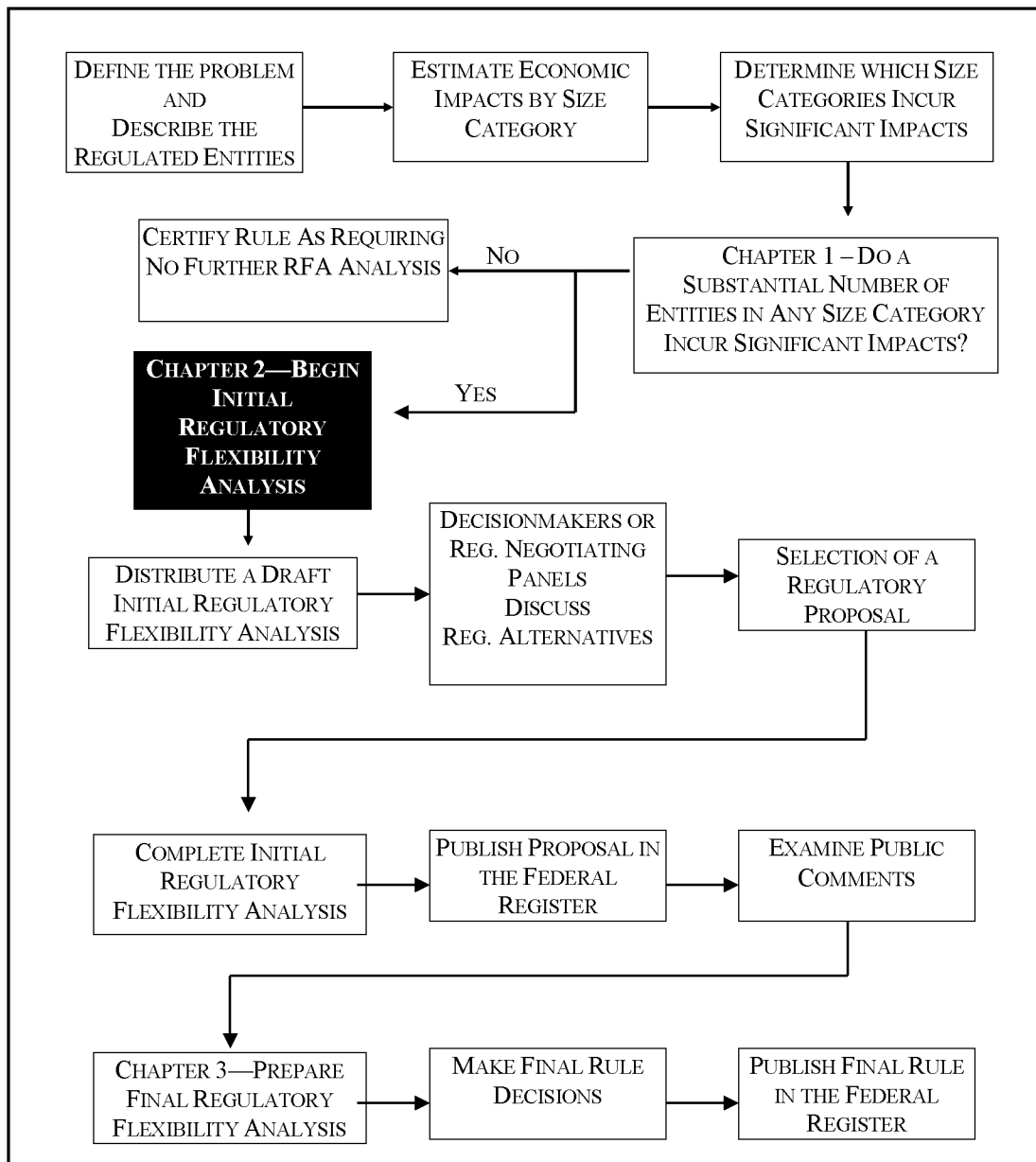
Continued on following page.

Table 1, Certification Checklist, *continued.*

<p>5. Criteria for “substantial number”</p>	<p>Look for:</p> <ul style="list-style-type: none"> √ The North American Industry Classification System (NAICS codes) of those regulated; √ A stratification of each industry by size, which should include the SBA size standard for each industry; √ Any alternative operational size definition used to tier requirements under the rule; √ Description of size categories demonstrating all entities within the category share similar economic characteristics; √ Whether a ‘percentage of entities significantly affected’ approach is used; √ Whether a ‘minimum number’ approach is used. (This is usually arbitrary and probably capricious); √ Justification of whatever criterion is used. <p>Typically, if an industry is properly segmented, analysis of a typical entity within the segment will indicate whether most or few will be significantly affected, as all within the segment should have similar economic characteristics.</p>
<p>6. Examination of industry segments with significant economic impacts</p>	<p>Look for:</p> <ul style="list-style-type: none"> √ An estimate of how many segments within an industry will experience significant impacts: if even one significant segment will, an IRFA is needed; √ An estimate of entities experiencing significant impacts. Other entities with similar economic characteristics should also be experiencing adverse impacts, and finding any with such adversely impacts tends to imply there is a segment that deserves special attention. The resulting IRFA should materially address the problems in that segment, recognizing the rest have few, if any impacts.
<p>7. Disclosure of assumptions</p>	<p>Look for:</p> <ul style="list-style-type: none"> √ A discussion on how sensitive underlying assumptions are to conclusions on whether there is no significant economic impact on a substantial number of small entities; √ A discussion on the uncertainty associated with the most significant underlying assumptions; √ A presentation on the range of potential findings, as reflects the underlying uncertainty in assumptions.
<p>8. Certification statement by the head of the agency</p>	<p>Look for:</p> <ul style="list-style-type: none"> √ A finding under 5 U.S.C. § 605, the Regulatory Flexibility Act, that “the proposed rule, if promulgated, will not have a significant economic impact on a substantial number of small entities.”

CHAPTER 2 PREPARING A PROPOSED RULE: THE INITIAL REGULATORY FLEXIBILITY ANALYSIS

Figure 2. The RFA decision process: IRFA



During the preparation of a proposed rule, an agency must prepare an initial regulatory flexibility analysis (IRFA) if it determines that a proposal may impose a significant economic impact on a substantial number of small entities.¹⁰² (If the agency determines that the proposed rule does not have such an impact, it should certify the rule as discussed in Chapter 1 of this guide.)

The RFA requires agencies to publish the IRFA, or a summary thereof, in the *Federal Register* at the same time it publishes the proposed rulemaking.¹⁰³ The IRFA must include a discussion of each element required by section 603 of the RFA, and the agency must also send a copy of the IRFA to the Chief Counsel for Advocacy.¹⁰⁴ Agencies are required to notify Advocacy when they submit a draft proposed or final rule to the Office of Information and Regulatory Affairs (OIRA) under Executive Order 12,866, or at a reasonable time prior to publication of the rule by the agency.¹⁰⁵ Moreover, the earlier a copy of the IRFA is provided to Advocacy, the more opportunity exists for constructive involvement and feedback to the agency. If an agency is preparing a series of closely related rules, it may, to avoid duplicative action, consider them one rule for the purposes of complying with the IRFA requirement.¹⁰⁶

Issues to be addressed in the analysis

Section 603 of the RFA requires agencies to perform a detailed analysis of the potential impact of the proposed rule on small entities.¹⁰⁷ In order to perform this analysis, an agency must enumerate the objectives and goals of the rule, as well any additional reasons the agency is pursuing the rule.

The agency then must examine the costs and other economic implications for the industry sectors targeted by the rule. When such data are unavailable, the agency should state why and request comments. Impacts include costs of compliance and economic implications that derive from additional compliance costs such as economic viability (including closure), competitiveness, productivity, and employment. The analysis should identify cost burdens for the industry sector and for the individual small entities affected. Costs might include engineering and hardware acquisition, maintenance and operation, employee skill and training, administrative practices (including recordkeeping and reporting), productivity, and promotion. The agency must also consider alternatives to the proposed regulation that would accomplish the agency's goals while not disproportionately burdening small businesses. As part of the discussion of the alternatives under section 603(c), it is recommended that the agency address, the costs, benefits, and other economic implications.

¹⁰² For a full discussion of "significant economic impact on a substantial number of small entities," and the requirements of a proper certification statement, see Chapter 1 of this guide.

¹⁰³ 5 U.S.C. § 603(a).

¹⁰⁴ *Id.*

¹⁰⁵ Exec. Order No. 13,272, § 3(b).

¹⁰⁶ 5 U.S.C. § 605(c).

¹⁰⁷ *Id.* at § 603(b)-(c).

Some of the important questions the agency should address in preparing an IRFA are:

- Should the agency redefine “small entity” for purposes of the IRFA?
- Which small entities are affected the most? Are all small entities in an industry affected equally or do some experience disparate impacts such that aggregation of the industry would dilute the magnitude of the economic effect on specific subgroups?¹⁰⁸
- Are all the required elements of an IRFA present, including a clear explanation of the need for and objectives of the rule?¹⁰⁹
- Has the agency identified and analyzed all major cost factors?
- Has the agency identified all significant alternatives that would allow the agency to accomplish its regulatory objectives while minimizing the adverse impact or maximizing the benefits to small entities?
- Can the agency use other statutorily required analyses to supplement or satisfy the IRFA requirements of the RFA?
- Are there circumstances under which preparation of an IRFA may be waived or delayed?
- What portion of the problem is attributable to small businesses (i.e., is regulation of small businesses needed to satisfy the statutory objectives)?
- Does the proposed solution meet the statutory objectives in a more cost-effective or cost-beneficial manner than any of the alternatives considered?

The results of the analysis should allow interested parties to compare the impacts of regulatory alternatives on the differing sizes and types of entities affected by the rule. It will enable direct comparison of small and large entities to determine the degree to which the alternatives chosen disproportionately affect small entities or a specific subset of small entities. Further, the analysis will examine whether the alternatives are effectively designed to achieve the statutory objectives.

The agency must balance the thoroughness of an analysis and practical limits of an agency's capacity to carry out the analysis with the significance of the rule and the expected economic impacts. Agencies should consult available information on how to conduct an economic analysis, such as the guidelines in OMB's *Economic Analysis of Federal Regulations under Executive Order 12,866* and should review small business data, including data referenced in Appendixes B and C.

If economic data are available, an agency should utilize the data in preparing an IRFA. When data are not readily available, the agency should consult with industry sources or other third parties to collect data. If the data collection is inadequate, then agencies should solicit the data as part of the proposed rulemaking.

¹⁰⁸ See discussion of this issue in Chapter 1.

¹⁰⁹ An agency may want to avoid repeating relevant text by cross-referencing the needs and objectives of the rule in its IRFA.

Elements of an IRFA

The preparation of an IRFA should be coordinated with the development of the data and analysis the agency will use in preparing the proposed rule under the requirements of the Administrative Procedure Act. In doing so, the agency should be mindful of the requirements of the RFA and collect data based on size. The development of a rational rule will require the acquisition of data that describe the scope of the problem, the entities affected, and the extent of those effects on the entities and the problem being addressed. Without such information, the agency will be unable to develop a rational rule.¹¹⁰

Under section 603(b) of the RFA, an IRFA must describe the impact of the proposed rule on small entities and contain the following information:

1. A description of the reasons why the action by the agency is being considered.
2. A succinct statement of the objectives of, and legal basis for, the proposed rule.
3. A description—and, where feasible, an estimate of the number—of small entities to which the proposed rule will apply.
4. A description of the projected reporting, recordkeeping, and other compliance requirements of the proposed rule, including an estimate of the classes of small entities that will be subject to the requirement and the types of professional skills necessary for preparation of the report or record.
5. An identification, to the extent practicable, of all relevant federal rules that may duplicate, overlap, or conflict with the proposed rule.

Section 603(c) is the key provision of the IRFA. It requires an agency to include a description of any significant alternatives to the proposed rule that minimize significant economic impacts on small entities while accomplishing the agency's objectives. The approach an agency takes while developing an IRFA depends on such factors as the quality and quantity of available information and the anticipated severity of a rule's impacts on small entities subject to the rule and the benefits yielded by each significant alternative. Section 607 of the RFA requires agencies to develop a quantitative analysis of the effects of a rule and its alternatives using available data. If quantification is not practicable or reliable, agencies may provide general descriptive statements regarding the rule's effects.¹¹¹ This second option is a last resort when it is not practicable for the agency to complete a significant quantitative analysis.

The new Section 603(d) of the RFA requires the CFPB to include a description of any projected increase in the cost of credit for small entities, as well as a description of significant alternatives which, while accomplishing the stated objectives, minimize any such increase, and the advice and recommendations of small entities with respect to these cost-of-credit issues.¹¹² The CFPB is required to identify small entity representatives in

¹¹⁰ *Bowen v. AHA*, 476 U.S. 610, 643 (1986); *National Ass'n of Home Health Agencies v. Schweiker*, 690 F.2d 932, 949 (D.C. Cir. 1982); *Chocolate Mfrs. Ass'n v. Block*, 755 F.2d 1098, 1103 (4th Cir. 1985).

¹¹¹ 5 U.S.C. § 607.

¹¹² 5 U.S.C. § 603(d), added to the RFA by the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010, Pub. Law 111-203 § 1100G(d)(1). This law also added the CFPB to the list of covered agencies—previously EPA and OSHA—that are required to hold small business review panels.

consultation with the Office of Advocacy and collect advice and recommendations about these cost-of-credit issues in addition to the issues raised by the proposed regulation.

The principal issues an agency should address in an IRFA are the impact of a proposed rule on small entities and the comparative effectiveness (benefits) and costs of alternative regulatory options. Each of the specific elements of the IRFA is discussed in turn below.

Reasons action is being considered

For the first element of the IRFA, the agency must discuss the reasons it is considering the proposed rule.¹¹³ The agency should list any issue to be addressed in the rulemaking and should be thorough in listing its reasons, as this section provides insight into the need for the rule.

Generally, the agency addresses this topic in the preamble to the proposed rule. The agency can summarize its discussion in the rulemaking, if the rulemaking addresses all the reasons the agency is considering the action. The discussion of the reasons leads directly into the objectives of the rule, the next element of the IRFA.

Objectives of the proposed rule

For the second element of the IRFA, the agency must list the objectives of the proposed rule.¹¹⁴ Again, the agency should be thorough when discussing its objectives, as this discussion conveys to the public the goals of the rulemaking and why the agency is taking specific actions contained within the proposed rule. This section provides the justification for the agency's actions, balancing the burden of the compliance requirements against the need for the rule. Such a discussion should include how the rule is achieving the statutory objectives. Compliance with this requirement should not be difficult since agencies are required to explain their proposed actions and the reasons underlying those proposed actions in order to elicit comment from the public as required by section 553 of the APA.¹¹⁵

As with the reasons for the proposed rule, the agency is likely to have addressed this topic in the rulemaking. The agency can draw from the language of the rulemaking to satisfy this section of the IRFA, as long as it lists all the objectives of the proposed rule that would entail compliance requirements with a significant economic impact on a substantial number of small entities.

¹¹³ *Id.* at § 603(b)(1).

¹¹⁴ *Id.* at § 603(b)(2).

¹¹⁵ See *Spartan Radiocasting v. FCC*, 619 F.2d 314, 321 (4th Cir. 1980).

Description and estimate of the number of small entities

The third element of the IRFA requires the agency to identify the classes of small entities affected by the proposed rule and provide an estimate of the number of small entities in each of those classes.¹¹⁶ In particular, the agency should pay special attention to small entities expected to face disproportionate impacts relative to other entities in the industry, whether those entities are large or small. Classification requires the development of a profile for the affected industry or industries and categorization by various size classes within each affected industry. It is crucial that the agency list all industry classes affected by the rule. Specifically, if the agency imposes a compliance requirement on a class of small entities, it must identify that class of small entities in this section of the IRFA. As a default, section 601 of the RFA requires agencies to use size standards set by the SBA in determining whether businesses are small businesses. SBA's Office of Size Standards set these standards using NAICS.¹¹⁷ Agencies must identify each of the affected classes according to their NAICS code. Once the agency has identified all the affected industries by code, it can use the NAICS code in combination with the U.S. Census data¹¹⁸ to gain an estimate of the number of entities in each class. To help agencies with this element of the IRFA, the Office of Advocacy provides a listing of NAICS codes along with links to the U.S. Census data for each class on its web page.¹¹⁹

If the agency determines that the existing SBA size standards for small businesses are not appropriate for RFA analysis purposes, the RFA permits the agency, after notice and comment, to establish one or more alternative definitions of a small entity that are appropriate for the rule.¹²⁰ The RFA requires an agency to consult with the Office of Advocacy when performing an RFA analysis using a different small business size standard than that provided by the SBA.¹²¹

Estimating compliance requirements

For the fourth element of the IRFA, the agency must describe and estimate the compliance requirements of the proposed rule.¹²² This is one of the two most important elements in the IRFA, because the alternatives the agency examines in the IRFA will be designed to minimize these compliance burdens. Provision of a list in the IRFA enables small entities to more easily identify potential burdens and tailor their comments in the rulemaking process to those burdens that most affect them without wading through many *Federal Register* pages.

As stated by the RFA, some of the costs the agency must describe in the IRFA include the costs of any recordkeeping; professional expertise, such as lawyer, accountant, or engineering, needed to comply with recordkeeping; and reporting requirements. Section

¹¹⁶ 5 U.S.C. § 603(b)(3).

¹¹⁷ See <http://www.sba.gov/size/>.

¹¹⁸ See <http://www.census.gov/>.

¹¹⁹ Office of Advocacy, Research and Statistics, <http://www.sba.gov/advocacy>.

¹²⁰ See the size standard discussion in Chapter 1.

¹²¹ 5 U.S.C. § 601(3).

¹²² *Id.* at § 603(b)(4).

603 also requires that the agencies examine other compliance requirements, which may include, for example, the following: (a) capital costs for equipment needed to meet the regulatory requirements; (b) costs of modifying existing processes and procedures to comply with the proposed rule; (c) lost sales and profits resulting from the proposed rule; (d) changes in market competition as a result of the proposed rule and its impact on small entities or specific submarkets of small entities; (e) extra costs associated with the payment of taxes or fees associated with the proposed rule; and (f) hiring employees dedicated to compliance with regulatory requirements.

Since all rules are different and impose different compliance requirements, the RFA contemplates that agencies will prepare analyses to determine all significant long- and short-term compliance costs. Agencies should list the compliance requirements separately to provide greater transparency.

The IRFA should also, to the extent practicable, compare the costs of compliance for small and large entities to determine whether the proposed rule affects small entities disproportionately, to analyze the ability of small entities to pass on these costs in the form of price increases or user fees, and to assess the effects on firms' profitability or their ability to provide services. This should be done in conjunction with an estimation of the costs of compliance relative to changes in market structure and the competitive status of various subclasses of small entities as well as the competitive positions of small entities in comparison with larger entities.¹²³

Significant alternatives considered

The keystone of the IRFA is the description of any significant alternatives to the proposed rule that accomplish the stated objectives of applicable statutes and that minimize the rule's economic impact on small entities.¹²⁴ The development and adoption of these alternatives provide regulatory relief to small entities.

Analyzing alternatives establishes a process for the agency to evaluate proposals that achieve the regulatory goals efficiently and effectively without unduly burdening small entities, erecting barriers to competition, or stifling innovation. This process provides an

¹²³ Competitive status is not relevant when the small entities regulated by the proposed rule are not-for-profit organizations or governmental jurisdictions. In regulations that are limited to nonprofits or governmental jurisdictions, changes in regulatory costs should not affect the competitive status of the entities. However, there are certain nonprofit and governmental jurisdictions that do compete with for-profit enterprises, such as electric cooperatives. In preparing an IRFA, the agency must be mindful of the type of small entity regulated and tailor its analytical requirements to those entities.

¹²⁴ 5 U.S.C. § 603(c). Since the RFA is an economically neutral statute, the IRFA should examine alternatives to ensure that the proposed rule is maximizing any beneficial impact on small entities. In the case of a rule that has a significant beneficial effect, the failure to consider alternatives that enhance the beneficial effect means that the agency has not examined alternatives that "minimize" the economic impact of the proposed rule. For example, if a rule increases revenue to a small entity by \$100 and an alternative exists that meets the statutory objective of the agency and increases revenue by \$200, then the agency has not complied with the RFA if it did not examine the second alternative. The failure to provide the small entity with a potential extra \$100 in revenue in essence does not minimize the economic impact on small entities.

additional filter by which the agency conducts rational rulemaking mandated by the APA. Rather than focus on the overall costs and benefits of a particular regulation (as might be required by statute, such as the best achievable control technology, or by the regulatory analysis requirements of E.O. 12,866), the RFA requires the agency to undertake an analysis in order to discover less costly methods of attaining the statutory objectives of the rulemaking agency. Instead of analyzing the impacts of its regulatory actions on all relevant sectors of the economy, the IRFA narrows the scope of the particular review to small entities. The premise underpinning the IRFA is that, everything else being equal, the most rational alternative is often the one that achieves the objective of the agency at the lowest cost. Since small entities typically constitute the vast majority of entities in a particular industry under the SBA size standards, it often makes the most economic sense to adopt the regulatory strategy that imposes the least cost on small entities because that generally would represent the most cost-effective strategy meeting the agency's statutory objectives.

The kinds of alternatives that are possible will vary based on the particular regulatory objective and the characteristics of the regulated industry. However, section 603(c) of the RFA gives agencies some alternatives that they must consider at a minimum:

1. Establishment of different compliance or reporting requirements for small entities or timetables that take into account the resources available to small entities.
2. Clarification, consolidation, or simplification of compliance and reporting requirements for small entities.
3. Use of performance rather than design standards.
4. Exemption for certain or all small entities from coverage of the rule, in whole or in part.

Additional alternatives include adopting different standards for the size of businesses or modifying the types of equipment that are required for large and small entities. In short, the agency should consider a variety of mechanisms to reach the regulatory objective without regard to whether that mechanism is statutorily permitted. In some cases, the identification of regulatory alternatives that would be beneficial to the economy but cannot be implemented because of a statutory directive provides Congress with a clear legislative path. It is critical to remember that the IRFA is designed to explore less burdensome alternatives and not simply those alternatives it is legally permitted to implement. Returning to the analogy between RFA and NEPA, Council on Environmental Quality regulations providing guidance on NEPA compliance expect the agency to examine a "no-action" alternative even if such alternative would violate the statutory mandate, such as the need to protect a threatened and endangered species pursuant to the Endangered Species Act. Similarly, an agency might examine an exemption of small businesses even if the statute does not permit it because that informs Congress, the public, and the courts that it understands the implications of its regulatory action and is taking a less desirable course of action than it wishes. Such an assessment follows the parallels between the RFA and NEPA while providing information to the regulated community and decisionmakers in other branches of the federal government.

Agencies are not limited to alternatives that minimize burdens only for small entities. As EPA's 1992 RFA guidance recognized, cost-effective alternatives for small entities often are cost-effective for all entities.¹²⁵ Agencies should identify regulatory alternatives at the earliest stage of rulemaking and not wait until after the proposed rule is finished to develop alternatives. This is crucial because otherwise the agency may have already bought into one particular regulatory solution without considering alternatives. Such predeterminations by the agency violate the basic tenet of rational rulemaking under the APA by making the notice and comment process irrelevant. Interpretations of the notice and comment provisions of the APA contemplate a dialogue between the agency and the regulated community.¹²⁶ An agency already predisposed to only one way of thinking undermines the notice and comment procedure, thereby leaving itself open to a finding by a court that the agency action was arbitrary, capricious, or otherwise not in accordance with the law under section 706 of the APA.¹²⁷ Thus, the development of alternatives in the RFA demonstrates to the court that an agency did not in the proposed rule have a predisposition to rule in a manner that eviscerates the notice and comment process. If an agency is unable to analyze small business alternatives separately, then alternatives that reduce the impact for businesses of all sizes must be considered.

In the memorandum on regulatory flexibility that accompanied President Obama's E.O. 13,563, the president expanded the existing requirement for an agency to document the decision to reject an alternative that may reduce regulatory burdens on small entities. The RFA had required agencies to explain in the final regulatory flexibility analysis accompanying final rules why significant alternatives were not selected.¹²⁸ President Obama directed that a similar explanation be provided for proposed rules.¹²⁹

Consistent with an agency's obligations under section 609 of the RFA, agencies should perform outreach to interested groups to help develop regulatory solutions. In doing so, agency personnel should recognize that different sectors of an industry may have very different perspectives on a particular regulatory approach. The agency, before adopting one approach, should ensure that it contacts small entities and their representatives as well as large entities and their representatives. This type of communication is not prohibited by the APA and will help the agency focus on potential benefits and costs of various approaches to small businesses. In practice, the best proposed rules have been developed through a robust pre-proposal exchange of specific rulemaking concepts with the stakeholders including small businesses.¹³⁰

¹²⁵ See Revised Interim Guidance for EPA Rulewriters, p. 18.

¹²⁶ See *Chocolate Mfrs. Ass'n v. Block*, 755 F.2d 1098, 1103 (4th Cir. 1985).

¹²⁷ See *McLouth Steel Prods. v. EPA*, 838 F.2d 1317, 1324 (D.C. Cir. 1988); *Levesque v. Block*, 723 F.2d 175, 187 (1st Cir. 1983); *United States Steel Corp. v. EPA*, 595 F.2d 207, 214 (5th Cir. 1979).

¹²⁸ 5 U.S.C. § 604(a)(6).

¹²⁹ Memorandum for the Heads of Executive Departments and Agencies, "Regulatory Flexibility, Small Business, and Job Creation" (76 Fed. Reg. 3827, January 21, 2011).

¹³⁰ Executive Order 13,563 restates the value of pre-proposal input from affected firms. Section 2(c) states: "Before issuing a notice of proposed rulemaking, each agency, where feasible and appropriate, shall seek the views of those who are likely to be affected, including those who are likely to benefit from and those who are potentially subject to such rulemaking."

In essence, this outreach is an informal approach to the advance notice of proposed rulemaking that agencies often undertake to flesh out the parameters of a particular rule. Except in cases of emergencies or statutory deadlines, the Office of Advocacy strongly recommends that agencies consider using advance notices of proposed rulemaking for the most significant rules to identify potentially interested small entities and obtain estimates of the costs and benefits to small entities of various regulatory options. In particular, advance notices of proposed rulemaking will be extremely useful in developing information on the economic and structural characteristics of the industry, the small entities within that industry, and alternatives that would minimize costs and maximize benefits. Where the agency does not use an advance notice of proposed rulemaking, it should consider requesting information in the proposal regarding the economic and structural characteristics of the industry, including such items as the typical firm size, typical profits and losses, and the marginal costs of production, and should solicit suggestions for cost-effective regulatory approaches.

Duplicative, overlapping, and conflicting rules

The sixth element of the IRFA is to identify any duplicative, overlapping, and conflicting federal rules.¹³¹ Rules are duplicative or overlapping if they are based on the same or similar reasons for the regulation, the same or similar regulatory goals, and if they regulate the same classes of industry. Rules are conflicting when they impose two conflicting regulatory requirements on the same classes of industry.¹³²

This section of the IRFA requires the agency to examine the potential conflicting and duplicative rules that can unnecessarily add cumulative regulatory burdens on small entities without any gain in regulatory benefits. By identifying overlapping, duplicative, or inconsistent regulations, the agency might be able to avoid adding an additional regulatory burden (even one as simple as an additional report that is already filed elsewhere).¹³³

Because of the breadth and volume of federal regulations, a review of all existing rules on a particular industry group can be an onerous task for a federal agency. Nevertheless, it is important that the agency try to identify potential conflicting, duplicative, and overlapping regulations. The IRFA should include a request for comments identifying such rules. At the very least, the agency should review its own rules and identify any rules that cover the same subject matter and affect the same classes of industry. In fact, the law already requires such a review under section 610 of the RFA.¹³⁴

¹³¹ 5 U.S.C. § 603(b)(5).

¹³² For example, under the repealed ergonomics rule, OSHA would have forced skilled nursing facilities to acquire mechanical lifts to move patients. On the other hand, regulations promulgated by the Centers for Medicare and Medicaid Services (CMS) mandated that patients have a right not to be moved using mechanical lifts. Thus, the OSHA and CMS regulations would have been at cross purposes with respect to providing ergonomic protection for employees.

¹³³ In 1999, EPA relieved hundreds of thousands of facilities—facilities that were already filing federal underground storage tank forms for gasoline and diesel fuel with local authorities—from filing very similar reports for the same fuels under the federal community right-to-know law.

¹³⁴ See Chapter 6 for more information on compliance with Section 610 of the RFA.

Using other analyses to satisfy the IRFA requirements

The RFA permits agencies to prepare IRFAs in conjunction with, or as a part of, other analyses required by law as long as the RFA's requirements are satisfied.¹³⁵ Agencies need to exercise caution when relying on other analyses to satisfy the RFA, as they may not necessarily be a complete substitute for a regulatory flexibility analysis. In fact, these other analyses will prove far more useful as sources for data to be used in the IRFA than as substitutes for an IRFA. For major rules that require the preparation of a regulatory impact analysis (RIA) under Executive Orders 12,866 and 13,563, agencies may prepare the RIA and the regulatory flexibility analyses together. Nevertheless, the agency must keep in mind that the RIA is a much broader analysis of benefits and costs and does not necessarily focus on the cost effectiveness of regulatory compliance for small entities. Thus, the focus of the RIA under the executive orders is not a substitute for the IRFA. Agencies can coordinate their preparation of regulatory flexibility analyses with any other analyses accompanying a rule.¹³⁶ In doing so, however, agencies should ensure that such analyses describe explicitly how the requirements of the Regulatory Flexibility Act are satisfied. Similarly, agencies can develop evaluations of administrative burdens associated with reporting and recordkeeping requirements in concert with the paperwork burden analysis prepared under the Paperwork Reduction Act. However, Paperwork Reduction Act analysis is not a substitute for RFA compliance analysis.

When an IRFA may be waived or delayed

Section 608 of the RFA provides that an agency may waive or delay the completion of some or all the requirements of section 603 regarding preparation of IRFAs if the agency is promulgating the rule in response to an emergency that makes compliance with the RFA impracticable.¹³⁷ Promulgating agencies must publish the waiver or delay in the *Federal Register* no later than the date of publication of the final rule. If a true emergency exists, the agency must explain clearly why the circumstances constitute an emergency.

The RFA does not specifically allow certifications of proposed (or final) rules issued pursuant to section 605(b) to be waived or delayed. Certifications must be published at the time of the proposed or final rule. As discussed in Chapter 1, federal agencies must make a threshold assessment regarding the impact of proposed rules on small entities. This assessment, if it results in a certification, is judicially reviewable.

¹³⁵ 5 U.S.C. § 605(a).

¹³⁶ Many requirements of Exec. Order No. 12,866 parallel those in the RFA. See a discussion in the Introduction. Executive Order 12,866 directs agencies to “assess both the costs and the benefits of the intended regulation...[and] propose or adopt a regulation only upon a reasoned determination that the benefits...justify the costs.” Further, E.O. 12,866 requires agencies to develop and analyze regulatory alternatives, including, where appropriate, small business alternatives that achieve statutory objectives. Thus, it is often most effective to coordinate or combine analytic products used to satisfy both the E.O. and the RFA.

¹³⁷ 5 U.S.C. § 608(a).

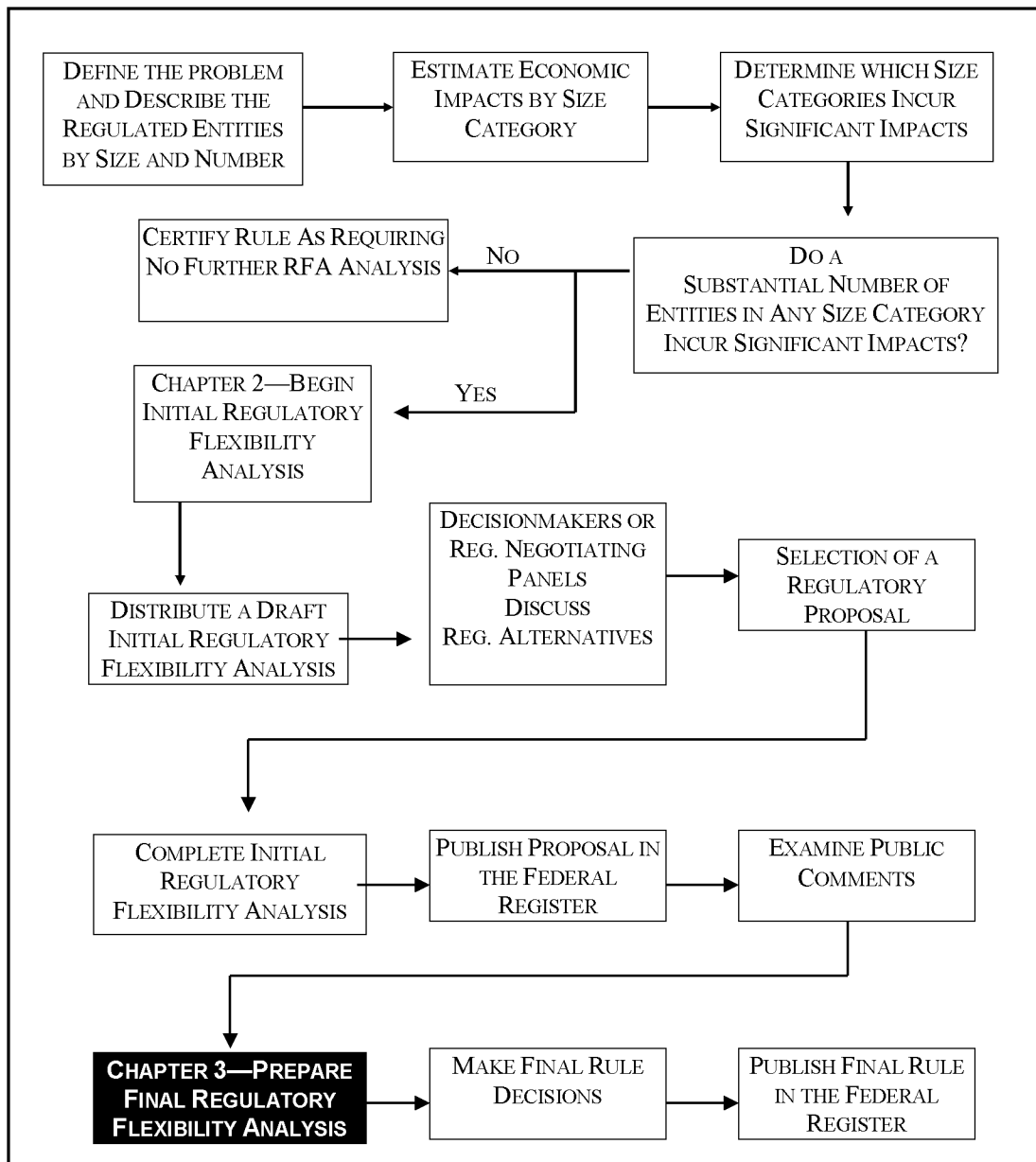
What an IRFA should look like: A real-life example

In Appendix L, a satisfactory IRFA by the Environmental Protection Agency contains the elements required by the RFA and a thorough analysis of the regulation's potential impact on small entities when insufficient data are available on cost or impact.¹³⁸

¹³⁸ For an example of a satisfactory IRFA when cost/impact data are available, see the CMS proposed rule on Medicare Program; Revisions to Payment Policies under the Physician Fee Schedule for Calendar 2003, 67 Fed. Reg. 43,846 (June 28, 2002), 43,865 ff. For another example, see U.S. Department of Transportation (DOT) proposed rule on Regulatory Assessment for Changes in Vessel and Facility Response Plans: 2003 Response Requirements, 67 Fed. Reg. 63,331, where DOT properly analyzed alternatives to the rule.

CHAPTER 3 PREPARING A FINAL RULE: THE FINAL REGULATORY FLEXIBILITY ANALYSIS

Figure 3. The RFA decision process: FRFA



When promulgating a final rule, agencies must prepare a final regulatory flexibility analysis (FRFA) unless the agency finds that the final rule will not have a significant economic impact on a substantial number of small entities or the final rule is issued under the APA provision allowing for good cause to forego notice and comment rulemaking.¹³⁹ When the agency publishes its final rule, it must also publish the FRFA, or a summary of the FRFA, in the *Federal Register*.¹⁴⁰ Draft final rules that are not certified must be submitted to Advocacy before publication in the *Federal Register*.¹⁴¹ The FRFA must include the agency's response to any comments filed by the Chief Counsel for Advocacy, including a detailed statement of any changes made to the proposed rule in the final rule as a result of such comments.¹⁴² The agency must also make copies of the FRFA available to the public. These published FRFAs are then subject to judicial review.¹⁴³

The RFA mandates that agencies revise their initial regulatory flexibility analysis based on the public comments received. Agencies routinely create a summary of the public's comments to be published along with the final rules. In developing this summary, the agency should specifically summarize comments from small entities even if the comments of the small entities do not relate to the RFA. This will help the agency prepare a more accurate FRFA or demonstrate support for a certification. Once the agency determines that it cannot certify the final rule under section 605(b), the agency must prepare a FRFA. If the agency determines that the rulemaking will not result in a significant economic impact on a substantial number of small entities, the head of the agency may so certify under section 605(b) of the RFA, and provide a copy of the certification to the Chief Counsel for Advocacy.¹⁴⁴

Issues to be addressed in the analysis

Section 604(a) of the RFA outlines the central issues the agency must address in the FRFA. In short, agencies must evaluate the impact of a rule on small entities and describe their efforts to minimize the adverse impact. To the extent that the final regulation has significant beneficial economic impacts, the agency should describe efforts to ensure that the benefits of the final rule maximize benefits to small businesses and minimize adverse economic impacts.

¹³⁹ 5 U.S.C. § 604 and 605(b). The APA provision is found in 5 U.S.C. § 553 (b)(B).

¹⁴⁰ *Id.* at § 604(b). Since the actual FRFA usually more accurately informs the public of the agency's efforts to analyze costs and alternatives, it is good practice to include the actual FRFA in the final rule preamble as published in the *Federal Register*.

¹⁴¹ Exec. Order No. 13,272, 67 Fed. Reg. 53,462 (Aug. 16, 2002).

¹⁴² 5 U.S.C. § 604(3).

¹⁴³ 5 U.S.C. § 611.

¹⁴⁴ As indicated earlier in the discussion concerning certifications, RFA § 605(b) requires that the certification appear in either the proposed or final rule. Although it is fairly clear that the certification must appear in the final rule if there is no certification in the proposed rule, it is not clear whether the certification must be duplicated in the final rule if it already appears in the proposed rule. The Office of Advocacy believes that, given the emphasis in the law on public notice, the certification should also appear in the final rule even though there may have already been a certification in the proposed rule. Doing so will help demonstrate the continued validity of the certification after receipt of public comments. In addition, significant changes between the proposed and final rule could warrant a change in the agency's certification evaluation for the final rule. For a more detailed discussion of certifications, see Chapter 1 of this guide.

The requirements for a FRFA are somewhat different from those for an IRFA. The requirements for the FRFA are very similar to the requirements that the courts impose on the development of a statement of basis and purpose for a final rule under section 553 of the APA.¹⁴⁵ The only additional requirements are those that relate to ensuring the items in the FRFA are easily identifiable to small entities without having to search the entire *Federal Register* notice. The agency should coordinate the preparation of the FRFA with development of the basis and purpose statement in the preamble. The preparation of a basis and purpose statement is not a substitute for a FRFA or for robust consideration of significant alternatives that are more cost-effective to small entities but still achieve the objectives of the agency. The requirements, outlined in seven provisions in section 604(a)(1)–(6), are highlighted in italics below:

1) *A statement of the need for, and objectives of, the rule.* The agency can cross-reference to a similar statement in the supplementary information if the cross reference enables small entities to easily identify the need for and objectives of the rule.

2) *A statement of the significant issues raised by the public comments in response to the IRFA, a statement of the assessment of the agency of such issues, and a statement of any changes made in the proposed rule as a result of such comments.* Under the APA, agencies are required to respond to comments addressing relevant statutory considerations.¹⁴⁶ Since the RFA constitutes a relevant statutory consideration, the agency is obligated under the APA to respond to comments on the RFA and relate how it changed the proposal, if at all, in response to the comments.

3) *The response of the agency to any comments filed by the Chief Counsel for Advocacy of the Small Business Administration in response to the proposed rule, and a detailed statement of any change made to the proposed rule in the final rule as a result of the comments.*¹⁴⁷

4) *A description of and an estimate of the number of small entities to which the rule will apply or an explanation of why no such estimate is available.*

5) *A description of the projected reporting, recordkeeping, and other compliance requirements of the rule, including an estimate of the classes of small entities which will be subject to the requirement and the type of professional skills necessary for preparation of the report or record.*

¹⁴⁵ *E.g.*, *Troy Corp. v. Browner*, 120 F.3d 277, 289 (D.C. Cir. 1997); *Lloyd Noland Hosp. v. Heckler*, 762 F.2d 1561, 1566-67 (11th Cir. 1985); *United States v. Nova Scotia Foods*, 568 F.2d 240, 252 (2d Cir. 1977); *Portland Cement Ass'n v. Ruckelshaus*, 486 F.2d 375, 393 (D.C. Cir. 1973), *cert. denied*, 417 U.S. 921 (1974); *Automotive Parts & Accessories Ass'n v. Boyd*, 407 F.2d 330, 338 (D.C. Cir. 1968).

¹⁴⁶ *Id.*

¹⁴⁷ The Small Business Jobs Act of 2010, Pub. Law 111-240 added this provision.

6) *A description of the steps the agency has taken to minimize the significant economic impact on small entities consistent with the stated objectives of applicable statutes, including a statement of the factual, policy, and legal reasons for selecting the alternative adopted in the final rule and why each of the other significant alternatives to the rule considered by the agency which affect the impact on small entities was rejected.* Again this requirement already is mandated by the rational rulemaking requirements of the APA.¹⁴⁸

6) *For a covered agency, as defined in section 609(d)(2), a description of the steps the agency has taken to minimize any additional cost of credit for small entities.*¹⁴⁹

As noted in the third provision above, section 1601 of the Small Business Jobs Act¹⁵⁰ further amended the final regulatory flexibility analysis (FRFA) section of the RFA by requiring agencies to respond to any comments filed by the Chief Counsel for Advocacy in response to a proposed rule and a detailed statement of any changes made in response to the comments.

Additional questions to be addressed in a FRFA

A number of important questions will assist the agency in preparing a FRFA:

Have all significant issues been assessed?

Have all significant issues raised in the public comments regarding the IRFA been summarized and assessed, and have any changes been made since the publication of the proposed rule as a result of those comments? The RFA does not require agencies to address every issue raised during the public comment period—only the significant ones. The RFA does require agencies to assess (and not just present) the significant issues raised by interested stakeholders. Agencies are also required to publish in the final rule the specific changes that were made to the proposed rule in response to the public comments, as well as comments from the SBA’s Chief Counsel for Advocacy. Although there is no requirement to do so, some agencies include in their FRFAs the number of times a particular comment was raised.

Has the number of small entities been estimated?

Is it possible to estimate the number of small entities to which the rule will apply? If not, why not? The RFA requires that during its IRFA preparation, the agency must estimate the number of small entities affected. An additional FRFA requirement is that if no estimates of the number of affected small entities are available, agencies must explain why. An agency must have a strong argument that it cannot estimate the number of small

¹⁴⁸ See *American Textile Mfrs. Inst. v. Donovan*, 452 U.S. 490, 539-41 (1981).

¹⁴⁹ The numbering is as shown—two paragraphs (6) were enacted. The Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010, Pub. Law 111-203, added this provision.

¹⁵⁰ Public Law 111-240.

entities, as in the case of a regulation affecting an emerging industry about which little is known.

If an agency is uncertain about how to proceed in the absence of firm data, Advocacy advises agencies to construct public records that reflect aggressive and meaningful public outreach. Agencies should compile economic data on the industries/organizational sectors to be regulated and the economic impacts on small entities within those sectors. If such efforts produce inconclusive data or fail entirely, the agency may demonstrate its efforts to comply with the requirements of the RFA and explain why such data were not available. Moreover, this will demonstrate to the courts that the agency was conducting rational rulemaking by determining the universe of affected entities.

Has the adverse economic impact on small entities been minimized?

Agencies must consider, and may adopt, one or more significant alternatives to minimize the rule's burden on small entities.¹⁵¹ Some of the traditional alternatives may include lengthening the time for compliance; tiering the compliance requirements based on the size of the business or degree to which small entities contribute to the problem; providing for exemptions for parts of the rule or the entire rule for small entities; timing compliance to correspond with other statutory deadlines with related requirements; allowing for increased flexibility in the methods used for achieving the agency's objectives (for example, using a performance standard instead of requiring a specific technology); making requirements less prescriptive; etc. Such alternatives also include providing regulatory relief to all regulated entities, such as lowering the overall stringency of a standard or changing the regulatory threshold. In the first instance, it remains the obligation of the agency to develop significant alternatives pursuant to the RFA. Otherwise the agency is transferring its statutory RFA mandate to those entities that can least afford or have the least expertise in rulemaking processes to craft alternatives—small entities. Even after the agency has crafted alternatives, it should, as a matter of course, in the proposed rule and IRFA, specifically request whether any other alternatives exist that the agency has not considered. Small entities may be able to provide additional alternatives based on the analysis already performed by the agency, i.e., the analysis may spark ideas that small entities may not have thought of absent such analysis. Adoption of this procedure will ensure that agencies have met their obligation to consider alternatives to the final regulatory solution as mandated by the RFA.

¹⁵¹ The outcome of a rulemaking would be superior if the agency adopted a standard that achieves its objectives but reduces burdens or increases benefits to small entities. Development of regulations that have small entity orientation will be beneficial in the long run to the agency. Since most regulated entities are small, rules that have a small entity orientation will likely garner greater support from that community, increased compliance, reduced penalties, and quicker achievement of the agency's statutory objective. A regulation that does not have such small entity orientation will face resistance from the regulated community, force the agency to increase enforcement, and delay accomplishment of whatever goal the agency was attempting to reach. For example, if the OSHA ergonomics rule had gone into effect in 2001, it is unlikely that many small entities could have complied. The Department of Labor would have expended scarce resources to obtain compliance without accomplishing the goal of increasing worker safety.

Have all significant alternatives been reviewed?

Has the statement of factual, policy, and legal reasons for selecting the alternative adopted in the final rule, and the reasons for rejecting other significant alternatives, been included or appropriately cross-referenced for easy identification by small entities? The Small Business Regulatory Enforcement Fairness Act (SBREFA)¹⁵² made significant changes to this section of the RFA with respect to compliance requirements. Prior to 1996, an agency needed only state the alternatives and the reason (or reasons) for rejecting a particular alternative. As a result of the amendments, an agency must now include a statement of the factual, policy, and legal reasons for selecting the alternative adopted in the final rule. This explanation already is required under the APA, and the FRFA will help the agency demonstrate compliance with the APA's rulemaking procedures through the clarification of the reasons for selecting or rejecting particular alternatives. In addition to educating the courts, the rationales might spur action by Congress to correct a flaw that the agency identified. Thus, the FRFA, if done correctly, can play a key role in the development of public policy. The agency must also detail for the public record why each of the other significant alternatives was rejected; again, this is a requirement of APA rulemaking requiring the agency to explain how it considered all relevant statutory criteria including those mandated by the RFA. The changes indicate that agencies were not providing specific explanations of their final actions. There should be significant articulable and supportable reasons for rejecting alternatives. President Obama reaffirmed the principle of documenting a decision to reject an alternative that may reduce regulatory burden for small entities.¹⁵³ The development and consideration of alternatives is subject to judicial review.¹⁵⁴

Permissible delays in publication; provision for lapse of final rule

Section 608(b) of the RFA provides that an agency may delay, but not waive, the completion of a FRFA if the rule is being promulgated in response to an emergency that makes compliance with the RFA impracticable. Under this provision, the agency must publish its reasons for the delay upon publication in the *Federal Register*. The delay may not exceed 180 days after the final rule is published; otherwise the rule lapses and has no effect. The rule cannot be re-promulgated until a FRFA has been completed. This section is also subject to judicial review.

¹⁵² 5 U.S.C. § 604(a)(6).

¹⁵³ Memorandum for the Heads of Executive Departments and Agencies, "Regulatory Flexibility, Small Business, and Job Creation" (76 Fed. Reg. 3827, January 21, 2011).

¹⁵⁴ See *National Ass'n of Psychiatric Health Sys. v. Shalala*, 120 F. Supp. 2d 33 (D.D.C. 2000), in which the court ordered HHS to complete a FRFA that discussed less burdensome alternatives considered and rejected in order to comply with the RFA.

What a FRFA should look like: A real-life example

In Appendix M is an example of a satisfactory FRFA released by the Environmental Protection Agency. This FRFA contains each of the elements required by the RFA and presents a thorough analysis of the regulation's impact on small entities.¹⁵⁵

¹⁵⁵ For an additional example of a satisfactory FRFA, see the Environmental Protection Agency final rule for Effluent Guidelines and Standards for the Organic Chemicals, Plastics, and Synthetic Fibers Industry, 58 Fed. Reg. 36,872 (July 9, 1993).

CHAPTER 4 SBREFA PANELS

In 1996, SBREFA amended the RFA to include a number of important provisions. One of those was section 609, which requires, among other things, that certain agencies conduct special outreach efforts to ensure that small entity views are carefully considered prior to the issuance of a proposed rule. This outreach is accomplished through the work of small business advocacy review panels, sometimes referred to as SBREFA or SBAR (small business advocacy review) panels.

In July 2010, the United States Congress passed the Dodd-Frank Wall Street Reform and Consumer Protection Act (Act).¹⁵⁶ Section 1011 of the act establishes the Consumer Financial Protection Bureau to supervise certain activities of financial institutions. Section 1100G, titled “Small Business Fairness and Regulatory Transparency,” amends 5 U.S.C. § 609(d), to require the CFPB to comply with the SBREFA panel process, making it the third agency with this responsibility, along with the Environmental Protection Agency and the Occupational Safety and Health Administration.

In addition to the regular requirements of the initial regulatory flexibility analysis (IRFA) found in 5 U.S.C. § 603, a CFPB IRFA must include “a description of (A) any projected increase in the cost of credit for small entities; (B) any significant alternatives to the proposed rule which accomplish the stated objectives of applicable statutes and which minimize any increase in the cost of credit for small entities; and (C) advice and recommendations of representatives of small entities relating to issues described in subparagraphs (A) and (B) and subsection (b).” When the Bureau produces a final regulatory flexibility analysis, it must include “a description of the steps the agency has taken to minimize any additional cost of credit for small entities.”

Who must hold SBREFA panels?

The statute requires that EPA, CFPB, and OSHA evaluate their regulatory proposals to determine whether SBREFA panels should be convened.¹⁵⁷ The requirement for SBREFA panels may appear to impose additional steps for these agencies in their rulemaking processes. However, the panel process only formalizes the outreach requirements and analyses that the Administrative Procedure Act and the RFA already mandate for all new rules that affect small businesses. Any additional work that may be needed in this special early outreach effort should be offset by time saved at the other end of the regulatory process. When problems are resolved before a proposed rule is published, objections from the public are reduced. Experience has shown that the panel process results in better rules, better compliance, and reduced litigation. In at least two instances, EPA withdrew a regulatory proposal based on work performed in connection with the panel process.¹⁵⁸

¹⁵⁶ Public Law 111-203.

¹⁵⁷ 5 U.S.C. § 609(d)(2).

¹⁵⁸ See EPA’s Effluent Limitations Guidelines for Industrial Laundries; 64 Fed. Reg. 45071 (Dec. 12, 1997), withdrawn by EPA on August 18, 1999, Effluent Limitations Guidelines for Construction and Development, 67 Fed. Reg. 42644 (June 24, 2002), withdrawn April 26, 2004.

How is the decision to hold a SBREFA panel made?

For each proposed rule, the RFA requires that an agency either certify that the proposal has no significant economic impact on a substantial number of small entities, or prepare an IRFA on the proposal.¹⁵⁹ Whenever EPA, CFPB, or OSHA determines that a regulatory proposal may have a significant economic impact on a substantial number of small entities, the law further requires that the agency convene a review panel. SBREFA panels are required for all EPA, CFPB, and OSHA rules for which an IRFA is required. Panel outreach must take place before the publication of the proposed rule. However, the Chief Counsel for Advocacy may waive the panel requirement upon the request of EPA, CFPB, or OSHA under certain conditions. To waive the panel requirement, the Chief Counsel must find that convening a panel would not advance the effective participation of small entities in the rulemaking process. Section 609(e) of the RFA lays out several factors in making this determination, including consideration of whether small entities have already been consulted in the rulemaking process and whether special circumstances warrant the prompt issuance of a rule.

How does a SBREFA panel work?

A SBREFA panel consists of a representative or representatives from the rulemaking agency, the Office of Management and Budget's Office of Information and Regulatory Affairs (OIRA) and the Chief Counsel for Advocacy.

The panel solicits information and advice from small entity representatives (SERs), who are individuals that represent small entities affected by the proposal. SERs help the panel better understand the ramifications of the proposed rule. Invariably, the participation of SERs provides extremely valuable information on the real-world impacts and compliance costs of agency proposals.

The law requires that a SBREFA panel be convened and complete its report with recommendations within a 60-day period. The formal panel process begins with the convening of the panel by the rulemaking agency. The date is normally fixed after consultation with both Advocacy and OIRA. Before convening, the three agencies work together to discuss regulatory alternatives and their advantages and disadvantages. The agencies also discuss what data, information, and regulatory alternatives will be presented to the SERs so that they can provide informed advice. As EPA advises in its SBREFA panel guidance, the agency “need(s) to describe in sufficient detail, including some analysis of the impact on small entities and environmental benefits, each significant regulatory alternative you have identified that accomplishes the statutory mandate.”¹⁶⁰ With this information, the small entity representatives will be able to provide informed advice to the panel. The rulemaking agency usually has preliminary discussions with small entities about its draft proposal before the panel is formally convened. These

¹⁵⁹ See Chapter 1 for a detailed discussion of how to certify a proposed rule and Chapter 2 on how to prepare an initial regulatory flexibility analysis.

¹⁶⁰ 2006 EPA Final Guidance, section 5.8.2. See section 5.8.2 for more guidance on what information should be provided to the panel and the small entity representatives.

preparations ensure that the panel process can be completed during the statutorily specified 60-day period.

The product of a SBREFA panel's work is its panel report on the regulatory proposal under review. The panel completes its final report, including its recommendations, early in a rule's developmental stages, so that the agency has the benefit of the report's findings prior to publication of a proposed rule. The panel report also becomes part of the official docket for the proposed rule.

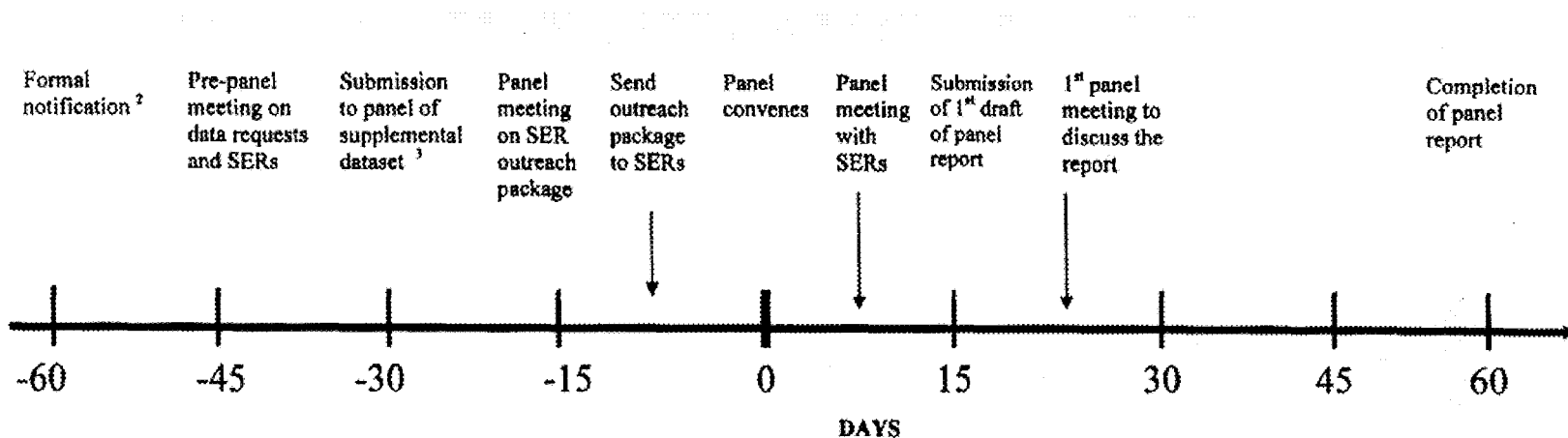
The purpose of the panel process is threefold. First, the panel process ensures that small entities that would be affected by a regulatory proposal are consulted about the pending action and offered an opportunity to provide information on its potential effects. Second, a panel can develop, consider, and recommend less burdensome alternatives to a regulatory proposal when warranted. Finally, the rulemaking agency has the benefit of input from both real-world small entities and the panel's report and analysis prior to publication.¹⁶¹

Suggested SBREFA panel timeline

The RFA provides that the formal panel process must be concluded within 60 days from the formal convening of the panel to the completion of its report. Experience has shown that the panel process works best if agencies and panel members accomplish as much preliminary work as possible before the formal convening of the panel. A suggested timeline is shown in Figure 4, although panel members have flexibility to adjust their pre-panel work schedules to ensure the best outcome for each individual rule.

The EPA procedure is to hold two meetings with the SERs, one preceding and one following the formal convening of the panel. There are two opportunities for oral exchanges with the panel members, followed by two opportunities for written comments 15 days after the meetings. The two sessions facilitate a robust discussion of the issues, and give the agency the ability to further refine its draft regulatory alternatives in light of the initial round of written SER comments. The timeline on the next page is based on the OSHA practice of a single SER meeting after convening; however, the OSHA practice is to start with a fully developed draft proposed rule and preamble.

¹⁶¹2006 EPA Final Guidance, Chapter 5, is a good source for effective panel procedures implementation, <http://www.epa.gov/rfa/documents/Guidance-RegFlexAct.pdf>.

Figure 4. Suggested SBREFA panel timeline¹

1. The suggested timeline for the panel process can be adjusted as necessary, except that the statute requires the panel's report to be completed within 60 days of the convening of the panel, Day 0 in this chart. Generally, as much preliminary work as possible should be done before Day -60.
2. The formal notifications by the convening agency to Advocacy and OIRA should include:
 - a description of the important components of the rule;
 - a description of the problem the rule is trying to solve and of the statutory obligations underlying the rule;
 - a quantitative or, if impracticable or unreliable, a qualitative description of the potential impacts;
 - a description of the types of entities likely to be affected by the proposed rule and of any small-entity stakeholder involvement in the process to date;
 - a description of any regulatory flexibility alternatives that are or have been under consideration;
 - a list of potential small entity representatives; and
 - a list of any other important documents or information that have already been developed to support the rulemaking.
3. The supplemental dataset should include a description of regulatory flexibility alternatives, information necessary to evaluate these alternatives or any other information that is reasonable to request, and the final list of SERs whom the Small Business Advocacy Review Panel Chairperson intends to select upon convening the panel.

CHAPTER 5 RFA LITIGATION: WHAT THE COURTS HAVE SAID

This chapter examines litigation regarding the Regulatory Flexibility Act and is organized in sections corresponding to those of the compliance guide overall. The section does not reflect the Office of Advocacy's opinion of the cases; rather, it is intended to provide the reader with information on specific case law and what the courts have held regarding agency compliance with the RFA.

Where do we begin? First steps of RFA rule analysis

Does the RFA apply?

An agency must first consider whether the RFA applies to the regulatory proposal at issue. An appropriate consideration begins with an examination of the Administrative Procedure Act (APA) as it relates to the RFA.

If, under the APA or any rule of general applicability governing federal grants to state and local governments, the agency is required to publish a general notice of proposed rulemaking (NPRM), the RFA must be considered.¹⁶² Significantly, some agencies, such as the Rural Utilities Service, have their own administrative rules that require notice and comment even though the agency's rules may be exempt from the APA. If an NPRM is not required, the RFA does not apply.¹⁶³ Pursuant to RFA section 601(2), the term "rule" does not include a rule of particular applicability to rates, wages, corporate or financial structures or reorganization thereof, prices, facilities, appliances, services, or allowances.

¹⁶² 5 U.S.C. § 604(a). See also *National Association of Home Builders v. Army Corps of Engineers*, 417 F.3d 1272 (D.C. Cir. 2005) where the plaintiffs challenged nationwide permits issued under the Clean Water Act by the Corps as violating, *inter alia*, the RFA, because the Corps did not conduct a flexibility analysis as required by the RFA. The Army Corps of Engineers argued that its permitting action did not constitute a "rule." It was an "order" because "order" included a "licensing" disposition and a "license" included a "permit." The court considered the argument an "elaborate statutory construction" and rejected it for a more straightforward one. The court found that the permitting action fit within the APA's definition of "rule" because each permit was a legal prescription of general and prospective applicability which the Corps issued to implement permitting authority that Congress entrusted to it pursuant to the Clean Water Act. As such, the action constituted a rule because it was an agency statement of general or particular applicability and future effect designed to implement, interpret, or prescribe law or policy. In addition, the court found that the Army Corps of Engineers action was a legislative rule because the permits authorized the discharge of certain materials, granted rights, imposed obligations, produced other significant effects on private interests. Accordingly, it was subject to the notice and comment requirements of the APA and to the requirements of the RFA.

¹⁶³In *Roche v. Evans*, 249 F. Supp. 2d 47 (D. Mass. 2003), the New England Fishery Management Council (Council) adopted an adjustment to the existing Northeast Multispecies Fisheries Management Plan (FMP) mandating that certain fishing areas would be closed to fishing for varying lengths of time. The court stated that the RFA does not apply to the adoption of such a framework adjustment to an FMP because, under the abbreviated framework adjustment procedure permitted under 50 C.F.R. § 648.90, there is no requirement that the Council "publish a general notice of proposed rulemaking." The court noted, "the whole purpose of the framework adjustment procedure is to dispense with that requirement." 249 F.Supp.2d at 57. With the trigger of notice and comment lacking, the court granted summary judgment in favor of the agency.

Further, only actions that qualify as rulemaking under the APA that affect small entities or small entity concerns trigger the protections of the RFA.¹⁶⁴ Small entities whose concerns must be accounted for include small businesses, small not-for-profit organizations, and small governmental jurisdictions—cities, counties, towns, townships, villages, school districts, or special districts, with a population of less than 50,000.¹⁶⁵

What qualifies as a rulemaking under the APA?

Rules are exempt from APA requirements, and therefore from the RFA requirements, when any of the following is involved:

1. Military or foreign affairs functions of the United States.¹⁶⁶
2. Matters relating to agency management or personnel or to public property, loans, grants, benefits, or contracts.¹⁶⁷

Also exempt from the APA requirement for notice and comment rulemaking are interpretative rules.¹⁶⁸ Interpretative rules generally require no judgments and little by the agency on implementation, but rather interpret the language or intent expressed by Congress. Legislative rules require judgments and great discretion; an example is setting a clean air standard for the nation.

Exemptions under the APA

The D.C. District Court has addressed exemptions under the APA in determining whether the action qualifies as a rulemaking requiring notice and comment. In the following cases the courts held that the RFA did not apply because the APA requirements for notice and comment are inapplicable:

¹⁶⁴ *Atlantic Fish Spotters Association v. Evans*, 206 F. Supp. 2F.d 81, 93 (D. Mass. 2002).

¹⁶⁵ 5 U.S.C. § 601(3)-(5). See also Chapter 1 of this guide for a discussion of what qualifies as a small entity. *State v. Centers for Medicare & Medicaid Services*, 2010 WL 1268090 (M.D. Ala. 2010) (memorandum and order) where the court held that the State of Alabama did not have standing as a small entity; *La Gloria Oil & Gas Co. v. United States*, 56 Fed. Cl. 211 (2003), the court declined to address the plaintiff's arguments concerning alleged violation of the RFA because plaintiff was not a small business; *Williams Alaska Petroleum v. United States*, 57 Fed. Cl. 789 (2003), the plaintiff was precluded from asserting a claim under the RFA because the plaintiff was not a small entity; *Navajo Refining Co. v. United States*, 58 Fed. Cl. 200 (2003), the court declined to address the plaintiffs' arguments concerning the defendant's alleged violation of the RFA because they were not small businesses and lacked standing to challenge the defendant's compliance with the RFA.

¹⁶⁶ APA § 553(a)(1) exempts from notice and comment rulemaking those rules involving "a military or foreign affairs function of the United States." The legislative history of § 553(a)(1) indicates the exception should be construed narrowly to include only those " 'affairs' which so affect relations with other governments that, for example, public rulemaking provisions would clearly provoke definitely undesirable international consequences." S.Rep. No. 752, 79th Cong., 1st Sess. 13 (1945). *Jean v. Nelson*, 711 F.2d 1455 (11th Circuit 1983).

¹⁶⁷ 5 U.S.C. § 553(a).

¹⁶⁸ SBREFA amended the RFA to bring certain interpretative rulemakings of the Internal Revenue Service within coverage of the RFA. The law now applies to those IRS rules published in the *Federal Register* that would normally be exempt from the RFA as interpretative rules, but that impose a "collection of information" requirement on small entities. For a more detailed discussion, see Chapter 1.

Military or foreign affairs functions of the United States. In reviewing the early RFA case, *In re Sealed Case*,¹⁶⁹ the D.C. District Court held that regulations such as those delineating the products subject to the ban on importation into the United States of uranium ore, uranium oxide, textiles, and coal from South Africa, fell under the foreign affairs function of the United States; thus, the provisions of the Administrative Procedure Act, 5 U.S.C. 553, requiring notice of proposed rulemaking and opportunity for public participation were inapplicable. Because a notice of proposed rulemaking is not required for this rule, the Regulatory Flexibility Act, 5 U.S.C. §601 *et seq.*, did not apply.¹⁷⁰

Interpretative rules. In *National Association for Home Care v. Shalala*,¹⁷¹ the plaintiffs argued that the Department of Health and Human Services failed to consider alternatives to the proposed rule as required by the RFA. The agency, however, asserted that the Balanced Budget Act (BBA) did not grant the Secretary any discretion in implementing the Interim Payment System (IPS). The court agreed, holding that the BBA was an interpretative rather than substantive rule, given its high degree of specificity regarding the implementation of the IPS. As an interpretative rule, the BBA need not comply with the RFA. The court stated generally that the RFA does not apply to interpretative rules which merely clarify or explain existing laws or regulations.¹⁷²

Publications not subject to the APA and rate exemptions. In *American Moving and Storage Association, Inc., v. DOD*,¹⁷³ the D.C. District Court examined a notice published in the *Federal Register* by the Department of Defense announcing a significant change in procurement policy regarding its source for distance calculations for payments and audits in its transportation programs from a previously used official mileage table to a new computer software program. The plaintiffs asserted that the change would have a significant economic impact on small carriers, requiring RFA compliance. DOD asserted that the policy change was not a “rule” as defined by the RFA, and therefore it did not have to comply with the RFA. The court agreed with the agency and held that the procurement policy change was not a “rule” for RFA purposes. The court further found that even if the RFA definition of a rule included some procurement policy changes, the calculations for payments and audits were exempt from the definition by the APA exception relating to rates.¹⁷⁴ As a result, the RFA did not apply.¹⁷⁵

Good Cause. In *Oregon Trollers Association v. Gutierrez*,¹⁷⁶ the Ninth Circuit upheld the lower court’s decision regarding NMFS’s invocation of the good cause exception to the APA’s notice and comment provisions in an action involving the management of the Chinook salmon season. The plaintiffs argued that NMFS failed to prepare the economic analyses required by the RFA. The RFA applies to any rule requiring notice and

¹⁶⁹ *In re Sealed Case*, 666 F. Supp. 231, 236 (D.D.C. 1987).

¹⁷⁰ *Id.* See also *Jean V. Nelson*, 711 F.2d 1455.

¹⁷¹ *National Association for Home Care v. Shalala*, 135 F. Supp. 2d 161, 165 (D.D.C. 2001).

¹⁷² *Id.* See also, *Cent. Tex. Tel. Coop. v. F.C.C.*, 402 F.3d 205 (D.C. Cir. 2005); *Broadgate, Inc. v. United States Citizenship & Immigration Services*, 730 F. Supp. 2d 240 (D.D.C. 2010).

¹⁷³ *American Moving and Storage Association v. DOD*, 91 F. Supp. 2d 132, 136 (D.D.C. 2000).

¹⁷⁴ 5 U.S.C. § 553(b) (1996).

¹⁷⁵ *Id.* at 136.

¹⁷⁶ *Oregon Trollers Association v. Gutierrez*, 452 F.3d 1104 (9th Cir. 2006).

comment under section 553(b) of the APA. The court held that NMFS’s invocation of the “good cause” exception to the RFA requirement was valid because the NMFS gave season-specific reasons for the exception. NMFS explained that management measures are based on data from the prior season, which are not available until January. Because the new season opens on May 1, the 60-day comment period is infeasible. The court added that as long as the NMFS provides fresh reasoning related to the season in which the exception applies, repeated invocation of the exception is not a problem.

The certification statement

The decision process

An agency may certify that no regulatory flexibility analysis is necessary when it determines that the rule will not have a significant economic impact on a substantial number of small entities that are subject to the requirements of the rule. However, an agency must provide a factual basis for the certification. A mere statement that there will be no effect is not sufficient. The agency must conduct an analysis demonstrating that it has considered the potential effects of the regulation.¹⁷⁷

Cases in which the certification violated the RFA. In a number of cases, the certification was found to have violated the RFA.

In *Northwest Mining Association v. Babbitt*,¹⁷⁸ the Bureau of Land Management (BLM) published a final rule in February 1997 that would impose a bonding requirement on hardrock mining. The rule was originally proposed in 1991. While the original proposal would have set a limit on bonding requirements, the final rule contained burdensome provisions not included in the proposal—provisions on which the public, therefore, had no opportunity to comment. The BLM certified that the rule would not have a significant economic impact on a substantial number of small entities. However, the agency failed to substantiate its conclusions. In remanding the rule, the court stated that the final rule’s certification violated the RFA because the factual basis for the certification that the agency provided failed to incorporate the correct definition of small entity.¹⁷⁹

In *North Carolina Fisheries Association v. Daley*,¹⁸⁰ the District Court for the Eastern District of Virginia found that NMFS violated the RFA when it certified that there would not be a significant economic impact on a substantial number of small entities, because the fishing quota would remain unchanged. The court remanded the matter to NMFS with instructions to perform a proper analysis because even though the quota was the same, the agency provided no data to show that the quota was still valid.¹⁸¹

¹⁷⁷ *North Carolina Fisheries Association v. Daley*, 16 F. Supp. 2d 647, 652 (E.D. Va. 1997).

¹⁷⁸ *Northwest Mining Association v. Babbitt*, 5 F. Supp. 2d 9, 14 (D.D.C. 1998).

¹⁷⁹ *Id.* at 652.

¹⁸⁰ *North Carolina Fisheries Association v. Daley*, 16 F. Supp. 2d 647 (E.D. Va. 1997).

¹⁸¹ See additional discussion of this case later in this chapter.

In *Harlan Land Co. v. United States Department of Agriculture*,¹⁸² the District Court for the Eastern District of California found the certification analysis performed by the Animal and Plant Health Inspection Services (APHIS) of the U.S. Department of Agriculture (USDA) was inadequate. APHIS had published a final rule allowing the importation of lemons, grapefruit, and oranges from various areas in Argentina. APHIS prepared an economic analysis of the rule and determined that the rule would not have a significant economic impact on a substantial number of small entities. Based on that determination, APHIS did not prepare an RFA analysis.¹⁸³ Citrus growers brought suit against the USDA and APHIS, arguing that the agency violated both the APA and the RFA in issuing the rule. The economic analysis in the final rule focused on the impact that the Argentine imports would have on the supply and prices of citrus fruit in the United States and the resulting costs and benefits to domestic growers, etc. The analysis failed to consider what the costs would be if Argentine plant pests were introduced into U.S. citrus orchards. The court found that APHIS's determination of a lack of significant economic impact on a substantial number of small entities was based on its conclusion that there was a negligible risk of pest introduction. The court considered the risk assessment to be flawed and thus remanded the final rule to the defendants for consideration of the economic impact that the importation of Argentine citrus will have on small businesses.

In *American Federation of Labor v. Chertoff*,¹⁸⁴ the Department of Homeland Security (DHS) promulgated a final rule titled Safe-Harbor Procedures for Employers Who Receive a No-Match Letter. Under the rule, an employer who received a “no-match letter” (indicating that an employee's name and social security number did not match) could take certain actions to avoid liability. The plaintiffs (union and business representatives) sought a preliminary injunction to bar enforcement of the rule, asserting that it was arbitrary and capricious in violation of the APA, and that promulgation of the rule violated the RFA. In promulgating the rule, DHS certified that the rule would not have a significant impact on small entities. However, in briefing, DHS claimed that an RFA analysis was unnecessary because the rule was voluntary, and that the RFA does not apply to interpretative rules. The court did not consider the post-rule rationalization that the rule was interpretive, focusing instead on DHS's first argument, which was that there was no impact on small entities because the rule was voluntary. The court was persuaded by the plaintiff's declarations that the rule could have significant costs, noting the potential costs of hiring human resources staff to track and solve mismatches, hiring legal services help, and training staff. The court decided that there were “serious questions [about] whether DHS violated the RFA,” and granted the plaintiff's motion for preliminary injunction.¹⁸⁵

Where the court found that certification was appropriate. In other cases, courts found that agencies properly certified rules.

¹⁸² *Harlan Land Co. v. United States Dept. of Agriculture*, 186 F. Supp. 2d 1076 (E.D. Cal. 2001).

¹⁸³ *Id.* at 1097.

¹⁸⁴ *American Federation of Labor v. Chertoff*, 552 F.Supp.2d 999, (N.D.Cal., 2007).

¹⁸⁵ *Id.* at 1013.

In *Associated Builders and Contractors, Inc., v. Herman*, the Department of Labor suspended a revised class of employees called “helpers” on federal construction sites in 1993 and reinstated former helper regulations pursuant to a congressional mandate.¹⁸⁶ Regarding the RFA, the Department of Labor certified that the rule would not have a significant economic impact on a substantial number of small entities. The court upheld the certification, because the rule preserved the status quo, and DOL estimated few firms would have taken advantage of the helper classifications during the interim period pending final rulemaking.¹⁸⁷

In *Environmental Defense Center. v. E.P.A.*,¹⁸⁸ EPA issued a rule, pursuant to the Clean Water Act, to control pollutants introduced into the nation’s waters by storm sewers. The rule mandated that discharges from small municipal storm sewers and construction sites sized 1-5 acres be subject to the permitting requirements of the National Pollutant Discharge Elimination System (NPDES). The EPA certified that the rule would not yield “significant impacts.” The plaintiffs argued that the EPA’s certification was erroneous because the EPA mislabeled significant costs as “not significant,” failed to account for the costs of all affected small entities, and failed to account for all significant costs to small entities. The Ninth Circuit agreed with the Natural Resources Defense Council’s view that “plain language of § 605(b) sets out a three-component test indicating that EPA need not perform a regulatory flexibility analysis if it finds that the proposed rule will not have: (1) “a significant economic impact” on (2) “a substantial number” of (3) “small entities.”¹⁸⁹ The Ninth Circuit determined that the EPA complied with the RFA and reasonably certified that the rule would not have a significant economic impact, but did not explain clearly its reasoning, beyond stating the legal test described above.

The court also noted that any procedural defect was harmless error because the EPA had already conducted the economic analyses the petitioners sought when they convened a small business advocacy review panel before publishing notice of the proposed rule. The EPA had followed the advice and recommendations of the panel and included provisions designed to minimize impacts on such entities, such as alternative compliance and reporting mechanisms responsive to the resources of small entities, simplified procedures, performance rather than design standards, and waivers. The court noted that “...the analyses required by RFA are essentially procedural hurdles; after considering the relevant impacts and alternatives, an administrative agency remains free to regulate as it sees fit.”¹⁹⁰

In *Cactus Corner v. U.S.D.A.*,¹⁹¹ USDA promulgated a rule allowing and setting conditions for resumption of the importation of Spanish clementines, following a ban after the discovery of live Mediterranean fruit fly (Medfly) larvae. Domestic fruit growers and packers sought declaratory and injunctive relief to set aside and hold the rule

¹⁸⁶ *Associated Builders and Contractors, Inc., v. Herman*, 976 F. Supp. 1 (D.D.C. 1997).

¹⁸⁷ *Id.*

¹⁸⁸ *Environmental Defense Center. v. E.P.A.*, 344 F.3d 832 (9th Cir. 2003).

¹⁸⁹ *Id.* at 879-880.

¹⁹⁰ 344 F.3d at 879.

¹⁹¹ *Cactus Corner v. U.S.D.A.*, 346 F. Supp. 2d 1075 (E.D. Cal. 2004).

unlawful, claiming, *inter alia*, that the rule violated the RFA because the agency had failed to prepare an initial or final regulatory flexibility analysis. It also sought to enjoin the defendant from implementing the rule or otherwise allowing the importation of clementines from Spain, and an award of costs, disbursements, and reasonable attorneys' fees. USDA conducted a regulatory impact analysis (RIA), which concluded that the regulatory benefits outweighed the regulatory costs associated with the implementation of the rule. Based on the RIA, the agency determined that the proposed rule would not have a significant economic impact on a substantial number of small entities. The court stated that, because the agency certified that the rule would "likely not have a significant economic impact on a substantial number of small Medfly host crop producers in the United States," initial and final regulatory flexibility analyses were not needed.¹⁹² It further stated that the certification was supported by an analytical statement including factors such as the relatively low percentage of income derived by small wholesalers from clementine sales, and that small importers and wholesalers would likely be "better off" under the proposed regulations when compared with their status under the current ban on the importation of clementines as well as compared with the less strict conditions imposed before the ban.¹⁹³ The court stated that the agency relied on other analyses supporting its overall conclusion that the rule itself will result in a sufficiently high probability that Medfly infestation will not occur to conclude that any impact the new rule will have on small entities will be positive rather than negative, negating the need for a regulatory flexibility analysis.¹⁹⁴

Size standards

It is important that an agency use the size standard contained in the Small Business Administration's small business size standard regulations,¹⁹⁵ promulgated by the SBA under the Small Business Act, or follow the consultation procedures outlined in section 601(3) of the RFA.

Incorrect size standard. In *Northwest Mining Association v. Babbitt*, discussed above, the court held that BLM violated the RFA because the agency failed to use the appropriate size standard as defined by the Small Business Administration (SBA). The court noted that "the RFA requires agencies to use the Small Business Administration's definition of small entity."¹⁹⁶ Continuing, the court stated that "section 601 of the RFA sets forth, in relevant part, '[f]or the purposes of this chapter ... the term 'small entity' shall have the same meaning as the term 'small business''"¹⁹⁷ The term "small business" has the same meaning as the term "small business concern" under section 3 of the Small Business Act.¹⁹⁸ The SBA publishes these small business definitions in 13 C.F.R. § 121.201. Division B of section 121.201 provides, in pertinent part, that mining

¹⁹² *Id.* at 1087.

¹⁹³ *Id.* at 1115.

¹⁹⁴ *Id.* at 1116.

¹⁹⁵ 13 C.F.R. § 121.201 (1996).

¹⁹⁶ *Northwest Mining Association*, 5 F. Supp. 2d at 15. See Chapter 1 for detail on exceptions to using SBA size standards.

¹⁹⁷ 5 U.S.C. § 601(6).

¹⁹⁸ 15 U.S.C. § 632; 5 U.S.C. § 601(3).

concerns must have 500 or fewer employees to be considered “small.”¹⁹⁹ Therefore, the standard for “small miner” which the BLM must use when performing an initial or final regulatory flexibility analysis or when certifying “no significant impact” is a 500 or fewer employee standard. By using a definition other than the SBA’s, the BLM violated the procedure of law mandated by the statute. The court found that the definitions section of the RFA uses phrases such as “ ‘small entity’ *shall have* the same meaning ...” and “ ‘small business’ *has* the same meaning ...”²⁰⁰ (emphasis added). The court concluded that words such as those do not leave room for alternate interpretations by the agency. The rule was remanded to the agency.

Use of incorrect size standard cured. In *Small Business in Telecommunications v. the Federal Communications Commission (FCC)*,²⁰¹ the FCC adopted its own definition of “small business” regarding its Lower Channel Report and Order concerning a regulatory scheme for specialized mobile radio (SMR) service in the 800 to 900 MHz range. The Court of Appeals for the District of Columbia Circuit held that although the FCC failed to seek approval from the SBA for its definition, the omission did not nullify the entire rulemaking, since SBA did ultimately approve the definition prior to commencement of the lower channel auction.²⁰² If the agency modifies a small business size standard in the implementation of a rule, it must seek approval from the SBA Administrator.²⁰³

The agency must conduct an adequate analysis before certifying

The landmark legal decision recognizing an agency’s failure to adequately examine the impact on affected entities before certification is the 1998 case, *Southern Offshore Fishing Association v. Daley*.²⁰⁴ In that matter, the National Marine Fisheries Service (NMFS) published a proposed rulemaking to institute a 50 percent reduction in the shark fishing industry. NMFS certified that the rule would not have a significant economic impact on a substantial number of small entities. Although the agency published a FRFA at the time it finalized the rule, the court found that the agency certified without making a “reasonable, good-faith effort,” prior to issuance of the final rule, to inform the public about the potential adverse effects of its proposals and about less harmful alternatives. The agency continued to deny that its proposal would likely have a significant impact on a substantial number of small entities after receiving public comments challenging the certification. The court concluded that the preparing of a FRFA constituted “an attempt to agreeably decorate a stubborn conclusion” that there was no significant impact on a substantial number of small entities. The court remanded the agency’s certification determination, requiring it to “undertake a rational consideration of the economic effects and potential [regulatory] alternatives.”²⁰⁵

¹⁹⁹ *Id.*

²⁰⁰ 5 U.S.C. § 601.

²⁰¹ *Small Businesses in Telecomm. v. FCC*, 251 F.3d 1015, 1025 (D.C. Cir. 2001).

²⁰² *Id.*

²⁰³ *Id.* at 1025.

²⁰⁴ *Southern Offshore Fishing*, 995 F. Supp. at 1437.

²⁰⁵ *Id.*

North Carolina Fisheries. The *North Carolina Fisheries* cases provide further guidance on what constitutes adequate analysis prior to certification that there will be no significant economic impact on a substantial number of small entities.

The first case arose in 1997.²⁰⁶ There, the National Marine Fisheries Service (NMFS) set the 1997 quota for flounder fishing by continuing the quota from the previous year. In doing so, NMFS did not perform a regulatory flexibility analysis. Instead, the agency certified that the rule would not have a significant impact on a substantial number of small businesses because the quota remained the same from 1996 to 1997. There was no record showing that the agency did any comparison between conditions in 1996 and 1997. The court stated that “a simple conclusory statement that, because the quota was the same in 1997 as it was in 1996, there would be no significant economic impact, is not an analysis.”²⁰⁷ The court remanded the issue to the agency with orders to “undertake enough analysis to determine whether the quota had a significant economic impact on the North Carolina Fishery.”²⁰⁸ The court further ordered the department to “include in [the] analysis whether the adjusted quota will have a significant economic impact on small entities in North Carolina.”²⁰⁹

The issue returned to the court in 1998.²¹⁰ The issue before the court on remand was whether the Secretary of Commerce had discharged his responsibilities under the RFA and under National Standard 8 of the Magnuson Act to perform an economic analysis.²¹¹ After review, the court concluded that “the Secretary of Commerce acted arbitrarily and capriciously in failing to give any meaningful consideration to the economic impact of the 1997 quota regulations on North Carolina fishing communities. Instead, the Secretary has produced a so-called economic report that obviously is designed to justify a prior determination.”²¹² The court further stated that as part of an adequate analysis before certification, the agency must consider alternatives less burdensome to small entities.²¹³ The court concluded that “Congress has not intended for administrative agencies to circumvent the fundamental purposes of the RFA by invocation of the certification provision.” The court felt that Secretary Daley’s certification in this instance amounted to an effort to avoid the requirements of the RFA, specifically the requirement to consider alternative ways to minimize economic impacts. Because the court found that the Secretary and the agency did not uphold their responsibilities under the law, it set aside the 1997 summer flounder quota and imposed a penalty against the NMFS.

Court cases have held that the agency must account for the public comments it received challenging the initial determination that no significant economic impact was likely.²¹⁴ In

²⁰⁶ *North Carolina Fisheries*, 16 F. Supp. 2d at 647.

²⁰⁷ *Id.* at 653.

²⁰⁸ *Id.*

²⁰⁹ *Id.*

²¹⁰ *North Carolina Fisheries Association v. Daley*, 27 F. Supp. 2d 650 (E.D. Va. 1998).

²¹¹ *Id.* at 660.

²¹² *Id.* at 668.

²¹³ *Id.* at 660.

²¹⁴ See generally, *National Truck Equip. Association v. NHTSA*, 919 F.2d 1148 (6th Cir. 1990); *Northwest Mining Association v. Babbitt*, 5 F. Supp. 2d 9 (D.D.C. 1998).

Northwest Mining Association v. Babbitt,²¹⁵ the court addressed the Bureau of Land Management (BLM) claims that the Northwest Mining Association (NWMA) did not have standing to object to its final rule under either the APA or the RFA because it did not submit comments during the notice and comment period. The NWMA asserted that it did not need to submit comments during the notice and comment period because the BLM's original rule proposal did not properly inform it that its interests were at stake. The court agreed with the NWMA, holding that because there was no way the NWMA could have submitted comments regarding issues on which it was not informed were at stake, the agency must consider even comments not submitted during the formal notice and comment period.²¹⁶

Bare certification not sufficient. In *Theiss v. Principi*,²¹⁷ the Veteran's Administration promulgated an amendment to define "educational institution," excluding home schools. The court determined that this was a substantive, legislative rule and was invalid for failure to comply with notice-and-comment procedures under the APA. The court warned that any future amendment should comply with the APA as well as with the provisions of the RFA and that a "bare certification" like the one in this case would likely be insufficient because it was not accompanied by a "statement providing the factual basis for such certification."²¹⁸

Direct versus indirect impact on small entities

Must the agency consider the indirect effects of the proposed regulation? It was first held in *Mid-Tex Electric Cooperative, Inc., v. Federal Energy Regulatory Commission (FERC)* that a regulatory flexibility analysis is required when an agency determines that the rule will have a significant economic impact on a substantial number of small entities that are subject to the requirements of the rule.²¹⁹ In that case, FERC proposed a rule that allowed electric utilities to include in their rate bases amounts equal to 50 percent of their investments in construction work in progress. In response to an argument that FERC "should have considered the impact of the proposed rule on wholesale and retail customers of the jurisdictional entities subject to rate regulation by the Commission," FERC stated that "the RFA does not require the Commission to consider the effect of this rule, a federal rate standard, on nonjurisdictional entities whose rates are not subject to the rule."²²⁰

The court agreed, reasoning that "Congress did not intend to require that every agency consider every indirect effect that any regulation might have on small businesses in any stratum of the national economy."²²¹ The court concluded that "an agency may properly certify that no regulatory flexibility analysis is necessary when it determines that the rule

²¹⁵ *Northwest Mining Association*, 5 F. Supp. 2d 9.

²¹⁶ *Id.* at 13.

²¹⁷ *Theiss v. Principi*, 18 Vet. App. 204 (2004).

²¹⁸ *Id.* at 214.

²¹⁹ *Mid-Tex Elec. Coop v. FERC*, 773 F.2d 327, 342 (D.C. Cir. 1985).

²²⁰ *Id.* at 341.

²²¹ *Id.*

will not have a significant economic impact on a substantial number of small entities that are subject to the requirements of the rule.”²²²

In viewing this decision, the same court later held in *United Distribution Companies v. FERC*²²³ that an agency is under no obligation to conduct a small entity impact analysis of effects on entities it does not regulate. Because in this case FERC had no jurisdiction to regulate the local distribution of natural gas, it could not be required to conduct a regulatory flexibility analysis for those entities engaged in the local distribution of the gas.²²⁴

Although *Mid-Tex* occurred prior to the passage of the Small Business Regulatory Enforcement Fairness Act (SBREFA) of 1996, post-SBREFA courts have upheld its reasoning. For example, in *Motor and Equipment Manufacturers Association v. Nichols*,²²⁵ the court found that because the deemed-to-comply rule did not subject any aftermarket businesses to regulation, EPA was not required to conduct a regulatory flexibility analysis as to small aftermarket businesses. It was only obliged to consider the impact of the rule on small automobile manufacturers subject to the rule, and it met that obligation. A number of other cases have held similarly.²²⁶

Likewise in *American Trucking Associations v. EPA*,²²⁷ EPA established a primary national ambient air quality standard (NAAQS) for ozone and particulate matter. At the time of the rulemaking, EPA certified the rule pursuant to 5 U.S.C. § 605(b). The basis of the certification was that EPA had concluded that small entities were not subject to the rule because the NAAQS only regulated small entities indirectly through the state implementation plans.²²⁸ Although the court remanded the rule to the agency for non-RFA reasons, the court found that EPA had complied with the requirements of the RFA.

Similarly, in *Michigan v. EPA*,²²⁹ EPA certified that its revised NAAQS would not have a significant economic impact within the meaning of the RFA. According to the EPA, the NAAQS itself imposed no regulations upon small entities. Instead, several states regulate small entities through the state implementation plans they are required by the Clean Air Act to develop. Because the NAAQS regulated small entities only indirectly—that is, insofar as it affected the planning decisions of the states—the EPA concluded that small entities were not “subject to the proposed regulation.” The court agreed, stating that states have broad discretion in determining the manner in which they will achieve compliance with the NAAQS. In conclusion, the court stated that “a State may, if it chooses, avoid

²²² *Id.* at 343.

²²³ *United Dist. Cos. v. FERC*, 88 F.3d 1105, 1170 (D.C. Cir. 1996).

²²⁴ *Id.*

²²⁵ *Motor and Equip. Mfrs. Association v. Nichols*, 142 F.3d 449, 467 (D.C. Cir. 1998).

²²⁶ See *American Trucking Associations v. EPA*, 175 F.3d at 1044; *Michigan v. EPA*, 213 F.3d 663, 689 (D.C. Cir. 2000); *Cement Kiln Recycling Coalition v. EPA*, 255 F.3d 855, 868 (D.C. Cir. 2001).

²²⁷ *American Trucking*, 175 F.3d at 1027.

²²⁸ *Id.*

²²⁹ *Michigan v. EPA*, 213 F.3d at 689.

imposing upon small entities any of the burdens of complying with a revised NAAQS.”²³⁰

The court in *Cement Kiln Recycling Coalition v. EPA*²³¹ further bolstered the notion that indirect impacts should be disregarded by noting that the RFA is not intended to apply to every entity that may be targeted by the proposed regulation. The fact that the rule will have economic impacts in many sectors of the economy does not change this. The court reasoned that “requiring an agency to assess the impact on all of the nation's small businesses possibly affected by a rule would be to convert every rulemaking process into a massive exercise in economic modeling, an approach we have already rejected.”²³²

An entity can otherwise experience indirect impacts through its dealings with the entity that experiences direct impacts, such as through increased after-market prices or newly required modifications to necessary equipment. Some courts have stated that this impact would likewise not require a regulatory flexibility analysis.²³³

In *White Eagle Cooperative Association v. Conner*,²³⁴ the plaintiffs, a cooperative of milk producers, brought action challenging USDA’s amendment of a regional milk marketing order. The United States District Court for the Northern District of Indiana entered summary judgment in government's favor, and the association appealed. Among other things, plaintiffs asserted that in adopting the amendments to the marketing order, USDA violated the RFA by failing to undertake an analysis and by employing the certification option without sufficient factual support. USDA asserted that plaintiffs could not challenge the agency’s RFA compliance because the order regulates handlers—not producers. Since the plaintiffs are an association of producers, not handlers, USDA argued that plaintiffs lacked standing to challenge the agency’s compliance. The court held that the association did not have standing to raise a challenge under the RFA because the impact was indirect.

Where an agency argued indirect impact unsuccessfully. In *Aeronautical Repair Station Association v. F.A.A.*²³⁵, the plaintiff challenged a final rule of the Federal Aviation Administration (FAA) which amended its drug and alcohol testing regulations to expressly mandate that air carriers require drug and alcohol tests of all employees of its contractors, including employees of subcontractors at any tier, who perform safety-related functions such as aircraft maintenance. The plaintiff challenged the rule on the

²³⁰ *Id.*; see also, *Nat’l Women, Infants, & Children Grocers Association v. Food & Nutr. Serv.*, 416 F. Supp. 2d 92 (D.D.C. 2006) where the court granted summary judgment to the defendant and denied the plaintiffs’ motion for summary judgment, holding the FNS certification proper because the interim rule regulated state agencies—the impact on small businesses was indirect. Furthermore, the court bolstered its reasoning by citing the fact that FNS stated in the *Federal Register* that it planned to use data collected from the interim rule to strengthen its ultimate FRFA.

²³¹ *Cement Kiln*, 255 F.3d at 868.

²³² *Id.*

²³³ See, e.g., *Nichols*, 142 F.3d at 467; *Cement Kiln*, 255 F.3d at 868.

²³⁴ *White Eagle Cooperative Association, et al., v. Charles F. Conner, Acting Secretary, United States Department of Agriculture*, 553 F.3d 467 (7th Cir. 2009).

²³⁵ *Aero. Repair Station Association v. F.A.A.* 494 F.3d 161 (D.C. Cir. 2007).

grounds that it impermissibly expanded the scope of employees tested in violation of the unambiguous statutory language of § 45102(a)(1), the Administrative Procedure Act, 5 U.S.C. §§ 701-06, and the Fourth and Fifth Amendments to the United States Constitution. In addition, it challenged the FAA's conclusion that it was not required to conduct a regulatory flexibility analysis. The court upheld the substance of the rule but rejected the FAA's RFA determination.

In the NPRM, the FAA performed a tentative RFA analysis and counted among RFA small entities both air carriers and Part 145 repair stations because it could not determine how many of the 2,412 Part 145 repair stations are considered small entities. In the second NPRM, the FAA determined that the small entity group is considered to be Part 145 repair stations, but it still could not determine how many of the Part 145 repair stations and their subcontractors were considered small entities. The FAA concluded that most, if not all of the noncertificated maintenance contractors would be considered small entities. Based on its calculation of annualized costs of less than 1 percent of annual median revenue, the FAA certified that the proposed action would not have a significant economic impact on a substantial number of small entities.

Although commentators raised RFA issues, in the final rule FAA disagreed and asserted that contractors were not among entities regulated under the testing regulations for the purpose of the RFA. The FAA noted that the directly regulated employers were air carriers operating under 14 CFR Parts 121 and 135, § 135.1(c) operators, and air traffic control facilities not operated by the FAA or under contract to the U.S. military, who must conduct drug and alcohol testing under the FAA regulations. For drug and alcohol testing purposes, certificated repair stations were contractors, and contractors were not regulated employers. Accordingly, the FAA concluded it was not required to conduct an RFA analysis, including considering significant alternatives, because contractors (including subcontractors at any tier) were indirectly regulated entities.

In making its determination, the FAA relied on the *Mid-Tex* case and other cases that held that under the RFA the regulating agency need consider only the economic impact on businesses directly affected and regulated by the subject regulations. The plaintiffs asserted that the FAA's determination was incorrect. The court found that, unlike the parties claiming economic injury in the cited cases, the contractors and subcontractors were directly affected and therefore regulated by the challenged regulations. Although the regulations immediately addressed the employer air carriers which were in fact the parties certified to operate aircraft, the regulations expressly required that the employees of contractors and subcontractors be tested. Thus, the contractors and subcontractors (at whatever tier) were entities subject to the proposed regulation.

The FAA also asserted that it had substantially complied with the RFA because it conducted initial evaluations and a final economic evaluation of the effects on the industry, responding to comments following the proposal. The court found that the final evaluation was not a FRFA because the FAA determined that contractors and subcontractors are not regulated entities for the purpose of the RFA. In addition, the FAA did not consider alternatives as required by the RFA. The court upheld the substance of

the FAA's 2006 final rule and remanded for the limited purpose of conducting the analysis required under the RFA, treating the contractors and subcontractors as regulated entities.

The initial regulatory flexibility analysis

Because an agency's initial regulatory flexibility analysis cannot be the subject of litigation,²³⁶ case law provides a detailed discussion only for the final regulatory flexibility analysis. It is important to note that although the IRFA is not judicially reviewable, a proper IRFA is necessary to provide the foundation for a good FRFA. An agency cannot develop an adequate FRFA if the IRFA did not lay the proper foundation for eliciting public comments and seeking additional economic data and information on the regulated industry's profile and regulatory impacts. Further, without an adequate IRFA, small entities cannot provide informed comments on regulatory alternatives that are not adequately addressed in the IRFA.²³⁷

In *Allied Local and Regional Manufacturers Caucus v. EPA*, paint manufacturers and associations of manufacturers and distributors of architectural coatings petitioned for review of EPA's regulations limiting the content of volatile organic compounds (VOCs) in consumer and commercial products such as architectural coatings, including paints.²³⁸ Plaintiffs alleged that EPA failed to comply with the RFA by failing to discuss the economic impact of "stigmatic harm" arising from the agency's suggestion that it may impose more stringent VOCs in the future, and of asset devaluation, in that the coatings rule allegedly will render existing product formulas valueless. The court ruled that section 603 of the RFA, which discusses IRFAs, was not subject to judicial review pursuant to section 611(c). However, the court did have the jurisdiction to determine whether the agency had met the overall requirement that the decisionmaking not be arbitrary and capricious. The court found that the EPA examined alternatives to product reformulation when creating regulations limiting content of VOCs in consumer and commercial products, and that its decisions were neither arbitrary nor capricious. The court, therefore, found that EPA had met its obligations under the RFA.

Similarly, in *U.S. Cellular Corp. v. FCC*,²³⁹ the court noted that an IRFA is not subject to judicial review. There, the FCC adopted an order requiring wireless carriers to bear financial responsibility for enhanced 911 implementation, rather than having local government guarantee costs. Plaintiffs argued that the FCC failed to issue an IRFA and that the FRFA did not contain a description of the steps the agency took to minimize the impact on small businesses, as required by the RFA. The court held that the RFA expressly prohibits courts from considering claims of noncompliance with RFA section 603's requirement to issue an IRFA.²⁴⁰

²³⁶ Because § 611 of the RFA does not mention § 603, the IRFA requirement, a court would consider a pre-promulgation challenge unripe.

²³⁷ *Southern Offshore Fishing*, 995 F. Supp. at 1434 and 1436 ("the agency could not possibly have complied with § 604 by summarizing and considering comments on an IRFA that NMFS never prepared").

²³⁸ *Allied Local and Reg'l Mfrs. Caucus v. EPA*, 215 F.3d 61 (D.C. Cir. 2000).

²³⁹ *U.S. Cellular Corp. v. FCC*, 254 F.3d 78, 89 (D.C. Cir. 2001).

²⁴⁰ *Id.*

The final regulatory flexibility analysis

General content

Section 604 of the RFA prescribes the content of the FRFA. Courts have found that an agency can satisfy the requirements of section 604 “as long as it compiles a meaningful, easily understood analysis that covers each requisite component dictated by the statute and makes the end product readily available to the public.”²⁴¹ For example, in *Associated Fisheries of Maine, Inc., v. Daley*, the court stated that the Secretary of Commerce had complied with FRFA requirements because the Secretary explicitly considered numerous alternatives, exhibited a fair degree of sensitivity concerning the need to alleviate the regulatory burden on small entities within the fishing industry, adopted some salutary measures designed to ease that burden, and satisfactorily explained reasons for adopting others. Similarly, in *Alenco Communications v. FCC*,²⁴² the court held that the regulatory analysis was compliant with the terms of the RFA where the agency provides a lengthy analysis of the economic impact of the proposed rule on small businesses and responds to comments submitted by the Office of Advocacy and other commenters.²⁴³

The court addressed the issue in *National Association of Mortgage Brokers v. Board of Governors of the Federal Reserve System*.²⁴⁴ In that case, the National Association of Independent Housing Professionals, Inc. (NAIHP) and the National Association of Mortgage Brokers (NAMB) “requested the Court to issue a temporary restraining order and preliminary injunction to enjoin the Board of Governors of the Federal Reserve System [Board] from implementing a Final Rule . . . that restricts certain compensation practices of loan originators relating to mortgage loans.” Among other claims, the Plaintiffs argued that the Board failed to comply with the RFA because they “[1] failed to provide a statement of need for or objectives of the rule; [2] failed to meaningfully analyze the Rule’s impact on small businesses; [3] failed to respond to public comments; and [4] failed to analyze alternatives to the proposed regulation.” The court disagreed, finding that the FRFA stated that the rule addressed problems that have been observed in the mortgage market in order to prohibit unfair and deceptive acts and practices in connection with mortgage loans, and recognized that the rule would have a “significant economic impact on a substantial number of small entities but the precise compliance costs were difficult to ascertain. The FRFA also discussed and rejected alternatives from public comments. The court stated that the Board did not need to address each portion of the rule challenged in the comments because it “addressed the effects of all of the Rule’s prohibitions regarding loan originator compensation collectively, and this satisfies the

²⁴¹ *Associated Fisheries of Maine, Inc., v. Daley*, 127 F.3d 104, 115 (1st Cir. 1997); *Grand Canyon Air Tour Coalition v. FAA*, 154 F.3d 455, 470 (D.C. Cir. 1998); *National Propane Gas Association v. DOT*, 43 F. Supp. 2d 665, 681 (N.D. Tex. 1999); *Associated Builders and Contractors, Inc., v. Herman*, 976 F. Supp. 1 (D.D.C. 1997).

²⁴² *Alenco Communications v. FCC*, 201 F.3d 608 (5th Cir. 2000).

²⁴³ *Id.* at 625.

²⁴⁴ *National Association of Mortgage Brokers v. Board. Of Governors of the Federal Reserve System*, 773 F. Supp. 2d 151 (D.D.C. 2011).

Board's obligations under 5 U.S.C. § 604(a).” In making its ruling the court reiterated that the RFA's requirements are purely procedural and although it directs agents to state, summarize, and describe, the RFA in and of itself imposes no substantive constraints on agency decisionmaking. Moreover, the agency does not need to present its FRFA in any particular mode of presentation, as long as the FRFA compiles a meaningful, easily understood analysis that covers each requisite component dictated by the statute and makes the end product readily available to the public. Finally, the court noted that failure to comply with the RFA may be, but does not have to be, grounds for overturning a rule. Additionally, while making a section 604 challenge, parties may raise the related but distinct claim that an agency did not reasonably address the rule's impact on small businesses and such challenges are evaluated under the arbitrary and capricious standard of review.²⁴⁵

Is a FRFA always required?

A FRFA is required in every instance where an agency finalizes a rule after being required to publish a general notice of proposed rulemaking under section 553 of the APA or any other law. The exception is when the agency certifies the rule will not have a significant economic impact on the affected entities, as discussed above.

However, in the event that the publication of an NPRM is impossible due to the emergency nature of the rule, the requirements of the RFA may be satisfied by publishing a FRFA subsequent to the rulemaking.²⁴⁶ In *National Propane Gas Association v. DOT*,²⁴⁷ the Department of Transportation's Research and Special Programs Administration (RSPA) instituted an emergency interim final rule to address concerns about the transportation of compressed gas on highways. RSPA later modified and adopted the interim final rule as the emergency discharge control regulation for loading or unloading of cargo tank motor vehicles. The regulation required vehicle operators to shut down immediately if they learned of a gas leakage.

Gas companies brought suit alleging various violations of the APA and RFA. Plaintiffs challenged the rule on the grounds that defendants failed to prepare a FRFA, as required by the RFA. RSPA argued that the rule was not subject to the RFA because the RFA applies only to the rules for which an agency is required to publish a notice of proposed rulemaking pursuant to section 553 of the APA. RSPA asserted that the APA did not require a notice of proposed rulemaking here because of the emergency nature of the rule. Nevertheless, RSPA claimed that in preparing preliminary and final regulatory evaluations under Executive Order 12,866, the agency did analyze the impact of the interim final rule and the final rule on all affected parties, including small businesses. The court agreed, and found that although the agency did not prepare a FRFA, all of the elements of a FRFA were available throughout their summary of such analysis published in the *Federal Register*. The court thus found that RSPA complied with each of the requirements found in the RFA, including responding to comments and consideration of

²⁴⁵ *Id.* at 178.

²⁴⁶ *National Propane Gas Association*, 43 F. Supp. 2d at 681.

²⁴⁷ *Id.*

alternatives. The court asserted that a preliminary regulatory evaluation was available in the docket for the public to provide comment, and it also found that to require an additional analysis by the agency would be duplicative.

Considering alternatives to the final rule

Section 604 of the RFA requires the agency to consider alternatives that would achieve the statutory objectives while lessening the regulatory burden on affected small entities. This involves making a “reasonable, good-faith effort to canvass major options and weigh their probable effects.”²⁴⁸

In *AML International, Inc., v. Daley*,²⁴⁹ the National Marine Fisheries Service implemented a management plan for the spiny dogfish industry that imposed quotas that effectively shut down the industry for the next five years. The plaintiffs asserted that NMFS failed to comply with the RFA because the NMFS failed to consider alternatives. The court found that NMFS’s consideration of alternatives was sufficient. NMFS considered and rejected alternatives because they did not meet the mandate of the Magnuson-Stevens Act or provide long-term economic benefits greater than those of the proposed action.²⁵⁰

Similarly, in *Ace Lobster Co. v. Evans*,²⁵¹ the Department of Commerce imposed limitations on the number of lobster traps that could be used in a particular area. Lobster fisherman and business owners alleged that the Department of Commerce implemented the regulations in violation of the APA, the Magnuson-Stevens Act, and the RFA. The basis for the assertion was that during the comment period, numerous commenters submitted information about an alternative plan for the lobster fishery, which was approved by the Lobster Conservation and Management Team and submitted for consideration as an alternative. The agency rejected the alternative because it would likely increase the number of lobster traps in offshore waters and increase the lobster mortality rate. Plaintiffs alleged that the defendant did not adequately analyze the selected alternative or consider the alternative that would mitigate the negative economic impacts on offshore fishing fleets, and that the agency’s concern for verification of prior fishing fleets was unfounded.²⁵² The court stated that under the standard for judicial review of compliance with the RFA, the court reviews only whether the agency conducted a complete IRFA and FRFA in which it described steps to minimize the economic impact of its regulations on small entities and discussed alternatives, providing a reasonable explanation for rejections. The RFA permits the agency to select an alternative that is more economically burdensome if there is evidence that other alternatives would not accomplish the objectives of the statute. Because the agency examined the alternative and decided that, while less onerous, it did not achieve the

²⁴⁸ National Association of Psychiatric Health Sys. v. Shalala, 120 F. Supp. 2d 33, 42 (D. D.C. 2000).

²⁴⁹ *AML Int’l v. Daley*, 107 F. Supp. 2d 90 (D. Mass. 2000).

²⁵⁰ *Id.* at 105.

²⁵¹ *Ace Lobster Co. v. Evans*, 165 F. Supp. 2d 148 (D. R.I. 2001).

²⁵² *Id.* at 185.

conservation goals, it met its obligations under the RFA. The court further found that there was sufficient analysis and explanation of the other rejected alternatives.²⁵³

What kinds of alternatives must the agency consider? In *Associated Fisheries of Maine*, the court first held that section 604 does not require that a FRFA address every alternative, only significant ones.²⁵⁴ The RFA does permit the agency to select an alternative that is more economically burdensome if there is evidence that other alternatives would not accomplish the stated objectives of the applicable statutes.²⁵⁵

What is a significant alternative? This question was clarified by the court in *Little Bay Lobster Co. v. Evans*.²⁵⁶ There, the court stated that “significant alternatives” are those with potentially lesser impacts on small entities (versus large-scale entities) as a whole, and not those that may lessen the regulatory burden on some particular small entity. Further, the agency is not obligated under the RFA to address alternatives that might have had lesser impacts on some small entities vis-à-vis other similarly affected small entities.²⁵⁷

In *Hall v. Evans*,²⁵⁸ the Department of Commerce determined that the monkfish fishery was overfished. To address the problem, the agency implemented a fishery management plan to prescribe landing limits for vessels holding limited access monkfish permits. The limits allowed categories A and C vessels using trawl gear to land up to 1,500 pounds of monkfish tailweight per day at sea, while vessels using any gear other than trawl or “mobile” gear may land up to 300 pounds of monkfish tailweight per day at sea. The plaintiffs filed suit asserting that the regulations violated the Magnuson Act and the RFA. The plaintiffs asserted that the defendant’s RFA analysis: (1) failed to recognize the costs of forcing closures of the directed monkfishing industry within 4 years, supposedly to allow the industry to receive positive revenue benefits after 20 years; (2) forced particularly harsh consequences on small businesses; and (3) failed to conduct an assessment of meaningful and more gradual restrictions in order to avoid severe costs to small businesses. Plaintiffs asserted that neither the IRFA nor the FRFA provided an assessment of the real economic impact on small entities in that the IRFA failed to assess the number and quality of vessels affected by the regulations and failed to address the disparity in landing allocations between different gear types. Although the regulations were set aside for violation of the Magnuson Act, the court found no violation of the RFA. With respect to the RFA allegations, the court found that there was enough evidence in the IRFA to show that the defendants considered both the economic effect of the fishery plan as a whole upon small entities and less onerous alternatives.²⁵⁹

²⁵³ *Id.*

²⁵⁴ *Associated Fisheries of Maine*, 127 F.3d at 115; See also *Grand Canyon Air Tour Coalition*, 154 F.3d at 470 and *Blue Water Fishermen’s Association v. Mineta*, 122 F. Supp. 2d 150, 178 (D. D.C. 2000).

²⁵⁵ *Associated Fisheries of Maine*, 127 F.3d at 114.

²⁵⁶ *Little Bay Lobster Co v. Evans*, 2002 WL 1005105, Slip. Op. (D. N.H. May 16, 2002).

²⁵⁷ *Id.* at 25.

²⁵⁸ *Hall v. Evans*, 165 F. Supp. 2d 114 (D.R.I. 2001).

²⁵⁹ *Id.* at 147.

What kind of description of the alternatives considered must the agency include in the FRFA? The RFA requires a statement of the factual, policy, and legal reasons for selecting the alternative adopted by the final rule and why each one of the other significant alternatives to the rule considered by the agency that affect the impact on small entities was rejected.

In *Ashley County Medical Center v. Thompson*,²⁶⁰ the Department of Health and Human Services imposed upper payment limit (UPL) regulations that would reduce the upper limit on what states could reimburse locally owned public hospitals for services to Medicaid beneficiaries. The plaintiffs alleged that the FRFA failed to describe the steps the agency had taken to minimize the significant economic impact on hospitals, and failed to discuss any affirmative steps the agency had taken or intended to take to mitigate the injury that the 2002 UPL rule would cause to public hospitals. The court, noting that the RFA requires only that the agency describe steps taken and not that the agency take any particular steps, stated that if there were no steps that could have been taken to minimize the impact on small businesses, then the statutory requirement would have been met simply by reporting that information. The court noted that the agency had provided a description of the alternatives considered and rejected in the *Federal Register*, and thus all the requirements of the RFA were clearly satisfied.²⁶¹

Conversely, in *Nat'l Assoc. of Psychiatric Health Sys. v. Shalala*,²⁶² the plaintiffs challenged an interim final rule promulgated by the Department of Health and Human Services (HHS) that required a face-to-face evaluation of patients within one hour after the patient has been placed in restraints or seclusion. The plaintiffs argued that the Secretary failed to conduct an adequate analysis before adopting the one-hour provision. The court agreed with the plaintiffs, stating that it could not find that the Secretary made a good-faith effort to canvass major alternatives and weigh their probable effects.²⁶³ Specifically, the Secretary did not obtain data or analyze available data on the impact of the final rule on small entities, nor did she properly assess the impact the final rule would have on small entities. The court stated that by these omissions the Secretary totally failed to comply with section 5 of section 604(a) of the RFA.²⁶⁴ The court thus remanded the matter to HHS for completion of a compliant FRFA.²⁶⁵

However, in *Southern Offshore Fishing Ass'n v. Daley*,²⁶⁶ the court stated that the agency's consideration of alternatives was inadequate. Particularly troublesome to the

²⁶⁰ *Ashley County Med. Ctr. v. Thompson*, 205 F. Supp. 2d 1026 (E.D. Ark. May 13, 2002).

²⁶¹ *Id.* See also *Nat'l Coal. for Marine Conservation v. Evans*, 231 F. Supp. 2d 119 (D.D.C. 2002) where plaintiffs argued that the Florida Closure violated the RFA, alleging that the DOC's analysis of the economic, social, and environmental effects of the closure as well as alternatives to minimize harm impacts Florida's fishing communities was flawed or superficial. The court held that the DOC considered alternatives, and granted the DOC's motion for summary judgment.

²⁶² *National Association of Psychiatric Health Sys. v. Shalala*, 120 F. Supp. 2d 33 (D.D.C. 2000).

²⁶³ *Id.* at 44.

²⁶⁴ *Id.* Note that because of renumbering resulting from an added provision in the RFA, section 5 is now section 6.

²⁶⁵ *Id.* at 42.

²⁶⁶ *Southern Offshore Fishing Association v. Daley*, 995 F. Supp. 1411 (M.D. Fla. 1998).

court was the “agency’s apparently superficial analysis of less restrictive alternatives to the quota reduction. After extensive discussion and summary of its statistical modeling, [the agency’s] report devotes only four of fifty pages to considering potential alternatives.”²⁶⁷

Exceptions to the requirement of considering alternatives:

- Where uniform requirements are mandated by statute, a statement to that effect by the implementing agency obviates the need to solicit or consider proposals which include differing compliance standards.²⁶⁸
- Where the Secretary is not granted the authority to examine alternatives in implementing the regulation.²⁶⁹

Analysis of the economic impact

What type of analysis must the agency conduct? It is now well established that the RFA does not require an economic modeling, per se.²⁷⁰ Rather, the RFA mandates only that the agency describe the steps it took “to minimize the economic impact on small entities consistent with the stated objectives of applicable statutes.”²⁷¹ Neither cost-benefit analysis nor economic modeling is specifically required,²⁷² as long as the agency compiles a meaningful, easily understood analysis that covers each requisite component dictated by the statute and makes the end product—whatever form it reasonably may take—readily available to the public.²⁷³ However, such an examination may be required by the underlying statute or E.O. 12,866, working in concert with the RFA.

An agency can satisfy the requirements of an economic impact analysis by providing either a quantifiable or numerical description of the effects of a proposed rule or alternatives to the proposed rule, or more general descriptive statements if quantification is not practicable or reliable.²⁷⁴ Courts have stated that sufficient analysis and

²⁶⁷ Southern Offshore Fishing Association, at 1437.

²⁶⁸ Greater Dallas Home Care Alliance v. United States, 10 F. Supp. 2d 638, 648 (N.D. Tex. 1998).

²⁶⁹ Greater Dallas Home Care Alliance v. United States, 36 F. Supp. 2d 765, 769 (N.D. Tex. 1999).

²⁷⁰ Alenco Communications, 201 F.3d at 625; see also Ashley County Med. Ctr 205 F. Supp. 2d at 1026; and Ace Lobster, 165 F. Supp. 2d at 184.

²⁷¹ Alenco Communications, 201 F.3d at 625.

²⁷² See National Telephone Cooperative Association v. Federal Communications Commission and United States of America, 563 F.3d 536, 385 U.S. App. D.C. 327, 47 Communications Reg. (P&F) 985 (C.A. D.C. 2009), where the court reiterated its previous finding that the RFA’s requirements are “purely procedural.” Though it directs the agencies to state, summarize, and describe, the act in and of itself imposes no substantive constraint on agency decisionmaking. The RFA requires agencies to publish analyses that address certain legally delineated topics. Because the analysis at issue addressed all of the legally mandated subject areas, it complied with the RFA. See also Association of American Physicians & Surgeons v. H.H.S., 224 F. Supp 2d 1115 (S.D. Tex. 2002).

²⁷³ National Association of Mortgage Brokers at 178.

²⁷⁴ Alenco Communications, 201 F.3d at 625.

explanations for the rejection of alternatives are all that is necessary to satisfy this requirement.²⁷⁵

Where the majority of businesses likely to experience impacts are deemed small, it follows that any attempt to reduce the adverse economic impacts of a regulation aimed at them is necessarily an attempt to minimize the negative effects of the regulation on small business.²⁷⁶

What is the relevant economic impact that agencies should consider? For the purpose of flexibility analysis, the relevant economic “impact” is the impact of compliance.²⁷⁷

The RFA requires only that the agency consider the economic effect on the entity, and not the effect on specific revenue earned.²⁷⁸ This means that the agency need not consider how one particular element of the affected entity’s business is affected. Rather, the agency should evaluate the regulation’s entire effect.

What type of information should the agency consider? The agency should consider economic data and information regarding the regulated industry’s profile and the anticipated regulatory impacts. The agency needs to consider the scope of the problem and the small business contribution to that problem. If necessary, the agency should seek additional information of this type through public comments, outside research, stakeholder meetings, etc.

It is important that the agency appropriately consider all relevant information. It has been held that although an agency has considerable discretion to act on the basis of less than perfect information when performing the analysis of the rule’s economic impact on small entities, it is not permissible to omit known information in order to skew the results.²⁷⁹

In *North Carolina Fisheries Ass’n v. Daley*,²⁸⁰ the court examined the agency’s economic analysis. In performing the analysis, the Secretary of Commerce utilized criteria employed internally by the National Marine Fisheries Service (NMFS) in evaluating the economic impacts of regulations under the RFA. Thus, the Secretary considered the following criteria:²⁸¹

Criterion 1: Does the action result in revenue loss of more than 5 percent for 20 percent or more of the participants?

Criterion 2: Does the action result in 2 percent of the entities ceasing operations?

²⁷⁵ *Ace Lobster*, 165 F. Supp. 2d at 185.

²⁷⁶ *Associated Fisheries of Maine*, 127 F.3d at 115.

²⁷⁷ *Mid-Tex*, 773 F.2d at 342.

²⁷⁸ *Washington v. Daley*, 173 F.3d 1158, 1170 (9th Cir. 1999).

²⁷⁹ *North Carolina Fisheries*, 27 F. Supp. 2d at 660.

²⁸⁰ *Id.*

²⁸¹ It should be noted that NMFS no longer uses these criteria for its RFA analyses.

Based on the NMFS's internal guidelines, the Secretary found that there would be no significant economic impact on a substantial number of small businesses arising from the 1997 summer flounder quota. In making this determination, the economic analysis used the total number of vessels to be issued moratorium permits as “the universe for the evaluation of impacts.” The small entities or communities studied constituted the whole state of North Carolina. Examining the unadjusted 1997 quota first, the economic analysis stated that it was “possible” that criterion 1 would be triggered by reducing the income of more than 20 percent of the entire North Carolina fleet by more than 5 percent. The economic analysis next considered the NMFS's criterion under the initial 1997 quota adjustment. Under the adjustment, the economic analysis determined that 57 percent of the vessels with home ports in North Carolina are projected to have revenue reductions of greater than 5 percent. The economic analysis further maintained that an additional 43 percent of North Carolina's flounder fleet may have reduced revenues by 25 percent or more. Despite this assessment, the economic analysis concluded that there were no significant economic impacts and asserted that any adverse effects arising from the initial 1997 quota adjustment were offset by previous revenues the fishermen had earned from overfishing.²⁸² The court concluded that the Secretary prepared an economic analysis utterly lacking in compliance with the requirements of the RFA. In the first place, the Secretary did not consider a community any smaller than the entire state of North Carolina. In the second place, the Secretary completely ignored readily available data that would have shown the number of fishing vessels likely to experience the impacts of the agency's regulatory actions. The agency's economic analysis indicating that there would be no significant economic impact on a substantial number of small entities was the result of impermissibly considering too large a community and ignoring readily available data.²⁸³

Public comments

Ordinarily, an agency must seek public comments regarding each proposal and the basis for the agency's decision in each case. The agency must be responsive to the comments it receives, accounting for the dismissal of significant alternatives proposed in the IRFA or by the commenters. Failure to seek public comments or to be responsive frustrates important public participation and will result in a breach of the RFA. An agency might consider eliciting information such as additional economic data, or information regarding the regulated industry's profile and regulatory impacts through public comments.

Must an agency always seek public comment? An agency need not seek comment on information that is supplementary to the decision. That is to say, an agency is entitled to rely on information not exposed to comment only as long as it is not substantially related to the agency's rationale.²⁸⁴ Any information relied on in the analytical process at all, however, must be included in the IRFA.

²⁸² North Carolina Fisheries, 27 F. Supp. 2d at 660.

²⁸³ Id.

²⁸⁴ National Mining Association v. Chao, 160 F. Supp. 2d 47, 88 (D.D.C. 2001).

Judicial review

The 1996 SBREFA amendment provides, for the first time, for judicial review of agency action under the RFA and allows the Chief Counsel for Advocacy to file as *amicus curiae* (friend of the court) in regulatory appeals. “In any such action, the Chief Counsel is authorized to present his or her views with respect to compliance with this chapter, the adequacy of the rulemaking record with respect to small entities, and the effect of the rule on small entities.”²⁸⁵ The standard of review is whether the agency acted in a manner that was arbitrary and capricious.²⁸⁶

²⁸⁵ 5 U.S.C. § 612(b).

²⁸⁶ National Association of Mortgage Brokers v. Board of Governors at 178.

CHAPTER 6 SECTION 610 REVIEW OF EXISTING RULES

Section 610 of the Regulatory Flexibility Act²⁸⁷ requires federal agencies to review regulations that have a significant economic impact on a substantial number of small entities²⁸⁸ within 10 years of their adoption as final rules. In Executive Orders 13,563 and 13,579, President Obama reaffirmed the need for agencies to carry out retrospective analyses of existing rules.²⁸⁹ For example, Executive Order 13,563 says that:

Within 120 days of the date of this order, each agency shall develop and submit to the Office of Information and Regulatory Affairs a preliminary plan, consistent with law and its resources and regulatory policies, under which the agency will periodically review its existing significant regulations to determine whether any such regulations should be modified, streamlined, expanded, or repealed so as to make the agency's regulatory program more effective or less burdensome in achieving the regulatory objectives.²⁹⁰

Executive Order 13,579 reiterates the provisions relating to retrospective analyses of existing rules, noting that independent agencies should, within 120 days of the date of the order, develop and release to the public a plan as described in E.O. 13,563.²⁹¹ President Obama issued a memorandum with the executive order, asking the independent agencies to reassess their regulations and to follow the principles of E.O. 13,563.

These periodic rule reviews are a mechanism for agencies to assess the impact of existing rules on small entities and to determine whether the rules should be continued without change, or should be amended or rescinded, consistent with the objectives of applicable statutes. Agency compliance with section 610's periodic review requirement has varied substantially from agency to agency since 1980.²⁹² While some agencies systematically review all of their existing rules, other agencies review few, if any, of their current rules. Agencies also vary considerably in the amount of public involvement they allow, and the amount of information they provide to the public about their reviews.

Statements made during the 1980 debate on the Regulatory Flexibility Act demonstrate that Congress intended for section 610 to be a mechanism that requires agencies to

²⁸⁷ 5 U.S.C. § 610 (2000).

²⁸⁸ "Small entities" include small businesses that meet the Small Business Administration size standard for small business concerns at 13 C.F.R. § 121.201, small governmental jurisdictions with a population of less than 50,000, and small organizations that are independently owned not-for-profit enterprises and which are not dominant in their field. See 5 U.S.C. §§ 601(3)-(5).

²⁸⁹ Exec. Order No. 13,563 § 6 and Exec. Order No. 13,579 § 2.

²⁹⁰ Exec. Order No. 13,563 § 6(b).

²⁹¹ Exec. Order No. 13,579 § 2(b).

²⁹² See, for example, Government Accountability Office, *Reexamining Regulations: Opportunities Exist to Improve Effectiveness and Transparency of Retrospective Reviews* (GAO-07-791), July 2007; General Accounting Office, *Regulatory Flexibility Act: Agencies' Interpretations Vary* (GAO/GGD-99-55) April 1999. See also Michael R. See, *Willful Blindness: Federal Agencies' Failure to Comply with the Regulatory Flexibility Act's Periodic Review Requirement – and Current Proposals to Invigorate the Act*, 33 *Fordham Urb. L.J.* 1199-1255 (2006).

periodically re-examine the regulatory burden of their rules vis-à-vis small entities, considered in the light of changing circumstances.²⁹³ This view was also reflected in Advocacy’s initial 1982 guidance explaining the then-new RFA, which stated that

The RFA requires agencies to review all existing regulations to determine whether maximum flexibility is being provided to accommodate the unique needs of small businesses and small entities. Because society is not static, changing environments and technology may necessitate modifications of existing, anachronistic regulations to assure that they do not unnecessarily impede the growth and development of small entities.²⁹⁴

RFA section 610 review

The objective of a section 610 review is like the goal of many other retrospective rule reviews:²⁹⁵ to determine whether an existing rule is actually working as it was originally intended and whether revisions are needed. Has the problem the rule was designed to address been solved? Are regulated entities (particularly small entities) able to comply with the rule as anticipated by the agency? Are the costs of compliance in line with the agency’s initial estimates? Are small businesses voicing continuing concerns about the difficulty they have complying with the rule? The section 610 review is an excellent way to address these questions.

Review of rules that were originally certified

In some cases, even if an agency was originally able to certify properly under section 605 of the RFA that a rule would not have a significant economic impact on a substantial number of small entities,²⁹⁶ changed conditions may mean that the rule now does have a significant impact and therefore should be reviewed under section 610. For

²⁹³ House Debate on the Regulatory Flexibility Act, 142 Cong. Rec. H24,575, H24,583-585 (daily ed. Sept. 8, 1980) (“At least once every 10 years, agencies must assess regulations currently on the books, with a view toward modification of those which unduly impact on small entities.” (Statement of Rep. McDade)) (“[A]gencies must review all regulations currently on the books and determine the continued need for any rules which have a substantial impact on small business.” (Statement of Rep. Ireland)). Similarly, the section-by-section analysis of the periodic review provision of S. 299, which became the RFA, notes that the required factors in a section 610 review mirror the evaluative factors in President Carter’s Executive Order 12,044, *Improving Government Regulations*. Exec. Order 12,044, 43 Fed. Reg. 12,661 (March 24, 1978). Pursuant to that Executive Order, President Carter issued a Memorandum to the Heads of Executive Departments and Agencies in 1979, further instructing federal agencies: “As you review existing regulatory and reporting requirements, take particular care to determine where, within statutory limits, it is possible to tailor those requirements to fit the size and nature of the businesses and organizations subject to them.” President Jimmy Carter, Memorandum to the Heads of Executive Departments and Agencies, November 16, 1979.

²⁹⁴ Office of Advocacy, *The Regulatory Flexibility Act* (October 1982).

²⁹⁵ Typical agency-initiated retrospective regulatory reviews include post-hoc validation studies, reviews conducted pursuant to petitions for rulemaking or reconsideration, paperwork burden reviews, and reviews undertaken to advance agency policies.

²⁹⁶ See 5 U.S.C. § 605(b).

example, many more small businesses may be subject to the rule now than when the rule was promulgated. The cost of compliance with a current rule may have increased sharply because of a required new technology. If there is evidence (such as new cost or burden data) that a rule is now having a significant economic impact on a substantial number of small entities, including small communities or small nonprofit organizations, Advocacy believes that the agency should conduct a section 610 review.

Advocacy is aware that some agencies interpret section 610 not to require the periodic review of rules that were originally certified when they were promulgated as having no significant economic impact on a substantial number of small entities. This narrow interpretation of the section 610 review requirements discounts several important considerations. First, evidence of significant current impacts to small entities from an existing rule may call into question the accuracy of the original determination that the rule would have no significant impact. Second, as time passes and the agency (along with regulated small entities) is better able to measure and understand the impacts of a regulation, it benefits the agency to use the periodic review process to update their rules and perform regulatory “housekeeping.” Third, limiting section 610 reviews only to rules that were found to have a significant economic impact on a substantial number of small entities at the time of promulgation would severely undercut the intent of section 610. EPA and OSHA, for example—which between them determine that at most one or two rules each year will have such an impact—will exclude from section 610 review each of the hundreds of other rules promulgated annually that may now have a significant impact on small entities. Given the legislative history of section 610, it is very difficult to believe that Congress intended this outcome. Finally, a reading of the plain language of section 610 supports Advocacy’s interpretation. If Congress meant to limit periodic reviews, it would have simply required agencies to review rules that originally had a significant impact, rather than rules that now have a significant impact.

An agency may learn about the current impacts of an existing rule through complaints from small entities or petitions for a section 610 or other retrospective review of the rule. If these complaints and/or petitions are founded on reliable cost and impact data, the agency will have a clear indication that the rule is now having an impact on small entities.

Scope of the review

Once an agency has determined that an existing rule has a significant economic impact on a substantial number of small entities at the present time, the agency’s section 610 review should, at a minimum, address each of the five factors listed in section 610(b)(1)-(5):

- Whether or not there a continuing need for this rule, consistent with the stated objectives of the applicable statutes;
- Whether the public has ever submitted comments or complaints about this rule;
- The degree of complexity of this rule;
- Whether some other federal or state requirement accomplishes the same regulatory objective as this rule; and

- The length of time since the agency has reviewed this rule, and/or the extent to which circumstances have changed which may affect regulated entities.

Particular attention should be paid to changes in technology, economic circumstances, competitive forces, and the cumulative burden faced by regulated entities. Has the impact of the rule on small entities remained the same?

Section 610(b) requires an agency to evaluate and minimize “any significant economic impact of a rule on a substantial number of small entities in a manner consistent with the stated objectives of applicable statutes.” To accomplish this, agencies may want to use an economic analysis similar to the initial regulatory flexibility analysis (IRFA) under section 603 of the RFA, taking into account the limitations on data availability and limited agency resources.²⁹⁷ Agencies have the discretion to place significant weight on other relevant factors, in addition to the types of economic data required by an IRFA. These other factors include an agency’s experience in implementing the rule, as well as the views expressed over time by the public, regulated entities, and Congress. With the benefit of actual experience with a rule, the agency and other interested parties should be in a good position to evaluate potential improvements to the rule. Several factors deserve attention here, such as the benefits achieved by the regulation, unintended market effects and market distortions, unusually high firm mortality rates in specific industry sub-sectors, and widespread noncompliance with reporting and other paperwork requirements. Thus, a useful review should go beyond obvious measures such as ensuring that regulatory requirements are expressed in plain language and that paperwork can be filed electronically. The analysis should be aimed at understanding and reducing burdens that unnecessarily have a significant impact on small entities.

As a matter of good practice, the section 610 analysis should be based on relevant data, public comments, and agency experience. The agency should make use of available information and data supplied by the public, and indicate the sources of the data. To the extent that an agency relies on specific data to reach a conclusion about the continuing efficacy of a rule, the agency should be able to provide that data. The agency should explain its assumptions so that stakeholders can understand its analysis.

Timing of the review

The language of section 610 specifies that the review should take place within 10 years after the date a rule is promulgated. While agencies need to gain some experience with a rule before undertaking a retrospective review, the review may take place prior to the 10-year mark. If an agency substantially revises a rule after its initial promulgation, it is

²⁹⁷ See 5 U.S.C. § 603. The legislative history of S.299, which became the RFA, notes that “[i]n reviewing existing rules, agencies should follow the procedures described in sections 602-609 [of the RFA] to the extent appropriate.” 142 Cong. Rec. H24,575, H24,583-585 (daily ed. Sept. 8, 1980). In the context of a section 610 review, the elements of an IRFA analysis that should be present include: a discussion of the number and types of small entities affected by the rule, a description of the compliance requirements of the rule and an estimate of their costs, identification of any duplicative or overlapping requirements, and a description of possible alternative regulatory approaches.

arguable whether the 610 review may be delayed to correspond to the revision date. Advocacy would not likely object to a revision of the date, but agencies should seek input from Advocacy on this point.

Section 610 does not specifically set a limit on the amount of time for a rule review. Some agencies have reported that they spend more than a year on each section 610 review. It is within an agency's discretion to determine how much time it needs to spend on retrospective rule reviews. Advocacy recognizes that section 610 reviews may take more than a year in order to permit adequate time to gather and analyze data, to allow public comment, and to consider those comments in the review. Of course, some reviews could take less time, based on the complexity of the issues and the nature of the regulated industry.

Agencies may wish to take advantage of the opportunity afforded in section 605(c) of the RFA to consider a series of "closely related rules" as one rule for periodic review purposes. An agency can accomplish a comprehensive section 610 review of closely related rules, satisfying the requirements of the RFA while potentially reducing the agency resources required.

Outreach to regulated small entities

Section 610(c) of the RFA requires agencies to publish in the *Federal Register* a list of the rules they plan to review in the upcoming year. Agencies use the *Unified Regulatory Agenda* for this purpose.²⁹⁸ This listing requirement is intended to give small entities early notice of the section 610 reviews so that they will be ready and able to provide the agency with comments about the rule under review. As a practical matter, however, agencies often give stakeholders no other information about the ongoing status of a section 610 review, what factors an agency is considering in conducting the review, how comments can be submitted to the agency, or the factual basis on which the agency made its section 610 review findings.

Agencies should communicate with interested entities about the status of ongoing section 610 reviews, as well as those they have completed, to enhance transparency. This information may be most efficiently communicated via an agency website or other electronic media, and should inform interested parties of their ability to submit comments, as well as the agency's commitment to consider those comments. Several agencies already utilize web-based communications as an outreach tool during section 610 reviews.²⁹⁹

Insights about an existing regulation received from regulated entities and other interested parties should be a key component of a retrospective rule review. By making the review process transparent and accessible, agencies are more likely to identify improvements that will benefit all parties at the conclusion of the review. Advocacy can help agencies who wish to communicate with small entity stakeholders by hosting roundtables, working

²⁹⁸ The *Unified Regulatory Agenda* can be accessed at <http://www.reginfo.gov>.

²⁹⁹ See, e.g., www.osha.gov, www.epa.gov, and www.dot.gov and search for "RFA section 610."

through trade groups, and getting a specific message to a targeted audience. Advocacy is ready to assist agencies in their outreach efforts.

Using other agency reviews to satisfy section 610

Agencies that undertake retrospective rule reviews to satisfy other agency objectives may also be able to satisfy the periodic review requirement of section 610, as long as the rule reviews are functionally equivalent. For example, agencies that evaluated a current regulation pursuant to Executive Order 13,563³⁰⁰, or earlier, the Office of Management and Budget's 2002 publicly nominated rule reform process³⁰¹ or OMB's manufacturing rule reform process³⁰² could qualify as section 610 reviews, if they otherwise met the criteria for section 610 review. Similarly, agencies that undertook retrospective reviews of their regulatory programs because of complaints or petitions from regulated entities could qualify as section 610 reviews – as long as the review includes the minimum factors required by section 610. The best way for agencies to get “credit” for a section 610 review in these circumstances is to communicate adequately with stakeholders, and with Advocacy.

Examples of successful retrospective rule reviews

Federal Railroad Administration's Section 610 review of railroad workplace safety.

On December 1, 2003, the Department of Transportation's Federal Railroad Administration (FRA) completed a section 610 review of its railroad workplace safety regulations. After determining that the workplace safety regulations had a significant economic impact on a substantial number of small entities, the FRA examined the rules in light of section 610's review factors. Although the FRA did not recommend any regulatory change as a result of this review, they provided a good description of its analysis of the workplace safety regulations under each review factor and the agency's conclusions.³⁰³

EPA's RCRA review. As a result of public nominations for reforms to the Environmental Protection Agency's hazardous waste management program under the Resource Conservation and Recovery Act (RCRA), EPA evaluated the program and identified duplicative requirements, such as forcing filers to submit reports to multiple locations when one location is adequate. By reducing or eliminating these procedures

³⁰⁰ See, for example, p. 61 of 2011 Report to Congress on the Costs and Benefits of Federal Regulations and Unfunded Mandates on State, Local, and Tribal Entities (2011) available at http://www.whitehouse.gov/sites/default/files/omb/inforeg/2011_cb/2011_cba_report.pdf.

³⁰¹ See Table 9, “New Reforms Planned or Underway – Regulations” and Table 10, “New Reforms Planned or Underway – Guidance Documents” in *Informing Regulatory Decisions: 2003 Report to Congress on the Costs and Benefits of Federal Regulations and Unfunded Mandates on State, Local, and Tribal Entities* (September 2003) at 26-34; available at [http://yosemite.epa.gov/SAB/sabcvpress.nsf/0/5143268e911789ba85256db900562c4b/\\$FILE/2003_cost-ben_final_rpt.pdf](http://yosemite.epa.gov/SAB/sabcvpress.nsf/0/5143268e911789ba85256db900562c4b/$FILE/2003_cost-ben_final_rpt.pdf).

³⁰² See *Regulatory Reform of the U.S. Manufacturing Sector* (2005) at http://georgewbush-whitehouse.archives.gov/omb/inforeg/reports/manufacturing_initiative.pdf.

³⁰³ See http://www.fra.dot.gov/downloads/safety/railroad_workplace_safety.pdf.

after public notice and comment, EPA enabled regulated entities to collectively save up to \$3 million per year while preserving the protections of the RCRA program. The retrospective review was successful because it involved a detailed review of the program's requirements and their costs, based on years of practical experience. The agency considered technical changes such as computerization that have made some of the older paperwork requirements redundant, and found ways to modernize the program to reflect current realities.³⁰⁴

OSHA excavations standard. In March 2007, the Occupational Safety and Health Administration completed a section 610 review of its rules governing excavations and trenches. These standards had been in place since 1989, and were designed to ensure that trenches do not collapse on workers and that excavated material does not fall back into a trench and bury workers. In the review, OSHA did a good job of seeking public input on how and whether the rule should be changed. While the agency ultimately decided that no regulatory changes to the standard were warranted, it did determine that additional outreach and worker training would help continue the downward trend of fewer deaths and injuries from trench and excavation work. OSHA concluded that its current excavations standard has reduced deaths from approximately 90 per year to about 70 per year.³⁰⁵

FCC Section 610 review of 1993-1995 rules. In May 2005, the Federal Communications Commission undertook a section 610 review of rules the Commission adopted in 1993, 1994, and 1995 which have, or might have, a significant economic impact on a substantial number of small entities. The FCC solicited public comment on the rules under review, explained the criteria it was using to review the rules, and gave instructions on where to file comments. This approach was transparent because the agency allowed adequate time for comments (three months) and gave interested parties sufficient information to prepare useful comments.³⁰⁶

Section 610 assistance from the Office of Advocacy

The Office of Advocacy is ready to assist agencies that are planning a retrospective review of their regulations, to ensure that the review fully meets the requirements of section 610. Discussions with the Office of Advocacy are confidential interagency communications, and the Advocacy staff is ready to assist. For more information about this guidance, or for other questions about compliance with section 610, contact Advocacy at (202) 205-6533.

³⁰⁴ 71 Fed. Reg. 16,862 (April 4, 2006).

³⁰⁵ 72 Fed. Reg. 14,727 (March 29, 2007).

³⁰⁶ 70 Fed. Reg. 33,416 (June 8, 2005).

CHAPTER 7 ADDITIONAL RFA AND SBREFA REQUIREMENTS

This chapter addresses additional agency responsibilities beyond the rulemaking process. Under the RFA and SBREFA, agencies have ongoing responsibilities toward small entities with respect to (1) providing notice of rulemakings, (2) developing compliance guides, (3) establishing penalty reduction policies, and (4) offering compliance assistance. In addition, SBREFA created a process for small businesses to report excessive federal agency enforcement actions.

Semi-annual regulatory agenda

Section 602 of the RFA requires federal agencies to publish a regulatory flexibility agenda in the *Federal Register* during April and October of each year.³⁰⁷ Each agency is required to list all rules it expects to propose or promulgate that are likely to have a significant economic impact on a substantial number of small entities. To be useful to small entities, the regulatory flexibility agenda should include a realistic assessment of the regulations under consideration by the agency for development in the coming year. Agencies generally prepare and publish their regulatory flexibility agenda with the unified regulatory agenda required by Executive Order 12,866.³⁰⁸

The regulatory flexibility agenda must contain:

- A brief description of the subject area of any rule the agency expects to propose or promulgate that is likely to have a significant economic impact on a substantial number of small entities. (See Chapter 1 of this guide for a discussion of how to certify a rule.)
- A summary of the nature of each such rule under consideration, the objectives and the legal basis for issuing each rule, and an approximate schedule for completing action on any rule for which an agency has issued a general notice of proposed rulemaking.
- The name and telephone number of an agency official knowledgeable about the rule.

The RFA requires agencies to endeavor to provide direct notification of the agenda to small entities or their representatives, or to publish the agenda in publications that small entities are likely to receive, and to invite comments in the agenda.³⁰⁹

³⁰⁷ 5 U.S.C. § 602(a).

³⁰⁸ Exec. Order 12,866 § 4(b).

³⁰⁹ See § 609 of the RFA regarding the outreach to small entities to obtain needed comment during agency rulemaking. An example of a useful outreach tool is the U.S. Department of Transportation's Docket Management System (DMS). DMS offers a service (listserv) to which a small entity can subscribe and tailor to receive notification when certain documents reach the DMS.

The law also requires each agency to transmit its regulatory flexibility agenda to the Chief Counsel for Advocacy for comment, if any. The Office of Advocacy welcomes the opportunity to provide an agency with input on a pre-publication draft of the agency's regulatory flexibility agenda. Advocacy will review the draft agenda and may provide comment on its completeness and the agency's assessment as to whether a given rule will or will not affect small entities. At a minimum, each agency must provide the Office of Advocacy with a copy of the regulatory flexibility agenda upon its publication. If the agenda is submitted upon publication, the Office of Advocacy will offer comments; however, the agency and the small entities reviewing the agenda will not receive the benefit of Advocacy's pre-publication review.

Small entity compliance guides

For each rule (or related series of rules) requiring a final regulatory flexibility analysis, section 212 of SBREFA requires the agency to publish one or more small entity compliance guides.³¹⁰ Agencies are required to publish the guides with publication of the final rule, post them to websites, distribute them to industry contacts, and report annually to Congress.³¹¹

Agency compliance with this requirement is varied.³¹² In other words, unless the agency is going to certify that the rule will not have a significant economic impact on a substantial number of small entities, the agency must issue a small entity compliance guide, and designate it as such. As appropriate to the rule, Advocacy urges agencies to write the small entity compliance guide in plain and simple language. It should be readily understandable from the perspective of small entities subject to the rule. The guide is to inform a small entity of its obligations and responsibilities under the rule. It may be appropriate to prepare separate guides for different classes or groups of small entities. The guides may cover federal and state requirements affecting the small entities subject to the rule.³¹³

In preparing a small business compliance guide, agencies should look to the small entity comments in the rulemaking record as one indicator of the type of questions to answer or issues to clarify in the compliance guide. In addition, it would be beneficial for the agency to contact small entities subject to the rule (or their trade associations) to solicit input on topics to address in the compliance guide. Agencies may engage the assistance of outside consultants and/or trade associations in the drafting and dissemination process. Small entities and their trade associations can also provide recommendations on the best venue for distribution of the compliance guides, through the agency website and/or through small business associations and organizations.

³¹⁰ Small Business Regulatory Enforcement Fairness Act, Pub. Law 104-121 § 212.

³¹¹ The Small Business and Work Opportunity Act of 2007 added these additional requirements for agency compliance to SBREFA.

³¹² See generally Regulatory Reform: Compliance Guide Requirement Has Had Little Effect on Agency Practices (GAO-02-172, December 2001).

³¹³ See § 215 of the Small Business Regulatory Enforcement Fairness Act of 1996, Pub. L. No. 104-121, 110 Stat. 857 (codified at 5 U.S.C. § 601 et seq.).

Most important, the agency must issue the compliance guide with the final rule, well before the deadline for small entity compliance. To accomplish this, an agency should include development of the compliance guide in the rule development timetable and planning process. As with the regulatory analyses required under the RFA, the agency should anticipate the need to allocate appropriate personnel and resources toward developing the compliance guide at the inception of the rule development process.

Although the compliance guide requirement under SBREFA is not specific in many regards as to what agencies are required to do, Advocacy has noted several instances in which agencies have failed to meet even the most basic requirements of the statute. For instance, the Federal Acquisition Regulation (FAR) Council³¹⁴ published a list of rules for which a FRFA was prepared. This is not a compliance guide.

Compliance guides issued pursuant to section 212 are not subject to judicial review under SBREFA; however, the content of the compliance guide may serve as evidence of the reasonableness or appropriateness of any proposed fines, penalties, or damages in a civil or administrative action against a small business for a violation.³¹⁵

Informal compliance assistance

Section 213 of SBREFA acknowledges the importance of compliance assistance and directs agencies that regulate small entities to establish a practice of answering inquiries from small entities. Agencies are to provide information and advice about compliance, helping small entities interpret and apply the law to specific facts provided by the small entity making the inquiry. As with the content of the compliance guides, guidance given by agencies on how the law is to be applied to a specific factual situation provided by the small entity may be considered evidence of the reasonableness or appropriateness of proposed fines, penalties, or damages imposed on the small entity. Under this section, and using existing resources as practicable, agencies are to institute a practice of providing informal compliance assistance. Agencies were required to establish a program to provide informal compliance assistance within one year of SBREFA's enactment in 1996 and to report to Congress on their programs no later than two years after enactment.³¹⁶

³¹⁴ The (FAR Council prepares and issues revisions to the uniform policies and procedures for acquisition by all executive agencies. The FAR Council does this in conjunction with the Defense Acquisition Regulations (DAR) Council and the Civilian Agency Acquisition (CAA) Council. 48 C.F.R § 1 (2000).

³¹⁵ Sections 231–233 of SBREFA amended the Equal Access to Justice Act (EAJA). These provisions expanded the ability of parties in litigation with the government to recover attorney fees under that law. In administrative and judicial proceedings, if the government's demand to enforce a party's compliance with a statutory or regulatory requirement is unreasonable when compared with the judgment or decision, the party may be entitled to attorney fees and other expenses related to defending against the action. SBREFA increased the allowable attorney fees from \$75 per hour to \$125 per hour.

³¹⁶ The Committee on Small Business and the Committee on Governmental Affairs of the U.S. Senate and the Committee on Small Business and the Committee on the Judiciary of the U.S. House of Representatives were to receive agency reports required under sections 213 and 223 of SBREFA.

Regulatory enforcement fairness

Section 222 of SBREFA establishes a process for small businesses to register complaints about excessive enforcement actions. Pursuant to the law, the Administrator of the U.S. Small Business Administration has designated a National Ombudsman and Assistant Administrator for Regulatory Enforcement Fairness and established Small Business Regulatory Fairness Boards in each of the SBA's 10 regions.

Each small business regulatory fairness board advises the Ombudsman on small business matters relating to agency enforcement activities and assists the Ombudsman with the preparation of the annual report to Congress. The fairness boards have the authority to hold hearings. Fairness board members are small business owners and operators appointed by the SBA Administrator after consultation with the chairperson and ranking minority members of the House and Senate Committees on Small Business.

The Ombudsman has established a process to receive comments from small businesses on agency enforcement activities and, when appropriate, the Ombudsman passes such comments on to the agency for review and response. The Ombudsman is required to report annually to Congress on agency enforcement efforts based on comments received from small business concerns and from the regulatory fairness boards.

For more information on the Ombudsman, visit <http://www.sba.gov/ombudsman/>.

Penalty reduction policies

Agencies regulating activities of small entities are required, under section 223 of SBREFA, to establish a policy or program to provide for the reduction (and, under appropriate circumstances, the waiver) of civil penalties for violations of a statutory or regulatory requirement by a small entity. SBREFA grants agencies broad discretion with respect to the scope of their penalty reduction and waiver policies.³¹⁷ Agencies were to implement their small entity penalty reduction and waiver programs within one year of the enactment of SBREFA in 1996 and to report on their programs to Congress one year later.³¹⁸ Under appropriate circumstances, an agency may consider the ability to pay as a factor in determining penalty assessments on small entities.

Policies or programs established by agencies should contain conditions or exclusions that may include, but are not limited to:

- Requiring a small entity to correct the violation within a reasonable period of time.

³¹⁷ See generally *Regulatory Reform: Implementation of Selected Agencies' Civil Penalty Relief Policies for Small Entities* (GAO-01-280, February 2001). The Office of Advocacy maintains that agencies should define small entities in accordance with section 601 of the RFA.

³¹⁸ Approximately 22 of the 77 agencies that assess penalties submitted a report pursuant to section 223 of SBREFA. House of Representatives Report 106-8, Part I, pp. 5-6.

- Limiting the applicability of the policy to violations discovered through participation by a small entity in a compliance assistance or audit program operated or supported by the agency or a state.
- Excluding small entities that have been subject to multiple enforcement actions by the agency.
- Excluding violations involving willful or criminal conduct.
- Excluding violations that pose serious health, safety, or environmental threats.
- Requiring a good-faith effort to comply with the law.

Congressional review

The Congressional Review Act, Section 251 of the Contract with America Advancement Act of 1996 (which also includes SBREFA), requires agencies to provide Congress with notice of final agency rulemaking actions and the opportunity to review a “major rule” before it becomes effective.³¹⁹ Before a final rule can become effective, the agency promulgating the rule must submit a report to the House of Representatives, the Senate, and the Comptroller General of the U.S. Government Accountability Office (GAO).³²⁰ The report must contain the following information:

- A copy of the rule.
- A concise general statement about the purpose of the rule, including whether it is a “major rule.”³²¹
- The proposed effective date of the regulation.

In addition, the agency is required to include with its report to the Comptroller General, and make available to both houses of Congress, the following information:

- A copy of the cost-benefit analysis of the rule, if any.
- The agency's actions relevant to sections 603, 604, 605, 607, and 609 of the RFA.
- The agency's actions relevant to sections 202, 203, 204, and 205 of the Unfunded Mandates Reform Act of 1995.³²²

Major rules cannot take effect until the end of a 60-legislative-day period beginning on the latter of: (1) the date Congress receives the agency’s report or (2) the date of the rule’s publication in the *Federal Register*. Congress may disapprove or rescind a rule by a joint resolution of disapproval, subject to a presidential veto.³²³

³¹⁹ Codified at Chapter 8 of title 5, United States Code.

³²⁰ 5 U.S.C. § 801. The GAO’s website, <http://www.gao.gov>, includes information on major rules, including a form for submitting a rule under the Congressional Review Act (<http://www.gao.gov/decisions/majrule/fedrule2.pdf>).

³²¹ A “major rule” is a rule that the Administrator of the Office of Information and Regulatory Affairs (OIRA) of the Office of Management and Budget (OMB) finds has resulted or is likely to result in an annual impact on the economy of \$100 million or more; have a major impact on an industry, government, or consumers; or have an effect on competition, productivity, or international trade. 5 U.S.C. § 804(2).

³²² 2 U.S.C. § 1501.

³²³ This congressional authority was first used in 2001 to prevent the Department of Labor’s Ergonomics Rule from taking effect (Pub. L. No. 107-5, 115 Stat. 7 [2001]). Subsequently, in 2017 Congress passed

CONCLUSION

The RFA does not seek preferential treatment for small entities, does not require agencies to adopt regulations that impose the least burden on small entities, and does not mandate exemptions for small entities.

Rather, as this guide has illustrated, the RFA establishes an analytical process for determining how public policy issues can best be achieved without erecting barriers to competition, stifling innovation, or imposing undue burdens on small entities. In so doing, it seeks a level playing field for small entities, not an unfair advantage.

This guide is designed to help institutionalize these concepts so that they become part of a regulatory agency's analytical fiber. The SBA's Office of Advocacy hopes that this guide helps to achieve this objective.

and the president signed 14 resolutions overturning the following agency rules: Disapproval of SEC Disclosure of Payments by Resource Extraction Issuers Rule, Pub. L. No. 115-4, 131 Stat. 9 (2017); Disapproval of Department of Interior Stream Protection Rule, Pub. L. No. 115-5, 131 Stat. 10 (2017); Disapproval of Social Security Administration Implementation of the NICS Improvement Amendments Act of 2007 Rule, Pub. L. No. 115-8, 131 Stat. 15 (2017); Disapproval of Federal Acquisition Regulation Fair Pay and Safe Workplaces Rule, Pub. L. No. 115-11, 131 Stat. 75 (2017); Disapproval of Department of Interior Rule Relating to Land Use Planning, Pub. L. No. 115-12, 131 Stat. 76 (2017); Disapproval of Department of Education Rule Relating to Accountability and State Plans under the Elementary and Secondary Education Act of 1965, Pub. L. No. 115-13, 131 Stat. 77 (2017); Disapproval of Department of Labor Drug Testing of Unemployment Compensation Applicants Rule, Pub. L. No. 115-17, 131 Stat. 81 (2017); Disapproval of Department of Interior Rule Relating to Non-Subsistence Take of Wildlife, and Public Participation and Closure Procedures, on National Wildlife Refuges in Alaska, Pub. L. No. 115-20, 131 Stat. 86 (2017); Disapproval of Department of Labor Rule Relating to Clarification of Employer's Continuing Obligation to Make and Maintain an Accurate Record of Each Recordable Injury and Illness, Pub. L. No. 115-21, 131 Stat. 87 (2017); Disapproval of Federal Communications Commission Rule Relating to Protecting the Privacy of Customers of Broadband and Other Telecommunications Services, Pub. L. No. 115-22, 131 Stat. 88 (2017); Disapproval of Department of Health and Human Services Rule on Compliance with Title X Requirements by Project Recipients in Selecting Subrecipients, Pub. L. No. 115-23, 131 Stat. 89 (2017); Disapproval of Department of Labor Rule Relating to Savings Arrangements Established by Qualified State Political Subdivisions for Non-governmental Employees, Pub. L. No. 115-24, 131 Stat. 90 (2017); Disapproval of Department of Labor Rule Relating to Savings Arrangements Established by States for Non-Governmental Employees, Pub. L. No. 115-35, 131 Stat. 848 (2017).

APPENDIX A THE REGULATORY FLEXIBILITY ACT

The following text of the Regulatory Flexibility Act of 1980, as amended, is taken from Title 5 of the United States Code, sections 601–612. The Regulatory Flexibility Act was originally passed in 1980 (P.L. 96-354). The act was amended by the Small Business Regulatory Enforcement Fairness Act of 1996 (P.L. 104-121), the Dodd-Frank Wall Street Reform and Consumer Protection Act (P.L. 111-203), and the Small Business Jobs Act of 2010 (P.L. 111-240).

Congressional Findings and Declaration of Purpose

(a) The Congress finds and declares that —

(1) when adopting regulations to protect the health, safety and economic welfare of the Nation, Federal agencies should seek to achieve statutory goals as effectively and efficiently as possible without imposing unnecessary burdens on the public;

(2) laws and regulations designed for application to large scale entities have been applied uniformly to small businesses, small organizations, and small governmental jurisdictions even though the problems that gave rise to government action may not have been caused by those smaller entities;

(3) uniform Federal regulatory and reporting requirements have in numerous instances imposed unnecessary and disproportionately burdensome demands including legal, accounting and consulting costs upon small businesses, small organizations, and small governmental jurisdictions with limited resources;

(4) the failure to recognize differences in the scale and resources of regulated entities has in numerous instances adversely affected competition in the marketplace, discouraged innovation and restricted improvements in productivity;

(5) unnecessary regulations create entry barriers in many industries and discourage potential entrepreneurs from introducing beneficial products and processes;

(6) the practice of treating all regulated businesses, organizations, and governmental jurisdictions as equivalent may lead to inefficient use of regulatory agency resources, enforcement problems and, in some cases, to actions inconsistent with the legislative intent of health, safety, environmental and economic welfare legislation;

(7) alternative regulatory approaches which do not conflict with the stated objectives of applicable statutes may be available which minimize the significant economic impact of rules on small businesses, small organizations, and small governmental jurisdictions;

(8) the process by which Federal regulations are developed and adopted should be reformed to require agencies to solicit the ideas and comments of small businesses, small organizations, and small governmental jurisdictions to examine the impact of proposed and existing rules on such entities, and to review the continued need for existing rules.

(b) It is the purpose of this Act [enacting this chapter and provisions set out as notes under this section] to establish as a principle of regulatory issuance that agencies shall endeavor, consistent with the objectives of the rule and of applicable statutes, to fit regulatory and informational requirements to the scale of the businesses, organizations, and governmental jurisdictions subject to regulation. To achieve this principle, agencies are required to solicit and consider flexible regulatory proposals and to explain the rationale for their actions to assure that such proposals are given serious consideration.

Regulatory Flexibility Act

- § 601 Definitions
- § 602 Regulatory agenda
- § 603 Initial regulatory flexibility analysis
- § 604 Final regulatory flexibility analysis
- § 605 Avoidance of duplicative or unnecessary analyses
- § 606 Effect on other law
- § 607 Preparation of analyses
- § 608 Procedure for waiver or delay of completion
- § 609 Procedures for gathering comments
- § 610 Periodic review of rules
- § 611 Judicial review
- § 612 Reports and intervention rights

§ 601 Definitions

For purposes of this chapter —

- (1) the term “agency” means an agency as defined in section 551(1) of this title;
- (2) the term “rule” means any rule for which the agency publishes a general notice of proposed rulemaking pursuant to section 553(b) of this title, or any other law, including any rule of general applicability governing Federal grants to State and local governments for which the agency provides an opportunity for notice and public comment, except that the term “rule” does not include a rule of particular applicability relating to rates, wages, corporate or financial structures or reorganizations thereof, prices, facilities, appliances, services, or allowances therefor or to valuations, costs or accounting, or practices relating to such rates, wages, structures, prices, appliances, services, or allowances;
- (3) the term “small business” has the same meaning as the term “small business concern” under section 3 of the Small Business Act, unless an agency, after consultation with the Office of Advocacy of the Small Business Administration and after opportunity for public comment, establishes one or more definitions of such term which are appropriate to the activities of the agency and publishes such definition(s) in the Federal Register;
- (4) the term “small organization” means any not-for-profit enterprise which is independently owned and operated and is not dominant in its field, unless an agency establishes, after opportunity for public comment, one or more definitions of such term which are appropriate to the activities of the agency and publishes such definition(s) in the Federal Register;
- (5) the term “small governmental jurisdiction” means governments of cities, counties, towns, townships, villages, school districts, or special districts, with a population of less than fifty thousand, unless an agency establishes, after opportunity for public comment, one or more definitions of such term which are appropriate to the activities of the agency and which are based on such factors as location in rural or sparsely populated areas or limited revenues due to the population of such jurisdiction, and publishes such definition(s) in the Federal Register;
- (6) the term “small entity” shall have the same meaning as the terms “small business,” “small organization” and “small governmental jurisdiction” defined in paragraphs (3), (4) and (5) of this section; and
- (7) the term “collection of information” —
 - (A) means the obtaining, causing to be obtained, soliciting, or requiring the disclosure to third parties or the public, of facts or opinions by or for an agency, regardless of form or format, calling for either —

(i) answers to identical questions posed to, or identical reporting or recordkeeping requirements imposed on, 10 or more persons, other than agencies, instrumentalities, or employees of the United States; or

(ii) answers to questions posed to agencies, instrumentalities, or employees of the United States which are to be used for general statistical purposes; and

(B) shall not include a collection of information described under section 3518(c)(1) of title 44, United States Code.

(8) Recordkeeping requirement — The term “recordkeeping requirement” means a requirement imposed by an agency on persons to maintain specified records.

§ 602. Regulatory agenda

(a) During the months of October and April of each year, each agency shall publish in the Federal Register a regulatory flexibility agenda which shall contain —

(1) a brief description of the subject area of any rule which the agency expects to propose or promulgate which is likely to have a significant economic impact on a substantial number of small entities;

(2) a summary of the nature of any such rule under consideration for each subject area listed in the agenda pursuant to paragraph (1), the objectives and legal basis for the issuance of the rule, and an approximate schedule for completing action on any rule for which the agency has issued a general notice of proposed rulemaking, and

(3) the name and telephone number of an agency official knowledgeable concerning the items listed in paragraph (1).

(b) Each regulatory flexibility agenda shall be transmitted to the Chief Counsel for Advocacy of the Small Business Administration for comment, if any.

(c) Each agency shall endeavor to provide notice of each regulatory flexibility agenda to small entities or their representatives through direct notification or publication of the agenda in publications likely to be obtained by such small entities and shall invite comments upon each subject area on the agenda.

(d) Nothing in this section precludes an agency from considering or acting on any matter not included in a regulatory flexibility agenda, or requires an agency to consider or act on any matter listed in such agenda.

§ 603. Initial regulatory flexibility analysis

(a) Whenever an agency is required by section 553 of this title, or any other law, to publish general notice of proposed rulemaking for any proposed rule, or publishes a notice of proposed rulemaking for an interpretative rule involving the internal revenue laws of the United States, the agency shall prepare and make available for public comment an initial regulatory flexibility analysis. Such analysis shall describe the impact of the proposed rule on small entities. The initial regulatory flexibility analysis or a summary shall be published in the Federal Register at the time of the publication of general notice of proposed rulemaking for the rule. The agency shall transmit a copy of the initial regulatory flexibility analysis to the Chief Counsel for Advocacy of the Small Business Administration. In the case of an interpretative rule involving the internal revenue laws of the United States, this chapter applies to interpretative rules published in the Federal Register for codification in the Code of Federal Regulations, but only to the extent that such interpretative rules impose on small entities a collection of information requirement.

(b) Each initial regulatory flexibility analysis required under this section shall contain —

(1) a description of the reasons why action by the agency is being considered;

(2) a succinct statement of the objectives of, and legal basis for, the proposed rule;

- (3) a description of and, where feasible, an estimate of the number of small entities to which the proposed rule will apply;
 - (4) a description of the projected reporting, recordkeeping and other compliance requirements of the proposed rule, including an estimate of the classes of small entities which will be subject to the requirement and the type of professional skills necessary for preparation of the report or record;
 - (5) an identification, to the extent practicable, of all relevant Federal rules which may duplicate, overlap or conflict with the proposed rule.
- (c) Each initial regulatory flexibility analysis shall also contain a description of any significant alternatives to the proposed rule which accomplish the stated objectives of applicable statutes and which minimize any significant economic impact of the proposed rule on small entities. Consistent with the stated objectives of applicable statutes, the analysis shall discuss significant alternatives such as —
- (1) the establishment of differing compliance or reporting requirements or timetables that take into account the resources available to small entities;
 - (2) the clarification, consolidation, or simplification of compliance and reporting requirements under the rule for such small entities;
 - (3) the use of performance rather than design standards; and
 - (4) an exemption from coverage of the rule, or any part thereof, for such small entities.
- (d) (1) For a covered agency, as defined in section 609(d)(2), each initial regulatory flexibility analysis shall include a description of—
- (A) any projected increase in the cost of credit for small entities;
 - (B) any significant alternatives to the proposed rule which accomplish the stated objectives of applicable statutes and which minimize any increase in the cost of credit for small entities; and
 - (C) advice and recommendations of representatives of small entities relating to issues described in subparagraphs (A) and (B) and subsection (b).
- (2) A covered agency, as defined in section 609(d)(2), shall, for purposes of complying with paragraph (1)(C)—
- (A) identify representatives of small entities in consultation with the Chief Counsel for Advocacy of the Small Business Administration; and
 - (B) collect advice and recommendations from the representatives identified under subparagraph (A) relating to issues described in subparagraphs (A) and (B) of paragraph (1) and subsection (b).

§ 604. Final regulatory flexibility analysis

- (a) When an agency promulgates a final rule under section 553 of this title, after being required by that section or any other law to publish a general notice of proposed rulemaking, or promulgates a final interpretative rule involving the internal revenue laws of the United States as described in section 603(a), the agency shall prepare a final regulatory flexibility analysis. Each final regulatory flexibility analysis shall contain —
- (1) a statement of the need for, and objectives of, the rule;
 - (2) a statement of the significant issues raised by the public comments in response to the initial regulatory flexibility analysis, a statement of the assessment of the agency of such issues, and a statement of any changes made in the proposed rule as a result of such comments;
 - (3) the response of the agency to any comments filed by the Chief Counsel for Advocacy of the Small Business Administration in response to the proposed rule, and a detailed statement of any change made to the proposed rule in the final rule as a result of the comments;
 - (4) a description of and an estimate of the number of small entities to which the rule will apply or an explanation of why no such estimate is available;

(5) a description of the projected reporting, recordkeeping and other compliance requirements of the rule, including an estimate of the classes of small entities which will be subject to the requirement and the type of professional skills necessary for preparation of the report or record;

(6) a description of the steps the agency has taken to minimize the significant economic impact on small entities consistent with the stated objectives of applicable statutes, including a statement of the factual, policy, and legal reasons for selecting the alternative adopted in the final rule and why each one of the other significant alternatives to the rule considered by the agency which affect the impact on small entities was rejected;

(6)¹ for a covered agency, as defined in section 609(d)(2), a description of the steps the agency has taken to minimize any additional cost of credit for small entities.

(b) The agency shall make copies of the final regulatory flexibility analysis available to members of the public and shall publish in the Federal Register such analysis or a summary thereof..

¹So in original. Two paragraphs (6) were enacted.

§ 605. Avoidance of duplicative or unnecessary analyses

(a) Any Federal agency may perform the analyses required by sections 602, 603, and 604 of this title in conjunction with or as a part of any other agenda or analysis required by any other law if such other analysis satisfies the provisions of such sections.

(b) Sections 603 and 604 of this title shall not apply to any proposed or final rule if the head of the agency certifies that the rule will not, if promulgated, have a significant economic impact on a substantial number of small entities. If the head of the agency makes a certification under the preceding sentence, the agency shall publish such certification in the Federal Register at the time of publication of general notice of proposed rulemaking for the rule or at the time of publication of the final rule, along with a statement providing the factual basis for such certification. The agency shall provide such certification and statement to the Chief Counsel for Advocacy of the Small Business Administration.

(c) In order to avoid duplicative action, an agency may consider a series of closely related rules as one rule for the purposes of sections 602, 603, 604 and 610 of this title.

§ 606. Effect on other law

The requirements of sections 603 and 604 of this title do not alter in any manner standards otherwise applicable by law to agency action.

§ 607. Preparation of analyses

In complying with the provisions of sections 603 and 604 of this title, an agency may provide either a quantifiable or numerical description of the effects of a proposed rule or alternatives to the proposed rule, or more general descriptive statements if quantification is not practicable or reliable.

§ 608. Procedure for waiver or delay of completion

(a) An agency head may waive or delay the completion of some or all of the requirements of section 603 of this title by publishing in the Federal Register, not later than the date of publication of the final rule, a written finding, with reasons therefor, that the final rule is being promulgated

in response to an emergency that makes compliance or timely compliance with the provisions of section 603 of this title impracticable.

(b) Except as provided in section 605(b), an agency head may not waive the requirements of section 604 of this title. An agency head may delay the completion of the requirements of section 604 of this title for a period of not more than one hundred and eighty days after the date of publication in the Federal Register of a final rule by publishing in the Federal Register, not later than such date of publication, a written finding, with reasons therefor, that the final rule is being promulgated in response to an emergency that makes timely compliance with the provisions of section 604 of this title impracticable. If the agency has not prepared a final regulatory analysis pursuant to section 604 of this title within one hundred and eighty days from the date of publication of the final rule, such rule shall lapse and have no effect. Such rule shall not be repromulgated until a final regulatory flexibility analysis has been completed by the agency.

§ 609. Procedures for gathering comments

(a) When any rule is promulgated which will have a significant economic impact on a substantial number of small entities, the head of the agency promulgating the rule or the official of the agency with statutory responsibility for the promulgation of the rule shall assure that small entities have been given an opportunity to participate in the rulemaking for the rule through the reasonable use of techniques such as—

(1) the inclusion in an advance notice of proposed rulemaking, if issued, of a statement that the proposed rule may have a significant economic effect on a substantial number of small entities;

(2) the publication of general notice of proposed rulemaking in publications likely to be obtained by small entities;

(3) the direct notification of interested small entities;

(4) the conduct of open conferences or public hearings concerning the rule for small entities including soliciting and receiving comments over computer networks; and

(5) the adoption or modification of agency procedural rules to reduce the cost or complexity of participation in the rulemaking by small entities.

(b) Prior to publication of an initial regulatory flexibility analysis which a covered agency is required to conduct by this chapter—

(1) a covered agency shall notify the Chief Counsel for Advocacy of the Small Business Administration and provide the Chief Counsel with information on the potential impacts of the proposed rule on small entities and the type of small entities that might be affected;

(2) not later than 15 days after the date of receipt of the materials described in paragraph (1), the Chief Counsel shall identify individuals representative of affected small entities for the purpose of obtaining advice and recommendations from those individuals about the potential impacts of the proposed rule;

(3) the agency shall convene a review panel for such rule consisting wholly of full time Federal employees of the office within the agency responsible for carrying out the proposed rule, the Office of Information and Regulatory Affairs within the Office of Management and Budget, and the Chief Counsel;

(4) the panel shall review any material the agency has prepared in connection with this chapter, including any draft proposed rule, collect advice and recommendations of each individual small entity representative identified by the agency after consultation with the Chief Counsel, on issues related to subsections 603(b), paragraphs (3), (4) and (5) and 603(c);

(5) not later than 60 days after the date a covered agency convenes a review panel pursuant to paragraph (3), the review panel shall report on the comments of the small entity representatives and its findings as to issues related to subsections 603(b), paragraphs (3), (4) and

(5) and 603(c), provided that such report shall be made public as part of the rulemaking record; and

(6) where appropriate, the agency shall modify the proposed rule, the initial regulatory flexibility analysis or the decision on whether an initial regulatory flexibility analysis is required.

(c) An agency may in its discretion apply subsection (b) to rules that the agency intends to certify under subsection 605(b), but the agency believes may have a greater than de minimis impact on a substantial number of small entities.

(d) For purposes of this section, the term “covered agency” means

(1) the Environmental Protection Agency,

(2) the Consumer Financial Protection Bureau of the Federal Reserve System; and

(3) the Occupational Safety and Health Administration of the Department of Labor.

(e) The Chief Counsel for Advocacy, in consultation with the individuals identified in subsection (b)(2), and with the Administrator of the Office of Information and Regulatory Affairs within the Office of Management and Budget, may waive the requirements of subsections (b)(3), (b)(4), and (b)(5) by including in the rulemaking record a written finding, with reasons therefor, that those requirements would not advance the effective participation of small entities in the rulemaking process. For purposes of this subsection, the factors to be considered in making such a finding are as follows:

(1) In developing a proposed rule, the extent to which the covered agency consulted with individuals representative of affected small entities with respect to the potential impacts of the rule and took such concerns into consideration.

(2) Special circumstances requiring prompt issuance of the rule.

(3) Whether the requirements of subsection (b) would provide the individuals identified in subsection (b)(2) with a competitive advantage relative to other small entities.

§ 610. Periodic review of rules

(a) Within one hundred and eighty days after the effective date of this chapter, each agency shall publish in the Federal Register a plan for the periodic review of the rules issued by the agency which have or will have a significant economic impact upon a substantial number of small entities. Such plan may be amended by the agency at any time by publishing the revision in the Federal Register. The purpose of the review shall be to determine whether such rules should be continued without change, or should be amended or rescinded, consistent with the stated objectives of applicable statutes, to minimize any significant economic impact of the rules upon a substantial number of such small entities. The plan shall provide for the review of all such agency rules existing on the effective date of this chapter within ten years of that date and for the review of such rules adopted after the effective date of this chapter within ten years of the publication of such rules as the final rule. If the head of the agency determines that completion of the review of existing rules is not feasible by the established date, he shall so certify in a statement published in the Federal Register and may extend the completion date by one year at a time for a total of not more than five years.

(b) In reviewing rules to minimize any significant economic impact of the rule on a substantial number of small entities in a manner consistent with the stated objectives of applicable statutes, the agency shall consider the following factors—

(1) the continued need for the rule;

(2) the nature of complaints or comments received concerning the rule from the public;

(3) the complexity of the rule;

(4) the extent to which the rule overlaps, duplicates or conflicts with other Federal rules, and, to the extent feasible, with State and local governmental rules; and

(5) the length of time since the rule has been evaluated or the degree to which technology, economic conditions, or other factors have changed in the area affected by the rule.

(c) Each year, each agency shall publish in the Federal Register a list of the rules which have a significant economic impact on a substantial number of small entities, which are to be reviewed pursuant to this section during the succeeding twelve months. The list shall include a brief description of each rule and the need for and legal basis of such rule and shall invite public comment upon the rule.

§ 611. Judicial review

(a) (1) For any rule subject to this chapter, a small entity that is adversely affected or aggrieved by final agency action is entitled to judicial review of agency compliance with the requirements of sections 601, 604, 605(b), 608(b), and 610 in accordance with chapter 7. Agency compliance with sections 607 and 609(a) shall be judicially reviewable in connection with judicial review of section 604.

(2) Each court having jurisdiction to review such rule for compliance with section 553, or under any other provision of law, shall have jurisdiction to review any claims of noncompliance with sections 601, 604, 605(b), 608(b), and 610 in accordance with chapter 7. Agency compliance with sections 607 and 609(a) shall be judicially reviewable in connection with judicial review of section 604.

(3) (A) A small entity may seek such review during the period beginning on the date of final agency action and ending one year later, except that where a provision of law requires that an action challenging a final agency action be commenced before the expiration of one year, such lesser period shall apply to an action for judicial review under this section.

(B) In the case where an agency delays the issuance of a final regulatory flexibility analysis pursuant to section 608(b) of this chapter, an action for judicial review under this section shall be filed not later than—

(i) one year after the date the analysis is made available to the public, or
(ii) where a provision of law requires that an action challenging a final agency regulation be commenced before the expiration of the 1-year period, the number of days specified in such provision of law that is after the date the analysis is made available to the public.

(4) In granting any relief in an action under this section, the court shall order the agency to take corrective action consistent with this chapter and chapter 7, including, but not limited to —

(A) remanding the rule to the agency, and

(B) deferring the enforcement of the rule against small entities unless the court finds that continued enforcement of the rule is in the public interest.

(5) Nothing in this subsection shall be construed to limit the authority of any court to stay the effective date of any rule or provision thereof under any other provision of law or to grant any other relief in addition to the requirements of this section.

(b) In an action for the judicial review of a rule, the regulatory flexibility analysis for such rule, including an analysis prepared or corrected pursuant to paragraph (a)(4), shall constitute part of the entire record of agency action in connection with such review.

(c) Compliance or noncompliance by an agency with the provisions of this chapter shall be subject to judicial review only in accordance with this section.

(d) Nothing in this section bars judicial review of any other impact statement or similar analysis required by any other law if judicial review of such statement or analysis is otherwise permitted by law.

§ 612. Reports and intervention rights

(a) The Chief Counsel for Advocacy of the Small Business Administration shall monitor agency compliance with this chapter and shall report at least annually thereon to the President and to the Committees on the Judiciary and Small Business of the Senate and House of Representatives.

(b) The Chief Counsel for Advocacy of the Small Business Administration is authorized to appear as amicus curiae in any action brought in a court of the United States to review a rule. In any such action, the Chief Counsel is authorized to present his or her views with respect to compliance with this chapter, the adequacy of the rulemaking record with respect to small entities and the effect of the rule on small entities.

(c) A court of the United States shall grant the application of the Chief Counsel for Advocacy of the Small Business Administration to appear in any such action for the purposes described in subsection (b).

APPENDIX B SMALL BUSINESS BY THE NUMBERS

Frequently Asked Questions About Small Business

1. What is a small business?

The Office of Advocacy defines a small business as an independent business having fewer than 500 employees. For the industry-level definitions of small business used in government programs and contracting, see www.sba.gov/content/small-business-size-standards.

2. How many small businesses are there in the U.S.?

In 2014, there were 29.6 million small businesses.

- Eighty percent, or 23.8 million, had no employees (termed “nonemployers”)
- Twenty percent, or 5.8 million, had paid employees
- There were 19,000 large businesses.

The number of small employers has increased after a decline during the recession. The number of nonemployers has gradually increased, from 15.4 million in 1997 to 23.8 million in 2014. (Figure 1).

Source: SUSB, NES

3. What is the role of small businesses in the economy?

Small businesses comprise:

- 99.9% of all firms
- 99.7% of firms with paid employees
- 97.6% of exporting firms (287,835 small exporters)
- 32.9% of known export value (\$440 billion out of \$1.3 trillion)
- 47.8% of private sector employees (58 million out of 121 million employees)
- 41.1% of private-sector payroll

Source: SUSB, NES (2014), ITA (2015)

4. What is the small business percent of net new jobs?

Small businesses accounted for 61.8% of net new jobs from the first quarter of 1993 until the third quarter of 2016. Figure 2 shows details from 1993 to 2016. The small business share of net job change was strongly positive for most of this 24-year time span, except during two recessionary periods.

Source: BED

5. What is the new business survival rate?

79.9% of establishments started in 2015 survived until 2016, the highest share since 2006. From 2005 to 2015, an average of 78.5% of new establishments survived one year.

- About half of all establishments survive five years or longer. In the past decade, this ranged from a low of 45.4% for establishments started in 2006, and a high of 51.0% for those started in 2011.
- About one-third of establishments survive 10 years or longer.

Although data is not available on firm survival rates, other data sources suggest that about two out of three establishment exits are the result of firm closures.

Source: BED, BDS; Office of Advocacy calculations

6. How can small businesses generate three-fifths of net new jobs, but their share of employment is less than 50%?

As firms grow, they change employment size classes. So as small firms grow, their growth counts toward small firm job gains; but if they pass the 500-employee mark, their employment is classified as large firm employment.

7. How many businesses open and close each year?

In 2014, there were about 404,000 startups (firms less than one year old) and 392,000 firm closures (Table 1). The share of businesses that were startups has hovered around 8% since 2010 (Figure 3).

Source: BDS

8. How many businesses do women own?

In 2012, there were 9.9 million women-owned firms, and 2.5 million firms owned equally by men and women (Table 2). This means that 12.3 million firms, or 45% of all classifiable firms, were at least 50% women-owned.

Source: SBO, "Women's Business Ownership: Data from the 2012 Survey of Business Owners," 2017. www.sba.gov/sites/default/files/advocacy/Womens-Business-Ownership-in-the-US.pdf

9. How many businesses do minorities own?

In 2012, 8 million businesses were minority-owned, or 29.3% of U.S. firms. Of these, 12% were Hispanic-owned, 10% were Black- or African American-owned, 7% were Asian-owned, 1% were owned by American Indians and Alaska Natives, and 0.2% were owned by Native Hawaiians and other Pacific Islanders (Table 2).

Source: SBO

10. How many businesses do veterans own?

In 2012, veterans owned 2.5 million businesses, or 9.3% of U.S. firms. About one-fifth of these firms, or 440,000, had paid employees (Table 2).

Source: SBO, "Veteran-Owned Businesses and Their Owners," 2017. www.sba.gov/advocacy/veteran-owned-businesses-and-their-owners.

11. What percent of entrepreneurs are immigrants? In which industries are immigrant-owned firms more common?

About one-seventh, or 14.4%, of business owners are immigrants. The industries with the greatest share of immigrant owners were accommodation and food services (29.1% of owners were foreign-born), and transportation and warehousing (27.5%).

Source: SBO

12. Is millennial entrepreneurship increasing?

Advocacy research shows that in 2014, millennials were less likely to be self-employed than older individuals. This research also shows that the rate of self-employment among individuals age 15 to 34 has been gradually declining since 1990.

Source: "The Missing Millennial Entrepreneurs," February 2016. www.sba.gov/sites/default/files/advocacy/Millennial_IB.pdf

13. What percent of firms are family-owned, and how does this compare to equally-owned firms?

About one in five firms (19.3%) are family-owned. Of these family-owned firms, about half are “equally-owned,” that is, 50% owned by one or more men, and 50% owned by one or more women. Hence, about one in 10 firms is both equally-owned and family-owned.

The industries with the highest share of family-owned firms are management of companies and enterprises (46.4% of firms in this industry are family-owned), real estate and rental and leasing (37.3%), and accommodation and food services (33.2%).

The industries with the highest share of equally-owned firms are real estate and rental and leasing (18.6% of firms in this industry are equally-owned), mining, quarrying, and oil and gas extraction (16.9%), and accommodation and food services (16.9%).

Source: SBO

14. How are most small businesses legally organized?

The majority of nonemployer establishments are sole proprietorships (86.4%), while only 14.4% of establishments at small employer firms are sole proprietorships. Nearly half of the establishments at small employer firms are S-corporations. Table 3 shows details.

Source: SUSB, NES

15. What percent of firms are home-based?

The share of businesses that are home-based has remained relatively constant over the past decade, at about 50% of all firms. More specifically, 60.1% of all firms without paid employees are home-based, as are 23.3% of small employer firms and 0.3% of large employer firms. The industries in which businesses are most likely to be home-based are information (70.0%), construction (68.2%), and professional, scientific, and technical services (65.3%). A home-based business is operated primarily out of one’s home, but business activities may take place at other locations as well.

Source: SBO

16. What percent of firms are franchises?

Overall, 2.9% of firms are franchises. More specifically, 2.3% of nonemployer firms are franchises, as are 5.3% of small employers and 9.6% of large employers.

Source: SBO

17. What is the status of the startup market?

Average employment at startups has fluctuated over the past decade, but reached a four-year high of 6.1 employees in 2014. Average employment at firms of all ages has increased slightly during this period, from 22.4 employees per firm in 2005 to 23.5 employees per firm in 2014 (Figure 4).

Source: BDS

18. How are small businesses financed?

The most common source of capital to finance business expansion is personal and family savings (21.9% of small firms), followed by business profits and assets (5.7%), business loans from financial institutions (4.5%), and business credit cards from banks (3.3%).

Source: SBO. For more information, see “Frequently Asked Questions about Small Business Finance,” www.sba.gov/sites/default/files/Finance-FAQ-2016_WEB.pdf.

19. What is the small business share of federal procurement?

In fiscal year 2016, 24.3% of contracting dollars went to small businesses, down from 25.8% in FY 2015 and 25.1% in FY 2014. Of agencies with at least \$1 billion in eligible contract dollars, the ones that awarded the highest share of contracting dollars to small businesses were the Departments of the Interior (59.8%), Agriculture (56.3%) and Transportation (52.0%).

Source: Small Business Dashboard, <https://smallbusiness.data.gov>

20. How many small businesses are in high-tech industries?

In 2014, there were 248,122 small employer firms in high-tech industries, representing 98.5% of all employer firms in these industries. The majority of these small firms provide services in either computer systems design or architecture and engineering (Figure 5). Among small firms, the industries with the highest growth from 2012 to 2014 were software publishers and computer systems design services (Table 4).

Note: This publication uses the Level I high-tech industries listed in Hecker's 2005 analysis, with the exception of 5161 and 5181, as no corresponding NAICS codes were available for 2012 or 2014 data. For the definition of high-tech industries, see www.bls.gov/opub/mlr/2005/07/art6full.pdf.

21. How are small businesses represented in high-patenting industries?

Small businesses represent about 95.9% of employer firms in high-patenting manufacturing industries, a percentage that remained constant from 2012 to 2014. During the same time period, small businesses' share of employment, payroll, and receipts decreased slightly (Table 5).

Source: SUSB

Data Sources

- BED Business Employment Dynamics, Bureau of Labor Statistics, US Department of Labor, www.bls.gov/bdm
- BDS Business Dynamics Statistics, US Census Bureau, US Department of Commerce, www.census.gov/ces/dataproducts/bds
- ITA International Trade Administration, US Department of Commerce, www.trade.gov/
- NES Nonemployer Statistics, US Census Bureau, US Department of Commerce, www.census.gov/econ/nonemployer
- SBO Survey of Business Owners, US Census Bureau, US Department of Commerce, www.census.gov/programs-surveys/sbo.html
- SUSB Statistics of US Businesses, US Census Bureau, US Department of Commerce, www.census.gov/programs-surveys/susb.html

Is there a PDF version of the FAQ?

Yes. The pdf version is located at www.sba.gov/advocacy/frequently-asked-questions-about-small-business

Updated August 2017

APPENDIX C SMALL BUSINESS STATISTICS FOR REGULATORY ANALYSIS

One of the key tasks in preparing an analysis for the Regulatory Flexibility Act is locating statistics on small business. The information in this appendix should help federal agencies identify data sources appropriate for regulatory analyses.

Ideally, the data used to analyze the costs and benefits of government regulation should be longitudinal firm-level micro-data—that is, data that can be used to trace the performance of a collection of individual firms over multiple years. Unfortunately, virtually all publicly available data on individual firms are subject to confidentiality restrictions.

Individual names and addresses not only cannot be disclosed, but data must also be presented so that individual firms cannot be identified or intuited, even by statistical manipulation. Therefore, most government agencies release summary information, grouping firm data by industry, size, and/or location. It is worth noting that the firms that make up each group change over time. For example, some firms start up while others go out of business; some firms expand into a higher size cohort while others decline into a smaller size category. It is difficult to distinguish between changes to firms that remain in the group and changes in the composition of the group.

The data sources listed here generally cover statistics on industry, employment, payroll, and receipts. Most databases available from government sources do not provide financial data, including the balance sheet and income statement information needed for analyses of the cost of regulations. This is the most sensitive type of information and is rarely available even in aggregate form. Profit information is usually unavailable as well. There is often a lag between the collection and release of government data, ranging from a few quarters to several years. Business data useful in regulatory analysis can be available through fee-based proprietary databases from private sector sources if agency resources permit.

Definitions

It is important to understand the differences between establishments and firms when using small business data:

Establishment: An establishment—a single physical location of a business—is the smallest unit at which business activity is conducted and on which statistical information is collected. Establishments may be branches of large firms or independently owned and operated businesses. Most small businesses consist of a single establishment, but some small businesses have multiple establishments.

Firm: A firm, or enterprise, consists of all establishments owned by a “parent” company. An enterprise may own subsidiaries, branches, and

unrelated establishments. It is a best practice to conduct regulatory analyses at the firm level in order to fully understand the small business impact.

Advocacy Economic Research and Data on Small Businesses

The SBA Office of Advocacy conducts research on a variety of small business topics, and disseminates data and statistics on small businesses. Advocacy research and small business data is available at: <http://www.sba.gov/advocacy>.

Statistics of U.S. Businesses, U.S. Census Bureau

Statistics of U.S. Businesses (SUSB) is an annual series that provides data on employer firms by firm size and industry at the national and subnational level. Beginning in late 1991, the SBA Office of Advocacy contracted with the Census Bureau to produce data on firms of different sizes. This data can be accessed on our website at <https://www.sba.gov/advocacy/firm-size-data>. The Office of Advocacy's data files include the number of establishments, firms, employment, annual payroll, and annual receipts by firm size, industry, and geographical region.

Data are generally available up to the six-digit North American Industry Classification System (NAICS) code level of industrial detail.

Other Federal Agency Data on Small Firms

Current Population Survey, U.S. Census Bureau

Each year, the Census Bureau's March Current Population Survey asks a series of expanded questions about self-employment as part of its firm-size supplement. These questions include the hours and weeks spent working in the business during the previous year, the income earned, the demographics of the business owner, whether the firm (owner) has or provides benefits and several related questions about the industry of the firm.

Data from the Current Population Survey can be used to describe the businesses' sources of capital, their profitability, employment, major industry, and home-based status of women and minority business owners. This data source provides some information on potential regulatory impacts on very small firms, particularly their ability to absorb the burden of federal regulation.

Survey of Business Owners, U.S. Census Bureau

Conducted every five years, the Survey of Business Owners (SBO) is the main source of nationally representative data on characteristics of both firms and their owners. Owner characteristics include demographic information, and firm characteristics include sales, export status, franchise status, owner, and sources of capital. The SBO includes data on firms with and without paid employees (i.e., employer and nonemployer firms). The new Annual Survey of Entrepreneurs is a smaller survey focused on employer businesses that provides additional data on the financing of businesses.

Annual Survey of Entrepreneurs, U.S. Census Bureau

The Annual Survey of Entrepreneurs is similar to the Survey of Business Owners, but smaller and focused on employer businesses. Performed annually starting in 2014, the ASE also collects owner and firm characteristics. The ASE collects data on business profitability and more detailed information about the financing of businesses.

Statistics of Income, Internal Revenue Service

Each quarter, the Statistics of Income (SOI) division of the Internal Revenue Service publishes the *SOI Bulletin*. This publication contains data for both households and businesses and is an invaluable source of statistical information. Data on business firms are generally classified by receipt size class for proprietorships, partnerships, and corporations. For sole proprietorships and partnerships, only data on net income are available.

For small business corporations, more data are available. The *IRS Source Book for Corporations* contains data for corporations by asset size class. Balance sheet and income statement information is available for corporations in different asset classes.

APPENDIX D EXECUTIVE ORDER 12,866

REGULATORY PLANNING AND REVIEW

PUBLIC PAPERS OF THE PRESIDENTS

SEPTEMBER 30, 1993

The American people deserve a regulatory system that works for them, not against them: a regulatory system that protects and improves their health, safety, environment, and well-being and improves the performance of the economy without imposing unacceptable or unreasonable costs on society; regulatory policies that recognize that the private sector and private markets are the best engine for economic growth; regulatory approaches that respect the role of State, local, and tribal governments; and regulations that are effective, consistent, sensible, and understandable. We do not have such a regulatory system today.

With this Executive order, the Federal Government begins a program to reform and make more efficient the regulatory process. The objectives of this Executive order are to enhance planning and coordination with respect to both new and existing regulations; to reaffirm the primacy of Federal agencies in the regulatory decision-making process; to restore the integrity and legitimacy of regulatory review and oversight; and to make the process more accessible and open to the public. In pursuing these objectives, the regulatory process shall be conducted so as to meet applicable statutory requirements and with due regard to the discretion that has been entrusted to the Federal agencies.

Accordingly, by the authority vested in me as President by the Constitution and the laws of the United States of America, it is hereby ordered as follows:

Section 1. Statement of Regulatory Philosophy and Principles. (a) The Regulatory Philosophy. Federal agencies should promulgate only such regulations as are required by law, are necessary to interpret the law, or are made necessary by compelling public need, such as material failures of private markets to protect or improve the health and safety of the public, the environment, or the well-being of the American people. In deciding whether and how to regulate, agencies should assess all costs and benefits of available regulatory alternatives, including the alternative of not regulating. Costs and benefits shall be understood to include both quantifiable measures (to the fullest extent that these can be usefully estimated) and qualitative measures of cost and benefits that are difficult to quantify, but nevertheless essential to consider. Further, in choosing among alternative regulatory approaches, agencies should select those approaches that maximize net benefits (including potential economic, environmental, public health and safety, and other advantages; distributive impacts; and equity), unless a statute requires another regulatory approach.

(b) The Principles of Regulation. To ensure that the agencies' regulatory programs are consistent with the philosophy set forth above, agencies should adhere to the following principles, to the extent permitted by law and where applicable:

- (1) Each agency shall identify the problem that it intends to address (including, where applicable, the failures of private markets or public institutions that warrant new agency action) as well as assess the significance of that problem.
- (2) Each agency shall examine whether existing regulations (or other law) have created, or contributed to, the problem that a new regulation is intended to correct and whether those regulations (or other law) should be modified to achieve the intended goal of regulation more effectively.
- (3) Each agency shall identify and assess available alternatives to direct regulation, including providing economic incentives to encourage the desired behavior, such as user fees or marketable permits, or providing information upon which choices can be made by the public.
- (4) In setting regulatory priorities, each agency shall consider, to the extent reasonable, the degree and nature of the risks posed by various substances or activities within its jurisdiction.
- (5) When an agency determines that a regulation is the best available method of achieving the regulatory objective, it shall design its regulations in the most cost-effective manner to achieve the regulatory objective. In doing so, each agency shall consider incentives for innovation, consistency, predictability, the costs of enforcement and compliance (to the government, regulated entities, and the public), flexibility, distributive impacts, and equity.
- (6) Each agency shall assess both the costs and the benefits of the intended regulation and, recognizing that some costs and benefits are difficult to quantify, propose or adopt a regulation only upon a reasoned determination that the benefits of the intended regulation justify its costs.
- (7) Each agency shall base its decisions on the best reasonably obtainable scientific, technical, economic, and other information concerning the need for, and consequences of, the intended regulation.
- (8) Each agency shall identify and assess alternative forms of regulation and shall, to the extent feasible, specify performance objectives, rather than specifying the behavior or manner of compliance that regulated entities must adopt.
- (9) Wherever feasible, agencies shall seek views of appropriate State, local, and tribal officials before imposing regulatory requirements that might significantly or uniquely affect those governmental entities. Each agency shall assess the effects of Federal regulations on State, local, and tribal governments, including specifically the availability of resources to carry out those mandates, and seek to minimize those burdens that uniquely or significantly affect such governmental entities, consistent with achieving regulatory objectives. In addition, as appropriate, agencies shall seek to harmonize Federal regulatory actions with related State, local, and tribal regulatory and other governmental functions.
- (10) Each agency shall avoid regulations that are inconsistent, incompatible, or duplicative with its other regulations or those of other Federal agencies.
- (11) Each agency shall tailor its regulations to impose the least burden on society, including individuals, businesses of differing sizes, and other entities (including small

communities and government entities), consistent with obtaining the regulatory objectives, taking into account, among other things, and to the extent practicable, the costs of cumulative regulations.

(12) Each agency shall draft its regulations to be simple and easy to understand, with the goal of minimizing the potential for uncertainty and litigation arising from such uncertainty.

Sec. 2. Organization. An efficient regulatory planning and review process is vital to ensure that the Federal Government's regulatory system best serves the American people. (a) The Agencies. Because Federal agencies are the repositories of significant substantive expertise and experience, they are responsible for developing regulations and assuring that the regulations are consistent with applicable law, the President's priorities, and the principles set forth in this Executive order.

(b) The Office of Management and Budget. Coordinated review of agency rulemaking is necessary to ensure that regulations are consistent with applicable law, the President's priorities, and the principles set forth in this Executive order, and that decisions made by one agency do not conflict with the policies or actions taken or planned by another agency. The Office of Management and Budget (OMB) shall carry out that review function. Within OMB, the Office of Information and Regulatory Affairs (OIRA) is the repository of expertise concerning regulatory issues, including methodologies and procedures that affect more than one agency, this Executive order, and the President's regulatory policies. To the extent permitted by law, OMB shall provide guidance to agencies and assist the President, the Vice President, and other regulatory policy advisors to the President in regulatory planning and shall be the entity that reviews individual regulations, as provided by the this Executive order.

(c) The Vice President. The Vice President is the principal advisor to the President on, and shall coordinate the development and presentation of recommendations concerning, regulatory policy, planning, and review, as set forth in this Executive order. In fulfilling their responsibilities under this Executive order, the President and the Vice President shall be assisted by the regulatory policy advisors within the Executive Office of the President and by such agency officials and personnel as the President and the Vice President may, from time to time, consult.

Sec. 3. Definitions. for purposes of this Executive order: (1) "Advisors" refers to such regulatory policy advisors to the President as the President and Vice President may from time to time consult, including, among the others: (1)the Director of OMB; (2) the Chair (or another member) of the Council of Economic Advisers; (3) the Assistant to the President for Economic Policy; (4) the Assistance to the President for Domestic Policy; (5) the Assistant to the President for National Security Affairs; (6) the Assistant to the President for Science and Technology; (7) the Assistant to the President for Intergovernmental Affairs; (8) the Assistant to the President and Staff Secretary; (9) the Assistant to the President and Chief of Staff to the Vice President; (10) the Assistant to the President and Counsel to the President; (11) the Deputy Assistant to the President and Director of the White House Office of Environmental Policy; and (12) the Administrator of OIRA, who also shall coordinate communications relating to this Executive order among the agencies, OMB, the other Advisors, and the Office of the Vice President.

(b) "Agency," unless otherwise indicated, means any authority of the United States that is an "agency" under 44 U.S.C. 3502(1), other than those considered to be independent regulatory agencies, as defined in 44 U.S.C. 3502(10).

(c) "Director" means the Director of OMB.

(d) "Regulation" or "rule" means an agency statement of general applicability and future effect, which the agency intends to have the force and effect of law, that is designed to implement, interpret, or prescribe law or policy or to describe the procedure or practice requirements of an agency. It does not, however, include:

(1) Regulations or rules issued in accordance with the formal rulemaking provisions of 5 U.S.C. 556, 557;

(2) Regulations or rules that pertain to a military or foreign affairs function of the United States, other than procurement regulations and regulations involving the import or export of non-defense articles and services;

(3) Regulations or rules that are limited to agency organization, management, or personnel matters; or

(4) Any other category of regulations exempted by the Administrator of OIRA.

(e) "Regulatory action" means any substantive action by an agency (normally published in the Federal Register) that promulgates or is expected to lead to the promulgation of a final rule or regulation, including notices of inquiry, advance notices of proposed rulemaking, and notices of proposed rulemaking.

(f) "Significant regulatory action" means any regulatory action that is likely to result in a rule that may:

(1) Have an annual effect on the economy of \$100 million or more or adversely affect in a material way the economy, a sector of the economy, productivity, competition, jobs, the environment, public health or safety, or State, local, or tribal governments or communities;

(2) Create a serious inconsistency or otherwise interfere with an action taken or planned by another agency;

(3) Materially alter the budgetary impact of entitlements, grants, user fees, or loan programs or the rights and obligations of recipients thereof; or

(4) Raise novel legal or policy issues arising out of legal mandates, the President's priorities, or the principles set forth in this Executive order.

Sec. 4. Planning Mechanism. In order to have an effective regulatory program, to provide for coordination of regulations, to maximize consultation and the resolution of potential conflicts at an early stage, to involve the public and its State, local, and tribal officials in regulatory planning, and to ensure that new or revised regulations promote the President's priorities and the principles set forth in this Executive order, these procedures shall be followed, to the extent permitted by law:

(a) Agencies' Policy Meeting. Early in each year's planning cycle, the Vice President shall convene a meeting of the Advisors and the heads of agencies to seek a common

understanding of priorities and to coordinate regulatory efforts to be accomplished in the upcoming years.

(b) Unified Regulatory Agenda. For purposes of this subsection, the term "agency" or "agencies" shall also include those considered to be independent regulatory agencies, as defined in 44 U.S.C. 3502(10). Each agency shall prepare an agenda of all regulations under development or review, at a time and in a manner specified by the Administrator of OIRA. The description of each regulatory action shall contain, at a minimum, a regulation identifier number, a brief summary of the action, the legal authority for the action, any legal deadline for the action, and the name and telephone number of a knowledgeable agency official. Agencies may incorporate the information required under 5 U.S.C. 602 and 41 U.S.C. 402 into these agendas.

(c) The Regulatory Plan. For purposes of this subsection, the term "agency" or "agencies" shall also include those considered to be independent regulatory agencies, as defined in 44 U.S.C. 3502(10). (1) As part of the Unified Regulatory Agenda, beginning in 1994, each agency shall prepare a Regulatory Plan (Plan) of the most important significant regulatory actions that the agency reasonably expects to issue in proposed or final form in that fiscal year or thereafter. The Plan shall be approved personally by the agency head and shall contain at a minimum:

- (A) A statement of the agency's regulatory objectives and priorities and how they relate to the President's priorities;
- (B) A summary of each planned significant regulatory action including, to the extent possible, alternatives to be considered and preliminary estimates of the anticipated costs and benefits;
- (C) A summary of the legal basis for each such action, including whether any aspect of the action is required by statute or court order;
- (D) A statement of the need for each such action and, if applicable, how the action will reduce risks to public health, safety, or the environment, as well as how the magnitude of the risk addressed by the action relates to other risks within the jurisdiction of the agency;
- (E) The agency's schedule for action, including a statement of any applicable statutory or judicial deadlines; and
- (F) The name, address, and telephone number of a person the public may contact for additional information about the planned regulatory action.

(2) Each agency shall forward its Plan to OIRA by June 1st of each year.

(3) Within 10 calendar days after OIRA has received an agency's Plan, OIRA shall circulate it to other affected agencies, the Advisors, and the Vice President.

(4) An agency head who believes that a planned regulatory action of another agency may conflict with its own policy or action taken or planned shall promptly notify, in writing, the Administrator of OIRA, who shall forward that communication to the issuing agency, the Advisors, and the Vice President.

(5) If the Administrator of OIRA believes that planned regulatory action of an agency may be inconsistent with the President's priorities or the principles set forth in this Executive order or may be in conflict with any policy or action taken or planned by another agency, the Administrator of OIRA shall promptly notify, in writing, the effected agencies, the Advisors, and the Vice President.

(6) The Vice President, with the Advisors' assistance, may consult with the heads of agencies with respect to their Plans and, in appropriate instances, request further consideration or inter-agency coordination.

(7) The Plans developed by the issuing agency shall be published annually in the October publication of the Unified Regulatory Agenda. This publication shall be made available to the Congress; State, local, and tribal governments; and the public. Any views on any aspect of any agency Plan, including whether any planned regulatory action might conflict with any other planned or existing regulation, impose any unintended consequences on the public, or confer any unclaimed benefits on the public, should be directed to the issuing agency, with a copy to OIRA.

(d) *Regulatory Working Group.* Within 30 days of the date of this Executive order, the Administrator of OIRA shall convene a Regulatory Working Group ("Working Group"), which shall consist of representatives of the heads of each agency that the Administrator determines to have significant domestic regulatory responsibility, the Advisors, and the Vice President. The Administrator of OIRA shall chair the Working Group and shall periodically advise the Vice President on the activities of the Working Group. The Working Group shall serve as a forum to assist agencies in identifying and analyzing important regulatory issues (including, among others (1) the development of innovative regulatory techniques, (2) the methods, efficacy, and utility of comparative risk assessment in regulatory decision-making, and (3) the development of short forms and other streamlined regulatory approaches for small businesses and other entities). The Working Group shall meet at least quarterly and may meet as a whole or in subgroups of agencies with an interest in particular issues or subject areas. To inform its discussions, the Working Group may commission analytical studies and reports by OIRA, the Administrative Conference of the United States, or any other agency.

(e) *Conferences.* The Administrator of OIRA shall meet quarterly with representatives of State, local, and tribal governments to identify both existing and proposed regulations that may uniquely or significantly affect those governmental entities. The Administrator of OIRA shall also convene, from time to time, conferences with representatives of businesses, nongovernmental organizations, and the public to discuss regulatory issues of common concern.

Sec. 5. Existing Regulations. In order to reduce the regulatory burden on the American people, their families, their communities, their State, local, and tribal governments, and their industries; to determine whether regulations promulgated by the executive branch of the Federal Government have become unjustified or unnecessary as a result of changed circumstances; to confirm that regulations are both compatible with each other and not duplicative or inappropriately burdensome in the aggregate; to ensure that all regulations are consistent with the President's priorities and the principles set forth in this Executive order, within applicable law; and to otherwise improve the effectiveness of existing regulations: (1) Within 90 days of the date of this Executive order, each agency shall submit to OIRA a program, consistent with its resources and regulatory priorities, under which the agency will periodically review its existing

significant regulations to determine whether any such regulations should be modified or eliminated so as to make the agency's regulatory program more effective in achieving the regulatory objectives, less burdensome, or in greater alignment with the President's priorities and the principles set forth in this Executive order. Any significant regulations selected for review shall be included in the agency's annual Plan. The agency shall also identify any legislative mandates that require the agency to promulgate or continue to impose regulations that the agency believes are unnecessary or outdated by reason of changed circumstances.

(b) The Administrative of OIRA shall work with the Regulatory Working Group and other interested entities to pursue the objectives of this section. State, local, and tribal governments are specifically encouraged to assist in the identification of regulations that impose significant or unique burdens on those governmental entities and that appear to have outlived their justification or be otherwise inconsistent with the public interest.

(c) The Vice President, in consultation with the Advisors, may identify for review by the appropriate agency or agencies other existing regulations of an agency or groups of regulations of more than one agency that affect a particular group, industry, or sector of the economy, or may identify legislative mandates that may be appropriate for reconsideration by the Congress.

Sec. 6. Centralized Review of Regulations. The guidelines set forth below shall apply to all regulatory actions, for both new and existing regulations, by agencies other than those agencies specifically exempted by the Administrator of OIRA:

(a) Agency Responsibilities. (1) Each agency shall (consistent with its own rules, regulations, or procedures) provide the public with meaningful participation in the regulatory process. In particular, before issuing a notice of proposed rulemaking, each agency should, where appropriate, seek the involvement of those who are intended to benefit from and those expected to be burdened by any regulation (including, specifically, State, local, and tribal officials). In addition, each agency should afford the public a meaningful opportunity to comment on any proposed regulation, which in most cases should include a comment period of not less than 60 days. Each agency also is directed to explore and, where appropriate, use consensual mechanisms for developing regulations, including negotiated rulemaking.

(2) Within 60 days of the date of this Executive order, each agency head shall designate a Regulatory Policy Officer who shall report to the agency head. The Regulatory Policy Officer shall be involved at each stage of the regulatory process to foster the development of effective, innovative, and least burdensome regulations and to further the principles set forth in this Executive order.

(3) In addition to adhering to its own rules and procedures and to the requirements of the Administrative Procedure Act, the Regulatory Flexibility Act, the Paperwork Reduction Act, and other applicable law, each agency shall develop its regulatory actions in a timely fashion and adhere to the following procedures with respect to a regulatory action:

(A) Each agency shall provide OIRA, at such times and in the manner specified by the Administrator of OIRA, with a list of its planned regulatory actions, indicating those which the agency believes are significant regulatory actions within the meaning of this

Executive order. Absent a material change in the development of the planned regulatory action, those not designated as significant will not be subject to review under this section unless, within 10 working days of receipt of the list, the Administrator of OIRA notifies the agency that OIRA has determined that a planned regulation is a significant regulatory action within the meaning of this Executive order. The Administrator of OIRA may waive review of any planned regulatory action designated by the agency as significant, in which case the agency need not further comply with subsection (a)(3)(B) or subsection (a)(3)(C) of this section.

(B) For each matter identified as, or determined by the Administrator of OIRA to be, a significant regulatory action, the issuing agency shall provide to OIRA:

(i) The text of the draft regulatory action, together with a reasonably detailed description of the need for the regulatory action and an explanation of how the regulatory action will meet that need; and

(ii) An assessment of the potential costs and benefits of the regulatory action, including an explanation of the manner in which the regulatory action is consistent with a statutory mandate and, to the extent permitted by law, promotes the President's priorities and avoids undue interference with State, local, and tribal governments in the exercise of their governmental functions.

(C) For those matters identified as, or determined by the Administrator of OIRA to be, a significant regulatory action within the scope of section 3(f)(1), the agency shall also provide to OIRA the following additional information developed as part of the agency's decision-making process (unless prohibited by law):

(i) An assessment, including the underlying analysis, of benefits anticipated from the regulatory action (such as, but not limited to, the promotion of the efficient functioning of the economy and private markets, the enhancement of health and safety, the protection of the natural environment, and the elimination or reduction of discrimination or bias) together with, to the extent feasible, a quantification of those benefits;

(ii) An assessment, including the underlying analysis, of costs anticipated from the regulatory action (such as, but not limited to, the direct cost both to the government in administering the regulation and to businesses and others in complying with the regulation, and any adverse effects on the efficient functioning of the economy, private markets (including productivity, employment, and competitiveness), health, safety, and the natural environment), together with, to the extent feasible, a quantification of those costs; and

(iii) An assessment, including the underlying analysis, of costs and benefits of potentially effective and reasonably feasible alternatives to the planned regulation, identified by the agencies or the public (including improving the current regulation and reasonably viable nonregulatory actions), and an explanation why the planned regulatory action is preferable to the identified potential alternatives.

(D) In emergency situations or when an agency is obligated by law to act more quickly than normal review procedures allow, the agency shall notify OIRA as soon as possible and, to the extent practicable, comply with subsections (a)(3)(B) and (C) of this section. For those regulatory actions that are governed by a statutory or court-

imposed deadline, the agency shall, to the extent practicable, schedule rulemaking proceedings so as to permit sufficient time for OIRA to conduct its review, as set forth below in subsection (b)(2) through (4) of this section.

(E) After the regulatory action has been published in the Federal Register or otherwise issued to the public, the agency shall:

(i) Make available to the public the information set forth in subsections (a)(3)(B) and (C);

(ii) Identify for the public, in a complete, clear, and simple manner, the substantive changes between the draft submitted to OIRA for review and the action subsequently announced; and

(iii) Identify for the public those changes in the regulatory action that were made at the suggestion or recommendation of OIRA.

(F) All information provided to the public by the agency shall be in plain, understandable language.

(b) OIRA Responsibilities. The Administrator of OIRA shall provide meaningful guidance and oversight so that each agency's regulatory actions are consistent with applicable law, the President's priorities, and the principles set forth in this Executive order and do not conflict with the policies or actions of another agency. OIRA shall, to the extent permitted by law, adhere to the following guidelines:

(1) OIRA may review only actions identified by the agency or by OIRA as significant regulatory actions under subsection (a)(3)(A) of this section.

(2) OIRA shall waive review or notify the agency in writing of the results of its review within the following time periods:

(A) For any notices of inquiry, advance notices of proposed rulemaking, or other preliminary regulatory actions prior to a Notice of Proposed Rulemaking, within 10 working days after the date of submission of the draft action to OIRA;

(B) For all other regulatory actions, within 90 calendar days after the date of submission of the information set forth in subsections (a)(3)(B) and (C) of this section, unless OIRA has previously reviewed this information and, since that review, there has been no material change in the facts and circumstances upon which the regulatory action is based, in which case, OIRA shall complete its review within 45 days; and

(C) The review process may be extended (1) once by no more than 30 calendar days upon the written approval of the Director and (2) at the request of the agency head.

(3) For each regulatory action that the Administrator of OIRA returns to an agency for further consideration of some or all of its provisions, the Administrator of OIRA shall provide the issuing agency a written explanation for such return, setting forth the pertinent provision of this Executive order on which OIRA is relying. If the agency head disagrees with some or all of the bases for the return, the agency head shall so inform the Administrator of OIRA in writing.

(4) Except as otherwise provided by law or required by a Court, in order to ensure greater openness, accessibility, and accountability in the regulatory review process, OIRA shall be governed by the following disclosure requirements:

(A) Only the Administrator of OIRA (or a particular designee) shall receive oral communications initiated by persons not employed by the executive branch of the Federal Government regarding the substance of a regulatory action under OIRA review;

(B) All substantive communications between OIRA personnel and persons not employed by the executive branch of the Federal Government regarding a regulatory action under review shall be governed by the following guidelines: (i) A representative from the issuing agency shall be invited to any meeting between OIRA personnel and such person(s);

(ii) OIRA shall forward to the issuing agency, within 10 working days of receipt of the communication(s), all written communications, regardless of format, between OIRA personnel and any person who is not employed by the executive branch of the Federal Government, and the dates and names of individuals involved in all substantive oral communications (including meetings to which an agency representative was invited, but did not attend, and telephone conversations between OIRA personnel and any such persons); and

(iii) OIRA shall publicly disclose relevant information about such communication(s), as set forth below in subsection (b)(4)(C) of this section.

(C) OIRA shall maintain a publicly available log that shall contain, at a minimum, the following information pertinent to regulatory actions under review:

(i) The status of all regulatory actions, including if (and if so, when and by whom) Vice Presidential and Presidential consideration was requested;

(ii) A notation of all written communications forwarded to an issuing agency under subsection (b)(4)(B)(ii) of this section; and

(iii) The dates and names of individuals involved in all substantive oral communications, including meetings and telephone conversations, between OIRA personnel and any person not employed by the executive branch of the Federal Government, and the subject matter discussed during such communications.

(D) After the regulatory action has been published in the Federal Register or otherwise issued to the public, or after the agency has announced its decision not to publish or issue the regulatory action, OIRA shall make available to the public all documents exchanged between OIRA and the agency during the review by OIRA under this section.

(5) All information provided to the public by OIRA shall be in plain, understandable language.

Sec. 7. Resolution of Conflicts. To the extent permitted by law, disagreements or conflicts between or among agency heads or between OMB and any agency that cannot be resolved by the Administrator of OIRA shall be resolved by the President, or by the Vice President acting at the request of the President, with the relevant agency

head (and, as appropriate, other interested government officials). Vice Presidential and Presidential consideration of such disagreements may be initiated only by the Director, by the head of the issuing agency, or by the head of an agency that has a significant interest in the regulatory action at issue. Such review will not be undertaken at the request of other persons, entities, or their agents.

Resolution of such conflicts shall be informed by recommendations developed by the Vice President, after consultation with the Advisors (and other executive branch officials or personnel whose responsibilities to the President include the subject matter at issue). The development of these recommendations shall be concluded within 60 days after review has been requested.

During the Vice Presidential and Presidential review period, communications with any person not employed by the Federal Government relating to the substance of the regulatory action under review and directed to the Advisors or their staffs or to the staff of the Vice President shall be in writing and shall be forwarded by the recipient to the affected agency(ies) for inclusion in the public docket(s). When the communication is not in writing, such Advisors or staff members shall inform the outside party that the matter is under review and that any comments should be submitted in writing.

At the end of this review process, the President, or the Vice President acting at the request of the President, shall notify the affected agency and the Administrator of OIRA of the President's decision with respect to the matter.

Sec. 8. Publication. Except to the extent required by law, an agency shall not publish in the Federal Register or otherwise issue to the public any regulatory action that is subject to review under section 6 of this Executive order until (1) the Administrator of OIRA notifies the agency that OIRA has waived its review of the action or has completed its review without any requests for further consideration, or (2) the applicable time period in section 6(b)(2) expires without OIRA having notified the agency that it is returning the regulatory action for further consideration under section 6(b)(3), whichever occurs first. If the terms of the preceding sentence have not been satisfied and an agency wants to publish or otherwise issue a regulatory action, the head of that agency may request Presidential consideration through the Vice President, as provided under section 7 of this order. Upon receipt of this request, the Vice President shall notify OIRA and the Advisors. The guidelines and time period set forth in section 7 shall apply to the publication of regulatory actions for which Presidential consideration has been sought.

Sec. 9. Agency Authority. Nothing in this order shall be construed as displacing the agencies' authority or responsibilities, as authorized by law.

Sec. 10. Judicial Review. Nothing in this Executive order shall affect any otherwise available judicial review of agency action. This Executive order is intended only to improve the internal management of the Federal Government and does not create any right or benefit, substantive or procedural, enforceable at law or equity by a party against the United States, its agencies or instrumentalities, its officers or employees, or any other person.

Sec. 11. Revocations. Executive Orders Nos. 12,291 and 12,498; all amendments to those Executive orders; all guidelines issued under those orders; and any exemptions from those orders heretofore granted for any category of rule are revoked.

William J. Clinton

The White House, September 30, 1993.

Filed with the Office of the Federal Register, 12:12 pm., October 1, 1993

APPENDIX E EXECUTIVE ORDER 13,272

TITLE 3--

THE PRESIDENT

EXECUTIVE ORDER 13,272 OF AUGUST 13, 2002

PROPER CONSIDERATION OF SMALL ENTITIES IN AGENCY RULEMAKING

By the authority vested in me as President by the Constitution and the laws of the United States of America, it is hereby ordered as follows:

Section 1. General Requirements. Each agency shall establish procedures and policies to promote compliance with the Regulatory Flexibility Act, as amended (5 U.S.C. 601 et seq.) (the "Act"). Agencies shall thoroughly review draft rules to assess and take appropriate account of the potential impact on small businesses, small governmental jurisdictions, and small organizations, as provided by the Act. The Chief Counsel for Advocacy of the Small Business Administration (Advocacy) shall remain available to advise agencies in performing that review consistent with the provisions of the Act.

Sec. 2. Responsibilities of Advocacy. Consistent with the requirements of the Act, other applicable law, and Executive Order 12,866 of September 30, 1993, as amended, Advocacy:

(a) shall notify agency heads from time to time of the requirements of the Act, including by issuing notifications with respect to the basic requirements of the Act within 90 days of the date of this order;

(b) shall provide training to agencies on compliance with the Act; and

(c) may provide comment on draft rules to the agency that has proposed or intends to propose the rules and to the Office of Information and Regulatory Affairs of the Office of Management and Budget (OIRA).

Sec. 3. Responsibilities of Federal Agencies. Consistent with the requirements of the Act and applicable law, agencies shall:

(a) Within 180 days of the date of this order, issue written procedures and policies, consistent with the Act, to ensure that the potential impacts of agencies' draft rules on small businesses, small governmental jurisdictions, and small organizations are properly considered during the rulemaking process. Agency heads shall submit, no later than 90 days from the date of this order, their written procedures and policies to Advocacy for comment. Prior to issuing final procedures and policies, agencies shall consider any such comments received within 60 days from the date of the submission of the agencies' procedures and policies to Advocacy. Except to the extent otherwise specifically provided by statute or Executive Order, agencies shall make the final

procedures and policies available to the public through the Internet or other easily accessible means;

(b) Notify Advocacy of any draft rules that may have a significant economic impact on a substantial number of small entities under the Act. Such notifications shall be made (i) when the agency submits a draft rule to OIRA under Executive Order 12,866 if that order requires such submission, or (ii) if no submission to OIRA is so required, at a reasonable time prior to publication of the rule by the agency; and

(c) Give every appropriate consideration to any comments provided by Advocacy regarding a draft rule. Consistent with applicable law and appropriate protection of executive deliberations and legal privileges, an agency shall include, in any explanation or discussion accompanying publication in the Federal Register of a final rule, the agency's response to any written comments submitted by Advocacy on the proposed rule that preceded the final rule; provided, however, that inclusion is not required if the head of the agency certifies that the public interest is not served thereby.

Agencies and Advocacy may, to the extent permitted by law, engage in an exchange of data and research, as appropriate, to foster the purposes of the Act.

Sec. 4. Definitions. Terms defined in section 601 of title 5, United States Code, including the term "agency," shall have the same meaning in this order.

Sec. 5. Preservation of Authority. Nothing in this order shall be construed to impair or affect the authority of the Administrator of the Small Business Administration to supervise the Small Business Administration as provided in the first sentence of section 2(b)(1) of Public Law 85-09536 (15 U.S.C. 633(b)(1)).

Sec. 6. Reporting. For the purpose of promoting compliance with this order, Advocacy shall submit a report not less than annually to the Director of the Office of Management and Budget on the extent of compliance with this order by agencies.

Sec. 7. Confidentiality. Consistent with existing law, Advocacy may publicly disclose information that it receives from the agencies in the course of carrying out this order only to the extent that such information already has been lawfully and publicly disclosed by OIRA or the relevant rulemaking agency.

Sec. 8. Judicial Review. This order is intended only to improve the internal management of the Federal Government. This order is not intended to, and does not, create any right or benefit, substantive or procedural, enforceable at law or equity, against the United States, its departments, agencies, or other entities, its officers or employees, or any other person.

GEORGE W. BUSH

THE WHITE HOUSE,

August 13, 2002.

APPENDIX F EXECUTIVE ORDER 13,563 AND MEMORANDA

IMPROVING REGULATION AND REGULATORY REVIEW

By the authority vested in me as President by the Constitution and the laws of the United States of America, and in order to improve regulation and regulatory review, it is hereby ordered as follows:

Section 1. General Principles of Regulation. (a) Our regulatory system must protect public health, welfare, safety, and our environment while promoting economic growth, innovation, competitiveness, and job creation. It must be based on the best available science. It must allow for public participation and an open exchange of ideas. It must promote predictability and reduce uncertainty. It must identify and use the best, most innovative, and least burdensome tools for achieving regulatory ends. It must take into account benefits and costs, both quantitative and qualitative. It must ensure that regulations are accessible, consistent, written in plain language, and easy to understand. It must measure, and seek to improve, the actual results of regulatory requirements.

(b) This order is supplemental to and reaffirms the principles, structures, and definitions governing contemporary regulatory review that were established in Executive Order 12,866 of September 30, 1993. As stated in that Executive Order and to the extent permitted by law, each agency must, among other things: (1) propose or adopt a regulation only upon a reasoned determination that its benefits justify its costs (recognizing that some benefits and costs are difficult to quantify); (2) tailor its regulations to impose the least burden on society, consistent with obtaining regulatory objectives, taking into account, among other things, and to the extent practicable, the costs of cumulative regulations; (3) select, in choosing among alternative regulatory approaches, those approaches that maximize net benefits (including potential economic, environmental, public health and safety, and other advantages; distributive impacts; and equity); (4) to the extent feasible, specify performance objectives, rather than specifying the behavior or manner of compliance that regulated entities must adopt; and (5) identify and assess available alternatives to direct regulation, including providing economic incentives to encourage the desired behavior, such as user fees or marketable permits, or providing information upon which choices can be made by the public.

(c) In applying these principles, each agency is directed to use the best available techniques to quantify anticipated present and future benefits and costs as accurately as possible. Where appropriate and permitted by law, each agency may consider (and discuss qualitatively) values that are difficult or impossible to quantify, including equity, human dignity, fairness, and distributive impacts.

Sec. 2. Public Participation. (a) Regulations shall be adopted through a process that involves public participation. To that end, regulations shall be based, to the extent feasible and consistent with law, on the open exchange of information and perspectives among State, local, and tribal officials, experts in relevant disciplines, affected stakeholders in the private sector, and the public as a whole.

(b) To promote that open exchange, each agency, consistent with Executive Order 12,866 and other applicable legal requirements, shall endeavor to provide the public with an opportunity to participate in the regulatory process. To the extent feasible and permitted by law, each agency shall afford the public a meaningful opportunity to comment through the Internet on any proposed regulation, with a comment period that should generally be at least 60 days. To the extent feasible and permitted by law, each agency shall also provide, for both proposed and final rules, timely online access to the rulemaking docket on regulations.gov, including relevant scientific and technical findings, in an open format that can be easily searched and downloaded. For proposed rules, such access shall include, to the extent feasible and permitted by law, an opportunity for public comment on all pertinent parts of the rulemaking docket, including relevant scientific and technical findings.

(c) Before issuing a notice of proposed rulemaking, each agency, where feasible and appropriate, shall seek the views of those who are likely to be affected, including those who are likely to benefit from and those who are potentially subject to such rulemaking.

Sec. 3. Integration and Innovation. Some sectors and industries face a significant number of regulatory requirements, some of which may be redundant, inconsistent, or overlapping. Greater coordination across agencies could reduce these requirements, thus reducing costs and simplifying and harmonizing rules. In developing regulatory actions and identifying appropriate approaches, each agency shall attempt to promote such coordination, simplification, and harmonization. Each agency shall also seek to identify, as appropriate, means to achieve regulatory goals that are designed to promote innovation.

Sec. 4. Flexible Approaches. Where relevant, feasible, and consistent with regulatory objectives, and to the extent permitted by law, each agency shall identify and consider regulatory approaches that reduce burdens and maintain flexibility and freedom of choice for the public. These approaches include warnings, appropriate default rules, and disclosure requirements as well as provision of information to the public in a form that is clear and intelligible.

Sec. 5. Science. Consistent with the President's Memorandum for the Heads of Executive Departments and Agencies,

"Scientific Integrity" (March 9, 2009), and its implementing guidance, each agency shall ensure the objectivity of any scientific and technological information and processes used to support the agency's regulatory actions.

Sec. 6. Retrospective Analyses of Existing Rules. (a) To facilitate the periodic review of existing significant regulations, agencies shall consider how best to promote retrospective analysis of rules that may be outmoded, ineffective, insufficient, or excessively burdensome, and to modify, streamline, expand, or repeal them in accordance with what has been learned. Such retrospective analyses, including supporting data, should be released online whenever possible.

(b) Within 120 days of the date of this order, each agency shall develop and submit to the Office of Information and Regulatory Affairs a preliminary plan, consistent with law and its resources and regulatory priorities, under which the agency will periodically review its existing significant regulations to determine whether any such

regulations should be modified, streamlined, expanded, or repealed so as to make the agency's regulatory program more effective or less burdensome in achieving the regulatory objectives.

Sec. 7. General Provisions. (a) For purposes of this order, "agency" shall have the meaning set forth in section 3(b) of Executive Order 12,866.

(b) Nothing in this order shall be construed to impair or otherwise affect:

(i) authority granted by law to a department or agency, or the head thereof; or

(ii) functions of the Director of the Office of Management and Budget relating to budgetary, administrative, or legislative proposals.

(c) This order shall be implemented consistent with applicable law and subject to the availability of appropriations.

(d) This order is not intended to, and does not, create any right or benefit, substantive or procedural, enforceable at law or in equity by any party against the United States, its departments, agencies, or entities, its officers, employees, or agents, or any other person.

BARACK OBAMA

THE WHITE HOUSE,
January 18, 2011.

The White House

Office of the Press Secretary

For Immediate Release

January 18, 2011

Presidential Memoranda - Regulatory Flexibility, Small Business, And Job Creation

Memorandum for the Heads of Executive Departments and Agencies

Subject: Regulatory Flexibility, Small Business, and Job Creation

Small businesses play an essential role in the American economy; they help to fuel productivity, economic growth, and job creation. More than half of all Americans working in the private sector either are employed by a small business or own one. During a recent 15-year period, small businesses created more than 60 percent of all new jobs in the Nation.

Although small businesses and new companies provide the foundations for economic growth and job creation, they have faced severe challenges as a result of the recession. One consequence has been the loss of significant numbers of jobs.

The Regulatory Flexibility Act (RFA), 5 U.S.C. 601-612, establishes a deep national commitment to achieving statutory goals without imposing unnecessary burdens on the public. The RFA emphasizes the importance of recognizing "differences in the scale and resources of regulated entities" and of considering "alternative regulatory approaches . . . which minimize the significant economic impact of rules on small businesses, small organizations, and small governmental jurisdictions." 5 U.S.C. 601 note.

To promote its central goals, the RFA imposes a series of requirements designed to ensure that agencies produce regulatory flexibility analyses that give careful consideration to the effects of their regulations on small businesses and explore significant alternatives in order to minimize any significant economic impact on small businesses. Among other things, the RFA requires that when an agency proposing a rule with such impact is required to provide notice of the proposed rule, it must also produce an initial regulatory flexibility analysis that includes discussion of significant alternatives. Significant alternatives include the use of performance rather than design standards; simplification of compliance and reporting requirements for small businesses; establishment of different timetables that take into account the resources of small businesses; and exemption from coverage for small businesses.

Consistent with the goal of open government, the RFA also encourages public participation in and transparency about the rulemaking process. Among other things, the statute requires

agencies proposing rules with a significant economic impact on small businesses to provide an opportunity for public comment on any required initial regulatory flexibility analysis, and generally requires agencies promulgating final rules with such significant economic impact to respond, in a final regulatory flexibility analysis, to comments filed by the Chief Counsel for Advocacy of the Small Business Administration.

My Administration is firmly committed to eliminating excessive and unjustified burdens on small businesses, and to ensuring that regulations are designed with careful consideration of their effects, including their cumulative effects, on small businesses. Executive Order 12,866 of September 30, 1993, as amended, states, "Each agency shall tailor its regulations to impose the least burden on society, including individuals, businesses of differing sizes, and other entities (including small communities and governmental entities), consistent with obtaining the regulatory objectives, taking into account, among other things, and to the extent practicable, the costs of cumulative regulations."

In the current economic environment, it is especially important for agencies to design regulations in a cost-effective manner consistent with the goals of promoting economic growth, innovation, competitiveness, and job creation.

Accordingly, I hereby direct executive departments and agencies and request independent agencies, when initiating rulemaking that will have a significant economic impact on a substantial number of small entities, to give serious consideration to whether and how it is appropriate, consistent with law and regulatory objectives, to reduce regulatory burdens on small businesses, through increased flexibility. As the RFA recognizes, such flexibility may take many forms, including:

- extended compliance dates that take into account the resources available to small entities;
- performance standards rather than design standards;
- simplification of reporting and compliance requirements (as, for example, through streamlined forms and electronic filing options);
- different requirements for large and small firms; and
- partial or total exemptions.

I further direct that whenever an executive agency chooses, for reasons other than legal limitations, not to provide such flexibility in a proposed or final rule that is likely to have a significant economic impact on a substantial number of small entities, it should explicitly justify its decision not to do so in the explanation that accompanies that proposed or final rule.

Adherence to these requirements is designed to ensure that regulatory actions do not place unjustified economic burdens on small business owners and other small entities. If regulations are preceded by careful analysis, and subjected to public comment, they are less likely to be based on intuition and guesswork and more likely to be justified in light of a clear understanding of the likely consequences of alternative courses of action. With that understanding, agencies will be in a better position to protect the public while avoiding excessive costs and paperwork.

This memorandum is not intended to, and does not, create any right or benefit, substantive or procedural, enforceable at law or in equity by any party against the United States, its departments, agencies, or entities, its officers, employees, or agents, or any other person. Nothing in this memorandum shall be construed to impair or otherwise affect the functions of the Director of the Office of Management and Budget relating to budgetary, administrative, or legislative proposals.

The Director of the Office of Management and Budget is authorized and directed to publish this memorandum in the *Federal Register*.

BARACK OBAMA

The White House

Office of the Press Secretary

For Immediate Release

January 18, 2011

PRESIDENTIAL MEMORANDA - REGULATORY COMPLIANCE

Memorandum for the Heads of Executive Departments and Agencies

Subject: Regulatory Compliance

My Administration is committed to enhancing effectiveness and efficiency in Government. Pursuant to the Memorandum on Transparency and Open Government, issued on January 21, 2009, executive departments and agencies (agencies) have been working steadily to promote accountability, encourage collaboration, and provide information to Americans about their Government's activities.

To that end, much progress has been made toward strengthening our democracy and improving how Government operates. In the regulatory area, several agencies, such as the Department of Labor and the Environmental Protection Agency, have begun to post online (at ogesdw.dol.gov and www.epa-echo.gov), and to make readily accessible to the public, information concerning their regulatory compliance and enforcement activities, such as information with respect to administrative inspections, examinations, reviews, warnings, citations, and revocations (but excluding law enforcement or otherwise sensitive information about ongoing enforcement actions).

Greater disclosure of regulatory compliance information fosters fair and consistent enforcement of important regulatory obligations. Such disclosure is a critical step in encouraging the public to hold the Government and regulated entities accountable. Sound regulatory enforcement promotes the welfare of Americans in many ways, by increasing public safety, improving working conditions, and protecting the air we breathe and the water we drink. Consistent regulatory enforcement also levels the playing field among regulated entities, ensuring that those that fail to comply with the law do not have an unfair advantage over their law-abiding competitors. Greater agency disclosure of compliance and enforcement data will provide Americans with information they need to make informed decisions. Such disclosure can lead the Government to hold itself more accountable, encouraging agencies to identify and address enforcement gaps.

Accordingly, I direct the following:

First, agencies with broad regulatory compliance and administrative enforcement responsibilities, within 120 days of this memorandum, to the extent feasible and permitted by law, shall develop plans to make public information concerning their regulatory compliance and enforcement activities accessible, downloadable, and searchable online. In so doing, agencies should prioritize making accessible information that is most useful to the general public and should consider the use of new technologies to allow the public to have access to real-time data. The independent agencies are encouraged to comply with this directive.

Second, the Federal Chief Information Officer and the Chief Technology Officer shall work with appropriate counterparts in each agency to make such data available online in searchable form, including on centralized platforms such as data.gov, in a manner that facilitates easy access, encourages cross-agency comparisons, and engages the public in new and creative ways of using the information.

Third, the Federal Chief Information Officer and the Chief Technology Officer, in coordination with the Director of the Office of Management and Budget (OMB) and their counterparts in each agency, shall work to explore how best to generate and share enforcement and compliance information across the Government, consistent with law. Such data sharing can assist with agencies' risk-based approaches to enforcement: A lack of compliance in one area by a regulated entity may indicate a need for examination and closer attention by another agency. Efforts to share data across agencies, where appropriate and permitted by law, may help to promote flexible and coordinated enforcement regimes.

This memorandum is not intended to, and does not, create any right or benefit, substantive or procedural, enforceable at law or in equity by any party against the United States, its departments, agencies, or entities, its officers, employees, or agents, or any other person. Nothing in this memorandum shall be construed to impair or otherwise affect the functions of the Director of the Office of Management and Budget relating to budgetary, administrative, or legislative proposals.

The Director of OMB is authorized and directed to publish this memorandum in the *Federal Register*.

BARACK OBAMA

APPENDIX G MEMORANDUM ON CUMULATIVE IMPACTS

EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF MANAGEMENT AND BUDGET
WASHINGTON, D.C. 20503

ADMINISTRATOR
OFFICE OF
INFORMATION AND
REGULATORY
AFFAIRS

March 20, 2012

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: Cass R. Sunstein
Administrator

SUBJECT: Cumulative Effects of Regulations

On January 18, 2011, the President issued Executive Order 13,563, “Improving Regulation and Regulatory Review,” which states that to the extent permitted by law, each agency must take into account “among other things, and to the extent practicable, the costs of cumulative regulations.” Executive Order 13,563 emphasizes that some “sectors and industries face a significant number of regulatory requirements, some of which may be redundant, inconsistent, or overlapping,” and it directs agencies to promote “coordination, simplification, and harmonization.” Executive Order 13,563 also states that to the extent permitted by law, each agency shall “propose or adopt a regulation only upon a reasoned determination that its benefits justify its costs.”

Executive Order 13,563 directs that regulations “shall be adopted through a process that involves public participation,” including an “open exchange of information and perspectives.” Public participation can and should be used to evaluate the cumulative effects of regulations, for example through active engagement with affected stakeholders well before the issuance of notices of proposed rulemaking. The President’s Council on Jobs and Competitiveness has emphasized the need for a smart and efficient regulatory system and has drawn particular attention to the cumulative effects of regulation. Cumulative burdens can create special challenges for small businesses and startups.

Consistent with Executive Order 13,563, and to the extent permitted by law, agencies should take active steps to take account of the cumulative effects of new and existing rules and to identify opportunities to harmonize and streamline multiple rules. The goals of this effort should be to simplify requirements on the public and private sectors; to ensure against unjustified, redundant, or excessive requirements; and ultimately to increase the net benefits of regulations.

To promote consideration of cumulative effects, and to reduce redundant, overlapping, and inconsistent requirements, agencies should carefully consider the following steps, where appropriate and feasible, and to the extent permitted by law:

- Early consultation with, advance notice to, and close engagement with affected stakeholders to discuss potential interactions between rulemakings under consideration and existing regulations as well as other anticipated regulatory requirements;
- Early engagement with state, tribal, and local regulatory agencies to identify opportunities for harmonizing regulatory requirements, reducing administrative costs, avoiding unnecessary or inconsistent requirements, and otherwise improving regulatory outcomes;
- Use of Requests for Information and Advance Notices of Proposed Rulemaking to obtain public input on potentially overlapping rulemakings and on rulemakings that may have significant cumulative effects;
- Specific consideration of the cumulative effects of regulations on small businesses and start-ups;
- Identification of opportunities to increase the net benefits of regulations and to reduce administrative and other costs, while meeting policy goals and legal requirements;
- Careful consideration, in the analysis of costs and benefits, of the relationship between new regulations and regulations that are already in effect;
- Identification of opportunities to integrate and simplify the requirements of new and existing rules, so as to eliminate inconsistency and redundancy;
- Coordination of timing, content, and requirements of multiple rulemakings that are contemplated for a particular industry or sector, so as to increase net benefits; and
- Consideration of the interactive and cumulative effects of multiple regulations affecting individual sectors as part of agencies' retrospective analysis of existing rules, consistent with Executive Order 13,563.

Where appropriate and feasible, agencies should consider cumulative effects and opportunities for regulatory harmonization as part of their analysis of particular rules, and should carefully assess the appropriate content and timing of rules in light of those effects and opportunities. Consideration of cumulative effects and of opportunities to reduce burdens and to increase net benefits should be part of the assessment of costs and benefits, consistent with the requirement of Executive Order 13,563 that, to the extent permitted by law, agencies must “select, in choosing among alternative regulatory approaches, those approaches that maximize net benefits.” Agencies should avoid unintentional burdens that could result from an exclusive focus on the most recent regulatory activities. As noted, the cumulative effects on small businesses and start-ups deserve particular attention.

This guidance is effective immediately.

APPENDIX H EXECUTIVE ORDER 13,579

REGULATION AND INDEPENDENT REGULATORY AGENCIES

By the authority vested in me as President by the Constitution and the laws of the United States of America, and in order to improve regulation and regulatory review, it is hereby ordered as follows:

Section 1. Policy. (a) Wise regulatory decisions depend on public participation and on careful analysis of the likely consequences of regulation. Such decisions are informed and improved by allowing interested members of the public to have a meaningful opportunity to participate in rulemaking. To the extent permitted by law, such decisions should be made only after consideration of their costs and benefits (both quantitative and qualitative).

(b) Executive Order 13,563 of January 18, 2011, "Improving Regulation and Regulatory Review," directed to executive agencies, was meant to produce a regulatory system that protects "public health, welfare, safety, and our environment while promoting economic growth, innovation, competitiveness, and job creation." Independent regulatory agencies, no less than executive agencies, should promote that goal.

(c) Executive Order 13,563 set out general requirements directed to executive agencies concerning public participation, integration and innovation, flexible approaches, and science. To the extent permitted by law, independent regulatory agencies should comply with these provisions as well.

Sec. 2. Retrospective Analyses of Existing Rules. (a) To facilitate the periodic review of existing significant regulations, independent regulatory agencies should consider how best to promote retrospective analysis of rules that may be outmoded, ineffective, insufficient, or excessively burdensome, and to modify, streamline, expand, or repeal them in accordance with what has been learned. Such retrospective analyses, including supporting data and evaluations, should be released online whenever possible.

(b) Within 120 days of the date of this order, each independent regulatory agency should develop and release to the public a plan, consistent with law and reflecting its resources and regulatory priorities and processes, under which the agency will periodically review its existing significant regulations to determine whether any such regulations should be modified, streamlined, expanded, or repealed so as to make the agency's regulatory program more effective or less burdensome in achieving the regulatory objectives.

Sec. 3. General Provisions. (a) For purposes of this order, "executive agency" shall have the meaning set forth for the term "agency" in section 3(b) of Executive Order 12,866 of September 30, 1993, and "independent regulatory agency" shall have the meaning set forth in 44 U.S.C. 3502(5).

(b) Nothing in this order shall be construed to impair or otherwise affect:

(i) authority granted by law to a department or agency, or the head thereof; or

(ii) functions of the Director of the Office of Management and Budget relating to budgetary, administrative, or legislative proposals.

(c) This order shall be implemented consistent with applicable law and subject to the availability of appropriations.

(d) This order is not intended to, and does not, create any right or benefit, substantive or procedural, enforceable at law or in equity by any party against the United States, its departments, agencies, or entities, its officers, employees, or agents, or any other person.

BARACK OBAMA

THE WHITE HOUSE,
July 11, 2011.

APPENDIX I EXECUTIVE ORDER 13,610

THE WHITE HOUSE

OFFICE OF THE PRESS SECRETARY

FOR IMMEDIATE RELEASE, MAY 10, 2012

EXECUTIVE ORDER

IDENTIFYING AND REDUCING REGULATORY BURDENS

By the authority vested in me as President by the Constitution and the laws of the United States of America, and in order to modernize our regulatory system and to reduce unjustified regulatory burdens and costs, it is hereby ordered as follows:

Sec. 1. Policy. Regulations play an indispensable role in protecting public health, welfare, safety, and our environment, but they can also impose significant burdens and costs. During challenging economic times, we should be especially careful not to impose unjustified regulatory requirements. For this reason, it is particularly important for agencies to conduct retrospective analyses of existing rules to examine whether they remain justified and whether they should be modified or streamlined in light of changed circumstances, including the rise of new technologies.

Executive Order 13,563 of January 18, 2011 (Improving Regulation and Regulatory Review), states that our regulatory system "must measure, and seek to improve, the actual results of regulatory requirements." To promote this goal, that Executive Order requires agencies not merely to conduct a single exercise, but to engage in "periodic review of existing significant regulations." Pursuant to section 6(b) of that Executive Order, agencies are required to develop retrospective review plans to review existing significant regulations in order to "determine whether any such regulations should be modified, streamlined, expanded, or repealed." The purpose of this requirement is to "make the agency's regulatory program more effective or less burdensome in achieving the regulatory objectives."

In response to Executive Order 13,563, agencies have developed and made available for public comment retrospective review plans that identify over five hundred initiatives. A small fraction of those initiatives, already finalized or formally proposed to the public, are anticipated to eliminate billions of dollars in regulatory costs and tens of millions of hours in annual paperwork burdens. Significantly larger savings are anticipated as the plans are implemented and as action is taken on additional initiatives.

As a matter of longstanding practice and to satisfy statutory obligations, many agencies engaged in periodic review of existing regulations prior to the issuance of Executive Order 13,563. But further steps should be taken, consistent with law, agency resources, and regulatory priorities, to promote public participation in retrospective review, to modernize our regulatory system, and to institutionalize regular assessment of significant regulations.

Sec. 2. Public Participation in Retrospective Review. Members of the public, including those directly and indirectly affected by regulations, as well as State, local, and tribal governments, have important information about the actual effects of existing regulations. For this reason, and consistent with Executive Order 13,563, agencies shall invite, on a regular basis (to be determined by the agency head in consultation with the Office of Information and Regulatory Affairs (OIRA)), public suggestions about regulations in need of retrospective review and about appropriate modifications to such regulations. To promote an open exchange of information, retrospective analyses of regulations, including supporting data, shall be released to the public online wherever practicable.

Sec. 3. Setting Priorities. In implementing and improving their retrospective review plans, and in considering retrospective review suggestions from the public, agencies shall give priority, consistent with law, to those initiatives that will produce significant quantifiable monetary savings or significant quantifiable reductions in paperwork burdens while protecting public health, welfare, safety, and our environment. To the extent practicable and permitted by law, agencies shall also give special consideration to initiatives that would reduce unjustified regulatory burdens or simplify or harmonize regulatory requirements imposed on small businesses. Consistent with Executive Order 13,563 and Executive Order 12,866 of September 30, 1993 (Regulatory Planning and Review), agencies shall give consideration to the cumulative effects of their own regulations, including cumulative burdens, and shall to the extent practicable and consistent with law give priority to reforms that would make significant progress in reducing those burdens while protecting public health, welfare, safety, and our environment.

Sec. 4. Accountability. Agencies shall regularly report on the status of their retrospective review efforts to OIRA. Agency reports should describe progress, anticipated accomplishments, and proposed timelines for relevant actions, with an emphasis on the priorities described in section 3 of this order. Agencies shall submit draft reports to OIRA on September 10, 2012, and on the second Monday of January and July for each year thereafter, unless directed otherwise through subsequent guidance from OIRA. Agencies shall make final reports available to the public within a reasonable period (not to exceed three weeks from the date of submission of draft reports to OIRA).

Sec. 5. General Provisions. (a) For purposes of this order, "agency" means any authority of the United States that is an "agency" under 44 U.S.C. 3502(1), other than those considered to be independent regulatory agencies, as defined in 44 U.S.C. 3502(5).

(b) Nothing in this order shall be construed to impair or otherwise affect:

(i) the authority granted by law to a department or agency, or the head thereof; or
(ii) the functions of the Director of the Office of Management and Budget relating to budgetary, administrative, or legislative proposals.

(c) This order shall be implemented consistent with applicable law and subject to the availability of appropriations.

(d) This order is not intended to, and does not, create any right or benefit, substantive or procedural, enforceable at law or in equity by any party against the United States, its departments, agencies, or entities, its officers, employees, or agents, or any other person.

BARACK OBAMA

APPENDIX J EXECUTIVE ORDER 13,771

FOR IMMEDIATE RELEASE, JANUARY 30, 2017

PRESIDENTIAL EXECUTIVE ORDER ON REDUCING REGULATION AND CONTROLLING REGULATORY COSTS

EXECUTIVE ORDER

REDUCING REGULATION AND CONTROLLING REGULATORY COSTS

By the authority vested in me as President by the Constitution and the laws of the United States of America, including the Budget and Accounting Act of 1921, as amended (31 U.S.C. 1101 et seq.), section 1105 of title 31, United States Code, and section 301 of title 3, United States Code, it is hereby ordered as follows:

Section 1. Purpose. It is the policy of the executive branch to be prudent and financially responsible in the expenditure of funds, from both public and private sources. In addition to the management of the direct expenditure of taxpayer dollars through the budgeting process, it is essential to manage the costs associated with the governmental imposition of private expenditures required to comply with Federal regulations. Toward that end, it is important that for every one new regulation issued, at least two prior regulations be identified for elimination, and that the cost of planned regulations be prudently managed and controlled through a budgeting process.

Sec. 2. Regulatory Cap for Fiscal Year 2017. (a) Unless prohibited by law, whenever an executive department or agency (agency) publicly proposes for notice and comment or otherwise promulgates a new regulation, it shall identify at least two existing regulations to be repealed.

(b) For fiscal year 2017, which is in progress, the heads of all agencies are directed that the total incremental cost of all new regulations, including repealed regulations, to be finalized this year shall be no greater than zero, unless otherwise required by law or consistent with advice provided in writing by the Director of the Office of Management and Budget (Director).

(c) In furtherance of the requirement of subsection (a) of this section, any new incremental costs associated with new regulations shall, to the extent permitted by law, be offset by the elimination of existing costs associated with at least two prior regulations. Any agency eliminating existing costs associated with prior regulations under this subsection shall do so in accordance with the Administrative Procedure Act and other applicable law.

(d) The Director shall provide the heads of agencies with guidance on the implementation of this section. Such guidance shall address, among other things,

processes for standardizing the measurement and estimation of regulatory costs; standards for determining what qualifies as new and offsetting regulations; standards for determining the costs of existing regulations that are considered for elimination; processes for accounting for costs in different fiscal years; methods to oversee the issuance of rules with costs offset by savings at different times or different agencies; and emergencies and other circumstances that might justify individual waivers of the requirements of this section. The Director shall consider phasing in and updating these requirements.

Sec. 3. Annual Regulatory Cost Submissions to the Office of Management and Budget.

(a) Beginning with the Regulatory Plans (required under Executive Order 12,866 of September 30, 1993, as amended, or any successor order) for fiscal year 2018, and for each fiscal year thereafter, the head of each agency shall identify, for each regulation that increases incremental cost, the offsetting regulations described in section 2(c) of this order, and provide the agency's best approximation of the total costs or savings associated with each new regulation or repealed regulation.

(b) Each regulation approved by the Director during the Presidential budget process shall be included in the Unified Regulatory Agenda required under Executive Order 12,866, as amended, or any successor order.

(c) Unless otherwise required by law, no regulation shall be issued by an agency if it was not included on the most recent version or update of the published Unified Regulatory Agenda as required under Executive Order 12,866, as amended, or any successor order, unless the issuance of such regulation was approved in advance in writing by the Director.

(d) During the Presidential budget process, the Director shall identify to agencies a total amount of incremental costs that will be allowed for each agency in issuing new regulations and repealing regulations for the next fiscal year. No regulations exceeding the agency's total incremental cost allowance will be permitted in that fiscal year, unless required by law or approved in writing by the Director. The total incremental cost allowance may allow an increase or require a reduction in total regulatory cost.

(e) The Director shall provide the heads of agencies with guidance on the implementation of the requirements in this section.

Sec. 4. Definition. For purposes of this order the term "regulation" or "rule" means an agency statement of general or particular applicability and future effect designed to implement, interpret, or prescribe law or policy or to describe the procedure or practice requirements of an agency, but does not include:

(a) regulations issued with respect to a military, national security, or foreign affairs function of the United States;

(b) regulations related to agency organization, management, or personnel; or

(c) any other category of regulations exempted by the Director.

Sec. 5. General Provisions. (a) Nothing in this order shall be construed to impair or otherwise affect:

(i) the authority granted by law to an executive department or agency, or the head thereof; or

(ii) the functions of the Director relating to budgetary, administrative, or legislative proposals.

(b) This order shall be implemented consistent with applicable law and subject to the availability of appropriations.

(c) This order is not intended to, and does not, create any right or benefit, substantive or procedural, enforceable at law or in equity by any party against the United States, its departments, agencies, or entities, its officers, employees, or agents, or any other person.

DONALD J. TRUMP

THE WHITE HOUSE,

January 30, 2017.

APPENDIX K EXECUTIVE ORDER 13,777

THE WHITE HOUSE

OFFICE OF THE PRESS SECRETARY

FOR IMMEDIATE RELEASE, FEBRUARY 24, 2017

PRESIDENTIAL EXECUTIVE ORDER ON ENFORCING THE REGULATORY REFORM AGENDA

EXECUTIVE ORDER

ENFORCING THE REGULATORY REFORM AGENDA

By the authority vested in me as President by the Constitution and the laws of the United States of America, and in order to lower regulatory burdens on the American people by implementing and enforcing regulatory reform, it is hereby ordered as follows:

Section 1. Policy. It is the policy of the United States to alleviate unnecessary regulatory burdens placed on the American people.

Sec. 2. Regulatory Reform Officers. (a) Within 60 days of the date of this order, the head of each agency, except the heads of agencies receiving waivers under section 5 of this order, shall designate an agency official as its Regulatory Reform Officer (RRO). Each RRO shall oversee the implementation of regulatory reform initiatives and policies to ensure that agencies effectively carry out regulatory reforms, consistent with applicable law. These initiatives and policies include:

- (i) Executive Order 13,771 of January 30, 2017 (Reducing Regulation and Controlling Regulatory Costs), regarding offsetting the number and cost of new regulations;
- (ii) Executive Order 12,866 of September 30, 1993 (Regulatory Planning and Review), as amended, regarding regulatory planning and review;
- (iii) section 6 of Executive Order 13,563 of January 18, 2011 (Improving Regulation and Regulatory Review), regarding retrospective review; and
- (iv) the termination, consistent with applicable law, of programs and activities that derive from or implement Executive Orders, guidance documents, policy memoranda,

rule interpretations, and similar documents, or relevant portions thereof, that have been rescinded.

(b) Each agency RRO shall periodically report to the agency head and regularly consult with agency leadership.

Sec. 3. Regulatory Reform Task Forces. (a) Each agency shall establish a Regulatory Reform Task Force composed of:

- (i) the agency RRO;
- (ii) the agency Regulatory Policy Officer designated under section 6(a)(2) of Executive Order 12,866;
- (iii) a representative from the agency's central policy office or equivalent central office; and
- (iv) for agencies listed in section 901(b)(1) of title 31, United States Code, at least three additional senior agency officials as determined by the agency head.

(b) Unless otherwise designated by the agency head, the agency RRO shall chair the agency's Regulatory Reform Task Force.

(c) Each entity staffed by officials of multiple agencies, such as the Chief Acquisition Officers Council, shall form a joint Regulatory Reform Task Force composed of at least one official described in subsection (a) of this section from each constituent agency's Regulatory Reform Task Force. Joint Regulatory Reform Task Forces shall implement this order in coordination with the Regulatory Reform Task Forces of their members' respective agencies.

(d) Each Regulatory Reform Task Force shall evaluate existing regulations (as defined in section 4 of Executive Order 13,771) and make recommendations to the agency head regarding their repeal, replacement, or modification, consistent with applicable law. At a minimum, each Regulatory Reform Task Force shall attempt to identify regulations that:

- (i) eliminate jobs, or inhibit job creation;
- (ii) are outdated, unnecessary, or ineffective;
- (iii) impose costs that exceed benefits;
- (iv) create a serious inconsistency or otherwise interfere with regulatory reform initiatives and policies;
- (v) are inconsistent with the requirements of section 515 of the Treasury and General Government Appropriations Act, 2001 (44 U.S.C. 3516 note), or the guidance issued pursuant to that provision, in particular those regulations that rely in whole or in part on data, information, or methods that are not publicly available or that are insufficiently transparent to meet the standard for reproducibility; or
- (vi) derive from or implement Executive Orders or other Presidential directives that have been subsequently rescinded or substantially modified.

(e) In performing the evaluation described in subsection (d) of this section, each Regulatory Reform Task Force shall seek input and other assistance, as permitted by law, from entities significantly affected by Federal regulations, including State, local, and tribal governments, small businesses, consumers, non-governmental organizations, and trade associations.

(f) When implementing the regulatory offsets required by Executive Order 13,771, each agency head should prioritize, to the extent permitted by law, those regulations that the agency's Regulatory Reform Task Force has identified as being outdated, unnecessary, or ineffective pursuant to subsection (d)(ii) of this section.

(g) Within 90 days of the date of this order, and on a schedule determined by the agency head thereafter, each Regulatory Reform Task Force shall provide a report to the agency head detailing the agency's progress toward the following goals:

(i) improving implementation of regulatory reform initiatives and policies pursuant to section 2 of this order; and

(ii) identifying regulations for repeal, replacement, or modification.

Sec. 4. Accountability. Consistent with the policy set forth in section 1 of this order, each agency should measure its progress in performing the tasks outlined in section 3 of this order.

(a) Agencies listed in section 901(b)(1) of title 31, United States Code, shall incorporate in their annual performance plans (required under the Government Performance and Results Act, as amended (see 31 U.S.C. 1115(b))), performance indicators that measure progress toward the two goals listed in section 3(g) of this order. Within 60 days of the date of this order, the Director of the Office of Management and Budget (Director) shall issue guidance regarding the implementation of this subsection. Such guidance may also address how agencies not otherwise covered under this subsection should be held accountable for compliance with this order.

(b) The head of each agency shall consider the progress toward the two goals listed in section 3(g) of this order in assessing the performance of the Regulatory Reform Task Force and, to the extent permitted by law, those individuals responsible for developing and issuing agency regulations.

Sec. 5. Waiver. Upon the request of an agency head, the Director may waive compliance with this order if the Director determines that the agency generally issues very few or no regulations (as defined in section 4 of Executive Order 13,771). The Director may revoke a waiver at any time. The Director shall publish, at least once every 3 months, a list of agencies with current waivers.

Sec. 6. General Provisions. (a) Nothing in this order shall be construed to impair or otherwise affect:

(i) the authority granted by law to an executive department or agency, or the head thereof; or

(ii) the functions of the Director relating to budgetary, administrative, or legislative proposals.

(b) This order shall be implemented consistent with applicable law and subject to the availability of appropriations.

(c) This order is not intended to, and does not, create any right or benefit, substantive or procedural, enforceable at law or in equity by any party against the United States, its departments, agencies, or entities, its officers, employees, or agents, or any other person.

DONALD J. TRUMP

THE WHITE HOUSE,

February 24, 2017

APPENDIX L EXAMPLE OF A SUCCESSFUL IRFA

SECTION NINE

INITIAL REGULATORY FLEXIBILITY ANALYSIS

9.1 THE REGULATORY FLEXIBILITY ACT (RFA) AS AMENDED BY THE SMALL BUSINESS REGULATORY ENFORCEMENT FAIRNESS ACT (SBREFA)

This section considers the effects that the proposed CAFO regulations may have on small livestock and poultry operations as required by the Regulatory Flexibility Act (RFA, 5 U.S.C et seq., Public Law 96-354) as amended by the Small Business Regulatory Enforcement Fairness Act of 1996 (SBREFA). The purpose of the RFA is to establish as a principle of regulation that agencies should tailor regulatory and informational requirements to the size of entities, consistent with the objectives of a particular regulation and applicable statutes. The RFA generally requires an agency to prepare a regulatory flexibility analysis of any rule subject to notice and comment rulemaking requirements under the Administrative Procedure Act or any other statute unless the agency certifies that the rule will not have a “significant impact on a substantial number of small entities.”¹ Small entities include small businesses, small organizations, and governmental jurisdictions.

For this proposed rulemaking, EPA could not conclude that costs are sufficiently low to justify “certification.” Instead, EPA complied with all RFA provisions and conducted outreach to small businesses, convened a Small Business Advocacy Review (SBAR) panel, and prepared an initial regulatory flexibility analysis (IRFA).² This analysis is detailed in this section and represents EPA’s assessment of the impacts of the proposed CAFO regulations on small businesses in the livestock and poultry sectors. Section 9.2 outlines EPA’s initial assessment of small businesses in the sectors affected by the proposed regulations. Section 9.3 presents EPA’s analysis (IRFA) and summarizes the steps taken by EPA to comply with the RFA. Section 9.4 presents the data, methodology, and results of EPA’s analysis of impacts to small businesses for this rulemaking.

9.2 INITIAL ASSESSMENT

EPA guidance on implementing RFA requirements suggests the following must be addressed in an initial assessment (USEPA, 1999i). First, EPA must indicate whether the proposal is a rule subject to notice-and-comment rulemaking requirements. EPA has determined

¹ The preparation of an IRFA for a proposed rule does not legally foreclose certifying no significant impact for the final rule (USEPA, 1999i).

²This analysis or a summary of the analysis must be published in the *Federal Register* at the time of publication of a proposal.

that the proposed CAFO regulations are subject to notice-and-comment rulemaking requirements. Second, EPA should develop a profile of the affected small entities. EPA has developed a profile of the livestock and poultry sectors, which includes all affected operations as well as small businesses. This information is provided in Section 2 and also in Sections 6, 7, and 8 of this EA. Much of the profile information covered in these sections of this report applies to small businesses. Additional information on small businesses in the livestock and poultry sectors is provided in Sections 9.2 and 9.3. Third, EPA's assessment needs to determine whether the rule would affect small entities and whether the rule would have an adverse economic impact on small entities.

Section 9.2.1 reviews the SBA definitions of small entities in the livestock and poultry industry and discusses a rationale for using an alternative definition of small business in one sector. Section 9.2.2 then uses the definitions of small entities laid out in Section 9.2.1 to estimate the number of operations that meet this small business definition. Finally, using the information developed in Sections 9.2.1 and 9.2.2, Section 9.2.3 presents the results of EPA's initial assessment. This assessment provides a first level screen of potential impacts to small CAFO businesses and serves as a signal for additional analysis.

9.2.1 Definition of Small CAFO Businesses

The RFA defines a "small entity" as a small not-for-profit organization, small governmental jurisdiction, or small business. There are no small governmental operations that operate CAFOs. There may be a few not-for-profit organizations that operate CAFOs, but complete information is not available to warrant inclusion of not-for-profit organizations in this analysis. This analysis therefore focuses only on small businesses that are defined or designated as CAFOs. (Section 3 describes the circumstances under which an AFO is defined or designated as a CAFO and is subject to the proposed regulations.)

The RFA requires, with some exception, that EPA define small businesses according to its size standards. SBA sets size standards for defining small businesses by number of employees or amount of revenues for specific industries. These size standards vary by North American Industry Classification System (NAICS) code. CAFOs are listed under NAICS 112, Animal Production.³

SBA's size standards differ from the revenue cutoff generally recognized by USDA, which has defined \$250,000 in gross sales as its cutoff between small and large family farms (USDA, 1998).

³ In September, 2000, SBA updated the basis for its size standard to NAICS codes from Standard Industrial Classification (SIC) codes (USGPO, 2000). By SIC code, these industries are listed under SIC 02, Livestock and Animal Specialties. The actual size standards for each sector, specified as an annual revenue threshold, did not change as a result of this update.

Table 9-1 shows SBA size standards by SIC code for each of the six livestock and poultry sectors, which are expressed in terms of average “annual receipts” (revenue). With one exception, current SBA standards define a “small business” within each of the main livestock and poultry sectors as an operation that generates average revenues ranging from less than \$0.5 million per year (for the hog, dairy, broiler, and turkey sectors) to less than \$1.5 million per year (for the beef feedlot sector), averaged over the most recent three fiscal years (USGPO, 1996; SBA, 1998). The exception is the revenue threshold for a small chicken egg operation (layer sector), which SBA has defined as a business that generates up to \$9 million annually.

Table 9-1. SBA Revenue Size Standards for Small Livestock and Poultry Operations

NAICS Code (SIC Code)	NAICS Industry Description	SBA Size Standard ^{a/}	EPA-Proposed Revenue Cutoff
112112 (0211)	Cattle Feedlots	\$1.5 million	same as SBA
11221 (0213)	Hog and pig farming	\$0.5 million	same as SBA
11212 (0241)	Dairy cattle and milk production	\$0.5 million	same as SBA
11232 (0251)	Broilers and other meat-type chickens	\$0.5 million	same as SBA
11231 (0252)	Chicken egg production	\$9.0 million	\$1.5 million
11233 0253	Turkey production	\$0.5 million	same as SBA

Source: SBA (1998); USGPO (1991a, 1991b and 1996); U.S. Census Bureau (2000).

^{a/}SBA Size Standards by NAICS code (13 CFR Part 121) correspond to classifications under SIC classification.

EPA believes that the definition of small business for the egg laying sector (revenues of \$9 million per year) might not truly characterize a small business in this sector. Therefore, EPA is proposing to use an alternative definition, as allowed by the RFA:

“...an agency, after consultation with the Office of Advocacy of the Small Business Administration and after the opportunity for public comment, establishes one or more definitions of such term which are appropriate to the activities of the agency and publishes such definition(s) in the Federal Register.” 5 U.S.C. §601(3).

EPA’s alternative definition identifies a small business for egg laying operations as any operation that generates up to \$1.5 million in annual revenue (see Table 9-1). Because this definition of a small business is not the definition established under the RFA, EPA is specifically seeking comment on the use of this alternative definition. EPA has also consulted with the SBA Chief Counsel for Advocacy on the use of this alternative definition (USEPA, 1999d). EPA believes this definition better reflects the agricultural community’s sense of what constitutes a small business and more closely aligns with the small business definitions codified by SBA for other animal operations.

There are four broad reasons why EPA believes that its alternative definition of small egg laying operations is more appropriate for the purpose of this rulemaking. These include: (1) EPA's definition is more consistent with size classes used by USDA and industry; (2) EPA's definition reflects the financial and institutional realities of the egg industry; (3) EPA's definition reflects similarities among the sectors of the poultry industry; and (4) EPA's definition captures the relevant segments of the industry (USEPA, 1999d). The four reasons for using the alternative definition of small egg laying operations are summarized below. Additional supporting data and analysis are provided in the rulemaking Record (USEPA, 1999d; USEPA, 2000f).

First, EPA's alternative definition is more consistent with size classes used by USDA (Madison, 1999) and more generally accepted by the regulated community (Gregory, 1999; Staples, 1998). USDA describes size classes reflective of farm level conditions at egg laying operations in terms of the number of houses, where a house has approximately 100,000 to 110,000 hens. Based on USDA's size classes, a small farm has a single house; a medium farm has two to five houses; and a large farm has more than five houses (i.e., more than 500,000 hens). Using USDA data, EPA estimates that a "small" egg operation by USDA standards generates approximately \$1.5 million in annual revenue (USEPA, 1999d and 2000f).⁴

In contrast, a definition of \$9 million in annual revenue fails to reflect farm level conditions based on USDA size classes and matching opinions from the farming community. Such an operation corresponds to an operation with more than six houses (with approximately 600,000 hens). EPA does not believe an operation with six chicken houses should be characterized as "small" for the proposed CAFO regulations. EPA visited one such facility. The facility resides on more than 200 acres and has an annual production of over 180 million eggs. The facility's extensive customer base includes three major supermarket chains and the U.S. military. Its distribution system spans four states. A facility with such a high production level and extensive customer base is not a small business.⁵ EPA's alternative definition would decrease confusion and facilitate communication with the regulated community (both large and small businesses) and with other stakeholders.

Second, EPA's alternative definition better reflects the financial and institutional realities of the egg industry. EPA focuses its regulatory analyses for the proposed CAFO regulations at the animal production level since it is the operator who directly incurs all costs associated with the management and disposal of manure generated from animals that are raised or housed onsite. EPA believes, based on a preliminary review of the background information supporting the SBA definition (USGPO, 1991a and 1991b) that the \$9 million definition applies to entities at a different level in the marketing chain—e.g., to large cooperatives or integrators, rather than farms. The alternative definition would allow EPA to better focus on the needs and concerns of those

⁴EPA estimates are derived using USDA-reported 1997 data: average yield of 255 eggs per layer per year (USDA/NASS, 1998b) and average annual producer price of 66.7 cents per dozen (USDA/NASS, 1998a).

⁵Information on EPA's farm site visits is in the rulemaking record.

businesses that are most likely to experience economic hardship associated with regulatory compliance.

Third, EPA's alternative definition better reflects similarities among the sectors of the poultry industry. EPA's analysis focuses on three sectors: egg laying, broiler, and turkey meat. The SBA definitions differ substantially between the egg laying sector and the other two sectors. As shown in Table 9-1, the small-business definition for layer operations is \$9 million in annual revenue; the small-business definition for both broiler and turkey operations is \$0.5 million. At the farm level, however, there are structural similarities among these three sectors, suggesting that small business definitions should not be so disparate for these operations. The sectors use similar technologies and similar manure management techniques. They have similar costs of production. They have similar industrial organization and marketing arrangements. Measured at the animal production level, the SBA definition of a small broiler or turkey operation is consistent with USDA's definition of a small- or medium-sized operation (based on the number of animals and housing structures, as discussed above).

In fact, prior to 1991, the SBA definition for layer operations was much closer to the definitions for the other two poultry sectors. The earlier SBA definition for layer operations was \$1.0 million. The definition was revised to \$7 million in 1991, and then escalated to \$9 million to account for inflationary changes (USGPO, 1991a and 1991b; Ray, 1999). One of the reasons cited for the 1991 increase was the "limited participation of small egg producers in government procurement" (USGPO, 1991a). For the regulatory flexibility assessment of the proposed CAFO regulations, EPA concludes that the alternative definition is more comparable to the definitions for other livestock sectors and is therefore more appropriate than the existing definition.

Finally, EPA's alternative definition is more appropriate in terms of capturing the relevant segments of the industry. Under EPA's alternative definition, small layer operations would account for roughly 60 percent of annual egg production (USEPA, 2000f). In contrast, under SBA's definition, small operations would account for approximately 90 percent of annual egg production. If EPA were to use SBA's definition, a very large share of total annual egg output would be generated from "small" operations. This would be inconsistent with the analysis of the broiler and turkey sector, where smaller operations represent roughly one-half of each sector's respective annual production. This would further contradict expectations by SBA in terms of the percent of sales attributable to small operations. According to SBA, about 99 percent of all farms in the economy are small and account for approximately 62 percent of sales (Perez, 2000; USEPA, 2000g). This agrees with the realities of the agricultural sector where the majority of farms are small, but account for a relatively small share of overall production. The trend in agriculture towards fewer, larger farms highlights that larger operations—while relatively few in number—represent a greater share of overall output.

EPA also considered another alternative definition for all six animal sectors based on the number of animals raised or housed at the CAFO site (USEPA, 2000e, 1999a, 1999l, and 1999n). Following discussions with representatives from both SBA and OMB during the SBAR Panel

process, EPA decided not to use this alternative definition for each of the animal sectors (USEPA, 2000g). A complete summary of EPA's correspondence with SBA on its proposal and use of an alternative definition is contained in the rulemaking record (see DCN 70509, DCN 70507, DCN 70473, DCN 70472, DCN 70511, DCN 70797, and DCN 93001).

9.2.2 Number of Small Businesses Affected by the Proposed CAFO Regulations

There are three steps for determining the number of small CAFO businesses that may be affected by the proposed regulations. First, EPA identifies small businesses in the relevant livestock and poultry sectors by equating SBA's annual revenue definition with the number of animals at an operation. Second, EPA estimates the total number of small businesses in these sectors using farm size distribution data from USDA. Third, based on the regulatory thresholds being proposed, EPA estimates the number of small businesses that would be subject to the proposed requirements. These steps are described in the following sections.

9.2.2.1 Equating SBA Size Standards with Animal Inventory

In the absence of entity level revenue data, EPA identifies small businesses in the livestock and poultry sectors by equating SBA's annual revenue definitions of "small business" to the number of animals at these operations (step 1). This step produces a threshold based on the number of animals that EPA uses to define small livestock and poultry operations and reflects the average farm inventory (number of animals) that would be expected at an operation with annual revenues that define a small business. This initial conversion is necessary because USDA data by farm size are not available by business revenue. With the exception of egg laying operations, EPA uses SBA's small business definition to equate the revenue threshold with the number of animals raised on site at an equivalent small business in each sector. For egg laying operations, EPA's alternative revenue definition of small business is used.

EPA estimates the number of animals at an operation to match SBA's definitions using SBA's annual revenue size standard (expressed as annual revenue per entity) and USDA-reported farm revenue data that are scaled on a per-animal basis (expressed as annual revenue per inventory animal for an average facility). (This calculation is shown below.) Per-animal financial data are calculated by multiplying the average value of the reported financial data per farm by the total number of farms and then dividing this by the total number of animals. (More information on this calculation is presented in Section 4.2.4.2 of this report.) The average per-animal revenues assumed for this analysis are shown in Table 9-2.

Financial data used by EPA are from the USDA's 1997 ARMS database. These data include farm financial data and corresponding summary information that match the reported average revenue to the total number of farms and the total number of animals in the sample set.

Table 9-2. Number of Small CAFOs That May Be Affected by the Proposed Regulations

Sector	Total Annual (Smillion) Revenue ^{a/} (x)	Revenue per Head ^{b/} (Avg. U.S.) (y)	Number of Animals at Small CAFO Businesses (z=x/y)	Estimated Number of Small AFOs	Two-Tier (500 AU) "Small" CAFO Businesses	Three-Tier "Small" CAFO Businesses
Cattle ^{c/}	\$1.5	\$1,060	1,400	106,450	2,280	2,600
Dairy	\$0.5	\$2,573	200	109,740	50	50
Hogs	\$0.5	\$363	1,400	107,880	300	300
Broilers	\$0.5	\$2	260,000	34,530	9,470	13,410
Egg Layers	\$9.0	\$25	365,000	ND	ND	ND
	\$1.5		61,000	73,710	200	590
Turkeys	\$0.5	\$20	25,000	12,320	0	500
All AFOs ^{d/}	NA	NA	NA	355,650	10,550	14,630

NA=Not Applicable. ND = Not Determined. "AFOs" have confined animals on-site.

^{a/} SBA Size Standards are at 13 CFR Part 121. EPA assumes an alternative definition of \$1.5 million in annual revenues for egg layers.

^{b/} Average revenue per head across all operations for each sector derived from data obtained from USDA's 1997 ARMS data (USDA/ERS, 1999a). See Section 4.

^{c/} Includes fed cattle, veal and heifers.

^{d/} Total adjusts for operations with mixed animal types and includes designated CAFOs (expressed over a 10-year period). See Section 2 of this document for estimates of the total number of AFOs.

These data were obtained with the assistance of staff at USDA's ERS (as described in Section 4.2.3.2).⁶ USDA's data report average national revenue for each sector, combining both livestock and nonlivestock farm revenue (income from crop sales and other farm-related income, including government payments). Use of total farm revenue corresponds to SBA's size standards that are expressed in terms of total annual business revenue (SBA, 1998; USGPO, 2000).

EPA uses the derived per-animal revenues shown in Table 9-2 to equate SBA's size standard (in revenues) with farm size based on the number of animals, as follows:

$$\frac{\text{Average \# Animals}}{\text{Farm}} = \frac{\text{SBA's Small Business Definition (\$ per year per farm)}}{\text{Average Total Revenue per head (\$/animal)}}$$

⁶As noted throughout this report, USDA periodically publishes aggregated data from the ARMS and Census databases and provides customized analyses of the data to members of the public and other government agencies. In providing such analyses, USDA maintains a sufficient level of aggregation to ensure the confidentiality of individual facility data.

The resultant number of animals represents the average animal inventory threshold for a small business. Estimated “small business” thresholds for each sector are shown in Table 9-2.

For the purpose of conducting its IRFA for this rulemaking, and based on the animal inventory thresholds discussed above, EPA is evaluating a “small business” for these sectors as an animal feeding operation that houses or confines less than: 1,400 fed beef cattle; 200 mature dairy cattle; 1,400 market hogs; 260,000 broilers; 61,000 layers; or 25,000 turkeys. Hereafter, all references to small CAFO businesses reflect the SBA definitions of “small” and the alternative definition proposed by EPA for small layer operations, applied on the basis of a calculated number of head.

9.2.2.2 Total Number of Operations that Match SBA Size Standards

Using the threshold sizes identified for small businesses in the livestock and poultry sectors (Table 9-2), EPA matches these thresholds with the number of operations associated with those size thresholds, based on available USDA data, to estimate the total number of small animal confinement operations in these sectors (step 2).

The 1997 Census constitutes the primary data source that EPA uses to match the small business thresholds to the number of operations by size. Other supplemental data used includes other published USDA data and information from industry and the state agriculture extension agencies. In some cases, EPA extrapolated between two size groupings to obtain an estimate of the number of small livestock and poultry operations. Additional information is also used to subdivide sector level data into subsectors. For example, the number of hog operations that are farrow-finish versus grow-finish are distinguished according to market share information (USDA/APHIS, 1995b). Information that differentiates the number of egg laying operations according to manure management system (wet versus dry) are approximated based on conversations with State Extension personnel for selected states, as described in the *Development Document* (USEPA, 2000a). The number of breeder and nursery pig operations and veal and heifer operations are approximated based on information obtained from state extension personnel and EPA’s farm site visits (USEPA, 2000a).

For many of the animal sectors, it is not possible to estimate from available U.S. farm data what proportion of total livestock and poultry operations have feedlots and what proportion are grazing operations only. For the beef and hog sectors, the USDA has limited data on the number of operations that are feedlot operations only (USDA/APHIS, 1995b; USDA/NASS, 1999a and 1999b). For analytical purposes, EPA has assumed that all dairy and poultry operations potentially are confinement operations. More information on the farm size distribution data that EPA uses to match the size thresholds to the number of poultry and livestock operations is documented in the *Development Document* (USEPA, 2000a).

Table 9-2 shows EPA's estimates of the total number of small livestock and poultry operations using this approach. As shown, an estimated 355,650 animal confinement operations meet SBA's small business definition. This is 95 percent of the estimated total number of animal feeding operations (375,700 operations).

EPA recognizes that this approach may not accurately portray actual small businesses in all cases across all sectors. On the one hand, the resulting small business estimate would suggest that a 10-house broiler operation with 260,000 birds would be a small business. Information from industry sources, however, suggest that a two-house broiler operation with roughly 50,000 birds is small (Madison, 1999; USEPA, 2000e). Therefore, it is likely that some medium- and large-size broiler operations are being considered small businesses (USEPA, 2000g).

On the other hand, it is possible that the resulting small business estimate may have failed to identify some small businesses as "small" in the other sectors. For example, EPA's approach identifies as a "small business" hog operations with less than 1,400 pigs and turkey operations with less than 25,000 turkeys, which account for less than 94 percent of all operations and less than 30 percent of sales in these sectors. These proportions are below SBA's presumed coverage rates that define as small about 99 percent of all operations that account for approximately 62 percent of sales (Perez, 2000). Therefore, it is likely that there are additional small hog and turkey businesses that are not captured under the revised methodology (USEPA, 2000g).

9.2.2.3 Total Number of Small CAFOs Subject to the Proposed Regulations

Based on the regulatory thresholds for each co-proposed alternative, EPA estimates the number of small businesses that will be subject to the proposed requirements (step 3).⁷ The 1997 Census constitutes the primary data source that EPA uses to match the small business thresholds (e.g., a small dairy operation has less than 200 milk cows) to the number of facilities that match that size group (e.g., the number of dairies with less than 200 cows, as reported by USDA). Other supplemental data used include other published USDA data and information from industry and the state extension agencies.

Table 9-2 shows the estimated total number of livestock and poultry operations that meet the SBA definition of a "small business" in each of the livestock and poultry sectors. Not all of small confinement operations would be subject to the proposed CAFO regulations, however. EPA's proposed regulations only apply to those operations that meet the regulatory definition of a CAFO or those that have been designated as a CAFO by the NPDES permitting authority due to risks posed to water quality and public health, as discussed in Section 3. The proposed changes *define* as a CAFO those operations that confine more than 300 or 500 AU (depending on co-

⁷In this section, EPA discusses numbers of affected CAFOs and impacts under the two-tier structure at 500 AU threshold (Scenario 4a) and three-tier structure (Scenario 3) only. "Two-tier structure" in this section refers to the 500 AU threshold, except where otherwise noted.

proposed scenario). The proposed requirements may also apply to an operation that confines fewer than 300 or 500 AU if it is *designated* as a CAFO by the NPDES permitting authority on a case-by-case basis, based on an on-site inspection.

Of the estimated 355,650 animal confinement operations that meet SBA's small business definition, EPA estimates that 10,550 operations that will be subject to the proposed requirements that are small businesses under the two-tier structure. Under the three-tier structure, an estimated 14,630 affected operations are small businesses. These estimates include expected designated facilities. The difference in the number of affected small businesses is among poultry producers, particularly broiler operations. See Table 9-2.

Table 9-3 presents the estimated number of livestock and poultry operations that may be subject to the proposed requirements under each co-proposed scenario that are also small businesses ("small CAFO businesses") by facility size category. The number of small CAFO businesses are shown as follows: (1) operations defined as CAFOs with more than 1,000 AU, (2) operations defined as CAFOs with between 300 to 1,000 AU or 500 to 1,000 AU, depending on scenario, and (3) operations that may be designated as CAFOs with fewer than 300 or 500 AU that may be designated (varies by co-proposed alternative). The number of small CAFO businesses in each of the three size categories is developed using the same data approach used to identify the total number of small operations, discussed in Section 9.2.2.2.

Based on estimates shown in Table 9-3, EPA estimates that there are 10,220 operations with more than 500 AU that may be defined as CAFOs that also meet the "small business" definition, under the two-tier structure. Under the three-tier structure, there are 14,530 operations with more than 300 AU that may be defined as CAFOs that are small businesses that meet the proposed risk-based conditions (described briefly in Section 3; more detail is provided in Section VII of the preamble). By broad facility size group, EPA estimates that about 4,000 operations have more than 1,000 AU, adjusting for operations with more than a single animal type. EPA estimates that about 6,000 operations have between 500 and 1,000 AU (two-tier structure) and about 10,000 operations have between 300 and 1,000 AU (three-tier structure), accounting for mixed operations. EPA's analysis assumes that all small businesses with 300 to 1,000 AU under the three-tier structure obtain a NPDES permit and that none certify out of the program.

Among operations that are defined as CAFOs, depending on co-proposed scenario, most small CAFO businesses are in the broiler and cattle sectors. As defined for this analysis, EPA expects that there are no small CAFO businesses in the dairy sector with more than 300 AU (see Section 9.2.2.1) and that small dairies will be subject to the regulations only if they are designated as a CAFO by the Permitting Authority. Also, as defined for this analysis, there are no small

Table 9-3. Total Number of Small CAFO Businesses Subject to Regulation

Sector	All "Small" AFOs		Two-Tier Structure			Three-Tier Structure		
	All	>1,000 AU	500-1,000 AU	<500 AU	Total	300-1,000 AU	<300 AU	Total
Fed Cattle	104,350	350	1,000	40	1,390	1,140	0	1,490
Veal	850	10	80	0	90	130	0	140
Heifers	1,250	300	500	0	800	680	0	980
Dairy	109,740	0	0	50	50	0	50	50
Hogs	107,800	0	100	200	300	250	50	300
Broilers	34,530	3,610	5,840	20	9,470	9,800	0	13,410
Layers	73,710	0	180	20	200	600	0	590
Turkeys	12,320	0	0	0	0	500	0	500
Sum Total	444,560	4,270	7,700	330	12,300	13,080	100	17,300
Total	355,565	4,060	6,160	330	10,550	10,470	100	14,630

Sources: Values presented in the table are EPA estimates, derived from published USDA data, including 1997 Census of Agriculture (USDA/NASS, 1999a) supplemented with other data, as described in the *Development Document* (USEPA, 2000a). All numbers are rounded to the nearest ten.

"Total" eliminates double counting of operations with mixed animal types. Based on survey level Census data, operations with mixed animal types account for roughly 25 percent of operations less than 1,000 AU; few operations with more than 1,000 AU have more than a single animal type.

grow-finish hog operations that may be defined as CAFO under either co-proposed scenario; also, there are no small CAFO businesses in the turkey sector under the two-tier structure (Table 9-3).

The majority (about 90 percent) of small confinement operations have fewer than 300 AU (Table 9-3). EPA's total estimate of small affected CAFOs includes an additional 330 small operations with fewer than 500 AU that may be designated as CAFOs under the two-tier structure over a 10-year period (consistent with the 10-year time frame used for EPA's financial model). As these facilities are designated, EPA did not adjust this total to reflect possible mixed animal operations. All of these operations are small businesses. Under the two-tier structure, designated operations are expected to consist of beef, dairy, hog, egg layer and broiler confinement operations that are located in more traditional farming regions and are determined to be significant

contributors of pollution.⁸ Under the three-tier structure, EPA expects that 100 dairy and hog operations will be designated as CAFO and, therefore, subject to the proposed regulations.

These estimates are based on farm data for 1997. Due to continued consolidation and facility closure since 1997, EPA's estimates may overstate the actual number of small businesses in these sectors. In addition, ongoing trends are causing some existing small- and medium-size operations to expand their inventories to achieve scale economies. Some of the CAFOs considered here as small businesses may no longer be counted as small businesses because they now have higher revenues.

9.2.3 Results of the Initial Assessment

Early on in the development of this rulemaking, EPA conducted a preliminary assessment of the potential impacts to small CAFO businesses based on the results of a costs-to-sales test for operations with more than 500 AU. This screening test indicated the need for additional analysis to characterize the nature and extent of impacts on small entities. This assessment is conducted for those CAFOs that are small businesses, as determined by EPA.

Table 9-4 presents the results of this screening test and indicates that about 80 percent (about 9,700) of the estimated number of small businesses with more than 500 AU that would be directly subject to the rule as CAFOs (two-tier) may incur costs in excess of three percent of sales. Compared to the total number of all small animal confinement facilities estimated by EPA (355,650 facilities), EPA estimates that operations that may incur costs in excess of three percent of sales comprise less than two percent of all small businesses in these sectors. (The cost and revenue data EPA uses for this assessment are presented in Section 9.4; more detailed information on these data is provided in Section 4 of this report.)

Based on the results of this initial assessment, EPA projected that the Agency would likely not certify that the proposal, if promulgated, would not impose a significant economic impact on a substantial number of entities. Therefore, EPA convened a Small Business Advocacy Review Panel and prepared an Initial Regulatory Flexibility Analysis (IRFA) pursuant to Sections 609(b) and 603 of the RFA, respectively, and prepared an economic analysis (see Sections 9.3 and 9.4).

⁸EPA expects that USDA will continue to provide voluntary assistance to those additional operations that are now defined as CAFOs under the current permitting requirements (300 AU to 500 AU) that are not covered by proposed CAFO revisions under the two-tier structure.

Table 9-4. EPA's Preliminary Assessment of Small Business Impacts using a Sales Test

Sector	Small AFOs	Small CAFOs (>500 AU)	Costs Exceed 3% of Revenues		
			#Small CAFOs	%Small CAFOS	%Small AFOs
Fed Cattle	104,350	1,350	80	6%	1%
Veal	850	90	10	1%	1%
Heifers	1,250	800	20	3%	2%
Dairy	109,736	0	0	0%	0%
Hog-FF	57,800	100	20	20%	1%
Hog-GF	50,000	0	0	0%	0%
Broilers	34,530	9,450	9,450	100%	28%
Layers-Wet	9,010	20	0	0%	0%
Layers-Dry	64,700	160	0	0%	0%
Turkeys	12,320	0	0	0%	0%
Sum Total	444,560	11,970	9,580	80%	2%

Source: USEPA. Total does not adjust for operations with mixed animal types, for comparison purposes. Includes CAFOs with more than 500 AU. Excludes designated operations. Sales test results are shown for the proposed BAT Option and NPDES Scenario 4a (described in Section 3).

9.3 EPA COMPLIANCE WITH RFA REQUIREMENTS

9.3.1 Outreach and Small Business Advocacy Review

As required by Section 609(b) of the RFA, as amended by SBREFA, EPA convened a Small Business Advocacy Review (SBAR) Panel for the proposed rule. The Panel was convened in December, 1999. Panel participants included representatives from EPA, the Office of Information and Regulatory Affairs within the Office of Management and Budget (OMB), and the Office of Advocacy of the Small Business Administration (SBA). "Small Entity Representatives" (SERs), who advised the Panel, included small livestock and poultry producers as well as representatives of the major commodity and agricultural trade associations. Throughout the development of these regulations, EPA conducted outreach to small businesses in the livestock and poultry sectors. EPA also consulted with SBA on the use of an alternative definition of small business for the egg laying sector.

Consistent with the RFA/SBREFA requirements, the Panel evaluated the assembled materials and small entity comments on issues related to the elements of the IRFA. The Panel's

activities and recommendations are summarized in the *Final Report of the Small Business Advocacy Review Panel on EPA's Planned Proposed Rule on National Pollutant Discharge Elimination System (NPDES) and Effluent Limitations Guideline (ELG) Regulations for Concentrated Animal Feeding Operations* (USEPA, 2000g), or "Panel Report." This document is included in the public record (DCN 93001). Section XII.G of the preamble provides a summary of the Panel's activities and recommendations and describes the subsequent action taken by the Agency. Section XII of the preamble also details various outreach activities conducted by EPA that include outreach to small businesses in these sectors.

9.3.2 EPA's Initial Regulatory Flexibility Analysis

As required by Section 603 of the RFA, as amended by SBREFA, EPA has conducted a initial regulatory flexibility analysis. The IRFA must include a discussion of the reason the agency is considering the proposed rule, as well as the objectives and legal basis for the proposal. It must also include a description and estimate of the number of small businesses that will be affected. It must describe the reporting, recordkeeping, and other compliance requirements of the proposed rule and must identify any federal rules that may duplicate, overlap, or conflict with the proposed rule. Finally, the IRFA must describe any significant regulatory alternatives to the rule that would accomplish the stated objectives of the applicable statutes and which minimize impacts to small businesses. Sections 9.3.2.1 through 9.3.2.6 below address each of these requirements of the IRFA that EPA has prepared to support the proposed CAFO regulations.

Section 607 of the RFA further notes that to comply with the IRFA requirements, the Agency must "provide either a quantifiable or numerical description of the effects of a proposed rule or alternatives to the proposed rule, or more general descriptive statements if quantification is not practicable or reliable." For this rulemaking, EPA has prepared an economic analysis of the impacts to small CAFO businesses. This analysis is provided in Section 9.4. Based on the results of this analysis, EPA has determined that the proposed regulations will result in financial stress to some affected small businesses, but not a substantial number of operations relative to the total number of affected small businesses in these sectors. Additional information and the detailed results of this analysis are presented in Section 9.4.2.

9.3.2.1 Reason EPA is Considering the Proposed Rule

Despite more than twenty years of regulation, there are persistent reports of discharge and runoff of manure and manure nutrients from livestock and poultry operations. The proposed revisions to the existing ELG and NPDES regulations for CAFOs are expected to mitigate future water quality impairment and the associated human health and ecological risks by reducing pollutant discharges from the animal production industry.

EPA's proposed revisions also address the changes that have occurred in the animal production industries in the United States since the development of the existing regulations. The continued trend toward fewer but larger operations, coupled with greater emphasis on more intensive production methods and specialization, is concentrating more manure nutrients and other animal waste constituents within some geographic areas. This trend has coincided with increased reports of large-scale discharges from these facilities and continued runoff that is contributing to the significant increase in nutrients and resulting impairment of many U.S. waterways.

EPA's proposed revisions of the existing regulations will make the regulations more effective in protecting or restoring water quality. The revisions will also make the regulations easier to understand and better clarify the conditions under which an AFO is a CAFO and, therefore, subject to the regulatory requirements.

Additional information on why EPA is revising the existing regulations is provided in Section IV of the preamble.

9.3.2.2 Objectives and Legal Basis for the Proposed Rule

A detailed discussion of the objectives and legal basis for the proposed CAFO regulations is presented in Sections I and III of the preamble.

9.3.2.3 Description and Estimate of Number of Small Entities Affected

As presented in Section 2, EPA estimates that there are about 375,700 livestock and poultry operations nationwide of which 355,650 (95 percent) are small (Table 9-2). Of these, the proposed CAFO regulations are expected to affect—and impose compliance costs on—approximately 10,550 operations or 14,630 operations (Table 9-3), depending on co-proposed scenario. Most (about 80 percent) of the estimated number of small CAFO businesses are in the poultry sectors, with the majority in the broiler sector. The cattle sector accounts for another 15 to 18 percent of small CAFO businesses, depending on tier structure. The remaining number of affected small CAFO businesses are in the hog and dairy sectors.

Tables 9-5 and 9-6 show the numbers of affected small businesses by EPA's model CAFO designation, which characterizes each of the small businesses by sector, size, and key production region. (Values shown in the tables do not adjust for operations with more than a single animal type.) These estimated CAFO numbers by model type are used to evaluate small business impacts, presented in Section 9.4 of this report.

Table 9-5. Numbers of Small CAFO Businesses by Sector, Size, and Region, Two-Tier Structure

Sector	Region	CAFOs <300AU	CAFOs "Medium 1"	CAFOs "Medium 2"	CAFOs "Large 1"	CAFOs "Large 2"
Fed Cattle	CE			160	70	
	MW		40	840	280	
Veal	MW			80	10	
Heifers	MW			500	300	
Dairy	MW	50				
	PA					
Hog: FF	MA					
	MW	50	150	100		
Hog: GF	MA					
	MW					
Layer: Wet	SO			40		
Layer: Dry	MW			60		
	SO			100		
Broiler	MA		740	1,190	980	70
	SO		1,280	2,650	2,300	260
Turkey	MA					
	MW					
Total		100	2,210	5,720	3,940	330

Source: USEPA. Size and region breakouts are based on 1997 Census data provided in the *Development Document* (USEPA, 2000a). Facility size and region definitions for model CAFOs are provided in Section 4, Table 4-1. Rounded to nearest ten. Numbers do not adjust for mixed animal types and include expected designated CAFOs (<500 AU under two-tier and <300 under three-tier structure) are included in the counts and are shown over a 10-year period. Shaded cells indicate that there are no small CAFO businesses that will be affected by the regulations that meet the SBA definition of a small business.

Table 9-6. Numbers of Small CAFO Businesses by Sector, Size, and Region, Three-Tier Structure

Sector	Region	CAFOs <300AU	CAFOs "Medium 1"	CAFOs "Medium 2"	CAFOs "Large 1"	CAFOs "Large 2"
Fed Cattle	CE		20	160	70	
	MW		120	840	280	
Veal	MW		50	80	10	
Heifers	MW		180	500	300	
Dairy	MW	50				
	PA					
Hog: FF	MA					
	MW	50	150	100		
Hog: GF	MA					
	MW					
Layer: Wet	SO		50	20		
Layer: Dry	MW		130	60		
	SO		230	100		
Broiler	MA		3210	1190	980	70
	SO		2750	2650	2,300	260
Turkey	MA		320			
	MW		180			
Total		100	7,390	5,700	3,940	330

Source: USEPA. See Table 9-5.

9.3.2.4 Description of the Proposed Reporting, Recordkeeping, and Other Requirements

The proposed CAFO regulations contain recordkeeping and reporting requirements. Costs associated with information collection include the recording of animal inventories, manure generation, findings from visual inspections of feedlot areas and fields, lagoon emptying, and other activities on a routine basis. Recordkeeping requirements also include collecting information on field application of manure and other nutrients (including amount, rate, method, incorporation, and dates), manure and soil analysis compilation, crop yield goals and harvested yields, crop rotations, tillage practices, rainfall and irrigation, and lime applications. Other

requirements include manure spreader calibration worksheets, manure application worksheets, maintenance logs, and soil and manure test results.

EPA has estimated the burden and costs associated with information collection imposed on CAFOs and states as a result of the proposed CAFO regulations. This analysis is provided in the Information Collection Request (ICR) document prepared by EPA (USEPA, 2000i). For the purpose of this analysis, “burden” means the total time, effort, or financial resources expended by persons to generate, maintain, retain, or disclose or provide information to or for a federal agency. This includes the time needed to review instructions; develop, acquire, install, and utilize technology and systems for the purposes of collecting, validating, and verifying information, processing and maintaining information, and disclosing and providing information; adjust existing procedures to comply with any previously applicable instructions and requirements; train personnel to be able to respond to a collection of information request; search data sources; complete and review the collection of information; and transmit or otherwise disclose the information.

EPA’s labor burden estimates for CAFO and state respondents are the hours of activity required to comply with changes to the NPDES CAFO program. For each activity, EPA estimates the burden in terms of the expected effort necessary to carry out these activities under normal conditions and reasonable labor efficiency. These activities and estimated burden and cost levels are described in more detail in the ICR (USEPA, 2000i). The ICR also contains a summary of wage rate information from USDA, state agricultural extension agencies, and the Bureau of Labor Statistics, compiled by EPA for the purpose of this analysis. Additional information on the ICR is provided in Section XIII.F of the preamble to this rulemaking. A summary of the analysis of impacts to CAFO operators is provided below. Additional information on the estimated burden and costs to states is provided in the ICR.

EPA identifies five burden activities to CAFO operators, including start-up activities, permit application, permit nutrient plan development, best available technology requirements, and ground water monitoring for new facilities. Start-up activities are steps that a CAFO owner or operator must take in preparation to comply with the information collection requirements of the proposed rule. Owners or operators that are potentially affected by the rule will need to familiarize themselves with the changes to the NPDES CAFO program to determine that they will need to apply for a permit (or certify out of the program, under three-tier structure only), develop a PNP, and implement the other BAT requirements. PNPs must be reviewed annually and rewritten every five years. Permit application activities involve completing and submitting either an NOI under a general permit or an application for an individual permit. These activities will be conducted once every five years.

PNP development and implementation will require owners or operators of CAFOs to apply for a permit and notify their permitting authority when the PNP has been developed or modified. This notice must include the number of animals covered by the plan, the number of acres receiving waste, the nutrient content of the manure, the application schedule and rate, and

the quantity that will be transferred off site. As part of their recordkeeping responsibilities, CAFO operators will be required to keep the plan on site for inspections and make it available to the permitting authority on request.

To meet the proposed BAT requirements, CAFO owners or operators will perform various activities which will need to be recorded, such as visual inspections of the feedlot facilities, testing or calibration of manure application equipment, collection of soil samples, recording of volume of manure and process wastewater produced as well as off-site transfer, and employee training. Existing beef and dairy sources as well as all NSPS have requirements will involve documentation of whether ground water is hydrologically linked to surface water at the CAFO site and, if it is, records of monitoring of ground water quality. Monitoring records must be maintained to demonstrate that no discharge has occurred.

In addition to recordkeeping costs, EPA estimates the capital and operation and maintenance (O&M) costs associated with these burden activities. A CAFO will incur capital costs when it purchases equipment or builds structures that are needed for compliance with the rule's reporting and recordkeeping requirements that the facility would not use otherwise. Consistent with the overall cost analysis for the proposed rule, capital costs are annualized assuming a 10-year amortization period and a 7 percent interest rate. Capital costs for the proposed rule include purchasing a soil auger to collect soil samples and a manure sampler. CAFOs applying manure on site (assumed to be 100 percent, although land application does not occur at 100 percent of CAFOs) will need to obtain a scale to calibrate the spreader. Some facilities will also need to install depth markers in their lagoons, and certain sources with ground water linked to surface water will need to install monitoring wells. EPA's estimates also include the one time cost for the nutrient management course in this cost category. A facility incurs O&M costs when it regularly uses services, materials, or supplies needed to comply with the rule's reporting and recordkeeping requirements that the facility will not use otherwise. Any cost for the operation and upkeep of capital equipment is considered an O&M cost. O&M costs may also be incurred on a non-annual basis, such as every three years. O&M costs include laboratory analysis of soil, manure, and ground water samples, training of person responsible for manure application, and maintenance of ground water monitoring wells.

EPA estimates that the public burden for this information collection request will require 1.2 to 1.6 million labor hours for all CAFO respondents to comply with the proposed regulations (USEPA, 2000i). Information collection at a CAFO is associated with permit application, PNP development, inspection and sampling, and ground water assessment. These estimates include the time required to review instructions, search existing data sources, gather and maintain all necessary data, and complete and review the information collection. EPA estimates total costs to regulated CAFOs associated with reporting and recordkeeping requirements under the proposed CAFO regulations at \$27 million annually (1999 dollars), under the two-tier structure. For the three-tier structure, EPA estimates costs to regulated CAFOs at \$35 million annually (USEPA, 2000i). This estimate excludes NPDES burden for CAFOs covered by other ICR estimates, as well as NPDES burden for co-permittees and off-site manure recipients.

Under the two-tier structure, EPA estimates that there will be approximately 7,300 CAFO respondents and an average of 80,700 CAFO responses per year. Under the three-tier structure, EPA estimates that there will be approximately 9,600 CAFO respondents and an average of 107,800 CAFO responses. Thus, the average burden per CAFO respondent is 163 to 166 hours and the average burden per CAFO response is 14 to 15 hours. For this analysis, EPA assumes that the administrative burden assumptions are generally the same regardless of CAFO size. Only soil sampling and PNP development burdens would differ by CAFO size. Costs are assessed using a weighted average acreage for all affected CAFOs and do not contain a breakdown for CAFOs with more than or less than 1,000 AU. This estimate likely overstates the time requirements at small CAFO businesses, since it is an average over all operations both large and small.

More detailed information on the burden and associated costs for each of the activities described above is provided in the ICR (USEPA, 2000i).

9.3.2.5 Identification of Relevant Federal Rules that May Duplicate, Overlap, or Conflict with the Proposed Regulations

For this analysis, EPA assumes that all CAFOs are already in compliance with existing federal and state regulations affecting animal production facilities. The Small Business Advocacy Review Panel did not identify any federal rules that duplicate or interfere with the requirements of the proposed rule (USEPA, 2000g).

9.3.2.6 Significant Regulatory Alternatives

EPA proposes to focus the regulatory revisions in this proposal on the largest operations, which present the greatest risk of causing environmental harm, and in so doing, has minimized the effects of the proposed regulations on small livestock and poultry operations. First, EPA is proposing to establish a two-tier structure with a 500 AU threshold. Unlike the current regulations, under which some operations with 300 to 500 AU are defined as CAFOs, operations of this size under the revised regulations would be CAFOs only by designation. Second, EPA is proposing to raise the size standard for defining egg laying operations as CAFOs. Third, EPA is proposing to eliminate the “mixed” animal calculation for operations with more than a single animal type for determining which AFOs are CAFOs.

Under the two-tier structure, EPA is proposing to revise the threshold for being defined as a CAFO down to 500 AU and eliminate the “middle category” for operations with between 300 and 1000 AU. This proposal would provide relief to small businesses by removing from the CAFO definition operations with between 300 AU to 500 AU that under the current rules are defined as CAFOs. EPA estimates that under the co-proposed alternatives, between 64 percent (two-tier) and 72 percent (three-tier) of all CAFO manure would be covered by the regulation.

(See Section 2 of this report.) Under the two-tier structure, the inclusion of all operations with more than 300 AU instead of operations with more than 500 AU, the CAFO definition would result in 13,800 additional operations being regulated, along with an additional 8 percent of all manure. An estimated 80 percent of these additional 13,800 CAFOs are small businesses (about 10,870 CAFOs). EPA estimates that by not extending the regulatory definition to operations with between 300 and 500 AU, these 10,870 small businesses will not be defined as CAFOs and will therefore not be subject to the proposed regulations. EPA estimates the additional costs of extending the regulations to these small CAFO businesses at almost \$150 million across all sectors. The difference in costs between the proposed BAT Option/Scenario and the proposed BAT Option and Scenario 4b combination may be approximated by comparing the estimated costs for these regulatory options, which are shown in Section 5.

Also, under the two-tier structure, EPA is proposing to raise the size standard for defining egg laying operations as CAFOs. This alternative would remove from the CAFO definition small egg laying operations with between 30,000 and 50,000 hens that under the current rules are defined as CAFOs, if they utilize a liquid manure management system. (The current regulations affects egg laying operations with more than 30,000 birds that use wet manure management systems only. Layer operations with dry manure systems are not covered by the regulations. EPA is proposing to regulate all layer operations of a certain size, regardless of the type of manure management systems used, as described in Section 3.) To provide relief to smaller operations, EPA is proposing to raise the size standard to apply to operations with more than 50,000 birds on-site. A higher size standard for egg laying operations is intended to avoid placing too much burden on small egg laying operations. These operations are virtually all small businesses (see Table 9-2). Most of these operations are concentrated in the Southern production regions. Data are not available to determine the number of egg laying operations with 30,000 to 50,000 layers. Therefore, EPA did not estimate the cost savings of raising the size standards for egg operations.

In addition, under both co-proposed alternatives, EPA is proposing to revise the threshold for being defined as a CAFO by eliminating the requirements for “mixed” operations (i.e., operations with more than a single animal type). Under the existing permit regulation, if a facility confines more than one animal type, each animal type is assigned a multiplication factor that is used to calculate the total number of animal units at the facility. Only poultry is excluded from this mixed animal type calculation under existing regulations. EPA is proposing to exclude mixed operations with more than a single animal type. The Agency determined that the inclusion of these operations would disproportionately burden small businesses while resulting in little additional environmental benefit. Since most mixed operations tend to be smaller in size, this exclusion represents important accommodations for small businesses. EPA expects that there are few large operations that confine more than a single animal type. If certain of these smaller operations are determined to be discharging to waters of the U.S., States can later designate them as CAFOs and subject them to the regulations. EPA’s decision not to include operations with more than a single animal type is also expected to simplify compliance and be more

administratively efficient, since the mixed operation multipliers were confusing to the regulated community and to enforcement personnel, and did not cover all animal types.

Overall, EPA's decision to mitigate the effects on small CAFO businesses through these scope considerations is intended to favor smaller—usually more traditional and often more sustainable—farm production systems where operators grow both livestock and crops and land apply manure nutrients. This is consistent with EPA's objectives under the USDA-EPA Unified National Strategy for Animal Feeding Operations, which targets only the largest operations since these pose the greatest *potential* risk to water quality and public health given the sheer volume of manure generated at these operations (USDA and USEPA, 1999). Larger operations that handle larger herds or flocks often do not have an adequate land base for manure disposal through land application. As a result, large facilities need to store significant volumes of manure and wastewater that have the potential, if not properly handled, to cause significant water quality impacts. In comparison, smaller operations manage fewer animals and tend to concentrate fewer manure nutrients at a single location. Smaller operations tend to be less specialized and are more diversified, engaging in both animal and crop production. These operations often have sufficient cropland and fertilizer needs to land apply manure nutrients generated at a livestock or poultry business.

9.4 EPA'S ANALYSIS OF SMALL BUSINESS IMPACTS

This section discusses the data and methodology EPA uses to assess economic impacts on small CAFO businesses (Section 9.4.1) and presents the results of this analysis (Section 9.4.2). This economic analysis supports the IRFA (Section 9.3) by quantifying the effects of the proposed CAFO regulations.

9.4.1 Data and Methodology

To examine the economic impacts of the proposed regulations on small CAFO businesses, EPA uses the same representative farm approach that is used to analyze impacts to all CAFOs (regardless of size), as described in Section 4 this EA. This approach evaluates impacts to select model CAFOs and extrapolates these results to the number of operations identified by each representative model, thus aggregating costs nationally across all sectors. Inputs for this analysis include the number of CAFOs represented by each model (see Section 9.3.3) and, for each model CAFO, the costs of the proposed regulations and selected financial characteristics (see Section 4).

EPA's analysis evaluates the economic achievability of the proposed regulatory options at small CAFO businesses based on changes in representative financial conditions across three criteria. These criteria are: a comparison of incremental costs to total revenue (sales test), projected post-compliance cash flow over a 10-year period, and an assessment of an operation's debt-to-asset ratio under a post-compliance scenario.

EPA determines economic impacts to small businesses by applying the proposed economic achievability criteria described in Section 4.2.5, which are used to divide the impacts of the proposed CAFO regulations into three categories (see Table 4-11). Accordingly, if an average model facility is determined to incur economic impacts under the proposed CAFO regulations that are regarded as “Affordable” or “Moderate,” then the results are considered to indicate economic achievability. “Moderate” impacts are not associated with operational change at the CAFO and are considered by EPA to indicate economic achievability. If an average operation is determined to incur “Stress,” this result is considered to potentially indicate that the proposed regulations might not be economically achievable, subject to other considerations. “Affordable” and “Moderate” impacts are associated with positive post-compliance cash flow over a 10-year period and a debt-to-asset ratio not exceeding 40 percent, in conjunction with a sales test result that shows that compliance costs are less than 5 percent of sales (“Affordable”) or between 5 and 10 percent of sales (“Moderate”). “Stress” impacts are associated with negative cash flow or a post-compliance debt-to-asset ratio exceeding 40 percent, or sales test results that show costs equal to or exceeding 10 percent of sales. More detail on this classification scheme, along with a discussion of the basis for EPA’s determination of these criteria for this analysis, is provided in Section 4.2.5.

Table 9-7 shows EPA’s estimated compliance costs for selected model CAFOs under the proposed BAT Option. Costs are not presented separately by facility model for each co-proposed scenario, since the only difference in costs between the two scenarios are associated with the difference in the numbers of regulated CAFOs. All costs shown are expressed on a per-animal basis and are differentiated by facility size, producing region, facility types, and other factors. Costs are reported in ranges across three types of land availability for manure application assumed for this analysis. These land availability types include: Category 1 farms, which have sufficient cropland for all on-farm nutrients generated; Category 2 farms, which have insufficient cropland; and Category 3 farms, which have no cropland. Ranges also reflect Option 3 and 3A costs.⁹ Section 4.2.1 provides additional information on EPA’s cost models. Unit costs shown in Table 9-7 are aggregated by the average number of animals assumed for each model CAFO to derive total entity compliance costs used in this analysis. Information on EPA’s model CAFOs used for this analysis is provided in Section 4.2 of this report.

⁹Option 3 assesses average costs to operations if there is no direct hydrologic connection to surface waters; Option 3A reflects costs to operations where there is a determined groundwater hydrologic connection (assumed at 24 percent of all affected operations).

Table 9-7. Estimated Per-Head Facility Costs (BAT Option/Co-Proposed Scenarios) for Model CAFOs

Sector	Region	Model CAFOs <300AU	Model CAFOs "Medium 1"	Model CAFOs "Medium 2"	Model CAFOs "Large 1"	Model CAFOs "Large 2"
			300 - 1,000 AU		>1,000 AU	
		(incremental compliance costs \$ per animal)				
Fed Cattle	CE		\$10.81-\$80.32	\$7.21-\$61.98	\$3.37-\$38.59	
	MW		\$15.87-\$67.24	\$11.31-\$50.56	\$7.12-\$34.65	
Veal	MW		\$2.65-\$7.78	\$2.54-\$4.75	\$2.50-\$4.75	
Heifers	MW		\$14.13-\$55.50	\$9.37-\$39.57	\$5.04-\$20.16	
	PA		\$10.63-\$60.86	\$6.86-\$43.96	\$3.51-\$27.14	
Dairy	MW	\$60.39-\$222.08				
	PA					
Hog: FF	MA					
	MW	\$5.80	\$6.03-\$7.45	\$4.35-\$5.65		
Hog: GF	MA					
	MW					
Layer: Wet	SO		\$0.83	\$0.39-\$0.60		
Layer: Dry	MW		\$0.02-\$0.27	\$0.02-\$0.23		
	SO		\$0.02-\$0.18	\$0.02-\$0.15		
Broiler	MA	\$0.07-\$0.13	\$0.07-\$0.13	\$0.07-\$0.12	\$0.07-\$0.12	\$0.05-\$0.10
	SO	\$0.07-\$0.15	\$0.07-\$0.15	\$0.07-\$0.13	\$0.06-\$0.13	\$0.05-\$0.11
Turkey	MA		\$0.07-\$0.71			
	MW		\$0.12-\$0.83			

Source: USEPA. Annualized costs are shown in Appendix A; actual costs are in the *Development Document* (USEPA, 2000a). Facility size and region definitions for model CAFOs are provided in Section 4, Table 4-1. Large operations roughly correspond to CAFOs with >1,000 AU and Medium operations correspond to CAFOs with 300-1,000 AU. Shaded cells indicate that there are no CAFOs that will be affected by the proposed regulations and that meet the SBA definition of a small business.

EPA also developed costs to confinement operations with less than 300 or 500 AU that may be designated as CAFOs by scaling the estimated compliance costs for the available “medium” and “large” CAFO models. (See Tables 9-5 and 9-6 for expected designated facilities under each co-proposed alternative.) The resulting costs—derived on a per-head basis—are adjusted by the average head counts at operations with fewer than 500 AU or 300 AU to derive the annualized per-facility compliance cost. EPA assumes that CAFOs with fewer than 500 AU or 300 AU have sufficient cropland for all on-farm nutrients generated (identified in the cost model as Category 1 costs). More detailed cost information is provided in the *Development Document* (USEPA, 2000a).

As explained in Section 4.2 of this report, EPA evaluates the effect of incurred compliance costs based on the total number of CAFOs in each sector, including mixed operations. This approach avoids understating costs at operations with more than one animal type that meets the size threshold for a CAFO or is designated as a CAFO by the Permitting Authority, and thus may incur costs to comply with the proposed requirements for each type of animal that is raised on site. Therefore, EPA’s compliance costs estimates likely represent the upper bound, since costs at facilities with more than a single animal type may, in some cases, be lower due to shared production technologies and practices across all animal types that are produced on site.

The financial data that EPA uses to analyze impacts on small CAFO businesses are from USDA’s ARMS database (see Section 4.2). These data are available for 1997 by commodity sector, facility size (animal inventory), and production region. Available 1997 financial data that are used to characterize average model CAFOs include gross farm revenue, net cash income (used to project cash flow), and baseline debt-to-asset ratios. Table 9-8 shows the gross revenue that EPA assumes for this analysis, expressed on a per-animal basis. Unit revenues shown in Table 9-8 are aggregated by the average number of animals assumed for each model CAFO to derive total entity revenue used in this analysis. Estimated cash flow and debt-to-asset ratios for CAFO models are provided in Section 4 of this report (Tables 4-5 and 4-7).

As Table 9-8 shows, USDA data indicate that operations with fewer than 300 AU, on average, have higher gross revenues when expressed on a per-animal basis than operations with more than 300 AU. This is explained by the fact that smaller farming operations tend to be more diversified and engage in both livestock and crop production. In general, larger businesses tend to be more specialized and concentrate on a single enterprise only. Consistent with SBA’s size standards that are expressed in terms of total annual business revenue (SBA, 1998), EPA assesses financial impacts at model CAFOs based on changes in *total* farm revenue. Total farm revenue, as reported in USDA’s ARMS database, includes gross cash income from both livestock and crop sales (including net Commodity Credit Corporation loans), government payments, and other farm-related income (income from machine-hire, custom work, livestock grazing, land rental, contract production fees, outdoor recreation, and other farm-related sources) (USDA/ERS, 1999a).

Table 9-8. Estimated Per-Head Facility Revenues for Model CAFOs

Sector	Region	Model CAFOs <300AU	Model CAFOs "Medium 1"	Model CAFOs "Medium 2"	Model CAFOs "Large 1"	Model CAFOs "Large 2"
			300 - 1,000 AU		>1,000 AU	
		(incremental compliance costs \$ per animal)				
Fed Cattle	CE			\$502	\$854	
	MW			\$535	\$862	
Veal	MW			\$535	\$862	
Heifers	MW			\$535	\$862	
	PA			\$502	\$854	
Dairy	MW	\$2,620				
	PA					
Hog: FF	MA					
	MW	\$606		\$304		
Hog: GF	MA					
	MW					
Layer: Wet	SO			\$25		
Layer: Dry	MW			\$25		
	SO			\$25		
Broiler	MA			\$1.5	\$1.1	
	SO			\$1.4	\$1.2	
Turkey	MA		\$11.2			
	MW		\$11.2			

Source: Derived from USDA/ERS, 1999a (see Section 4.2.4). Facility size and region definitions for model CAFOs are provided in Section 4, Table 4-1. Large operations roughly correspond to CAFOs with >1,000 AU and Medium operations correspond to CAFOs with 300-1,000 AU. Shaded cells indicate that there are no CAFOs that will be affected by the proposed regulations and that meet the SBA definition of a small business.

Higher total farm revenues per animal at smaller-sized farms (due to the inclusion of revenue from all farm-related sources) is demonstrated in the original USDA ARMS data that are presented in the individual subcategory sections of this report, including Section 6 (poultry), Section 7 (hogs), and Section 8 (cattle and dairy). Derived on a per animal basis, these data show that operations with less than 300 AU tend to generate a larger share of total revenue from other secondary sources, including other secondary livestock revenue as well as revenue from crop sales. Other sources of farm-related revenue that tend to be greater at operations with less than 300 AU, compared to operations with more than 300 AU, include other farm-related revenue, such as government payments and nonfarm income. Since EPA's small business analysis considers a business' total entity revenue, with SBA size standards, the derived per-unit revenues are relatively lower per-unit for model CAFOs with more than 300 AU compared to model CAFOs with fewer than 300 AU. EPA's analysis does not consider sources of non-farm revenue in its analysis, even though data from USDA indicate that nonfarm revenue often constitutes a significant share of total operating income (USDA/ERS, 2000d, 1996a and 1999a).

The same ARMS financial data, however, consistently indicate that per-unit cash expenses tend to be greater among smaller producers than among larger operations. This is consistent with expectations of economies of size in agricultural production. A review of the agricultural literature suggests that there may be a statistically positive relationship between farm size and per-unit production costs, such that as farm size (number of animals) increases, per-unit costs are lower (ERG, 2000d; Lazarus, et al., 1999). This may result in lower per-unit capital costs and create a competitive advantage among larger-sized operations relative to smaller ones. This literature review is provided in the rulemaking record (ERG, 2000d—see DCN 70641).

9.4.2 Economic Analysis Results

Using the proposed economic achievability criteria, discussed in Section 9.4.1, EPA's economic analysis indicates that the proposed regulations will not impose financial stress on a substantial number of operations, relative to the total number of affected confinement operations in these sectors. The results of this analysis are presented in Table 9-9 for each of the co-proposed tier structures. (Results for Scenario 5 (two-tier structure at 750 AU threshold) and Scenario 6 are not determined, but fall within the range of the results presented.)

Under both the two-tier and three-tier structures, EPA's analysis indicates that the proposed requirements will not impose stress impacts on any affected small businesses in the veal, dairy, hog, egg laying, and turkey sectors. Under the two-tier structure, the proposed requirements will not result in financial stress to affected small operations in the heifer sector. Operations in these sectors are expected to be able to absorb the costs associated with the

Table 9-9. Results of EPA’s Small Business Analysis

Sector	Number of Small CAFOs	Affordable	Moderate	Stress	Affordable	Moderate	Stress
		Zero Cost Passthrough					
		(Number of Operations)			(% Affected Operations)		
Two-Tier Structure (Proposed BAT Option/Scenario 4a)							
Fed Cattle	1,390	1,130	250	10	81%	18%	1%
Veal	90	90	0	0	100%	0%	0%
Heifer	800	680	120	0	85%	15%	0%
Dairy	50	40	10	0	80%	20%	0%
Hogs	300	300	0	0	100%	0%	0%
Broilers	9,470	1,860	7,460	150	20%	79%	2%
Layers	200	200	0	0	100%	0%	0%
Turkeys	0	0	0	0	NA	NA	NA
TOTAL	10,550	4,300	7,840	160	41%	74%	2%
Three-Tier Structure (Proposed BAT Option/Scenario 3)							
Fed Cattle	1,490	1,100	380	10	74%	26%	1%
Veal	140	140	0	0	100%	0%	0%
Heifer	980	800	150	30	82%	15%	3%
Dairy	50	40	10	0	80%	20%	0%
Hogs	300	300	0	0	100%	0%	0%
Broilers	13,410	1,910	11,220	280	14%	84%	2%
Layers	590	590	0	0	100%	0%	0%
Turkeys	500	460	40	0	92%	8%	0%
TOTAL	14,630	5,340	11,800	320	37%	81%	2%

Source: USEPA. Impact estimates shown include impacts to designated operations. Option/Scenario definitions provided in Table 3-1. Category definitions (“Affordable,” “Moderate” and “Stress”) are provided in Table 4-13. Numbers may not add due to rounding. NA = Not Applicable.

Number of operations does not adjust for operations with mixed animal types, for comparison purposes, to avoid understating costs at operations with more than one animal type that may incur costs to comply with the proposed requirements for each type of animal that is raised on-site. The number of CAFOs includes designated facilities.

proposed CAFO regulations without having to rely on cost passthrough. EPA's analysis shows that operations across most sectors may experience moderate financial impacts (Table 9-9). Moderate impacts are not associated with operational change at the CAFO (i.e., will not result in facility or product line closure) and are considered by EPA to be economically achievable.

In the cattle and broiler sectors, however, EPA's analysis indicates that each of the co-proposed tier structures will result in financial stress on some small businesses in the fed cattle and broiler sectors, as will the three-tier structure on some small heifer operations. These small businesses may be vulnerable to closure. Overall, operations that may experience financial stress comprise about 2 percent of all affected small CAFO businesses. For the two-tier structure, EPA estimates that 10 small beef operations and 150 small broiler operations will experience financial stress. For the three-tier structure, EPA estimates that 40 small beef and heifer operations and 280 small broiler operations will experience financial stress. No designated operations under either co-proposed scenario are estimated to experience financial stress. Small broiler facilities with stress impacts are larger operations with more than 1,000 AU under both tier structures. Small cattle and heifer operations with stress impacts are those that have a ground water link to surface water. This analysis is conducted assuming that no costs are passed through between the CAFO and processor segments of these industries. Based on the results of this analysis, EPA is proposing that the proposed regulations are economically achievable to small businesses in these sectors.

EPA believes that the estimated financial impacts shown in Tables 9-9 are worst-case. These reasons are summarized below.

First, all results are estimated assuming no costs can be passed through between CAFOs and the processing sectors. As discussed in Section 5 of this report, if modest levels of cost passthrough are assumed in the broiler sectors, then the proposed regulations are affordable to all small broiler operations. EPA did not evaluate economic impacts to cattle operations under a cost passthrough scenario; however, it is expected that long-run market and structural adjustment by producers in this sector will diminish the estimated impacts. Even without assumptions of cost passthrough, EPA's analysis shows that adverse impacts will not be experienced by a substantial number of operations, as compared to the number of affected operations in these sectors. EPA has conducted an extensive literature review of issues concerning cost passthrough. Based on the results of the available empirical research on market power and price transmission in these industries, EPA believes that there is little evidence to support that increased production costs may not be passed through the market levels. A summary of this literature review is provided in the rulemaking record (ERG, 2000c — DCN 70640).

Second, as noted in the Panel Report, EPA believes that the number of small broiler operations is overestimated. In the absence of business level revenue data, EPA estimates the number of "small businesses" using the approach described in Section 9.2. Using this approach, virtually all (>99.9 percent) broiler operations are considered "small" businesses. This categorization may not accurately portray actual small operations in this sector since it classifies a

10-house broiler operation with 260,000 birds as a small business. Information from industry sources suggests that a two-house broiler operation with roughly 50,000 birds is more appropriately characterized as a small business in this sector (Madison, 1999; Staples, 1998). Therefore, it is likely that the number of small broiler operations may reflect a number of medium and large size broiler operations being considered as small entities. As discussed in Section 9.2.1, EPA consulted with SBA on the use of an alternative definition for small businesses in all affected sectors based on animal inventory at an operation during the development of the rulemaking.

Third, EPA believes that the use of a costs-to-sales comparison is a crude measure of impacts on small business in sectors where production contracting is commonly used, such as in the broiler sector (and also in the turkey, egg, and hog sectors, though to a lesser extent). As discussed in Section 4.2.4.5, lower reported operating revenues in the broiler sector reflect the predominance of contract growers in this sector. Contract growers receive a pre-negotiated contract price that is lower than the USDA-reported producer price, thus contributing to lower gross revenues at these operations (USDA, 1999). Lower producer prices among contract growers are often offset by lower overall production costs at these operations, since the affiliated processor firm pays for a substantial portion of the grower's annual variable cash expenses. Inputs supplied by the integrator may include feeder pigs or chicks, feed, veterinary services and medicines, technical support, and transportation of animals (USDA, 1996b). These variable cash costs comprise a large component of annual operating costs, averaging more than 70 percent of total variable and fixed costs at livestock and poultry operations (USDA, 1999). The contract grower also faces reduced risk because the integrator guarantees the grower a fixed output price (see Section 2.3.1 for more details on contracting in animal agriculture). Because production costs at a contract grower operation are lower than at an independently owned operation, a profit test (costs-to-profit comparison) is a more accurate measure of impacts at grower operations. However, financial data are not available that differentiate between contract grower and independent operations.

Fourth, EPA's initial regulatory flexibility analysis also does not consider a range of potential cost offsets available to most farms. One source of cost offset is manure sales, particularly of relatively higher value dry poultry litter. EPA estimates that sales of dry poultry litter could offset the costs of meeting the regulatory requirements on the order of more than 50 percent. This reduction alone exceeds the level of cost passthrough (42 percent) assumed for the cost impact analysis of the broiler sector. Details on how EPA calculated these manure sale offsets and how they would reduce the economic impacts at poultry operations are presented in Section 6.

Another source of potential cost offset is cost share and technical assistance available to farmers for on-farm improvements from various state and federal programs, such as the Environmental Quality Incentives Program (EQIP) administered by USDA. The EQIP program provides cost-share assistance to all livestock and poultry operations, regardless of size, for terraces, filter strips, and runoff trenches, as well as technical assistance in formulating conservation plans. More importantly, operations with 1,000 or fewer AU in confinement, which

make up the majority of small CAFO businesses, are also eligible to receive funding for construction of animal waste storage and treatment facilities (e.g., lagoons, holding tanks). Additionally, many poultry operations with more than 1,000 AU are considered small under SBA definitions, fall below the EQIP size threshold, and are eligible for waste storage and treatment funding (e.g., poultry operations with less than 455,000 broilers or less than 250,000 layers). Although funding may be limited, it is expected that the majority of funds are likely to go to operations eligible for waste storage and treatment funding (ERG, 2000a).

Many other state and federal cost share programs base eligibility not on size thresholds but on priority watersheds (e.g., USDA's Small Watershed Program; the New York City Watershed Program), priority contaminants (e.g., Kansas Non-Point Source Pollution Control Fund), or proposed waste management practices (e.g., Maryland, Minnesota, Missouri, Nebraska, and North Carolina state programs). However, technical assistance under most programs is available to all operations, regardless of watershed, contaminants, proposed practices, or size (ERG, 2000a). A review of cost-share and technical assistance programs available to animal feeding operations is provided in the rulemaking record (ERG, 2000a — DCN 70130).

Finally, this analysis does not take into account certain noneconomic factors that may influence an operation's decision to weather the boom and bust cycles that are commonplace in agricultural markets. Farm typology data from USDA indicate that a large share of farming operations (more than 90 percent) have annual sales of less than \$250,000 and are considered "small family farms" by USDA (USDA/ERS, 2000d and 2000e). Of these, the majority (about 60 percent) are "limited-resource," "retirement," or "residential" operations where farming is not the primary source of income (USDA/ERS, 2000e and 1999a). In many cases, these operations have negative annual income supplemented by sources of off-farm income that subsidize the farming operation (USDA/ERS, 2000d and 1996a).

USDA's ERS (1996a) reports that about 60 percent of farm operators reporting negative net income had nonfarm occupations. About 75 to 80 percent of farms rely on some nonfarm income, and even in the largest operations nonfarm income can be a significant portion of total household income (USDA/ERS, 1996a). More than 90 percent of farm operators with negative net income had nonfarm income averaging more than \$35,700 per year; even farms with positive net income rely somewhat on nonfarm income (Heimlich and Barnard, 1995; USDA/ERS, 1996a).

When farm income is negative over a period of time, sales tests can be very difficult to interpret (Heimlich and Barnard, 1995). One reason that incomes can remain negative over several years is that operators can supplement farm income with nonfarm income, and these losses can be used to reduce total income tax liabilities while the real estate value of the farm property appreciates. Additional noneconomic factors might also include the satisfaction of working for oneself, the ability to employ family members, a sense of tradition and the ability to pass on that tradition to future generations, and the fact that the operation is both a home and a livelihood. These and other noneconomic factors may influence the decision to close a livestock or poultry operation cannot be adequately addressed in an economic model. To the extent that these factors

play a role in that decision, EPA's economic model may overstate the possibility of closure among small businesses.

USDA's farm financial data include operations where farming is part-time and not the primary occupation, but excludes sources of nonfarm income at these operations. As noted in Section 4.2, the inclusion of these operations may result in lower average data values than would be the case if these operations were excluded from the analysis. EPA believes that the inclusion of these operations may tend to overstate impacts. Previous analyses by USDA and EPA have also noted the potential effect on average farm data of including these operations and have regarded these part-time business more as "hobbies or recreational activities" (Heimlich and Barnard, 1995; DPRA, 1995). Heimlich and Barnard (1995) further indicate that considering non-farm income in addition to farm income may provide a more appropriate comparison to the costs of required measures where the motivation for staying in business is not necessarily purely economic.

Overall, EPA expects that the proposed CAFO regulations will benefit the smallest businesses in these sectors, since the regulations may create a comparative advantage for smaller operations (less than 300 or 500 AU), especially those operations that are not subject to the regulations. Except for the few AFOs that are designated as CAFOs, these smaller operations will not incur costs associated with the proposed requirements and may benefit from eventual higher producer prices as these markets adjust to higher production costs in the longer term.

APPENDIX M EXAMPLE OF A SUCCESSFUL FRFA

SECTION FOUR

FINAL REGULATORY FLEXIBILITY ANALYSIS

This section considers the effects of the CAFO regulations on small businesses in the livestock and poultry industries. Section 4.1 discusses EPA's requirements under the Regulatory Flexibility Act. Section 4.2 outlines EPA's initial assessment of small businesses in the sectors affected by the regulations. Section 4.3 presents EPA's final regulatory flexibility analysis and summarizes other steps taken by the Agency to comply with the RFA. Section 4.4 presents the data, methodology, and results of EPA's analysis of impacts on small businesses for this rulemaking.

4.1 THE REGULATORY FLEXIBILITY ACT AS AMENDED BY THE SMALL BUSINESS REGULATORY ENFORCEMENT FAIRNESS ACT

The Regulatory Flexibility Act (RFA, 5 U.S.C. *et seq.*, Public Law 96-354), as amended by the Small Business Regulatory Enforcement Fairness Act of 1996 (SBREFA) generally requires an agency to prepare a regulatory flexibility analysis describing the impact of the regulatory action on small entities as part of the rulemaking. This analysis is required for any rule subject to notice-and-comment rulemaking requirements under the Administrative Procedure Act or any other statute unless the agency certifies that the rule will not have a "significant impact on a substantial number of small entities." Small entities include small businesses, small organizations, and governmental jurisdictions. Because the CAFO regulations could have a significant economic impact on a substantial number of small entities, EPA has prepared this final regulatory flexibility analysis (FRFA).

In addition to the preparation of an analysis, the RFA, as amended by SBREFA, imposes certain responsibilities on EPA when the Agency proposes rules that might have a significant impact on a substantial number of small entities. These include requirements to consult with representatives of small entities about the proposed rule. The statute requires that, where EPA has prepared an initial regulatory flexibility analysis (IRFA), the Agency must convene a Small Business Advocacy Review (SBAR) Panel for the proposed rule to seek the advice and recommendations of small entities concerning the rule. The panel is composed of employees from EPA, the Office of Information and Regulatory Affairs within the Office of Management and Budget, and the Office of Advocacy of the Small Business Administration (SBA).

4.2 INITIAL ASSESSMENT

Prior to the 2001 Proposal, EPA conducted an initial assessment according to Agency guidance on implementing RFA requirements (USEPA, 1999i). First, EPA must indicate whether the proposal is a rule subject to notice-and-comment rulemaking requirements. EPA determined that the proposed CAFO regulations were subject to notice-and-comment rulemaking requirements. Second, EPA should develop a profile of the affected small entities. EPA has developed such a profile of the livestock and poultry sectors, which includes all affected operations as well as small businesses. This information is provided in Section 2 and other sections of the Proposal EA (USEPA, 2001a). Third, EPA's assessment needs to

determine whether the rule would affect small entities and whether the rule would have an adverse economic impact on small entities.

For the proposed rulemaking, EPA could not conclude that costs are sufficiently low to justify “certification” that the regulations would not impose a significant economic impact on a substantial number of entities. Instead, EPA complied with all RFA provisions and conducted outreach to small businesses, convened an SBAR Panel, and prepared an IRFA. That analysis described EPA’s assessment of the impacts of the proposed CAFO regulations on small businesses in the livestock and poultry sectors. A summary of this analysis was published in the *Federal Register* at the time of publication of the 2001 Proposal (66 FR 3099-3103, see: USGPO, 2001a). More detailed information on EPA’s IRFA is provided in Section 9 of the Proposal EA. EPA’s Proposal EA also describes other requirements of EPA’s initial assessment of small businesses and summarizes the steps taken by EPA to comply with the RFA.

Since proposal, EPA has received new information and data related to small business in the livestock and poultry industries, including revisions to the SBA’s definition of “small business” in these sectors and updates to EPA’s estimate of the number of affected operations to reflect USDA estimates. This information was presented in the 2001 Notice (66 FR 58556; USGPO, 2001b). Section 4.2.1 of this report reviews SBA’s revised definitions of small entities in the livestock and poultry industry and discusses a rationale for using an alternative definition of small business in one sector. Section 4.2.2 then uses the definitions of small entities laid out in Section 4.2.1 to estimate the number of operations that meet this small business definition. Section 4.2.3 presents the results of the initial assessment EPA conducted for the 2001 Proposal, which provides a first level screen of potential impacts on small business CAFOs and serves as a signal for additional analysis.

4.2.1 Definition of Small CAFO Businesses

The RFA defines a “small entity” as a small not-for-profit organization, small governmental jurisdiction, or small business. No small governmental operations operate CAFOs. A few not-for-profit organizations might operate CAFOs, but complete information is not available to warrant including not-for-profit organizations in this analysis. The analysis therefore focuses only on small businesses that are defined or designated as CAFOs. (Section 1 of this report describes the circumstances under which an AFO is defined or designated as a CAFO and is subject to the final regulations.)

The RFA requires, with some exceptions, that EPA define small businesses according to SBA’s size standards. SBA sets size standards for defining small businesses by number of employees or amount of revenues for specific industries. These size standards vary by North American Industry Classification System (NAICS) code. CAFOs are listed under NAICS 112, Animal Production.³¹

Table 4-1 shows SBA size standards by SIC code for each of the six livestock and poultry sectors, which are expressed in terms of average “annual receipts” (revenue). With one exception,

³¹ In September 2000, SBA updated the basis for its size standard from Standard Industrial Classification (SIC) codes to NAICS codes (USGPO, 2000; U.S. Census Bureau, 2000). By SIC code, these industries are listed under SIC 02, Livestock and Animal Specialties. The actual size standards for each sector, specified as an annual revenue threshold, did not change as a result of this update.

current SBA standards define a “small business” within each of the main livestock and poultry sectors as an operation that generates average revenues ranging from less than \$0.75 million per year (hog, dairy, broiler, and turkey sectors) to less than \$1.5 million per year (beef feedlot sector), averaged over the three most recent fiscal years (USGPO, 2000; SBA, 1998). The exception is the revenue threshold for a small chicken egg operation, which SBA has defined as a business that generates up to \$9 million annually. For reasons outlined in Section 9.2.1 of the Proposal EA, EPA believes that SBA’s definition of small business for the egg laying sector (revenues of \$9 million per year) does not truly characterize a small business in this sector. As discussed extensively in documentation supporting the 2001 Proposal, EPA is using an alternative definition of \$1.5 million annually for this analysis. Refer to the Proposal EA (USEPA, 2001a) and docket materials cited in that document, and the proposal itself (USGPO, 2001a).

SBA’s size standards differ from the revenue cutoff generally recognized by USDA, which has set \$250,000 in gross sales as its cutoff between small and large family farms (USDA, 1998).

As discussed in the 2001 Notice (66 FR 58570-58571; see USGPO, 2001b), recent revisions to SBA’s small business definitions for some sectors necessitate changes to EPA’s estimate of the number of AFOs that are potentially defined as CAFOs and subject to the final requirements. Prior to June 2001, SBA defined a “small business” for the dairy, hog, broiler, and turkey sectors as an operation with annual sales of less than \$0.5 million per year. On June 7, 2001, SBA raised the size standards for these four sectors to \$0.75 million per year. SBA’s notice of this change is at 66 FR 30646 (USGPO, 2001c). Although SBA did not revise its small business definition for the beef feedlot and egg laying sectors, updates to USDA estimates of the number of AFOs that are potentially defined as CAFOs also require changes to EPA’s overall estimates of the number of small businesses affected by the rulemaking. EPA’s revised estimates of the number of affected small businesses are presented in Section 4.2.2.

Table 4-1. SBA Revenue Size Standards for Small Livestock and Poultry Operations

NAICS Code (SIC Code)	NAICS Industry Description	SBA Size Standard ^{a/}	EPA-Assumed Revenue Cutoff
112112 (0211)	Beef Cattle Feedlots	\$1.5 million	same as SBA
112111 (0241/0212)	Beef Cattle Ranching and Farming	\$0.75 million	same as SBA
11221 (0213)	Hog and pig farming	\$0.75 million	same as SBA
11212 (0241)	Dairy cattle and milk production	\$0.75 million	same as SBA
11232 (0251)	Broilers and other meat-type chickens	\$0.75 million	same as SBA
11231 (0252)	Chicken egg production	\$9.0 million	\$1.5 million
11233 0253	Turkey production	\$0.75 million	same as SBA

Source: SBA, 1998; USGPO, 1991a, 1991b, 1996, 2000, 2001c; U.S. Census Bureau, 2000.

^{a/}SBA Size Standards by NAICS code (13 CFR Part 121) correspond to classifications under SIC classification.

4.2.2 Number of Affected Small Businesses

EPA uses three steps to determine the number of small businesses that might be affected by the CAFO regulations. First, EPA identifies small businesses in the relevant livestock and poultry sectors by equating SBA's annual revenue definition with the number of animals at an operation. Second, EPA estimates the total number of small businesses in these sectors using farm size distribution data from USDA. Third, based on the regulatory thresholds being promulgated, EPA estimates the number of small businesses that would be subject to the final requirements. These steps are summarized below. More detailed information on this approach is presented in Section 9.2.2 of the Proposal EA.

In the absence of entity level revenue data, EPA identifies small businesses in the livestock and poultry sectors by equating SBA's annual revenue definitions of "small business" to the number of animals at these operations (step 1). This step produces a threshold based on the number of animals that EPA uses to define small livestock and poultry operations and reflects the average farm inventory (number of animals) that would be expected at an operation with annual revenues that define a small business. This initial conversion is necessary because USDA data by farm size are not available by business revenue. With the exception of egg laying operations, EPA uses SBA's small business definition to equate the revenue threshold with the number of animals raised onsite at an equivalent small business in each sector (shown in Table 4-1). For egg laying operations, EPA uses an alternative revenue definition of small business, discussed in Section 4.2.1.

EPA estimates the number of animals at an operation to match SBA's small business definitions based on annual revenue size standard (expressed as annual revenue per entity) and USDA-reported farm revenue data that are scaled on a per-animal basis (expressed as annual revenue per inventory animal for an average facility). Financial data used for this calculation are from USDA's 1997 ARMS database (USDA/ERS, 1999a). USDA's data report average national revenue for each sector, combining both livestock and nonlivestock farm revenue (income from crop sales and other farm-related income, including government payments). EPA uses the derived per-animal revenues shown in Table 4-2 to equate SBA's size standard (in revenues) with farm size based on the number of animals, as follows:

$$\frac{\text{Average Number of Animals}}{\text{Farm}} = \frac{\text{SBA's small business definition (\$ per year per farm)}}{\text{average total revenue per head (\$/animal)}}$$

The resultant number of animals represents the average animal inventory threshold for a small business. Estimated "small business" thresholds for each sector are shown in Table 4-2. Additional information on this approach and the data used for this calculation are outlined in Section 4.2.2 of the Proposal EA. The resultant size threshold represents an average animal inventory for a small business.

For the purpose of conducting its FRFA for this rulemaking, EPA is defining "small business" for these sectors as an operation that houses or confines less than the following: 1,400 fed beef cattle (includes fed beef, veal, and heifers); 300 mature dairy cattle; 2,100 market hogs; 37,500 turkeys; 61,000 layers; or 375,000 broilers (Table 4-2). As shown in Table 4-2, with the exception of dairy and some poultry operations, SBA's small business definition for these sectors more or less corresponds to operations with fewer than 1,000 AU being considered small businesses.

EPA then estimates the total number of small businesses in these sectors using facility size distribution data from USDA (step 2). Using the threshold sizes identified for small businesses in the

livestock and poultry sectors (Table 4-2), EPA matches these thresholds with the number of operations associated with the size thresholds, based on available USDA data, to estimate the total number of small animal confinement operations in these sectors. EPA's estimates of the number of potential CAFOs, derived from these USDA data (Kellogg, 2002), are presented in Section 3.1 of this report. This constitutes the primary data source that EPA uses to match the small business thresholds corresponding to SBA's definitions.

Because the USDA data are organized by broad AU groupings—operations with more than 1,000 AU, 750 AU, 500 AU and 300 AU—EPA has matched the animal thresholds above to the closest available AU grouping as follows. For hogs, EPA assumes that data reported for the 1,000 AU threshold (about 2,500 hogs) provide a close approximation of the 2,100 hog threshold to determine the number of small businesses in this sector. For dairies, EPA assumes that the 500 AU threshold (about 350 dairy cows) approximates the 300 dairy cow threshold. For turkey and egg laying operations with dry manure systems, EPA assumes that the 750 AU threshold (about 61,500 layers and 38,500 turkeys) approximates the 61,000 layers and 37,500 turkeys threshold. Because egg-laying operations with wet manure systems are regulated based on a different AU threshold (1,000 AU is equivalent to 30,000 birds), EPA assumes that all estimated operations for this category are small businesses. The resultant estimates of the number of small businesses in these sectors derived under these assumptions, in conjunction with available USDA data (Kellogg, 2002), are presented in Table 4-2.

For both cattle and broilers, EPA also relies on data on operations with more than 1,000 AU (corresponding to operations with 1,000 beef, veal, and heifers, and about 125,000 broilers), but uses these data as a starting point to assess the total number of small businesses in these sectors. To further determine the number of small businesses with more than 1,000 AU (corresponding to operations with less than 1,400 cattle and 375,000 broilers, as shown in Table 4-1), EPA assumes that, for cattle, about 40 percent of operations with more than 1,000 AU are potentially small businesses. This assumption is based on available USDA data on the share of feedlots with between 1,000 and 2,000 head, calculated as a share of all operations with more than 1,000 AU (Krause, 1991). For broilers, EPA assumes that nearly all operations are small businesses, with the exception of the largest 330 operations, which EPA assumes have more than 375,000 birds. This assumption is consistent with that assumed for the 2001 Proposal and is consistent with USDA broiler sales data and information (USDA/NASS, 2000a). The resultant estimates of the number of small businesses in these sectors using this approach, in conjunction with USDA data (Kellogg, 2002) are presented in Table 4-2.

USDA estimates that there were approximately 238,000 animal confinement facilities in 1997 (see Section 3). Table 4-2 presents EPA's estimates of the total number of small livestock and poultry operations that are potentially small businesses. Using the approach outlined in this section, EPA estimates that about 227,000 operations (95 percent of all operations) are small businesses. However, not all of these operations would be affected by the CAFO regulations.

EPA recognizes that this approach might not accurately portray actual small businesses in all cases across all sectors. On the one hand, the resulting small business estimate would suggest that a 15- to 20-house broiler operation with 375,000 birds would be a small business. Information from industry sources, however, suggests that a two-house broiler operation with roughly 50,000 birds is small (Madison, 1999; USEPA, 2000d). Therefore, it is likely that some medium- and large-size broiler operations are being considered small businesses (USEPA, 2000e).

On the other hand, it is possible that the resulting small business estimate might have failed to identify some small businesses in the other sectors as "small." For example, EPA's approach identifies

as a “small business” hog operations with fewer than 2,500 pigs and turkey operations with fewer than 41,250 turkeys, which account for less than 93 percent and 80 percent of all operations, respectively, and less than 40 percent of sales in these sectors (Kellogg, 2002). These proportions are below SBA’s presumed coverage rates, which define as small about 99 percent of all operations and account for approximately 62 percent of sales (Perez, 2000). Therefore, it is likely that there are additional small hog and turkey businesses that are not captured under the revised methodology (USEPA, 2000e).

Table 4-2. Number of Small CAFOs That Might Be Affected by the CAFO Regulations

Sector	Total Annual Revenue ^{a/} (x)	Revenue per Head ^{b/} (Avg. U.S.) (y)	Number of Animals at Small CAFO Businesses (z = x/y)	Estimated Number of AFOs	Total “Small” AFOs	Small Business CAFOs
Cattle ^{d/}	\$1.5	\$1,060	1,400	21,800	20,430	1,200
Dairy	\$0.75	\$2,573	300	94,800	91,360	1,294
Hogs	\$0.75	\$363	2,100	51,800	47,850	1,485
Broilers	\$0.75	\$2	375,000	17,800	17,450	1,822
Egg Layers	\$9.0	\$25	365,000	ND	ND	ND
	\$1.5		61,000	6,400	5,460	486
Turkeys	\$0.75	\$20	37,500	3,300	2,660	27
All AFOs	NA	NA	NA	237,800 ^{e/}	227,120 ^{e/}	6,314

NA=Not Applicable. ND = Not Determined. “AFOs” have confined animals on-site.

^{a/}SBA size standards are at 13 CFR Part 121. EPA assumes an alternative definition of \$1.5 million in annual revenues for egg laying operations.

^{b/}Average revenue per head across all operations for each sector derived from data obtained from USDA’s 1997 ARMS data (USDA/ERS, 1999a). For more information, see Section 4 of EPA’s Proposal EA (USEPA, 2001a).

^{c/}Total small business CAFOs do not include estimates of designated CAFOs.

^{d/}Includes fed cattle, veal and heifers.

^{e/}USDA total include estimates of the number of operations with “cattle other than fattened cattle or milk cows” and also adjusts for double counting, accounting for roughly 42,000 operations (Kellogg, 2002). See Section 3. EPA’s total for broilers and egg layers also differs because of differing 1,000 AU definitions (see Section 3).

The final step (step 3) in EPA’s approach is to estimate the number of small businesses subject to the CAFO regulations based on the regulatory thresholds being promulgated, as discussed in Section 3 of this report. Not all small confinement operations would be subject to the CAFO regulations. The final regulations apply only to those operations that meet the regulatory definition of a CAFO or those that have been designated as CAFOs by the NPDES permitting authority because of risks posed to water quality and public health, as discussed in Section 1. The regulations *define* as a CAFO those operations that confine more than 1,000 AU, as well as a subset of operations with between 300 and 1,000 AU. The final regulations may also apply to an operation that is *designated* as a CAFO by the NPDES permitting authority on a case-by-case basis, based on an on-site inspection. As described in this section, EPA’s estimates of the number of operations is based on USDA information for 1997 (Kellogg, 2002), which

constitutes the primary data source that EPA uses to determine the number of potential small businesses that might be subject to the regulations.

Table 4-3 presents the estimated number of livestock and poultry operations that might be subject to the CAFO regulations and are also small businesses ("small business CAFOs") by facility size category. EPA estimates that of the approximately 238,000 animal confinement facilities in 1997 roughly 95 percent are small businesses. Not all of these operations would be affected by the final rule. Table 4-3 shows EPA's estimates of the number of small business CAFOs that are expected to be affected by this rule. For this analysis, EPA estimates that about 6,200 affected CAFOs across all size categories are small businesses, accounting for more than 40 percent of the estimated 14,515 affected facilities. EPA estimates that among CAFOs with more than 1,000 AU about 2,330 operations are small businesses (accounting for about one-fourth of all CAFOs in this size category). Most affected small businesses are in the broiler sector. Among CAFOs with between 300 and 1,000 AU, EPA estimates about 3,830 operations are small businesses, with most of the affected small businesses are in the hog, dairy, and broiler sectors.

These estimates are based on USDA data for 1997. Because of continued consolidation and facility closure since 1997, EPA's estimates might overstate the actual number of small businesses in these sectors. Ongoing trends are causing some existing small and medium operations to expand their inventories to achieve economies of scale. Some of the CAFOs considered here as small businesses might no longer be counted as small businesses because they now have higher revenues. Furthermore, some CAFOs might now be owned by a larger, vertically integrated firm and might no longer be small businesses. EPA expects that there are few such operations, but it does not have data or information to reliably estimate the number of CAFOs that meet this description. In addition, for reasons noted in the record, EPA believes that the number of small broiler operations is overestimated and might actually include a number of medium and large broiler operations that should not be considered small businesses.

Table 4-3 also shows the expected number of small businesses that may be designated as CAFOs and subject to the rule. EPA estimates that about 172 operations will be designated as CAFOs. This estimate is expressed over the 5-year permit period (that is, assumes that roughly 35 operations will be designated annually). Among these, an estimated 160 operations are in the 300 to 1,000 AU size category; about 12 operations have fewer than 300 AU. (See Table 3-1.) EPA assumes that all of these operations are small businesses.³² For analysis purposes, EPA also assumes that these operations are located in more traditional production regions and are characterized by operations with available land for land application of manure but also high technology needs (see discussion in Section 4.4).

³²EPA expects that USDA will continue to provide voluntary assistance to those additional operations that are now defined as CAFOs under the current permitting requirements and are not covered by the final regulations.

Table 4-3. Total Number of Small Business CAFOs Subject to Regulation

Sector	All AFOs	Total Small Business AFOs	Small Business CAFOs >1,000 AU (Defined)	Small Business CAFOs 300-1000 AU (Defined)	Small Business CAFOs (Designated)
			(Number of operations)		
Fed Cattle	17,800	16,570	538	174	15
Veal	3,840	160	5	7	0
Heifers	170	3,700	97	230	3
Dairy	94,790	91,360	0	1,330	30
Hogs	51,770	47,850	0	1,485	52
Broilers	17,780	17,450	1,303	520	52
Layers	6,450	5,460	383	48	10
Turkeys	3,310	2,660	0	31	10
Total	237,820 ^{a/}	227,120 ^{b/}	2,326	3,825	172 ^{b/}

Sources: Values presented in the table are EPA estimates, derived from published USDA data (Kellogg, 2002). All numbers are rounded to the nearest ten.

^{a/}USDA total include estimates of the number of operations with “cattle other than fattened cattle or milk cows” and also adjusts for double counting, accounting for roughly 42,000 operations (Kellogg, 2002). See Section 3. EPA’s total for broilers and egg layers also differs because of differing 1,000 AU definitions (see Section 3).

^{b/}Number of designated facilities shown over 5-year permit period. EPA assumes all estimated designated facilities are small businesses.

4.2.3 Results of the Initial Assessment for the 2001 Proposal

For past regulations, EPA has often analyzed the potential impacts to small businesses by evaluating the results of a costs-to-sales test, measuring the number of operations that will incur compliance costs at varying threshold levels (including ratios where costs are less than 1 percent, between 1 and 3 percent, and greater than 3 percent of gross income).³³ EPA conducted such an analysis at the time of the 2001 proposal, indicating that about 80 percent of the estimated number of small businesses directly subject to the rule as CAFOs might incur costs in excess of three percent of sales. These results were based on an assessment of the potential impacts on small CAFO businesses based on the results of a sales test for all operations with more than 500 AU. This screening test indicated the

³³ EPA believes that its more refined analysis used for its general analysis (presented in Section 3 of this EA) better reflects the potential impacts to regulated small businesses.

need for additional analysis to characterize the nature and extent of impacts on small entities. This assessment is conducted for those CAFOs that are small businesses, as determined by EPA.

The results of this screening test for the 2001 Proposal indicate that, measured against all confinement operations with more than 500 AU, about 80 percent of the estimated number of small businesses could incur costs in excess of 3 percent of sales. Compared to the total number of all small animal confinement facilities estimated by EPA, operations that are estimated to incur costs in excess of three percent of sales constitute less than two percent of all small businesses in these sectors. The results of this analysis are presented in Section 9.2.3 of the Proposal EA.

Based on the results of this initial assessment, EPA projected that it would likely not certify that the regulations would not impose a significant impact on a substantial number of entities. This is because EPA's initial assessment indicates that the regulations could impose a significant economic impact on a substantial number of entities. Therefore, prior to the 2001 Proposal, EPA convened a SBAR Panel and prepared an initial regulatory flexibility analysis (IRFA) pursuant to Sections 609(b) and 603 of the RFA, respectively, and prepared an economic analysis. Sections 4.3 and 4.4 of this report present the results of EPA's final regulatory flexibility analysis (FRFA).

4.3 EPA COMPLIANCE WITH RFA REQUIREMENTS

4.3.1 Outreach and Small Business Advocacy Review

As required by Section 609(b) of the RFA, as amended by SBREFA, EPA convened a SBAR Panel for the proposed rule. See 66 FR 3121-3124; 3126-3128 (January 12, 2001). The Panel was convened in December 1999. Panel participants included representatives from EPA, the Office of Information and Regulatory Affairs within the Office of Management and Budget (OMB), and the Office of Advocacy of the Small Business Administration (SBA). "Small Entity Representatives" (SERs), who advised the Panel, included small livestock and poultry producers as well as representatives of the major commodity and agricultural trade associations. Throughout the development of these regulations, EPA conducted outreach to small businesses in the livestock and poultry sectors. EPA also consulted with SBA on the use of an alternative definition of small business for the egg laying sector.

Consistent with the RFA/SBREFA requirements, the Panel evaluated the assembled materials and small entity comments on issues related to the elements of the IRFA. The Panel's activities and recommendations are summarized in the *Final Report of the Small Business Advocacy Review Panel on EPA's Planned Proposed Rule on National Pollutant Discharge Elimination System (NPDES) and Effluent Limitations Guideline (ELG) Regulations for Concentrated Animal Feeding Operations* (April 7, 2000), or "Panel Report" (USEPA, 2000e). This document is included in the public record (DCN 93001). Section 12 of the preamble to the 2001 Proposal provides a summary of the Panel's activities and recommendations and describes the subsequent action taken by the Agency (see 66 FR 3121-3124). Section 12 of the preamble to the 2001 Proposal also details various outreach activities conducted by EPA, which include outreach to small businesses in these sectors.

For the 2001 Proposal, EPA prepared an economic analysis of the impacts on small businesses, which is provided in Section 9.4 of the Proposal EA. EPA's economic analysis supporting the final regulations is provided in Section 4.4 of this report.

For all final regulations for which an FRFA is prepared, Section 212 of the RFA requires that the Agency also issue a small entity compliance guide providing a plain language explanation of how to comply with the final regulations. EPA's small entity compliance guide for the CAFO regulations will be issued following promulgation.

4.3.2 EPA's Final Regulatory Flexibility Analysis

For the proposed regulations, EPA has conducted an IRFA, as required by Section 603 of the RFA, as amended by SBREFA. The IRFA must contain the following: (1) a description of the reasons why action by the agency is being considered; (2) a succinct statement of the objectives of, and legal basis for, the proposed rule; (3) a description of and, where feasible, an estimate of the number of small entities to which the proposed rule will apply; (4) a description of the projected reporting, recordkeeping, and other compliance requirements of the proposed rule, including an estimate of the classes of small entities that will be subject to the requirement and the type of professional skills necessary for preparation of the report or record; and (5) identification, to the extent practicable, of all relevant Federal rules that might duplicate, overlap or conflict with the proposed rule. The IRFA shall also contain a description of any significant alternatives to the proposed rule that accomplish the stated objectives of applicable statutes and that minimize any significant economic impact of the proposed rule on small entities. Sections 9.3.2.1 through 9.3.2.6 of the Proposal EA show how EPA addressed each of these requirements in the IRFA it prepared to support the 2001 Proposal. EPA also prepared an economic analysis of the impacts on small CAFO businesses, which is provided in Section 9.4 of the Proposal EA (USEPA, 2001a).

For the final regulations, EPA has conducted an FRFA, as required by Section 604 of the RFA, as amended by SBREFA. The FRFA addresses the issues raised by public comments on the IRFA, which was part of the proposal of this rule. The FRFA must contain: (1) a succinct statement of the need for, and objectives of, the rule; (2) a summary of the significant issues raised by the public comments in response to the initial regulatory flexibility analysis, a summary of the assessment of the agency of such issues, and a statement of any changes made in the proposed rule as a result of such comments; (3) a description of and an estimate of the number of small entities to which the rule will apply or an explanation of why no such estimate is available; (4) a description of the projected reporting, recordkeeping, and other compliance requirements of the rule, including an estimate of the classes of small entities that will be subject to the requirement and the type of professional skills necessary for preparation of the report or record; and (5) a description of the steps the agency has taken to minimize the significant economic impact on small entities consistent with the stated objectives of applicable statutes, including a statement of the factual, policy, and legal reasons for selecting the alternative adopted in the final rule and why each one of the other significant alternatives to the rule considered by the agency that affect the impact on small entities was rejected. Sections 4.3.2.1 through 4.3.2.5 of this report address each of these FRFA requirements.

4.3.2.1 Need for and Objectives of the CAFO Regulations

A detailed discussion of the need for the regulation is presented in Section 4 of the 2001 Proposal (66 FR 2293-2972-2976). A summary is also provided in Sections 1 and 10 of the Proposal EA. In summary, EPA's rationale for revising the existing regulations include the following: address reports of continued discharge and runoff from livestock and poultry operations in spite of the existing

requirements; update the existing regulations to reflect structural changes in these industries over the past few decades; and improve the effectiveness of the existing regulations.

Despite nearly 30 years of regulation, there are persistent reports of discharge and runoff of manure and manure nutrients from livestock and poultry operations. Revisions to the existing ELG and NPDES regulations for CAFOs are expected to mitigate future water quality impairment and the associated human health and ecological risks by reducing pollutant discharges from the animal production industries.

EPA's revisions also address the changes that have occurred in the animal production industries in the United States since the development of the existing regulations. The continued trend toward fewer but larger operations, coupled with greater emphasis on more intensive production methods and specialization, is concentrating more manure nutrients and other animal waste constituents within some geographic areas. This trend has coincided with increased reports of large-scale discharges from these facilities and continued runoff that is contributing to the significant increase in nutrients and resulting impairment of many U.S. waterways.

EPA's revisions to the existing regulations will make the regulations more effective in protecting or restoring water quality. The revisions will also make the regulations easier to understand and will clarify the conditions under which an AFO is a CAFO and, therefore, subject to the regulatory requirements.

A detailed discussion of the objectives and legal basis for these regulations is presented in Sections 1 and 3 of the preamble to the final rule and also the 2001 Proposal (see: 66 FR 2959 or USGPO, 2001a).

4.3.2.2 Significant Comments in Response to the IRFA

The significant issues raised by public comments on the IRFA address exemptions for small businesses, disagreement with SBA definitions and guidance on how to define small businesses for these sectors, and general concerns about EPA's financial analysis and whether the analysis adequately captures potential financial effects on small businesses.

Commenters generally recommend that EPA exempt all small businesses from regulation, arguing in some cases that regulating small businesses could affect competition in the marketplace, discourage innovation, restrict improvements in productivity, create entry barriers, and discourage potential entrepreneurs from introducing beneficial products and processes. Several commenters claimed that EPA had misrepresented the number of small businesses. In particular, several commenters objected to SBA's small business definition for dairy operations, claiming it understates the number of small businesses in this sector (see, for example, NMPF, 2001). One commenter claimed that EPA's estimate of the total number of operations is understated and therefore must understate the number of small businesses (Department of Agriculture, 2001). Some commenters objected to the consideration of total farm-level revenue to determine the number of small businesses because this approach understates the number of small businesses, despite SBA guidance that bases its definitions on total entity revenue for purposes of defining a small business (NCBA, 2001). Other commenters, however, claimed that EPA's approach does not truly capture operations that are, in fact, small businesses but reflect larger corporate operations (see, for example, Citizens Against Poultry Pollution, 2001). Another commenter

recommended that EPA simply consider any operation with fewer than 1,000 AU as small businesses (Wyoming Office of Federal Land Policy, 2001). EPA also received comments requesting that EPA consider use of regional-specific definitions of small business because of concerns that the revenue-based SBA definition might not be applicable to operations in Hawaii since producers in that State generally face higher cost of production and also higher producer prices relative to revenue and cost conditions at farms in the contiguous 48 States. Comments from SBA recommended that EPA adopt the Panel's recommendation not to consider changing the designation criteria for operations with fewer than 300 animal units as a means to provide relief to small businesses (SBA, 2001). SBA also recommends that EPA adopt the SBAR Panel's approach and allow permitting authorities to focus resources where there is greatest need (SBA, 2001). Finally, some commenters generally questioned the results of EPA's financial analysis, giving similarly stated concerns about EPA's financial data and models used for its main analysis (see, for example, NCBA, 2001).

In response, EPA notes that the projected impacts of today's final regulations on small businesses are lower than the projected impacts of the proposed rule. For example, the final rule does not extend the effluent guideline regulations to CAFOs with between 300 and 1,000 AU, as was proposed in the 2001 proposal. Instead, EPA is retaining the existing regulatory threshold, applying the effluent guideline to CAFOs with more than 1,000 AU only. Requirements for CAFOs with between 300 and 1,000 AU will continue to be subject to the BPJ requirements as determined by the permitting authority, thus requiring that fewer small businesses adopt the effluent guideline standards. More information on this topic is available in section IV of this preamble. Section 4 of the final rule preamble discusses other regulatory changes since the 2001 proposal, indicating greater alignment with SBAR Panel recommendations. Refer to Section 4 of the preamble for more information on the comments and EPA's responses to those comments, as well as EPA's justification for final decisions on these options.

EPA received two comments from one commenter requesting that EPA not use the alternative definition for egg-laying operations but instead consider regional-specific conditions for determining the number of small businesses. The commenter expressed concern that SBA's revenue-based definition might not be applicable to operations in Hawaii since producers in that State generally face higher cost of production and also higher producer prices relative to revenue and cost conditions at farms in the contiguous 48 States. There are a number of reasons why EPA did not use a regional-specific definition of small business for egg operations. First, as instructed under the Regulatory Flexibility Act (RFA), EPA uses small business definitions as defined by the Small Business Administration (SBA) for all sectors (except for the egg-laying sector). Since size standards set by the Small Business Administration (SBA) do not vary by region, EPA follows SBA's lead. Second, the regulations set requirements by the number of animal units at a farm, not the revenues associated with those animal units. A 1,000 AU egg-laying operation in the Midwest will be subject to the same effluent limitations guidelines as a 1,000 AU egg-laying operation in Hawaii and the territories. Third, the economic analysis, uses a representative farm approach. Only the broadest regional information could be obtained through USDA and other sources. Although some small subregions or localities might face unique issues, without performing a Section 308 survey of all regulated entities EPA must rely on the representative farm approach. (See also response to comment DCN CAFO201246C-6 regarding EPA's use of a representative farm approach, which is consistent with longstanding practices at USDA and the land grant universities.) Fourth, very few impacts are seen in the egg-laying sector, regardless of size. Even if EPA had classified the majority of egg-laying operations with less than 1,000 AU as small businesses, this would not have changed the outcome of the Agency's small business analysis in any material way. Finally, even if EPA were to classify all operations as small businesses in areas outside the contiguous 48 States (including Hawaii and Alaska), this would only raise the total number of small business by less than 10 operations. See

response to comment DCN CAFO NODA600053-5 regarding EPA's consideration of regional-specific definition of small business for the regulated sectors.

Regarding EPA's estimate of the number of small businesses, the Agency continues to follow SBA guidance and SBA definitions on how to define small businesses for these sectors. However, EPA has made substantial changes to the financial data and models used for its main analysis, which is also used to evaluate financial effects on small businesses. Both the 2001 Notice (66 FR 58556) and the 2002 Notice (67 FR 48099) describe the public comments received by EPA on the baseline financial data and the methodological approach developed by the Agency to evaluate financial effects. These comments and how EPA has addressed them are discussed more fully throughout this report. EPA's detailed responses to comments, and the comments themselves, are contained in the Agency's comment response document (see, for example, DCN CAFO200179D-3).

4.3.2.3 Description and Estimate of Number of Small Entities Affected

The small entities subject to this rule are small businesses. No nonprofit organizations or small governmental operations operate CAFOs. As discussed in section 7 of the preamble to the final rule, to estimate the number of small businesses affected by this final rule, EPA relied on the SBA size standards for these sectors, with the exception of size definitions for the egg sector. SBA defines a "small business" in these sectors as an operation with average annual revenues of less than \$0.75 million for dairy, hog, broiler, and turkey operations; \$1.5 million in revenue for beef feedlots; and \$9.0 million for egg operations. The definitions of small business for the livestock and poultry industries are in SBA's regulations at 13 CFR 121.201. For this rule, EPA proposed and solicited public comment on and is using an alternative definition for small business for the egg-laying operations. EPA defines a "small" egg laying operation for purposes of its regulatory flexibility assessments as an operation that generates less than \$1.5 million in annual revenue. EPA consulted with SBA on the use of this alternative definition, as documented in the rulemaking record for the 2001 proposal. Given these considerations, EPA evaluates "small business" for this rule as an operation that houses or confines fewer than 1,400 fed beef cattle (includes fed beef, veal, and heifers); 300 mature dairy cattle; 2,100 market hogs; 37,500 turkeys; 61,000 layers; or 375,000 broilers. The approach used to derive these estimates is described in Section 4.2.

Using these definitions and available data from USDA and industry, EPA estimates that about 6,200 affected CAFOs across all size categories are small businesses. Among CAFOs with more than 1,000 AU, EPA estimates that about 2,330 operations are small businesses. Among CAFOs with between 300 and 1,000 AU, EPA estimates that about 3,870 operations are small businesses. Table 4.3 shows EPA's estimates of the number of regulated small businesses across all industry sectors. Table 4.4 provides this information by sector and by representative CAFO model.

Table 4-4. Numbers of Small CAFO Businesses by Sector, Size Grouping, and Region

Sector	Region	Total	CAFOs "Large 1"	CAFOs "Large 2"	CAFOs "Medium 3"	CAFOs "Medium 2"	CAFOs "Medium 1"
Fed Cattle	All	858	538		49	95	176
	CE	605	342		39	79	145
	MW	209	170		8	11	20
	PA	28	23		1	1	3
	SO	1	0		0	0	1
	MA	15	3		1	4	7
Veal	MW	12	5		1	2	4
Heifers	All	329	97		29	70	133
	MW	127	0		6	14	107
	CE	120	58		14	35	13
	PA	82	39		9	21	13
Dairy	All	1,294					1,294
	CE	276					276
	MW	361					361
	PA	227					227
	SO	97					97
	MA	333					333
Hog: FF	All	916			123	252	541
	MA	811			105	216	490
	MW	105			18	36	51
Hog: GF	All	568			103	174	291
	MA	500			88	148	264
	MW	68			15	26	27
Broiler	All	1,822	1,034	269	184	197	138
	MA	1,223	698	182	120	133	90
	SO	599	336	87	64	64	48
Layer: Dry	All	59				21	38
	MW	29				10	19
	SO	30				11	19
Layer: Wet	SO	431	383		48		
Turkey	All	26				9	17
	MA	14				5	9
	MW	12				4	8

Source: USEPA. Preliminary based on estimates associated with the August 4, 2002, cost estimates. Size and region breakouts are based on an analysis of the 1997 Census data by USDA (Kellogg, 2002). Facility size and region definitions for model CAFOs are provided in Section 2, Table 2-1. Shaded cells indicate that there are no small CAFO businesses that will be affected by the regulations and that meet the SBA definition of a small business. Estimates do not include number of designated CAFOs.

4.3.2.4 Description of the Reporting, Recordkeeping, and Other Requirements

The final regulations would require all AFOs that meet the CAFO definition to apply for a permit, develop and implement a nutrient management plan, collect and maintain records required by applicable technology-based effluent discharge standards, and submit an annual report to the responsible NPDES permitting authority. (No nonprofit organizations or small governmental operations operate CAFOs.) All CAFOs would also be required to maintain records of off-site transfers of manure. Record-keeping and reporting burdens include the time to record and report animal inventories, manure generation, field application of manure (amount, method, date, weather conditions), manure and soil analysis results, crop yield goals, findings from visual inspections of feedlot areas, and corrective measures. Records may include manure spreader calibration worksheets, manure application worksheets, maintenance logs, and soil and manure test results. EPA believes the owner/operator has the skills necessary to keep these records and make reports to the permitting authority.

State permitting authorities will incur reporting burdens when they update their NPDES programs to incorporate the regulatory changes in the final rule. They will incur record keeping burdens as they implement the final rule. Data collection and record keeping activities include reviewing CAFO permit applications and periodic reports, and tracking compliance through on-site inspections.

EPA has estimated the burden and costs associated with information collection imposed on CAFOs, including small businesses, and also States as a result of the CAFO regulations. This analysis is provided in the Information Collection Request for the Final NPDES and ELG Regulatory Revisions for Concentrated Animal Feeding Operations (EPA ICR NO. 1989.02) prepared by EPA (USEPA, 2002j), which updates an analysis conducted for the 2001 Proposal (USEPA, 2000f).

For the purpose of this analysis, "burden" means the total time, effort, or financial resources expended by persons to generate, maintain, retain, or disclose or provide information to or for a federal agency. This includes the time needed to review instructions; develop, acquire, install, and utilize technology and systems for the purposes of collecting, validating, and verifying information, processing and maintaining information, and disclosing and providing information; adjust existing procedures to comply with any previously applicable instructions and requirements; train personnel to be able to respond to a collection of information request; search data sources; complete and review the collection of information; and transmit or otherwise disclose the information.

EPA's labor burden estimates for CAFO and State respondents are the hours of activity required to comply with changes to the NPDES CAFO program. For each activity, EPA estimates the burden in terms of the expected effort necessary to carry out these activities under normal conditions and reasonable labor efficiency. These activities and estimated burden and cost levels are described in more detail in the Supporting Statement for the ICR (USEPA, 2002j). The ICR also contains a summary of wage rate information from USDA and the Bureau of Labor Statistics, compiled by EPA for the purpose of this analysis. Additional information on the ICR is provided in the preamble supporting the final regulations. A summary of the analysis of impacts to CAFO operators is provided below. Additional information on the estimated burden and costs is provided in the ICR (USEPA, 2002j).

EPA identifies four burden categories to CAFO operators, including start-up activities, permit application, best available technology requirements, and NPDES record keeping and reporting requirements.

Start-up activities are steps that a CAFO owner or operator must take in preparation to comply with the information collection requirements of the final rule. Owners or operators that are potentially affected by the rule will need to familiarize themselves with the changes to the NPDES CAFO program to determine that they will need to apply for a permit.

Permit application activities include completing and submitting either an NOI to obtain coverage under a general permit or an application for an individual permit. These activities will be conducted once every five years. The final rule requires that the following information be provided on the application forms: the name of the owner or operator; facility address and mailing address; latitude and longitude of the production area; a topographic map; the type and number of animals in open confinement and housed under a roof; the type of containment and total capacity for manure, litter, and wastewater storage; the number of acres for land application; the estimated amount of manure generated per year; and estimated amount of manure transferred off site each year. As part of their record keeping responsibilities, CAFO operators will be required to keep the plan on site for inspections and make it available to the permitting authority on request. Nutrient management plans must be reviewed and rewritten at least every five years.

CAFO owners or operators will perform and record various activities to meet the BAT requirements such as visual inspections of the feedlot facilities, inspections of manure application equipment, collection of soil samples, and recording of volume of manure and process wastewater produced. CAFOs with more than 1,000 AU will also be required to record information for transfers of manure, litter, and process wastewater to other people.

In addition to the labor costs associated with these activities, EPA estimates the capital and operation and maintenance (O&M) costs incurred to collect data and keep records. A CAFO will incur capital costs when it purchases equipment or builds structures that are needed for compliance with the rule's reporting and record keeping requirements that the facility would not use otherwise. Consistent with the overall cost analysis for the final rule, capital costs are annualized assuming a 10-year amortization period and a 7 percent interest rate. Capital costs for the final rule include such items as purchasing a soil auger to collect soil samples and a manure sampler. Some facilities will also need to install depth markers in their lagoons. A facility incurs O&M costs when it regularly uses services, materials, or supplies needed to comply with the rule's reporting and record keeping requirements that the facility will not use otherwise. Any cost for the operation and upkeep of capital equipment is considered an O&M cost. O&M costs may also be incurred on a non-annual basis, such as every five years for a soil analysis. O&M costs include laboratory analysis of soil and manure.

EPA estimates that the public burden for this information collection request will require 1.6 million labor hours for all CAFO respondents to comply with the final regulations and 0.3 million labor hours for State permitting authority respondents (USEPA, 2002j). Information collection and reporting at a CAFO is associated with applying for permits, developing nutrient management plans, conducting site inspections, tracking land application and off-site manure transfers. These estimates include the time required to review instructions, search existing data sources, gather and maintain all necessary data, and complete and review the information collection. EPA estimates costs to regulated CAFOs at \$29 million annually, which includes \$25 million in labor costs and \$4 million capital and O&M expenditures; annual State costs of \$10 million include \$8.6 million in labor costs and \$1.6 million in O&M expenditures (USEPA, 2002j). This estimate excludes NPDES burden for CAFOs covered by other ICR estimates.

Under the final rule, EPA estimates that there will be an annual average of 11,712 CAFO respondents and an annual average of 82,705 CAFO responses, which includes multiple responses per CAFO. Thus, the annual average burden per CAFO respondent is 138 hours and the average burden per CAFO response is 19 hours. For this analysis, EPA assumes that the administrative burden assumptions are generally the same regardless of CAFO size. The annual average burden per State respondent is 10,152 and the average burden per response is 16 hours (USEPA, 2002j).

More detailed information on the burden and associated costs for each of the activities described above is provided in the ICR (USEPA, 2002j). Section 10 of the final rule preamble further summarizes the expected reporting and record-keeping requirements under the final regulations based on information compiled as part of the *Information Collection Request for the Final NPDES and ELG Regulatory Revisions for Concentrated Animal Feeding Operations* (OMB ICR NO. 2040-0250) prepared by EPA.

4.3.2.5 Steps Taken to Minimize Significant Economic Impacts on Small Entities

For the final regulations, EPA has adopted an approach for a regulatory program that mitigates impacts on small business, recognizes and promotes effective non-NPDES State programs, and works in partnership with USDA to promote environmental stewardship through voluntary programs, and financial and technical assistance. EPA's proposal included many options that were not finally adopted in deference to these principles.

Because of the estimated impacts on small entities EPA is not certifying that this rule will not impose a significant impact on a substantial number of small entities. EPA has complied with all RFA provisions and conducted outreach to small businesses, convened a SBAR panel, prepared an Initial Regulatory Flexibility Analysis (IRFA) and a Final Regulatory Flexibility Analysis (FRFA), and also prepared an economic analysis. The Agency's actions include the following efforts to minimize impacts on small businesses:

- Retained structure of existing regulations, which allows EPA and states to focus on the largest producers;
- Retained existing designation criteria and process;
- Retained existing definition of an AFO;
- Retained conditions for being defined as a Medium CAFO;
- Eliminated the "mixed" animal calculation for operations with more than a single animal type for determining which AFOs are CAFOs;
- Raised the duck threshold for dry manure handling duck operations; and
- Adopted a dry-litter chicken threshold higher than proposed.

EPA went to some length to explore and analyze a variety of ELG regulatory alternatives to minimize impacts on small businesses. The record for today's rule includes extensive discussions of the alternatives, EPA's analysis of those alternatives, and the rationale for the Agency's decisions. In large part, the Agency incorporated most of the alternative considerations to reduce the burden to small businesses. By way of example, today's regulations will affect fewer small businesses at significantly reduced costs, as compared to the estimates of the number of operations and expected costs to those affected entities based on the requirements set forth in the 2001 proposal. For more information on EPA's option selection rationale, see Section 4 of the preamble to the final rule.

4.4 EPA'S ANALYSIS OF SMALL BUSINESS IMPACTS

This section discusses the data and methodology EPA uses to assess economic impacts on small CAFO businesses (Section 4.4.1) and presents the results of this analysis (Section 4.4.2). This economic analysis supports the FRFA (Section 4.3) by quantifying the effects of the CAFO regulations. Based on the results of this analysis, EPA has determined that the CAFO regulations would result in financial stress to some affected small businesses, but not a substantial number of operations relative to the total number of affected small businesses in these sectors.

4.4.1 Data and Methodology

To examine the economic impacts of the final regulations on small CAFO businesses, EPA uses the same representative farm approach that is used to analyze impacts on all CAFOs (regardless of size), as described in Section 2 of this EA.³⁴ This approach evaluates impacts on select model CAFOs and extrapolates these results to the number of operations identified by each representative model, thus aggregating costs nationally across all sectors. Inputs for this analysis include the number of CAFOs represented by each model (see Section 4.3.3) and, for each model CAFO, the costs of the final regulations and selected financial characteristics (see Section 2).

EPA's analysis evaluates the economic achievability of the final regulatory options at small CAFO businesses based on financial criteria established for this analysis. These criteria reflect a combination of both farm level and enterprise level criteria. Three farm level criteria are assessed: (1) a comparison of incremental costs to total revenue (sales test), (2) projected post-compliance cash flow over a 10-year period, and (3) an assessment of an operation's debt-asset ratio under a post-compliance scenario. Projected post-compliance cash flow over a 10-year period is also assessed at the enterprise level in order to evaluate the potential effects at a facility's livestock or poultry enterprise, apart from the effects assessed for the entire facility.

EPA used the results from these analyses to divide affected CAFOs into three financial impact categories: Affordable, Moderate, and Stress. CAFOs experiencing affordable or moderate impacts are considered to experience some financial impact on operations, but EPA does not expect the costs of complying with this rule to make such operations vulnerable to closure. EPA considers that for CAFOs in both the "Affordable" and "Moderate" impact categories the final requirements are economically achievable. Operations experiencing financial stress, however, are considered to be vulnerable to closure because of the costs of this rule. EPA considers that for CAFOs in the "Stress" impact category, the final requirements might not be economically achievable, subject to other considerations. For more information on this decision framework, see Table 2-8 and Figure 2-1.

EPA conducted its analysis first at the farm level based on data reflecting financial conditions for the entire farm operation (e.g., reflecting income and cost information spanning the entire operation, thus

³⁴ For past regulations, EPA has often analyzed the potential impacts to small businesses by evaluating the results of a costs-to-sales test, measuring the number of operations that will incur compliance costs at varying threshold levels. EPA conducted such an analysis at the time of the 2001 proposal, but believes that its more refined analysis used for its general analysis better reflects the potential impacts to regulated small businesses.

considering the operation's primary livestock production, along with other income sources such as secondary livestock and crop production, government payments, and other farm-related income). Based on the farm level results, EPA also assessed the financial effects on CAFOs at the enterprise level (e.g., limiting the scope of the assessment to the operation's livestock or poultry enterprise, and excluding other non CAFO-related sources of income from the analysis).

Starting with the farm level analysis, EPA considers the regulations to be economically achievable for a representative model CAFO if the average operation has a post-compliance sales test estimate within an acceptable range, a positive post-compliance cash flow over a 10-year period, and a post-compliance debt-to-asset ratio not exceeding a benchmark value. Specifically, if the sales test shows that compliance costs are less than 3 percent of sales, or if post-compliance cash flow is positive and the post-compliance debt-to-asset ratio does not exceed a benchmark (depending on the baseline data) and compliance costs are less than 5 percent of sales, EPA considers the options to be "Affordable" for the representative CAFO group. (Although a sales test result of less than 3 percent does indicate "Affordable" in the farm level analysis, further analysis is conducted to determine the effects at the operation's livestock or poultry enterprise.) The benchmark values assumed for the debt-asset test are sector-specific. EPA assumes a 70 percent benchmark value for the debt-asset test to indicate financial stress in the hog and dairy sectors, and an 80 percent benchmark for the debt-asset test to indicate financial stress in the beef cattle sector. These benchmark values address public comment received and alternative debt and asset data submitted for the livestock sectors. For the poultry sectors, however, EPA did not obtain alternative debt and asset data and continues to evaluate data used for proposal against a 40 percent benchmark value.

A sales test of greater than 5 percent but less than 10 percent of sales with positive cash flow and a debt-to-asset ratio of less than these sector-specific debt-asset benchmark values is considered indicative of some impact at the CAFO level, but at a level not as severe as those indicative of financial distress or vulnerability to closure. These impacts are labeled "Moderate" for the representative CAFO group. EPA considers both the "Affordable" and "Moderate" impact categories to be economically achievable by the CAFO, subject to the enterprise analysis (see below). If, with a sales test of greater than 3 percent, post-compliance cash flow is negative or the post-compliance debt-to-asset ratio exceeds these sector-specific debt-asset benchmarks, or if the sales test shows costs equal to or exceeding 10 percent of sales, EPA considers the final regulations to be associated with potential financial stress for the entire representative CAFO group. In such cases, each of the operations represented by that group might be vulnerable to closure. For operations that are determined to experience financial "Stress" at the farm level, the final requirements are likely not economically achievable.

The enterprise level analysis builds on the farm level analysis, evaluating effects at a farm's livestock or poultry enterprise. If the farm level analysis shows that the regulations impose "Affordable" or "Moderate" effects on the operation, the enterprise level analysis is conducted to determine whether the enterprise's cash flow is able to cover the cost of regulations. This analysis uses a discounted cash flow approach similar to that used to assess the farm level effects, in which the net present value of cash flow is compared to the net present value of the total cost of the regulatory options over the 10-year time frame of the analysis. Over the analysis period, if an operation's livestock or poultry enterprise maintains a cash flow stream that both exceeds the cash costs of the rule (operating and maintenance costs plus interest) and covers the net present value of the principal payments on the capital, EPA concludes that the enterprise will likely not close because of the CAFO rule. This analysis is conducted on a pass/fail basis. If the net present value of cash flow minus the net present value of the rule's costs is greater than zero, the enterprise passes the test and the enterprise is assumed to continue to operate. EPA

considers these results to indicate that the final requirements are economically achievable. If the net present value of cash flow is not sufficient to cover the net present value of the cost of the rule, EPA assumes that the CAFO operator would consider shutting down the livestock or poultry enterprise. That is, if an operation fails the enterprise level analysis, these operations are determined to experience financial “Stress” and the final requirements are likely not economically achievable.

More detail on the classification scheme established for this analysis, along with a discussion of the basis for EPA’s use of these criteria, is provided in Section 2. Section 2.3. presents the baseline (farm and enterprise level) financial data that EPA uses to analyze impacts on small CAFO businesses.

Appendix B shows EPA’s estimated compliance costs for selected model CAFOs under the final BAT Option. These costs reflect the range of facility level costs for model CAFOs based on estimated per-unit costs aggregated by the average number of animals assumed for each model. All costs shown are expressed on a per-animal basis and are differentiated by facility size, producing region, facility types, and other factors. Costs are reported in ranges across three types of land availability for manure application and also across three types of technology needs assumed for model CAFOs for the purpose of this analysis. The land availability types include: Category 1 farms, which have sufficient cropland for all on-farm nutrients generated; Category 2 farms, which have insufficient cropland; and Category 3 farms, which have no cropland. USDA data/information grouping facilities into the categories of technology adoption and use are: “least needs” and “most needs” operations (assumed to account for 25 percent each of all facilities) and also “average needs” (assumed to account for 50 percent of all operations). These groupings are based on available USDA data; detailed information is available in the *Development Document* supporting the proposed regulations (USEPA, 2002). Section 2 provides a summary on EPA’s engineering cost models.

To estimate financial effects on operations with between 300 and 1,000 AU that may be defined as CAFOs under the NPDES permit regulations, EPA assumes that the estimated costs for CAFOs with between 300 and 1,000 AU to comply with the effluent guideline regulations are similar to the costs that will be incurred by sized operations of that size to comply with BPJ requirements under the revised NPDES regulations. Because the costs to comply with the effluent guideline represent the likely high end of the possible cost range, estimated impacts on operations in this size range might be overstated.

To estimate financial effects on expected designated facilities, EPA uses the same general approach described in Section 2 of this report to assess impacts on an estimated 344 designated facilities over a 10-year period.³⁵ For this analysis EPA uses estimated costs for the smallest size model CAFO among operations with between 300 and 1,000 AU (“Medium” operations)³⁶ for model CAFOs developed for operations located in the more traditional production regions (Midwest for the livestock and turkey operations and South for the broiler and egg-laying operations; Table 2-1 shows these definitions). For example, EPA assumes that operations characterized as having available land for land application of manure (Category 1 model facilities) and high technology needs (“most needs” or Category H) may be characterized as Category 1H models for purposes of costing across the range of

³⁵ As shown in Table 3-3, EPA estimates 172 designated facilities over a 5-year permit period. For the purpose of this analysis, EPA assumes that half are expected to be designated during the first 5 years, and the other half, in the second 5 years.

³⁶Medium 3 for wet layer operations.

land base and technology needs cost models. These cost estimates are shown in Appendix B. More detailed cost information is provided in the *Development Document* (USEPA, 2002).

For CAFOs with between 300 to 1,000 AU, operations are distributed in the key regions (Midwest or South) in the key category group (Category 1H) across the Medium 1, 2, and 3 model sizes. For all sectors excluding hog, the farm counts are distributed evenly across these three size groups. The hog models are more complex, because the engineering costs are divided by size, region, operation type (farrow finish and grow finish) and manure process (liquid and pit for Medium models), and the financial models are divided by size, region, and contract versus independent, and so forth, leading to a much larger matrix of models than those for other sectors. The designated counts were distributed in a ratio of NPDES farm counts over the Medium 1, 2, and 3 models for liquid and pit manure processes, by contract vs. independent and by farrow finish and grow finish in the Midwest 1H categories. For designated CAFOs with fewer than 300 AU, operations are placed in one model for each sector, with the exception of hog facilities. Hog operations are distributed evenly among the model types (manure process by grow finish or farrow finish and by contract versus independent). Costs for these "Small" models are developed using the Medium 1, Category 1H costs for each sector (or in the case of hog, each process and operation type). The cost per head for the Medium 1, 1H operation was applied to an assumed 300 AU number of head to estimate an annualized compliance cost per "small" facility. Because there was no Medium 1 size for wet layer, the Medium 3 size group per head cost was applied to the number of head associated with 300 AU.

4.4.2 Economic Analysis Results

Using the economic achievability criteria established for this analysis, discussed in Section 4.4.1, EPA's economic analysis indicates that the CAFO regulations will not impose financial stress on a substantial number of operations, relative to the total number of affected confinement operations in these sectors. The results of this analysis are presented in Table 4-5.

EPA estimates that about 6,200 small business CAFOs would be affected by this rule. For this analysis, EPA estimates that about 6,200 affected CAFOs are small businesses, consisting of about 2,330 operations with more than 1,000 AU and about 3,830 operations with between 300 and 1,000 AU. Most of these affected small businesses are in the hog, dairy, and broiler sectors.³⁷

In examining the effects on small businesses for the final rule, EPA followed the same approach used to evaluate the impacts on existing CAFOs, as described in Section ES.2. For the purposes of this analysis, EPA assumes that small business CAFOs with between 300 and 1,000 AU would incur costs similar to those estimated for CAFOs with more than 1,000 AU (although these smaller-sized operations will be subject to BPJ and not the ELG requirements under the revised NPDES requirements). These upper end cost estimates could, therefore, overstate the financial effects for small businesses in this size category. For past regulations, EPA has often analyzed the potential impacts to small businesses by evaluating the results of a costs-to-sales test, measuring the number of operations that will incur compliance costs at varying threshold levels (including ratios where costs are less than 1 percent,

³⁷ For reasons noted in the record, EPA believes that the number of small broiler operations is overestimated and might actually include a number of medium and large broiler operations that should not be considered small businesses.

between 1 and 3 percent, and greater than 3 percent of gross income). EPA conducted such an analysis at the time of the 2001 proposal, indicating that about 80 percent of the estimated number of small businesses directly subject to the rule as CAFOs might incur costs in excess of three percent of sales. EPA believes that its more refined analysis used for its general analysis (presented here) better reflects the potential impacts to regulated small businesses.

Using this approach, EPA's analysis indicates that the final rule could cause financial stress to some small businesses, making these businesses vulnerable to closure. These results are presented in Table 4-5a (Option 1) and Table 4-5b (Option 2).

For Option 1, the analysis indicates that, among all small business CAFOs in the veal, dairy, hog, turkey, and egg-laying sectors, the impacts due to this rule can be characterized as "Affordable" or "Moderate." EPA estimates that a total of 172 small businesses (3 percent of all small business CAFOs with more than 300 AU) would experience financial stress and might be vulnerable to closure. By sector, these closures are comprised of about 131 small businesses in the beef sector, 38 businesses in the heifer sector, and 3 businesses in the broiler sector. Most of these (nearly 90 percent) are operations with fewer than 1,000 AU. For Option 2, the analysis indicates that, among all small business CAFOs in the veal, dairy, hog, turkey, and egg-laying sectors, the impacts due to this rule can be characterized as "Affordable" or "Moderate." EPA estimates that a total of 262 small businesses (4 percent of all small business CAFOs with more than 300 AU) would experience financial stress and might be vulnerable to closure. By sector, these closures are comprised of about 183 small businesses in the beef sector, 50 businesses in the heifer sector, and 19 businesses in the broiler sector. Nearly 90 percent of these potential closures are operations with fewer than 1,000 AU.

These estimates of the number of potential CAFO closures are cumulative and reflect the results of both the farm level analysis and the enterprise level analysis. These results are based on an analysis that does not consider the longer term effects on market adjustment and also available cost-share assistance from Federal and State farm conservation programs. EPA believes that such adjustments could lessen the economic impacts of the final regulations over time.

Table 4-5 shows the results of this analysis aggregated across all estimated designated operations with less than 1,000 AU, indicating that nearly one-half of all designated operations may go out of business. Closures among designated operations are all in the broiler, beef, and heifer sectors.

Table 4-5a. Results of EPA's Small Business Analysis (Option 1)

Sector	Number of Small CAFOs	Affordable	Moderate	Stress	Affordable	Moderate	Stress
		(Number)			(Percent of Total Operations)		
CAFOs >1000 AU (excl. designated operations)							
Fed Cattle	712	581	0	131	82%	0%	18%
Veal	12	12	0	0	100%	0%	0%
Heifer	327	289	0	38	88%	0%	12%
Dairy	1,330	1,330	0	0	100%	0%	0%
Hogs	1,485	1,485	0	0	100%	0%	0%
Broilers	1,823	1,395	424	3	77%	23%	0%
Layers: Dry	24	24	0	0	100%	0%	0%
Layers: Wet	407	407	0	0	100%	0%	0%
Turkeys	31	31	0	0	100%	0%	0%
Total	6,151	5,554	424	172	90%	7%	3%
CAFOs >1,000 AU							
Fed Cattle	538	533	0	5	99%	0%	1%
Veal	5	5	0	0	100%	0%	0%
Heifer	97	97	0	0	100%	0%	0%
Dairy	0	--	--	--	--	--	--
Hogs	0	--	--	--	--	--	--
Broilers	1,303	1,065	234	3	82%	18%	0%
Layers: Dry	0	--	--	--	--	--	--
Layers: Wet	383	383	0	0	100%	0%	0%
Turkeys	0	--	--	--	--	--	--
Total	2,326	2,083	234	8	90%	10%	0%

Table 4-5a. Results of EPA's Small Business Analysis (Option 1)

Sector	Number of Small CAFOs	Affordable	Moderate	Stress	Affordable	Moderate	Stress
		(Number)			(Percent of Total Operations)		
Operations 300 - 1,000 AU (Defined as CAFOs)							
Fed Cattle	174	48	0	126	27%	0%	73%*
Veal	7	7	0	0	100%	0%	0%
Heifer	230	192	0	38	83%	0%	17%
Dairy	1,330	1,330	0	0	100%	0%	0%
Hogs	1,485	1,485	0	0	100%	0%	0%
Broilers	520	330	190	0	63%	37%	0%
Layers: Dry	24	24	0	0	100%	0%	0%
Layers: Wet	24	24	0	0	100%	0%	0%
Turkeys	31	31	0	0	100%	0%	0%
Total	3,825	3,471	190	164	91%	5%	4%
Operations <1,000 AU (Designated as CAFOs)							
Fed Cattle	30	4	0	26	13%	0%	87%
Veal	0	0	0	0	NA	NA	NA
Heifer	6	0	0	6	0%	0%	100%
Dairy	60	60	0	0	100%	0%	0%
Hogs	104	104	0	0	100%	0%	0%
Broilers	104	0	0	104	0%	0%	100%
Layers: Dry	4	4	0	0	100%	0%	0%
Layers: Wet	16	16	0	0	100%	0%	0%
Turkeys	20	20	0	0	100%	0%	0%
Total	344	208	0	136	61%	0%	40%

Source: USEPA. May not add due to rounding. Does not include the number of CAFOs includes designated facilities. Assumes that the costs that will be incurred by those sized operations to comply with BPJ-based limitations under the revised NPDES regulations are similar to the estimated costs that would be incurred if Medium CAFOs had to comply with the ELG.

"Layers: dry" are operations with dry manure systems. "Layers: wet" are operations with liquid manure systems.

Table 4-5b. Results of EPA's Small Business Analysis (Option 2)

Sector	Number of Small CAFOs	Affordable	Moderate	Stress	Affordable	Moderate	Stress
		(Number)			(Percent of Total Operations)		
CAFOs >1000 AU (excl. designated operations)							
Fed Cattle	712	529	0	183	74%	0%	26%
Veal	12	12	0	0	100%	0%	0%
Heifer	327	277	0	50	85%	0%	15%
Dairy	1,330	1,306	24	0	98%	2%	0%
Hogs	1,485	1,483	2	0	100%	0%	0%
Broilers	1,823	1,026	780	19	56%	43%	1%
Layers: Dry	24	24	0	0	100%	0%	0%
Layers: Wet	407	407	0	0	100%	0%	0%
Turkeys	31	31	0	0	100%	0%	0%
Total	6,151	5,129	806	262	83%	13%	4%
CAFOs >1,000 AU							
Fed Cattle	538	522	0	16	97%	0%	3%
Veal	5	5	0	0	100%	0%	0%
Heifer	97	88	0	9	91%	0%	9%
Dairy	0	--	--	--	--	--	--
Hogs	0	--	--	--	--	--	--
Broilers	1,303	763	532	9	58%	41%	1%
Layers: Dry	0	--	--	--	--	--	--
Layers: Wet	383	383	0	0	100%	0%	0%
Turkeys	0	--	--	--	--	--	--
Total	2,326	1,795	532	34	76%	23%	1%

Table 4-5b. Results of EPA's Small Business Analysis (Option 2)

Sector	Number of Small CAFOs	Affordable	Moderate	Stress	Affordable	Moderate	Stress
		(Number)			(Percent of Total Operations)		
Operations 300 - 1,000 AU (Defined as CAFOs)							
Fed Cattle	174	7	0	167	4%	0%	96%
Veal	7	7	0	0	100%	0%	0%
Heifer	230	189	0	41	82%	0%	18%
Dairy	1,330	1,306	24	0	98%	2%	0%
Hogs	1,485	1,483	2	0	100%	0%	0%
Broilers	520	263	248	10	51%	48%	1%
Layers: Dry	24	24	0	0	100%	0%	0%
Layers: Wet	24	24	0	0	100%	0%	0%
Turkeys	31	31	0	0	100%	0%	0%
Total	3,825	3,334	274	228	87%	7%	6%
Operations <1,000 AU (Designated as CAFOs)							
Fed Cattle	30	4	0	26	13%	0%	87%
Veal	0	0	0	0	NA	NA	NA
Heifer	6	0	0	6	0%	0%	100%
Dairy	60	60	0	0	100%	0%	0%
Hogs	104	104	0	0	100%	0%	0%
Broilers	104	0	0	104	0%	0%	100%
Layers: Dry	4	4	0	0	100%	0%	0%
Layers: Wet	16	16	0	0	100%	0%	0%
Turkeys	20	20	0	0	100%	0%	0%
Total	344	208	0	136	61%	0%	40%

Source: USEPA. May not add due to rounding. Does not include the number of CAFOs includes designated facilities. Assumes that the costs that will be incurred by those sized operations to comply with BPJ-based limitations under the revised NPDES regulations are similar to the estimated costs that would be incurred if Medium CAFOs had to comply with the ELG.

"Layers: dry" are operations with dry manure systems. "Layers: wet" are operations with liquid manure systems.

EPA believes that the estimated financial impacts shown in Tables 4-5(a) and 4.5(b) represent the worst case. The reasons are summarized below.

First, all results are estimated assuming no costs can be passed through between CAFOs and the processing sectors. As discussed in Section 3 of this report, if modest levels of cost passthrough are assumed in the broiler sectors, the BAT requirements are affordable to all small broiler operations. EPA did not evaluate economic impacts on cattle operations under a cost passthrough scenario; however, it is expected that long-run market and structural adjustment by producers in this sector will diminish the estimated impacts. Even without an assumption of cost passthrough, EPA's analysis shows that adverse impacts will not be experienced by a substantial number of operations, as compared to the number of affected operations in these sectors. EPA has conducted an extensive literature review of issues concerning cost passthrough. Based on the results of the available empirical research on market power and price transmission in these industries, EPA believes that there is little evidence to support the position that increased production costs may not be passed through the market levels. A summary of this literature review is provided in the rulemaking record (ERG, 2000c — DCN 70640).

Second, as noted in the SBAR Panel Report, EPA believes that the number of small broiler operations is overestimated. In the absence of business level revenue data, EPA estimates the number of "small businesses" using the approach described in Section 4.2. Using this approach, virtually all (>99.9 percent) broiler operations are considered "small" businesses. This categorization may not accurately portray actual small operations in this sector because it classifies a 15- to 20-house broiler operation with 375,000 birds as a small business. Information from industry sources suggests that a two-house broiler operation with roughly 50,000 birds is more appropriately characterized as a small business in this sector (Madison, 1999; Staples, 1998). Therefore, it is likely that the number of small broiler operations might include a number of medium and large size broiler operations being considered small entities. As discussed in Section 9.2.1 of the Proposal EA, EPA consulted with SBA on the use of an alternative definition for small businesses in all affected sectors based on animal inventory at an operation during the development of the rulemaking.

Third, EPA believes that a costs-to-sales comparison is a crude measure of impacts on small business in sectors where production contracting is commonly used, such as in the broiler sector (and also in the turkey, egg, and hog sectors, though to a lesser extent). As discussed in Section 4.2.4.5 of the Proposal EA, lower reported operating revenues in the broiler sector reflect the predominance of contract growers in this sector. Contract growers receive a prenegotiated contract price that is lower than the USDA-reported producer price, thus resulting in lower gross revenues at these operations (USDA/ERS, 1996b; Perry et al., 1999; Farm Journal, 1998). Lower producer prices among contract growers are often offset by lower overall production costs at these operations, because the affiliated processor firm pays for a substantial portion of the grower's annual variable cash expenses. Inputs supplied by the integrator may include feeder pigs or chicks, feed, veterinary services and medicines, technical support, and transportation of animals (USDA, 1996a). These variable cash costs compose a large component of annual operating costs, averaging more than 70 percent of total variable and fixed costs at livestock and poultry operations (USDA/ERS, 1999a). The contract grower also faces reduced risk because the integrator guarantees the grower a fixed output price (see Section 2 of the Proposal EA for more information on contracting in animal agriculture). Because production costs at a contract grower operation are lower than that at an independently owned operation, a profit test (costs-to-profit comparison) is a more accurate measure of impacts at grower operations. However, financial data are not available that differentiate between contract grower and independent operations.

Fourth, EPA's initial regulatory flexibility analysis also does not consider a range of potential cost offsets available to most farms. As discussed in Section 2.4 of this report, one source of potential cost offset is cost share and technical assistance available to farmers for on-farm improvements from various State and Federal programs, such as the Environmental Quality Incentives Program (EQIP) administered by USDA. Cost sharing for eligible producers under EQIP may cover up to 75 percent of the costs of certain conservation practices, such as grassed waterways, filter strips, manure management facilities, capping of abandoned wells, and other practices important to improving and maintaining the health of natural resources in the area. Technical assistance is also available for formulating conservation plans. In the Spring of 2002, new Farm Bill legislation passed by Congress might significantly raise government expenditures for this program. Total EQIP authorization for FY 2002 to FY 2007 is \$5.8 billion, ranging from \$400 million to \$1.3 billion per year over the period. This compares to current authorized levels of about \$200 million per year. The new legislation targets 60 percent of available EQIP funds to livestock and poultry producers, including confinement and grass-based systems. The new legislation also removed the previous EQIP eligibility requirements that restricted funding for certain structural practices to operations with fewer than 1,000 AU (as measured by USDA), replacing this restriction with an overall payment limitation of \$450,000 per producer over the authorized life of the 2002 Farm Bill. Many other State and Federal cost share programs base eligibility not on size thresholds but on priority watersheds (e.g., USDA's Small Watershed Program; the New York City Watershed Program), priority contaminants (e.g., Kansas Non-Point Source Pollution Control Fund), or proposed waste management practices (e.g., Maryland, Minnesota, Missouri, Nebraska, and North Carolina state programs). However, technical assistance under most programs is available to all operations, regardless of watershed, contaminants, proposed practices, or size (ERG, 2000a). A review of cost-share and technical assistance programs available to AFOs is provided in the rulemaking record (ERG, 2000a—DCN 70130).

Section 2.4 also describes another source of potential cost offset, which is manure sales, particularly of relatively higher value dry poultry litter. EPA estimates that sales of dry poultry litter could offset the costs of meeting the regulatory requirements on the order of more than 50 percent. As illustrated in the Proposal EA, this reduction alone exceeds the level of cost passthrough (42 percent) assumed at proposal for the cost impact analysis of the broiler sector. Details on how EPA calculated these manure sale offsets and how they would reduce the economic impacts at poultry operations are presented in Section 6 of the Proposal EA.

Finally, this analysis does not take into account certain noneconomic factors that might influence an operation's decision to weather the boom and bust cycles that are commonplace in agricultural markets. Farm typology data from USDA indicate that a large share of farming operations (more than 90 percent) have annual sales of less than \$250,000 and are considered "small family farms" by USDA (USDA/ERS, 2000d, 2000e). Of these, about 60 percent are "limited-resource," "retirement," or "residential" operations where farming is not the primary source of income (USDA/ERS, 2000e, 1999a). In many cases, these operations have negative annual income supplemented by sources of off-farm income that subsidize the farming operation (USDA/ERS, 2000d and 1996a).

USDA's ERS (1996a) reports that about 60 percent of farm operators reporting negative net income had nonfarm occupations. About 75 to 80 percent of farms rely on some nonfarm income, and even in the largest operations nonfarm income can be a significant portion of total household income (USDA/ERS, 1996a). More than 90 percent of farm operators with negative net income had nonfarm income averaging more than \$35,700 per year; even farms with positive net income rely somewhat on nonfarm income (Heimlich and Barnard, 1995; USDA/ERS, 1996a).

When farm income is negative over a period of time, sales tests can be very difficult to interpret (Heimlich and Barnard, 1995). One reason that incomes can remain negative over several years is that operators can supplement farm income with nonfarm income, and these losses can be used to reduce total income tax liabilities while the real estate value of the farm property appreciates. Additional noneconomic factors might also include the satisfaction of working for oneself, the ability to employ family members, a sense of tradition and the ability to pass on that tradition to future generations, and the fact that the operation is both a home and a livelihood. These and other noneconomic factors might influence the decision to close a livestock or poultry operation cannot be adequately addressed in an economic model. To the extent that these factors play a role in that decision, EPA's economic model might overstate the possibility of closure among small businesses.

USDA's farm financial data include operations where farming is part-time and not the primary occupation, but exclude sources of nonfarm income at these operations. As noted in Section 4.2 of the Proposal EA, the inclusion of these operations may result in lower average data values than would be the case if these operations were excluded from the analysis. EPA believes that including of these operations might tend to overstate impacts. Previous analyses by USDA and EPA have also noted the potential effect on average farm data of including these operations and have regarded these part-time business more as "hobbies or recreational activities" (Heimlich and Barnard, 1995; DPRA, 1995). Heimlich and Barnard (1995) further indicate that considering non-farm income in addition to farm income may provide a more appropriate comparison to the costs of required measures where the motivation for staying in business is not necessarily purely economic.

Overall, EPA expects that the CAFO regulations will benefit the smallest businesses in these sectors, because the regulations might create a comparative advantage for small operations that are not subject to the regulations. Except for the few AFOs that are designated as CAFOs, these small operations will not incur costs associated with the final requirements and may benefit from eventual higher producer prices as these markets adjust to higher production costs in the long term.

APPENDIX N ABBREVIATIONS USED IN THIS GUIDE

Advocacy	Office of Advocacy, U.S. Small Business Administration
AFO	animal feeding operation
ANPRM	advance notice of proposed rulemaking
APHIS	Animal and Plant Health Inspection Service
APA	Administrative Procedure Act
ATA	American Trucking Association
AU	animal unit
BLM	Bureau of Land Management
CAFO	concentrated animal feeding operation
CFPB	Consumer Financial Protection Bureau
C.F.R.	Code of Federal Regulations
CMS	Centers for Medicare and Medicaid Services
DOC	Department of Commerce
DOD	Department of Defense
DOT	Department of Transportation
DHS	Department of Homeland Security
E.O.	Executive Order
EPA	Environmental Protection Agency
EQIP	Environmental Quality Assessment Program
ERS	Economic Research Service (USDA)
FAA	Federal Aviation Administration
FAR	Federal Acquisition Regulation
FCC	Federal Communications Commission
FERC	Federal Energy Regulatory Commission
FR	Federal Register
FRA	Federal Railroad Administration
FRFA	final regulatory flexibility analysis
GAO	General Accounting Office, now Government Accountability Office
HHS	U.S. Department of Health and Human Services
ICR	information collection request
IPS	Interim Payment System
IRFA	initial regulatory flexibility analysis
IRS	Internal Revenue Service
MMS	Minerals Management Service
NAAQS	national ambient air quality standard
NAICS	North American Industry Classification System
NAIHP	National Association of Independent Housing Professionals
NAMB	National Association of Mortgage Brokers
NEPA	National Environmental Policy Act
NMFS	National Marine Fisheries Service
NOAA	National Oceanic and Atmospheric Administration
NPDES	National Pollutant Discharge Elimination System
NPRM	notice of proposed rulemaking
NWMA	Northwest Mining Association

OIRA	Office of Information and Regulatory Affairs
OMB	Office of Management and Budget
OSHA	Occupational Safety and Health Administration
P.L.	Public Law
RCRA	Resource Conservation and Recovery Act
RFA	Regulatory Flexibility Act
RIA	regulatory impact analysis
RSPA	Research and Special Programs Administration
RUS	Rural Utilities Service
SBA	Small Business Administration
SBIC	small business investment company
SBJA	Small Business Jobs Act
SBREFA	Small Business Regulatory Enforcement Fairness Act
SERS	small entity representatives
SIC	Standard Industrial Classification system
SMR	specialized mobile radio
USDA	U.S. Department of Agriculture
U.S.C.	United States Code
UPL	upper payment limit
VOCs	volatile organic compounds

APPENDIX O OFFICE OF ADVOCACY STAFF

UPDATED AUGUST 2017

Office of the Chief Counsel

Major Clark, III	Acting Chief Counsel	202-205-6533	Major.Clark@sba.gov
Claudia Rodgers	Senior Counsel	202-205-6804	Claudia.Rodgers@sba.gov

Office of Interagency Affairs (Legal Staff)

Charles Maresca	Director	202-205-6978	Charles.Maresca@sba.gov
Kevin Bromberg	Environment	202-205-6964	Kevin.Bromberg@sba.gov
Major Clark, III	Procurement, Small Business	202-205-7150	Major.Clark@sba.gov
Bruce Lundegren	Safety, Transportation, Security	202-205-6144	Bruce.Lundegren@sba.gov
Linwood Rayford	Food, Drugs and Health	202-401-6880	Linwood.Rayford@sba.gov
Janis Reyes	Labor and Immigration	202-619-0312	Janis.Reyes@sba.gov
Zvi Rosen	Intellectual Property	202-205-6948	Zvi.Rosen@sba.gov
David Rostker	Environment, Regulatory Reform	202-205-6966	David.Rostker@sba.gov
Jamie Saloom	Telecom, Technology, and Intellectual Property	202-205-6890	Jamie.Saloom@sba.gov
Prianka Sharma	Agriculture, Energy, Natural Resources	202-205-6938	Prianka.Sharma@sba.gov
Jennifer Smith	Economic Regulation	202-205-6943	Jennifer.Smith@sba.gov
Joseph Sobota	Veterans	202-205-6952	Joseph.Sobota@sba.gov
Rosalyn Steward	International Regulation, Energy	202-205-7013	Rosalyn.Steward@sba.gov
Dillon Taylor	Tax, Securities, Pensions	202-401-9798	Dillon.Taylor@sba.gov
Tayyaba Waqar Zeb	Environment	202-205-6790	Tayyaba.Waqar@sba.gov

Office of Economic Research

Patrick Delehanty	Acting Director, Regulatory Economist	202-205-7123	Patrick.Delehanty@sba.gov
Lindsay Abate	Regulatory Economist	202-205-6537	Lindsay.Abate@sba.gov
Elizabeth Glass	Research Fellow	202-205-9512	Elizabeth.Glass@sba.gov
Brian Headd	Economist	202-205-6952	Brian.Headd@sba.gov
Michael McManus	Regulatory Economist	202-205-7168	Michael.McManus@sba.gov
Jonathan Porat	Regulatory Economist	202-205-6973	Jonathan.Porat@sba.gov
Richard Schwinn	Research Economist	202-205-6958	Richard.Schwinn@sba.gov
Victoria Williams	Economist	202-205-6191	Victoria.Williams@sba.gov
Daniel Wilmoth	Research Economist	202-205-6878	Daniel.Wilmoth@sba.gov

Office of Information

Jason Doré	Assistant Chief Counsel for External Affairs	202-205-6934	Jason.Dore@sba.gov
Victoria Carlborg	Writer/Editor	202-205-7990	Victoria.Carlborg@sba.gov
Sarah Coleman	Graphic Designer	202-205-6835	Sarah.Coleman@sba.gov
Rebecca Krafft	Senior Editor	202-205-6224	Rebecca.Krafft@sba.gov
Emily Theroux	Public Affairs Assistant	202-205-6936	Emily.Theroux@sba.gov
Emily M. Williams	Outreach and Events Specialist	202-205-6949	Emily.Williams@sba.gov

Administrative Support Branch

Luciette Wren	Director	202-205-7749	Luciette.Wren@sba.gov
Shawn Fouladi	Program Assistant	202-205-6947	Shawn.Fouladi@sba.gov
Nicholas Ivory	Program Assistant	202-205-9972	Nicholas.Ivory@sba.gov
Akira Loring	Program Assistant	202-619-2310	Akira.Loring@sba.gov
Kena Morgan-Nicholson	Program Assistant	202-205-6768	Kena.Morgan-Nicholson@sba.gov
Natalyn Tart	Administrative Officer	202-205-6181	Natalyn.Tart@sba.gov

Regional Advocates

vacant	Region 1 Advocate		
vacant	Region 2 Advocate		
vacant	Region 3 Advocate		
vacant	Region 4 Advocate		
Les Davis	Region 5 Advocate	614-469-6860	Leslie.Davis@sba.gov
vacant	Region 6 Advocate		
Nancy Zurbuchen	Region 7 Advocate	314-539-6615	Nancy.Zurbuchen@sba.gov
vacant	Region 8 Advocate		
Philip Lovas	Region 9 Advocate	602-745-7231	Philip.Lovas@sba.gov
R. Apollo Fuhriman	Region 10 Advocate	206-553-5231	AppolloF@sba.gov

INDEX

- Ace Lobster Co. v. Evans*, 71
- Administrative Procedure Act, 9, 10
 - and RFA litigation, 55ff
- adverse v. beneficial impact of regulations, 23
- Aeronautical Repair Station Association v. FAA*, 66
- Alenco Communications v. FCC*, 69
- Allied Local and Regional Manufacturers Caucus v. EPA*, 68
- alternatives to regulations, 37
- American Federation of Labor v. Chertoff*, 59
- American Moving and Storage Association, Inc. v. DOD*, 57
- American Trucking Associations v. EPA*, 23, 65
- AML International, Inc., v. Daley*, 71
- Ashley County Medical Center v. Thompson*, 73
- Associated Builders and Contractors, Inc. v. Herman*, 60
- Associated Fisheries of Maine, Inc., v. Daley*, 69, 72
- Atlantic Fish Spotters Association v. Evans*, 56n
- beneficial impact of regulations, 23
- Cactus Corner v. U.S.D.A.*, 60
- Cement Kiln Recycling Coalition v. EPA*, 66
- Certification, 11
 - adequate vs. inadequate, 24
 - checklist, 29
 - factual basis for, 13
 - litigation regarding, 58
- compliance assistance, 89
- compliance guides, 88
- Congressional Review Act, 91
- Consumer Financial Protection Bureau and Regulatory Flexibility Act, 2
 - and SBREFA panels, 51
- costs of regulatory compliance, 1
- cumulative effects of rules, 135
 - see also* Section 610 review
- data resources, 106, 109
- direct vs. indirect impact of regulations, 22, 64
- Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010, 2, 51, 103
- economic impact analysis, 74
- economic research on small business, 109
- Environmental Defense Center v. E.P.A.*, 60
- Environmental Protection Agency,
 - and Regulatory Flexibility Act, 2
 - and SBREFA panels, 51
- Executive Order 12,866, 4, 107
- Executive Order 13,272, 3, 119
- Executive Order 13,563, 4, 79, 121
- Executive Order 13,579, 5, 79, 130
- Executive Order 13,610, 132
- Executive Order 13,771, 134
- Executive Order 13,777, 137
- exemptions from RFA, 9, 56
- existing rules, review of, 79
- factual basis for certification, 13
- frequently asked questions about small business, 106
- final regulatory flexibility analysis, 43
 - adverse economic impact minimized, 47
 - delays in publication of, 48
 - example of, 171
 - issues to address, 44
 - litigation concerning, 69
 - small entities estimated, 46
- foreign affairs functions exemption from RFA, 56, 57
- Friends of Fiery Gizzard v. Farmers Home Administration*, 24
- “Good cause” exception, 57
- Hall v. Evans*, 72
- Harlan Land Co. v. United States Department of Agriculture*, 59
- indirect impact of regulations, 22
 - litigation concerning, 64
- Industry analysis, 16, 18
- initial regulatory flexibility analysis, 31
 - compliance requirements, 36
 - duplicative rules considered, 40
 - elements of, 33
 - example of, 139
 - issues to address, 32
 - litigation concerning, 68
 - objectives of rule, 35
 - and other analyses, 41
 - reasons for rule, 35
 - significant alternatives, 37
 - small entities affected, 36
 - threshold analysis, 11
 - waiver of, 41
- Internal Revenue Service rules and RFA, 10
- interpretative rules, 9, 10, 57
- Judicial review provision of RFA, 77
- LaGloria Oil & Gas Co. v. United States*, 56n

legislative rules and the RFA, 9
litigation and the RFA, 55
Little Bay Lobster Co. v. Evans, 72

Michigan v. EPA, 65
Mid-Tex Electric Cooperative, Inc. v. FERC, 22, 64, 65, 67, 75n
military functions exemption, 9, 56, 57
Motor and Equipment Manufacturers Association v. Nichols, 65

National Association for Home Care v. Shalala, 57
National Association of Home Builders v. Army Corps of Engineers, 55n
National Association of Mortgage Brokers v. Board of Governors of the Federal Reserve System, 69, 77n
National Association of Psychiatric Health Systems v. Shalala, 73
National Environmental Policy Act, 2, 7
National Mining Association v. Chao, 76n
National Propane Gas Association v. DOT, 70
National Truck Equipment Association v. NHTSA, 63n
National Women, Infants & Children Grocers Association v. Food and Nutrition Service, 66n
Navajo Refining Co. v. United States, 56n
North Carolina Fisheries Association v. Daley, 58, 63, 75, 76n
Northwest Mining Association v. Babbitt, 58, 61, 63n, 64

Occupational Safety and Health Administration and Regulatory Flexibility Act, 2 and SBREFA panels, 51
Oregon Trollers Association v. Gutierrez, 57

penalty reduction policies, 90

rates and the RFA, 10, 57
regulatory agendas, 87
regulatory enforcement fairness, 90
Regulatory Flexibility Act, 93

- additional requirements of, 87
- adverse v. beneficial impact, 23
- analysis under, 16
- applicability of, 9
- certification of rules under, 11, 29
- direct vs. indirect impact, 22
- exemptions, 9, 56
- final regulatory flexibility analysis, 43
- history of, 1
- impact assessment under, 15
- initial regulatory flexibility analysis, 31
- judicial review provision, 77
- litigation concerning, 55
- purpose of, 7
- recent developments concerning, 102
- retrospective review, 79
- SBREFA panels, 51
- Section 610 review, 79
- semi-annual regulatory agendas, 87
- significant impact determination, 18
- substantial number determination, 21
- threshold analysis under, 11, 12

Roche v. Evans, 55n

SBREFA panels, 51

- timeline for, 54

Section 610 review of existing rules, 79

- see also* cumulative effects of rules

semi-annual regulatory agendas, 87
Significant impact, 18
Size standards, 15, 17

- litigation concerning, 61

Small Business and Work Opportunity Act of 2007, 2, 102
Small business definition, 14
Small Business in Telecommunications v. FCC, 62
Small Business Jobs Act of 2010, 3, 103
Small Business Regulatory Enforcement Fairness Act

- background and provisions of, 2
- and interpretative rules, 10
- SBREFA panels, 51

Small entity definition, 14
Small governmental jurisdiction definition, 15
Small organization definition, 14
Southern Offshore Fishing Association v. Daley, 25, 62, 73
State v. Centers for Medicare & Medicaid Services, 56n
Statistics for regulatory analysis, 109
Substantial number, 21

Theiss v. Principi, 64

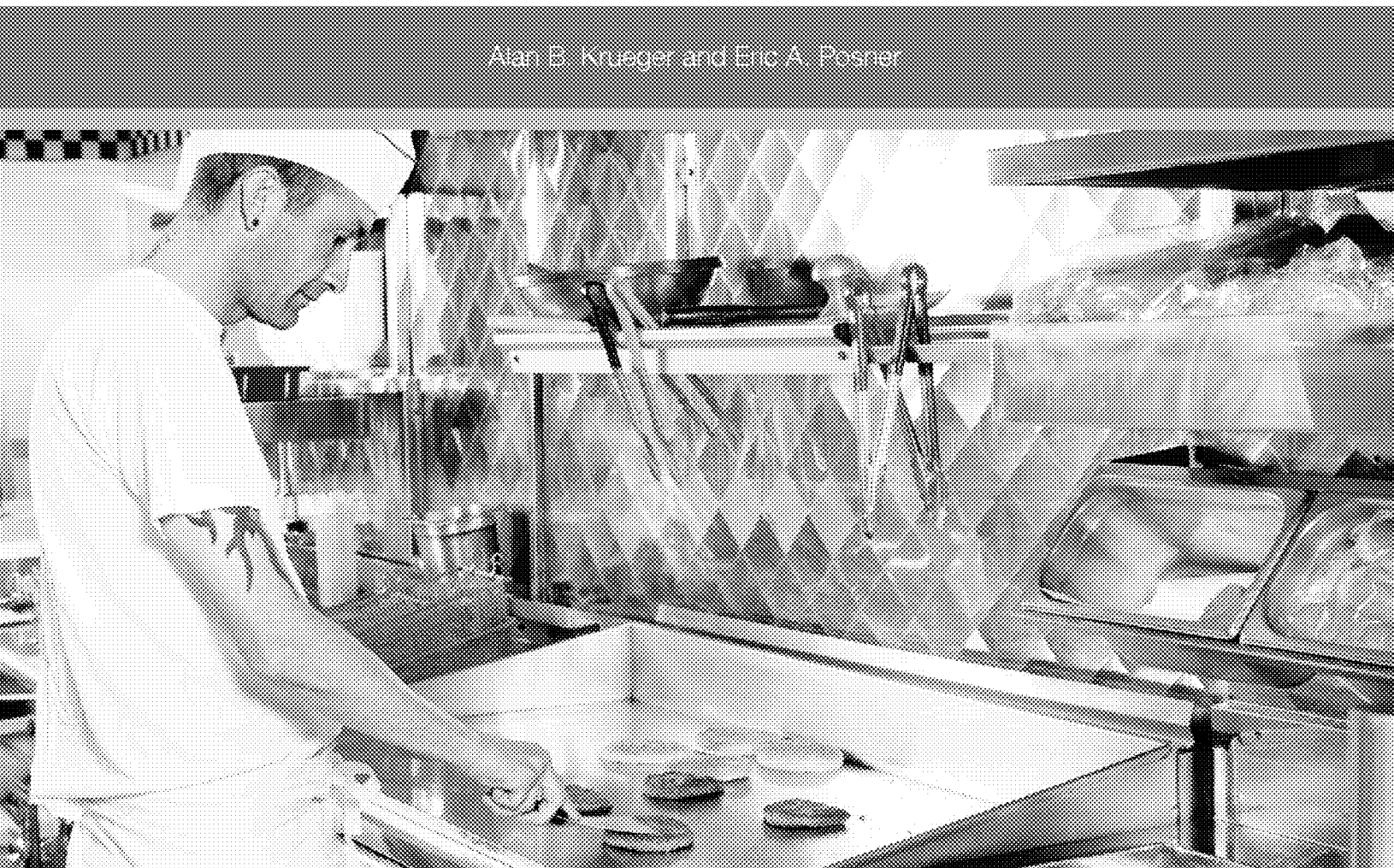
U.S. Cellular Corporation v. FCC, 68
Unfunded Mandates Act, 91
United Distribution Companies v. FERC, 65

Washington v. Daley, 75n
White Eagle Cooperative Association v. Conner, 66
Whitman v. American Trucking Association, 9
Williams Alaska Petroleum v. United States, 56n

Updated August 2017

A Proposal for Protecting Low-Income Workers from Monopsony and Collusion

Alan B. Krueger and Eric A. Posner



MISSION STATEMENT

The Hamilton Project seeks to advance America's promise of opportunity, prosperity, and growth.

We believe that today's increasingly competitive global economy demands public policy that commensurate with the challenges of the 21st Century. The Project's economic strategy reflects a conviction that long-term prosperity is best achieved by fostering economic growth and broad participation in that growth, by enhancing individual economic security, and by embracing a role for effective government in making needed public investments.

Our strategy calls for continuing public investment, a secure social safety net, and fiscal discipline. In that framework, the Project puts forward innovative proposals from leading economic thinkers — based on credible evidence and experience, not ideology or doctrine — to introduce new and effective policy options into the national debate.

The Project is named after Alexander Hamilton, the nation's first Treasury Secretary, who laid the foundation for the modern American economy. Hamilton stood for sound fiscal policy, believed that broad-based opportunity for advancement would drive American economic growth, and recognized that " prudent aids and encouragements on the part of government" are necessary to enhance and guide market forces. The guiding principles of the Project remain consistent with these views.





A Proposal for Protecting Low-Income Workers from Monopsony and Collusion

Alan B. Krueger
Princeton University

Eric A. Posner
The University of Chicago Law School

FEBRUARY 2018

This policy proposal is a proposal from the author(s). As emphasized in The Hamilton Project's original strategy paper, the Project was designed in part to provide a forum for leading thinkers across the nation to put forward innovative and potentially important economic policy ideas that share the Project's broad goals of promoting economic growth, broad-based participation in growth, and economic security. The author(s) are invited to express their own ideas in policy papers, whether or not the Project's staff or advisory council agrees with the specific proposals. This policy paper is offered in that spirit.

BROOKINGS

Abstract

New evidence that labor markets are being rendered uncompetitive by large employers suggests that the time has come to strengthen legal protections for workers. Labor market collusion or monopsonization—the exercise of employer market power in labor markets—may contribute to wage stagnation, rising inequality, and declining productivity in the American economy, trends which have hit low-income workers especially hard. To address these problems, we propose three reforms. First, the federal government should enhance scrutiny of mergers for adverse labor market effects. Second, state governments should ban non-compete covenants that bind low-wage workers. Third, no-poaching arrangements among establishments that belong to a single franchise company should be prohibited.

Table of Contents

ABSTRACT	2
INTRODUCTION	4
THE CHALLENGE	6
A NEW APPROACH	12
QUESTIONS AND CONCERNS	14
CONCLUSION	15
AUTHORS AND ACKNOWLEDGEMENTS	16
ENDNOTES	17
REFERENCES	18

Introduction

In recent decades, rising income inequality and stagnating wages among all but the highest-paid workers have raised alarms about the health of the U.S. labor market and its capacity to provide workers with the means to adequately support themselves. Alongside the familiar explanations, including automation and foreign competition, a new and perhaps surprising one has emerged: monopsonization of, or collusion in, labor markets. As firms have grown in size, they have become capable of dominating local labor markets—a phenomenon referred to as monopsonization—and of using their market power to suppress wages.¹ There is also evidence that some firms have colluded, entering into no-poaching and similar arrangements that restrict workers’ choices among employers. Various impediments to perfect competition, including reluctance among many workers to relocate to change jobs, have added to this problem.

The problem has been serious enough to draw the attention of the U.S. government. In 2016 the White House and the Department of Treasury issued reports critical of non-compete agreements (U.S. Department of the Treasury 2016; White House 2016). In the same year, the Department of Justice (DOJ) and the Federal Trade Commission (FTC) together issued a guidance document advising human resource professionals that it is illegal under the antitrust laws for rival firms to agree not to hire each other’s workers or to compete on wages (DOJ and FTC 2016). DOJ has brought lawsuits against firms that have allegedly engaged in such arrangements, including a hospital association in Arizona, and technology companies, including Apple and Google. The FTC has brought cases against firms that tried to collude in the labor market for nurses and fashion models (FTC 1995).² In 2017 DOJ noted that it was conducting several investigations of labor market collusion that might lead to criminal prosecutions (Nylen 2017).

But given the scale of the problem and burdens of litigation, ad hoc legal interventions based on existing antitrust law will not be enough to solve it. To prevail in litigation, plaintiffs must offer proof about complex economic phenomena, such as the scope of markets and the relationship between wages and market power, which can be difficult to evaluate. Furthermore, antitrust authorities have limited resources. For these reasons, new approaches are needed for protecting workers from wage suppression and similar anticompetitive behavior.

We focus on three types of business behavior that have contributed to the current problems in the labor market. First, a combination of several decades of mergers and growth in industries where network effects tilt toward one dominant firm have created massive employers who apparently enjoy market power in various labor markets (Autor et al. 2017). While it is illegal for firms to merge for the purpose of dominating a labor market, the government does not focus on labor market effects when it screens mergers under the Horizontal Merger Guidelines (DOJ and FTC 2010). We propose a beefed-up screening procedure that alerts regulators of the risk that a merger will create anticompetitive effects in labor markets.

Second, it has recently become clear that firms use non-compete agreements to suppress labor market competition among low-wage workers. In a non-compete agreement (also called a covenant not to compete), the worker agrees that he or she will not work for competing employers for a period of time after termination. In principle, a non-compete agreement could violate antitrust law if it is used to enhance or exploit market power, but non-compete agreements are almost never the subject of antitrust litigation.

There are limits to the enforceability of non-compete agreements in the common law. If a non-compete agreement is not “reasonable” in the light of legitimate business goals—such as recovering the cost of training or preventing the disclosure of trade secrets—then a court will refuse to enforce it.³ The practical effect of this rule is that if a worker knows his or her legal rights, or can afford a lawyer to explain them and defend him or her in court, then the non-compete agreement may not be harmful, and could enhance efficiency.³ For example, the risk of turnover can result in insufficient investment in firm-specific training. But with non-competes a worker and firm can jointly reach a bargain in which the firm pays the cost of industry-specific training and shares some of the return from that investment in exchange for the worker agreeing to refrain from moving to another firm in the industry. The problem is that, typically, only high-level executives and professionals can afford a lawyer to review such agreements and ensure that the worker’s interests are fully represented. And even in these cases, there is a concern that in “thin” labor markets for critical talent, an employer can use non-compete agreements to bind workers and discourage competitors from entering the market because

they will face a scarcity of available labor. Many employers use non-competes for low-wage jobs (Starr, Prescott, and Bishara 2017), where workers do not know their rights, cannot afford lawyers, receive little training, and are susceptible to threats from their former employers. Accordingly, we propose that non-compete agreements involving low-wage workers be banned or heavily restricted. A handful of states have recently been considering such actions.

Third, new evidence suggests that franchise companies have used no-poaching agreements to suppress labor market competition. In a no-poaching agreement, two or more employers agree that they will not hire each other's employees. When these agreements are made between independent companies, they clearly run afoul of the antitrust laws, as

DOJ and FTC guidance makes clear. However, in recent years no-poaching agreements have increasingly been included in franchisors' contracts with their franchisees, where antitrust law is harder to enforce. When a franchisor requires the different franchisees within its chain not to poach each other's workers, a claim can be made that the antitrust laws do not apply because the rules are internal to a single organization, while antitrust laws apply to the relationships among independent firms. However, if more than one franchisee exists in a single labor market, and those franchisees are collectively a dominant employer in that labor market, the no-poaching agreement is anticompetitive, and will tend to suppress the wages of workers. We argue that no-poaching agreements in franchises should be banned.

The Challenge

THE ECONOMICS OF LABOR MARKET MONOPSONIZATION AND COLLUSION

Under perfect competition, workers are paid the value of their contribution to output. A perfectly competitive labor market requires that workers can move freely to seek the most desirable opportunities for which they are qualified, and that neither employers nor employees have the ability to set pay. If employers have market power, however, they can pay workers less than the value of their contribution to output. The Joan Robinson (1969) variant of monopsony occurs when there is a single employer in a labor market. In this situation, the employer faces the market supply curve for labor, and must pay a higher wage to hire additional labor. The profit-maximizing decision for such a monopsonist is to hire less than the quantity of labor that would be hired under perfect competition, and pay workers below the value of marginal product of the last worker hired. A monopsonist makes do with unfilled jobs, which typically appear as vacancies; it is unable to find workers at the low wages it offers and unwilling to raise pay to attract more workers.

Burdett and Mortenson (1998), Manning (2003), and others show that a similar situation arises even if there are many small employers competing for labor in an otherwise competitive market, to the extent that labor market frictions—for example, from turnover and recruitment costs—cause employers to face a rising cost of labor.

These forms of monopsony power arise by natural forces, and are not a legal cause of action, much as a firm that achieves monopoly pricing power in the product market because of scale economies is not in violation of antitrust laws. Historically, labor unions played a greater role in counterbalancing such monopsony power, but with only 7 percent of private sector workers unionized, unions play a much smaller role today.

Employers can exert monopsony power through deliberate means, however, by restricting competition for labor or by colluding with other employers to suppress pay or benefits below the competitive level. These cases are of much greater concern for the law. The notion that employers have an interest in manipulating the labor market and restricting competition is hardly new. In *The Wealth of Nations*, for example, Adam Smith (1776, 81) observed, “[Employers] are always and everywhere in

a sort of tacit, but constant and uniform combination, not to raise the wages of labour above their actual rate.” If employers act in concert to suppress wages below the prevailing level, then they jointly act as a monopsonist, which reduces pay and employment for workers. Likewise, if employers restrict their employees’ outside options by pressuring or deceiving them to sign non-compete clauses, they can reduce worker mobility and suppress wages below the competitive level. If a labor market is already concentrated, non-compete agreements between incumbent firms and workers may deter new firms from entering the market and bidding up wages by depriving those firms of a ready source of labor. And agreements among employers to not hire or recruit from other employers—so-called no-poaching agreements—are a form of collusive behavior that restricts competition and suppresses pay and employment opportunities.

EVIDENCE

Collusion and Monopsonization in the Labor Market

Until recently economists assumed that labor markets are fairly competitive. The company towns of the past are long gone, and the vast majority of workers live in urban areas where employers are plentiful. But recent events—including agreements among technology companies not to poach engineers and among hospitals not to poach nurses—have led many economists and government officials to question this assumption (Council of Economic Advisers [CEA] 2016). Of course, such cases are hardly new, but legal scrutiny of them remains relatively rare. We have found fewer than two dozen cases since 2000 where courts have considered allegations of improper use of labor market monopsony power or collusion, most of them involving specialized settings such as sports leagues.⁵

However, the most powerful evidence for increased monopsony power relates to broad changes in the labor market. CEA (2016) provides a thorough summary of evidence regarding monopsony power in the labor market. Among the evidence that CEA cites are these: (1) Firm concentration has increased in recent years. (2) Labor market dynamism and geographic mobility have trended down in recent decades, enabling noncompetitive wage differentials to persist with less external pressure from worker mobility. (3) Other forces that tend to

counteract monopsony power and collusion are weaker than has historically been the case in the United States, due to the decline in the real value of the minimum wage and the decline in the fraction of workers represented by labor unions. (4) And, in the current recovery, wage growth has not been stronger in industries that have experienced greater job openings. Next we provide evidence on two types of contractual practices that support employer monopsony power: non-compete agreements and no-poaching agreements.

Non-Compete Agreements

Non-compete agreements are contracts or clauses in contracts that prohibit an employee from working for a competitor after the employee separates from the employer. In an employment contract, a non-compete clause may prohibit the employee from working for a rival firm when employment terminates (i.e., the employee quits and/or is fired). An employee might also sign a non-compete agreement at the time of termination in return for consideration such as money. A typical non-compete specifies the relevant industry in which the employee is prohibited from finding employment, the time period during which the noncompetition obligation remains in effect, and the geographic scope of the noncompetition obligation. For example, a non-compete for a salesperson who specializes in business software might specify that the person may not work as a salesperson for firms that sell business software, for a period of one year, and in the area in which the employer operates, such as a county or state. The scope of non-compete clauses varies significantly from industry to industry, and even

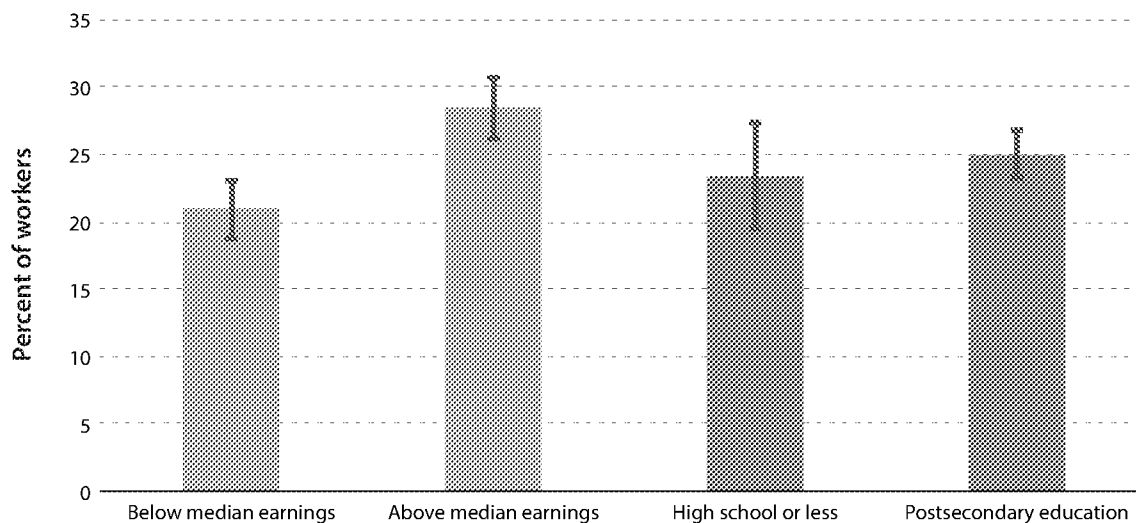
within industries, and from place to place. Some are written narrowly and some are written broadly.

Until recently, academic and policy discussion about non-competes presumed that they were used only for high-skill workers. But in 2014 it was revealed that Jimmy John's, a fast-food franchise, required low-level employees to sign contracts with non-competes that prohibited them from taking jobs at any business that obtained more than 10 percent of its revenue from "selling submarine, hero-type, deli-style, pita and/or wrapped or rolled sandwiches" within two (later extended to three) miles of any franchise, anywhere in the United States (Jamieson 2014). The non-compete covenant extended for two years. Its effect would have been to prevent a worker from obtaining a new job as a sandwich maker in large areas, including the entire city of Chicago.

Anecdotal evidence suggests that Jimmy John's practice—since discontinued—is not uncommon (Dougherty 2017a). And survey data reported in a recent paper by Starr, Prescott, and Bishara (2017) indicate that 12 percent of low-income workers—those lacking a college education with incomes less than \$40,000 per year—were subject to non-competes in 2014. Over all income levels, Starr, Prescott, and Bishara estimate that one in five workers was bound by a non-compete clause.

To supplement these findings, we contracted with Survey Sampling Inc. (SSI) to conduct a short internet survey of 919 workers in February 2017 to assess the extent to which workers are covered by non-compete clauses. After deleting

FIGURE 1.
Share of Workers Covered by a Non-Compete Agreement in Current or Former Job, by Weekly Earnings and Education



Source: Authors' calculations based on SSI survey; see text.
Note: The length of the error bars indicate the standard errors of the respective estimates.

responses by self-employed individuals, we have a sample of 795 employees. We derived sampling weights for respondents based on their income, race, sex, education, and age to make the weighted sample representative of the U.S. workforce. Specifically, workers were asked, “Does your employment relationship restrict you in any way from taking another job, such as through a non-compete clause or no-raid pact?” If they answered in the affirmative, they were asked whether a non-compete clause, no-raid pact, or other arrangement was the source of the restriction.

In the weighted sample, 15.5 percent of workers responded they were currently covered by a non-compete clause. This figure is similar to Starr, Bishara, and Prescott’s (2017) estimate before they made an adjustment for underreporting. The percentage of workers who said they were covered by a non-compete clause was slightly higher for those with a high school diploma or less (17.5 percent) than for workers with post-high school education (14.6 percent), on average.

For those who responded that their employment relationship does not restrict them in any way from taking another job, we asked, “Have you ever worked for a company that restricted where you could work after you left that company because of a non-compete clause or some other reason?” Taking into account previous employment as well as current employment, 24.5 percent of the workforce is bound by a non-compete restriction on their current job, or was bound by a non-compete from a previous job. Figure 1 displays the proportion of workers who are restricted by a non-compete

agreement in their current job or have been so restricted in a former job, disaggregated by earnings (above or below the median weekly earnings) and education (high school or less versus some postsecondary education or more). As one would expect, higher-income workers are more likely to be covered by non-compete agreements, but a remarkably high 21 percent of workers who earn less than the median salary are currently or have been restricted by a non-compete agreement. And workers with a high school diploma or less are almost equally likely to be covered by a non-compete agreement in a current or former job as are workers with some postsecondary education.

Franchise No-Poaching Agreements

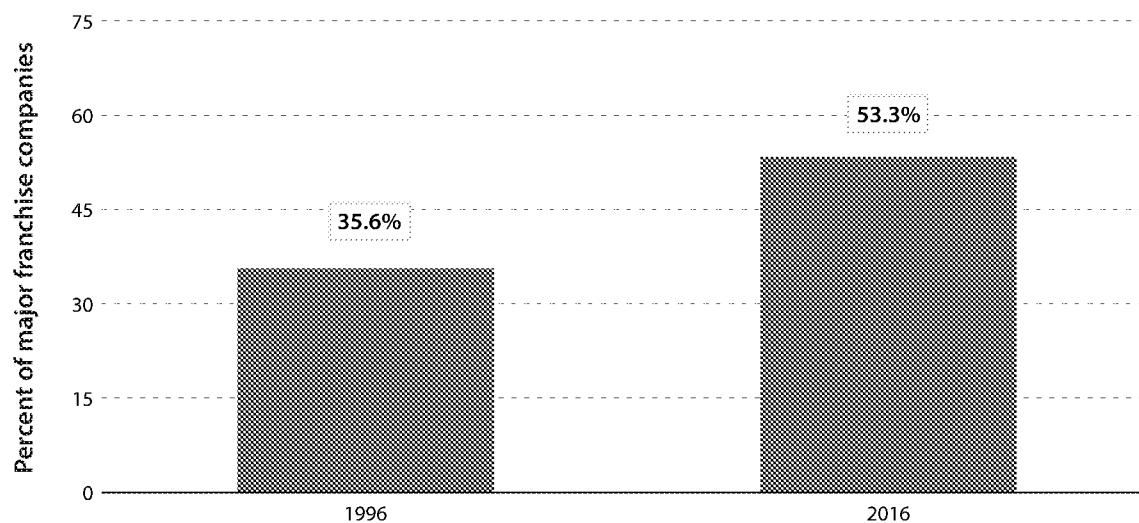
Like non-competes, no-poaching agreements went unnoticed by many labor market observers until recently. There was little evidence that companies used them, and in any event no one challenged that they were illegal. But in 2017 employees of McDonald’s sued the company under the antitrust laws for subjecting its franchisees to a no-poaching arrangement.⁶

Since at least 1987 until early in 2017, McDonald’s has included the following no-poaching clause in its standard franchise contract:

Interference With Employment Relations of Others. During the term of this Franchise, Franchisee shall not employ or seek to employ any person who is at the time employed by McDonald’s, any of its subsidiaries, or by any person who is at the time operating a McDonald’s restaurant or otherwise induce, directly or indirectly,

FIGURE 2.

Share of Major Franchise Companies with a No-Poaching Clause, 1996 and 2016



Source: Based on data provided by FRANData (frandata.com); and Krueger and Ashenfelter 2017.



such person to leave such employment. This paragraph 14 shall not be violated if such person has left the employ of any of the foregoing parties for a period in excess of six (6) months.⁷

This clause was dropped from McDonald's franchise contract in early 2017, around the time that CKE Restaurants Holdings was sued for having a similar clause in its Carl's Jr. franchise contract.

By examining franchise disclosure documents for 156 franchisors with more than 500 franchise units operating in the United States in 2016, Krueger and Ashenfelter (2017) show that 56 percent of major franchisors have no-poaching agreements in their franchise contracts. They provide an illustrative calculation indicating how no-poaching agreements within franchisors can greatly increase the effective Herfindahl-Hirschman index—a measure of industry concentration used to evaluate market competitiveness—and create employer market power over workers. In essence, if all units of a franchise chain act as if they are one company in terms of hiring practices, then an otherwise competitive labor market can become much more concentrated.

To determine whether this practice has increased or decreased over time, we obtained franchise disclosure documents filed in 1996 for the 45 largest franchisors in 2016 that were in operation in 1996 from the same source used by Krueger and Ashenfelter (2017). Figure 2 reports the share of these franchise chains with a no-poaching agreement in 1996 and in 2016. Over the past 20 years the share of major franchise companies that included a no-poaching covenant in their standard franchise agreement increased from just over one-third to slightly more than half.⁸ An example of a chain that added a no-poaching clause in the past twenty years is the International House of Pancakes, which currently requires the following of its franchisees:

Non-Solicitation. During the Term of this Agreement and for one year following the expiration or termination and each Assignment, Franchisee shall not, without the prior written consent of Franchisor, directly or indirectly: (a) employ or attempt to employ any person who at that time is employed by Franchisor, an Affiliate of Franchisor, or any other Franchisee or area developer of Franchisor, including, without limitation, any manager or assistant manager; (b) employ or attempt to employ any person who within six months prior thereto had been employed by Franchisor, an Affiliate of Franchisor, or any other Franchisee or area developer of Franchisor; or (c) induce or attempt to induce any person to leave his or her employment with Franchisor, an Affiliate of Franchisor, or any franchisee or area developer of Franchisor.⁹

In all likelihood, the proliferation of no-poaching agreements has increased franchise companies' monopsony power over workers in recent decades.

THE LIMITS OF THE LAW: WHY A NEW APPROACH IS NEEDED

Collusion and Monopsonization

Labor market concentration poses a difficult challenge to antitrust enforcement. A firm that enjoys monopsony power over a labor market and uses that power to pay its workers below the competitive rate is not liable under the antitrust laws, as long as the firm did not take intentional actions to obtain that power. For example, if a large factory dominates the labor market of a small town because other factories in the area have shut down, the factory owner is free to pay below-market wages without violating antitrust laws.

In contrast, when firms achieve labor market power through mergers or collusion—such as through no-poaching agreements—they do violate the antitrust laws. Firms obtain labor market power through merger when two employers who compete for workers combine into a single entity. If the labor market is already relatively concentrated or the firms are large employers, the increase in labor market power may be significant. Firms can obtain market power even without merging by agreeing to not compete over labor. They can do this in many ways—for example, agreeing not to hire away each other's workers, agreeing to draw from different pools of labor, coordinating on wages and benefits, sharing information, and so on.

Firms that obtain labor market power in these ways violate the antitrust laws. The problem lies in enforcement. Firms accused of violating the antitrust laws can defend themselves by arguing that apparently anticompetitive behavior allows them to lower prices by exploiting economies of scale. Anticompetitive behavior can result from hard-to-prove, and not always illegal, tacit coordination rather than explicit agreement. Thus, even when firms do not enter no-poaching agreements, firms may be able to coordinate wages without entering into explicit agreements, for example, through sharing of information about compensation, or adopting parallel practices of not raiding each other's workforce (DOJ and FTC 2016). When firms engage in these more ambiguous types of activities, plaintiffs will have trouble persuading courts that their actions are illegal.

An additional hurdle to antitrust enforcement is the cost of bringing lawsuits. Individual employees will almost never have the resources or incentives to sue employers for antitrust violations because of the vast cost of an antitrust suit along with the relatively small sums at stake. Private wage suppression suits therefore require a class action, which imposes considerable costs and risks on law firms. While the

government can bring such suits, and has in a few cases, it faces a similar problem of limited resources and high risk. In contrast, product-market antitrust claims are often brought by large firms that are harmed by the alleged anticompetitive practices.

Non-Compete Agreements

Common Law

In the common law, courts make an exception to the principle of freedom of contract and refuse to enforce non-compete agreements that are “unreasonable.”¹⁰ To determine whether a non-compete clause is unreasonable, a court typically asks whether the clause is broader than necessary to protect the employer’s legitimate business interest. Accordingly, a court might determine that the geographic scope of a non-compete clause is too broad if the employee works in a much smaller area, or the industry scope is too broad if not all employers within the designated industry actually compete with the employer in question.

Employers usually argue that the clause is needed to protect trade secrets, such as client lists, or to protect their investment in the employee, who may have received training. The worry is that if employees are permitted to work for rivals of their employers, then they will be able to transfer information to those rivals, which would discourage employers from sharing information with employees, force them to use elaborate firewalls and other protections, or refuse to invest in trade secrets in the first place. Employers might also underinvest in their employees if employees can take their new skills to rivals.

While the courts’ approach to non-compete agreements may provide some protection to low-income workers, it is plainly inadequate. First, employees frequently do not read or understand employment agreements because they are long and complex, and the workers do not have the means to hire a lawyer to interpret the contract for them. Poorly educated workers who can command only low wages are at a greater-than-usual disadvantage. In some cases, employees may be first informed of the non-compete clause after they begin work or when they quit. Second, the remedy for an unreasonable non-compete clause is generally either nonenforcement or reformation of the clause so that it is less broad; the employer is not penalized or forced to pay damages to the employee. This means that employees threatened with a lawsuit if they try to work for a rival firm will not be able to attract a lawyer to defend them. Lawyers must be paid, and low-wage workers cannot afford to pay lawyers; since they will not receive damages, lawyers cannot be paid out of any recovery. Given the frequency of the practice, employers appear to understand that they face no sanction if they insert unenforceable non-compete clauses in contracts even if the clauses enable the employers to intimidate the employees. Finally, because of the vagueness of the legal standard that governs non-compete

clauses, it is always possible that an employee will lose a case. This will further deter an employee from seeking legal relief, and a lawyer from helping him or her.

Another problem with the common law approach to noncompetition agreements is that these agreements might have significant anticompetitive effects even when they are permissible. Imagine that a monopsonistic employer requires all employees to sign non-competes as a condition of employment. The non-competes may be deemed reasonable under the common law because of their limited scope and duration, but nonetheless deter other employers from entering the market for labor because they fear that they will not be able to find enough employees to run their businesses. From a social standpoint, it may be optimal to prohibit such non-competes because of their collective anticompetitive effect even though they are individually reasonable.

Legislation

In most states, non-compete agreements are mainly governed by the common law only. But in California, North Dakota, and Oklahoma, non-competes are generally prohibited by statute.¹¹ In recent years several state legislatures, including those of Hawaii, New Mexico, Oregon, and Utah, have considered or passed legislation that puts limits on non-competes (Lohr 2016). Notably, in 2016 Illinois passed a law banning non-competes for low-wage workers, defined as those who earn no more than \$13 per hour or the relevant legal minimum wage, whichever is higher.¹²

Maine, Maryland, Massachusetts, and New Hampshire are currently considering legislation to restrict non-compete clauses, particularly with respect to low-wage workers (Beck 2017; Quinton 2017). The bills vary greatly, but some of them entail fairly sweeping changes. For example, one bill being considered in Massachusetts tightens the common law analysis of all non-compete agreements, while also prohibiting their use for low-wage workers (nonexempt workers under the Fair Labor Standards Act, who are lower-income and paid on a wage basis). For all non-competes, the bill requires employers to give workers notice of non-competes, to supply additional consideration when non-competes are created after employment begins, to review the agreement with the worker every three years, and to notify the worker of the agreement at termination. It also tightens the common law limits on duration, geographic scope, and industry scope.¹³ Going in the other direction, Idaho recently passed a law that makes it more difficult for employees to challenge a non-compete (Dougherty 2017b).

Overall, the legal regime is insufficient to address the antitrust problems posed by non-competes for several reasons. First, the common law and much of the statutory law do not address problems of market power in an adequate way. When

employers enjoy monopsony power, this type of law offers no protection to workers who must either accept unfavorable terms or do without wages. Second, the remedies are too weak. Even when non-competes are illegal, the normal remedy is simply nonenforcement. This means that employers have nothing to lose from inserting non-competes into contracts. Since employers may be able to deter workers from quitting and finding new jobs in the same industry simply by pointing out the existence of the clauses in the contracts, the law does nothing to deter employers from using the clauses. Third, while some states have taken strides to restrict non-competes for low-wage workers, these types of agreements remain lawful nearly everywhere. Fourth, while non-competes can be challenged under the antitrust laws, which provide for significant remedies, defendants can often avoid liability by showing that the non-competes serve a reasonable business purpose.¹⁴

No-Poaching Agreements within Franchises

When firms are independent, no-poaching and related agreements are clear violations of antitrust law.¹⁵ Antitrust law forbids independent firms from agreeing not to compete, and in a no-poaching agreement firms agree not to compete for workers.

However, no-poaching agreements remain common and have grown in usage in franchise contracts, as we show above. The difference is that typically a single franchisor enters an agreement with each individual franchisee under which the franchisee promises the franchisor that it will not poach employees from other franchisees or company-owned

units. This type of arrangement does not as clearly run afoul of antitrust law for two reasons. First, the components of a franchise may be considered a “single economic entity,” in which case antitrust law does not apply. Second, the agreement in the franchise setting is technically a “vertical” rather than a “horizontal” agreement, which is evaluated under a more generous standard in antitrust law. In *Williams v. I. B. Fischer Nevada*, a court recognized both of these issues in the course of holding that a no-poaching agreement between the Jack in the Box franchise and each of its franchisees did not violate section 1 of the Sherman Act.¹⁶ It is unclear whether this holding remains good law after the Supreme Court narrowed the definition of a “single economic entity” in 2010, making it easier for courts to see franchisees as independent companies that may enter conspiracies in violation of the Sherman Act.¹⁷

Nonetheless, franchisors who enter no-poaching agreements with franchisees face little risk of antitrust liability. The law remains unsettled; even if it becomes clear that the single economic entity rule has been relaxed for franchises, it will remain difficult for victims of no-poaching agreements to win cases because of the complexity of the rule-of-reason analysis applied to vertical agreements. As in the case of non-competes, workers who seek to vindicate possible legal claims face fundamental logistical problems. Because antitrust cases are complex, expensive, and risky, and no-poaching agreements may be secret, it may not be worth the time and money to bring lawsuits. Class actions remain possible but they, too, pose considerable risk to the lawyers who bring them.¹⁸ In addition, in recent years the Supreme Court has erected new barriers to class actions by workers against employers.¹⁹

A New Approach

HORIZONTAL MERGER GUIDELINES

DOJ and the FTC review mergers between large firms under the Horizontal Merger Guidelines (DOJ and FTC 2010). The Guidelines focus on the problem of product market competition, and provide rules that help regulators determine whether a merger will have anticompetitive effects in such markets. While the Guidelines acknowledge that regulators should also be on guard against mergers that enhance market power for buyers vis-à-vis suppliers, they do not address the special issues that arise when those suppliers supply labor rather than other inputs (DOJ and FTC 2010). This omission needs to be corrected.

The Guidelines (DOJ and FTC 2010) should include a new section that directs the government to screen mergers based on their likely effects on labor markets. Such an analysis can be based on the normal approach to analyzing the effects of mergers on product markets. First, the agency should define the labor activity—for example, sandwich maker, waiter, barista, or retail clerk. It may be appropriate to use very broad definitions in some cases (e.g., unskilled labor). The frequency of movements of workers between occupations—which is informative about the similarity of tasks involved in various occupations—could be a useful guide for defining the scope of labor activity.

Second, the agency should identify the various labor markets affected by the mergers. These are geographic areas that encompass the commuting range of workers of the relevant skill level. Some labor markets are national in scope (e.g., skilled professionals) and some are more limited.

Third, the agency should assess the effect of the merger on concentration in the labor market. Specifically, the agency would calculate the premerger and postmerger Herfindahl-Hirschman index levels of the labor market, and recognize a presumption against a merger if the postmerger absolute level of concentration and/or the increase indicate too high a risk of wage suppression.

Fourth, merging firms should be allowed to rebut this presumption by showing special characteristics of the labor market, such as high worker mobility, or evidence that the merger will create significant benefits—economies of scale, for example—that sufficiently offset any losses to workers.

Under our proposal, the regulators would be on guard against effects on both product market competition and labor market competition. The two are obviously different. Imagine that two manufacturers seek to merge, and that they both sell goods into a national market in which many other competitors are involved. The merger would pass the Guidelines as currently written. But imagine that the factories of the two competitors are located in the same town, and those factories are the largest employers of the town's low-skill workers. The merger should be blocked because of its negative labor market effects unless the merging companies can show that the labor market will remain competitive or that there are other significant benefits from the merger.

Because this proposal may require more analysis by the Antitrust Division at DOJ, we also suggest that the resources of this department be expanded, with special attention to hiring labor market economists. This would also provide more capacity to investigate wage collusion or no-poaching agreements.

NON-COMPETE AGREEMENTS

Non-compete agreements may be justified when employers heavily invest in training employees, or trust them with valuable information, including trade secrets, but this is rarely the case with unskilled or low-skilled workers. In these cases, the most plausible explanation for non-competes is their anticompetitive value for employers. Moreover, because many low-income workers rarely read and understand their employment contracts, the risk of harm is far greater than in other contexts. Accordingly, we believe that states should pass laws, modeled on Illinois' laws, that flatly ban non-competes for workers earning less than \$13 per hour. Specifically, we propose that non-competes be uniformly unenforceable and banned if they govern a worker who earns less than the median wage in her state.

It is possible to argue that such an approach is too crude. Some low-income workers are given significant training, and some are entrusted with trade secrets. It could be argued that employers should be allowed to use non-competes—if not too strict in terms of geographic scope, industry definition, and duration—when they can show the non-compete advances

these interests. But this would just duplicate current law, which is plainly inadequate, and in any event trade secrets are protected by another area of the law that we would leave undisturbed. Experience in California, where Silicon Valley flourishes despite (or perhaps in part due to) the unenforceability of non-competes, suggests that the strong claims made on behalf of the value of non-competes are greatly exaggerated (Fallick, Fleischman, and Rebitzer 2005; Gilson 1999). Accordingly, we believe that the best approach is a flat ban of the kind we describe.

A further problem needs to be addressed, which is the deterrent effect of even unenforceable non-competes against workers who lack the resources and sophistication to challenge them in court. To address this problem, states should pass laws that require firms to delete from employment contracts non-competes that are legally unenforceable; and to pay penalties if the firms incorrectly tell employees that they are governed by non-competes and threaten to sue them if they quit and accept jobs elsewhere in the industry. The latter types of action can be likened to fraudulent conduct and business torts that are already illegal. The regulation we advocate can also be seen as akin to the type of disclosure rules that require employers to inform workers of their employment and labor rights.

NO-POACHING AGREEMENTS

Employers sometimes defend no-poaching agreements on the grounds that they allow employers to protect their investments in employees. This is simply not an accepted view in antitrust law. There are more-efficient ways to protect investments—for example, by offering employees bonuses if they stay with the employer—that do not pose such a significant risk to labor market competition.

The same logic holds for no-poaching agreements between franchisors and franchisees. While franchisors sometimes argue that within-franchise no-poaching agreements lead to more-specific training, that training would not be lost to the franchise if no-poaching agreements were illegal; there is even less economic justification for a no-poaching agreement among franchisees in the same chain than among other unrelated employers.

Accordingly, we propose a per se rule against no-poaching agreements regardless of whether they are used outside or within franchises. In other words, no-poaching agreements would be considered illegal regardless of the circumstances of their use.

Questions and Concerns

1. Are problems with non-competes really a matter of inadequate information (e.g., Marx and Fleming 2012) rather than a problem of labor market concentration? If so, isn't the appropriate remedy a disclosure rule?

The problem with disclosure rules is that they rarely work as intended, likely because of information overload. In the context of consumer protection, study after study shows that consumers ignore or misunderstand information that is disclosed as a result of legal mandates (Ben-Shahar and Schneider 2014). This problem is especially acute for people with little education and who are often desperate for work.

2. Isn't market power more of a problem with high-skill and hence high-income workers than with low-skill workers?

Sandwich makers might be indifferent between taking a job at another sandwich shop and at any other employer of low-skill workers, e.g., a warehouse or factory. If so, the non-compete that is limited to the sandwich industry will not prevent them from switching jobs. In contrast, computer programmers whose skills and training are specific to that industry might have trouble finding new positions if they are subject to a non-compete.

We focus on the case of low-income workers because it has been overlooked and the hardship is greater. If labor markets for low-wage workers are at least somewhat disconnected from each other, then restricting mobility will suppress low-wage workers' ability to move to higher-paying jobs. Moreover, even if all employers offered low-wage workers the same pay, non-competes could depress the entire wage scale by crowding low-wage workers into certain sectors. The fact that employers at Jimmy John's and other franchises use (or have used) non-competes suggests that they think that it increases their market power over workers. In addition, low-skilled workers are less likely to move across geographic boundaries than high-wage workers, which gives employers local monopsony power over low-wage workers. Finally, if monopsony power and anticompetitive practices suppress pay, low-wage sectors may, in fact, be a manifestation of such features of the labor market.

3. Are there less-aggressive, more-tailored measures to address the problems we identify (including disclosure rules, as discussed above)?

There may be, but it is important to note the considerable confusion over whether non-competes are enforceable, as well as widespread employer abuse of the practice. We argue that a simple, easily understood rule, such as an outright ban of non-competes for workers earning less than the state median wage, is likely to be effective and ultimately more efficient than a more tailored approach that in principle could be economically efficient, but in practice would be very complicated to administer and follow. The fact that some states, like Illinois, have begun to ban non-competes is a sign that political economy forces are aligned behind this approach, because of its simplicity, popularity, and efficacy.

4. Is there a federal remedy for problems of employer wage collusion, non-competes, and no-poaching agreements?

If states do not adequately regulate non-competes and no-poaching agreements, then the federal government should step in. Congress could pass laws that ban these practices. In addition, under its existing legal authority, the FTC could likely ban non-competes and no-poaching agreements as unfair trade practices. While federal regulation can be applied only to "interstate commerce," that term has been interpreted broadly by the courts, so that a federal intervention would likely be valid and effective.

5. If these proposals are implemented, won't employers find other ways to exercise monopsony power?

Even if non-competes and no-poaching agreements are prohibited, and mergers are subjected to greater scrutiny, employers likely will seek out new ways of extending and exercising monopsony power. But it is doubtful that these other methods are equally effective substitutes for the practices that we seek to constrain. In any event, we advocate additional research and, if appropriate, legal regulation to address these other practices.

Conclusion

The problems we have focused on—mergers, non-competes, and no-poaching agreements—are part of a much larger problem: employer concentration and market power within labor markets. While the exact contours of the problem remain obscure, there is little doubt that shifting market power has contributed to income inequality, wage stagnation, and sluggish economic growth. Even if our solutions are adopted, we expect that labor market concentration and unequal bargaining power will continue to be a problem as employers find new ways to enhance their market power.

We hope, then, to stimulate reflection on this larger problem. There seem to be three general avenues for future research and policy. First, it may be necessary to strengthen and reorient antitrust law so that it is more usable for labor market concentration than it currently is. Merger screening is only one part of this process. There may be other commonly used practices—like information sharing, coordination of hiring

through headhunters and networks, and so on—that facilitate coordination on wages and hiring, or enable monopsonists to extend their market power.

Second, researchers should also evaluate anew employment regulations that may enhance workers' bargaining power. While a great deal of attention has been devoted to minimum wage laws, other laws that control aspects of the employment relationship—including hours, working conditions, and benefits—may have desirable competitive effects by offsetting unequal employer bargaining power. Contract terms (beyond non-competes) that reduce worker mobility also may be a matter of concern.

Third, there are broad public-policy strategies that might meaningfully improve the bargaining power of workers. These include public infrastructure, which can increase the size of labor markets by reducing commute times; education; immigration policy; and union regulation.

Authors

Alan Krueger

*Bendheim Professor of Economics and Public Affairs,
Princeton University*

Alan Krueger has published widely on the economics of education, unemployment, labor demand, income distribution, social insurance, labor market regulation, terrorism, and environmental economics. Since 1987 he has held a joint appointment in the Economics Department and Woodrow Wilson School at Princeton University. He is the founding Director of the Princeton University Survey Research Center. He is the author of *What Makes A Terrorist: Economics and the Roots of Terrorism* and *Education Matters: A Selection of Essays on Education*, co-author of *Myth and Measurement: The New Economics of the Minimum Wage*, and co-author of *Inequality in America: What Role for Human Capital Policies?*

Krueger served as Chairman of President Barack Obama's Council of Economic Advisers and a Member of the Cabinet from 2011 to 2013. He also served as Assistant Secretary for Economic Policy and Chief Economist of the U.S. Department of the Treasury in 2009–10 and as Chief Economist at the U.S. Department of Labor in 2004–05.

He is currently Vice President of the American Economic Association, and has been a member of the Executive Committee of the American Economic Association (2005–07) and International Economic Association. Since February 2017, he has been a member of the Board of Directors of BNP Paribas USA. He has been a member of the Board of Directors of the Russell Sage Foundation, MacArthur Foundation, and the American Institutes for Research, as well as a member of the editorial board of *Science* (2001–09), editor of the *Journal of Economic Perspectives* (1996–2002) and co-editor of the *Journal of the European Economic Association* (2003–05).

He received a BS degree (with honors) from Cornell University's School of Industrial & Labor Relations in 1983, an AM in Economics from Harvard University in 1985, and a PhD in Economics from Harvard University in 1987.

Eric Posner

*Kirkland & Ellis Distinguished Service Professor of Law,
The University of Chicago Law School*

Eric Posner is Kirkland and Ellis Distinguished Service Professor of Law, University of Chicago. His research interests include financial regulation, international law, and constitutional law. His books include *Radical Markets* (Princeton, forthcoming) (with Glen Weyl); *Last Resort: The Financial Crisis and the Future of Bailouts* (University of Chicago Press, forthcoming); *The Twilight of International Human Rights* (Oxford, 2014); *Economic Foundations of International Law* (with Alan Sykes) (Harvard, 2013); *Contract Law and Theory* (Aspen, 2011); *The Executive Unbound: After the Madisonian Republic* (with Adrian Vermeule) (Oxford, 2011); *Climate Change Justice* (with David Weisbach) (Princeton, 2010); *The Perils of Global Legalism* (Chicago, 2009); *Terror in the Balance: Security, Liberty and the Courts* (with Adrian Vermeule) (Oxford, 2007); *New Foundations of Cost-Benefit Analysis* (with Matthew Adler) (Harvard, 2006); *The Limits of International Law* (with Jack Goldsmith) (Oxford, 2005); *Law and Social Norms* (Harvard, 2000); *Chicago Lectures in Law and Economics* (editor) (Foundation, 2000); *Cost-Benefit Analysis: Legal, Economic, and Philosophical Perspectives* (editor, with Matthew Adler) (University of Chicago, 2001). He is a fellow of the American Academy of Arts and Sciences and a member of the American Law Institute.

Acknowledgments

We thank the staff of The Hamilton Project for helpful feedback, Orley Ashenfelter for comments, and Kyle Trevett for valuable research assistance.

Endnotes

1. For evidence on the effect of employer concentration on wages, see Azar, Marinescu, and Steinbaum (2017). For evidence on growing firm concentration in the labor market, see Autor et al. (2017).
2. See also U.S. and State of Arizona v. Arizona Hospital and Health Care Association & AzHHA Service Corp., No. CV07-1030-PHX (D. Ariz. Final Judgment filed September 12, 2007), www.justice.gov/atr/case/us-and-state-arizona-v-arizona-hospital-and-healthcare-association-and-azhha-service-corp.
3. States vary substantially in terms of what they consider to be a reasonable non-compete agreement, and how they approach the enforcement of non-competes more generally. For example, some states will allow a court to enforce a modified version of a contract that is otherwise unenforceable, while other states do not permit this.
4. Non-competes can nevertheless still be damaging for workers with adequate legal representation and knowledge, as the examples in Dougherty (2017a) suggest.
5. For some notable cases involving more general settings, see Todd v. Exxon, 275 F.3d 191 (2d Cir. 2001) (petrochemical companies shared salary information of certain employees); Nobody in Particular Presents, Inc. v. Clear Channel Commc'n, 311 F.Supp. 2d 1048 (D. Co. 2004) (DJs); Jung v. Ass'n of Am. Med. Coll., 300 F.Supp.2d 119 (D. D.C. 2004) (physicians); In re Animation Workers Antitrust Litig., 123 F.Supp.3d 1175 (N.D. Cal. 2015) (animation workers).
6. Class Action Complaint, Deslandes v. McDonald's USA, LLC et al, No. 1:17-cv-04857 (N.D. Ill. filed June 28, 2017).
7. Deslandes v. McDonald's USA, 18.
8. The 18-percentage-point increase in the share of major franchise chains with a no-poaching restriction over the past two decades was unlikely to have occurred by chance; a paired t-test of no change has a p-value of 0.004.
9. Section 16.16 of the International House of Pancake 2017 Franchise Agreement, registered with the Wisconsin Department of Financial Institutions, <https://www.wdfr.org/apps/FranchiseSearch/details.aspx?id=615829&hash=177165780&search=external&type=GENERAL> on January 1, 2018.
10. There is considerable variation in the relevant common law across states. The discussion abstracts away from the many differences in law.
11. Cal. Business & Professions Code § 16600; N.D. Cent. Code § 9-08-06; OK Stat. § 15-219A.
12. Illinois Public Act 099-0860 (2016).
13. H.2366, 2017 Gen. Court, 190th Sess. (Mass. 2017).
14. 15 U.S.C. § 1-2. Under standard antitrust analysis, plaintiffs can prevail either by showing that the non-compete was the result of a conspiracy (§ 1) or that it furthered an effort to monopolize (or monopsonize) (§ 2). But an ordinary non-compete clause is not a conspiracy, because it involves an agreement between the employee and the employer, who are not competitors, rather than between two firms. And Section 2 can usually be enforced only against firms that achieve or attempt to achieve significant market dominance, and not in the case that concerns us, where common usage of non-competes across firms create labor market frictions that enhance employers' bargaining power without giving them full-blown monopsonies. For an attempt to challenge a fairly significant non-compete arrangement that failed because a court was persuaded that it served legitimate business purposes, see Eichorn v. AT&T Corp., 248 F.3d 131 (3rd Cir. 2001).
15. In 2010 Adobe Systems, Apple, Google, Intel, Intuit, and Pixar entered a consent decree after the government accused them of entering no-poaching agreements in violation of antitrust law. United States v. Adobe Sys., Inc., No. 1:10-cv-01629 (D.D.C. filed Sept. 24, 2010); United States v. Lucasfilm Ltd., No. 1:10-cv-02220 (D.D.C. filed Dec. 21, 2010); DOJ 2010.
16. 999 F.2d 445, 447-448 (9th Cir. 1993).
17. Am. Needle, Inc. v. Nat'l Football League, 560 U.S. 183, 186-187 (2010). For a discussion, see Lindsay and Santon (2012).
18. See e.g., Weisfeld v. Sun Chem. Corp., 84 Fed.Appx. 257 (3rd Cir. 2004), which provides a vivid illustration of the difficulties that lawyers face in constructing a class of workers. To obtain class certification, a plaintiff must show that the alleged wrongful conduct affected all members of the class in a similar way. The Court held that the plaintiff could not make such a showing because of variation among putative class members, including:
 - whether a covenant not to compete was included in a particular employee's contract; the employee's salary history, educational and other qualifications; the employer's place of business; the employee's willingness to relocate to a distant competitor; and [employees'] ability to seek employment in other industries in which their skills could be utilized (e.g., pharmaceuticals, cosmetics).*Id.*, citing Weisfeld v. Sun Chemical Corp., 210 F.R.D. 136, 144 (D.N.J.2002).
19. Wal-Mart v. Dukes, 564 U.S. 338 (2011).

References

- Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen. 2017, May. “The Fall of the Labor Share and the Rise of Superstar Firms.” Working Paper 23396, National Bureau for Economic Research, Cambridge, MA.
- Azar, José, Ioana Marinescu, and Marshall I. Steinbaum. 2017, December. “Labor Market Concentration.” Working Paper 24147, National Bureau for Economic Research, Cambridge, MA.
- Beck, Russell. 2017, October. “Massachusetts Noncompete Reform: What You Need to Know—by October 31.” Fair Competition Law, Boston, MA.
- Ben-Shahar, Omri, and Carl E. Schneider. 2014. *More Than You Wanted to Know: The Failure of Mandated Disclosure*. Princeton, NJ: Princeton University Press.
- Burdett, Kenneth, and Dale T. Mortenson. 1998. “Wage Differentials, Employer Size, and Unemployment.” *International Economic Review* 39 (2): 257–73.
- Council of Economic Advisers (CEA). 2016, October. “Labor Market Monopsony: Trends, Consequences, and Policy Responses.” Issue Brief, White House, Washington, DC.
- Dougherty, Conor. 2017a, May 13. “How Noncompete Clauses Keep Workers Locked In.” *New York Times*.
- . 2017b, July 14. “Noncompete Pacts, Under Siege, Find Haven in Idaho.” *New York Times*.
- Fair Labor Standards Act of 1938 (FLSA) 29 U.S.C. 201, et seq.
- Fallick, Bruce, Charles A. Fleischman, and James B. Rebitzer. 2005. “Job-Hopping in Silicon Valley: Some Evidence Concerning the Micro-Foundations of a High Technology Cluster.” Federal Reserve Board, Washington, DC.
- Federal Trade Commission (FTC). 1995, June. “Council of Fashion Designers of America.” Press Release, Federal Trade Commission, Washington DC.
- Gilson, Ronald J. 1999. “The Legal Infrastructure of High Technology Industrial Districts: Silicon Valley, Route 128, and Covenants Not to Compete.” *New York University Law Review* 74 (3): 575–629.
- Jamieson, Dave. 2014, October. “Jimmy John’s Makes Low-Wage Workers Sign ‘Oppressive’ Noncompete Agreements.” Huffington Post.
- Krueger, Alan B. and Orley Ashenfelter. 2017, September. “Theory and Evidence on Employer Collusion in the Franchise Sector.” Working Paper 614, Princeton University, Princeton, NJ.
- Lindsay, Michael, and Katherine Santon. 2012. “No Poaching Allowed: Antitrust Issues in Labor Markets.” *Antitrust* 26 (3): 73–75.
- Lohr, Steve. 2016, June 28. “To Compete Better, States are Trying to Curb Noncompete Pacts.” *New York Times*.
- Manning, Alan. 2003. *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton, NJ: Princeton University Press.
- Marx, Matt, and Lee Fleming. 2012. “Non-compete Agreements: Barriers to Entry . . . and Exit?” In *Innovation Policy and the Economy*, vol. 12., ed. Josh Lerner and Scott Stern (39–64). Chicago, IL: University of Chicago Press.
- Nylen, Leah. 2017, September. “DOJ Cases Challenging No-Poach, Wage-Fixing Likely Coming Soon.” MLex Market Insight, Washington, DC.
- Quinton, Sophie. 2017, May. “Why Janitors Get Noncompete Agreements, Too.” The Pew Charitable Trusts, Research & Analysis, Stateline.
- Robinson, Joan. 1969. *The Economics of Imperfect Competition*. London, UK: Palgrave Macmillan.
- Sherman Antitrust Act of 1890, 26 Stat. 209 (1890).
- Smith, Adam. 2007, January. *An Inquiry into the Nature and Causes of the Wealth of Nations*. Reprint of the 1776 London edition, University of Michigan Library. <http://name.umdl.umich.edu/004861571.0001.001>.
- Starr, Evan P., J. J. Prescott, and Norman Bishara. 2017, December. “Noncompetes in the U.S. Labor Force.” Available online at SSRN.
- U.S. Department of Justice (DOJ). 2010, September. “Justice Department Requires Six High Tech Companies to Stop Entering into Anticompetitive Employee Solicitation Agreements.” Press Release, Office of Public Affairs, U.S. Department of Justice, Washington, DC.
- U.S. Department of Justice (DOJ) and Federal Trade Commission (FTC). 2010. “Horizontal Merger Guidelines.” U.S. Department of Justice and Federal Trade Commission, Washington, DC.
- . 2016. “Antitrust Guidance for Human Resource Professionals.” U.S. Department of Justice and Federal Trade Commission, Washington, DC.
- U.S. Department of the Treasury. 2016, March. “Non-Compete Contracts: Economic Effects and Policy Implications.” Office of Economic Policy, U.S. Department of the Treasury, Washington, DC.
- White House. 2016, May. “Non-Compete Agreements: Analysis of the Usage, Potential Issues, and State Responses.” White House, Washington, DC.

THE
HAMILTON
 PROJECT

ADVISORY COUNCIL

GEORGE A. AKERLOF
 University Professor
 Georgetown University

ROBERT C. ALTMAN
 Founder & Senior Chairman
 Frameco

KAREN ANDERSON
 Senior Director of Policy and Communications
 Bureau of Economic Analysis
 Research in Economics
 The University of Chicago

ALAN S. BLINDER
 Gordon G. Ross Chair Memorial Professor of
 Economics & Public Affairs
 Provost University
 National & Senior Fellow
 The Brookings Institution

ROBERT C. BOYD
 Professor of International
 Government University

STEPHEN A. BRIDGEMAN
 Chairman
 General Motors

JOHN G. BURCH
 Program Director
 Manufacturing Institute of Technology

CHRISTOPHER C. CHURCH
 Co-Founder and Co-Founder
 The Community Institute

BLAIR CHRISTIAN
 Partner
 Carmichael Partners LLC

ANDREW W. CLINEBARGER
 Dean & D. H. Price Professor
 of Public Policy
 Harvard Kennedy School

ED CURRAN
 Professor & Partner Dean
 McQuar School of Public Policy
 Georgetown University

ROLAND D. DAVIS
 Henry Lee Professor of Economics
 Harvard University

JASON D. DEBBAN
 Professor of the Practice of
 Economic Policy
 Harvard Kennedy School
 Senior Counselor
 The Hamilton Project

MARK T. DILLON
 Chairman & Managing Director
 Cornerstone Partners

TED DUNN
 Vice President of Financial
 Economic Studies
 The Brookings Institution

THOMAS F. DUBOIS
 President
 Warming Climate

DEBORAH CERRIACOFI
 President & Chief Executive Officer
 Capital Group Corporation Africa

ROBERT GREENSTEIN
 Founder & President
 Center for Budget and Policy Priorities

MICHAEL GREENSTONE
 Milton Friedman Professor of Economics
 Director of the Becker-DeMunnich Institute for
 Research in Economics
 Director of the Energy Policy Institute
 University of Chicago

GLENN H. HUBBARD
 Co-Founder
 HealthFund
 Co-Founder
 Bluebird

HEIDI A. JENSEN
 Chairman
 American Capital Partners

LANCELOT F. JAY
 Elizabeth Allen Professor of Economics
 Harvard University

MICHAEL S. KEATING
 Professor of Economics
 University of Maryland
 Nease Senior Fellow
 The Brookings Institution

JEFF KATON
 Managing Partner
 Bank of America Global

HOWARD J. MARAS
 Co-Founder
 Oakley Capital Management, L.P.

MARIE MOGENSEN
 Partner, Senior Lecturer, W. Bush
 Co-Founder, No Labels

BRIAN MURPHY
 Chief Executive Officer & Founder
 Blue Park Capital Management

ALY MURAS
 Former Head of American Private Equity
 KPC
 Founder
 NextCo Holdings

GLENN W. NICKERSON
 Former Vice Chairman
 German Swiss Group, Inc.

PETER ORNSTEIN
 Vice Chairman of Investment Banking
 Managing Director and
 Chief Economist of Health
 Care
 Nease Senior Fellow
 The Brookings Institution

NICHOLAS PERINI
 Managing Partner &
 Chief Executive Officer
 Rely Capital

HENRY PREDERER
 Chairman
 RBF Partners

MERCHANTELLITY
 Managing Director
 The Hamilton Project
 Ford McDevitt Guest Foundation

ROBERTO PEREIRA
 Chief Executive Officer, Policy & President Emeritus
 United Medical

GEORGE H. RIVER
 Senior Fellow, Economic Studies
 Center for Health Policy
 The Brookings Institution

DARWIN RUBENSTEIN
 Co-Founder &
 Co-Chief Executive Officer
 The Center Group

TERENCE RYAN
 Former U.S. Treasury Secretary
 Co-Chair Emeritus
 Council on Foreign Relations

EDWARD S. SACHS
 Senior Counsel
 Chief Counsel, Steven F. Hamilton LLP

SHARON SAMPSON
 Chief Operating Officer
 Facebook

LIANE WHITMORE STEINBERGER
 Maxwell Walker Alexander Professor
 Director
 The Institute of Policy Studies
 Harvard Kennedy School
 Nease Senior Fellow
 The Brookings Institution

RALPH W. STEINBERGER
 President & Chief Executive Officer
 Ustream

ERIC STEINER
 Technical Advisor
 Al Johnson Inc.

ERIC STEINER
 Chairman and CEO
 3d West Holdings

THOMAS F. STEVEN
 Business Leader and Philanthropist

LAWRENCE H. STEINBERG
 Charles H. Busch Distinguished Professor
 Harvard University

LALITA SANDHYA TILSON
 Professor of Business Administration and
 Economics Director
 Institute for Business & Society Impact
 Berkeley-Haas School of Business

RAY SHAMBALKIN
 Chairman

Highlights

New evidence that labor markets are being rendered uncompetitive by employers suggests that the time has come to strengthen legal protections for workers. Labor market collusion or monopsonization—the exercise of employer market power in labor markets—may contribute to wage stagnation, rising inequality, and declining productivity in the American economy, trends which have hit low-income workers especially hard. Alan Krueger and Eric Posner propose three reforms to address these problems.

The Proposal

Enhance scrutiny of mergers for adverse labor market effects. The authors propose that the Horizontal Merger Guidelines include a new section that directs the government to screen mergers based on their likely effects on labor markets.

Ban non-compete covenants that bind low-wage workers. The authors propose that non-competes be prohibited for workers who earn less than the median wage in their state.

Prohibit no-poaching arrangements among establishments that belong to a single franchise company. The authors propose that no-poaching agreements between franchisors and franchisees be uniformly banned.

Benefits

Mergers that reduce labor market competition, non-compete agreements, and no-poaching agreements are part of a much larger problem: employer concentration and market power within labor markets. While the exact contours of the problem remain unclear, there is little doubt that shifting market power has contributed to income inequality, wage stagnation, and sluggish economic growth. The policies in this proposal would limit some of the more harmful employer practices.



1775 Massachusetts Ave., NW
Washington, DC 20036

(202) 797-6464

BROOKINGS



Printed on recycled paper.

WWW.HAMILTONPROJECT.ORG

FTC_AR_0000733

The growing use of mandatory arbitration

Access to the courts is now barred for more than 60 million American workers

Report • By Alexander J.S. Colvin • September 27, 2017

Executive summary

In a trend driven by a series of Supreme Court decisions dating back to 1991, American employers are increasingly requiring their workers to sign mandatory arbitration agreements. Under such agreements, workers whose rights are violated can't pursue their claims in court but must submit to arbitration procedures that research shows overwhelmingly favor employers.

In reviewing the existing literature on the extent of this practice, we found that the share of workers subject to mandatory arbitration had clearly increased in the decade following the initial 1991 court decision: by the early 2000s, the share of workers subject to mandatory arbitration had risen from just over 2 percent (in 1992) to almost a quarter of the workforce. However, more recent data were not available. In order to obtain current data for our study, we conducted a nationally representative survey of nonunion private-sector employers regarding their use of mandatory employment arbitration.

This study finds that since the early 2000s, the share of workers subject to mandatory arbitration has more than doubled and now exceeds 55 percent. This trend has weakened the position of workers whose rights are violated, barring access to the courts for all types of legal claims, including those based on Title VII of the Civil Rights Act, the Americans with Disabilities Act, the Family and Medical Leave Act, and the Fair Labor Standards Act.

In October 2017, the Supreme Court will hear a set of consolidated cases challenging the inclusion of class action waivers in arbitration agreements. Class action waivers bar employees from participating in class action lawsuits to address widespread violations of workers' rights in a workplace. The Court will rule on whether class action waivers are a violation of the National Labor Relations Act; their decision could have wide-reaching implications for workers' rights going forward.

Key findings of this study

- ✦ More than half—53.9 percent—of nonunion private-sector employers have mandatory arbitration

SECTIONS

1. Executive summary • 1
2. Introduction • 2
3. Background: The Supreme Court's role in the increased use of mandatory employment arbitration agreements • 3
4. Existing research on the extent of mandatory employment arbitration • 4
5. Findings of this study • 4
6. Conclusion: Mandatory arbitration is a growing threat to workers' rights • 7

About the author • 8

Methodological appendix • 8

Endnotes • 9

References • 10