Proceedings of the 2004 Winter Simulation Conference
R .G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, eds.

OUTPATIENT CLINIC SCHEDULING – A SIMULATION APPROACH

Ming Guo

Division of Human Resources
Cincinnati Children's Hospital
Medical Center
3333 Burnet Ave.
Cincinnati, OH 45229, U.S.A

Michael Wagner

Division of Biomedical Informatics
Cincinnati Children's Hospital
Medical Center
3333 Burnet Ave.
Cincinnati, OH 45229, U.S.A.

Constance West

Division of Pediatric Ophthalmology
Cincinnati Children's Hospital
Medical Center
3333 Burnet Ave.
Cincinnati, OH 45229, U.S.A

ABSTRACT

The process by which outpatients are scheduled for a doctor's visit is a crucial determinant of the overall efficiency of the patient flow. The problem at hand consists of determining prioritization (triage) rules so that adequate patient care is guaranteed, resources (provider schedules) are utilized efficiently and a service guarantee can be ensured. We present a simulation framework for the evaluation and optimization of scheduling rules. We outline the basic ingredients of our model, illustrate the kinds of analyses it has enabled us to perform and summarize our experience with a preliminary implementation for the Division of Pediatric Ophthalmology at Cincinnati Children's Hospital Medical Center. Challenges for adaptations to other settings are also outlined.

1 INTRODUCTION

In the face of continuously rising health care costs, various initiatives have been started to increase the operational efficiency and cost effectiveness of the health care delivery process (see, e.g., the AHRQ website in the first reference for a wealth of information on health care cost management). In particular, and especially amid diminished capacity, there is a clear need for analytical tools that can provide insights into the dynamics of patient flows in clinics and hospitals. Variability in both supply and demand, when left unmanaged, necessarily produce crowding, staff overloads, unmet patient needs and general frustration (McManus et. al. 2003).

The scheduling of outpatient appointments, i.e., the process performed by customer service representatives in call centers and their systems by which slots on providers' scheduled are assigned to incoming requests for appointments, is an integral component of the overall management of patient flow and an important factor for the overall operational efficiency of any outpatient clinic. It can be viewed as the point where supply meets demand in a clinic. The main contributing factors to the complexity of this scheduling problem are the randomness of patient demand, substantial no-show rates in certain population segments, the large number of diagnosis types resulting in different follow-up patterns and the highly variable nature of the providers' schedules which effectively cause the supply to exhibit severe temporary bottlenecks.

Our aims are three-fold. By developing a modeling strategy we want to foster a deeper understanding of operational variables that affect key performance measures such as patients' waiting times for appointments and effective schedule utilization. Additionally, we aim to provide a computational test bed which can be used to optimize scheduling strategies implemented in the call center. Finally, we aim to provide clinic management with a decision support tool that can be utilized, e.g., to justify hiring decisions or operational changes.

A number of related papers have appeared in the literature. (Isken, Ward and McKee 1999) outlined a general framework for modeling outpatient clinics with the purpose of exploring questions related to demand, appointment scheduling, patient flow patterns and staffing. They assume a fully loaded one week appointment book as input for their simulation. (Ho and Lau 1992) present theoretical models of detailed daily operations with patient arrivals and resource constraints as well as the impact on staff idle time. (Harper and Gamlin 2003) and references therein discuss simulation approaches to designing detailed daily schedules, e.g., to minimize waiting times for patients after they arrive in the clinic. The problem we discuss here is distinct but complementary since its focus lies on a higher level: our primary aims are to minimize the delays for patients to get an appointment while simultaneously maximizing provider utilization and overall clinic efficiency. We envision future versions of our simulation model which integrate aspects of daily operations into the overall framework.

As an example of an application of our model we discuss a pilot implementation within the Division of Pediatric Ophthalmology at Cincinnati Children's Hospital Medical Center (CCHMC). This outpatient clinic specializes in diagnosing and treating all types of eye disorders, including those systemic diseases that also affect the eye, in children. It serves as a regional referral center for all major eye diseases and trauma. Pediatric care ranges from routine eye exams to very complex diagnoses. Over the past three years, it has experienced over 50% of growth in patient flow. The clinic is currently staffed by four pediatric ophthalmologists (MDs) and two optometrists (ODs). Until recently the clinic, like so many others, suffered from a long scheduling backlog, which resulted in long waiting times for new appointments. The current appointment system is a fragmented one, with a mix of manual and computerized systems being used. The soon-to-be realized introduction of a new call-center software package (Tempus Software, Jacksonville, FL) which allows for the definition of flexible triage and scheduling rules was the initial motivation for this study. However, we want to stress that the modeling approach is general and applicable to other clinics and settings.

The remainder of the paper is organized as follows: Section 2 provides an overview of the system characteristics and design; in Section 3 we detail our data sources as well as our implementation. Section 4 describes several analysis tools we developed to evaluate scheduling strategies. Section 5 concludes with future extensions and a summary.

2 OVERVIEW OF THE SYSTEM

Our Patient Scheduling Simulation Model (PSSM) captures four components of outpatient clinic scheduling systems: external demand for appointments, supply of provider time-slots, the patient flow logic (which effectively also characterizes internally generated demand) and the scheduling algorithm. The first three components need to be represented in the model with sufficient accuracy as to result in a realistic representation of true system dynamics. The last component is the target of optimization. We proceed by discussing our approaches to each of these components.

2.1 Demand

Demand is realized by the (stochastic) arrival of calls from patients requesting appointments. Our model is patient-centric, i.e., individual patients and their characteristics are simulated explicitly instead of being aggregated into flows. The essential characteristics by which patients are described are their diagnosis class (i.e., the type of appointment slot they require), their preference for a provider, their follow-up and call behavior as well as their no-show probability.

The variability of the demand stream hence can be attributed to several different factors. The number of calls for appointments from new patients varies from day to day. Desired follow-up intervals (patient initiated or provider mandated) are variable, the total number of clinic visits per patient is stochastic, and, last but not least, patients have different habits as to when they schedule follow-up appointments. PSSM accounts for all of these factors explicitly. Another source of complexity in patient scheduling (and a reason that traditional process simulation approaches are of limited use in this context) is that incoming requests for appointments can be for slots that are weeks or months in the future. At the same time, a significant proportion of appointments are scheduled "at the last minute", be it because they are true urgencies or because they are scheduled shortly before the patient is asked to come in. In order to tend to the true urgencies and to provide a certain service guarantee (e.g., that all new patients can be seen within a week if they so desire), this implies that parts of the schedule need to be kept open (carved out). Demand that is manifested far in advance on the other hand, is less time-sensitive and can be effectively used to smooth out the schedule utilization.

PSSM explicitly distinguishes between patients with commercial insurance coverage and Medicaid/self-pay patients primarily for two reasons. First, we determined that the insurance type has a high correlation with (and thus is a good predictor of) the no-show rate, which is a major contributor towards the variation in scheduling efficiency. (We currently assume that no-show rates are independent of the time that appointments are scheduled in advance, an extension to a model that takes the postulated direct functional relationship between no-show rate and the waiting time into account is planned.) Secondly, the reimbursement structures for the two types of insurance differ significantly, so identifying the insurance type enables later extension of the model to analyze the financial impact of different scheduling policies. An analysis of historical data revealed that roughly 65% of all calls come from patients covered by Medicaid.

Currently, we distinguish the nine different appointment types enumerated in Table 1. There is an obvious tradeoff between complexity and accuracy when deciding on the level of aggregation for appointment types, in fact one of our aims is for PSSM to provide a suitable platform which can be used to experiment with different levels of aggregation. The basic requirement is that appointments that are grouped into one category be sufficiently homoge-

Table 1: Appointment Types in PSSM

DL	Dilated "regular" appointment
FU	Non-dilated follow-up appointment
DM	Dilated regular appointment (Medicaid/self-pay)
RM	Non-dilated follow-up appointment (Medicaid/self-pay)
ER	Emergency patient appointment
PO	Pre- or post-surgery checkup appointment
AN	Adult patient dilated appointment
AF	Adult patient non-dilated appointment
RO	Specialty appointment for ROP patients

neous in terms of capacity utilization (i.e., they should be similar in terms of required provider face time so that they can be easily interchanged in any given schedule).

Emergency appointments require immediate care and include physician and in-house referrals. PO slots are reserved for the typically very short pre- or post surgerycheckups. Note that we do not account for the actual surgeries in the current version of this model, this is under consideration for future versions. ROs are specialty appointments for premature neonates with retinopathies (retrolental fibroplasias).

2.2 Patient Flow Logic

The patient flow, i.e., the sequence of appointments each patient goes through in the model, is largely determined by the clinical diagnosis, which in our case is necessarily represented by the appointment type. In particular, we used historical data to determine distributions of follow-up patterns (number and type of follow-ups) for each appointment type. PSSM assumes that patients should preferentially be seen by the same doctor for each visit to the clinic. Some proportion of the population of new “regular” patients do express a preference for a particular doctor and are willing to wait for that doctor to be available (we currently take this to be about 40%), while the rest will want to be seen by any doctor as early as possible. Emergencies are attended by any available physician.

Patients differ in their habits of when they tend to call in to schedule appointments. PSSM assumes that new patients will want to schedule appointments as early as possible (e.g., right after they call in). Follow-up appointments, however, are sometimes scheduled well in advance of the actual date. PSSM models this “call behavior” explicitly by using any pre-specified distribution of call-ahead times. For example, we currently assume 50% of patients schedule a follow-up appointment immediately after exiting the previous one, 30 % call 2 weeks in advance and 20% call wanting to schedule a follow-up immediately before wanting to be seen. The overall patient flow logic is summarized in Figures 1 and 3.

1. Arrival of new patient call
2. Patient characteristics are drawn from distributions (appointment type, insurance, etc.)
3. Appointment is scheduled
4. Delay until appointment day
5. Does patient show up for appointment? If not, go to 6, otherwise go to 7
6. Does patient call for rescheduling? If not, exit, otherwise go to 3.
7. Does patient need a follow-up appointment in the same appointment category? If so, then go to 8, if not, then go to 9.
8. Delay until patient calls for follow-up appointment, then go to 3.
9. Does patient need a “regular” follow-up appointment? If not, then exit system.
10. Delay until patient calls for follow-up appointment, then go to 3.

Figure 1: Basic Patient Flow Logic

2.3 Supply

CCHMC’s Ophthalmology clinic currently employs six providers, four MDs and two ODs, with one OD having been hired in July 2003. As is common practice, the provider schedules are encoded by templates, that is, daily specifications of the numbers of appointments of different types each doctor aims to fill (see Table 2). In particular, this accommodates different productivities of the providers as well as different specializations that will result in different proportions of appointment types.

Because of various scheduling requirements, vacation times, research time and other commitments it turns out (surprisingly) that overall provider availability is highly variable, with the total number of weekly slots among all doctors varying between 137 and 622. Figure 2 illustrates this by showing a time series of the aggregate weekly capacity for follow-up appointments (FU) for all providers in the timeframe 9/02-7/04.

Table 2: Extract From a Template Schedule Specification (Only 2 out of 6 Providers Shown)

Provider	MD1	MD1	MD2	MD2	MD2	…
Weekday	4	5	1	2	4	…
DL	13	20	9	10	10	…
DM	8	2	10	10	6	…
FU	8	4	8	8	9	...
RM	6	2	12	10	6	...
ER	4	0	2	2	2	…
PO	6	0	4	4	2	…
AN	0	0	0	0	0	…
AF	0	0	1	0	0	…
RO	0	0	0	0	0	…

The templates currently used by the Division of Pediatric Ophthalmology are relatively rigid, in the sense that slots allocated for a certain appointment type will, in general, only be allowed to be filled by a patient that fits the description. This can be viewed as a carve-out model, where capacity is rigidly carved out for certain appointment types. The only exceptions are emergencies (ER and, to a lesser extent, PO and RO appointments), where the urgency of the condition takes highest precedence and which can be overbooked into routine slots (see the following section).

One of the motivations for this modeling exercise was to see whether these rigid templates are in fact sufficiently efficient in handling stochastic demand, or whether a more

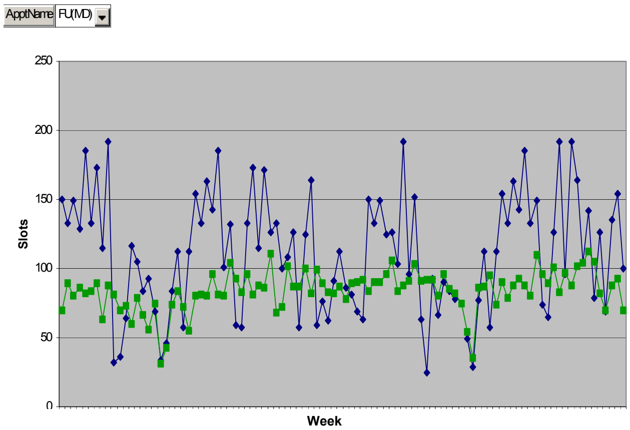

Figure 2: Weekly Supply (Blue Line) and Demand (Green Line) for Follow-up (FU) Slots for a 20 Month Period

sophisticated strategy which frees up capacity a few days in advance would result in better schedule utilization. We will report our conclusions in a future paper.

2.4 Scheduling Rules

As mentioned earlier, the Ophthalmology division currently uses a mixed manual/computerized scheduling system but is planning to move their scheduling operations to a centralized call center. Part of the motivation for this work was to be able to predict the effect that more rigid scheduling rules implemented in the call center software will have on overall patient flow in the clinic. Hence instead of trying to mimic the complexities and many arbitrary decisions made in manual scheduling, we specified an algorithm by which open slots are assigned to patient demands for appointments which could well be implemented in the call center.

The main criterion currently used in practice by the schedulers in this process is the level of urgency of the appointment. It determines the scheduling flexibility (i.e., the timeframe in which the appointment must be fulfilled), whether or not an appointment may be overbooked, and whether the appointment is specific to a particular provider or whether any available doctor should provide the necessary care. Table 3 summarizes our current implementation of scheduling rules, in order of decreasing urgency.

3 DATA SOURCES, IMPLEMENTATION AND VALIDATION

Our aim is to create as realistic a simulation model as reasonably possible. In order to populate our model we used 2-year historical data provided by the Division of Ophthalmology from the KIDS (Kids Inpatient Database System) hospital information system. In particular, we used the KIDS data to estimate empirical distributions for the following:

- The number of new patient calls requesting the various appointment types.
- The number of visits per patient for different appointment types
- The time to follow-up appointments for different diagnosis types.
- The overall proportion of commercial patients was estimated to be roughly 40%.
- No-show rates were estimated to be 5% for commercial patients whereas Medicaid/self-pay patients have a 20% no-show rate for new appointments and a 50% no-show rate for follow-up appointments.

Table 3: Overbooking/Scheduling Flexibility for Different Appointment Types

Appointment Type	Overbook if needed?	Scheduling Flexibility?	Provider Flexibility?
ER	Yes	No, same day only.	Yes, take first available
ROP	Yes	Yes, ±3 days	Yes, any MD..
PO	Yes	Yes, ±2 days	No.
New patients (routine)	No	First available	Yes (for most patients)
Follow-up patients (routine)	No	First avail. after desired date	No.

Only a few parameters had to be estimated for lack of data. The probability that a patient reschedules an appointment after a no-show was set to 50%, we subsequently used parameters like this number to fine-tune the simulation results to match observed historical behavior.

On the supply side we used the actual template schedules for the six providers in the clinic. This provides a very realistic picture of the supply side and revealed the (somewhat surprising) high variability in the number of total weekly available slots. This variance stems from the fact that the doctors are all involved in research, teaching and other activities as university faculty, which makes their presences in the clinic irregular. Additionally, vacations, holidays and travel are explicitly accounted for.

PSSM was implemented using a combination of tools. The stochastic arrivals and the patient flow logic was implemented in a straightforward way using the Arena 8.0 (Rockwell Software, West Allis, WI) simulation software package (see Figure 3). The scheduling process is implemented using a Microsoft Visual Basic module that queries and modifies a Microsoft Access database table with the doctor template schedules. Additionally, all patient appointments (whether realized or only scheduled) are recorded in a table which is built up during the run and contains a complete simulation record which can be analyzed.

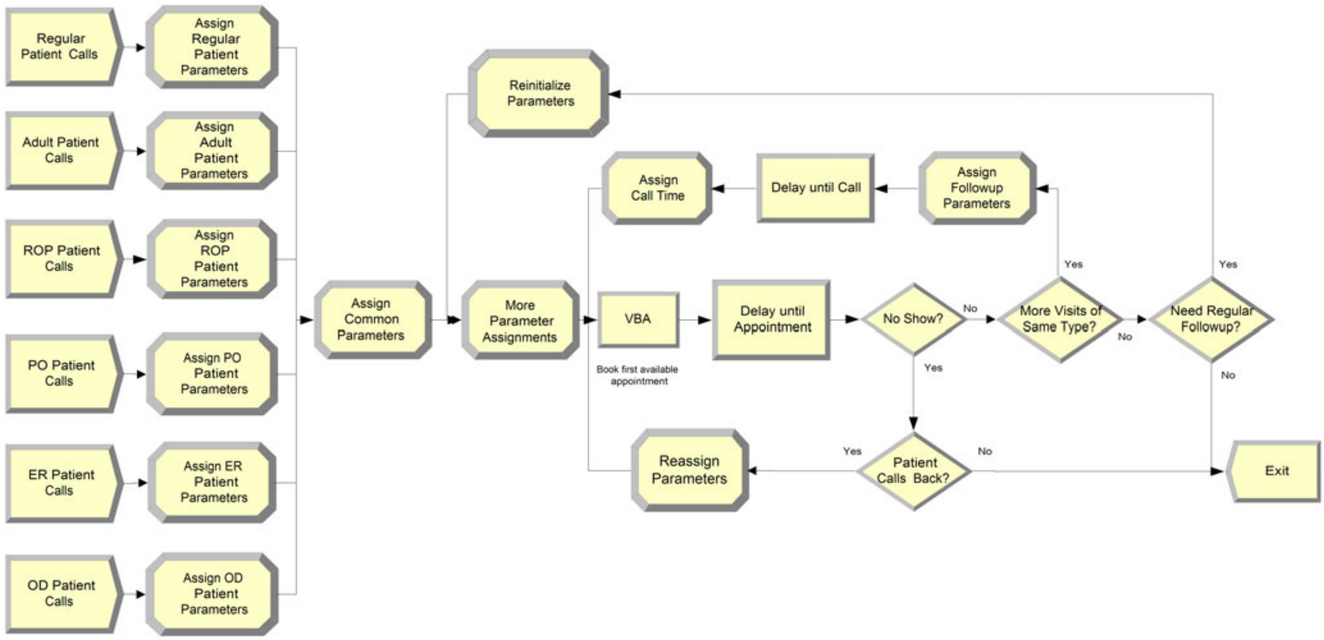

Figure 3: Arena Flow Chart for the Patient Scheduling Simulation Model

As a result, we are able to provide many rich analyses of the output by simply querying the database table containing all scheduled appointments over the period that the model was run.

Since many routine follow-up appointments are scheduled a year in advance we decided to let the system warm up for 15 months, that is we started our simulation runs in July 2001 but will only analyze the system behavior from September 2002 onwards. Sample runs on empty schedules confirmed that this is a reasonable warm-up period for the simulation.

Validating a complex model like PSSM is difficult. We verified that the total number of patients seen in the model in our reference timeframe of September 2002 – November 2003 was within 5% of the actual number of patients seen (roughly 14000), which is encouraging. Furthermore, the model confirmed that the optometrist was generally overbooked until August 2003, which is when a second optometrist was hired for relief (see also

Figure 6). Overall the resulting schedules have a realistic "feel" to them; once call center data becomes available we will be in an excellent position to apply statistical validation procedures to our model. For now we feel that our results will serve well as a reference baseline which can be used to benchmark different scheduling strategies.

4 ANALYSIS TOOLS

When evaluating a scheduling algorithm we decided to focus on three high-level characteristics of the resulting schedules that indicate how successful a given strategy is with respect to the most important goals of patient flow management.

As an institution with a public service mission and also in order to ensure patient satisfaction, the Division of Ophthalmology aims to provide prompt service to all patients who need to be seen by an ophthalmologist. As part of the "Pursuing Perfection" project at CCHMC, it has become stated policy that 95% of patients should be seen within a week of when they want to be seen. One benefit of the PSSM simulation model is that one can easily track waiting times in the system and monitor the 95th percentile of the resulting waiting times distribution for the various appointment types. Figure 4 shows the weekly maximum and average waiting times for a two year period for follow-

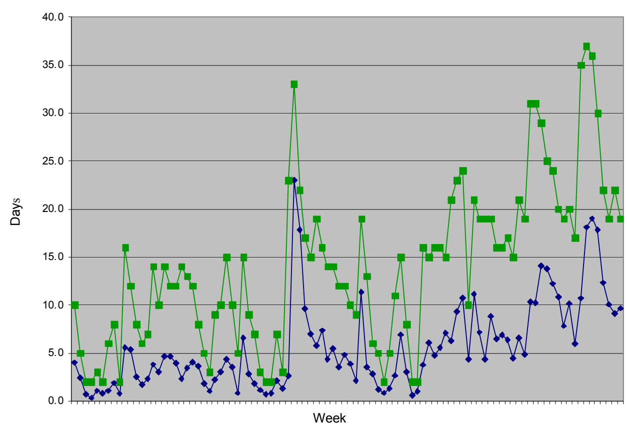

Figure 4: Maximum (Green) and Average Waiting Times for Follow-up (FU) Appointments for One Provider (MD)

up appointments with one particular provider. In this case we observe an increasing trend which calls for corrective action, e.g., by changing the template (e.g., by shifting capacity from another, underutilized appointment type).

As a second measure of the quality of a particular scheduling strategy we monitor the number of "busy" days (days with > 95% real utilization) and the number of "quiet" days (days with < 75% real utilization) for each provider (see Figure 5). This provides insight into how over-bookings affect day-to-day operations and whether the bottom line of clinic operations is affected by low capacity utilization (which in practice often lead to clinic cancellations).

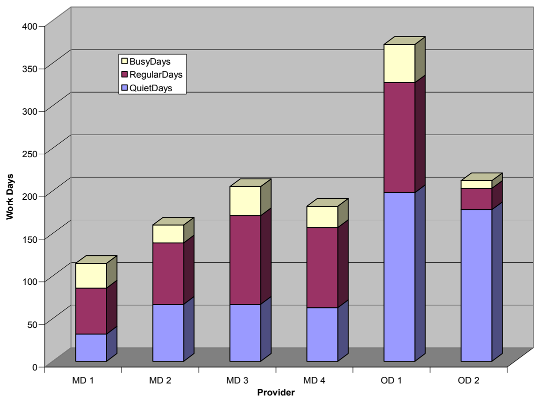

Figure 5: Breakdown of Types of Workdays for the Providers

Finally, we display overall utilization rates for each doctor and each appointment type. We distinguish between scheduled utilization (number of booked appointment slots divided by total number of available slots) and real utilization (number of patients that showed up for their appointment vs. maximum number of slots the providers expected to be filled). This is helpful for the design of templates as well as for overbooking strategies and to guide hiring decisions. Our analysis sheets allow for detailed looks at the fluctuations in scheduled vs. real utilization on a weekly basis, enabling us for example to fine-tune scheduling strategies in the event a doctor goes on an extended leave.

Figure 6 shows the real weekly utilization rates for an optometrists over a 2 year period. A second optometrist was hired in July 2003, reducing the workload to more acceptable levels and allowing for an expansion of business.

Figure 7 illustrates the significant variability in utilization of one particular appointment type (DL slots in this case). This variability is due to no-shows as well as fluctuations in both supply and demand and is also observed in practice. Utilization rates over 100% are due to routine overbooking of urgent patients.

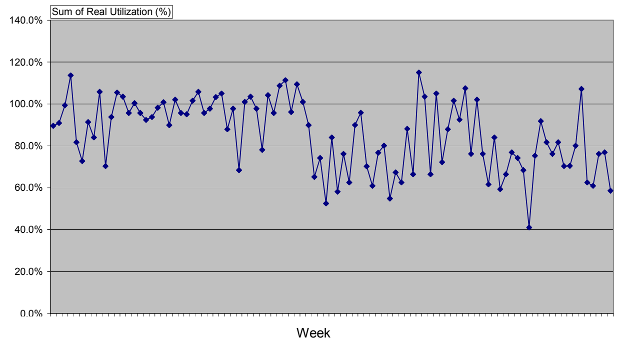

Figure 6: Real Utilization for One of the Two Optometrists

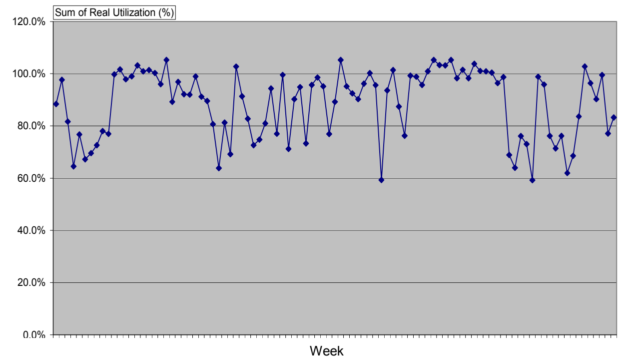

Figure 7: Weekly Average Real Utilization of DL Appointment Slots for all MDs

5 FUTURE WORK AND CONCLUSIONS

Our current implementation is still preliminary and specific to the Division of Ophthalmology. Once the division moves from manual scheduling to the new call-center-based scheduling system and implements rigid scheduling rules with the Tempus Software, as planned for later this year, we expect to gain access to significantly more detailed and better data, including data about how patients reschedule appointments and what their behavior is with respect to calling in for appointments. The move to a call-center based system will imply that the scheduling decisions done in real life will resemble those in our simulation model more closely since overbooking rules will be more closely enforced by systems software.

We expect to extend this model to other clinics (notably the Endocrinology and Gastroenterology Clinics) at Cincinnati Children's, and we expect new insights into the generalizability and extensibility of our model, ultimately resulting in a completely general scheduling simulation framework for outpatient clinics. We are also contemplating an addition of financial measures to further provide insight as to how different scheduling strategies are likely to impact the bottom line of clinic operations.

Patient scheduling is a crucial determinant of the flow through any medical clinic and as such an important influence on patient satisfaction, provider satisfaction and operational cost-effectiveness. There is a need for models that appropriately represent the complexities and dynamics involved in this process, and we believe that our PSSM system is a first step in this direction. By implementing a simulation platform we provide decision makers in clinics with a powerful test bed for optimizing scheduling strategies as well as a decision support tool to identify bottlenecks and to justify, e.g., hiring decisions. We foresee the final result to be a generic scheduling simulation platform with wide applicability.

REFERENCES

AHRQ website: <http://www.ahrq.gov/news/ulp/costs/ulpcosts.htm>

Harper, P.R. and Gamlin, H.M. (2003) Reduced outpatient waiting times with improved appointment scheduling: a simulation modeling approach. *OR Spectrum* 25:207-222

Ho, C.J. and Lau, H.S. (1992) Minimizing Total Cost in Scheduling Outpatient Appointments, *Management Science*, .38 (12): 1750-1764

Isken, M.W., Ward, T.J. and McKee, T.C. (1999) Simulating Outpatient Obstetrical Clinics, *Proceedings of the 1999 Winter Simulation Conference*, ed. P.A. Farrington, H.B. Nembhard, D.T. Sturrock, and G.W. Evans,,1557-1563. Piscataway, New Jersey.

McManus, M.L., Long, M.C., Cooper, A., Mandell, J., Berwick, D.M., Pagano, M. and Litvak, E. (2003). Variability in Surgical Caseload and Access to Intensive Care Services, *Anesthesiology*, 98**:**6

AUTHOR BIOGRAPHIES

MING GUO is a statistician in the Human Resources Department at Cincinnati Children's Hospital Medical Center. Her interests include process modeling strategies for the health care industry. She received a Master of Business Administration from the University of Akron in 2001.

MICHAEL WAGNER is an assistant professor of Biomedical Informatics in the Cincinnati Children's Hospital Research Foundation. His research interests include the application of operations research techniques to hospital operations, as well as bioinformatics and machine learning. He received a Ph.D. in operations research from Cornell University in 2000.

CONSTANCE WEST is the director of the Division of Pediatric Ophthalmology and an Associate Professor of Pediatrics at Cincinnati Children's Hospital. She obtained her MD degree from University of Massachusetts Medical School.