Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Bayesian Sequential Learning for Clinical Trials of Multiple Correlated Medical Interventions

Stephen E. Chick, Noah Gans, Özge Yapar

To cite this article:
Stephen E. Chick, Noah Gans, Özge Yapar (2021) Bayesian Sequential Learning for Clinical Trials of Multiple Correlated Medical Interventions. Management Science

Published online in Articles in Advance 05 Nov 2021

. https://doi.org/10.1287/mnsc.2021.4137

informs
http://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–20
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Bayesian Sequential Learning for Clinical Trials of Multiple Correlated Medical Interventions

Stephen E. Chick,[a] **Noah Gans,**[b] **Özge Yapar**[c]

[a] Technology & Operations Management Area, INSEAD, 77305 Fontainebleau, France; [b] Operations, Information and Decisions Department, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104; [c] Operations & Decision Technologies Department, Kelley School of Business, Indiana University, Bloomington, Indiana 47405

Contact: stephen.chick@insead.edu, https://orcid.org/0000-0002-8026-1571 (SEC); gans@wharton.upenn.edu, https://orcid.org/0000-0002-5839-7122 (NG); oyapar@iu.edu, https://orcid.org/0000-0002-4208-9221 (OY)

Received: May 31, 2018
Revised: August 31, 2020; February 8, 2021
Accepted: May 26, 2021
Published Online in Articles in Advance:
November 5, 2021

https://doi.org/10.1287/mnsc.2021.4137

Abstract. We propose and analyze the first model for clinical trial design that integrates each of three important trends intending to improve the effectiveness of clinical trials that inform health-technology adoption decisions: adaptive design, which dynamically adjusts the sample size and allocation of interventions to different patients; multiarm trial design, which compares multiple interventions simultaneously; and value-based design, which focuses on cost-benefit improvements of health interventions over a current standard of care. Example applications are to seamless phase II/III dose-finding trials and to trials that test multiple combinations of therapies. Our objective is to maximize the expected population health-economic benefit of health-technology adoption decisions less clinical trial costs. We show that unifying the adaptive, multiarm, and value-based approaches to trial design can reduce the cost and duration of multiarm trials with efficient adaptive look ahead policies that focus on value to patients and account for correlated rewards across arms. Features that differentiate our approach from much other work on stochastic optimization include stopping times that balance sampling costs and the expected value of information of those samples, performance guarantees offered by new asymptotic convergence proofs, and the modeling of arms' potentially different sampling costs. Our proposed solution can be computed feasibly and can randomize patients. The class of trials for the base model assumes that health-economic data are collected and observed quickly. Related work from Bayesian optimization can enable the further inclusion of trials with intermediate duration delays between the time of treatment initiation and observation of outcomes.

History: Accepted by Stefan Scholtes, healthcare management.

Funding: This work was supported by the Wharton Risk Management Center's Russell Ackoff Doctoral Student Fellowship and the Wharton–INSEAD Alliance. This research was supported in part by Lilly Endowment, Inc., through its support for the Indiana University Pervasive Technology Institute. S. E. Chick acknowledges the support of the European Union Marie Curie Initial Training Network [Marie Skłodowska-Curie actions - European Sepsis Academy - Innovative Training Network Project 676129].

Supplemental Material: The online appendix is available at https://doi.org/10.1287/mnsc.2021.4137.

Keywords: Bayesian • clinical trials • health economics • sequential experimentation • ranking and selection • simulation optimization • dynamic programming

1. Introduction

The UK National Institutes of Health Research (NIHR) (2020) states that it is "keen to see the design, development and delivery of more efficient, faster, innovative studies to provide robust evidence to inform clinical practice and policy." Other regulators, researchers, and funders of trials also call for more efficient clinical trial designs with a goal of expanding and better allocating the financial and clinical resources available to assess health technologies (EU 2014, FDA 2016, Hudson et al. 2016, EMA 2017).

Concerns regarding ineffective use of resources have many facets and range from the time it takes to assess new health technologies to the mismatch between the statistical criteria used to evaluate the medical efficacy of treatments and the cost-effectiveness measures that drive adoption decisions and clinical guidance (NICE 2014) to the challenge of recruiting patients to multiple, related two-arm trials, each including a separate control arm, among others. These concerns are relevant for high-stakes drug trials that cost more than a billion dollars (DiMasi et al. 2016), as well as for public sector-funded research that assesses nonpharmaceutical treatment options and that may cost a few millions (NIHR 2018, Forster et al. 2021).

A variety of innovative approaches to trial design have been proposed to address these challenges, including adaptive trial designs, multiarm trials, and value-based trials. This paper proposes what appears to be a first framework to bring together all three of these trends and applies it to an interesting class of trials: those for which the delay between the time a treatment is administered and the time an outcome is observed is short to moderate, treatment costs and health-economic metrics are estimable during the trial, and public health or related cost-benefit criteria are of interest (NIHR 2020). We consider these three approaches in turn and then discuss our model.

Adaptive clinical trials use information accumulated during a trial to modify the experimental design as the trial progresses. Here, the adoption of multiple stages of sampling allows the designer to modify the sample size of the trial, to adapt the allocation of patients among arms, and/or to drop inferior arms based on the outcomes observed so far, with the goal of improving a trial's cost, duration, or information gain. The advantages of adaptive trials are discussed widely (Berry 2012, Chow 2014, Ahuja and Birge 2016, Ellenberg and Ellenberg 2017, Pallmann et al. 2018).

Multiarm trials allow for the comparison of multiple interventions, rather than a single alternative, with a common placebo or control, such as the current standard of care, and can reduce the number of patients required for a trial. For example, the European Medicines Agency (EMA) and the U.S. Food and Drug Administration (FDA) have advocated the use of multiarm trials for pediatric rare diseases, for which patient recruitment is difficult (EMA 2017).

Multiarm, multistage (MAMS) trials use the first two approaches together (Sydes et al. 2012, Wason and Jaki 2012, Jaki and Hampson 2016, Boeree et al. 2017). Arms can represent individual treatments or combinations of options (Cai et al. 2013), with or without controls (Magaret et al. 2016), as well as the dose levels considered in phase II/III dose-finding trials (Huang et al. 2015). Platform trials allow arms to dynamically enter and leave trials (Adaptive Platform Trials Coalition et al. 2019). New information systems are being developed to support data needs of adaptive trials (Lock 2019).

Value-based trials respond to calls, such as that from the National Health Service (NHS) England (2017) to, "examine how best to 'bake in' an assessment of value and real world cost as an integral and default part of future NHS research studies rather than see this a separate 'optional extra.'" Value-based trials focus on incorporating health benefits and costs into the design of the trials themselves using the expected value of sample information (EVSI). EVSI has been used in health economics, for example, to assess the expected benefit of collecting additional data to improve a health-technology adoption decision or to further prioritize research for health-resource allocation decisions (Claxton and Posnett 1996, Fenwick et al. 2020). Such benefits are typically based on population-level health benefits relative to treatment costs, often using quality-adjusted life-year (QALY) criteria rather than traditional hypothesis tests of the significance of differences in average clinical effectiveness. Value-based trials can be used to rank the expected health-economic merit of research proposals (Berry and Ho 1988, Lewis et al. 2007), and EVSI techniques have been developed to drive the design of fully adaptive, value-based, two-arm trials (Pertile et al. 2014, Chick et al. 2017, Forster et al. 2021).

Value-based trials use real-valued outcomes to model the cost-QALY assessments used in health-technology adoption decisions (Chick et al. 2017). Although Flight et al. (2019) report that many adaptive trials collect these data, they find that few actually use such value-based criteria in trials design. Nevertheless, they identify researchers who suggest designing trials with value-based criteria in mind and join others in recommending the use of cost-effectiveness in clinical trial design (Nixon et al. 2009, Meltzer and Smith 2011, Draper 2013, NIHR 2020). Williamson and Villar (2020) further argue for the ability to incorporate real-valued, rather than Bernoulli, outcomes in the MAMS setting and note the large fraction of trials with continuous outcomes.

The operations research community has worked to improve clinical trials with various approaches related to the value-based approach here (e.g., Kouvelis et al. 2017, Villar and Rosenberger 2018, Alban et al. 2020, Bastani and Bayati 2020, Anderer et al. 2021, Bravo et al. 2021).

Although there is much active research regarding adaptive, multiarm, and value-based trials, current work typically studies only one or two of the three approaches at a time. This paper brings all three together, first for a more restricted class of trials (a) with short to moderate delay between the time a treatment is administered to a patient and the time outcomes are observed, (b) that include QALY and treatment costs as end points, and (c) that takes the cost-benefit perspective of a social planner (NIHR 2020). We then follow with extensions that address a wider range of settings, as well as practical issues that arise in trials.

1.1. Overview

Section 2 presents the base model: a fully sequential, value-based trial with multiple, potentially correlated arms. Fully sequential means that that the trial design permits decisions regarding which treatment to allocate and whether to stop the trial after each subject's data are observed. Correlated arms allow, for example, for similar doses to have similar average treatment effects. Its objective is to maximize the expected health benefit, less clinical trial and patient treatment costs. Our design balances the marginal cost of enrolling additional patients in the trial with the expected value of the information to be gained from that patient's outcome.

Section 3 characterizes the optimal trial design for either discounted or undiscounted rewards. Computing the optimal solution suffers from a curse of dimensionality, however.

Section 4 introduces a pair of heuristic policies that do not suffer from the curse of dimensionality, one for sequentially allocating treatments to patients who enroll in the trial and another for choosing when to stop the trial as patient data are observed. Both heuristics make decisions on the basis of indices that we calculate using the EVSI of dynamic, forward-looking, potentially adaptive sampling plans, an approach that has proven useful in other settings, with arms that are not correlated (Chick and Gans 2009, Chick and Frazier 2012, Smith and Villar 2018). The new policies also extend the fixed look ahead approach of the so-called correlated knowledge gradient (cKG), which has proven useful in Bayesian optimization (Frazier et al. 2009), to adaptive look ahead policies.

For the case of undiscounted rewards, Section 4 also provides theoretical results regarding the asymptotic consistency of the allocation heuristic as the sample size grows without bound. The proof provides a novel variation of that in Xie et al. (2016) (hereafter referred to as XFC), for the asymptotic consistency of certain fixed step look ahead indices, that applies to adaptive look ahead indices.

Although the heuristic policies developed in Section 4 eliminate the curse of dimensionality, their indices remain computationally intensive to evaluate, and Section 5 describes two additional sets of policies that require less computation and serve to benchmark their performance. A first set of comparators follows from easily calculated lower and upper bounds on the heuristics' indices, bounds that can be used to more efficiently implement the core heuristics and be used, in their own right, as the indices of more quickly computed policies. A second set of comparators follows other policies found in the literature. The most natural of these is the cKG policy of Frazier et al. (2009), which selects the arm by maximizing the EVSI of a related fixed look ahead-based index. Section 5 also discusses additional comparators and implementation issues.

Section 6 discusses the specification of a prior distribution for the vector of the arms ′ mean rewards. It proposes a practical way to use pilot study data to specify a prior for phase II/III dose-finding trials.

Section 7 reports the results of simulations that assess the performance of our heuristics and their comparators in multiarm settings. The first set of experiments shows that, for fixed sample sizes, our new allocation indices outperform those that assume independent arms or that focus on estimation, rather than optimizing reward, and they perform similarly to cKG-type allocation policies that account for correlation. The second set of experiments shows that it is important for stopping times to account for the potential of further sampling from multiple arms when setting a sample size and that response-adaptive stopping holds promise. The third set of experiments illustrates how to use pilot study data to develop a prior distribution for a phase II/III dose-finding trial that is robust to a potential misspecification of the prior in a representative study.

The second and third sets of experiments also assess a simple, forward simulation-based method of selecting a fixed stopping time that performs surprisingly well when used in conjunction with adaptive allocation policies. Although the systematic analysis and optimization of this scheme as its own dynamic stopping time are beyond the scope of this paper, the results suggest that this approach merits additional research.

The paper's core analysis is applicable to trials with short to moderate delay between the time a treatment is administered to a patient and the time outcomes are observed, that include QALY and treatment costs as end points, and that take the cost-benefit perspective of a social planner. Section 8 extends the paper's approach to a broader set of settings and addresses a number of important practical considerations. It provides approaches for addressing the delays noted and the need to randomize the allocation of subjects to treatment arms and offers a discussion about the breadth of application of real-valued outcomes and other topics.

Online Appendix A summarizes notation. Online Appendix B gives proofs of mathematical claims. Online Appendix C presents implementation details. Online Appendix D discusses additional issues about clinical trials.

2. Model for Fully Sequential Trials with Multiple Correlated Arms

We present a Bayesian, decision-theoretic model of an adaptive clinical trial that compares multiple interventions. We seek a sequential sampling policy that dynamically decides the interventions to which patients should be allocated, as well as the time at which the trial should stop, in order to maximize the expected monetary value of health benefits generated for the target population less the cost of the trial and any costs incurred in health-technology adoption. We use correlation among beliefs regarding the population-mean rewards of the interventions to capture potential similarities among the alternatives.

2.1. Trial Design and Outcomes

We consider a clinical trial that evaluates $M \geq 2$ alternative interventions, which we refer to as *arms*. For a controlled trial, the standard of care intervention and/or a placebo can be included among the M arms. The arm selected for implementation at the end of the trial will be used to treat P patients. The value of P may represent many years' worth of patients treated over the arm's useful life, and we assume P is fixed, as is the case for fixed horizon market exclusivity and many nonpharmaceutical health-technology adoption decisions, although the model is amenable to more general P. (See Online Appendix D.2.) We let $\mathcal{M} = \{1, 2, \ldots, M\}$ denote the set of arms in the trial.

Before selecting an arm to implement, we can sample from it to obtain information regarding trial subjects' outcomes. Sampling from an arm requires money and time. There is a cost c_i per observation for arm $i \in \mathcal{M}$. We model sampling as occurring sequentially at equally spaced times, $t = 0, 1, \ldots$. The index t also represents the total number of patient observations seen. We let T denote the (potentially random) time at which we stop the trial and select an arm to implement.

At time $t = 0, 1, \ldots, T-1$, we decide which arm patient t + 1 will receive and can use the data observed from the first t patients in doing so. We denote by E_i^t the random variable whose realization is the effectiveness of arm $i \in \mathcal{M}$ observed for patient $t = 1, 2, \ldots, T$. We assume that the clinical effectiveness of each arm can be expressed in monetary terms so that E_i^t describes the effect that arm i has on the clinical condition of the tth patient converted to a financial value, for example with QALY data and willingness to pay parameters (e.g., 20,000£/QALY) (NICE 2014). We let the random variable C_i^t denote the patient-level cost of arm i, which may include medical procedures, drugs, and the effect of potential complications, again converted to a financial value. Then, the net monetary benefit (NMB) of arm i for patient t is

$$Y_i^t = E_i^t - \mathbb{1}_{CE} C_i^t, \tag{1}$$

where $\mathbb{1}_{CE} = 1$ if the trial evaluates cost-effectiveness and $\mathbb{1}_{CE} = 0$ if it evaluates only effectiveness (see also Section 8.3). Here, c_i is the marginal cost for an extra patient in the trial, whereas C_i is the marginal cost of treatment and complications whether the patient is in the trial.

2.2. Bayesian Prior and Inference

Each Y_i^t has an unknown mean θ_i and a known sampling variance λ_i. Here, θ_i can be interpreted as the population mean of the treatment effect for arm i, and λ_i captures random differences in how individual patients respond to arm i. We assume that observations are independent and normally distributed, conditional on the unknown mean, so that $Y_i^t \mid \theta_i \sim \mathcal{N}(\theta_i, \lambda_i)$ for $t = 1, 2, \ldots$ and $i \in \mathcal{M}$. Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_M)^{\mathsf{T}}$ be the vector of unknown means, and let $\boldsymbol{\Lambda} = diag(\lambda_1, \ldots, \lambda_M)$ be the positive definite, diagonal matrix of sampling variances. We will derive results assuming the λ_i values are known.

A prior distribution or *prior* for $\boldsymbol{\theta}$ describes our initial uncertainty about the M arms' mean effectiveness. This initial belief about $\boldsymbol{\theta}$ is distributed according to a multivariate normal prior, with $\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}^0, \boldsymbol{\Sigma}^0)$. We assume that $\boldsymbol{\Sigma}^0$ is positive definite. It can be nondiagonal, which would imply that initial beliefs about the means are correlated. We discuss the inference process in Section 4.1, the specification of a prior and plug-in estimators for $\boldsymbol{\Lambda}$ in Section 6, and their use in Section 7.4.

2.3. Decisions and Timeline: Allocation, Stopping, and Selection

At each time t, we observe the outcome of the tth patient's intervention, and we use this observation to update our beliefs about the mean arm effects. Then, we choose either to stop the trial or to continue and include one more patient. A decision to stop is followed by the selection of an arm for implementation, and a continuation decision requires the choice of an arm to allocate to the next patient. For now, we assume the time to observe the results from a patient to be short enough that it does not delay the decision for the next patient. We discuss delayed observations in Section 8.1.

To track our choices, we define a number of variables. At each time t, we choose an action u^t from the set of available actions $\mathcal{U} = \{1, 2, \ldots, 2M\}$, with M actions for continuation and M actions for stopping. We let $u^t \in \{i \mid i = 1, 2, \ldots, M\}$ denote the allocation of arm u^t to the t + 1st patient, and we let $u^t \in \{i \mid i = M+1, M+2, \ldots, 2M\}$ denote the stopping of the trial and selection of arm $\mathcal{D} = u^t - M$ for implementation. After an action $u^t \in \{i \mid i = 1, 2, \ldots, M\}$ is chosen, the observation $Y_{u^t}^{t+1}$ is realized before the next period's decision. We then compute the belief about unknown means, $\boldsymbol{\theta}$, at time t + 1, $(\boldsymbol{\mu}^{t+1}, \boldsymbol{\Sigma}^{t+1})$, using the belief at time t, $(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t)$, the observation $Y_{u^t}^{t+1}$, and Bayes' rule, as Section 4.1 discusses. At the first occurrence of $u^t \in \{i \mid i = M+1, M+2, \ldots, 2M\}$, the trial stops, so that the stopping time is $T = t$.

Formally, we define $\mathcal{H}_t = \{\boldsymbol{\mu}^0, \boldsymbol{\Sigma}^0, u^0, Y^1_{u^0}, u^1, Y^2_{u^1}, \ldots, u^{t-1}, Y^t_{u^{t-1}}\}$ to be the *history* of a trial up to time t. In turn, we define a nonanticipating *policy* $\pi : \mathcal{H}_t \to \mathcal{M}$ to be a function that maps the history at each time $t = 0, 1, \ldots$ to an action, u^t, that either continues the trial at time t and tests arm u^t or stops the trial at time $T = t$ and implements arm $\mathcal{D} = u^t - M$. We define Π to be the set of all such nonanticipating policies.

Informally, a policy has three features. A *stopping time* specifies if sampling is to stop at a given time ($T = t$) or not ($T > t$). An *allocation policy* specifies which arm to assign if sampling continues. A *selection decision* specifies which arm to select for implementation when sampling stops.

2.4. The Optimal Multiarm Fully Sequential Value-Based Trial Design Problem

We seek to maximize the expected NMB to the population of P patients to be treated by arm $\mathcal{D}$ upon stopping the trial, net of the costs of sampling and any potential fixed costs of implementing arm $\mathcal{D}$, given the results of the trial. To describe this objective, we need some additional notation. Let $\Delta \in (0,1]$ be a discount factor. Let I_i denote the expected value of a one-time fixed cost associated with implementing arm $i \in \mathcal{M}$ at the end of a trial. It is the sum of all investment costs required to implement the chosen arm, such as capital, training, and infrastructure costs for the healthcare system. Let $\mathbb{E}_\pi$ be the expectation induced by π.

Given the prior distribution $\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}^0, \boldsymbol{\Sigma}^0)$ and a policy $\pi \in \Pi$, this objective function is

$$V^\pi(\boldsymbol{\mu}^0,\boldsymbol{\Sigma}^0)=\mathbb{E}_\pi\left[\sum_{t=0}^{T-1}-\Delta^t c_{u^t}+\Delta^T\Big(P\mathbb{E}\Big[Y^{T+1}_{\mathcal{D}}\Big|\boldsymbol{\mu}^T,\boldsymbol{\Sigma}^T\Big]-I_{\mathcal{D}}\Big)\;\Big|\;\boldsymbol{\mu}^0,\boldsymbol{\Sigma}^0\right], \tag{2}$$

where the random stopping time $T \geq 0$ equals zero if it is optimal to stop immediately, rather than sampling, in which case the sum of the discounted sampling costs is defined to be zero.

We focus on the problem of choosing a policy $\pi^* \in \Pi$ that maximizes (2), the expected discounted value when $\Delta < 1$, or the expected net reward when $\Delta = 1$. To ensure that we will not sample costlessly over an infinite horizon, we require that $\Delta < 1$, that all $c_i > 0$, or both. We call this problem the *optimal multiarm fully sequential value-based trial design problem*:

$$V^*(\boldsymbol{\mu}^0, \boldsymbol{\Sigma}^0) = \sup_{\pi\in\Pi} V^\pi(\boldsymbol{\mu}^0, \boldsymbol{\Sigma}^0). \tag{3}$$

This problem is a multiarmed stoppable bandit with correlated mean rewards. We write $V^{\pi^*}(\boldsymbol{\mu}^0, \boldsymbol{\Sigma}^0)$ as $V^*(\boldsymbol{\mu}^0, \boldsymbol{\Sigma}^0)$ to simplify notation. We may also write $V^*(\boldsymbol{\mu}^0, \boldsymbol{\Sigma}^0; \Xi)$ to emphasize the role of parameters Ξ (say, c_i, P, Δ, or the set Π for the supremum) in determining V^*.

This model can help assess the maximum amount one should pay for a fully adaptive trial rather than for a trial in the set Π_{fix} of traditional designs with a fixed sample size ($T = T_{\text{fix}}$). The variable costs of observations, $\mathbf{c} = (c_1, c_2, \ldots, c_M)$, for an adaptive trial may exceed those of a fixed sample size trial, $\mathbf{c}_{\text{fix}}$. The maximum that one should pay for a fully adaptive trial the cost of such a traditional fixed sample size trial is $V^*(\boldsymbol{\mu}^0, \boldsymbol{\Sigma}^0; \mathbf{c}, \Pi) - V^*(\boldsymbol{\mu}^0, \boldsymbol{\Sigma}^0; \mathbf{c}_{\text{fix}}, \Pi_{\text{fix}})$.

3. Theoretically Optimal Multiarm Fully Sequential Trial

Proposition 1 uses Bellman's equation to characterize the optimal allocation policy, stopping time, and selection decision for the multiarm, fully sequential, value-based trial design problem in (3).

Proposition 1. *If $c_i > 0$ for all $i \in \mathcal{M}$, $\Delta < 1$, or both, then there exists a Markov policy $\pi^* \in \Pi$ that is optimal, and*

$$V^*(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t)=\max\left\{\max_{j\in\mathcal{M}}-c_j+\Delta\mathbb{E}\Big[V^*(\boldsymbol{\mu}^{t+1},\boldsymbol{\Sigma}^{t+1})\,|\,\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t;u^t=j\Big],\right.$$
$$\left.\max_{j\in\mathcal{M}}\Big\{P\mu^t_j-I_j\Big\}\right\}. \tag{4}$$

Bellman's Equation (4) implies that, at each time t, we compare the expected value of $2M$ potential actions. Each of the first M potential actions, in the left maximization within the curly braces in (4), represents the expected value of sampling once for a given arm and then implementing the optimal policy from $t + 1$ forward. The second M potential actions, in the right maximization within (4), represent the expected values of immediately stopping and implementing a given arm.

Proposition 1 shows that there exist an optimal allocation policy, stopping time, and selection decision that satisfy Bellman's equation. Bellman's equation, in turn, motivates the use of indices to guide allocation and stopping decisions. When one of the right maximands, $\max_{j\in\mathcal{M}}\{P\mu^t_j - I_j\}$, maximizes (4), it is optimal to stop, and the *optimal selection decision* picks the maximizer at random if there is more than one such arm. We can therefore rewrite the problem in (3) as

$$V^*(\boldsymbol{\mu}^0,\boldsymbol{\Sigma}^0)=\sup_{\pi\in\Pi}\mathbb{E}_\pi\left[\sum_{t=0}^{T-1}-\Delta^t c_{u^t}+\Delta^T\max_{j\in\mathcal{M}}\Big\{P\mu^T_j-I_j\Big\}\;\Big|\;\boldsymbol{\mu}^0,\boldsymbol{\Sigma}^0\right]. \tag{5}$$

In contrast, if at least one of the left maximands of (4) exceeds the right maximand, then it is optimal to continue and to allocate the next observation to the maximizing arm, at random if there is more than one such arm.

We can also define allocation and stopping indices that reflect the *expected value of information* (EVI). At any time $t = 0, 1, 2, \ldots,$ we can define an EVI-based *stopping index* to be

$$\mathrm{EVI}^{*}(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t) = \sup_{\pi \in \Pi} \mathbb{E}_{\pi}\left[\sum_{r=0}^{T-1} -\Delta^r c_{u^{t+r}} + \Delta^T \max_{j \in \mathcal{M}} \left\{P\mu_j^{t+T} - I_j\right\} \;\middle|\; \boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t\right] - \max_{j \in \mathcal{M}} \left\{P\mu_j^t - I_j\right\}, \tag{6}$$

the expected incremental value obtained from additional sampling rather than stopping immediately and selecting the arm with the greatest expected reward. We use EVI rather than EVSI as we account for the cost of sampling, whereas EVSI traditionally does not, and we allow for a potentially response-adaptive number of further observations rather than a fixed step look ahead.

Here, $\mathrm{EVI}^{*}(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t) \geq 0$ by construction, and $\mathrm{EVI}^{*}(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t) > 0$ if and only if it is optimal to continue sampling rather than stopping immediately. Thus, the associated *optimal stopping time* falls before or at time t if and only if $\mathrm{EVI}^{*}(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t) = 0$.

The EVI of allocating *at least* one observation to arm $i \in \mathcal{M}$ and then proceeding optimally is straightforward to define using the notation of the Bellman's equation:

$$-c_i + \Delta \mathbb{E}[V^{*}(\boldsymbol{\mu}^{t+1}, \boldsymbol{\Sigma}^{t+1}) \mid \boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t; u^t = i] - \max_{j \in \mathcal{M}} \left\{P\mu_j^t - I_j\right\}. \tag{7}$$

The allocation index analogue to (6) is more delicate to construct. To ensure that the first decision allocates an observation to i, it requires that $T \geq 1$ and that $u^t = i$.

The associated *optimal allocation policy* allocates the next sample to $i \in \mathcal{M}$ that maximizes the index in (7), with ties broken at random. We note that the term in (7) can be negative, when it is strictly preferable to stop rather than sample from i.

4. Heuristic for Approximating Optimal Sequential Sampling

Numerical evaluation of the optimal indices (6) and (7) is challenging. Although $\max_{j \in \mathcal{M}}\{P\mu_j^t - I_j\}$ is straightforward to calculate, the computation of the expected value of allocating one patient to arm $i \in \mathcal{M}$ and continuing optimally, $-c_i + \Delta \mathbb{E}[V^{*}(\boldsymbol{\mu}^{t+1}, \boldsymbol{\Sigma}^{t+1}) \mid \boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t; u^t = i]$, suffers from a curse of dimensionality; there are M^T possible sequences for sampling M arms over T time steps.

To address this challenge, we replace the original, optimal indices with heuristic ones. Section 4.1 recalls the inference process for sampling from the ith arm that our analysis requires. Section 4.2 describes our main heuristic index and how to compute it. This computation involves solving a so-called free boundary problem for a heat equation, a type of partial differential equation. Thus, we will refer to it as the correlation partial differential equation (cPDE) index, where "c" stands for correlation. Section 4.3 provides asymptotic consistency results for cPDE.

4.1. Updating Equations for Sampling from a Single Arm

Measurements from arm i impact the posterior distribution of all arms' means. Consider the case of allocating τ patients to arm i, starting at time t, and let $Y_i^{t+1}, Y_i^{t+2}, \ldots, Y_i^{t+\tau}$ denote the τ observations. Each observation is normally distributed with mean θ_i and variance λ_i, and the mean of these observations is $\overline{Y}_i^{\tau} = \sum_{r=t+1}^{t+\tau} Y_i^r / \tau \sim \mathcal{N}(\theta_i, \lambda_i / \tau)$.

Bayes' rule provides the following posterior mean and covariance (Frazier et al. 2009):

$$\boldsymbol{\mu}^{t+\tau} = \boldsymbol{\mu}^t + \frac{\overline{Y}_i^{\tau} - \mu_i^t}{\lambda_i/\tau + \Sigma_{i,i}^t} \boldsymbol{\Sigma}^t \mathbf{e_i}, \text{ and}$$

$$\boldsymbol{\Sigma}^{t+\tau} = \boldsymbol{\Sigma}^t - \frac{\boldsymbol{\Sigma}^t \mathbf{e_i} \mathbf{e_i}^{\mathsf{T}} \boldsymbol{\Sigma}^t}{\lambda_i/\tau + \Sigma_{i,i}^t}, \tag{8}$$

where $\mathbf{e_i}$ is a $M \times 1$ vector with a one in row i and zeros elsewhere.

It will be useful to define $Z_i^{\tau} \equiv \mu_i^{t+\tau} - \mu_i^t$ to express the change in the mean belief of arm i:

$$\boldsymbol{\mu}^{t+\tau} = \boldsymbol{\mu}^t + \frac{Z_i^{\tau}}{\Sigma_{i,i}^t} \boldsymbol{\Sigma}^t \mathbf{e_i}. \tag{9}$$

For notational simplicity, let the effective sample size for arm i at time t be defined as $n_i^t = \lambda_i / \Sigma_{i,i}^t$. The distribution of Z_i^{τ} for a given τ and information at time t is then (DeGroot 2004)

$$Z_i^{\tau} \sim \mathcal{N}\left(0, \sigma_{Z_i^{\tau}}^2\right) \text{ where } \sigma_{Z_i^{\tau}}^2 = \frac{\lambda_i \tau}{n_i^t (n_i^t + \tau)}. \tag{10}$$

4.2. Adaptive Sampling from a Single Arm Before Selection: cPDE Heuristic

We approximate the optimal index for each arm i in (6) with that associated with a heuristic index that is based on the solution to an optimal stopping problem, which at each time $t = 0, 1, 2, \ldots,$ maximizes the EVI of repeatedly sampling *only* from arm i before selecting a best arm to implement. Our heuristic index combines the benefits of analogues that have been shown to be effective in related problems with correlated arms but whose values depend on an a priori fixed number of samples (Frazier et al. 2009) with those of indices that, like ours, are based on optimal stopping but with independent arms (Chick and Gans 2009, Chick and Frazier 2012). This section defines our main heuristic index and establishes some properties that simplify its calculation.

Let Π_i denote the set of all policies that sample only from arm i before an arm is selected as best, and consider the problem starting at time t, with the prior belief $(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t)$. Each $\pi_i \in \Pi_i$ has an associated relative stopping time T_i and an absolute stopping time of $t + T_i$. Such "adaptive look ahead" stopping times for indices have subscripts (e.g., $T_i, T_{i,l}$) to distinguish them from the stopping time for the overall clinical trial T, which does not have a subscript. Under the assumption that we are sampling only from arm i, starting at time t, the problem in (4) and (5) then becomes

$$V_i^*(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t) = \max\left\{-c_i + \Delta\mathbb{E}\left[V_i^*(\boldsymbol{\mu}^{t+1},\boldsymbol{\Sigma}^{t+1}) \;\middle|\; \boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t\right], \max_{j\in\mathcal{M}}\left\{P\mu_j^t - I_j\right\}\right\} \tag{11}$$

$$= \sup_{\pi_i\in\Pi_i} \mathbb{E}_{\pi_i}\left[\sum_{r=0}^{T_i-1} -\Delta^r c_i + \Delta^{T_i} \max_{j\in\mathcal{M}}\left\{P\mu_j^{t+T_i} - I_j\right\} \;\middle|\; \boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t\right], \tag{12}$$

where the index of summation r covers the $T_i - 1$ periods starting at t, over which we (conceptually) continue to allocate patients to arm i, and an optimal adaptive look ahead of $T_i^* = 0$ implies that we stop immediately and select the alternative with the greatest expected value.

4.2.1. Scale Invariance. In the remainder of this section, we assume for simplicity that $P = 1$ and $I_j = 0$ for all $j \in \mathcal{M}$. Proposition 2 shows that our results can easily be scaled for different values of P and I_j.

Proposition 2. *Let* $\boldsymbol{I} = [I_1, I_2, \ldots, I_M]$ *and* $\boldsymbol{0} = [0, 0, \ldots, 0]$. *Let* $V_i^*(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t; c_i, P, \boldsymbol{I}, \Delta)$ *denote the problem in* (12) *with its parameters explicitly stated. Then,*

$$V_i^*(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t; c_i, P, \boldsymbol{I}, \Delta) = P \times V_i^*\left(\boldsymbol{\mu}^t - \frac{\boldsymbol{I}}{P}, \boldsymbol{\Sigma}^t; \frac{c_i}{P}, 1, \boldsymbol{0}, \Delta\right). \tag{13}$$

Following the development in Section 3, we define a *stopping index* for policies $\pi_i \in \Pi_i$ to be the expected value of sampling from arm i, beyond stopping and selecting the current best, $V_i^{\pi_i}(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t) - \max_{j\in\mathcal{M}}\{\mu_j^t\}$. Subtraction of a constant $\max_{j\in\mathcal{M}}\{\mu_j^t\}$ does not impact the structure of the optimal solution for arm i. Given $P = 1$ and $I_j = 0$, we can use (9) to rewrite (12) so that posterior means at the adaptive look ahead time $T_i \geq 0$ are expressed as a function of $Z_i^{T_i}$ and that the EVI of sampling from arm i is

$$\mathrm{EVI}_i^*(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t) = \sup_{\pi_i} \mathbb{E}_{\pi_i}\left[\sum_{r=0}^{T_i-1} -\Delta^r c_i + \Delta^{T_i} \max_{j\in\mathcal{M}}\left\{\mu_j^t + \frac{\Sigma_{i,j}^t}{\Sigma_{i,i}^t} Z_i^{T_i}\right\} \;\middle|\; \boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t\right] - \max_{j\in\mathcal{M}} \mu_j^t. \tag{14}$$

Our proposed *cPDE stopping time* continues sampling at time t (declares $T > t$) if $\mathrm{EVI}_i^*(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t) > 0$ for at least one arm, $\max_{i\in\mathcal{M}} \mathrm{EVI}_i^*(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t) > 0$, and stops (declares $T = t$) otherwise.

We similarly define the *allocation index* for i to be a version of (7) that is restricted to sampling from arm i and is normalized by dividing by the sampling cost c_i,

$$\nu_i^*(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t) \equiv \frac{1}{c_i}\left(-c_i + \Delta\mathbb{E}\left[V_i^*(\boldsymbol{\mu}^{t+1},\boldsymbol{\Sigma}^{t+1}) \mid \boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t; u^t = i\right]\right)$$
$$= \frac{1}{c_i}\Delta\mathbb{E}\left[V_i^*(\boldsymbol{\mu}^{t+1},\boldsymbol{\Sigma}^{t+1}) \mid \boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t; u^t = i\right] - 1, \tag{15}$$

to reflect the expected value per monetary unit of observation when c_i values differ across arms.

Our proposed *cPDE allocation policy* allocates one observation at time t to the arm $i \in \mathcal{M}$ that maximizes $\nu_i^*(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t)$, with ties broken randomly. Note that, although the index $\nu_i^*(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t)$ reflects the possibility that $T_i > 1$, the *allocation policy* samples the index-maximizing arm *only once* before recalculating the arms' $\nu_i^*(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t)$ values, a distinction that also arises in related cKG policies.

4.2.2. Useful Simplifications. We now provide two simplifications for (14) and (15) that will be useful to prove claims of asymptotic consistency in Section 4.3. The first is based on an idea of Frazier et al. (2009). To this end, we define intercept and slope parameters for the linear functions in (14):

$$a_j = \mu_j^t \text{ and } b_j = \Sigma_{i,j}^t \;/\Sigma_{i,i}^t. \tag{16}$$

We denote by M' the number of *undominated* arms whose functions, $a_j + b_j z$, are maximal for some value of $z \in \mathbb{R}$ and by (l) the arm that has the lth lowest slope among undominated arms. Frazier et al. (2009) shows that without loss of generality the ordering is strict. We let $g(z)$ be a function that returns the index with the largest $a_{(l)} + b_{(l)}z$ value when evaluated at $z \in \mathbb{R}$. We break ties by choosing the largest (ordered) index. The new ordering implies that $\sup\{z \mid g(z) = (l)\} = \min\{z \mid g(z) = (l+1)\}$, and we express the intersection point, in z, between arms (l) and $(l+1)$ as

$$d_{(l)} = (a_{(l)} - a_{(l+1)}) \;/(b_{(l+1)} - b_{(l)}). \tag{17}$$

This allows the maximization of M terms in (14) to be replaced with a sum of $M' \leq M$ terms:

$$\mathrm{EVI}_i^*(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t) = \sup_{\pi_i} \mathbb{E}_{\pi_i}\left[\sum_{r=0}^{T_i-1} -\Delta^r c_i + \Delta^{T_i}\left(a_{g(0)} + b_{g(0)} Z_i^{T_i} + \sum_{l=1}^{M'-1} (b_{(l+1)} - b_{(l)})\left(-\mid d_{(l)}\mid + Z_i^{T_i}\right)^+\right) \;\middle|\; \boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t\right] - a_{g(0)}. \tag{18}$$

The second simplification applies for undiscounted rewards ($\Delta = 1$) and proceeds in three steps. (1) The $a_{g(0)}$ terms in (18) are constants and cancel out. (2) Let T_i^* be the optimal adaptive look ahead time determined by the solution to (18). Because we temporarily constrain the allocation decision to sample only from arm i, the expectation over π_i^* can be replaced with one over T_i^*. (3) Proposition 3 shows we can eliminate the term $\mathbb{E}_{T_i^*}[b_{g(0)} Z_i^{T_i^*} \mid \boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t]$.

Proposition 3. *Z_i^τ is a uniformly integrable martingale. $\mathbb{E}_{T_i}[Z_i^{T_i} \mid \boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t] = 0$ for any stopping time T_i.*

Thus, when $\Delta = 1$, the *stopping index* for arm i becomes

$$\mathrm{EVI}_i^*(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t) = \sup_{T_i \ge 0} \mathbb{E}_{T_i}\left[-c_i T_i + \sum_{l=1}^{M'-1} (b_{(l+1)} - b_{(l)})\left(-|d_{(l)}| + Z_i^{T_i}\right)^+ \mid \boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t\right], \quad (19)$$

and taking the expectation over $T_i \ge 1$ and dividing by c_i, the *allocation index* is then

$$\nu_i^*(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t) = \sup_{T_i \ge 1} \mathbb{E}_{T_i}\left[-T_i + \frac{1}{c_i} \sum_{l=1}^{M'-1} (b_{(l+1)} - b_{(l)})\left(-|d_{(l)}| + Z_i^{T_i}\right)^+ \mid \boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t\right]. \quad (20)$$

When $\Delta = 1$, we can explicitly compute (19) by changing the terminal reward function in Chick and Frazier (2012), $\max\{0, Z_i\}$, to account for correlation, $\sum_{l=1}^{M'-1}(b_{(l+1)} - b_{(l)})(-|d_{(l)}| + Z_i^{T_i})^+$. For $\Delta < 1$, a similar substitution for the terminal reward in Chick and Gans (2009) applies to (18).

4.3. Asymptotic Properties of the New cPDE Allocation Policy

For the case of undiscounted rewards ($\Delta = 1$), we can demonstrate that the cPDE allocation policy is asymptotically consistent. That is, we show that the selection, at each t, of an arm with maximal index $\nu_i^*(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t)$ in (19) appropriately converges as a fixed stopping time, $T = T_{\text{fix}}$, becomes arbitrarily large.

Our proof takes advantage of machinery that is used in XFC to prove the consistency of certain cKG allocation policies. Moreover, its approach uses a novel argument for indices that can be "sandwiched" between related cKG allocation indices and allows us to show that a broader set of allocation policies is asymptotically consistent, including related adaptive polices in Sections 5.1 and 8.2.

To state the conditions of our proofs precisely, we make explicit the assumptions and conditions related to those stated in XFC. In particular, the assumptions allow smaller subsets of arms, $\mathcal{M}_t \subseteq \mathcal{M}$, to be considered for allocation in any given period t.

Assumption 1. (i) *$\boldsymbol{\mu}^0$, $\boldsymbol{\Sigma}^0$, and $\boldsymbol{\Lambda}$ are known.* (ii) *$\boldsymbol{\Sigma}^0$ and $\boldsymbol{\Lambda}$ are positive definite.* (iii) $\mathbb{P}\left\{\lim_{T\to\infty} \sum_{t=1}^{T} \mathbf{1}\{j \in \mathcal{M}_t\} = \infty\right\} = 1$ *for each $j \in \mathcal{M}$.* (iv) *$\Delta = 1$ and $c_j > 0$ for each $j \in \mathcal{M}$.*

Given these explicit assumptions, we are ready to state our consistency result for cPDE.

Theorem 1. *Suppose Assumption 1 holds. Then,* (i) *$\lim_{T\to\infty} \Sigma_{i,i}^T = 0$ almost surely for each i;* (ii) *$\lim_{T\to\infty} \nu_i^*(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t) = -1$ almost surely for each i;* (iii) *$\lim_{T\to\infty} \mu_i^T = \theta_i$ almost surely and in L^2 for each i; and* (iv) *$\lim_{T\to\infty} \arg\max\{\mu_i^T\} = \arg\max\{\theta_i\}$ almost surely for each i.*

The convergence of $\nu_i^*(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t)$ to -1 implies that the EVI per unit cost of sampling converges to 0. The -1 arises from subtracting the normalized cost per sample in (15).

5. Assessing cPDE: Comparator Policies

This section presents comparator policies for use in the numerical experiments we conduct in Section 7. Section 5.1 presents new adaptive look ahead policies based on bounds on cPDE's indices. Sections 5.2 and 5.3 recall existing comparator polices from the literature. Section 5.4 summarizes the results of preliminary numerical tests of the quality and computational intensity of these comparators.

5.1. New Adaptive Look Ahead Comparators Based on Bounds for cPDE

We introduce upper and lower bounds on the cPDE indices in (19) and (20) given undiscounted rewards ($\Delta = 1$). The bounds prove to be useful for implementing cPDE, and they also provide computationally efficient indices that serve as basis of their own allocation policies and stopping times. We present the bounds and then define the allocation policies and stopping times that they determine.

5.1.1. Useful Bounds for cPDE Indices. The second term of the expectations in (19) and (20) is piecewise linear, and at each intersection point, $d_{(l)}$, $l = 1, 2, \ldots, M' - 1$, there is a kink. We can exploit this structure by allowing the supremum to differ for each piece, l, of the sum and let $T_{i,l}$ denote the (adaptive look ahead) stopping time for piece l. If we consider one piece at a time, we obtain $M' - 1$ lower bounds for the associated index, the maximum of which is the tightest of the lower bounds.

Proposition 4. *$\underline{\mathrm{EVI}}_i(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t) \le \mathrm{EVI}_i^*(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t)$ and $\underline{\nu}_i(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t) \le \nu_i^*(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t)$, where*

$$\underline{\mathrm{EVI}}_i(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t) \equiv \max_{l=1,2,\ldots,M'-1} \left\{ \sup_{T_{i,l} \ge 0} \mathbb{E}_{T_{i,l}}\left[-c_i T_{i,l} + (b_{(l+1)} - b_{(l)})\left(-|d_{(l)}| + Z_i^{T_{i,l}}\right)^+ \;\middle|\; \boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t\right]\right\}, \quad (21)$$

and $\underline{\nu}_i(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t)$ is constructed as in (21), taking stopping times to be $T_{i,l} \ge 1$ and dividing by c_i.

Alternatively, we can allocate the sampling cost of arm i among the $M' - 1$ pieces and sum their values after maximizing each separately. This leads to a class of upper bounds for $\mathrm{EVI}_i^*(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t)$ and $v_i^*(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t)$ that use a sum of $M'-1$ optimizations with allocated costs.

Proposition 5. *Let $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\dots,\alpha_{M'-1})$ be vector weights to be used to allocate samples, such that $\alpha_l \ge 0$ for all l and $\sum_{l=1}^{M'-1}\alpha_l = 1$. Then, $\overline{\mathrm{EVI}}_{i,\boldsymbol{\alpha}}(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t) \ge \mathrm{EVI}_i^*(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t)$ and $\overline{v}_{i,\boldsymbol{\alpha}}(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t) \ge v_i^*(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t)$, where*

$$\overline{\mathrm{EVI}}_{i,\boldsymbol{\alpha}}(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t) \equiv \sum_{l=1}^{M'-1} \sup_{T_{i,l}\ge 0} \mathbb{E}_{T_{i,l}}\left[-c_i\alpha_l T_{i,l} + (b_{(l+1)}-b_{(l)})\left(-\,|\,d_{(l)}\,|+Z_i^{T_{i,l}}\right)^+ \;\Big|\; \boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t\right], \tag{22}$$

and $\overline{v}_{i,\boldsymbol{\alpha}}(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t)$ is constructed as in (22), taking stopping times to be $T_{i,l} \ge 1$ and dividing by c_i.

The structure of an arbitrary subproblem l from (21) or (22) is the same as that of a stopping problem in which a single arm with unknown mean reward is compared with a known standard. It can be computed with standard techniques (Chick and Gans 2009, Chick and Frazier 2012).

Proposition 6. $\overline{\mathrm{EVI}}_{i,\boldsymbol{\alpha}}(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t)$ *and* $\overline{v}_{i,\boldsymbol{\alpha}}(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t)$ *are convex in* $\boldsymbol{\alpha}$.

Proposition 6 says that there is a least upper bound $\overline{\mathrm{EVI}}_{i,\boldsymbol{\alpha}^*}(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t)$ that can be found by minimizing over $\boldsymbol{\alpha}$. That said, we use equal weights, $\alpha_l = 1/(M'-1)$, in all experiments unless otherwise specified. Numerical results in Online Appendix C.1 provide the rationale for our choice.

5.1.2. Comparator Allocation Policies and Stopping Times.

The bounds on $v_i^*(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t)$ and $\mathrm{EVI}_i^*(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t)$ can themselves be used as indices that form the basis of associated allocation policies and stopping times. The *cPDELower allocation policy* allocates the next observation to the arm with the greatest lower bound, $\underline{v}_i(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t)$, and the *cPDELower stopping time* continues sampling if and only if the greatest of the arms' lower bounds, $\max_{i\in\mathcal{M}}\underline{\mathrm{EVI}}_i(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t)$, is strictly positive. The *cPDEUpper allocation policy* and *cPDEUpper stopping time* use the upper bounds, $\overline{v}_{i,\boldsymbol{\alpha}}(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t)$ and $\overline{\mathrm{EVI}}_{i,\boldsymbol{\alpha}}(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t)$, to make analogous decisions, where $\boldsymbol{\alpha}$ implicitly has equal weights unless otherwise specified. We use these bounds and policies to help us analyze our main index, cPDE.

5.1.3. Consistency.

The bounding strategy of Theorem 1 in Section 4.3 is robust and is easily adapted to prove analogous consistency results for the new cPDELower and cPDEUpper allocation policies.

Corollary 1. *The consistency results for cPDE, with index $v_i^*(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t)$, stated in Theorem 1 also hold for cPDELower, with index $\underline{v}_i(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t)$, and for cPDEUpper, with index $\overline{v}_{i,\boldsymbol{\alpha}}(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t)$.*

5.2. Adaptation of the Correlated Knowledge Gradient

The cPDE allocation policy and stopping time of Section 4.2 and the cPDELower and cPDEUpper policies of Section 5.1 rely on indices that reflect optimal, adaptive solutions to stopping problems. In contrast, the cKG approach of Frazier et al. (2009) computes allocation indices based on a fixed duration stopping time and is a useful comparator. In our notation, the cKG approach sets T_i in (19) to a specified τ and bases its indices on the incremental expected value of taking exactly τ more sample(s) and then stopping. This leads to a family of cKG-type lower bounds for $\mathrm{EVI}_i^*(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t)$ that account for sampling cost:

$$\mathrm{EVI}_i^{\mathrm{cKG}_\tau}(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t) \equiv -c_i\tau + \sum_{l=1}^{M'-1}(b_{(l+1)}-b_{(l)})\,\sigma_{Z_i^\tau}\,\psi\!\left(|d_{(l)}|/\sigma_{Z_i^\tau}\right), \tag{23}$$

where $\psi(x) \equiv \mathbb{E}[(X-x)^+]$ is the loss function of a standard normal random variable $X \sim \mathcal{N}(0,1)$.

In Section 7, we will compare cPDE and its variants to two cKG-type policies. The first sets $\tau = 1$, and the second optimizes over τ and is denoted as $\mathrm{EVI}_i^{\mathrm{cKG}_*}(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t) \equiv \sup_{\tau\ge 1}\mathrm{EVI}_i^{\mathrm{cKG}_\tau}(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t)$.

Proposition 7. $\mathrm{EVI}_i^*(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t) \ge \mathrm{EVI}_i^{\mathrm{cKG}_*}(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t) \ge \mathrm{EVI}_i^{\mathrm{cKG}_1}(\boldsymbol{\mu}^t,\boldsymbol{\Sigma}^t)$.

When implementing allocation policies, the KG approach uses the ratio of the EVI to the sampling cost rather than their difference. For consistency with extant KG literature, we also use a ratio, suitably shifted so it can be compared with zero as with our new indices defined. Thus, we use the following cKG_τ allocation index for the cKG_τ allocation policy:

$$\mathrm{cKG}_\tau \equiv \frac{1}{c_i\tau}\left\{\sum_{l=1}^{M'-1}(b_{(l+1)}-b_{(l)})\,\sigma_{Z_i^\tau}\,\psi\!\left(|d_{(l)}|/\sigma_{Z_i^\tau}\right)\right\} - 1. \tag{24}$$

In turn, we define the cKG_* allocation index with respect to the supremum of the index cKG_τ in (24) over $\tau \ge 1$. Although the knowledge gradient is typically used for allocation indices, we also assess the performance of cKG_* as a stopping time.

5.3. Other Existing Comparator Policies

In gauging the performance of cPDE policy, we also consider other allocation policies that exist in the literature. To assess the benefit of accounting for correlation among unknown means, we define so-called ESPB indices that (incorrectly) assume the arms are statistically independent. That is, they also use (19) and (20) but with an initial prior distribution that has $\Sigma^0_{i,j} = 0$ for $i \neq j$.

To more broadly assess the benefit of information-based indices, we also define policies that do not make use of the expected value of information. The *Equal allocation policy* assigns arms in round robin fashion, and the *Random allocation policy* picks arms randomly with equal probability. The *Variance allocation policy* selects the arm with the largest posterior variance for its unknown mean. It effectively seeks to estimate the mean reward of each arm.

Finally, the *Fixed stopping time* always stops after a given predetermined number of samples, T_{fix}. One can use optimization to choose T_{fix} to maximize $V^{\pi}(\boldsymbol{\mu}^0, \boldsymbol{\Sigma}^0)$ over sets of policies with fixed sample size, and we do so in Section 7.

5.4. Quality and Computational Issues for EVI Approximations

5.4.1. Quality of Stopping Indices to Approximate EVI^*_i. We find that the stopping index for cKG$_1$ severely underestimates and that for cKG$_*$ somewhat underestimates EVI^*_i for all mean prior values. We find that $\overline{\text{EVI}}_{i,\alpha}$ overestimates EVI* only when $M' > 2$, whereas $\underline{\text{EVI}}_i$ offers a good estimate for EVI* *except* when $M' > 2$ and the prior means of arms are close in value. See Online Appendix C.2.

5.4.2. Computational Speed. Experiments reported in Online Appendix C.3 show that ν^*_i takes two orders of magnitude more time to compute than do $\underline{\nu}_i$ and $\overline{\nu}_{i,\alpha}$. These latter allocation indices in turn require an order of magnitude longer to compute than that for cKG$_*$ and two orders of magnitude more time than that for cKG$_1$. These policies therefore represent a broad cross-section of speed-accuracy trade-offs. Online Appendix C.3 also reports how the bounds on EVI^*_i defining cPDEUpper and cPDELower can be used to dramatically reduce the time to compute the cPDE stopping time.

6. Prior Distribution for Phase II/III Dose-Finding Trials

The performance of allocation policies and stopping times depends on the specified prior distribution. In general, prior means, μ^0_i, variances and covariances, $\Sigma^0_{i,j}$, for the unknown means, $\boldsymbol{\theta}$, can be elicited from medical experts or derived from the results of earlier phases of a trial. One may choose to articulate covariances with a parametric family as in Section 6.1. Section 6.2 suggests a method for using pilot data to develop an empirical Bayes estimator for the prior and sampling distributions for a phase II/III dose-finding trial to identify the most cost-effective dose. We assess the use of these priors with numerical experiments in Section 7. An approach applicable to factorial trial designs has been described elsewhere (Chick et al. 2019).

6.1. Covariances for the Unknown Mean Rewards

Here, we focus on converting knowledge about the relationship among arms' mean rewards into a prior covariance matrix, $\boldsymbol{\Sigma}^0$, and explain an approach that applies to dose-finding trials that compare multiple dose levels of the same drug.

In our example, we include the current standard of care as a control, indexed by $i = 1$, and other dose levels by $i = 2, 3, \ldots, M$. The dose level of arm i is denoted by h_i for all $i \in \mathcal{M}$, so the control has dose level $h_1 = 0$. In the context of (5), our formulation of dose finding optimizes the expected net benefit of an intervention, net of costs, rather than identifying the so-called ED$_{95}$, the smallest effective dose level at which 95% of subjects have a desired clinical effect.

The efficacy of different dose levels of a drug is often modeled using a dose-response curve. (For an example, see Section 7.4.) In our context, the problem of finding a maximal θ_i can be seen as that of maximizing a discretized version of an unknown continuous function.

A common approach to the generation of such an unknown function assumes that it is a realization of a Gaussian process and parameterizes its priors by the choice of a mean function and a covariance function (Frazier et al. 2009). We use a squared exponential (Gaussian) kernel to model the covariance; for given parameters σ^2 and ζ, the covariance between arms i and j is

$$\Sigma^0_{i,j} = \sigma^2 \exp\left\{-\zeta(h_i - h_j)^2\right\}. \tag{25}$$

Here, σ^2 is the variance of each of the unknown means and is the same for all arms. The smoothness of the inferred function for the mean, from which our θ_i values are discretized, is determined by ζ. Large values of ζ imply that the correlation among arms is low and that the function value can change quickly. Other kernels can also be employed (Rasmussen and Williams 2006, Chen et al. 2013).

6.2. Manipulating Prior Distribution for Robustness

This section proposes the use of pilot study data and Gaussian process regression (GPR) to obtain parameter estimates for a model such as the one in Section 6.1. We then manipulate that prior, with a goal of supporting a decision maker who is concerned with the potential for misspecification of the prior distribution.

Let $\mathcal{M}_0 \subseteq \mathcal{M}$ be the set of arms sampled during the pilot study. We assume the pilot study initially obtains $n_{0,j}$ observations from each arm $j \in \mathcal{M}_0$ in the pilot study. We use the simulated pilot data, GPR with the covariance kernel in (25), and Matlab's fitrgp function to estimate the prior mean $\boldsymbol{\mu}^{0,\mathrm{GPR}}$, covariance $\boldsymbol{\Sigma}^{0,\mathrm{GPR}}$, and sampling variance Λ^{GPR}. If diagnostics suggest that a few more observations would benefit the quality of the estimates, we add some or check an alternative method of fitting the parameters. When the pilot is completed, we let $\overline{y}_{pilot,j}$ denote the sample mean for arm $j \in \mathcal{M}_0$. We call this the *GPR estimate* for the pilot data and can use it as a prior distribution for the adaptive trial. See Online Appendix C.4 for details.

Next, we explore two manipulations of the initial prior distribution to assess their potential to safeguard trial performance if the prior is misspecified. The *first manipulation* draws upon two observations of Powell and Ryzhov (2012). They note that, for reward maximization problems with the cKG_1 allocation policy, a prior mean that is manipulated to be too high, rather than too low, induces initial sampling for a wider range of arms, as cKG_1 is forced to check whether arms are optimal. They also note that, if a prior has been misspecified, a systematic increase in prior variances can induce more sampling, reducing the risk of premature stopping and a poorly selected arm. We formalize these observations by defining a *Robust prior* whose mean is a constant, at the maximum of the sample means from the pilot and of the GPR estimate, plus a "fudge" factor for uncertainty, z_α, times a standard error:

$$\begin{aligned}\mu_i^{0,\mathrm{ROB}} &= \max\{\max_{j_1 \in \mathcal{M}_0} \overline{y}_{pilot,j_1}, \max_{j_2 \in \mathcal{M}} \mu_{j_2}^{0,\mathrm{GPR}})\} \\ &\quad + z_\alpha \mathrm{sqrt}(\max_{j_3 \in \mathcal{M}} (\Sigma_{j_3,j_3}^{0,\mathrm{GPR}})),\end{aligned}$$

for all $i \in \mathcal{M}$. We also double the range of uncertainty for the Robust prior, with $\boldsymbol{\Sigma}^{0,\mathrm{ROB}} = 4\boldsymbol{\Sigma}^{0,\mathrm{GPR}}$.

We call the *second manipulation* of the original prior distribution the *Tilted prior*. It is similar to the Robust prior but has slightly more elevated means for low-dose levels and slightly less elevated means for higher-dose levels. Specifically, the Tilted prior has

$$\begin{aligned}\mu_i^{0,\mathrm{TILT}} &= \max\{\max_{j_1 \in \mathcal{M}_0} \overline{y}_{pilot,j_1}, \max_{j_2 \in \mathcal{M}} \mu_{j_2}^{0,\mathrm{GPR}})\} \\ &\quad + 2z_\alpha(1 - i/M)\mathrm{sqrt}(\max_{j_3 \in \mathcal{M}} (\Sigma_{j_3,j_3}^{0,\mathrm{GPR}}))\end{aligned}$$

for all $i \in \mathcal{M}$, and we set $\boldsymbol{\Sigma}^{0,\mathrm{TILT}} = 4\boldsymbol{\Sigma}^{0,\mathrm{GPR}}$. The Tilted prior encourages initial sampling at lower doses before higher doses, a feature that can help to address safety concerns in early-stage trials (Huang et al. 2015, Wheeler et al. 2019).

7. Numerical Assessment of cPDE for Clinical Trials

We use numerical experiments to assess the effectiveness of cPDE, our main heuristic index for the value of sampling information. We compare its performance with that of our other new allocation and stopping indices from Section 5.1 and with that of the indices introduced in Sections 5.2 and 5.3. Our experiments use examples calibrated for comparison with previous research, as well as those estimated from published data from a dose-finding trial.

Section 7.1 describes the performance metrics we use and how they were estimated. Section 7.2 compares the performance of various *allocation policies* for a given sample size. It shows that both the cPDE and cKG family, all EVI-based policies that account for correlation among arms, perform well as compared with allocation policies that do not model correlation or that focus on minimizing the variance of posterior means. Section 7.3 explores the effectiveness of the new *stopping times* and finds that response-adaptive stopping times based on multiarm multistep look aheads can be beneficial for maximizing our trial design objective in (5). Section 7.4 demonstrates the manipulations of the Gaussian process regression model in Section 6.2 that help select a prior distribution for a phase II/III dose-finding trial.

7.1. Metrics for Analysis and Experimental Details

Our experiments assess how the cPDE family of indices performs with respect to other allocation and stopping indices. In doing so, we allow the allocation and stopping indices of a policy to be based on different criteria. For example, the policy π that combines the cPDE allocation policy and the cPDEUpper stopping time is referred to as the cPDE-cPDEUpper policy, and it has expected reward $V^{\,\mathrm{cPDE-\,cPDEUpper}}(\boldsymbol{\mu}^0, \boldsymbol{\Sigma}^0)$. To emphasize a specific fixed value of $T_{\mathrm{fix}} = \tau$ with the Fixed stopping time, we may refer to the cPDE-τ policy or its expected reward as $V^{\,\mathrm{cPDE}-\tau}(\boldsymbol{\mu}^0, \boldsymbol{\Sigma}^0)$.

We use three main undiscounted $(\Delta = 1)$ metrics to measure the performance of a policy π: the *expected sample size* $E[T]$, opportunity cost, and total cost. The *expected opportunity cost* measures the regret of the selection decision $\mathcal{D}$, $E[OC] = \mathbb{E}_\pi\big[\max_j\{P\theta_j - I_j\} - (P\theta_{\mathcal{D}} - I_{\mathcal{D}}) \mid \boldsymbol{\mu}^0, \boldsymbol{\Sigma}^0\big]$. The *expected total cost*, $E[TC] = \mathbb{E}_\pi[\sum_{t=0}^{T-1} c_{u^t} \mid \boldsymbol{\mu}^0, \boldsymbol{\Sigma}^0] + E[OC]$, is the sum of the expected sampling cost and the expected opportunity cost. Minimizing E[TC] is equivalent to maximizing our main objective function, $V^\pi(\boldsymbol{\mu}^0, \boldsymbol{\Sigma}^0)$, in (3). We may also report the probability of correctly selecting the "true best," P(CS), and the average CPU time required to compute a policy's indices.

We estimate expected values for each metric using Monte Carlo simulation. Each replication of the simulation is a sample path within which allocation and stopping decisions are based on the policies and the specified prior distribution.

For each experiment, we estimate expected values by calculating the average and standard error over 1,000 (or more) sample-path replications. We use common random numbers (CRNs) to sharpen comparisons for the difference in performance between pairs of policies. Namely, we used CRN to match any unknown parameters (e.g., $\boldsymbol{\theta}$) for all policies and to match the samples $Y_i^t \mid \boldsymbol{\theta}$ for a given $\boldsymbol{\theta}$ and for all policies as well. We denote the sample average of simulation replications, which estimates the expectation, by $\mathrm{E}[\cdot]$.

7.2. Do the New Allocation Policies Improve the Speed of Learning?

This section focuses on assessing the effectiveness of each allocation policy as a function of sample size by pairing all allocation policies with the Fixed stopping time. We use the synthetic problem setup of Frazier et al. (2009), who first proposed cKG, to facilitate comparisons with prior literature on sequential optimization with correlation across arms. Although that problem is not explicitly a phase II/III dose-finding trial, it corresponds to a linear structure of M arms and is therefore amendable to the assessment described in Section 6 for such trials. We do so here.

7.2.1. Experimental Setup. We run four experiments with $M = 80$ arms in which we set $c_i = 1$, $I_i = 0$, and $\lambda_i = 0.01$ for all i. (Sampling costs were not explicitly modeled in that earlier paper.) In all four experiments, the prior mean equals zero for all arms, and $\sigma^2 = 0.5$. The four experiments vary the strength of the covariance across arms, with ζ equal to either $16/(80-1)^2$ or $100/(80-1)^2$ in (25), and the adopting population size, with P equal to either 10^6 or 10^8.

7.2.2. Numerical Comparison of the Allocation Policies. Figure 1 displays the results of pairing each allocation policy with the Fixed stopping time for $P = 10^6$. Results with $P = 10^8$, a larger adopting population size, are similar to those reported here (data not shown). The left panel of Figure 1 plots results for a lower correlation, $\zeta = 100/(80-1)^2$, and the right panel plots results for a higher correlation, $\zeta = 16/(80-1)^2$. The horizontal axes mark the sample size of the Fixed stopping time. The vertical axes display the $\log_{10}$ of the average opportunity cost for that sample size.

In both panels of Figure 1, pairwise differences among cKG_1, cKG_*, cPDEUpper, cPDELower, and cPDE are not statistically significant (95% confidence interval (CI)) after $T = 100$ observations. Variance and Random are also not significantly different from each other at $T = 100$. All other pairwise differences are significant at $T = 100$. The drop in the curve for Equal allocation policy as the sample size approaches 80 occurs as each arm is sampled exactly once.

Figure 1 provides two key insights. First, all EVI-based allocation policies that model correlation perform well, and performance improves with higher correlation. Second, all other allocation policies perform less well. ESPB, an EVI-based policy that assumes independence across arms, performs poorly when correlation is present, as do Equal and Random, which use roughly balanced sampling. For example, cPDE needs 20 observations to obtain the same amount of information as is acquired by ESPB with 100+ observations. Variance's focus on estimation, through minimization of the posterior variance of all unknown means, also yields poorer performance.

7.3. Do the New Stopping Times Improve the Total Value of the Trial?

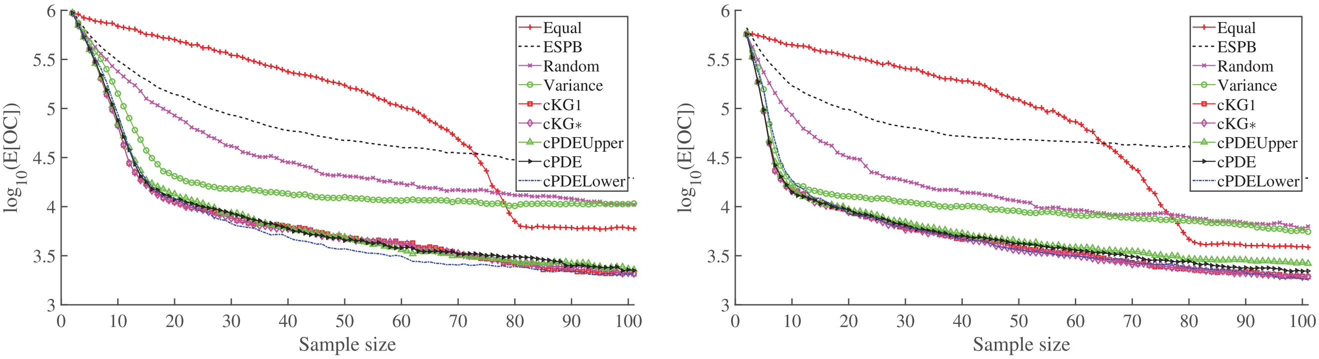
Section 7.2 showed that EVI-based indices that account for correlation perform well when used in allocation policies. We now address whether cPDE or its variations work well as stopping times, with an ability to effectively balance the expected value of information with the cost of sampling, to improve the total value of the trial in (5). For the special case of two arms, (a) the cPDELower, cPDE, and cPDEUpper stopping times are equivalent, and (b) these adaptive stopping times are optimal for the special case of two arm trials with patients allocated in pairs to those arms (Chick et al. 2017). In particular, the use of cPDE-type stopping times results in a higher expected reward than that does any fixed sample size policy in such cases.

For multiarm trials, however, it may be that a fixed length trial design (with stopping time T_{fix}) with response-adaptive sampling can provide a larger expected reward than does a design with a stopping time that looks at the EVI of future sampling from one arm (such as cKG-type or cPDE-type stopping times). This would occur if value of adaptively sampling from multiple arms can outweigh the benefit of optimal sampling with adaptive look ahead from one arm.

7.3.1. Experimental Setup. The experimental setup is the same as in the preceding section, with a population of $P = 10^6$ patients and the relatively lower level of correlation, $\zeta = 100/(80-1)^2$, in (25). These parameters obtain average sample sizes of 100–200 for most policies, which is not atypical for a preliminary phase II dose-finding trial (Huang et al. 2015).

7.3.2. Numerical Comparison. Table 1 presents Monte Carlo estimates of E[T], E[OC], and E[TC] for policies that combine cKG-type and cPDE-type indices and four fixed length policies. The Fixed stopping time with sample size 200 approximates the average sample size of the adaptive policy with the lowest E[TC]

Figure 1. (Color online) Log of the Expected Opportunity Cost, E[OC], for Several Policies with a Fixed Stopping Time for $\zeta = 100/(80-1)^2$ (Left Panel) and $\zeta = 16/(80-1)^2$ (Right Panel), with $P = 10^6$



in this experiment, the cPDELower-cPDEUpper policy. Similarly, the Fixed stopping times with sample sizes 130 and 150 approximate the average sample sizes of the cPDE stopping time for each allocation index. The Fixed stopping time with sample size of 493 is found by optimizing $V^{\text{cKG}_1 - T_{\text{fix}}}(\boldsymbol{\mu}^0, \boldsymbol{\Sigma}^0)$ over policies with fixed sample size $T_{\text{fix}} \in [1, 10{,}000]$ and the cKG$_1$ allocation policy (using Monte Carlo). For each allocation policy, we order stopping times by E[TC]. Results with $\zeta = 16/(80-1)^2$ are similar to those reported here and lead to the same qualitative conclusions (data not shown). Although we did not test cPDE allocation because it requires orders of magnitude more CPU time to compute than do the cKG$_1$ and cPDELower allocation policies, the heuristic presented in Online Appendix C.3 allows us to speed up computation of the cPDE stopping time to more easily assess it here.

Although the standard errors for individual policies in Table 1 may seem large, our use of CRN reduces the standard errors of differences between pairs of policies. The E[TC] values for the cKG$_1$-200, cKG$_1$-cPDEUpper, cPDELower-cPDEUpper, and cKG$_1$-cPDE policies are not statistically different from each other. The other policies perform significantly worse than the best five (95% CI).

The first, and perhaps most important, observation from Table 1 for this experiment with 80 arms is that there is clearly a benefit from optimizing a fixed stopping time with an adaptive allocation policy (here with cKG$_1$ and $T_{\text{fix}} = 493$), rather than using a cPDE-type or cKG-type stopping time, to reduce E[TC]. Thus, an adaptive sample size based on optimal adaptive look ahead from only one arm at a time is insufficient in this multiarm context, even though it is optimal (up to a diffusion approximation) for the two-arm contexts noted.

This result suggests a new stopping time; at each time t, one checks if there is a fixed number of additional observations, τ, such that the value of continuing to sample is positive. Namely, one continues to sample beyond time t if there is a $\tau \geq 1$ such that $V^{\text{cKG}_1 - \tau}(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t) > 0$, and one stops otherwise. Similar

Table 1. The Expected Sample Size, E[T]; the Expected Opportunity Cost, E[OC]; and the Expected Total Cost, E[TC], for Several Policies

Allocation	Stopping	E[T]	S.E.	E[OC]	S.E.	E[TC]	S.E.	P(CS)	CPU
cKG$_1$	Fixed-493	493.00	0.00	316.16	59.20	809.16	59.20	0.94	4.78
cKG$_1$	Fixed-200	200.00	0.00	860.05	64.93	1,060.05	64.93	0.89	2.88
cKG$_1$	cPDEUpper	201.53	4.81	970.52	101.48	1,172.05	101.41	0.92	43.21
cKG$_1$	cPDE	129.01	2.49	1,209.29	107.88	1,338.30	107.84	0.89	105.67
cKG$_1$	Fixed-130	130.00	0.00	1,481.04	118.90	1,611.04	118.90	0.85	2.50
cKG$_1$	cPDELower	105.80	2.78	1,679.28	162.51	1,785.07	162.40	0.87	20.66
cKG$_1$	cKG$_*$	44.57	1.67	4,055.20	435.52	4,099.70	435.50	0.76	8.38
cPDELower	cPDEUpper	205.27	4.89	863.78	88.44	1,069.05	88.45	0.91	154.12
cPDELower	cPDE	150.06	3.10	1,022.07	91.65	1,172.13	91.64	0.89	309.70
cPDELower	Fixed-200	200.00	0.00	1,182.68	116.86	1,382.68	116.86	0.87	118.61
cPDELower	cPDELower	111.59	2.72	1,570.55	141.91	1,682.14	141.85	0.87	82.73
cPDELower	Fixed-150	150.00	0.00	1,592.77	190.85	1,742.77	190.85	0.85	89.39
cPDELower	cKG$_*$	58.77	2.31	2,835.05	301.35	2,893.81	301.20	0.79	49.58
Variance	Fixed-200	200.00	0.00	7,387.00	598.42	7,587.00	598.42	0.66	0.11
Variance	Fixed-150	150.00	0.00	8,212.30	670.81	8,362.30	670.81	0.66	0.10

Note. Also shown are the standard error (S.E.) of the Monte Carlo estimate of those quantities, the probability of correct selection, P(CS), and average CPU time (seconds per simulated trial).

stopping times would check if $V^{\text{cPDE}-\tau}(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t) > 0$ or $V^{\text{cPDELower}-\tau}(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t) > 0$ for each t but may require more CPU time to compute.

A further exploration of such new adaptive stopping times is an area for future research and beyond the scope of this paper. However, Table 1 provides some additional observations, which might inform further work. First, as with Section 7.2, there is clearly a benefit to use an adaptive allocation policy that seeks to optimize (e.g., cKG$_1$, cPDELower) as opposed to estimate (e.g., Variance), for a given stopping time.

Second, for each adaptive stopping time tested here, the cPDELower allocation policy was more effective than the cKG$_1$ allocation policy. Thus, the cPDELower allocation policy may be useful when an adaptive stopping time is used. The cKG$_1$ allocation policy is much faster to compute and was therefore more practical to assess empirically here.

Third, we compare the best of the policies with adaptive sample sizes with the best of the policies with a similar fixed sample size. Specifically, cPDELower-cPDEUpper has an average sample size (E[T] ≈ 205) similar to that of cKG$_1$ -200. We note that cPDELower-cPDEUpper has a very similar expected total cost to cKG$_1$ -200 ($1,069 > 1,060$) and a (statistically significantly) better P(CS) ($0.912 > 0.887$), even though value-based heuristics seek to optimize E[TC] rather than P(CS). Similarly, cPDELower-cPDE has a better E[TC] and P(CS) than cPDELower-150 ($1,172 < 1,742$ and $0.895 > 0.851$), and cKG$_1$ -cPDE has a better E[TC] and P(CS) than cKG$_1$ -130 ($1338 < 1611$ and $0.893 > 0.850$). Thus, a response adaptive sample size may be valuable relative to trials with fixed sample sizes, for E[TC], E[OC], and P(CS), for a given expected sample size.

In summary, there is value in allowing stopping times to be response adaptive. The results indicate that there is more value in finding a stopping time with optimal fixed look ahead in a way that allows for multiple arms to be response adaptively sampled, when compared with (a) optimal adaptive look ahead that allows for sampling from only one arm or (b) optimal fixed look ahead with multiple arms allocated to improve response estimation rather than response optimization. These comments assume that one is confident that the prior distribution is specified well.

7.4. How to Pragmatically Choose a Prior Distribution for a Dose-Finding Trial?

Section 7.3 made numerical assessments for synthetic experiments under the assumption that the problem configurations were sampled from the trial manager's prior distribution. In theory, the trial manager's beliefs about the unknown mean rewards should determine the prior. In practice, a trial designer may seek reassurance about selecting a prior for the unknown mean rewards so that it is robust to the risk of misspecification, as discussed in Section 6.2.

This section explores pragmatic ways to specify a prior distribution that has "good" results when applied to a representative phase II/III dose-finding trial. Our experiments illustrate how to apply our proposed trial design and are not intended to make clinical recommendations.

7.4.1. Experimental Setup. Here, we presume that arm 1 is a control, with a dose of zero. We then have arms 2, 3, …, 17 that represent 16 positive dose levels, with doses proportional to $2^{(i-1)/2}$, so that doses are evenly spaced on a log scale. It is known that health benefits may increase in effectiveness as effective levels are achieved and then may decrease as toxic levels are achieved (Dimmitt et al. 2017). Such effectiveness benefits and toxicity levels are often described individually with logistic curves (Gadagkar and Call 2015), and we use logistic curves here for illustrative purposes.

Specifically, we assume the ground truth for arm 17, with dose level $2^{(17-1)/2} = 256$, corresponds to the maximum dose identified from a previously conducted phase I/IIa dose-ranging trial. Such trials study safety and tolerability (Berry et al. 2002, Bornkamp et al. 2007). We may use such dose-ranging data to guide our simulated seamless phase II/III dose-finding trial. For that trial, observations are simulated according to a ground truth that is unknown to the modeler. For the unknown ground truth, we assume that the true ED50 (dose level effective for 50% of patients) is four, so that the maximum dose tested is $2^{8-4} = 16$ times stronger than the ED50 dose. We then compute the difference of those logistic functions, with assumed parameters so that toxicity effects of particularly high doses result in deleterious outcomes:

$$\theta_i = \frac{4500}{1 + \exp[-2 \times ((i-1)/2 - 4)]} - \frac{7000}{1 + \exp[-1.5 \times ((i-1)/2 - 8)]}.$$

We assume the unknown true sampling standard deviation is $\sqrt{\lambda_i} = \text{US\$}4240$ for all i. We set the population size to be $P = 2 \times 10^5$, and we assume $I_i = \text{US\$}0$ and $c_i = \text{US\$}8500$ for all i. The true optimal of these dose levels is then arm 12, which has health benefit $\frac{4500}{1+\exp[-2\times((12-1)/2-4)]} \approx 4286$, which is $4286/4500 \approx 95\%$ of the maximum effective benefit. Although this model can be further refined for applications, its structure is appropriate for generating relevant insights.

7.4.2. Simulated Pilot Study. We use the empirical Bayes approach of Section 6.2 to obtain the GPR prior distribution based on simulated pilot study data. We then run our seamless phase II/III dose-finding trial to maximize (5).

To develop the GPR prior, we chose the set $\mathcal{M}_0$ of arms to sample during the simulated pilot study to be the control (dose level 0) and four other doses (2.0, 4.0, 6.0, and 8.0). One replication of a simulated pilot study was generated by 10 initial observations each from the assumed ground truth for the control (dose level 0) and the four other doses (2.0, 4.0, 6.0, and 8.0). We then derive the two manipulations of that prior, the Robust and Tilted priors of Section 6.2.

Figure 2 depicts the GPR estimate from a representative simulated pilot study (left panel), along with its associated Robust prior (center panel) and Tilted prior (right panel). Here, we set $z_\alpha = 0.5$. In each panel, the horizontal axis shows the dose levels of the arms, and the vertical axis shows the dollar value of the mean response. The dashed lines represent the prior means, μ_i^0, and the gray bands surrounding the means display a confidence band that is ± 1 sqrt($\Sigma_{i,i}^0$) wide.

To estimate the effectiveness of our policies for phase II/III dose-finding trials with these priors, we use the ground truth, θ_i, to simulate 1,000 sets of pilot study data. For each set of pilot data, we compute its GPR estimate and obtain each realized prior distribution (GPR, Robust, and Tilted). We then compute each policy 's performance characteristics for each prior by averaging over simulated pilots and trials. We use CRN to sharpen comparisons among policies.

7.4.3. Results.

Table 2 summarizes a subset of results. The table lists policies with the lowest E[TC] first for a given prior distribution, so that the best policies appear first for a given prior. We do not display results with the cKG$_*$ stopping time as results were poor in comparison with other stopping times. As in Section 7.3, we assessed certain fixed sample sizes with adaptive allocations. The optimal τ to minimize E[TC] over cKG$_1$ -τ policies depends upon the prior distribution. Thus, the Fixed sample size reported in Table 2 differs for the GPR, Robust, and Tilted priors. The average E[TC] values for the GPR and Tilted priors are displayed in Figure 3, which shows the expected cost of sampling, E[OC], and E[TC] as a function of the sample size (averaged over 4,000 random pilot studies). Note that E[TC] is relatively flat near the optimal fixed sample size, so minor variation in the optimal τ for a given prior is not likely to drastically change results.

Our first observation is that most insights from Section 7.3 apply here as well. For example, response adaptive stopping rules with a fixed sample size chosen to optimize $V^{\text{cKG}_1-\tau}$ performed best for each of the three prior distributions tested. We found two differences from the insights of Section 7.3. The first is that the cPDE stopping time tended to be as good or better than the cPDEUpper stopping time. The second is that the Variance allocation policy performed relatively better here, as compared with expectations from Section 7.3, although adaptive allocation policies still perform best. These mild deviations might be explained because Section 7.3 averaged results over randomly sampled dose-response curves, whereas this section assumes a single underlying ground truth that is representative of a single unimodal dose-response curve.

Our second observation is that the Robust and Tilted priors provided an effective reduction in E[TC] for each given policy in Table 2. For example, the E[TC] with cKG$_1$ -cPDE is approximately 1.96×10^7 with the Robust and Tilted priors and is 2.47×10^7 with the GPR prior, a 20% decrease in expected total cost. Similar results were obtained with fixed stopping times, such as for cKG$_1$ -1,100 and cPDELower-1,100 (data not shown). We also observed that the Tilted prior was best at encouraging lower doses to be tested initially more frequently than did the Robust prior, which in turn, favored lower initial doses more than the GPR prior, as expected, for all policies with adaptive allocation policies (data not shown).

In summary, the Robust and Tilted priors, which were manipulations of a GPR prior estimated from pilot data, resulted in a higher expected reward (lower

Figure 2. Actual Dose-Response Curve, $\boldsymbol{\theta}$, Together with the GPR Estimate from One Simulated Pilot Study (Left Panel) as well as Its Associated Robust Prior (Center Panel) and Tilted Prior (Right Panel)

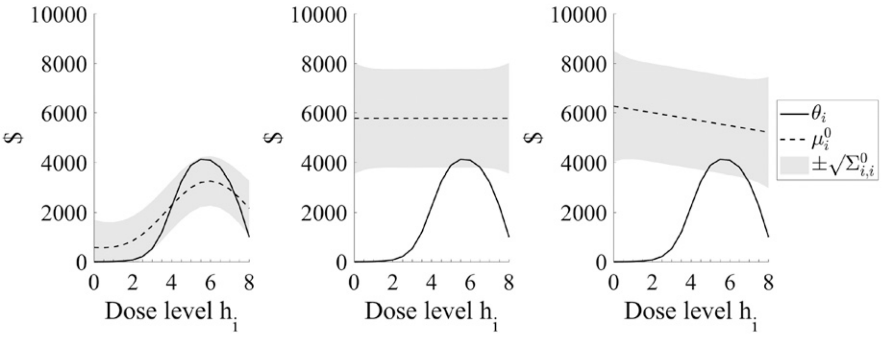


Table 2. The Expected Sample Size, E[T]; the Expected Opportunity Cost, E[OC]; and the Expected Total Cost, E[TC], for Several Policies Using the Gaussian Process Regression Prior, Robust Prior, and Tilted Prior Distribution

Prior	Allocation	Stopping	E[T]	S.E.	E[OC]	S.E.	E[TC]	S.E.	P(CS)
GPR	cKG$_1$	Fixed-1,060	1,060.0	0.0	1.12E+07	1.85E+06	2.02E+07	1.85E+06	0.52
GPR	cPDELower	Fixed-1,060	1,060.0	0.0	1.14E+07	1.88E+06	2.04E+07	1.88E+06	0.51
GPR	Variance	Fixed-1,060	1,060.0	0.0	1.41E+07	1.57E+06	2.31E+07	1.57E+06	0.55
GPR	cKG$_1$	cPDE	541.4	12.8	2.01E+07	3.09E+06	2.47E+07	3.07E+06	0.54
GPR	cKG$_1$	cPDEUpper	1,104.8	25.1	1.60E+07	2.69E+06	2.54E+07	2.67E+06	0.51
GPR	cPDELower	cPDEUpper	1,081.6	24.5	1.66E+07	2.76E+06	2.58E+07	2.73E+06	0.51
GPR	cPDELower	cPDE	601.5	15.7	2.10E+07	3.19E+06	2.61E+07	3.18E+06	0.51
Robust	cKG$_1$	Fixed-565	565.0	0.0	1.11E+07	1.65E+06	1.59E+07	1.65E+06	0.46
Robust	cPDELower	Fixed-565	565.0	0.0	1.29E+07	1.87E+06	1.77E+07	1.87E+06	0.45
Robust	cKG$_1$	cPDE	689.3	15.1	1.38E+07	2.38E+06	1.96E+07	2.37E+06	0.49
Robust	Variance	Fixed-565	565.0	0.0	1.72E+07	1.72E+06	2.20E+07	1.72E+06	0.52
Robust	cPDELower	cPDE	825.3	20.0	1.54E+07	2.61E+06	2.25E+07	2.59E+06	0.49
Robust	cKG$_1$	cPDEUpper	1,202.1	27.6	1.29E+07	2.23E+06	2.31E+07	2.21E+06	0.45
Robust	cPDELower	cPDEUpper	1,158.4	26.7	1.33E+07	2.32E+06	2.32E+07	2.29E+06	0.46
Tilted	cKG$_1$	Fixed-595	595.0	0.0	0.95E+07	1.33E+06	1.46E+07	1.33E+06	0.46
Tilted	cPDELower	Fixed-595	595.0	0.0	1.24E+07	1.87E+06	1.75E+07	1.87E+06	0.44
Tilted	cKG$_1$	cPDE	684.6	15.1	1.38E+07	2.38E+06	1.96E+07	2.37E+06	0.48
Tilted	Variance	Fixed-595	595.0	0.0	1.61E+07	1.57E+06	2.12E+07	1.57E+06	0.52
Tilted	cKG$_1$	cPDEUpper	1,207.4	27.7	1.13E+07	2.04E+06	2.15E+07	2.02E+06	0.45
Tilted	cPDELower	cPDE	835.3	19.7	1.57E+07	2.60E+06	2.28E+07	2.58E+06	0.45
Tilted	cPDELower	cPDEUpper	1,168.0	25.9	1.36E+07	2.38E+06	2.36E+07	2.35E+06	0.45

Note. Also shown are the standard error (S.E.) of the Monte Carlo estimate of those quantities and the probability of correct selection, P(CS).

E[TC]) in this example with a representative dose-response curve. These results suggest that the Robust or Tilted priors may be useful manipulations, along with an adaptive allocation policy and fixed stopping time τ optimized to maximize the value of that allocation policy. Here, we chose τ to optimize the predicted expected value of a cKG$_1$ -τ policy for reasons of computational speed in computing rewards. Future work would include assessing whether optimizing τ for the cPDELower-τ policy or whether using the response adaptive stopping times suggested in Section 7.3 might further improve rewards.

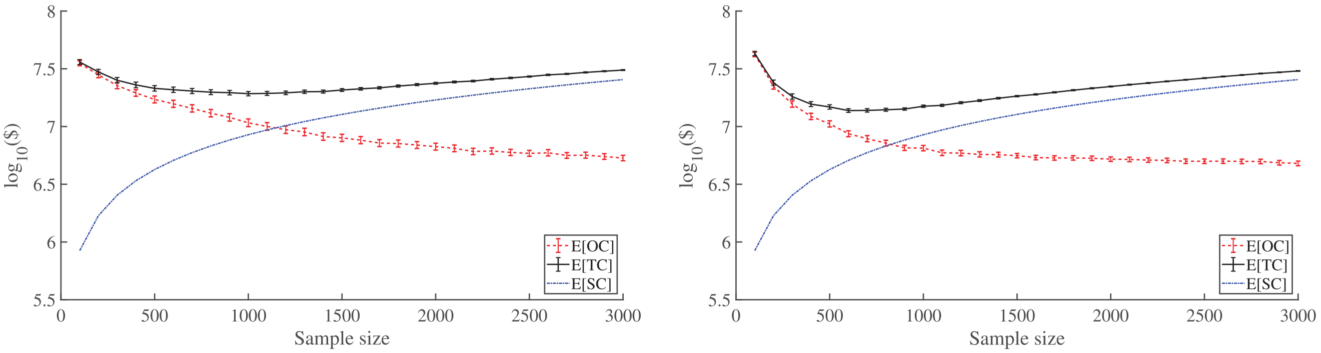
8. Additional Considerations for Clinical Trial Applications

This section discusses some important practical considerations for clinical trials. In many trials, there are delays between treatment initiation and the time at which outcomes are observed. Section 8.1 proposes approaches to handle such delayed observations. Section 8.2 describes practical ways to manage randomization, an important technique to manage bias or confounding (Piantadosi 1997). Section 8.3 notes that our proposed allocation policies also apply to trials with normally distributed observations, even if QALY data are not available. Online Appendix D addresses the evaluation of frequentist power curves and issues to further enhance the framework for clinical trials.

8.1. Delay Between Treatment Initiation and Observation of Outcome

We have assumed so far that each patient's outcome is observed before the next patient arrives. Although many trials satisfy this property (Flight et al. 2017),

Figure 3. (Color online) The Expected Opportunity Cost, E[OC]; the Expected Total Cost, E[TC]; and the Expected Sampling Cost, E[SC], with cKG$_1$ Allocation Index and Fixed Stopping Rule, as a Function of Sample Size for Dose-Response Application in Section 7.4 (for GPR Prior in the Left Panel and for Tilted Prior in the Right Panel)



many trials do not. For the case of two-armed trials with a protocol that specifies a fixed duration delay, Chick et al. (2017) solve for fully sequential stopping times with correlated arms by randomizing patients in pairs. For multiarm trials with delayed observations, randomizing M patients to all M arms may be inefficient.

For $M > 2$ arms, we propose a simple heuristic to account for delays based on batch allocations. Suppose there are B_t "pipeline" patients whose treatment has started but whose outcomes are not yet observed at time t. One can approximate the EVI of an allocation for $B_t + 1$ patients, given $(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t)$ and the existing assignment for the B_t "pipeline" patients, with related Bayesian ranking and selection techniques with batch allocation policies based on EVI-type criteria. This entails the assessment of M approximations for the EVI of those M ways of assigning an arm to the B_{t+1} st patient. There are several approximations available for mean rewards of correlated arms (Chick and Inoue 2001, Fu et al. 2007, Wu and Frazier 2016). Experiments in those papers suggest this approach may be useful and effective in our context, at a loss of the benefit of adaptive look ahead.

8.2. Randomized Allocation Policies and Their Asymptotic Properties

The base models for cPDE-type and cKG$_1$ allocation policies do not randomize. We present three ways to introduce randomization in our framework and then comment on how such randomization might help address statistical issues of confounding.

Our first type of randomized allocation adapts an idea of Williamson et al. (2017) for two-armed trials with Bernoulli outcomes. Instead of action i representing the choice of arm i, we let it represent a probabilistic assignment, which favors but does not always choose arm i. Specifically, at each time t, we allocate to each arm via random draw with probability p/M, and with additional probability $(1-p)$, we allocate to an arm with the maximal allocation index, with ties for the maximizer broken randomly. These randomized versions of our cPDE family of allocation policies possess the desirable asymptotic consistency properties established for their nonrandomized analogues.

Corollary 2. *The consistency results stated in Theorem 1 also hold for the randomized versions of the cPDE, cPDE-Lower, and cPDEUpper allocation policies.*

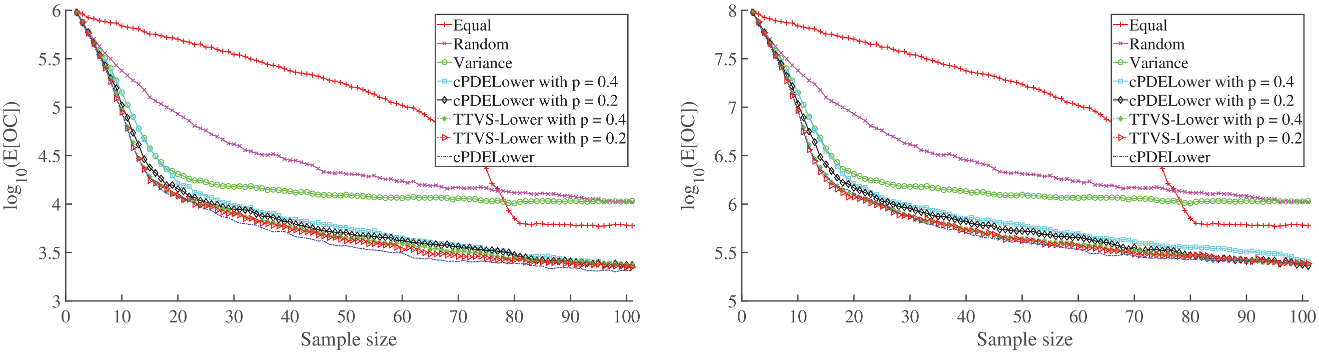
Our second type of randomized allocation is motivated by top-two value-sampling (TTVS) allocation (Russo 2020). TTVS adaptively and randomly allocates among the two arms that have the highest expected value of perfect information for each t. TTVS does not consider the cost of sampling, which would be infinite to obtain perfect information. Instead, we implement TTVS for cPDE-type indices that account for sampling costs and allow them to differ across arms by sampling the arm with the largest allocation index with probability $1-p$ and the arm with second largest index with probability p. The resulting allocation policies also inherit the asymptotic convergence properties of Corollary 2. (See Online Appendix B.11.) Here, we denote by *TTVS-Lower* the allocation policy that uses TTVS randomization for the cPDELower allocation index.

Figure 4 compares the performance of a randomized cPDELower allocation policy for p = 0, 0.2, and 0.4 against the Equal, Random, Variance, and TTVS-Lower (for p = 0.2 and 0.4) allocation policies, all with the Fixed stopping time. It is based on the same problem setup as that in Figure 1 in Section 7.2, and the key features of two graphs are also the same. Figure 4 shows that the performance of the randomized cPDE-Lower allocation policy is not strongly impaired when an arm is randomly selected with $p \in \{0.2, 0.4\}$. Randomized cPDELower (p = 0.2) required approximately 25%–30% more observations than cPDELower to achieve similar levels of log(E[OC]). Results for the randomized cPDEUpper and cKG$_*$ allocation policies are similar to those for the randomized cPDELower allocation. For a given randomization probability, p, results for the randomized TTVS-Lower are even less impaired than is the case for randomized cPDELower. Randomized TTVS-Lower (p = 0.2) required approximately 7%–15% more observations than cPDELower to achieve similar levels of log(E[OC]). This is because the randomized arm selected with TTVS-Lower is the second most informative arm, whereas randomization with cPDELower can pick any arm. Thus, randomization maintains asymptotic consistency and can work well with small sample sizes.

A third type of randomized allocation selects a batch of treatments for randomization across a batch of patients. For example, the cKG$_\tau$ allocation of XFC can be used to force an allocation of two arms to two patients at a time. To randomize $q \geq 2$ arms to q patients at a time, one can adapt the techniques mentioned in Section 8.1 for delays, which allows for allocation of arms in batches.

These three randomization techniques can be adapted to help balance covariates, addressing some issues of bias or confounding (Piantadosi 1997). For example, in the first model one can reserve some fraction of the randomization probability to allocate with probabilities that are proportional to those for known randomization techniques, such as propensity score methods (PSMs), to achieve desired balance (Rosenbaum and Rubin 1983, Li and Li 2019). The randomization probability $p < 1$ can be spread flexibly over the M arms and vary over t. A sufficient condition for convergence to hold is that the p_{it} with $\sum_{j=1}^{M} p_{jt} = p$ are chosen to fulfill

Figure 4. (Color online) Log of the Expected Opportunity Cost, E[OC], for Several Policies with a Fixed Stopping Time for $P = 10^6$ (Left Panel) and $P = 10^8$ (Right Panel)



part (iii) of Assumption 1. Alternatively, if a batch of q arms is to be allocated to a batch of patients (third type of randomization), one can allocate them to help address bias using PSM or group-sequential techniques (Villar et al. 2018). Or, one might adapt TTVS by randomizing arms to q patients with probability roughly proportional to the EVSI per unit sampling cost, $\nu_i^* + 1 + \epsilon$, for some small $\epsilon > 0$, a regret-weighted variation on Thompson sampling (Russo 2020). Such issues of potential confounding are applicable to adaptive trials more broadly (Berry 2011, Lipsky and Greenland 2011), and a study of how results for those other trials apply to our proposal is an area for future research.

8.3. Trials with Real-Valued Outcomes but Without QALY Information

Our proposed allocation policies can be applied to other clinical trials with real-valued outcomes, even if cost and QALYs are not observed, if they can be modeled by a normal distribution. This can be done by setting $\mathbb{1}_{CE} = 0$ in (1). Our allocation policies would then seek to minimize for any modeler-specified stopping rule, such as a traditional fixed sample size, the E[OC] from the trial. The numerical results for allocation policies in Section 7.2 therefore are relevant to nonvalue-based trials with other such real-valued outcomes. Settings where health outcomes cannot be converted to monetary values would seem to prevent the application of our value-based stopping rules.

9. Discussion and Conclusions

This paper responds to calls from regulators and funders of clinical trials that seek innovative, efficient trial designs (EU 2014, FDA 2016, Hudson et al. 2016, EMA 2017, NHS England 2017, NIHR 2020). It formulates a novel Bayesian, decision-theoretic model for *fully sequential sampling* for adaptive *multiarm clinical trials with correlated mean rewards*. It adopts a *value-based framework* to help answer the following questions: When beliefs regarding mean outcomes from arms are correlated, to which arms should patients be allocated as the trial evolves? When should one stop patient recruitment in an adaptive, value-based trial and implement the selected arm?

We provided structural results that characterize the optimal solution for such fully sequential, value-based trials with multiple correlated arms. We constructed the cPDE index, a heuristic approximation of the optimal EVI of optimal sampling, which combines the benefits of modeling correlation and optimal stopping times for valuing additional information. The cPDE and associated allocation policies possess asymptotic convergence and strong empirical performance.

It is known that cPDE-type policies are optimal adaptive stopping times when *allocating patient pairs to two arms* (Chick et al. 2017). It is also known that *allocation policies* that are based on the EVI of further sampling from one arm can be very effective (Frazier et al. 2009, Russo 2020, this paper). Section 7 suggests that, for multiarm trials such as dose-finding trials, optimal stopping requires evaluating the EVI of potential further adaptive sampling from multiple arms over multiple steps. Also, the cKG_1 or cPDELower allocation, with a dynamically optimized sample size, and a manipulated prior distribution have desirable performance in our numerical experiments. The present work extends past KG work by accounting for practical issues in multiarm trials such as delayed observations, randomization, stopping times, and the development of practical methods for constructing a prior distribution based on pilot study data. It also contributes to an understanding of potential future developments for multistep adaptive look ahead stopping times that balance the cost of learning and the expected benefits of that learning.

The types of trials handled by our base model are those for which the outcomes are observed quickly after treatment initiation, whose data allow for cost-benefit information to be accumulated during the trial, and which take a social planner's perspective. We extend our base model to account for randomization and provide a heuristic to handle short to intermediate delays in observing outcomes. The online companion presents theory and practical implementation details.

Acknowledgments

The authors thank Paul Shaman for discussions about factorial design as well as Andres Alban, Martin Forster, Sofia Villar, Peter Jacko, and Spyros Zoumpoulis for feedback on an earlier version of the manuscript. The authors also thank the review team for helpful, constructive feedback.

References

Adaptive Platform Trials Coalition (2019) Adaptive platform trials: Definition, design, conduct and reporting considerations. *Nature Rev. Drug Discovery* 18(10):797–807.

Ahuja V, Birge JR (2016) Response-adaptive designs for clinical trials: Simultaneous learning from multiple patients. *Eur. J. Oper. Res.* 248(2):619–633.

Alban A, Chick SE, Forster M (2020) Value-based clinical trials: Selecting trial lengths and recruitment rates in different regulatory contexts. Discussion Paper No. 20/01, Department of Economics, University of York, York, United Kingdom.

Anderer A, Bastani H, Silberholz J (2021) Adaptive clinical trial designs with surrogates: When should we bother? *Management Sci.* Forthcoming.

Bastani H, Bayati M (2020) Online decision-making with high-dimensional covariates. *Oper. Res.* 58(1):276–294.

Berry DA (2011) Adaptive clinical trials: The promise and the caution. *J. Clinical Oncology* 29(6):606–609.

Berry DA (2012) Adaptive clinical trials in oncology. *Nature Rev. Clinical Oncology* 9(4):199–207.

Berry DA, Ho CH (1988) One-sided sequential stopping boundaries for clinical trials: A decision-theoretic approach. *Biometrics* 44(1):219–227.

Berry DA, Mueller P, Grieve AP, Smith M, Parke T, Blazek R, Mitchard N, Krams M (2002) Adaptive Bayesian designs for dose-ranging drug trials. Gatsonis C, Kass RE, Carlin B, Carriquiry A, Gelman A, Verdinelli I, West M, eds. *Case Studies in Bayesian Statistics, Lecture Notes in Statistics*, vol. 162 (Springer, New York), 99–181.

Boeree MJ, Heinrich N, Aarnoutse R, Diacon AH, Dawson R, Rehal S, Kibiki GS, et al. (2017) High-dose rifampicin, moxifloxacin, and SQ109 for treating tuberculosis: A multi-arm, multi-stage randomised controlled trial. *Lancet Infectious Diseases* 17(1): 39–49.

Bornkamp B, Bretz F, Dmitrienko A, Enas G, Gaydos B, Hsu C-H, König F, et al. (2007) Innovative approaches for designing and analyzing adaptive dose-ranging trials. *J. Biopharmaceutical Statist.* 17(6):965–995.

Bravo F, Corcoran T, Long E (2021) Flexible drug approval policies. *Manufacturing Service Oper. Management*, ePub ahead of print March 23, https://doi.org/10.1287/msom.2020.0963.

Cai C, Yuan Y, Johnson VE (2013) Bayesian adaptive phase II screening design for combination trials. *Clinical Trials* 10(3): 353–362.

Chen X, Ankenman BE, Nelson BL (2013) Enhancing stochastic kriging metamodels with gradient estimators. *Oper. Res.* 61(2): 512–528.

Chick SE, Frazier PI (2012) Sequential sampling for selection with economics of selection procedures. *Management Sci.* 58(3): 550–569.

Chick SE, Gans N (2009) Economic analysis of simulation selection problems. *Management Sci.* 55(3):421–437.

Chick SE, Inoue K (2001) New procedures for identifying the best simulated system using common random numbers. *Management Sci.* 47(8):1133–1149.

Chick SE, Forster M, Pertile P (2017) A Bayesian decision-theoretic model of sequential experimentation with delayed response. *J. Roy. Statist. Soc. B* 79(5):1439–1462.

Chick SE, Gans N, Yapar O (2019) Sequential, value-based designs for certain clinical trials with multiple arms having correlated rewards. Mustafee N, Bae K-HG, Lazarova-Molnar S, Rabe M, Szabo C, Haas P, Son Y-J, eds. *Proc. 2019 Winter Simulation Conf. (WSC)* (IEEE, Piscataway, NJ), 1032–1043.

Chow SC (2014) Adaptive clinical trial design. *Annual Rev. Medicine* 65:405–415.

Claxton K, Posnett J (1996) An economic approach to clinical trial design and research priority-setting. *Health Econom.* 5(6):513–524.

DeGroot MH (2004) *Optimal Statistical Decisions* (John Wiley & Sons, Hoboken, NJ).

DiMasi J, Grabowski H, Hansen R (2016) Innovation in the pharmaceutical industry: New estimates of R&D costs. *J. Health Econom.* 47:20–33.

Dimmitt S, Stampfer H, Martin JH (2017) When less is more-efficacy with less toxicity at the ED50. *British J. Clinical Pharmacology* 83(7):1365–1368.

Draper D (2013) Discussion on 'Group sequential tests for delayed responses' (by L. Hampson and C. Jennison). *J. Roy. Statist. Soc. B* 75(1):48.

Ellenberg SS, Ellenberg JH (2017) Proceedings of the University of Pennsylvania ninth annual conference on statistical issues in clinical trials: Where are we with adaptive clinical trial designs? *Clinical Oncology* 14(5):415–416.

EMA (2017) Paediatric Gaucher disease: A strategic collaborative approach from EMA and FDA. European Medicines Agency Working Paper No. EMA/237265/2017, European Medicines Agency, Amsterdam.

EU (2014) Clinical trial regulation EU no. 536/2014. European Union, Amsterdam. Accessed February 7, 2021, https://www.ema.europa.eu/en/human-regulatory/research-development/clinical-trials/clinical-trial-regulation.

FDA (2016) Adaptive designs for medical device clinical studies. US Food and Drug Administration, Rockville, MD. Accessed May 7, 2018, https://www.fda.gov/media/92671/download.

Fenwick E, Steuten L, Knies S, Ghabri S, Basu A, Murray JF, Koffijberg HE, Strong M, Sanders Schmidler GD, Rothery C (2020) Value of information analysis for research decisions-an introduction: Report 1 of the ISPOR value of information analysis emerging good practices task force. *Value Health* 23(2):139–150.

Flight L, Julious SA, Goodacre S (2017) Can emergency medicine research benefit from adaptive design clinical trials? *Emergency Medicine J.* 34(4):243–248.

Flight L, Arshad F, Barnsley R, Patel K, Julious S, Brennan A, Todd S (2019) A review of clinical trials with an adaptive design and health economic analysis. *Value Health* 22(4):391–398.

Forster M, Brealey S, Chick S, Keding A, Corbacho B, Alban A, Rangan A, (2021) A Bayesian decision-theoretic model of a sequential clinical trial: Application to the ProFHER pragmatic trial. Clinical Trials: *J. Soc. Clinical Trials.* https://doi.org/10.1177/17407745211032909.

Frazier PI, Powell W, Dayanik S (2009) The knowledge-gradient policy for correlated normal beliefs. *INFORMS J. Comput.* 21(4): 599–613.

Fu MC, Hu JQ, Chen CH, Xiong X (2007) Simulation allocation for determining the best design in the presence of correlated sampling. *INFORMS J. Comput.* 19(1):101–111.

Gadagkar SR, Call GB (2015) Computational tools for fitting the hill equation to dose-response curves. *J. Pharmacological Toxicological Methods* 71:68–76.

Huang JH, Su QM, Yang J, Lv YH, He YC, Chen JC, Xu L, Wang K, Zheng QS (2015) Sample sizes in dosage investigational clinical trials: A systematic evaluation. *Drug Design Development Therapy* 9:305–312.

Hudson KL, Lauer MS, Collins FS (2016) Toward a new era of trust and transparency in clinical trials. *JAMA* 316(13):1353–1354.

Jaki T, Hampson LV (2016) Designing multi-arm multi-stage clinical trials using a risk-benefit criterion for treatment selection. *Statist. Medicine* 35(4):522–533.

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

Lewis RJ, Lipsky AM, Berry DA (2007) Bayesian decision-theoretic group sequential clinical trial design based on a quadratic loss function: A frequentist evaluation. *Clinical Trials* 4(1):5–14.

Li F, Li F (2019) Propensity score weighting for causal inference with multiple treatments. *Ann. Appl. Statist.* 13(4):2389–2415.

Lipsky A, Greenland S (2011) Confounding due to changing background risk in adaptively randomized trials. *Clinical Trials* 8(4): 390–397.

Lock S (2019) Matching clinical trials with unmet clinical need. *BioScience Today* 16(Spring):24–25.

Magaret A, Angus DC, Adhikari NK, Banura P, Kissoon N, Lawler JV, Jacob ST (2016) Design of a multi-arm randomized clinical trial with no control arm. *Contemporary Clinical Trials* 46:12–17.

Meltzer DO, Smith PC (2011) Theoretical issues relevant to the economic evaluation of health technologies. Pauly MV, McGuire TG, Barros PP, eds. *Handbook of Health Economics*, vol. 2 (Elsevier, New York), 433–469.

NHS England (2017) *Twelve Actions to Support and Apply Research in the NHS* (NIHR, London). https://www.england.nhs.uk/publication/12-actions-to-support-and-apply-research-in-the-nhs/.

NICE (2014) *Developing NICE Guidelines: The Manual* (UK National Institute for Health and Care Excellence, London). https://www.nice.org.uk/process/pmg20/.

NIHR (2018) *NIHR Efficacy and Mechanism Evaluation* (UK National Institute for Health Research, London). https://www.nihr.ac.uk/explore-nihr/funding-programmes/efficacy-and-mechanism-evaluation.htm.

NIHR (2020) *Annual Efficient Studies Funding Calls for CTU Projects* (UK National Institute for Health Research, London). https://www.nihr.ac.uk/documents/ad-hoc-funding-calls-for-ctu-projects/20141.

Nixon R, O'Hagan A, Oakley J, Madan J, Stevens JW, Bansback N, Brennan A (2009) The rheumatoid arthritis drug development model: A case study in Bayesian clinical trial simulation. *Pharmaceutical Statist.* 8(4):371–389.

Pallmann P, Bedding AW, Choodari-Oskooei B, Dimairo M, Flight L, Hampson LV, Holmes J, et al. (2018) Adaptive designs in clinical trials: Why use them, and how to run and report them. *BMC Medicine* 16(1):29.

Pertile P, Forster M, Torre DL (2014) Optimal Bayesian sequential sampling rules for the economic evaluation of health technologies. *J. Roy. Statist. Soc. Series A* 177(2):419–438.

Piantadosi S (1997) *Clinical Trials: A Methodologic Perspective* (John Wiley & Sons, Hoboken, NJ).

Powell WB, Ryzhov IO (2012) *Optimal Learning* (John Wiley & Sons, Hoboken, NJ).

Rasmussen CE, Williams CK (2006) *Gaussian Processes for Machine Learning* (MIT Press, Cambridge, MA).

Rosenbaum PR, Rubin DR (1983) The central role of the propsensity score in observational studies for causal effects. *Biometrika* 70(1): 41–55.

Russo D (2020) Simple Bayesian algorithms for best arm identification. *Oper. Res.* 68(6):1625–1647.

Smith AL, Villar SS (2018) Bayesian adaptive bandit-based designs using the Gittins index for multi-armed trials with normally distributed endpoints. *J. Appl. Statist.* 45(6):1052–1076.

Sydes MR, Parmar MKB, Mason MD, Clarke NW, Amos C, Anderson J, de Bono J, et al. (2012) Flexible trial design in practice–stopping arms for lack-of-benefit and adding research arms mid-trial in STAMPEDE: A multi-arm multi-stage randomized controlled trial. *Trials* 13(1):168.

Villar SS, Rosenberger WF (2018) Covariate-adjusted response-adaptive randomization for multi-arm clinical trials using a modified forward looking Gittins index rule. *Biometrics* 74(1): 49–57.

Villar SS, Bowden J, Wason J (2018) Response-adaptive designs for binary responses: How to offer patient benefit while being robust to time trends? *Pharmaceutical Statist.* 17(2):182–197.

Wason J, Jaki T (2012) Optimal design of multi-arm multi-stage trials. *Statist. Medicine* 31(30):4269–4279.

Wheeler G, Mander A, Bedding A, Brock K, Cornelius V, Grieve AP, Jaki T, et al. (2019) How to design a dose-finding study using the continual reassessment method. *BMC Medicine Res. Methodology* 19(1):18.

Williamson SF, Villar SS (2020) A response-adaptive randomization procedure for multi-armed clinical trials with normally distributed outcomes. *Biometrics* 76(1):197–209.

Williamson SF, Jacko P, Villar SS, Jaki T (2017) A Bayesian adaptive design for clinical trials in rare diseases. *Comput. Statist. Data Anal.* 113:136–153.

Wu J, Frazier P (2016) The parallel knowledge gradient method for batch Bayesian optimization. *Adv. Neural Inform. Processing Systems* 29:3126–3134.

Xie J, Frazier PI, Chick SE (2016) Bayesian optimization via simulation with pairwise sampling and correlated prior beliefs. *Oper. Res.* 64(2):542–559.