Submitted to *Operations Research*
manuscript

Adaptive Design of Clinical Trials: A Sequential Learning Approach

Zhengli Wang
Stanford Graduate School of Business, Stanford, CA 94305, wzl@stanfrod.edu

Stefanos Zenios
Stanford Graduate School of Business, Stanford, CA 94305, stefzen@stanford.edu

Adaptive design of clinical trials hold great promise for reducing the cost of clinical trials without affecting their type I or type II error. In this paper, we propose a tractable Bayesian framework that allows the derivation of optimal policies: the clinical trial designer can either choose among different experiments with different information, or to terminate the trial; there are rewards associated with making the correct decision and penalties from incorrect ones. We show that the log-likelihood ratio (LLR) converges to a diffusion process via a limiting approximation in which the distributions under the null and the alternative converge in symmetric Kullback–Leibler (KL) divergence. The resulting asymptotic stochastic control problem is then solved analytically. We present numerical results based on a real-world clinical trial setting (the Phoenix Champion Trial), in which we compare the asymptotically optimal policy derived with our framework to simpler policies that either have been used in practice or are optimized using our framework. The result shows that in a wide range of scenarios, the asymptotically optimal policy outperforms all the other policies considered. In particular, across all scenarios considered, the percentage improvement of economic benefit of the asymptotically optimal policy versus the adaptive policy used in practice ranges from 0.4% to 8.7% (with a median of 4.7%). When comparing the asymptotically optimal policy versus optimized two-stage or multi-period policies, the percentage is smaller but still significant (with a median of 3.2% and 1.0% respectively). The results suggest that optimized multi-period adaptive clinical trial designs can significantly improve the economic value generated by such trials.

Key words: clinical trial, adaptive design, sequential hypothesis testing, healthcare, stochastic control.

1. Introduction

Clinical trials have been central to the improvement of healthcare over the past decades. As part of the R&D process required to gain regulatory approval, firms engage in clinical trials on human subjects to evaluate the safety and efficacy of drugs or treatment for disease. The trials are costly: it is estimated that the average total cost of the clinical trial needed to secure regulatory approval reaches $1.46 billion and only about 11.8% of drugs entering clinical testing are eventually approved (DiMasi et al. (2016)). Consequently, there has been immense pressure on pharmaceutical firms and regulatory bodies alike to reduce costs associated with each trial while maintaining the same level of accuracy.

A solution is adaptive clinical trial design, which holds great promise for reducing the cost of clinical trials (Zang and Lee (2014), Kruizinga et al. (2019)). An adaptive trial involves multi periods and the sample size or even the experimental design can be adjusted between periods based on the strength of the evidence collected. The advantages of adaptive clinical trials are discussed widely (Pallmann et al. (2018), Ryan et al. (2020)) and regulatory bodies such as the US FDA are interested in the use of adaptive trials and encourage the development of methodologies to support their design and adoption in practice.

In this paper, we propose a framework to guide the development of an optimal adaptive clinical trial design and to assist trial designers on how they could use experiments to dynamically conduct the trials. The framework is a Bayesian decision-theoretical one that builds on the sequential hypothesis testing paradigm: in each period, the trial designer can choose to either terminate the trial or continue for one more period and can select among several alternative experiments to perform. There are three novel features in our framework: a) different experiments are allowed in each period, b) the log-likelihood ratio is approximated by a diffusion process that represents an asymptotic environment in which the experiments are becoming increasingly uninformative (in the sense that the distributions under the null and the alternative converge in symmetric Kullback–Leibler (KL) divergence), and c) the optimal solution to the resulting stochastic control problem is derived

analytically. The framework allows specified costs and rewards for erroneously and correctly making decisions. The optimal policy, which identifies the optimal choice of experiment in each period and the optimal boundaries for terminating the trial and either accepting or rejecting the null, is derived. The resulting optimal type I and II errors can be derived from the optimal boundaries. While previous papers have combined Bayesian decision analysis and adaptive design methods (Cheng et al. (2003), Willan and Kowgier (2008), Ahuja and Birge (2016)), these had to rely on simulations for more than two periods due to their models' intractability. In contrast, this paper generates tractable analytical results for multiple periods.

We also present numerical results based on a real-world clinical trial setting (the Phoenix Champion Trial), in which the asymptotically optimal policy derived from our framework is compared to simpler policies that have either been used in practice or are optimized using our framework. The result shows that in a wide range of scenarios, the asymptotically optimal policy outperforms all the other policies considered. This happens when the prior lies in the intermediate range between the boundaries set by the various alternatives, the difference between the null and the alternative is moderate, and the population baseline adverse rate is not too small. The results also show that there are some scenarios where the asymptotically optimal policy performs poorly because the asymptotic approximation is not valid.

Our Contributions. Our paper makes the following contributions. First, we present an analytically tractable Bayesian formulation of the multi-period clinical trial design problem with multiple types of experiments. Second, we demonstrate that the log-likelihood ratio of the multi-period problem converges to a diffusion process, whose drift and variance depend on the underlying null and alternative hypothesis; a similar asymptotic result was derived in Araman and Caldentey (2021), but our results involve a strictly weaker assumption. Third, we show that under the diffusion approximation, the problem of multi-period clinical trial design becomes a tractable stochastic control problem, for which we analytically derive the optimal solution. This formulation generalizes the reward function in the models of Harrison and Sunar (2015) and Kwon and Lippman (2011).

Adapting the solution approach from Harrison and Sunar (2015), we show that this generalization leads to a broader class of optimal policies, which involve using more informative controls when we are more certain, and using less informative controls when we are less certain. Fourth, we show how our formulation can be adapted to accommodate constraints on the type I error, the type II error and the expected termination time. Fifth, we validate our framework using a real-world clinical trial setting, in which we compare the asymptotically optimal policy to simpler policies that either have been used in practice or are optimized using our framework, and identify the wide range of scenarios that the asymptotically optimal policy has superior performance.

The Organization of the Paper. §2 provides a review of the literature and highlights the paper's contribution. §3 formulates the multi-period clinical trial problem. §4 demonstrates that the log-likelihood ratio converges to a diffusion process via a limiting approximation in which the designer runs a series of increasingly less informative experiments. §5 shows that in the limiting approximation, the continuous-time problem is tractable with an analytically derived optimal solution. §6 shows that our formulation can be adapted to accommodate constraints on the type I and II errors as well as on the expected termination time. §7 performs a numerical study based on a real-world clinical trial setting and §8 concludes.

2. Literature Review

Our paper builds on three streams of literature: sequential hypothesis testing and adaptive clinical trial design, Bayesian decision analysis, and stochastic control in applied probability (or more broadly, decision theory under uncertainty). It also contributes to the literature on learning in OM (see Figure 1). In this section, we will review the most relevant papers in each literature stream, and highlight several key papers that lie on the interface of these streams.

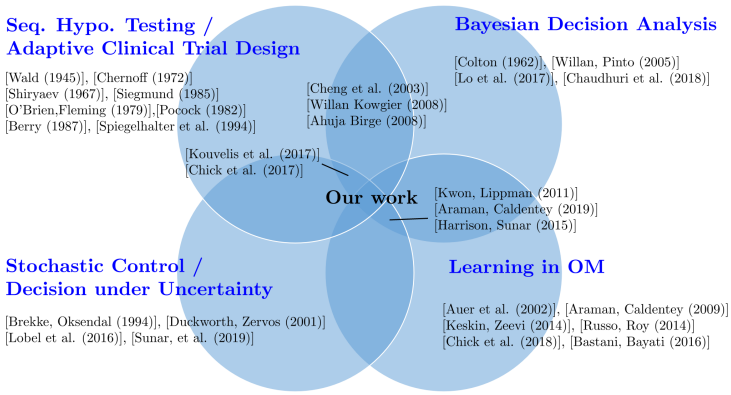
Sequential Hypothesis Testing and Adaptive Clinical Trial Design. The literature on sequential hypothesis testing dates back to Wald (1945), who formulated the discrete-time version of the sequential hypothesis testing framework (and Shiryaev (1967) and Siegmund (1985) developed the continuous-time versions). Chernoff (1959) proposed an asymptotic policy where in each period

the decision maker determines which state is more likely by maximum likelihood estimates, and then select the control with the highest expected KL-information number based on that state. More recently, there are papers that propose asymptotic policies to refine Chernoff (1959)'s. One such paper is Naghshvar and Javidi (2013), in which the proposed policies have a similar structure and consist of two phases: in the early phase select experiments that better discriminate all hypothesis pairs, in the latter phase when the posterior belief of a hypothesis passes some threshold, select experiments that favor that hypothesis. Relative to Chernoff (1959) and Naghshvar and Javidi (2013), our paper is different in the following: 1). our model can incorporate discounting in the objective whereas theirs cannot; 2) our experiment costs may be different for each experiment, whereas in their settings all experiments have the same costs; 3) our model have both rewards and penalties whereas theirs only have penalties (which they call "risks"); 4) our optimal policy can be analytically characterized whereas their policies cannot.

The sequential hypothesis testing framework has been widely applied in clinical settings, which evolved into the vast literature on adaptive design of clinical trials. The adaptive design methods were developed to address the issue of fixed-sample size in the conventional approach, as they allow greater flexibility and can be applied to multi-period clinical trial settings. The adaptation can be based on a frequentist approach (Pocock (1982), O'Brien and Fleming (1979), Gordon Lan and DeMets (1983), Pawitan and Hallstrom (1990)) or on a Bayesian one (Berry (1987), Spiegelhalter et al. (1994)). The flexibility of these adaptive design methods can lead to many benefits such as smaller expected sample sizes (Pocock (1982), O'Brien and Fleming (1979)), flexibility in adjustment of each period's sample sizes (Gordon Lan and DeMets (1983)), inference without knowing the test statistics (Pawitan and Hallstrom (1990), and the ability to monitor trials Spiegelhalter et al. (1994)). However, in most of these papers the type I and II errors need to be specified in advance. This is contrary to our framework, in which these errors are endogenous variables in the model.

Bayesian Decision Analysis. Another very relevant stream of literature is the one on Bayesian decision analysis, which is a well-known quantitative framework advocated and recommended by

Figure 1 The multiple streams of literature this paper builds on.



the U.S. Food and Drug Administration (Food and Administration (2010)). The Bayesian decision analysis framework computes the optimal type I and type II errors using cost-minimization, utility-maximization or other similar objectives, thus foregoing the need to specify arbitrary type I and type II errors as in the conventional approach. Notable papers include Colton (1963), Willan and Pinto (2005), Montazerhodjat et al. (2017) and Chaudhuri et al. (2018). However, these papers focus on only one period of optimization and do not address clinical trials that are conducted in multiple periods, which is often the case in adaptive trials. Moreover, most of these papers do not present an analytical form for the optimal type I and type II errors. The framework in our paper addresses multi-period clinical trials and presents analytical expressions for the optimal type I and type II errors. While there are papers similar to ours that combine Bayesian decision analysis and adaptive design methods (Cheng et al. (2003), Willan and Kowgier (2008), Ahuja and Birge (2016)), most of these papers rely on simulations for more than two periods due to their models' intractability, while our framework leads to analytical results for the multi-period case. In addition, unlike these earlier models, our model also allows discounting.

Stochastic Control in Applied Probability and Decision Theory under Uncertainty. Our paper also builds on the literature in applied probability involving stochastic control and the literature on decision theory under uncertainty. We leverage the techniques used from this stream of literature to solve complex stochastic control and optimal stopping problems. Classical papers include [(Brekke and Øksendal (1994))], Davis and Zervos (1994), Karatzas and Sudderth

(1999) and Duckworth and Zervos (2001). More recently, such techniques have been employed to solve problems in Operations Research and Management Science. For instance, Lobel et al. (2016) analyzes the problem where a firm optimizes the launch times and the price of their product assuming the technological process follows a diffusion process, Matoglu et al. (2015) analyzes the problem of managing capacity by modeling the inventory control problem as a Brownian drift control problem, and Sunar et al. (2019) studies how to develop and price a product optimally by controlling the product launching time. Most of these papers begin by assuming the underlying process to be a geometric or standard Brownian motion, and derive their results based on this assumption. A notable feature of our paper is that we provide a generating mechanism for the underlying diffusion process (or observation process), besides deriving the results generated from this process. This offers a more comprehensive and holistic approach to addressing the statistical problem of interest.

Papers that Lie On the Interface. Kouvelis et al. (2017) combines adaptive clinical trial design and Bayesian theoretical decision-making, and uses a diffusion process to approximate the patient enrollment level. The authors analyze a clinical trial design problem for drug development, where a firm needs to decide in each period the patient recruitment rate and how many sites to open. Similar to the formulation in our framework, their objective is to maximize the total expected economic benefit. Our modeling framework differs from theirs in three ways. First, the total patient sample size is fixed in their case while in ours it is not. Second, their control at each period is the patient enrollment rate, while ours can be very general and can represent more complex decisions or experiments. Third, they do not continuously update their beliefs, and their optimal control at the beginning of each period does not depend on trial outcomes in the previous periods, while this is not the case in our model.

Anderer et al. (2022) suggests a novel Bayesian adaptive clinical trial design that combines data from both surrogate and direct outcomes to improve decision making, aiming to address the limitations of using direct outcomes alone in drug approval decisions, with potential significant

cost reduction and maintaining similar error rates. Our modeling framework differs from theirs in the following ways. First, the structure of the optimal policy in our framework can be explicitly characterized. Second, our framework applies to clinical trials where the outcome types are binary or continuous valued, but not time-to-event ones. Third, we consider discounting in the objective, while their framework abstracts away from discounting.

Chick et al. (2017) uses Bayesian theoretical decision-making in an adaptive trial design setting. Similar to our framework, they use a diffusion process to approximate state evolution, which represents the relative benefit of new treatment compared to the conventional one. The optimal stopping boundary can be obtained by solving the heat equation. Again the main difference is that the control in their settings represents only sampling allocation while ours can be very general. Moreover, our model allows for the control of type I and II errors as well as expected termination time.

There are three papers which lie on the interface of sequential hypothesis testing and stochastic control that are very relevant to ours. Kwon and Lippman (2011) explores when it is optimal for a firm to abandon or invest in a Bayesian sequential decision setting. Relative to their model, we allow multiple experiments in our framework. Araman and Caldentey (2021) analyzes a sequential sampling problem with applications in assortment selection and new product introduction. They develop a diffusion approximation regime similar to ours. Relative to their model, our approximation allows strictly weaker assumptions (see §Appendix EC.4) and we allow each experiment to be associated with a cost, which results in a non-static optimal policy. Harrison and Sunar (2015) studies a problem with multiple controls involving profit maximization of a firm, and our model heavily utilizes and generalizes their framework. Our extensions relative to Harrison and Sunar (2015) are that we show how the diffusion process they assumed is a result of an asymptotic approximation of the LLR test in a so-called *heavy experimentation regime* (see §4), and we generalize their reward function, which leads to a completely new class of optimal policies that require different proof techniques. More specifically, both the set of controls that may be used in

the optimal policy and the nature of the optimal policy change. Interested readers may find the relevant details in the Electronic Companion (§Appendix EC.5).

Learning in Operations Management (OM). Lastly, our paper is also connected to the literature on learning theory in OM, where policies for adaptive design problems on bandits are extensively analyzed. Lai et al. (1985) and Auer et al. (2002) represent two seminal studies. More recent work includes Araman and Caldentey (2009), Keskin and Zeevi (2014), Russo and Van Roy (2014), Bastani and Bayati (2020), Chick et al. (2022). Most of these papers propose and analyze policies that guarantee good performance only when the time horizon is long, because in many scenarios an exact optimal solution to the problem is nearly impossible to find. Unlike these papers, the optimal policy characterized in our problem is an exact one that is independent of time horizon.

3. Problem Formulation

Let $\theta \in \{0, a\}$ represent the true state of the world. A decision maker is testing between two hypotheses $H_0 : \theta = 0$ and $H_a : \theta = a$. At each time period, the decision maker can either stop and accept one of the hypotheses, or incur some cost to run an experiment and obtain signals that will inform him or her about θ. There are N types of experiments, or controls, each associated with a different cost. Control j costs $c(j)$ (or abbreviated as c_j) per unit time. Throughout the paper, we will use the terms "experiment" and "control" interchangeably.

To be more precise, let $e_1, e_2, ...$ be the sequence of controls chosen by the decision maker, $X_1, X_2, ...$ be the sequence of signals acquired, and $\mathcal{S} = \{1, 2, 3, ..., N\}$ denote the control set. Given a control $e_i \in \mathcal{S}$, $X_i \sim p_0(\cdot|e_i)$ or $p_a(\cdot|e_i)$ if the true state of the world is 0 or a respectively. We assume the X_i's are independent across time periods, and only attain values on a discrete set Ω (which we will call the signal set). Without loss of generality we assume $p_\theta(x|e) > 0$ for all $x \in \Omega, e \in \mathcal{S}$ and $\theta \in \{0, a\}$. Let T denote the non-anticipating stopping time (we will also call it the stopping rule or termination time), $r \in (0, 1)$ the discount rate and $\pi_0 = P(\theta = a)$ the decision maker's prior belief. Depending on whether the final decision is correctly or incorrectly made, there are pre-determined

rewards or penalties. Assume the decision maker makes a terminal decision $\hat{\theta} \in \{0, a\}$ (and recall the true state of the world is $\theta \in \{0, a\}$). The terminal payoff is given by $H(\theta, \hat{\theta})$, where

$$H(\theta, \hat{\theta}) = \begin{cases} R_0 \text{ , if } \hat{\theta} = \theta = 0, \\ R_a \text{ , if } \hat{\theta} = \theta = a, \\ -\kappa_0 \text{ , if } \hat{\theta} = 0, \theta = a, \\ -\kappa_a \text{ , if } \hat{\theta} = a, \theta = 0. \end{cases}$$

The decision maker maximizes the total expected economic benefit

$$E\left[\left(\sum_{i=1}^{T} -\frac{c(e_i)}{(1+r)^{i-1}}\right) + \frac{H(\theta, \hat{\theta})}{(1+r)^T}\right], \tag{1}$$

where the first term represents the cumulative experimental costs and the second represents the terminal reward.

Remark 1. In a clinical trial setting, different experiments can correspond to different patient sample sizes (Chaudhuri et al. (2018)). Given a particular sample size (i.e. an experiment), half of the patients are allocated to the treatment group and the other half to the control group but this can be generalized to uneven allocations. The signal from the experiment will be the difference of outcomes between the two groups.

In certain scenarios, the decision maker may also wish to control for the type I and II errors and the expected termination time. For the convenience of presenting the result, we will first focus on maximizing (1) without considering these complications, and then show how our formulation can be adapted to accommodate these constraints.

Given the decision maker's policy $\{e_i\}$ and the resulting signals $\{X_i\}$, we denote the log-likelihood ratio (LLR)

$$L(k) = \sum_{i=1}^{k} \ln \frac{p_a(X_i|e_i)}{p_0(X_i|e_i)}, \tag{2}$$

and denote $L_j(k)$ to be the corresponding quantity when $e_i = j$ for all $1 \le i \le k$, i.e.

$$L_j(k) = \sum_{i=1}^{k} \ln \frac{p_a(X_i|j)}{p_0(X_i|j)}. \tag{3}$$

We also introduce the following notations associated with the LLR.

$$\mu_0(j) = E\left[ln\frac{p_a(X|j)}{p_0(X|j)} \;\middle|\; H_0\right],\ \mu_a(j) = E\left[ln\frac{p_a(X|j)}{p_0(X|j)} \;\middle|\; H_a\right], \tag{4}$$

$$\sigma_0(j) = \mathrm{SD}\left[ln\frac{p_a(X|j)}{p_0(X|j)} \;\middle|\; H_0\right],\ \sigma_a(j) = \mathrm{SD}\left[ln\frac{p_a(X|j)}{p_0(X|j)} \;\middle|\; H_a\right], \tag{5}$$

and note that $\mu_a(j) > 0 > \mu_0(j)$ due to Gibbs' inequality.

Remark 2. In many clinical trial settings, the decision maker keeps track of a different test statistics such as the difference in outcomes between patients in the treatment group and control group. This is equivalent to keeping track of the log-likelihood ratio (see §Appendix EC.7 for more details).

4. Asymptotic Approximation: the Limiting Control Problem

As the discrete nature of the formulation renders our problem intractable, we will adopt a continuous-time approximation of the LLR process. We will first present a high-level intuition of the idea and then design a sequence of systems to show rigorously how the LLR process weakly converges to a diffusion process.

High-level Intuition. The key idea is to approximate $L(k)$, $L_j(k)$ with diffusion processes $L(t)$, $L_j(t)$ with matched 1^{st} and 2^{nd} moments under H_0 and H_a. That is, informally for any experiment j,

$$\text{under } H_0,\ L_j(t) \approx B(t; \mu_0(j), \sigma_0^2(j)), \tag{6}$$

$$\text{under } H_a,\ L_j(t) \approx B(t; \mu_a(j), \sigma_a^2(j)), \tag{7}$$

where $B(t; \mu, \sigma^2)$ represents a Brownian motion with drift μ and variance σ^2.

In fact, it turns out that we can prove a stronger result in which the drift and variance of the diffusion process under the null and alternative are related to each other. Specifically, if we let for each control j, $\eta_j = \mu_a(j) - \mu_0(j)$ (which we will call the *information quality*), then we will show that $L_j(t)$ can be approximated by $B(t; -\frac{\eta_j}{2}, \eta_j)$ and $B(t; \frac{\eta_j}{2}, \eta_j)$ under H_0 and H_a respectively.

Formal Presentation. To present the approximation more formally, we consider a *heavy experimentation* regime in which there is a sequence of closely related systems indexed by K. In the K-th

system, in each period instead of performing 1 experiment, the decision maker sequentially runs K less informative ones, and as $K \to \infty$, the LLR converges to the approximating diffusion model. In other words, in the K-th system, each original experiment is replaced with K independent and identically distributed ones. In each of these new experiments, the corresponding probability distributions that generate the signals are denoted by $p_0^{(K)}(\cdot \mid j)$ and $p_a^{(K)}(\cdot \mid j)$ respectively, and we define $\mu_0^{(K)}$, $\mu_a^{(K)}$, $\sigma_0^{(K)}$, $\sigma_a^{(K)}$ accordingly as in (4) and (5), and $L^{(K)}(k)$, $L_j^{(K)}(k)$ as in (2) and (3). With a slight abuse of notation, we also define the continuous-time log-likelihood ratio processes $L^{(K)}(t) = L^{(K)}(\lfloor Kt \rfloor)$, $L_j^{(K)}(t) = L_j^{(K)}(\lfloor Kt \rfloor)$ and their associated limiting counterparts $L(t) = \lim_{K\to\infty} L^{(K)}(t)$, $L_j(t) = \lim_{K\to\infty} L_j^{(K)}(t)$ (where the limit represents weak convergence, and the existence will be shown in Theorem 1). We will assume that for each $j \in \mathcal{S}$, as $K \to \infty$ the two distributions converge to each other in the following manner

$$\lim_{K\to\infty} K\left[\mu_a^{(K)}(j) - \mu_0^{(K)}(j)\right] = \eta_j, \tag{8}$$

$$\frac{p_a^{(K)}(x|j)}{p_0^{(K)}(x|j)} \to 1, \text{ for all } x \in \Omega. \tag{9}$$

(8) essentially says that the K experiments which replace the original one in the K-th system generate the same information in terms of the difference in total expected log-likelihood ratios under the alternative and the null, and (9) says that the LLR converges to 1 (i.e. the signals become increasingly uninformative). We note that (8) and (9) are conditions that are strictly weaker than Assumption 1 of Araman and Caldentey (2021) (a rigorous statement can be seen at §Appendix EC.4). Under (8), (9) we can show their first two moments satisfy the following asymptotic behavior (c.f. lemma EC.1):

$$\lim_{K\to\infty} K\mu_0^{(K)}(j) = -\frac{\eta_j}{2}, \ \lim_{K\to\infty} \sqrt{K}\sigma_0^{(K)}(j) = \sqrt{\eta_j}, \tag{10}$$

$$\lim_{K\to\infty} K\mu_a^{(K)}(j) = \frac{\eta_j}{2}, \ \lim_{K\to\infty} \sqrt{K}\sigma_a^{(K)}(j) = \sqrt{\eta_j}. \tag{11}$$

Define an *admissible strategy* as a non-anticipating right-continuous process $M = \{M_t, t \geq 0\}$ taking values in $\mathcal{S}$, then we have the following:

THEOREM 1. *Suppose for all experiments $j \in \mathcal{S}$ for each experiment j, we construct a sequence of experiments $\{p_0^{(K)}(\cdot \mid j),\ p_a^{(K)}(\cdot \mid j)\}_{K=1}^{\infty}$ such that (8) and (9) hold. Then for any $j \in \mathcal{S}$,*

$$\text{under } H_0,\ L_j^{(K)}(t) \Rightarrow B(t; -\frac{\eta_j}{2}, \eta_j) \text{ as } K \to \infty, \tag{12}$$

$$\text{under } H_a,\ L_j^{(K)}(t) \Rightarrow B(t; \frac{\eta_j}{2}, \eta_j) \text{ as } K \to \infty, \tag{13}$$

where $B(t; \mu, \sigma^2)$ represents a Brownian motion with drift μ and variance σ^2. Moreover, given any admissible strategy $M = \{M_t, t \geq 0\}$, the log-likelihood ratio process $L(t)$ is determined by

$$L(0) = 0,\ L(t) = \int_0^t \mu_\theta(M_s)ds + \sigma(M_s)dB_s, \tag{14}$$

where $\mu_a(j) = \frac{\eta_j}{2}, \mu_0(j) = -\frac{\eta_j}{2}$ and $\sigma(j) = \sqrt{\eta_j}$.

The result of Theorem 1 provides the underlying mechanism for the initial assumptions introduced in a diverse stream of literature, including Shiryaev (1967), Siegmund (1985), Peskir and Shiryaev (2006), Kwon and Lippman (2011), Harrison and Sunar (2015) (henceforth referred to as "HS"), Dyrssen and Ekström (2018) and Henry and Ottaviani (2019). It essentially says that when experiment j is used, $L(t)$ will evolve as a diffusion process approximately with variance η_j and drift $-\frac{\eta_j}{2}$ or $\frac{\eta_j}{2}$ under the null or the alternative. A detailed proof of Theorem 1, which employs the tool of Functional Central Limit Theorem (FCLT), is presented in the Electronic Companion (§Appendix EC.1).

In the remainder, we will refer to η_j as the *information quality* of experiment j, and we will say that an experiment with a higher *information quality* is more informative. Moreover, unless otherwise specified, we will use the terms $L(t)$ or L_t interchangeably.

5. Formulation in Continuous Time

Recall that the decision maker's *prior belief* is $\pi_0 = P(\theta = a) \in (0, 1)$. Denote $\mathcal{F}_t^L$ as the filtration generated by L, and define the *posterior belief* $\pi_t = P(\theta = a | \mathcal{F}_t^L)$. We have the following (equivalent to Proposition 2 in Araman and Caldentey (2021)).

LEMMA 1. *The posterior belief satisfies the stochastic differential equation*

$$d\pi_t = \sqrt{\eta(M_t)}\pi_t(1-\pi_t)dB_t, \tag{15}$$

where $B = \{B_t, t \geq 0\}$ is a standard Brownian motion with respect to $\{\mathcal{F}_t^L, t \geq 0\}$. An *admissible policy* is defined as a pair of (M, T) where M is an admissible strategy and T is a finite stopping time with respect to $\mathcal{F}_t^L$. Because the posterior belief captures all the information that is needed for the decision maker, we restrict the set of admissible policies under consideration to be Markovian (i.e. M_t is a function of π_t), in the forms of a natural class of policies called *interval policies*.

DEFINITION 1. A set $\Pi = \{M, m(\cdot), \{i_k\}_{k=0}^{D}, \{\zeta_k\}_{k=0}^{D+1}, l, u\}$ is an *interval policy* (see Figure 2) if

$$M = \{M_t, t \geq 0\} \text{ is an } \textit{admissible strategy} \text{ s.t. } M_t = m(\pi_t), \tag{16}$$

$$0 < l = \zeta_0 < \zeta_1 < ... < \zeta_D < \zeta_{D+1} = u < 1, \tag{17}$$

$$m(y) = \begin{cases} i_k, \text{ if } y \in [\zeta_k, \zeta_{k+1}), i_k \in \mathcal{S}, k = 0, ..., D, \\ STOP, \text{ if } y \in (0, l) \cup [u, 1), \end{cases} \tag{18}$$

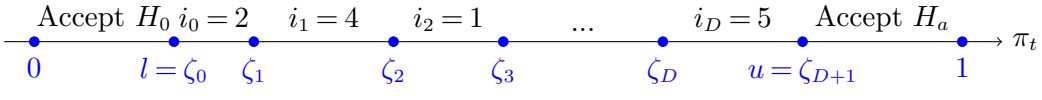
$$i_k \neq i_{k+1}, \ k = 0, ..., D, \tag{19}$$

where we call ζ_k's the *switching beliefs*, i_k's the *intermediate controls*, l and u the *lower* and *upper critical belief*, and $m(\cdot)$ the *control function*, respectively. An *interval policy* is associated with the natural stopping rule

$$T = \inf\{t \geq 0 \mid \pi_t \in (0, l) \cup [u, 1)\}. \tag{20}$$

In other words, an *interval policy* is a policy where the decision maker divides the interval $[l, u)$ into $D+1$ sub-intervals for some D by $\{\zeta_k\}_{k=0}^{D+1}$ such that the same control i_k is employed whenever $\pi_t \in [\zeta_k, \zeta_{k+1})$, $k = 0, 1, ..., D$ (see Figure 2). The decision maker stops and accepts H_0 or H_a whenever π_t reaches below l or above u, respectively.

Figure 2 Illustration of an interval policy.



Given an interval policy Π with the associated stopping rule T, the decision maker's economic benefits (the continuous-time version of (1)) are defined by

$$E\left[e^{-\lambda T}G(\pi_T) - \int_0^T e^{-\lambda t}c(M_t)dt \;\middle|\; \pi_0\right], \tag{21}$$

where $\lambda = -ln(\frac{1}{1+r})$ is the corresponding continuous-time discount rate and

$$G(y) = \max\{-\kappa_0 y + R_0(1-y), R_a y - \kappa_a(1-y)\} \tag{22}$$

represents the conditional expected net reward when the decision maker stops. Intuitively, $G(\cdot)$ can be interpreted as follows. Suppose the decision maker stops at time T (with a posterior belief π_T), the expected terminal payoff from accepting H_0 or H_a is $-\kappa_0\pi_T + R_0(1-\pi_T)$ or $R_a\pi_T - \kappa_a(1-\pi_T)$ respectively, and he or she will choose the maximum of the two. The optimization problem resembles that of HS, with the important difference that the boundary condition $G(\cdot)$ is different due to the presence of R_0 and $-\kappa_0$.

The decision maker wants to maximize the objective (21). We define the decision maker's optimal profit function (or value function) over $(0,1)$ as

$$V(y) = \max_{\Pi} E\left[e^{-\lambda T}G(\pi_T) - \int_0^T e^{-\lambda t}c(M_t)dt \;\middle|\; \pi_0 = y\right]. \tag{23}$$

5.1. The Ordering of Controls

We are now in a position to present the optimal policy. At this point, the first question that the reader may wonder is, out of the N possible controls available to the decision maker (and N can be very large), how we are going to determine which controls will potentially be used. It turns out that only a subset of the N controls will be used in the optimal policy. This subset constitutes what we will call *the efficient frontier*, the notion of which was first introduced in HS. In the remainder of the section, we will first demonstrate how we can find *the efficient frontier* in our framework, and then illustrate how we can characterize an optimal interval policy from it. The *efficient frontier* in our model is different from that of HS, and we compare and contrast them in detail in the Electronic Companion (§Appendix EC.5).

Recall that each control $i \in \mathcal{S}$ has a corresponding cost rate c_i and information quality η_i, and we can represent the set of controls by $\{(\eta_i, c_i)\}_{i=1}^{N}$. It is obvious that there exists an index $\tilde{N} \leq N$ and a numbering of the controls such that (see Figure 3)

$$0 < \eta_1 < ... < \eta_{\tilde{N}}, \tag{24}$$

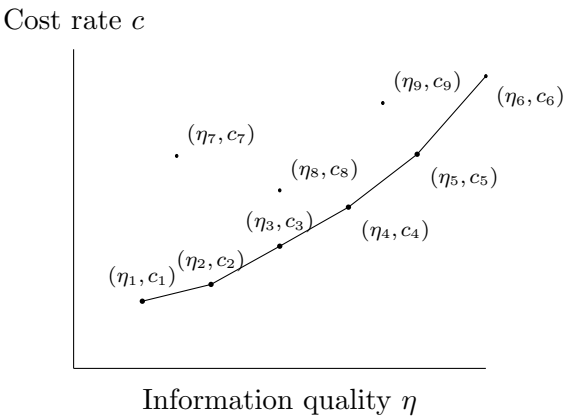
$$\frac{c_2 - c_1}{\eta_2 - \eta_1} < \frac{c_3 - c_2}{\eta_3 - \eta_2} < ... < \frac{c_{\tilde{N}} - c_{\tilde{N}-1}}{\eta_{\tilde{N}} - \eta_{\tilde{N}-1}}, \tag{25}$$

$$c_i = \phi(\eta_i), \text{ for } i = 1, 2, ..., \tilde{N}, \tag{26}$$

$$c_i \geq \phi(\eta_i), \text{ for } i = \tilde{N}+1, ..., N, \tag{27}$$

where $\phi(\cdot)$ is the strictly increasing, piece-wise linear and convex function that connects (η_1, c_1), ..., $(\eta_{\tilde{N}}, c_{\tilde{N}})$. In other words, $\phi(\cdot)$ is *the efficient frontier* and an example of it is shown in Figure 3 (in which $\tilde{N} = 6$ and $N = 9$). Controls $1, 2, ..., \tilde{N}$ are either the left endpoint, the right endpoint, or a point on *the efficient frontier* at which the slope changes.

Figure 3 An example of $N = 9$ controls plotted on the (η, c)-plane with control i represented by (η_i, c_i).



Note. In this example, $\tilde{N} = 6$ and the solid line represents the *efficient frontier*, i.e. the piecewise linear function $\phi(\cdot)$ that satisfies (26) and (27).

Intuitively, *the efficient frontier* consists of controls that are in a sense "un-dominated". For controls that are not on *the efficient frontier*, each of them is essentially "dominated" by a control, or a pair of controls, on *the efficient frontier*. To illustrate using the example in Figure 3, control 8 is "dominated" by control 3: they both have the same information quality, but control 8 has a

larger cost. Suppose an interval policy uses control 8, then we can replace it by control 3 to achieve a better objective value. Similarly, control 9 is "dominated" by a combination of control 4 and 5. We can show that in the continuous-time formulations, if a policy uses control 9, then similarly we can replace it by a combination of control 4 and 5 to achieve a better objective value.

5.2. The Optimal Interval Policy (OIP)

We are now in a position to characterize the optimal interval policy in terms of a sequence of controls on *the efficient frontier* and the optimal lower boundary l^* and upper boundary u^*.

THEOREM 2. *Let the set of controls* $\{(\eta_i, c_i)\}_{i=1}^N$ *be labelled by (24), (25), (26) and (27). Then for (23), there exists an optimal interval policy* $(M^*, m^*(\cdot), \{i_k^*\}_{k=0}^D, \{\zeta_k^*\}_{k=0}^{D+1}, l^*, u^*)$ *with* $\{i_k^*\}$ *satisfying*

i) $i_k^* \in \{1, 2, ..., \tilde{N}\}$,

ii) $|i_k^* - i_{k-1}^*| = 1$ *for all* $k = 1, ..., D$, *if* $D \geq 1$,

iii) $\exists k' \in \{0, ..., D\}$ *such that* $i_k^* > i_{k+1}^*$ *when* $0 \leq k \leq k' - 1$ *and* $i_k^* < i_{k+1}^*$ *when* $k' \leq k \leq D - 1$,

and whose value function $V(\cdot)$ *is twice continuously differentiable over* (l^*, u^*) *and satisfies*

$$-c_i + \eta_i y^2 (1-y)^2 V''(y)/2 \leq \lambda V(y), \text{ for all } y \in (l^*, u^*) \text{ and all } i \in \{1, 2, ..., N\}, \tag{28}$$

$$-c_{m^*(y)} + \eta_{m^*(y)} y^2 (1-y)^2 V''(y)/2 = \lambda V(y), \text{ for all } y \in (l^*, u^*). \tag{29}$$

Moreover under this policy, the optimal action is to reject H_a *when* $\pi_t \in [0, l^*]$, *to reject* H_0 *when* $\pi_t \in [u^*, 1]$, *and to continue experimenting when* $\pi_t \in (l^*, u^*)$.

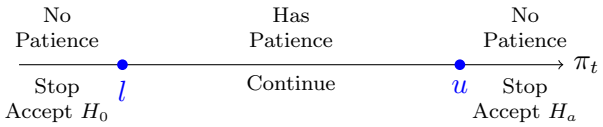
The key statement of Theorem 2 is that there exists an optimal interval policy (OIP) that uses *unimodally consecutive* controls on the *efficient frontier* and which is characterized by the two sets $\{i_k^*\}_{k=0}^D$ and $\{\zeta_k^*\}_{k=0}^{D+1}$. In particular, $\zeta_1^*, \zeta_2^*, ..., \zeta_D^*$ divide the interval $[l^*, u^*)$ into $D+1$ disjoint sub-intervals and when the belief π_t is within the sub-interval $[\zeta_k^*, \zeta_{k+1}^*)$, the decision maker uses the same control i_k^*. The decision maker stops whenever $\pi_t < l^*$ or $\pi_t \geq u^*$: in the former case accepts H_0 and in the latter case accepts H_a. The set of intermediate controls $\{i_k^*\}_{k=0}^D$ satisfies the

properties i), ii) and iii) as listed in Theorem 2. i) states that when the decision maker has not yet stopped (i.e. $\pi_t \in [l^*, u^*)$), the optimal control to be employed comes from *the efficient frontier*. ii) asserts that between adjacent disjoint intervals specified by $\{\zeta_k^*\}_{k=0}^{D+1}$, the controls correspond to two *adjacent* points on *the efficient frontier* at which its slope changes. iii) specifies that the sequence form what we will call a *unimodally consecutive* sequence. Essentially there are four cases. *Case 1*: The OIP policy is monotonically increasing ($D \geq 1$, $k' = 0$). This happens when R_0, κ_0 are relatively small and R_a, κ_a are relatively large, which is essentially the scenario analyzed in HS's formulation. *Case 2*: The OIP policy is monotonically decreasing ($D \geq 1$, $k' = D$). This is the reverse of case 1, which happens when R_0, κ_0 are relatively large and R_a, κ_a are relatively small. *Case 3*: The OIP policy is a singleton ($D = 0$). This happens when all of R_0, κ_0, R_a and κ_a are relatively small and of similar magnitude. *Case 4*: The OIP policy is strictly unimodal ($D \geq 2$, $1 \leq k' \leq D-1$). This is the new case revealed by the analysis of this paper, and it occurs when all of R_0, κ_0, R_a and κ_a are relatively large and of similar magnitude.

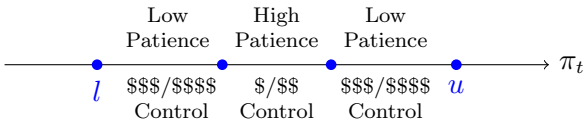
The intuition behind the structure of the OIP in *Case 4* is as follows. At any point in time, the decision maker can terminate and achieve a potential payoff with the current belief (higher when the belief is more extreme), or delay the final decision later to get the payoff with an updated belief via experimentation, but pays both a price to experiment and a price of discounting the payoff. It is intuitively obvious that the stopping region is located at the extreme ends of the belief space while the experimentation region is at the middle (see Panel (a) of Figure 4). Within the experimentation region, the decision maker typically employs relatively higher-cost more informative controls when the belief is more extreme (i.e. when $\pi_t \in [l^*, u^*)$ and close to either l^* or u^*), and relatively lower-cost less informative controls when the belief is less extreme (i.e. when $\pi_t \in [l^*, u^*)$ is close to neither l^* nor u^*). This is because in the former case, the potential payoff is relatively high, and thus the decision maker feels impatient and wants to employ a relatively higher-cost more informative experiment so that he or she can reach the nearer critical boundary faster to obtain the payoff without it being discounted too much. In the latter case, the potential payoff is relatively moderate

or low, the decision maker becomes relatively more patient, and will employ a moderately expensive or a relatively less expensive experiment so as not to incur too much present cost (see Panel (b) of Figure 4). In essence the decision maker faces a trade-off between speed and cost, favors speed by making a faster decision when he or she is more certain, and favors cost by choosing a less costly control when he or she is less certain. It is precisely the balance of this trade-off that leads to the strictly unimodal control sequence in the OIP.

Figure 4 The intuition of why the control sequence of an OIP is *unimodally consecutive*.



(a) The decision maker has no patience to experiment in the stopping region, and has patience to experiment in the continuation region.



(b) Within the continuation region, the decision maker has low patience near the critical boundaries, and has more patience to experiment in the middle.

We note that we can easily identify the controls on the *efficient frontier* (i.e. the controls that will potentially be used in the OIP), which are indexed $1, 2, ..., \tilde{N}$. To find the actual controls used in the OIP, or the set of switching beliefs $\{\zeta_k^*\}_{k=0}^{D+1}$, we refer the reader to the numerical procedure in the Electronic Companion (§Appendix EC.3).

6. Adaptation of Optimal Policy

We note that in a given trial, there may be other important quantities of interest and concern: the type I and II errors α, β and the expected termination time $E(T)$. The corresponding α, β and $E(T)$ associated with the OIP may be large. In reality, the decision maker may also wish to control either the former or the latter, or both. In the remainder of the section, we present adaptations of our formulation to accommodate such a requirement.

6.1. Control of Type I and II Errors

In certain cases the designers of a clinical trial require a pre-specified type I and type II error. This can be achieved in our formulation by fixing the lower and upper critical beliefs, which are

functions of the tolerated errors. The resulting stochastic control problem can then be solved with these beliefs fixed.

To be more specific, let the decision maker's maximum tolerated level of α, β be α_0, β_0. Given any interval policy Π, the resulting type I error $\alpha(\Pi)$ and type II error $\beta(\Pi)$ are defined as

$$\alpha(\Pi) = Pr(\text{Accept } H_a \mid \Pi, H_0),$$

$$\beta(\Pi) = Pr(\text{Accept } H_0 \mid \Pi, H_a).$$

In our continuous-time formulation, the type I and II errors are uniquely determined by the policy's lower and upper critical beliefs.

LEMMA 2. *Let π_0 be fixed. Given any interval policy Π, the following relationship between its lower and upper critical beliefs l, u and the type I and II errors α, β holds*

$$\alpha = \frac{\frac{1}{1-\pi_0} - \frac{1}{1-l}}{\frac{1}{1-u} - \frac{1}{1-l}}, \ \beta = \frac{\frac{1}{\pi_0} - \frac{1}{u}}{\frac{1}{l} - \frac{1}{u}}. \tag{30}$$

In other words, for a given interval policy and a prior belief π_0, the type I and type II errors depend only on l, u and do not depend on the switching beliefs and intermediate controls (l and u here are analogous to the O'Brien-Fleming boundary in the classical hypothesis testing literature). This consequence arises from Theorem 1, where under either hypothesis, the diffusion process L_t always has a variance proportional to its drift. Subsequently, it can be shown that starting at π_0 the probability of hitting either l or u is the same no matter which control is being used.

Denote $\mathcal{E} = (\alpha_0, \beta_0)$ and let $l^{\mathcal{E}}, u^{\mathcal{E}}$ be the corresponding lower and upper critical beliefs given by (computed by inverting (30))

$$l^{\mathcal{E}} = \frac{\beta_0\pi_0}{\beta_0\pi_0 + (1-\alpha_0)(1-\pi_0)}, \ u^{\mathcal{E}} = \frac{(1-\beta_0)\pi_0}{(1-\beta_0)\pi_0 + (1-\pi_0)\alpha_0}. \tag{31}$$

The idea is that we will fix the continuation region to be $(l^{\mathcal{E}}, u^{\mathcal{E}})$ and solve for the optimal interval policy of the following stochastic control problem

$$V_{\mathcal{E}}(y) = \max_{\Pi | l(\Pi) = l^{\mathcal{E}}, u(\Pi) = u^{\mathcal{E}}} E\left[e^{-\lambda T} G(\pi_T) - \int_0^T e^{-\lambda t} c(M_t) dt \;\middle|\; \pi_0 = y \right], \tag{32}$$

where $l(\Pi)$ and $u(\Pi)$ denote the lower and upper critical beliefs of the *interval policy* Π. Then we can obtain the following result that is analagous to Theorem 2.

THEOREM 3. *Let the set of controls* $\{(\eta_i, c_i)\}_{i=1}^{N}$ *be labelled by (24), (25), (26) and (27). Then for (32), there exists an optimal interval policy* $(\tilde{M}^*, \tilde{m}^*(\cdot), \{\tilde{i}_k^*\}_{k=0}^{D}, \{\tilde{\zeta}_k^*\}_{k=0}^{D+1}, l^{\mathcal{E}}, u^{\mathcal{E}})$ *with* $\{\tilde{i}_k^*\}$ *satisfying*

i) $\tilde{i}_k^* \in \{1, 2, ..., \tilde{N}\}$,

ii) $|\tilde{i}_k^* - \tilde{i}_{k-1}^*| = 1$ *for all* $k = 1, ..., D$, *if* $D \geq 1$,

iii) $\exists k' \in \{0, ..., D\}$ *such that* $\tilde{i}_k^* > \tilde{i}_{k+1}^*$ *when* $0 \leq k \leq k' - 1$ *and* $\tilde{i}_k^* < \tilde{i}_{k+1}^*$ *when* $k' \leq k \leq D - 1$,

and whose value function $V_{\mathcal{E}}(\cdot)$ *is twice continuously differentiable over* $(l^{\mathcal{E}}, u^{\mathcal{E}})$ *and satisfies*

$$-c_i + \eta_i y^2 (1-y)^2 V_{\mathcal{E}}''(y)/2 \leq \lambda V_{\mathcal{E}}(y), \text{ for all } y \in (l^{\mathcal{E}}, u^{\mathcal{E}}) \text{ and all } j \in \{1, 2, ..., N\}, \tag{33}$$

$$-c_{\tilde{m}^*(y)} + \eta_{\tilde{m}^*(y)} y^2 (1-y)^2 V_{\mathcal{E}}''(y)/2 = \lambda V_{\mathcal{E}}(y), \text{ for all } y \in (l^{\mathcal{E}}, u^{\mathcal{E}}). \tag{34}$$

Moreover under this policy, the optimal action is to reject H_a *when* $\pi_t \in [0, l^{\mathcal{E}}]$, *to reject* H_0 *when* $\pi_t \in [u^{\mathcal{E}}, 1]$, *and to continue experimenting when* $\pi_t \in (l^{\mathcal{E}}, u^{\mathcal{E}})$.

Theorem 3 is analogous to Theorem 2 in the sense that in both cases the resulting optimal policies are *unimodally consecutive* and use controls from the same *efficient frontier*. The difference is that the optimal lower and upper critical beliefs are determined exogenously in the former and endogenously in the latter.

6.2. Control of Expected Termination Time

The resulting OIP from the continuous-time formulation (21) does not control for the expected termination time $E(T)$: it is possible that within a certain belief range, the OIP uses relatively uninformative ("slow") controls that result in π_t taking a long time to reach the critical belief thresholds. Suppose the decision maker wishes to place a constraint $E(T) \leq \bar{t}$. We propose two heuristics below.

The first approach involves successively re-optimizing the model by removing elements of $\mathcal{S}$ one by one (from controls that are less informative to those that are more informative). The second, faster, approach does not need to re-optimize the model many times. Instead it first eliminate all

controls that, if used alone, produce an expected termination time greater than $\hat{t}$. More specifically, denote $E_{\pi_0,l,u}(T;j)$ to be the expected termination time associated with the interval policy that always uses control j and has critical beliefs l,u (and prior π_0). The following lemma provides a closed-form expression for $E_{\pi_0,l,u}(T;j)$.

LEMMA 3.

$$E_{\pi_0,l,u}(T;j) = \frac{2}{\eta_j(u-l)}\left[(u-\pi_0)\psi_l(\pi_0)To + (\pi_0-l)\psi_u(\pi_0)\right], \tag{35}$$

where $\psi_z(y) = -(2y-1)\ln\left(\frac{y}{1-y}\right) + (2z-1)\ln\left(\frac{z}{1-z}\right)$.

To control the OIP's expected termination time, we first use (l^*,u^*) from the original OIP as a reference, and eliminate the set of relatively uninformative controls $\mathcal{U} = \{j \in \mathcal{S} \mid E_{\pi_0,l^*,u^*}(T;j) > \bar{t}\}$. Suppose $\mathcal{S}\backslash\mathcal{U} = \emptyset$, then $\bar{t}$ is probably too small and we may want to increase it and make the constraint on $E(T)$ less restrictive. Suppose $\mathcal{S}\backslash\mathcal{U} \neq \emptyset$, then we find the new OIP with controls in $\mathcal{S}\backslash\mathcal{U}$. The following lemma shows that the new OIP's expected termination time is controlled as desired.

LEMMA 4. *Let* Π^*, $\hat{\Pi}^*$ *denote the OIP associated with control set* $\mathcal{S}$ *and* $\mathcal{S}\backslash\mathcal{U}$ *respectively, then*

$$l(\Pi^*) \le l(\hat{\Pi}^*),\ u(\Pi^*) \ge u(\hat{\Pi}^*), \tag{36}$$

$$E(T;\tilde{\Pi}^*) \le \bar{t}. \tag{37}$$

Lemma 4 essentially says that the new OIP's continuation region $(l(\hat{\Pi}^*), u(\hat{\Pi}^*))$ is narrower compared to the old one's. This implies that $E_{\pi_0,l(\hat{\Pi}^*),u(\hat{\Pi}^*)}(T;j) \le \bar{t}$ for all $j \in \mathcal{S}\backslash\mathcal{U}$, and hence the new OIP (which comprises these controls) also has an expected termination time not exceeding $\bar{t}$.

7. Numerical Study

In this section, we compare the dynamic adaptive policy OIP developed in this paper to other simpler adaptive policies that are used in practice. We do this in the context of a real world adaptive clinical trial design: the Phoenix Champion Trial NCT01156571 (Bhatt et al. (2013)), a phase-III

trial for the small molecule drug cangrelor that helps reduce the rate of ischemic complications during percutaneous coronary interventions (PCI).

The primary efficacy endpoint was the composite rate of death from causes such as myocardial infarction, ischemia-driven revascularization or stent thrombosis in the 48 hours after undergoing PCI. The Phoenix Trial used a two-stage adaptive design with a maximum patient sample size of 10,000, with 70% of the patients recruited in the first stage. Based on the result of the first stage, it would then be decided whether the additional 30% of the patients would be recruited. The recruited patients were separated into either the control group that used clopidogrel or the treatment group that used cangrelor.

Below, §7.1 provides an overview of the relevant policies to be analyzed and §7.2 describes the relevant controls used in the trial. §7.3 present the parameter estimates and §7.4 presents the result. All simulations and numerical analyses are done in MATLAB.

7.1. Overview of Relevant Policies

We analyze the following policies: OIP (our dynamic policy), 2PA70-30 (2-period adaptive, with the 1^{st} and 2^{nd} period using 70% and 30% of total patient size replicating the actual trial design), 2PABest (2-period adaptive, with percentage of the total patient size used in the 1^{st} and 2^{nd} period optimized), Multi (multi-period adaptive, derived from 2PABest) and 1NA (1-period non-adaptive).

OIP. To derive the OIP policy, we compute the information quality of each control, compute the *efficient frontier* and after that compute the OIP policy using the numerical procedure specified in the Electronic Companion §Appendix EC.3.

2PA70-30 & 2PABest. 2PA70-30 is the policy used in the Phoenix Champion Trial, where given a total patient sample size, the decision maker uses 70% of it in the first period. Once the results are obtained, the prior is updated and if it drops below a lower threshold the null is accepted. If it increased above an upper threshold, the null is rejected. Otherwise. the remaining 30% is recruited in the second period, after which the decision makers accepts H_0 if the posterior

is below a threshold and rejects it otherwise. The two thresholds in the first period and the single threshold in the second period, along with the sample size optimized to maximize total economic benefit. 2PABest represents the best two-stage dynamic adaptive policy which is similar to 2PA70-30, with the difference that the percentages of sample size used in the first and second period are also optimized.

Multi. Multi represents a multi-period adaptive heuristic that uses same sample size as 2PABest, and for each period also uses the lower and upper belief thresholds from 2PABest's second period. In each period, if the updated posterior belief drops below the lower threshold the null is accepted. If it increased above an upper threshold, the null is rejected. Otherwise. continue experimenting with the same sample size in the next period.

1NA. The 1NA policy is the 1-period non-adaptive policy where the total patient sample size is given. After the first period the trial ends, and the decision maker reviews the result and decide whether to accept H_0 or H_a. The total patient sample size is optimized. This represents a traditional clinical trial design.

7.2. Description of Controls

The Phoenix Trial had two stages. For each stage, the decision maker selected different patient sample sizes (no. of patients to recruit). Moreover, the decision maker decided on how many clinical sites to open to accommodate the patients recruited and whether to add more sites over time. In the numerical study, we assume each stage lasts for a year. This is because the Phoenix Trial had two stages and took about two years to complete (from Sep. 2010 to Oct. 2012). We assume each site has a capacity of 100 patients. This is because the Phoenix Trial initially planned about $10,000$ patients for about 100 sites. Thus, for the numerical study, a control in a particular stage can be represented by (n_P, n_S), where n_P and n_S refer to the number of patients recruited and the number of sites opened up for that stage. If a control (n_P, n_S) is chosen by the decision maker, three types of costs are incurred: a per patient recruitment cost c_P multiplied by n_P, a per site retaining cost c_S (includes cost of keeping the site open, such as cost of maintenance and employee

training) multiplied by n_S, and a miscellaneous cost c_O (includes overhead cost, data management cost, coordination cost, IRB approval and amendment costs, etc) that is independent of n_P and n_S. For computational tractability, we assume n_P is a multiple of 500, and the decision maker can set up at most 100 sites in each stage (i.e. $n_P \leq 100 n_S$, $n_S \leq 100$, $n_P \in \{500, 1000, ..., 10000\}$). The average recruiting cost per patient varies across sites, reflecting the situation in which different sites have different advertising, recruiting, lab and hospital costs. Let μ_P^n denote the average recruiting cost per patient, when the firm enrolls a total of n patients to take part in the trial for one stage. The values for μ_P^n are derived from the estimates in Table 2 of Kouvelis et al. (2017), with details described in the Electronic Companion (see §Appendix EC.6). The resulting values range from $0.002M$ to $0.04M$ and they satisfy the following inequality: $\mu_P^{100} < \mu_P^{200} < ... < \mu_P^{10000}$. These values are also in line with those from Sertkaya et al. (2016) and Moore et al. (2018). In addition, we assume each site incurs a site cost per stage of $\mu_S = 2.25M$, which is derived from Sertkaya et al. (2016) (see §Appendix EC.6).

7.3. Study Description and Parameters

In the baseline scenario, we perform the numerical study with the parameters based on the Phoenix Trial. From earlier studies, we assume that the probabilities that a randomly chosen patient fails to survive are $\rho_a = 0.039$ and $\rho_0 = 0.051$ under the treatment group (cangrelor group) and control group (clopidogrel group) respectively (Bhatt et al. (2013)). We assume cangrelor has a true probability of efficacy of $\pi_0 = 0.5$, because this value represents historically the percentage of phase-III clinical trials that succeed (Pretorius (2016)). We also assume an annual discount rate of 0.1, a common assumption in many clinical trial studies (Sertkaya et al. (2016), Chaudhuri et al. (2018) and Woo et al. (2019)).

To select the parameters $(R_0, \kappa_0, R_a, \kappa_a)$, we attempt to estimate the parameters implied by the actual design used in the Phoenix Champion Trial. To do this, we search over the range of possible parameters, and for each choice considered we compute the lower and upper boundaries to maximize total economic benefit and then derive the type I and type II errors. Finally, we

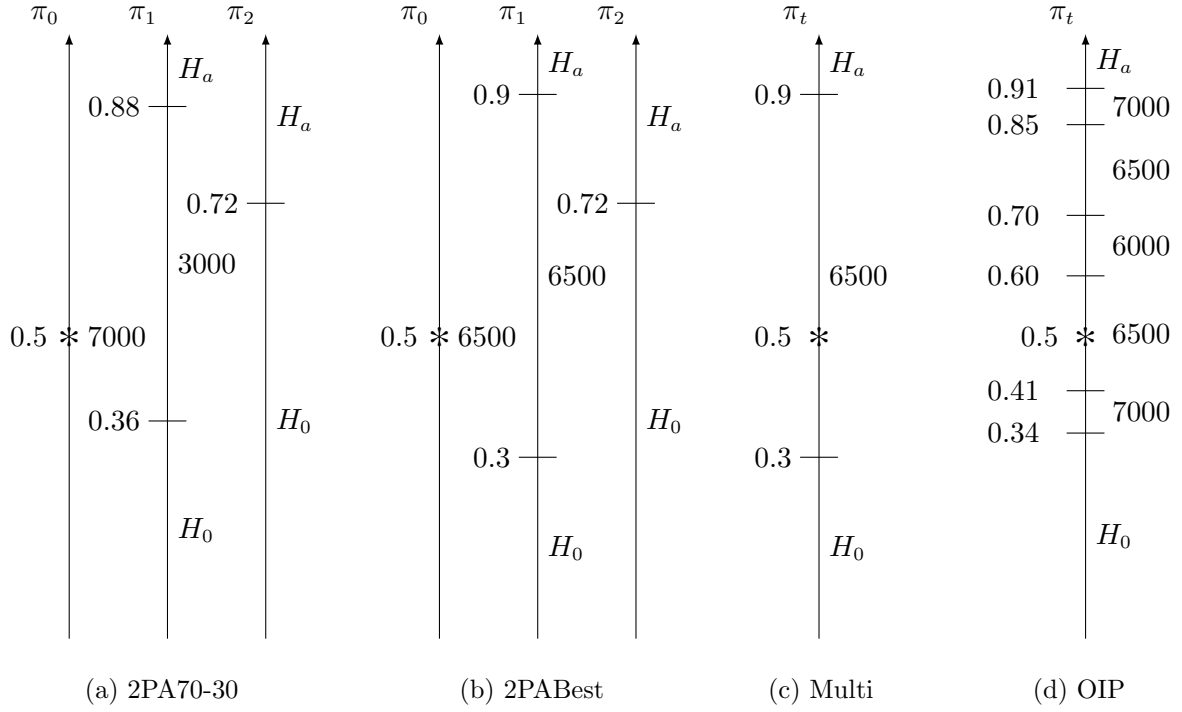
select the parameters that most closely match the actual type I and type II errors: $\alpha = 0.05, \beta = 0.15$. We find that $(R_0, \kappa_0, R_a, \kappa_a) = (3000, 0, 2700, 4000)$ generates the desired type I and type II errors. This approach leads into a Bayesian framework that generates the type I and type II errors required by the frequentist design used in the actual clinical trials. It is based on the methodology outlined in Silva (2018) which calibrates non adaptive Bayesian and frequentist tests to generate concordant conclusions. The values obtained through this method suggest that the trial designer assigns roughly a 33% higher cost of incorrectly accepting the alternative compared to the reward of correctly accepting either the null or the alternative, and assigns zero cost to incorrectly accepting the null. This implies that the cost of introducing an ineffective drug to the market is perceived to be $1/3^{rd}$ larger than the benefit of introducing an effective drug, and that there is no loss in incorrectly concluding that an effective drug is ineffective.

Starting from this baseline scenario, we then evaluate the performance of OIP relative to the other policies over a variety range of parameters. First, we vary $\rho_0 - \rho_a$ over a range of 0.01 to 0.02 (with ρ_a from 0.02 to 0.45). Second, we vary π_0 over a range of 0.4 to 0.7. The range for $\rho_0 - \rho_a$ is chosen as such because when it is less than 0.01, the experiments are not informative enough and hence all policies would never run any experiment and would immediately accept H_0; when it is more than 0.02, the approximation behind our dynamic policy OIP becomes increasingly worse and there is a point where OIP's performance becomes worse than those of the simpler policies. We are going to explore the value where that switch happens, but most of the informative variations of performance between the three policies happen when $\rho_0 - \rho_a$ is between 0.01 and 0.02, which is also the range of values that is very similar to that of the actual clinical trial (0.012) and is a clinically relevant range. The range for π_0 is chosen because this represents the different estimates of the probability that a phase-III clinical trial succeeds (Pretorius (2016)).

7.4. Result

In this subsection, we will present results on the following: a) Performance of all policies in four scenarios - a baseline scenario that represents the Phoenix Trial, a best-case scenario in which the

OIP attains the greatest performance improvement, an intermediate scenario in which the OIP attains a moderate performance improvement and a worst-case scenario that is representative of OIP's performance relative to the other policies' among all parameters considered; b) Aggregate performance improvement of 2PABest, 2PA70-30, Multi and OIP relative to 1NA; c) Performance improvements of OIP relative to the other policies as ρ_a and ρ_0 vary.



(a) 2PA70-30 (b) 2PABest (c) Multi (d) OIP

Figure 5 Explicit policies of 2PA70-30, 2PABest, Multi and OIP in the baseline scenario. 2PA70-30 & 2PABest: first axis presents #patients recruited in the 1^{st} stage; second axis presents the lower / upper critical beliefs and #patients recruited in the 2^{nd} stage; third axis shows the terminal belief threshold. OIP: axis shows the belief thresholds and #patients recruited. Multi: axis shows the belief thresholds and #patients recruited.

Baseline Scenario: Figure 5 shows the explicit policies of 2PA70-30, 2PABest, Multi and OIP. For 2PA70-30 and 2PABest, the first axis presents the number of patients recruited in the 1^{st} stage. In particular, 2PA70-30 recruits 7000 patients and 2PABest recruits 6500. The second axis presents the lower critical belief (below which H_0 is accepted) and the upper critical belief (above which H_a is accepted) thresholds, as well as the number of patients recruited in the 2^{nd} stage (if the posterior

belief falls between the lower and upper thresholds). For instance, 2PA70-30 recruits 3000 patients in the 2^{nd} stage if the posterior belief is between 0.36 and 0.88. The third axis shows the terminal belief threshold, below which H_0 is accepted and above which H_a is accepted. For both 2PA70-30 and 2PABest this thresholds is 0.72 (the value can be shown to be $(R_0+\kappa_a)/(R_0+\kappa_0+R_a+\kappa_a)$). For OIP, the only axis presents the belief thresholds and the number of patients recruited for a posterior belief that falls between two adjacent thresholds. For instance, when the posterior belief at any particular stage is between 0.41 and 0.6, OIP recruits 6500 patients. For Multi, the only axis presents the identical information as that of OIP.

We observe that in the first period, all the adaptive policies begin with approximately the same sample size. In the second period, they have similar lower and upper critical beliefs, but 2PABest, Multi and OIP approximately double the sample size. Because 2PABest, Multi and OIP adopt very similar sample sizes and for both OIP and Multi most decisions are made within 2 periods, these policies perform almost the same (results to be shown next).

Table 1 Average economic benefits of all policies in the baseline scenario ($\pi_0=0.5$, $\rho_a=0.039$, $\rho_0-\rho_a=0.012$). Brackets show the percentage change relative to 1NA.

	1NA	2PA70-30	2PABest	Multi	OIP
Econ. Benefit ± SE (Δ%)	1663.5±45.4 (0%)	1789.3±39.7 (7.6%)	1869.1±33.3 (12.4%)	1877.9±30.8 (12.9%)	1876.4±30.9 (12.8%)
T ± SE (Δ%)	1±0 (0%)	1.23±0.01 (22.7%)	1.31±0.01 (31.3%)	1.42±0.02 (41.8%)	1.39±0.02 (39.4%)
Sample Size ± SE (Δ%)	8000±0 (0%)	7681±39.8 (-4%)	8534.5±95.4 (6.7%)	9217±143.6 (15.2%)	9089±142 (13.6%)
Cost of Experiments ± SE (Δ%)	440.3±0 (0%)	379.9±1.3 (-13.7%)	410.3±4.3 (-6.8%)	437.7±6.1 (-0.6%)	432.9±6.1 (-1.7%)
α ± SE (Δ%)	0.09±0.01 (0%)	0.06±0.01 (-27.3%)	0.04±0.01 (-52.3%)	0.04±0.01 (-59.1%)	0.03±0.01 (-61.4%)
β ± SE (Δ%)	0.17±0.02 (0%)	0.15±0.02 (-11.8%)	0.1±0.01 (-43.5%)	0.06±0.01 (-63.5%)	0.08±0.01 (-55.3%)

The performance of all policies in the baseline scenario (which represents a reduction of 23% in the relative risk of adverse events) is shown in Table 1. The result shows that OIP performs better than the policies used in practice (with 12.8% improvement compared to 1NA, and additional 5.2% improvement compared to 2PA70-30). However OIP's improvement achieved relative to 2PA70-30 is almost the same as 2PABest's and Multi's, that is, the same as if we would optimize the two-period policy and make it multi-period. Note that this conclusion can already be deduced from Figure 5. All adaptive policies have termination greater than 1NA, as they may continue experimenting after observing the result from the first period, with the termination time of OIP

comparable to that of 2PABest and Multi, and slightly larger than that of 2PA70-30. In terms of sample size, 2PABest, Multi and OIP have larger values than 1NA, whereas 2PA70-30 has a smaller value than 1NA. In terms of experiment cost, all adaptive policies achieve a smaller value than 1NA, with 2PA70-30 having the smallest value. In terms of type-I and II errors, 2PABest, Multi and OIP achieve lower values compared to 2PA70-30. To summarize, in the baseline scenario that is representative of the actual Phoenix Trial, OIP performs better than the adaptive policy used in practice (2PA70-30) and the non-adaptive one (1NA), but is almost indistinguishable to the two-period optimized policy 2PABest and the simpler multi-period policy (Multi).

Best Case Scenario: Here we present a scenario that represents the largest performance gap (among all scenarios considered) between OIP and the remaining policies. The parameters are $\pi_0 = 0.6$, $\rho_0 - \rho_a = 0.014$, $\rho_a = 0.2$. This is a scenario in which the primary endpoint occurs 20% of the time and the treatment arm represents a reduction of 7% in the relative risk of adverse events. Table 2 shows the result, where OIP achieves an additional 9.7%, 8.9% and 4.4% economic benefit relative to 2PA70-30, 2PABest and Multi, respectively. In this scenario, π_0 lies relatively in the middle of experimentation interval for all policies, and the information quality is not very large (both ρ_a and $\rho_0 - \rho_a$ are in the intermediate range). OIP performs the best because in this set of parameters, the expected duration of a trial is long on average (1.7, longer than 1.4 in the baseline scenario), and the asymptotic regime in our framework is relatively more accurate.

Table 2 Average economic benefits of all policies. Brackets show the percentage change relative to 1NA, for $\pi_0 = 0.6$, $\rho_0 - \rho_a = 0.014$, $\rho_a = 0.2$.

	1NA	2PA70-30	2PABest	Multi	OIP
Econ. Benefit ± SE (Δ%)	1308.5±52.4 (0%)	1316.4±49.6 (0.6%)	1327.3±48.8 (1.4%)	1385.2±44.4 (5.9%)	1443.9±42.9 (10.3%)
T ± SE (Δ%)	1±0 (0%)	1.32±0.01 (31.5%)	1.4±0.02 (40.1%)	1.62±0.03 (61.6%)	1.7±0.03 (69.8%)
Sample Size ± SE (Δ%)	7000±0 (0%)	8739.5±48.5 (24.9%)	9106.5±100.8 (30.1%)	10504±190.3 (50.1%)	9619±190.5 (37.4%)
Cost of Experiments ± SE (Δ%)	357.7±0 (0%)	449.6±1.6 (25.7%)	435.9±4.5 (21.8%)	490.5±7.8 (37.1%)	415.4±7.3 (16.1%)
α ± SE (Δ%)	0.13±0.02 (0%)	0.13±0.02 (-1.9%)	0.12±0.02 (-9.4%)	0.09±0.01 (-30.2%)	0.08±0.01 (-39.6%)
β ± SE (Δ%)	0.38±0.02 (0%)	0.29±0.02 (-24.1%)	0.3±0.02 (-21.9%)	0.23±0.02 (-38.2%)	0.26±0.02 (-32.5%)

Intermediate Scenario: Here we present a scenario that represents an average performance gap (among all scenarios considered) between OIP and the remaining policies, where in the alternative, the drug reduces the adverse events by 12%. The parameters are $\pi_0 = 0.6$, $\rho_a = 0.09$ and $\rho_0 - \rho_a =$

0.012. Table 3 shows the result, where OIP achieves an additional 8.5%, 6.5% and 5.5% economic benefit relative to 2PA70-30, 2PABest and Multi, respectively. Similarly as in the best case scenario, the expected duration of a trial is long on average (1.65, longer than 1.4 in the baseline scenario), and the asymptotic regime in our framework is relatively accurate.

Table 3 Average economic benefits of all policies. Brackets show the percentage change relative to 1NA, for $\pi_0 = 0.6$, $\rho_0 - \rho_a = 0.012$, $\rho_a = 0.09$.

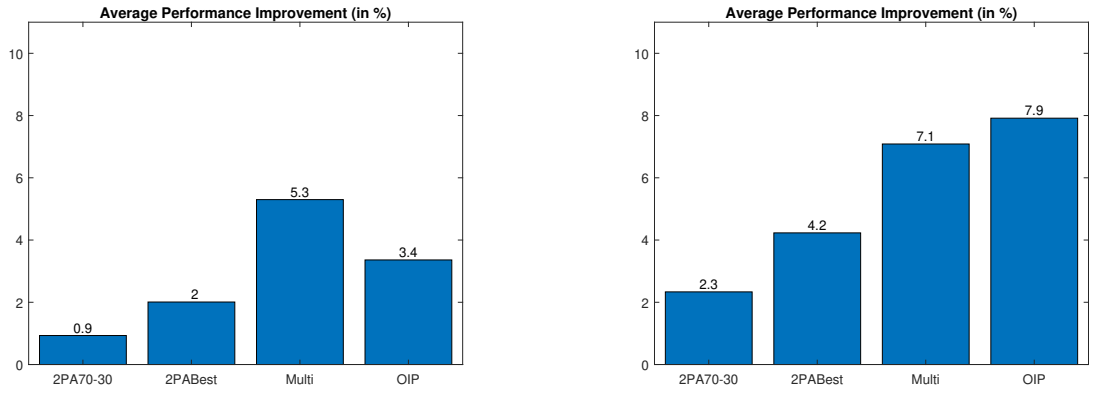
	1NA	2PA70-30	2PABest	Multi	OIP
Econ. Benefit ± SE (Δ%)	1394.8±50.1 (0%)	1441.8±46.5 (3.4%)	1470.2±46.3 (5.4%)	1484.6±43.6 (6.4%)	1561±40.2 (11.9%)
T ± SE (Δ%)	1±0 (0%)	1.28±0.01 (28.1%)	1.38±0.02 (37.9%)	1.59±0.03 (59.2%)	1.65±0.03 (64.7%)
Sample Size ± SE (Δ%)	8000±0 (0%)	8627.3±46.9 (7.8%)	8963.5±99.8 (12%)	10348±199.1 (29.3%)	10034±196 (25.4%)
Cost of Experiments ± SE (Δ%)	440.3±0 (0%)	445.8±1.6 (1.2%)	429.5±4.5 (-2.5%)	482.7±8 (9.6%)	450.8±7.7 (2.4%)
α ± SE (Δ%)	0.13±0.02 (0%)	0.11±0.02 (-14%)	0.11±0.02 (-12%)	0.09±0.01 (-26%)	0.07±0.01 (-42%)
β ± SE (Δ%)	0.28±0.02 (0%)	0.24±0.02 (-12%)	0.22±0.02 (-21.6%)	0.17±0.02 (-37.7%)	0.17±0.02 (-39.5%)

Worst Case Scenario: Finally we present a scenario that represents the worst performance gap (among all scenarios considered) between OIP and the remaining policies, where in the alternative, the treatment reduces the relative risk of the adverse events by 50%. The result (see Table 4) is that OIP achieving about 7% less economic benefit than 2PABest and Multi, and about 6% less than 2PA70-30. In this scenario, π_0 lies relatively near the end of the experimentation range for all policies, and the information quality is relatively large (ρ_a is relatively small, $\rho_0 - \rho_a$ is relatively large, and both are at the extreme end of the range). OIP performs the worst because in this set of parameters, the expected duration of a trial is short on average (1.04, shorter than 1.4 in the baseline scenario), and the asymptotic regime in our framework becomes relatively less accurate.

Table 4 Average economic benefits of all policies. Brackets show the percentage change relative to 1NA, for $\pi_0 = 0.4$, $\rho_0 - \rho_a = 0.02$, $\rho_a = 0.02$.

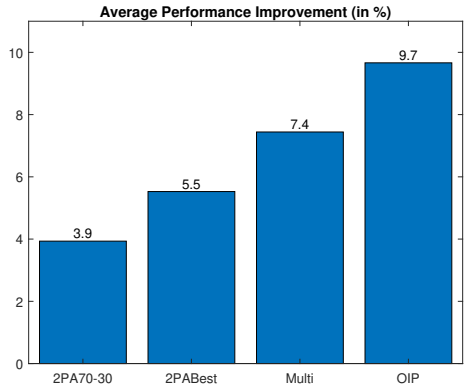
	1NA	2PA70-30	2PABest	Multi	OIP
Econ. Benefit ± SE (Δ%)	2305.6±21 (0%)	2336.2±16 (1.3%)	2354.8±13.5 (2.1%)	2359.7±11.8 (2.3%)	2192.6±7.4 (-4.9%)
T ± SE (Δ%)	1±0 (0%)	1.08±0.01 (8.3%)	1.12±0.01 (12%)	1.13±0.01 (13.2%)	1.04±0.01 (4%)
Sample Size ± SE (Δ%)	5000±0 (0%)	5074.1±18.3 (1.5%)	5041.4±46.3 (0.8%)	5095.4±53.6 (1.9%)	7804.1±48.8 (56.1%)
Cost of Experiments ± SE (Δ%)	216.6±0 (0%)	215.5±0.5 (-0.5%)	206.6±1.7 (-4.6%)	208.5±2 (-3.7%)	412.8±2.4 (90.6%)
α ± SE (Δ%)	0.013±0.004 (0%)	0.008±0.003 (-36.8%)	0.005±0.002 (-59.2%)	0.003±0.002 (-73.7%)	0.001±0.001 (-93.4%)
β ± SE (Δ%)	0.049±0.007 (0%)	0.02±0.004 (-59.5%)	0.011±0.003 (-78.5%)	0.007±0.003 (-85.1%)	0.001±0.001 (-99%)

In summary, these scenarios suggest that OIP tends to perform well when the prior belief π_0 lies in the intermediate range between the lower and upper boundaries (0.5 - 0.7), ρ_a is not of the same order of magnitude as $\rho_0 - \rho_a$, and $\rho_0 - \rho_a$ lies between 0.01 and 0.02.

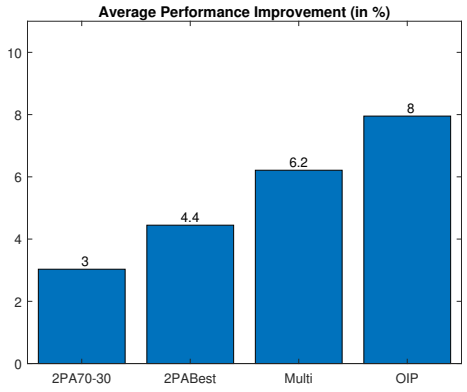


(a) $\pi_0 = 0.4$

(b) $\pi_0 = 0.5$



(c) $\pi_0 = 0.6$

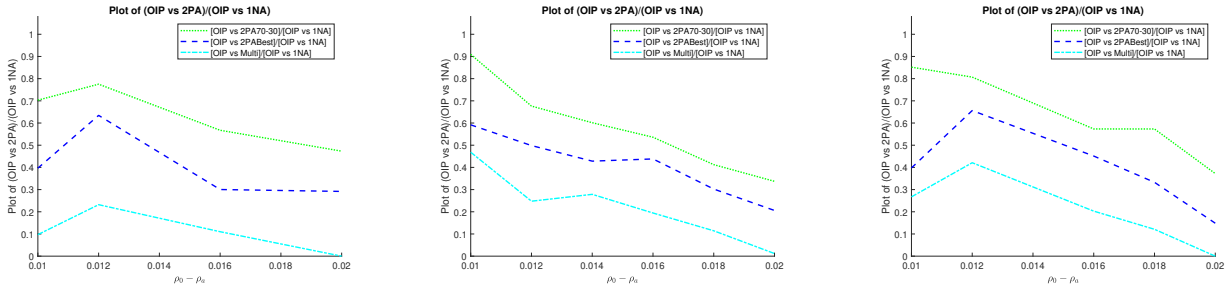


(d) $\pi_0 = 0.7$

Figure 6 Aggregate performance improvement of 2PA70-30, 2PABest, Multi and OIP relative to 1NA for $\pi_0 = 0.4, 0.5, 0.6$ and 0.7.

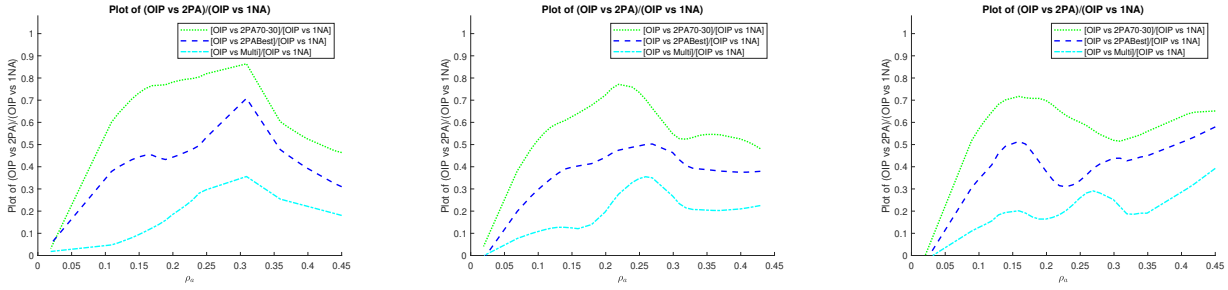
Aggregate Performance Improvement: Next, we evaluate the aggregate performance improvement of 2PA70-30, 2PABest, Multi and OIP relative to 1NA for $\pi_0 = 0.4, 0.5, 0.6$ and 0.7. Specifically, for each value of π_0 considered, we simulate the performance of all policies holding $\rho_0 - \rho_a$ fixed and ρ_a ranging from 0.02 to 0.45 (in increment of 0.01). We repeat this for $\rho_0 - \rho_a$ ranging from 0.01 to 0.02 (in increment of 0.002) and then compute the aggregate performance improvement of all adaptive policies relative to 1NA. Figure 6 shows the result, with the 4 panels corresponding to $\pi_0 = 0.4, 0.5, 0.6$ and 0.7. We see that when the prior is very close to the lower boundary (i.e. $\pi_0 = 0.4$), the simple multi-period policy performs the best. However, when the prior moves away from the boundary, OIP performs the best. The largest performance gap between OIP

and the remaining policies occurs when the prior is in the middle of the range considered (i.e. $\pi_0 = 0.6$). In this case, we see an additional economic benefit improvement of 1.6%, 1.9% and 2.3%, when the policy transitions from 2PA70-30 to 2PABest, from 2PABest to Multi, and from Multi to OIP respectively. In this case, the improvement of OIP comes from all three of its adaptive features: multi periods, changing sample sizes between periods, and adjusting the sample size based on the strength of the evidence.



(a) $\pi_0 = 0.5$ (b) $\pi_0 = 0.6$ (c) $\pi_0 = 0.7$

Figure 7 Plot of OIP's incremental benefit (relative to 2PA70-30, 2PABest and Multi) against $\rho_0 - \rho_a$ for $\pi_0 = 0.5, 0.6$ and 0.7.



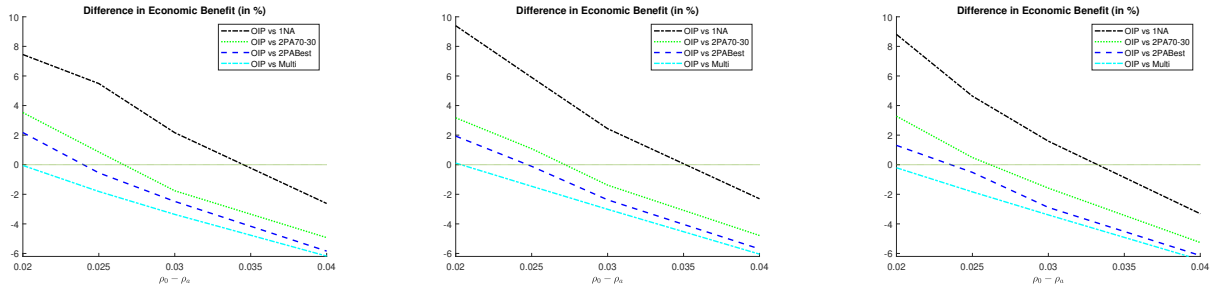
(a) $\pi_0 = 0.5$ (b) $\pi_0 = 0.6$ (c) $\pi_0 = 0.7$

Figure 8 Plot of OIP's incremental benefit (relative to 2PA70-30, 2PABest and Multi) against ρ_a for $\pi_0 = 0.5, 0.6$ and 0.7.

Factors Driving the Performance of OIP. To further explore the drivers of the performance gains of OIP, we focus on the π_0 values where OIP achieves a significant improvement relative to

1NA ($\pi_0 = 0.5$, 0.6 and 0.7) and provide more detailed data on the incremental benefit of OIP's performance improvement relative to 2PA70-30, 2PABest and Multi. Furthermore, to determine how much of the improvement in OIP is due to simple adaptive elements (e.g. changing sample size from period to period, allowing multiple periods) versus more complex elements (adapting the sample size to the strength of evidence), we express the improvement of OIP versus the three adaptive policies as a percentage of the improvement of the OIP relative to 1NA. More specifically, we compute (Economic Benefit of OIP − Economic Benefit of 2PA70-30) / (Economic Benefit of OIP − Economic Benefit of 1NA) and we repeat for 2PABest and Multi.

Figure 7 plots the three metrics against $\rho_0 - \rho_a$, and Figure 8 plots them against ρ_a. In both Figures, we can see that the relative improvement can be quite significant. In Figure 7, for OIP versus 2PA70-30 the relative improvement can be as high as 90% (see Panel (b)) and the average across all panels is 65.1%; for OIP versus 2PABest it can be as high as 65% (see Panel (c)) with an average of 45.3%; for OIP versus Multi it can be as high as 45% (see Panel (b)) with an average of 23.8%. In Figure 8, for OIP versus 2PA70-30 the relative improvement can be as high as 85% (see Panel (a)) with an average of 56.6%; for OIP versus 2PABest it can be as high as 65% (see Panel (a)) with an average of 40.7%; for OIP versus Multi it can be as high as 35% (see Panel (b)) with an average of 21.6%. These results put in perspective the improvement of OIP versus the other adaptive policies: while in absolute percentage terms the improvement is in the single digits, as a percentage of the maximum possible improvement (measured as the improvement of OIP versus 1NA), it can be significant and as high as 90%. These results also demonstrate how different aspects of the adaptive policy contribute to the economic benefit gains: In Figure 7, the non-optimized two stage adaptive policy used in practice attains on average 34.9% of the maximum improvement attainable by the OIP, the optimized two stage policy attains on average 54.7% of the maximum improvement and the multi-period policy attains 76.2% improvement. The remaining 25.8% of the attainable improvement can be attributed to the OIP's adaptation of the sample size in each period based on the strength of the evidence. The corresponding numbers for Figure 8 are 45.4%, 59.3 %, 78.4% and 21.6%.



(a) $\pi_0 = 0.5$ (b) $\pi_0 = 0.6$ (c) $\pi_0 = 0.7$

Figure 9 Plot of difference in economic benefit (OIP versus 1NA, 2PA70-30, 2PABest and Multi) against $\rho_0 - \rho_a$ for $\pi_0 = 0.5, 0.6$ and 0.7.

Validity of Asymptotic Approximation. Finally, we explore what happens when the asymptotic approximation no longer applies by examining how the performance of OIP deteriorates as $\rho_0 - \rho_a$ becomes greater than 0.02. Figure 9 shows this observation: for example given $\pi_0 = 0.5$, when $\rho_0 - \rho_a$ is about 0.02 OIP starts to perform worse than Multi; when $\rho_0 - \rho_a$ is about 0.023 OIP starts to perform worse than 2PABest; and when $\rho_0 - \rho_a$ is about 0.025 OIP starts to perform worse than 2PA70-30. A similar pattern can be observed for $\pi_0 = 0.6$ and $\pi_0 = 0.7$.

In conclusion, these results demonstrate that for a wide range of parameters that are typical of real life phase III trials, the proposed dynamic OIP policy provides significant improvement in economic benefit. In many scenarios, most of the improvement can be attained by simple optimized adaptive policies but not always. The greatest improvement happens when π_0 is in the intermediate range, $\rho_0 - \rho_a$ is relatively moderate, and ρ_a is relatively not too small. Although for the baseline parameters in the context of the Phoenix Trial, the benefit is small compared to the optimized two-stage adaptive policy and the simpler multi-period policy, for a population where the baseline rate of adverse event is 10% (doubles that of the actual trial, e.g. the intermediate scenario) or higher (e.g. the best scenario), the benefit can be much more significant. There are also scenarios where the asymptotically optimal policy performs poorly compared to the other policies. This is because in these scenarios the underlying asymptotic approximation is not valid. This suggests that computing the exactly optimal adaptive design (not asymptotically optimal) can lead to further improvements.

8. Discussion

Adaptive clinical trial is a special form of adaptive experimentation that is used for hypothesis testing. Beyond drug discovery, our framework for adaptive experimentation can be applied to other areas, such as adaptive A/B testing, randomized control trials in development economics and new product testing. With regard to adaptive A/B testing, it can help technology companies determine how they can improve key customer metrics. For example, a high growth company may want to test alternative ways of reducing the customer churn rate, from 20% to 18%. While 2% may seem like a small improvement, it can have a potentially large impact on the company's profitability. Our results from the numerical study, in particular, the best case scenario and intermediate scenario from Table 2 and Table 3, show that our adaptive policy can provide significant economic benefit to companies performing A/B testing. With respect to development economics, our model can be directly applied to trials testing whether micro-lending helps lift people out of poverty or makes them more indebted, and trials testing whether changing the structure of the voting forms positively affects voter turnouts in elections. Most of these experiments involve pilot programs to estimate how effective a given intervention policy is, before irreversibly launching the policy in large scale. In the abstraction of our model, the controls can represent the number of communities to run pilot program on, and accepting H_0 or H_a corresponds to abandoning the policy or launching the policy at scale. Our framework is also relevant to the entrepreneurship literature involving product development. In the real world, an entrepreneur tries to determine whether a novel product will be successful by performing a series of experiments that inform him or her about the prospects of the product, and eventually makes a decision whether to develop the product or not. Depending on whether the final decision is correctly or incorrectly made, there is an associated reward or penalty. In the abstraction of our model, the controls represent different experiments the entrepreneurs can run, and accepting H_0 or H_a corresponds to abandoning the product or investing on the product development.

Our modeling framework has several limitations. First, the sequential hypothesis testing framework does not control for the realized termination time. Although a heuristic is presented to control

for the expected termination time, it is possible that in some realizations of signal sequences, the log-likelihood ratio stays within the continuation region for a very long period of time. This may be undesirable from a practical point of view. Second, the terminal rewards and penalties $R_0.R_a, \kappa_0, \kappa_a$ are assumed to be constant independent of the termination time T. It is plausible that in some scenarios or applications, these may be dependent. For instance, a large T may lead to a lower reward R_a due to a shorter patent protection period. Third, all experiments have the same duration. This implies follow-up times are the same for all patients in the trial. The framework, for example, does not allow for clinical trials where the key outcome is time-to-event (TTE). Fourth, patient outcomes are perfectly observed. Specifically, the framework assumes no patients quit due to unforeseen circumstances during the trials. Fifth, our model does not allow testing of multiple (≥ 3) hypotheses, as this may involve generalizing the one-dimensional belief space to multi-dimensional which may be relatively more challenging to characterize and solve. Lastly, our model does not consider switching cost when the decision maker changes experiments during the trial. Extensions that address any of these limitations present opportunities for future development.

References

Ahuja V, Birge JR (2016) Response-adaptive designs for clinical trials: Simultaneous learning from multiple patients. *European Journal of Operational Research* 248(2):619–633.

Anderer A, Bastani H, Silberholz J (2022) Adaptive clinical trial designs with surrogates: When should we bother? *Management Science* 68(3):1982–2002.

Araman VF, Caldentey R (2009) Dynamic pricing for nonperishable products with demand learning. *Operations research* 57(5):1169–1188.

Araman VF, Caldentey RA (2021) Diffusion approximations for a class of sequential experimentation problems. *Management Science* .

Auer P, Cesa-Bianchi N, Fischer P (2002) Finite-time analysis of the multiarmed bandit problem. *Machine learning* 47(2):235–256.

Bastani H, Bayati M (2020) Online decision making with high-dimensional covariates. *Operations Research* 68(1):276–294.

Berry DA (1987) Interim analysis in clinical trials: the role of the likelihood principle. *The American Statistician* 41(2):117–122.

Bhatt DL, Stone GW, Mahaffey KW, Gibson CM, Steg PG, Hamm CW, Price MJ, Leonardi S, Gallup D, Bramucci E, et al. (2013) Effect of platelet inhibition with cangrelor during pci on ischemic events. *New England Journal of Medicine* 368(14):1303–1313.

Brekke KA, Øksendal B (1994) Optimal switching in an economic activity under uncertainty. *SIAM Journal on Control and Optimization* 32(4):1021–1036.

Chaudhuri SE, Ho MP, Irony T, Sheldon M, Lo AW (2018) Patient-centered clinical trials. *Drug discovery today* 23(2):395–401.

Cheng Y, Su F, Berry DA (2003) Choosing sample size for a clinical trial using decision analysis. *Biometrika* 90(4):923–936.

Chernoff H (1959) Sequential design of experiments. *The Annals of Mathematical Statistics* 30(3):755–770.

Chick S, Forster M, Pertile P (2017) A bayesian decision theoretic model of sequential experimentation with delayed response. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 79(5):1439–1462.

Chick SE, Gans N, Yapar Ö (2022) Bayesian sequential learning for clinical trials of multiple correlated medical interventions. *Management Science* 68(7):4919–4938.

Colton T (1963) A model for selecting one of two medical treatments. *Journal of the American Statistical Association* 58(302):388–400.

Davis MH, Zervos M (1994) A problem of singular stochastic control with discretionary stopping. *The Annals of Applied Probability* 226–240.

DiMasi JA, Grabowski HG, Hansen RW (2016) Innovation in the pharmaceutical industry: new estimates of r&d costs. *Journal of health economics* 47:20–33.

Duckworth K, Zervos M (2001) A model for investment decisions with switching costs. *Annals of Applied probability* 239–260.

Dyrssen H, Ekström E (2018) Sequential testing of a wiener process with costly observations. *Sequential Analysis* 37(1):47–58.

Food U, Administration D (2010) Guidance for industry and fda staff: Guidance for the use of bayesian statistics in medical device clinical trials.

Gordon Lan K, DeMets DL (1983) Discrete sequential boundaries for clinical trials. *Biometrika* 70(3):659–663.

Harrison JM, Sunar N (2015) Investment timing with incomplete information and multiple means of learning. *Operations Research* 63(2):442–457.

Henry E, Ottaviani M (2019) Research and the approval process: The organization of persuasion. *American Economic Review* 109(3):911–55.

Karatzas I, Sudderth WD (1999) Control and stopping of a diffusion process on an interval. *Annals of Applied Probability* 188–196.

Keskin NB, Zeevi A (2014) Dynamic pricing with an unknown demand model: Asymptotically optimal semi-myopic policies. *Operations research* 62(5):1142–1167.

Kouvelis P, Milner J, Tian Z (2017) Clinical trials for new drug development: Optimal investment and application. *Manufacturing & Service Operations Management* 19(3):437–452.

Kruizinga MD, Stuurman FE, Groeneveld GJ, Cohen AF (2019) The future of clinical trial design: the transition from hard endpoints to value-based endpoints. *Concepts and Principles of Pharmacology* 371–397.

Kwon HD, Lippman SA (2011) Acquisition of project-specific assets with bayesian updating. *Operations research* 59(5):1119–1130.

Lai TL, Robbins H, et al. (1985) Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics* 6(1):4–22.

Lobel I, Patel J, Vulcano G, Zhang J (2016) Optimizing product launches in the presence of strategic consumers. *Management Science* 62(6):1778–1799.

Matoglu MO, Vate JV, Wang H (2015) Solving the drift control problem. *Stochastic Systems* 5(2):324–371.

Montazerhodjat V, Chaudhuri SE, Sargent DJ, Lo AW (2017) Use of bayesian decision analysis to minimize harm in patient-centered randomized clinical trials in oncology. *JAMA oncology* 3(9):e170123–e170123.

Moore TJ, Zhang H, Anderson G, Alexander GC (2018) Estimated costs of pivotal trials for novel therapeutic agents approved by the us food and drug administration, 2015-2016. *JAMA internal medicine* 178(11):1451–1457.

Naghshvar M, Javidi T (2013) Active sequential hypothesis testing. *The Annals of Statistics* 41(6):2703–2738.

O'Brien PC, Fleming TR (1979) A multiple testing procedure for clinical trials. *Biometrics* 549–556.

Pallmann P, Bedding AW, Choodari-Oskooei B, Dimairo M, Flight L, Hampson LV, Holmes J, Mander AP, Odondi L, Sydes MR, et al. (2018) Adaptive designs in clinical trials: why use them, and how to run and report them. *BMC medicine* 16(1):1–15.

Pang G, Talreja R, Whitt W (2007) Martingale proofs of many-server heavy-traffic limits for markovian queues. *Probability Surveys* 4:193–267.

Pawitan Y, Hallstrom A (1990) Statistical interim monitoring of the cardiac arrhythmia suppression trial. *Statistics in medicine* 9(9):1081–1090.

Peskir G, Shiryaev A (2006) *Optimal stopping and free-boundary problems* (Springer).

Pocock SJ (1982) Interim analyses for randomized clinical trials: the group sequential approach. *Biometrics* 153–162.

Pretorius S (2016) Phase iii trial failures: costly, but preventable. *Applied Clinical Trials* 25(8/9):36.

Russo D, Van Roy B (2014) Learning to optimize via posterior sampling. *Mathematics of Operations Research* 39(4):1221–1243.

Ryan EG, Brock K, Gates S, Slade D (2020) Do we need to adjust for interim analyses in a bayesian adaptive trial design? *BMC medical research methodology* 20(1):1–9.

Sertkaya A, Wong HH, Jessup A, Beleche T (2016) Key cost drivers of pharmaceutical clinical trials in the united states. *Clinical Trials* 13(2):117–126.

Shiryaev AN (1967) Two problems of sequential analysis. *Cybernetics* 3(2):63–69.

Siegmund D (1985) *Sequential analysis: tests and confidence intervals* (Springer Science & Business Media).

Silva IR (2018) On the correspondence between frequentist and bayesian tests. *Communications in Statistics-Theory and Methods* 47(14):3477–3487.

Spiegelhalter DJ, Freedman LS, Parmar MK (1994) Bayesian approaches to randomized trials. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 157(3):357–387.

Sunar N, Birge JR, Vitavasiri S (2019) Optimal dynamic product development and launch for a network of customers. *Operations Research* 67(3):770–790.

Wald A (1945) Sequential method of sampling for deciding between two courses of action. *Journal of the American Statistical Association* 40(231):277–306.

Willan A, Kowgier M (2008) Determining optimal sample sizes for multi-stage randomized clinical trials using value of information methods. *Clinical Trials* 5(4):289–300.

Willan AR, Pinto EM (2005) The value of information and optimal clinical trial design. *Statistics in medicine* 24(12):1791–1806.

Woo J, Kim E, Sung TE, Lee J, Shin K, Lee J (2019) Developing an improved risk-adjusted net present value technology valuation model for the biopharmaceutical industry. *Journal of Open Innovation: Technology, Market, and Complexity* 5(3):45.

Zang Y, Lee JJ (2014) Adaptive clinical trial designs in oncology. *Chinese clinical oncology* 3(4).

The Electronic Companion to
Bayesian Adaptive Design of Clinical Trials: A Sequential Learning Approach

Appendix EC.1. Proof of Theorem 1

Proof of Theorem 1. Suppose (8) and (9) hold. Define

$$W_j^{(K)}(t) = \frac{L_j^{(K)}(t) - \mu_0^{(K)}(j)\,\lfloor Kt \rfloor}{\sigma_0^{(K)}(j)\sqrt{K}}. \tag{38}$$

Then it can be easily verified that under the null hypothesis, $W_j^{(K)}(t)$ is a martingale and $E\left[W_j^{(K)}(t)^2 \;\middle|\; H_0\right] = \frac{1}{\sigma_0^{(K)}(j)^2 K}(\lfloor Kt \rfloor\, \sigma_0^{(K)}(j)^2) \to t$ as $K \to \infty$. By applying the martingale FCLT (Pang et al. (2007) Theorem 8.1(ii)) to this process, we have

$$\text{Under } H_0,\ W_j^{(K)}(t) \Rightarrow B(t;0,1) \text{ as } K \to \infty. \tag{39}$$

Under the alternative hypothesis,

$$\begin{aligned} W_j^{(K)}(t) &= \frac{L_j^{(K)}(t) - \mu_a^{(K)}(j)\,\lfloor Kt \rfloor}{\sigma_0^{(K)}(j)\sqrt{K}} + \frac{(\mu_a^{(K)}(j) - \mu_0^{(K)}(j))\,\lfloor Kt \rfloor}{\sigma_0^{(K)}(j)\sqrt{K}} \\ &= \frac{\sigma_a^{(K)}(j)}{\sigma_0^{(K)}(j)} \frac{L_j^{(K)}(t) - \mu_a^{(K)}(j)\,\lfloor Kt \rfloor}{\sigma_a^{(K)}(j)\sqrt{K}} + \frac{(\mu_a^{(K)}(j) - \mu_0^{(K)}(j))\,\lfloor Kt \rfloor}{\sigma_0^{(K)}(j)\sqrt{K}}. \end{aligned}$$

Thus under H_a, $\tilde{W}_j^{(K)}(t) \triangleq W_j^{(K)}(t) - \frac{(\mu_a^{(K)}(j) - \mu_0^{(K)}(j))\lfloor Kt \rfloor}{\sigma_0^{(K)}(j)\sqrt{K}}$ is a martingale. Since $\lim_{K\to\infty}(\frac{\sigma_a^{(K)}(j)}{\sigma_0^{(K)}(j)}) = 1$ (c.f. Lemma EC.1), $E\left[\tilde{W}_j^{(K)}(t)^2 \;\middle|\; H_a\right] \to t$ as $K \to \infty$. Applying the martingale FCLT Theorem again we have

$$\text{Under } H_a,\ W_j^{(K)}(t) \Rightarrow B(t;\sqrt{\eta_j},1) \text{ as } K \to \infty. \tag{40}$$

Note that by (38),

$$L_j^{(K)}(t) = \sigma_0^{(K)}(j)\sqrt{K} W_j^{(K)}(t) + \mu_0^{(K)}(j)\,\lfloor Kt \rfloor, \tag{41}$$

and by Lemma EC.1 we have

$$\sqrt{K}\sigma_0^{(K)}(j) = \sqrt{\eta_j}, \tag{42}$$

$$K\mu_0^{(K)}(j) = -\frac{\eta_j}{2}. \tag{43}$$

Under H_0 we have (39). This together with (41), (42), (43) proves (12). Similarly under H_a we have (40). This together with (41), (42), (43) shows (13). □

LEMMA EC.1. *For any $j \in \mathcal{S}$ denote $\eta_j \triangleq \mu_a(j) - \mu_0(j)$ and suppose (8), (9) hold, then*

$$\lim_{K\to\infty} K\mu_a^{(K)}(j) = \frac{\eta_j}{2},\ \lim_{K\to\infty} K\mu_0^{(K)}(j) = -\frac{\eta_j}{2},\ \lim_{K\to\infty} K\sigma_a^{(K)2}(j) = \eta_j,\ \lim_{K\to\infty} K\sigma_0^{(K)2}(j) = \eta_j, \tag{44}$$

$$\lim_{K\to\infty} \sqrt{K}\frac{\left[\mu_a^{(K)}(j) - \mu_0^{(K)}(j)\right]}{\sigma_0^{(K)}(j)} = \sqrt{\eta_j}. \tag{45}$$

Proof of Lemma EC.1. Fix an experiment j and denote its signal set $\Omega = \{1, 2, 3, ...\}$. Without loss of generality assume $p_0^{(K)}(\cdot|j) \to \tilde{p}(\cdot|j) > 0$ uniformly. Denote for each $x \in \Omega$, $h_x^{(K)} = p_0^{(K)}(x|j)$ and $k_x^{(K)} = p_a^{(K)}(x|j) - p_0^{(K)}(x|j)$. The two sets $\{k_x^{(K)}\}_{x,K}$ and $\{h_x^{(K)}\}_{x,K}$ satisfy

$$\sum_x h_x^{(K)} = 1,\ h_x^{(K)} \in (0,1) \text{ for all } x, K, \tag{46}$$

$$\sum_x k_x^{(K)} = 0,\ h_x^{(K)} + k_x^{(K)} \in (0,1) \text{ for all } x, K. \tag{47}$$

Since (9) holds, for every $\epsilon > 0$ there exists k' such that whenever $K \geq k'$, $\sup_x |\frac{k_x^{(K)}}{h_x^{(K)}}| \leq \epsilon$. When this happens, we have

$$\mu_a^{(K)}(j) = \sum_x (h_x^{(K)} + k_x^{(K)}) \log \frac{h_x^{(K)} + k_x^{(K)}}{h_x^{(K)}} = \left[-(\sum_x k_x^{(K)}) + \frac{1}{2}(\sum_x \frac{k_x^{(K)2}}{h_x^{(K)}}) + O(\epsilon^3) \right] \tag{48}$$

$$\overset{(47)}{=} \frac{1}{2}(\sum_x \frac{k_x^{(K)2}}{h_x^{(K)}}) + O(\epsilon^3), \tag{49}$$

$$\mu_0^{(K)}(j) = \sum_x h_x^{(K)} \log \frac{h_x^{(K)} + k_x^{(K)}}{h_x^{(K)}} = -\frac{1}{2}(\sum_x \frac{k_x^{(K)2}}{h_x^{(K)}}) + O(\epsilon^3), \tag{50}$$

$$\sigma_a^{(K)2}(j) = \left[\sum_x (h_x^{(K)} + k_x^{(K)}) \left(\log \frac{h_x^{(K)} + k_x^{(K)}}{h_x^{(K)}} \right)^2 - \mu_a^{(K)}(j)^2 \right] = (\sum_x \frac{k_x^{(K)2}}{h_x^{(K)}}) + O(\epsilon^3), \tag{51}$$

$$\sigma_0^{(K)2}(j) = \left[\sum_x h_x^{(K)} \left(\log \frac{h_x^{(K)} + k_x^{(K)}}{h_x^{(K)}} \right)^2 - \mu_0^{(K)}(j)^2 \right] = (\sum_x \frac{k_x^{(K)2}}{h_x^{(K)}}) + O(\epsilon^3). \tag{52}$$

Hence as $K \to \infty$,

$$\frac{\mu_a^{(K)}(j)}{\mu_0^{(K)}(j)} \to 1,\ \frac{\mu_a^{(K)}(j)}{\sigma_a^{(K)2}(j)} \to \frac{1}{2},\ \frac{\mu_a^{(K)}(j)}{\sigma_0^{(K)2}(j)} \to \frac{1}{2}. \tag{53}$$

If (8) holds, then

$$\lim_{K\to\infty} K\mu_a^{(K)}(j) = \frac{\eta_j}{2},\ \lim_{K\to\infty} K\mu_0^{(K)}(j) = -\frac{\eta_j}{2},\ \lim_{K\to\infty} K\sigma_a^{(K)2}(j) = \eta_j,\ \lim_{K\to\infty} K\sigma_0^{(K)2}(j) = \eta_j,$$

and

$$\lim_{K\to\infty}\sqrt{K}\frac{(\mu_a^{(K)}(j)-\mu_0^{(K)}(j))}{\sigma_0^{(K)}(j)}=\frac{K(\mu_a^{(K)}(j)-\mu_0^{(K)}(j))}{\sqrt{K}\sigma_0^{(K)}(j)}=\sqrt{\eta_j}.$$

□

Appendix EC.2. Proof of Lemma 1, 2, 3 and 4

Proof of Lemma 1.

Without loss of generality, we can assume $H_0: dL_t=\sqrt{\eta}dB_t$, $H_a: dL_t=\eta t+\sqrt{\eta}dB_t$.

Suppose prior is $\pi_0=Pr(H_a)$ and we observe L_t at time t. Denote $\phi_t=\frac{Pr(L_t|H_a)}{Pr(L_t|H_0)}$ to be the likelihood ratio, then the posterior belief π_t satisfies

$$\pi_t=\frac{\frac{\pi_0}{1-\pi_0}\phi_t}{1+\frac{\pi_0}{1-\pi_0}\phi_t}. \tag{54}$$

By an application of Ito's formula, we have

$$d\pi_t=\pi_t(1-\pi_t)dL_t-\pi_t^2(1-\pi_t)\eta dt. \tag{55}$$

Note that $dL_t=\eta\pi_t dt+\sqrt{\eta}dB_t$, and hence

$$=\pi_t(1-\pi_t)(\eta\pi_t dt+\sqrt{\eta}dB_t)-\pi_t^2(1-\pi_t)\eta dt \tag{56}$$

$$=\sqrt{\eta}\pi_t(1-\pi_t)dB_t. \tag{57}$$

□

Proof of Lemma 2. By a standard result in the sequential hypothesis testing paradigm

$$\log\left(\frac{\pi_t}{1-\pi_t}\right)=L_t+\log\left(\frac{\pi_0}{1-\pi_0}\right),$$

i.e. $\pi_t=\frac{\pi_0 e^{L_t}}{(1-\pi_0)+\pi_0 e^{L_t}}$ and $L_t=ln\frac{1-\pi_0}{\pi_0}\frac{\pi_t}{1-\pi_t}$. We further define the following two quantities

$$\hat{u}=ln\frac{1-\pi_0}{\pi_0}\frac{u}{1-u},\ \hat{l}=ln\frac{1-\pi_0}{\pi_0}\frac{l}{1-l}. \tag{58}$$

Suppose H_0 is true, i.e. $dL_t=-\frac{\eta_j}{2}dt+\sqrt{\eta_j}dB_t$. Denote the type I and II errors as

$$\mathcal{P}_\alpha(y)=Pr(L_t \text{ hits } \hat{u} \text{ before } \hat{l}\mid H_0, L_0=y), \tag{59}$$

$$\mathcal{P}_\beta(y)=Pr(L_t \text{ hits } \hat{l} \text{ before } \hat{u}\mid H_a, L_0=y). \tag{60}$$

Using a standard first-order analysis we have

$$\mathcal{P}_\alpha(L_t) = E_{dL_t}\mathcal{P}_\alpha(L_t + dL_t) \tag{61}$$

$$= \mathcal{P}_\alpha(L_t) + E[\mathcal{P}'_\alpha(L_t)dL_t] + E[\mathcal{P}''_\alpha(L_t)dL_t^2/2] + o(dt) \tag{62}$$

$$= \mathcal{P}_\alpha(L_t) + \mathcal{P}'_\alpha(L_t)(-\eta_j/2)dt + \mathcal{P}''_\alpha(L_t)((\sqrt{\eta_j})^2/2)dt + o(dt), \tag{63}$$

where the last inequality follows from Ito's formula (under H_0, $dL_t = -(\eta_j/2)dt + \sqrt{\eta_j}dB_t$).

Thus $\mathcal{P}_\alpha(y)$ satisfies the following ODE

$$-\mathcal{P}'_\alpha(y) + \mathcal{P}''_\alpha(y) = 0, \tag{64}$$

$$\mathcal{P}_\alpha(\hat{l}) = 0,\ \mathcal{P}_\alpha(\hat{u}) = 1. \tag{65}$$

which has the solution $\mathcal{P}_\alpha(y) = \frac{e^y - e^{\hat{l}}}{e^{\hat{u}} - e^{\hat{l}}}$. Similarly we have $\mathcal{P}_\beta(y) = \frac{e^{-y} - e^{-\hat{u}}}{e^{-\hat{l}} - e^{-\hat{u}}}$. Substituting $\hat{l}, \hat{u}$ as in (58) gives the desired result. □

Proof of lemma 3. Given l, u fixed, for a given control j and current belief $y \in (l, u)$, we will show that expected termination time $E_{y,l,u}(T; j)$ satisfies

$$E_{y,l,u}(T; j) = \frac{2}{\eta_j(u-l)}\left[(u-y)\psi_l(y) + (y-l)\psi_u(y)\right], \tag{66}$$

where $\psi_z(y) = -(2y-1)\ln\left(\frac{y}{1-y}\right) + (2z-1)\ln\left(\frac{z}{1-z}\right)$.

Denote $\mathcal{T}(y) = E_{y,l,u}(T; j)$. Then $\mathcal{T}(y)$ satisfies the following ordinary differential equation

$$1 + \frac{\eta_j y^2(1-y)^2}{2}\mathcal{T}''(y) = 0. \tag{67}$$

The solution to (67) gives (where C_1, C_2 are constants to be determined)

$$\mathcal{T}(y; C_1, C_2) = \frac{1}{\eta_j}(-4y+2)ln(\frac{y}{1-y}) + \frac{4}{\eta_j} + C_1 y + C_2. \tag{68}$$

Applying the boundary conditions $\mathcal{T}(l) = \mathcal{T}(u) = 0$, we have

$$\mathcal{T}(y) = \frac{2}{\eta_j(u-l)}\left[-(2y-1)ln\left(\frac{y}{1-y}\right)(u-l) + (2l-1)(u-y)ln\left(\frac{l}{1-l}\right) + (2u-1)(y-l)ln\left(\frac{u}{1-u}\right)\right] \tag{69}$$

$$= \frac{2}{\eta_j(u-l)}\left[-(2y-1)ln\left(\frac{y}{1-y}\right)(u-y+y-l) + (2l-1)(u-y)ln\left(\frac{l}{1-l}\right) + (2u-1)(y-l)ln\left(\frac{u}{1-u}\right)\right] \tag{70}$$

$$= \frac{2}{\eta_j(u-l)}\left[(u-y)\psi_l(y) + (y-l)\psi_u(y)\right]. \tag{71}$$

□

Proof of lemma 4. Let $V_{\Pi^*}(\cdot), V_{\hat{\Pi}^*}(\cdot)$ denote the value function associated with $\Pi^*, \hat{\Pi}^*$ respectively. Suppose $l(\Pi^*) > l(\hat{\Pi}^*)$, then this means for $\epsilon > 0$ sufficiently small, $V_{\hat{\Pi}^*}(l(\hat{\Pi}^*) + \epsilon) > V_{\Pi^*}(l(\hat{\Pi}^*) + \epsilon)$, a contradiction. Similarly the same conclusion can be deduced for $u(\Pi^*)$ and $u(\hat{\Pi}^*)$.

Appendix EC.3. Proof of Theorem 2 and 3

In this section, we will prove Theorem 2 by constructing the OIP policy. The proof is organized into four main steps. First, we define some relevant quantities that will be useful in the proof. Second, we show we can construct a value function given any *lower critical belief* that lies within certain ranges. Third we show that the set of value functions constructed in this way possess certain desired properties and are ordered. Lastly combining all the previous steps, we show that there exists an *interval policy* which is the required OIP. The proof of Theorem 3 mimics the exact same steps and is omitted.

EC.3.1. Step I: Notations and Definitions

To express compactly the differential equations that appear frequently in the remainder of the section, we will denote the differential operator

$$\Gamma f(y) \triangleq \frac{y^2(1-y)^2 f''(y)}{2}, \tag{72}$$

whenever $f(\cdot)$ is twice continuously differentiable. For the convenience of presenting the results, we will also define the following terms "partially underneath", "over-touch" and "lie above". These definitions are intuitive and they are illustrated in Figure 10.

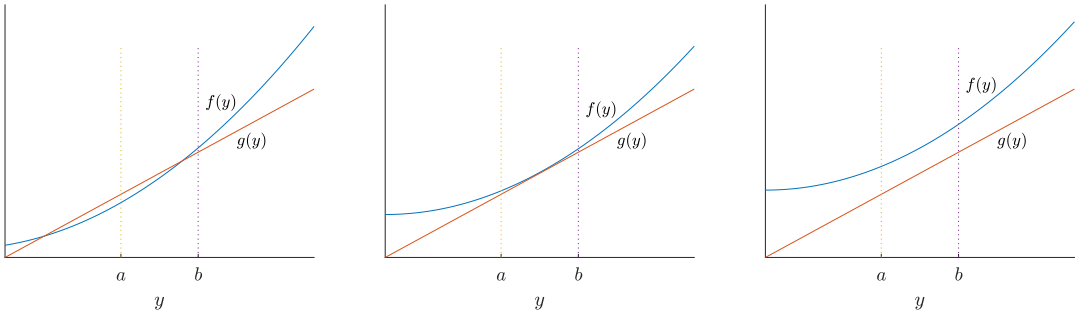
Definition 2. Let $f(y)$ and $g(y)$ be two continuous functions over an interval $A = (a, b) \subset \mathcal{R}$

i) $f(y)$ is *partially underneath* $g(y)$ over A if $f(y) < g(y)$ over some sub-interval $A_s \subset A$.

ii) $f(y)$ *over-touches* $g(y)$ over A if $f(y) \geq g(y)$ over A and there exists at least one $y_0 \in A$ s.t. $f(y_0) = g(y_0)$.

iii) $f(y)$ *lies above* $g(y)$ if $f(y) > g(y)$ over A.

Figure 10 (Color) Definitions of *partially underneath, over-touch, and lie above.*



(a) $f(\cdot)$ is partially underneath $g(\cdot)$ over (a,b).

(b) $f(\cdot)$ over-touches $g(\cdot)$ over (a,b).

(c) $f(\cdot)$ lies above $g(\cdot)$ over (a,b).

Moreover, in the remainder of the section, we will assume the set of controls $\{(\eta_i, c_i)\}_{i=1}^{N}$ is numbered as in (24), (25), (26) and (27) with $\tilde{N}$ corresponding to the control with the highest information quality on the *efficient frontier*. To proceed further, it will be convenient to define the following set of thresholds $\Psi = \{\xi_i\}_{i=0}^{\tilde{N}}$. The thresholds are chosen to ensure that the resulting constructed value function $V_l(\cdot)$ is twice continuously differentiable at the switch points (to be identified), and hence twice continuously differentiable over the whole interval $[l,1)$ (shown later in Lemma EC.3).

Definition 3.

$$\xi_i = \begin{cases} \frac{1}{\lambda}\left(\frac{c_{i+1}\eta_i - c_i\eta_{i+1}}{\eta_{i+1}-\eta_i}\right) \text{ , if } i \in \{1,2,...,\tilde{N}-1\}, & (73) \\ -\frac{c_1}{\lambda} \text{ , if } i = 0, & (74) \\ \infty \text{ , if } i = \tilde{N}, & (75) \end{cases}$$

and it can be shown that this set of thresholds satisfies the following.

Lemma EC.2. *Let the set $\Psi = \{\xi_i\}_{i=0}^{\tilde{N}}$ be defined as in (73), (74) and (75). Then we have*

$$-\frac{c_1}{\lambda} = \xi_0 < \xi_1 < ... < \xi_{\tilde{N}-1} < \xi_{\tilde{N}} = \infty. \tag{76}$$

Proof of Lemma EC.2. $\xi_{i+1} > \xi_i$ if and only if $\frac{c_{i+2}\eta_{i+1} - c_{i+1}\eta_{i+2}}{\eta_{i+2}-\eta_{i+1}} > \frac{c_{i+1}\eta_i - c_i\eta_{i+1}}{\eta_{i+1}-\eta_i}$ and we will verify that the latter holds. Since (η_{i+2}, c_{i+2}) is on the *efficient frontier*, we have $c_{i+2} > c_{i+1} + (\eta_{i+2} - \eta_{i+1})\frac{c_{i+1}-c_i}{\eta_{i+1}-\eta_i}$, and thus

$$\frac{c_{i+2}\eta_{i+1} - c_{i+1}\eta_{i+2}}{\eta_{i+2}-\eta_{i+1}} > \frac{\left[c_{i+1} + (\eta_{i+2}-\eta_{i+1})\frac{c_{i+1}-c_i}{\eta_{i+1}-\eta_i}\right]\eta_{i+1} - c_{i+1}\eta_{i+2}}{\eta_{i+2}-\eta_{i+1}} \tag{77}$$

$$= \frac{c_{i+1}\eta_i - c_i\eta_{i+1}}{\eta_{i+1} - \eta_i}. \tag{78}$$

Moreover since $\eta_2 > \eta_1$, $\xi_1 > -\frac{c_1}{\lambda}$ if and only if $\frac{\eta_2 c_1 - \eta_1 c_2}{\eta_2 - \eta_1} < c_1$, which is equivalent to $c_2 > c_1$. □

Lastly, the following two quantities will be useful in presenting the proofs.

Definition 4.

$$L_c = \arg\min_{\{y\in[0,1]\}} G(y),$$
$$l_{max} = \min\{\frac{-\xi_0 + R_0}{R_0 + \kappa_0}, L_c\}.$$

where $G(\cdot)$ is defined as in (22). Note that $\frac{-\xi_0+R_0}{R_0+\kappa_0}$ is equivalent to $\frac{c_1+R_0\lambda}{\lambda(R_0+\kappa_0)}$.

EC.3.2. Step II: Construction of $V_l(\cdot)$ for a given l

Given l s.t. $0 < l \le l_{max}$, we will now construct a trial function $V_l(\cdot)$ such that it is C^2 over $(l, 1)$ and satisfies

$$V_l(l) = -(R_0 + \kappa_0)l + R_0,\ V_l'(l) = -(R_0 + \kappa_0), \tag{79}$$

$$V_l''(y) \ge 0 \text{ for all } y \in (l, 1), \tag{80}$$

and later we will show that we can choose an appropriate value of l s.t. $V_l(\cdot)$ solves

$$\max_{j\in\{1,2,\ldots,N\}} \left\{-c_j - \lambda V(y) + \frac{1}{2}\eta_j y^2(1-y)^2 V''(y)\right\} = 0, \tag{81}$$

and over-touches $G(\cdot)$ at l and some u with $0 < l < u < 1$.

The construction process starts by taking an initial guess of l. Then, it sequentially identifies switch points and controls, based on whether the function's value crosses the thresholds in Ψ. Given l, we set $\zeta_0 = l$ and define (see Figure 11)

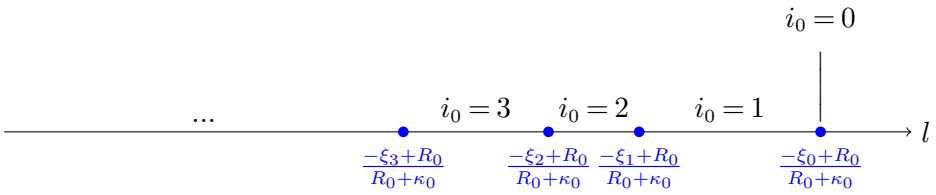
$$i_0 = \begin{cases} j \text{ , if } l \in \left[\frac{-\xi_j + R_0}{R_0 + \kappa_0}, \frac{-\xi_{j-1} + R_0}{R_0 + \kappa_0}\right) \text{ for some } j \text{ s.t. } 1 \le j \le n, & (82) \\ 0 \text{ , if } l = \frac{-\xi_0 + R_0}{R_0 + \kappa_0}. & (83) \end{cases}$$

If $l=\frac{-\xi_0+R_0}{R_0+\kappa_0}$ (i.e. $V_l(l)=\xi_0=-\frac{c_1}{\lambda}$), we trivially construct $V_l(\cdot)$ as $V_l(y)=-(R_0+\kappa_0)y+R_0$ over $[l,1)$, and the construction process for $V_l(\cdot)$ is complete. If $l<\frac{-\xi_0+R_0}{R_0+\kappa_0}$, then (82) is equivalent to (c.f. Lemma EC.2)

$$i_0=\max\left\{i\in\{1,2,...,\tilde{N}\}\mid -(R_0+\kappa_0)l+R_0>\xi_{i-1}\right\}, \tag{84}$$

and it can be easily verified that $i_0\geq 1$.

Figure 11 The Thresholds that determine i_0.



Let $f_0(\cdot)$ be C^2 over $(l,1)$ that solves

$$f_0(l)=-(R_0+\kappa_0)l+R_0, f_0'(l)=-(R_0+\kappa_0), \tag{85}$$

$$-c_{i_0}+\eta_{i_0}y^2(1-y)^2f_0''(y)/2=\lambda f_0(y). \tag{86}$$

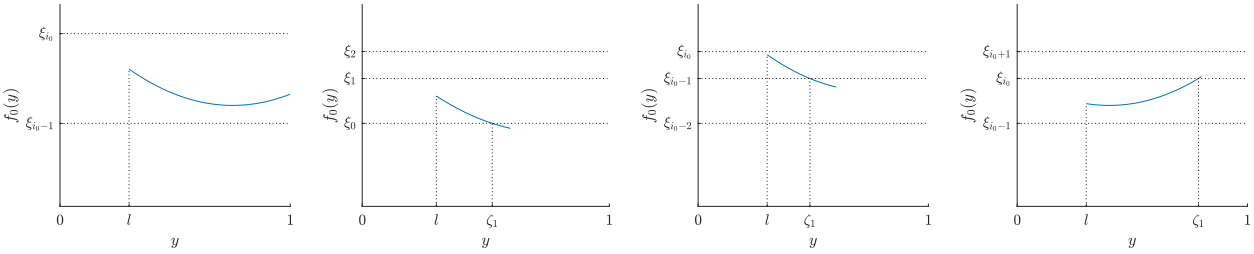
Since $f_0'(l)<0$, there exists $\delta_0>0$ s.t. $\{f_0(y),y\in(l,l+\delta_0)\}\cap\Psi=\emptyset$ and $f_0(l+\delta_0)\in(\xi_{i_0-1},\xi_{i_0})$. We identify the first switch point (where by default we denote $\inf\{\emptyset\}=+\infty$)

$$\zeta_1\triangleq\inf\{y\in(l+\delta_0,1)\mid f_0(y)\leq\xi_{i_0-1}\text{ or }f_0(y)\geq\xi_{i_0},f_0'(y)\neq 0\}. \tag{87}$$

There are four possible cases (see Figure 12): i.1) $f_0(y)$ never falls out of the range (ξ_{i_0-1},ξ_{i_0}), in which case $\zeta_1=\infty$ (note that $\xi_{i_0}=+\infty$ when $i_0=\tilde{N}$); ii.1) $f_0(y)$ first reaches ξ_{i_0-1} and $i_0=1$; iii.1) $f_0(y)$ first reaches ξ_{i_0-1} and $i_0\geq 2$ and iv.1) $f_0(y)$ first reaches ξ_{i_0}. In particular, for i,1) and ii.1) the construction of $V_l(\cdot)$ is complete and for iii.1) and iv.1) the construction is not yet complete and we will need to identify the next switch point (if it exists). More specifically in each case, we will proceed as follows:

i.1) Suppose $\zeta_1=\infty$ then we set $V_l(y)=f_0(y)$ for all $y\in[l,1)$ and the construction of $V_l(\cdot)$ is complete.

Figure 12 (Color) Four cases of the first switch point ζ_1



i.1), $\zeta_1 = \infty$. ii.1), $f_0(\zeta_1) = \xi_{i_0-1}$, $i_0 = 1$. iii.1), $f_0(\zeta_1) = \xi_{i_0-1}$, $i_0 \geq 2$. iv.1), $f_0(\zeta_1) = \xi_{i_0}$.

ii.1) Suppose $\zeta_1 = \inf\{y \in (l,1) \mid f_0(y) \leq \xi_{i_0-1}\} < 1$ (this happens if and only if $f_0'(y) < 0$) and $i_0 = 1$, then let $i_1 = 0$ and define $f_1(\cdot)$ over $[\zeta_1, 1)$ as

$$f_1(\zeta_1) = f_0(\zeta_1),\ f_1'(\zeta_1) = f_0'(\zeta_1), \tag{88}$$

$$f_1(y) = f_0'(\zeta_1)(y - \zeta_1) + f_0(\zeta_1). \tag{89}$$

Set $V_l(y) = f_0(y)$ on $[l, \zeta_1)$, $V_l(y) = f_1(y)$ on $[\zeta_1, 1)$ and the construction of $V_l(\cdot)$ is complete.

iii.1) If $\zeta_1 = \inf\{y \in (l,1) \mid f_0(y) \leq \xi_{i_0-1}\} < 1$ (this happens if and only if $f_0'(y) < 0$) and $i_0 \geq 2$, then let $i_1 = i_0 - 1$ and define $f_1(\cdot)$ over $[\zeta_1, 1)$ as

$$f_1(\zeta_1) = f_0(\zeta_1),\ f_1'(\zeta_1) = f_0'(\zeta_1), \tag{90}$$

$$-c_{i_1} + \eta_{i_1}\Gamma f_1(y) = \lambda f_1(y),\ y \in (\zeta_1, 1). \tag{91}$$

iv.1) If $\zeta_1 = \inf\{y \in (l,1) \mid f_0(y) \geq \xi_{i_0}\} < 1$ (this happens if and only if $f_0'(y) > 0$), then let $i_1 = i_0 + 1$ and define $f_1(\cdot)$ over $[\zeta_1, 1)$ as

$$f_1(\zeta_1) = f_0(\zeta_1),\ f_1'(\zeta_1) = f_0'(\zeta_1), \tag{92}$$

$$-c_{i_1} + \eta_{i_1}\Gamma f_1(y) = \lambda f_1(y),\ y \in (\zeta_1, 1). \tag{93}$$

In iii.1) or iv.1), since $f_1'(\zeta_1) = f_0'(\zeta_1) \neq 0$, there exists $\delta_1 > 0$ s.t. $\{f_0(y), y \in (\zeta_1, \zeta_1 + \delta_1)\} \cap \Psi = \emptyset$ and $f_1(\zeta_1 + \delta_1) \in (\xi_{i_1-1}, \xi_{i_1})$. Define the next switch point

$$\zeta_2 \triangleq \inf\{y \in (\zeta_1 + \delta_1, 1) \mid f_1(y) \leq \xi_{i_1-1} \wedge f_1(y) \geq \xi_{i_1}, f_1'(y) \neq 0\}, \tag{94}$$

and repeat the same process. Continuing in the same way, we iteratively identify $\{\zeta_k\}_{k\geq 0}, \{i_k\}_{k\geq 0}$ and functions $\{f_k(\cdot)\}_{k\geq 0}$ that satisfy

$$0 < l \triangleq \zeta_0 < \zeta_1 < \zeta_2 < ...\zeta_{\tilde{D}} < \zeta_{\tilde{D}+1} \triangleq 1, \tag{95}$$

$$i_k - i_{k-1} = \begin{cases} -1, \text{ if } f'_{i_{k-1}}(z_k) < 0, \\ 1, \text{ if } f'_{i_{k-1}}(z_k) > 0, \end{cases} \tag{96}$$

$$-c_{i_k} + \eta_{i_k}\Gamma f_k(y) - \lambda f_k(y) = 0, \ y \in (\zeta_k, 1), \tag{97}$$

$$f_k(\zeta_k) = f_{k-1}(\zeta_k), \ f'_k(\zeta_k) = f'_{k-1}(\zeta_k), \tag{98}$$

where $\tilde{D} \geq 0$ is the total number of switch points arisen from the above construction process, with $\zeta_{\tilde{D}+1}$ defined to be 1. Moreover, define for $k \in \{0, 1, ..., \tilde{D}\}$ the associated control function $m_l(\cdot)$ as

$$m_l(y) = i_k \text{ , if } y \in [\zeta_k, \zeta_{k+1}), \tag{99}$$

and we will omit the subscript l (i.e. denote it as $m(\cdot)$) whenever doing so does not cause confusion. We have thus constructed $V_l(\cdot)$ for any given $l \in (0, l_{max}]$ to be

$$V_l(y) = f_k(y) \text{ , if } y \in [\zeta_k, \zeta_{k+1}). \tag{100}$$

EC.3.3. Step III: The Properties & Orderings of $V_l(\cdot)$

Now we are ready to prove some properties of $V_l(\cdot)$ and show that the set $\{V_l(\cdot)\}_{l\in(0,l_{max}]}$ is ordered. The former is given by Lemma EC.3 and the latter by Lemma EC.4.

LEMMA EC.3 **(Properties of V_l)**.

Fix $l \in (0, l_{max}]$ and let V_l be constructed as in Step II (§EC.3.2), then

a) On $(l, 1)$, $V_l(\cdot)$ is twice differentiable. The sequence of switch points $\{\zeta_k(l)\}$ is finite. The sequence $\{i_k(l)\}$ is unimodally consecutive.

b) $V_l''(y) \geq 0$ for all $y \in (l, 1)$, i.e. $V_l(\cdot)$ is convex. If $0 \notin \{i_k(l)\}_{k\geq 1}$, then in $(l, 1)$ V_l has at most one local minimum and cannot have any local maximum.

c) If at any $y_0 \in (l,1)$, $m(y_0) \geq 2$ *or* $m(y_0) = 1$ *and* $V_l(y_0) > -\frac{c_1}{\lambda}$, *then* $V_l''(y_0) > 0$. *If at any* $y_0 \in (l,1)$, $m(y_0) = 0$, *then* $V_l''(y) = 0$ *for all* $y \geq y_0$.

d) For $y \in (l,1)$ *s.t.* $\Gamma V_l(y) \notin \Psi$, *we have* $m(y) = j$, *where* $j \in \{1,2,...,\tilde{N}\}$ *satisfies*

$$\frac{c_j - c_{j-1}}{\eta_j - \eta_{j-1}} < \Gamma V_l(y) < \frac{c_{j+1} - c_j}{\eta_{j+1} - \eta_j}. \tag{101}$$

Moreover, for $k \in \{1,2,3,...,\tilde{D}\}$ *the* $\zeta_k(l)$*'s satisfy*

$$\Gamma V_l(\zeta_k) = \frac{c_{i_{k-1}} - c_{i_k}}{\eta_{i_{k-1}} - \eta_{i_k}}. \tag{102}$$

e) If $0 \in \{i_k(l)\}_{k=0}^{\tilde{D}}$, *we must have*

$$i_{\tilde{D}}(l) = 0,\ \zeta_{\tilde{D}}(l) = \inf\{\Gamma V_l(y) \leq 0\},\ \Gamma V_l(\zeta_{\tilde{D}}(l)) = 0,\ V_l'(\zeta_{\tilde{D}}(l)) \leq 0. \tag{103}$$

Next we will show that the set $\{V_l(\cdot)\}_{l\in(0,l_{max}]}$ can be ordered. Because the $V_l(\cdot)$'s are defined over different domains, we extend all of them to be defined over $(0,1)$ so that their values can be directly compared. Specifically, for each $l \in (0, l_{max}]$, $V_l(\cdot)$ will be extended to $\hat{V}_l(\cdot)$, defined below.

$$\hat{V}_l(y) \triangleq \begin{cases} -(R_0 + \kappa_0)y + R_0 \text{ , if } y \in (0,l), & (104) \\ V_l(y) \text{ , if } y \in (l,1), & (105) \end{cases}$$

and Lemma EC.4 states that the $\hat{V}_l(\cdot)$'s are ordered as in Figure 13, with a larger l giving a uniformly smaller function $\hat{V}_l(\cdot)$.

LEMMA EC.4 **(Ordering of** $\hat{V}_l$**)**. *Fix* $l \in (0, l_{max}]$ *and let* $V_l(\cdot)$ *be constructed as in Step II (§EC.3.2),* $\hat{V}_l(\cdot)$ *be defined as in (104), then*

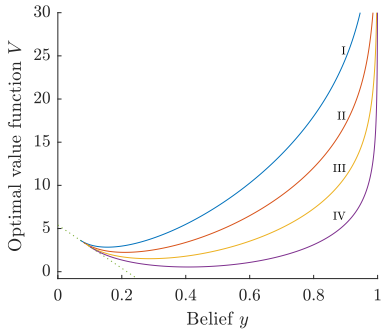
a) $V_l(\cdot)$ *lies above* $G(\cdot)$ *over* $[L_c, 1)$ *for* l *sufficiently small,* $V_{l_{max}}(\cdot)$ *is partially underneath* $G(\cdot)$ *over* $[L_c, 1)$. *If* $0 \in \{i_k\}_{k\geq 1}$, *then* $V_l(\cdot)$ *is partially underneath* $G(\cdot)$ *over* $[L_c, 1)$.

b) For a given $l \in (0, l_{max})$, $V_l''(y) > 0$ *for all* $y \in (l, l_{max}]$.

c) The functions $\{\hat{V}_l(\cdot)\}_{l\in(0,l_{max}]}$ *are ordered, with a larger* l *producing a uniformly smaller function* $\hat{V}_l(\cdot)$. *That is, for any* l_1, l_2 *s.t.* $0 < l_1 < l_2 \leq l_{max}$, $\hat{V}_{l_1}(y) > \hat{V}_{l_2}(y)$ *for all* $y \in [l_2, 1)$.

The proofs of Lemma EC.3 and EC.4 require additional supplementary Lemmas, and we defer them to §EC.3.5. The last step (§EC.3.4) combines all the results we have so far to prove Theorem 2.

Figure 13 (Color) Ordering of $\hat{V}_l(\cdot)$: curves I through IV represent $l=0.07$, $l=0.08$, $l=0.09$, $l=0.1$, respectively.



Note. The data used are: $\{\eta_i\}_{i=1}^6=\{10,20,30,40,50,60\}, \{c_i\}_{i=1}^6=\{8,14,24,35,48,64\}, \lambda=5, R_0=5.4, \kappa_0=19.6$ (R_a and κ_a's values are not needed to produce the Figure).

EC.3.4. Step IV: Proof of Theorem 2

Proof of Theorem 2. We will now prove Theorem 2. In particular, we will show that there exists an *interval policy* $(M, m(\cdot), \{i_k\}_{k=0}^D, \{\zeta_k\}_{k=0}^{D+1}, l, u)$ with $\{i_k\}$ *unimodally consecutive*, whose value function $V(\cdot)$ is continuously differentiable at l and u (the policy's lower and upper critical beliefs), $V(\cdot)\in C^2$ over $(0,1)\backslash\{l,u\}$, and further satisfies the followings

$$V(y)\geq G(y), \tag{106}$$

$$V(l)=-(R_0+\kappa_0)l+R_0,\ V'(l)=-(R_0+\kappa_0),\ V(u)=(R_a+\kappa_a)u-\kappa_a,\ V'(u)=R_a+\kappa_a, \tag{107}$$

$$-c_j+\eta_j y^2(1-y)^2V''(y)/2\leq \lambda V(y)\ \text{, for all } y\in(0,1)/\{l,u\} \text{ and all } j\in\{1,2,...,N\}, \tag{108}$$

$$-c_j+\eta_j y^2(1-y)^2V''(y)/2=\lambda V(y)\ \text{, for all } y\in(l,u) \text{ and some } j\in\{1,2,...,N\}. \tag{109}$$

Moreover, we will show that the above interval policy specified is optimal. Under this policy, the optimal action is to reject H_a when $\pi_t\in[0,l)$, to reject H_0 when $\pi_t\in[u,1]$, and to continue experimenting when $\pi_t\in[l,u)$.

By Lemma EC.3 (a) and Lemma EC.4 (a),(c), there exists a value l s.t. its associated control sequence $\{i_k(l)\}$ is *unimodally consecutive*, $0\notin\{i_k(l)\}$ and its associated value function $\hat{V}_l(\cdot)$ satisfies the following

$$\hat{V}_l(l)=-(R_0+\kappa_0)l+R_0,\ \hat{V}_l'(l)=-(R_0+\kappa_0)$$

$$\hat{V}_l(u) = (R_a + \kappa_a)u - \kappa_a, \hat{V}_l'(u) = R_a + \kappa_a.$$

i.e. $\hat{V}_l(\cdot)$ over-touches $G(\cdot)$ for some $u \in (l,1)$, $\hat{V}_l(\cdot)$ is continuously differentiable at l and u (the policy's *lower* and *upper critical beliefs*), is twice differentiable at each of the switching beliefs $\zeta_1(l), \zeta_2(l), ..., \zeta_K(l)$. By Lemma EC.3 (b), $\hat{V}_l(y) \geq G(y)$ over $[L_c, 1)$. We will now establish

$$-c_j + \eta_j y^2(1-y)^2 \hat{V}_l''(y)/2 \leq \lambda \hat{V}_l(y) \text{ , for all } y \in (0,1)\backslash\{l,u\} \text{ and all } j \in \{1,2,...,N\}. \tag{110}$$

For convenience, we shall use the following terminology: we say that control j dominates control i at y if

$$-c_i + \eta_i y^2(1-y)^2 \hat{V}_l''(y)/2 \leq -c_j + \eta_j y^2(1-y)^2 \hat{V}_l''(y)/2. \tag{111}$$

Let F_k be the collection of intervals of y s.t. $m(y) = k$, i.e. $-c_k + \eta_k \Gamma V_l(y) - \lambda \hat{V}_l(y) = 0$. By Lemma EC.3 (d), for all $y \in F_k$ we have

$$\frac{c_k - c_{k-1}}{\eta_k - \eta_{k-1}} \leq \Gamma \hat{V}_l(y) \leq \frac{c_{k+1} - c_k}{\eta_{k+1} - \eta_k}. \tag{112}$$

Fix $y \in F_k$. For (η_i, c_i) on the *efficient frontier* s.t. $1 \leq i < k$, we have $c_i < c_k$, $\eta_i < \eta_k$ and

$$-c_i + \eta_i \Gamma \hat{V}_l(y) - \lambda \hat{V}_l(y) \leq 0 \Leftrightarrow -c_i + \eta_i \Gamma \hat{V}_l(y) - \lambda \hat{V}_l(y) \leq -c_i + \eta_i \Gamma \hat{V}_l(y) - \lambda \hat{V}_l(y) \tag{113}$$

$$\Leftrightarrow \frac{c_k - c_i}{\eta_k - \eta_i} \leq \Gamma \hat{V}_l(y), \tag{114}$$

which is true by (112) and the fact that (η_i, c_i) is on the *efficient frontier*.

For (η_i, c_i) on the *efficient frontier* s.t. $i > k$, we have $c_i > c_k$, $\eta_i > \eta_k$ and

$$-c_i + \eta_i \Gamma \hat{V}_l(y) - \lambda \hat{V}_l(y) \leq 0 \Leftrightarrow -c_i + \eta_i \Gamma \hat{V}_l(y) - \lambda \hat{V}_l(y) \leq -c_i + \eta_i \Gamma \hat{V}_l(y) - \lambda \hat{V}_l(y) \tag{115}$$

$$\Leftrightarrow \frac{c_i - c_k}{\eta_i - \eta_k} \geq \Gamma \hat{V}_l(y), \tag{116}$$

which is again true by (112) and the fact that (η_i, c_i) lies on the *efficient frontier*.

If $i \in \{1, 2, ..., N\}$ is not on the *efficient frontier*, then it can be easily verified that for every $y \in (l, 1)$, there exists at least one $j \in \{1, 2, ..., \tilde{N}\}$ s.t. control i is dominated by j at y. Thus (110) is established. Now define

$$V(y) \triangleq \begin{cases} \hat{V}_l(y) \text{ , if } y \in [0, u), \\ (R_a + \kappa_a)y - \kappa_a \text{ , if } y \in [u, 1), \end{cases} \tag{117}$$

$$D \triangleq \max\{k \in \{0, 1, ..., \tilde{D}\} \mid \zeta_k < u\}, \tag{118}$$

$$\zeta_{D+1} = u. \tag{119}$$

The value function $V(\cdot)$ is the desired value function that satisfies (106), (107), (108) and (109).

□

EC.3.5. Supplementary: Proof of Lemma EC.3 and EC.4

Proof of Lemma EC.3. First we show a part of (a): $V_l(\cdot)$ is twice differentiable. The construction process of $V_l(\cdot)$ in Step II (§EC.3.2) ensures that $V_l(\cdot)$ is C^1 over $(l, 1)$. To show V_l is twice differentiable over $(l, 1)$, it suffices to only look at V_l at the z_k's. By definition

$$\Gamma f_{k-1}(z_k) = \frac{c_{i_{k-1}} - c_{i_k}}{\eta_{i_{k-1}} - \eta_{i_k}} \Rightarrow \frac{c_{i_{k-1}} - \lambda f_{k-1}(\zeta_k)}{\eta_{i_{k-1}}} = \frac{c_{i_{k-1}} - c_{i_k}}{\eta_{i_{k-1}} - \eta_{i_k}} \tag{120}$$

$$\Rightarrow \lambda f_{k-1}(\zeta_k) = c_{i_{k-1}} - \eta_{i_{k-1}} \frac{c_{i_{k-1}} - c_{i_k}}{\eta_{i_{k-1}}} \tag{121}$$

$$\Rightarrow \lambda f_k(\zeta_k) = c_{i_{k-1}} - \eta_{i_{k-1}} \frac{c_{i_{k-1}} - c_{i_k}}{\eta_{i_{k-1}}} \tag{122}$$

$$\Rightarrow \Gamma f_k(z_k) = \frac{c_{i_{k-1}} - c_{i_k}}{\eta_{i_{k-1}} - \eta_{i_k}} \tag{123}$$

Thus $f'_{k-1}(z_k) = f'_k(z_k)$ and V_l is twice differentiable.

To prove (b), note that by construction $V_l''(y) \geq 0$ for all $y \in (l, 1)$ (This is because $\Gamma V_l(y) = c_j + \lambda V_l(y)$ for some j, and by construction, $V_l(y) \geq -\frac{c_1}{\lambda} \geq -\frac{c_j}{\lambda}$ for all j). If $0 \notin \{i_k\}_{k \geq 1}$, then $V_l''(\cdot)$ can be equal to 0 at at most one point in $(l, 1)$. Hence it has at most one local minimum and cannot have any local maximum in $(l, 1)$.

We shall next prove (c). If at any $y \in (l,1), m(y) \geq 2$, then by definition of the construction process, $\Gamma V_l(y) \geq \frac{c_i - c_{i-1}}{\eta_i - \eta_{i-1}} > 0$, which implies $V_l''(y) > 0$. If at any $y \in (l,1), m(y) = 1$ and $V_l(y) > -\frac{c_1}{\lambda}$, then $\Gamma V_l(y) = \frac{c_1 + \lambda V_l(y)}{\eta_1} > 0$. Note that (d) is trivially true by the continuity of $\Gamma V_l(\cdot)$.

We shall next prove (e). If $0 \in \{i_k\}_{k \geq 0}$, this means $i_{\tilde{D}} = 0, \zeta_{\tilde{D}} = \inf\{\Gamma V_l(y) \leq 0\}$ and $\Gamma V_l(\zeta_{\tilde{D}}) = 0$, i.e. $V_l''(\zeta_{\tilde{D}}) = 0$ and V_l runs control 1 over $(\zeta_{\tilde{D}-1}, \zeta_{\tilde{D}})$. We claim that $V_l'(\zeta_{\tilde{D}}) \leq 0$. Suppose not, i.e. $V_l'(\zeta_{\tilde{D}}) > 0$, then there exists $\delta > 0$ s.t. $\delta < \zeta_{\tilde{D}} - \zeta_{D-1}$ and $V_l(\zeta_{\tilde{D}} - \delta) < V_l(\zeta_{\tilde{D}})$. Since V_l runs control 1 at $\zeta_{\tilde{D}} - \delta$, this means $\Gamma V_l(\zeta_{\tilde{D}} - \delta) < \Gamma V_l(\zeta_{\tilde{D}}) = 0$, which implies $\zeta_{\tilde{D}} \neq \inf\{\Gamma V_l(y) \leq 0\}$, a contradiction.

Lastly we prove the remaining part of (a). The sequence $\{i_k\}$ is a singleton if and only if $\tilde{D} = 0$, i.e. the sequence $\{\zeta_k\}$ is empty, and is monotone if and only if $\tilde{D} = 1$, i.e. the sequence $\{\zeta_k\}$ consists only of one element. Suppose $\tilde{D} \geq 2$ then we have two possibilities:

i) If $0 \in \{i_k\}_{k \geq 0}$, by (e) $V_l''(y) > 0$ for $y \in (l, \zeta_{\tilde{D}})$ and $V_l'(\zeta_{\tilde{D}}) \leq 0$. This implies that

$$V_l'(y) < 0 \text{ for all } y \in (l, \zeta_{\tilde{D}}) \tag{124}$$

Hence $V_l(\cdot)$ is strictly decreasing over $(l, \zeta_{\tilde{D}})$. By (100), $-c_{i_k} + \eta_{i_k} \Gamma V_l(y) - \lambda V_l(y) = 0$ for $y \in [\zeta_k, \zeta_{k+1}), k \in \{0, 1, ...\tilde{D}\}$, i.e. $\Gamma V_l(y) = \frac{c_{i_k} + \lambda V_l(y)}{\eta_{i_k}}$ for $y \in [\zeta_k, \zeta_{k+1})$. Hence $\Gamma V_l(\cdot)$ follows the monotonicity of $V_l(\cdot)$ in each interval $[\zeta_k, \zeta_{k+1})$, and by the continuity of $\Gamma V_l(\cdot)$, it follows the monotonicity of $V_l(\cdot)$ for $y \in [l,1)$. By (124), $\Gamma V_l(\cdot)$ is monotonically decreasing and hence the sequence $\{i_k\}$ is monotonically decreasing.

ii) If $0 \notin \{i_k\}_{k \geq 0}$, then by the same argument, $\Gamma V_l(\cdot)$ follows the monotonicity of $V_l(\cdot)$ for $y \in [l,1)$. By (b), $V_l(\cdot)$ has at most one local minimum and cannot have any local maximum, and hence this holds for $\Gamma V_l(\cdot)$ over $(l,1)$ as well. Hence, $\{i_k\}$ can only be monotonically decreasing, monotonically increasing or monotonically decreasing then monotonically increasing.

In addition, since $\{i_k\}$ is *unimodally consecutive*, the sequence of switch points is finite. □

Proof of Lemma EC.4. First we prove (a). We note that if $0 \in \{i_k\}_{k \geq 0}$, then by Lemma EC.3 (e)

$$V_l'(\zeta_{\tilde{D}}) \leq 0, \; V_l(\zeta_{\tilde{D}}) = -\frac{c_1}{\lambda}. \tag{125}$$

where $\zeta_{\bar{D}} < 1$. This implies $V_l(y) \le -\frac{c_1}{\lambda}$ for $y \in [\zeta_{\bar{D}}, 1)$ and hence is partially underneath $G(y)$ over $[L_c, 1)$.

By definition, $V_{l_{max}}$ is a line with negative slope, and hence is partially underneath $G(y)$ for some $y \in [L_c, 1)$.

To prove $V_l(y)$ lies above $G(y)$ for l sufficiently small, let $j = \sup\{i \mid \xi_i < R_0\}$. Then there exists $\bar{l} > 0$ sufficiently small such that $G(\bar{l}) > 0$, and we can choose $l > 0, \epsilon > 0$ sufficiently small that together satisfy

$$l + \epsilon < \bar{l},$$

$$m(y) = j + 1, \text{ for all } y \in (l, l + \epsilon),$$

$$V_l''(y) > 0, \text{ for all } y \in (l, \hat{l}).$$

Since $V_l'(l) = -(R_0 + \kappa_0)$, by (126) we have $V_l'(y) > -(R_0 + \kappa_0)$ for $y \in (l, l + \epsilon)$. This implies $V_l(l+\epsilon) > V_l(l) + V_l'(l) \cdot \epsilon = R_0 - (R_0 + \kappa_0)l - (R_0 + \kappa_0)\epsilon = G(\bar{l}) > 0$. In other words, $V_l(y) > G(\bar{l}) > 0$ for $y \in (l, l + \epsilon)$, which implies $\Gamma V_l(y) > \frac{\lambda G(\bar{l}) + c_{j+1}}{\eta_{j+1}} > 0$ for $y \in (l, l + \epsilon)$. As a result, $V_l''(y) > \frac{\delta}{y^2}$ for some δ independent of l for $y \in (l, l + \epsilon)$. Hence

$$V_l'(l+\epsilon) = V_l'(l) + \int_l^{l+\epsilon} V_l''(s)ds \ge -(R_0 + \kappa_0) + \int_l^{l+\epsilon} \frac{\delta}{s^2} ds = -(R_0 + \kappa_0) + 2\delta(\frac{1}{l} - \frac{1}{l+\epsilon}). \quad (126)$$

By taking $l > 0$ sufficiently small and by choosing an appropriate $\epsilon > 0$ (such that $l + \epsilon < L_c$), $V_l'(l + \epsilon) > R_0 + \kappa_0$ and hence by Lemma EC.3 (b), $V_l'(y) > V_l'(l + \epsilon) > R_0 + \kappa_0$ for $y \in (l + \epsilon, 1)$. This implies $V_l(l + \epsilon) > G(l + \epsilon)$ and $V_l(\cdot)$ lies above $G(\cdot)$ over $[L_c, 1)$.

To prove (b), we fix $l \in (0, l_{max})$, and let its associated control set be $\{i_k\}_{k \ge 0}$ with $i_0 \ge 1$ (see Figure 11). Suppose $1 \notin \{i_k\}_{k \ge 0}$, then by Lemma EC.3 (c) we are done. Suppose $1 \in \{i_k\}_{k \ge 0}$, then let ζ_k be the switch point that V_l first switches to control 1 (and let $k = 0$ if $i_0 = 1$). If $\zeta_k \ge l_{max}$ we are also done. If $\zeta_k < l_{max}$, by Lemma EC.3 (c) we must have

$$V_l(\zeta_k) \ge -(R_0 + \kappa_0)\zeta_k + R_0, \quad (127)$$

$$V_l'(\zeta_k) \ge -(R_0 + \kappa_0), \quad (128)$$

$$-c_1 + \eta_1 \Gamma \tilde{V}(y) - \lambda \tilde{V}(y) = 0, \text{ for } y \in [\zeta_k, \zeta_{k+1}). \quad (129)$$

If $l_{max} \in [\zeta_k, \zeta_{k+1}]$, by Lemma EC.6 we are done. If $l_{max} > \zeta_{k+1}$, by Lemma EC.6 we must have $V_l''(\zeta_{k+1}-) > 0$ and hence by Lemma EC.3 (e) we must have $i_{k+1} = 2$. Hence $V_l''(y) > 0$ for $y \in [\zeta_{k+1}, l_{max}]$ and we are done.

The following two supplementary Lemmas will be useful in the proof of Lemma EC.3 and EC.4.

LEMMA EC.5. *Suppose $c_1 + \lambda z_0 > 0$ and let $f_1(\cdot)$ and $f_2(\cdot)$ be defined by the following*

$$\begin{cases} f_1(w_0) = z_0, \\ f_1'(w_0) = z_1, \end{cases} \quad \begin{cases} f_2(w_0) = z_0 + h_0, \\ f_2'(w_0) = z_1 + h_1, \end{cases} \tag{130}$$

$$-c_1 + \eta_1 \Gamma f_i(y) - \lambda f_i(y) = 0, \ i \in \{1, 2\}. \tag{131}$$

for some $h_0 \geq 0, h_1 \geq 0$. Then $f_i''(w_0) > 0, i \in \{1, 2\}$. If either $h_0 > 0$ or $h_1 > 0$, then we have

$$f_2(y) > f_1(y), \ f_2'(y) > f_1'(y), \ f_2''(y) > f_1''(y), \tag{132}$$

for all $y \in (w_0, 1)$.

Proof of Lemma EC.5. Note that since $\Gamma f_i(y) = \frac{c_1 + \lambda z_0}{\eta_1}$, the condition $c_1 + \lambda z_0 > 0$ implies that $f_i''(w_0) > 0, i \in \{1, 2\}$.

If either $h_0 > 0$ or $h_1 > 0$, there exists $y_0 > w_0$ s.t. $f_2(y_0) > f_1(y_0), f_2'(y_0) > f_1'(y_0), f_2''(y_0) > f_1''(y_0)$. For $y > y_0$,

$$f_2(y) > f_1(y) \overset{(131)}{\Leftrightarrow} \Gamma f_2(y) > \Gamma f_1(y) \Leftrightarrow f_2''(y) > f_1''(y)$$

□

LEMMA EC.6. *Let $w_0 \in (0, \frac{c_1 + R_0\lambda}{\lambda(R_0 + \kappa_0)})$ and let $f(\cdot)$ be defined by the following*

$$f(w_0) \geq -(R_0 + \kappa_0)w_0 + R_0, \tag{133}$$

$$f'(w_0) \geq -(R_0 + \kappa_0), \tag{134}$$

$$-c_1 + \eta_1 \Gamma f(y) - \lambda f(y) = 0. \tag{135}$$

Then we have $f''(y) > 0$ for all $y \in [w_0, \frac{c_1 + R_0\lambda}{\lambda(R_0 + \kappa_0)}]$.

Proof of Lemma EC.6. $w_0 < \frac{c_1+R_0\lambda}{\lambda(R_0+\kappa_0)}$ implies $c_1 + \lambda[-(R_0+\kappa_0)w_0 + R_0] > 0$. By Lemma EC.5, $f''(w_0) > 0$. By the continuity of $f''(\cdot)$, there exists $\delta > 0$ s.t.

$$f''(y) > 0 \text{ for } y \in (w_0, w_0 + \delta). \tag{136}$$

Suppose there exists $y_0 \in [w_0, \frac{c_1+R_0\lambda}{\lambda(R_0+\kappa_0)}]$ s.t. $f''(y_0) \leq 0$. Let $t \triangleq \inf\{y \in [w_0, \frac{c_1+R_0\lambda}{\lambda(R_0+\kappa_0)}] \mid f''(y) \leq 0\}$ (and it can be easily verified that $t \leq y_0 \leq \frac{c_1+R_0\lambda}{\lambda(R_0+\kappa_0)}$). Since $f''(y)$ is continuous, we must have $f''(t) = 0$, which by (135) implies that $\frac{c_1+\lambda f(t)}{\eta_1} = 0$, i.e.

$$f(t) = -\frac{c_1}{\lambda}. \tag{137}$$

By the definition of t, for $y \in [w_0, t], f''(y) \geq 0$. Moreover, by (136) we have $f'(y) > -(R_0+\kappa_0)$ for $y \in (w_0+\delta, t)$. Thus,

$$f(t) = f(w_0) + \int_{w_0}^{t} f'(s)ds > -(R_0+\kappa_0)w_0 + R_0 + \int_{w_0}^{t} -(R_0+\kappa_0)ds = R_0 - (R_0+\kappa_0)t \geq -\frac{c_1}{\lambda}, \tag{138}$$

i.e. $f(t) > -\frac{c_1}{\lambda}$ which contradicts equation (137). □

We will now proceed to prove (c), i.e. to show $\{\hat{V}_l, l \in (0, l_{max}]\}$ is ordered. Suppose $0 < l_1 < l_2 \leq l_{max}$, by the proofs in the previous part of this Lemma we know that $V''_{l_1}(y) > 0$ for all $y \in (l_1, l_2]$, which by (85) and (100) implies that

$$V_{l_1}(l_2) > V_{l_2}(l_2) - -(R_0+\kappa_0)l_2 + R_0, \tag{139}$$

$$V'_{l_1}(l_2) > V'_{l_2}(l_2) = -(R_0+\kappa_0). \tag{140}$$

We shall claim and show that for $y \in [l_2, 1)$, $V_{l_1}(y) > V_{l_2}(y)$ implies $V''_{l_1}(y) > V''_{l_2}(y)$. If this is true, by (140) we have $V_{l_1}(y) > V_{l_2}(y)$ for $y \in [l_2, 1)$. Moreover, it is obvious that $V_{l_1}(y) > -(R_0+\kappa_0)y$ for $y \in (l_1, l_{max})$, so the Lemma is proven.

We shall now show out claim above. Take any $y \in [l_2, 1)$ and it suffices to consider the case where neither $m_{l_1}(y) \neq 0$ nor $m_{l_2}(y) \neq 0$ uses control 0 at y. Suppose $V_{l_1}(y) > V_{l_2}(y)$. If $m_{l_1}(y) = m_{l_2}(y)$, then $\Gamma V_{l_1}(y) > \Gamma V_{l_2}(y)$ which implies $V''_{l_1}(y) > V''_{l_2}(y)$. So it only remains to consider the case where

at y, $m_{l_1}(y) = j_1 \neq j_2 = m_{l_2}(y)$. Consider the case where $\Gamma V_l(y) \notin \Psi$ (the case where $\Gamma V_l(y) \in \Psi$ can be trivially shown and is omitted).

Suppose $j_2 \geq 2$. Since $V_{l_1}(y) > V_{l_2}(y)$, by Lemma EC.2 we have

$$j_2 \triangleq \max\{i \mid \Gamma V_{l_2}(y) \geq \frac{c_i - c_{i-1}}{\eta_i - \eta_{i-1}}\} = \max\{i \mid V_{l_2}(y) \geq -\gamma_{i-1}\} \tag{141}$$

$$\leq \max\{i \mid V_{l_1}(y) \geq -\gamma_{i-1}\} = \max\{i \mid \Gamma V_{l_1}(y) \geq \frac{c_i - c_{i-1}}{\eta_i - \eta_{i-1}}\} \triangleq j_1. \tag{142}$$

Suppose $j_2 = 1$, then obviously we still have $j_2 \leq j_1$. Since $j_2 \neq j_1$ we must have $\Gamma V_{l_1}(y) > \Gamma V_{l_2}(y)$, i.e. $V''_{l_1}(y) > V''_{l_2}(y)$. The Lemma is thus proven. □

Appendix EC.4. Proof that (8) and (9) are strictly weaker than Assumption 1 of Araman and Caldentey (2021)

LEMMA EC.7. *Given two sequences of probability distributions* $\{p_0^{(K)}(X \mid j)\}_{K=1}^{\infty}$ *and* $\{p_a^{(K)}(X \mid j)\}_{K=1}^{\infty}$ *(where* $j \in \mathcal{S}$*), Suppose there exists* $p(X \mid j)$, $\alpha_a(X,j)$, $\alpha_0(X,j)$ *such that*

$$\sqrt{K}\left(\frac{p_a^{(K)}(X \mid j)}{p(X \mid j)} - 1\right) \to \alpha_a(X,j), \quad \sqrt{K}\left(\frac{p_0^{(K)}(X \mid j)}{p(X \mid j)} - 1\right) \to \alpha_0(X,j), \tag{143}$$

which represents Assumption 1 of Araman and Caldentey (2021). Then there exists η_j *such that (8) and (9) hold. On the other hand, there exist sequences of probability distributions such that (8) and (9) hold but (143) does NOT hold.*

Proof. First we note that by (143), we have

$$\sqrt{K}\left(\frac{p_a^{(K)}(X \mid j)}{p_0^{(K)}(X \mid j)} - 1\right) = \sqrt{K}\left(\frac{\frac{\alpha_a(X,j)}{\sqrt{K}} + 1}{\frac{\alpha_0(X,j)}{\sqrt{K}} + 1} - 1\right) \tag{144}$$

$$\to \alpha_a(X,j) - \alpha_0(X,j), \tag{145}$$

and hence (8) holds. To prove (9), we will define $\alpha(X,j) \triangleq \alpha_a(X,j) - \alpha_0(X,j)$. By definition,

$$\begin{aligned}\mu_0^{(K)}(j) &= \sum_X \ln\left(\frac{p_a^{(K)}(X \mid j)}{p_0^{(K)}(X \mid j)}\right) p_0^{(K)}(X \mid j) \\ &= \sum_X \left(\frac{p_a^{(K)}(X \mid j)}{p_0^{(K)}(X \mid j)} - 1 + O\left(\frac{p_a^{(K)}(X \mid j)}{p_0^{(K)}(X \mid j)} - 1\right)^2\right) p_0^{(K)}(X \mid j)\end{aligned}$$

$$\mu_a^{(K)}(j) = \sum_X \ln\left(\frac{p_a^{(K)}(X \mid j)}{p_0^{(K)}(X \mid j)}\right) p_a^{(K)}(X \mid j)$$
$$= \sum_X \left(\frac{p_a^{(K)}(X \mid j)}{p_0^{(K)}(X \mid j)} - 1 + O\left(\frac{p_a^{(K)}(X \mid j)}{p_0^{(K)}(X \mid j)} - 1\right)^2\right) p_a^{(K)}(X \mid j)$$

Hence we have,

$$K\left(\mu_a^{(K)}(j) - \mu_0^{(K)}(j)\right) \to \sum_X \left(\sqrt{K}\left(\frac{p_a^{(K)}(X \mid j)}{p_0^{(K)}(X \mid j)} - 1\right)\right) \sqrt{K}\left(p_a^{(K)}(X \mid j) - p_0^{(K)}(X \mid j)\right) \tag{146}$$
$$= \sum_X \left(\sqrt{K}\left(\frac{p_a^{(K)}(X \mid j)}{p_0^{(K)}(X \mid j)} - 1\right)\right)\left(\sqrt{K}\left(\frac{p_a^{(K)}(X \mid j)}{p_0^{(K)}(X \mid j)} - 1\right)\right) p_0^{(K)}(X \mid j) \tag{147}$$

$$\to \mathbb{E}_{p_0^{(K)}(\cdot|j)}[\alpha(X,j)^2] \tag{148}$$

$$\to \mathbb{E}_{p(\cdot|j)}[\alpha(X,j)^2], \tag{149}$$

and hence (9) holds.

Next we will show that there exist sequences of probability distributions such that (8) and (9) hold but (143) does NOT hold. We assume there is only a single control and omit the notation j. Let

$$p_0^{(K)}(\cdot) = p(\cdot) \sim Bernoulli(1/2),\ p_a^{(K)}(\cdot) \sim Bernoulli(1/2 + \frac{1}{\sqrt{K}}(-1)^K). \tag{150}$$

Then in this case, we have

$$\frac{p_a^{(K)}(X=1)}{p_0^{(K)}(X=1)} - 1 = \frac{(1/2) + \frac{1}{\sqrt{K}}(-1)^K}{1/2} - 1 = \frac{2}{\sqrt{K}}(-1)^K, \tag{151}$$

which implies (8) holds and

$$\sqrt{K}[\frac{p_a^{(K)}(X=1)}{p_0^{(K)}(X=1)} - 1] = \sqrt{K}[\frac{p_a^{(K)}(X=1)}{p(X=1)} - 1] = 2(-1)^K, \tag{152}$$

which does NOT converge as $K \to \infty$ (i.e.(143) does NOT hold). Moreover,

$$\lim_{K\to\infty} K[\frac{p_a^{(K)}(X=1)}{p_0^{(K)}(X=1)} - 1]^2 = \lim_{K\to\infty} K[\frac{2}{\sqrt{K}}(-1)^K]^2 = 4(-1)^{2K} = 4, \tag{153}$$

i.e. $\alpha^2(X=1)=4$ is well-defined, and

$$\lim_{K\to\infty} K[\frac{p_a^{(K)}(X=0)}{p_0^{(K)}(X=0)}-1]^2=\lim_{K\to\infty} K[-\frac{2}{\sqrt{K}}(-1)^K]^2=4(-1)^2(-1)^{2K}=4, \tag{154}$$

i.e. $\alpha^2(X=0)=4$ is also well-defined. By (148),

$$K\left(\mu_a^{(K)}(j)-\mu_0^{(K)}(j)\right)\to\mathbb{E}_{p_0^{(K)}}[\alpha(X,j)^2]=\frac{1}{2}\cdot 4+\frac{1}{2}\cdot 4=4, \tag{155}$$

i.e. (9) holds. □

Appendix EC.5. Comparison of OIP's Efficient Frontier with HS's

The *efficient frontier* in our model, introduced in §5, is different from that of HS's, where they analyze a model with no reward or penalty when $\theta=0$. In this section we will compare the *efficient frontiers* from both models. Using the same notation as before, let the set of controls be $\{(\eta_i,c_i)\}_{i=1}^N$ and let the controls be numbered as in (24), (25), (26) and (27). HS's *efficient frontier* is the function $\phi_{HS}(\cdot)$ given by

$$0<\eta_n<...<\eta_{\tilde{N}}, \tag{156}$$

$$\frac{c_n}{\eta_n}<\frac{c_{n+1}-c_n}{\eta_{n+1}-\eta_n}<...<\frac{c_{\tilde{N}}-c_{\tilde{N}-1}}{\eta_{\tilde{N}}-\eta_{\tilde{N}-1}}, \tag{157}$$

$$c_i=\phi_{HS}(\eta_i),\text{ for } i=n,n+1,...,\tilde{N}, \tag{158}$$

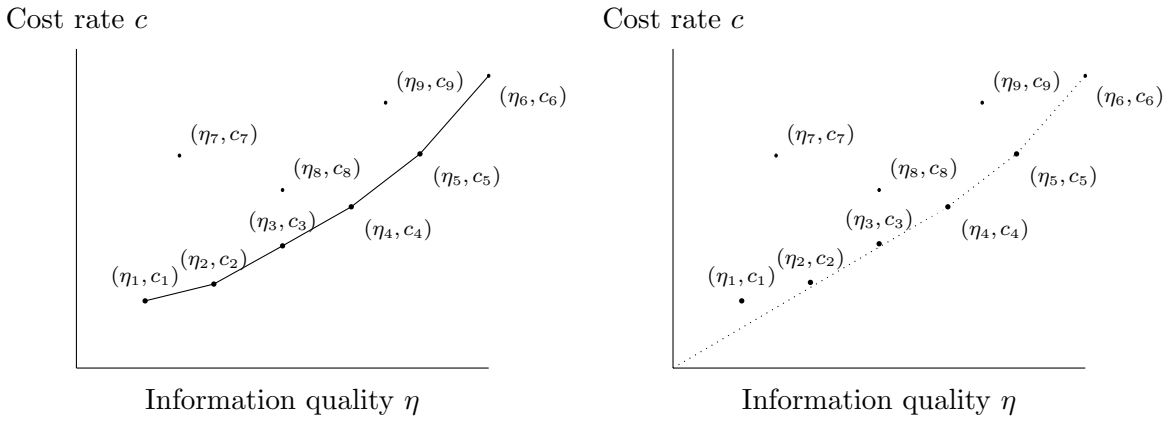
$$c_i\geq\phi(\eta_i),\text{ for } i=1,2,...,n-1,\tilde{N}+1,...,N, \tag{159}$$

where $\phi_{HS}(\cdot)$ is the strictly increasing, piece-wise linear and convex function that connects $(0,0),(\eta_n,c_n),(\eta_{n+1},c_{n+1})$, ..., $(\eta_{\tilde{N}},c_{\tilde{N}})$.

Figure 14 shows the two main differences between $\phi(\cdot)$ and $\phi_{HS}(\cdot)$. First, $\phi_{HS}(\cdot)$ passes through $(0,0)$ but $\phi(\cdot)$ does not. Second, points on $\phi_{HS}(\cdot)$ must be on $\phi(\cdot)$, but not vice versa. In other words, the points on HS's *efficient frontier* are a subset of the points on ours.

In HS's model structure where there is no reward or penalty when $\theta=0$, the optimal policy *only* uses the controls on $\phi_{HS}(\cdot)$, in increasing order of their index as the posterior belief increases. In our model with a more general reward structure (where when $\theta=0$ there is either a reward or a penalty, or both), the optimal policy shifts to one which may also incorporate controls on $\phi(\cdot)$, and the optimal control sequence of the controls may no longer be increasing but becomes *unimodally consecutive* (see Theorem 2).

Figure 14 The difference between the *efficient frontier* in our model and that in HS's model.



(a) The *efficient frontier* in our model.

(b) The *efficient frontier* in HS's model.

Appendix EC.6. Derivation of Patient Recruiting Cost and Site Retaining Cost from Past Literature

Patient Recruiting Cost. The values for μ_P^n are derived from the estimates in Table 2 (Chantix/Champix) of Kouvelis et al. (2017). The weekly patient enrollment rate of 58, 93, 140 is equivalent to to an annual enrollment rate (a year has 52 weeks) of 3016, 4836, 7280, and has a per patient cost of $0.01M$, $0.02M$, $0.03M$ respectively. Therefore based on the above numbers, we assume that when the number of patients recruited n_P in a stage is 2500, 5000, 7500, 10000, the per patient cost is to $0.01M$, $0.02M$, $0.03M$, $0.04M$ respectively. Otherwise, we approximate the value of the per patient cost using linear interpolation of the nearest two cost rates (and assuming enrolling 0 patients costs 0).

Site Cost. From table 2 of Sertkaya et al. (2016), the total retaining cost for a site for a typical phase III clinical trial is $(0.395 + 1.31 + 2.32 + 1.62)/4 = 5.6M$. Again from Sertkaya et al. (2016), the average duration of a typical phase III trial is about 2.5 years. Therefore, the retaining cost per site per year is $5.6/2.5 = 2.25M$.

Appendix EC.7. Mapping of Difference in Patient Outcomes between the Treatment Group and Control Group to Log-likelihood Ratio

Assume the decision maker is testing between $H_0 : \mu_Y - \mu_X = 0$ and $H_a : \mu_Y - \mu_X = \Delta$, where μ_Y and μ_X represent the patient outcome in the treatment group and control group, and Δ represents the treatment effect to be detected.

Let Y_i and X_i denote the realized outcome for one patient in the treatment and control group, and assume $Y_i \sim N(\mu_Y, s^2)$, $X_i \sim N(\mu_X, s^2)$. Assume an experiment consists of recruiting K patients, with half allocated in treatment group and half in control group. The difference in outcome is $\hat{Z} = \sum_{i=1}^{K/2} Y_i - \sum_{i=1}^{K/2} X_i$. Note that

$$\hat{Z} \sim N\left(\frac{K}{2}(\mu_Y - \mu_X), 2(\frac{K}{2})s^2\right), \tag{160}$$

and the log-likelihood ratio L can be computed by

$$ln\frac{p_a(\hat{Z})}{p_0(\hat{Z})} = ln\frac{\frac{1}{\sqrt{K}s\sqrt{2\pi}}e^{\frac{-(\hat{Z}-\frac{K}{2}\Delta)^2}{Ks^2}}}{\frac{1}{\sqrt{K}s\sqrt{2\pi}}e^{\frac{-\hat{Z}^2}{Ks^2}}} \tag{161}$$

$$= ln\frac{e^{\frac{-(\hat{Z}-\frac{K}{2}\Delta)^2}{Ks^2}}}{e^{\frac{-\hat{Z}^2}{Ks^2}}} \tag{162}$$

$$= ln\left(e^{\frac{-(\hat{Z}-\frac{K}{2}\Delta)^2+\hat{Z}^2}{Ks^2}}\right) \tag{163}$$

$$= \frac{\frac{K}{2}\Delta(2\hat{Z} - \frac{K}{2}\Delta)}{Ks^2} \tag{164}$$

i.e. keeping track of $\hat{Z}$ and L is equivalent.