Service Science

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

Optimal Stopping of Adaptive Dose-Finding Trials

Amir Ali Nasrollahzadeh, Amin Khademi

To cite this article:
Amir Ali Nasrollahzadeh, Amin Khademi (2020) Optimal Stopping of Adaptive Dose-Finding Trials. Service Science 12(2-3):80-99. https://doi.org/10.1287/serv.2020.0261

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

SERVICE SCIENCE
Vol. 12, No. 2-3, June–September 2020, pp. 80–99
ISSN 2164-3962 (print), ISSN 2164-3970 (online)

Optimal Stopping of Adaptive Dose-Finding Trials

Amir Ali Nasrollahzadeh,[a] **Amin Khademi**[b]

[a] Department of Mechanical and Aerospace Engineering, University of California, Davis, Davis, California 95616; [b] Department of Industrial Engineering, Clemson University, Clemson, South Carolina 29634

Contact: anasr@ucdavis.edu, https://orcid.org/0000-0002-8023-4629 (AAN); khademi@clemson.edu, https://orcid.org/0000-0002-5281-8715 (AK)

Received: June 18, 2019
Revised: March 31, 2020; August 10, 2020
Accepted: August 28, 2020
Published Online in Articles in Advance: October 7, 2020

https://doi.org/10.1287/serv.2020.0261

Abstract. The primary objective of this paper is to develop computationally efficient methods for optimal stopping of an adaptive Phase II dose-finding clinical trial, where the decision maker may terminate the trial for efficacy or abandon it as a result of futility. We develop two solution methods and compare them in terms of computational time and several performance metrics such as the probability of correct stopping decision. One proposed method is an application of the one-step look-ahead policy to this problem. The second proposal builds a diffusion approximation to the state variable in the continuous regime and approximates the trial's stopping time by optimal stopping of a diffusion process. The secondary objective of the paper is to compare these methods on different dose-response curves, particularly when the true dose-response curve has no significant advantage over a placebo. Our results, which include a real clinical trial case study, show that look-ahead policies perform poorly in terms of the probability of correct decision in this setting, whereas our diffusion approximation method provides robust solutions.

Funding: This research is supported by the National Science Foundation, Division of Civil, Mechanical and Manufacturing Innovation [Award 1651912].
Supplemental Material: The online supplement is available at https://doi.org/10.1287/serv.2020.0261.

Keywords: optimal stopping • adaptive clinical trials • dynamic learning • knowledge gradient • diffusion approximation

1. Introduction

The cost of inventing, developing, and introducing a new drug to market has surpassed $2.6 billion (Tufts Center for the Study of Drug Development 2014). The biggest drivers of this high cost are clinical trials, the cost of which depends on several factors such as the number of participants, locations of research facilities, and complexity of the trial protocol (Roy 2012). In fact, the total cost can reach $300–$600 million for large clinical trials (Griffin et al. 2010). The main goal of dose-finding clinical trials is to identify a "target dose," which is used in later stages with more patients to confirm its effects and is considered a critical step in the drug development process (Bornkamp et al. 2007). This is because a poor selection of the target dose may cause the Food and Drug Administration (FDA) to disapprove Phase III, the next phase, which is the most costly phase in drug development, as a result of insignificant positive evidence (futility; see Snapinn et al. (2006)) or exposure to unnecessary risk (adverse effects). In fact, during 2000–2012, failure to select optimal drug doses was a leading factor for the delay or denial of drug submissions in the first submission round by the FDA (Sacks et al. 2014). In particular, several studies have shown that more than 50% of Phase III clinical trials fail to establish the efficacious benefits of the treatment (see, e.g., Grignolo and Pretorius (2016)). This high attrition rate is of particular concern because it demonstrates that more than half of the trials that clear Phase II by indicating a positive evidence of efficacy fail to produce similar outcomes in Phase III. Grignolo and Pretorius (2016) suggested that inadequate Phase II studies and suboptimal dose selection policies are among the major factors contributing to this high failure rate. Considering the high attrition rate and costs of Phase III clinical trials motivates a new approach that is able to detect futility in earlier stages (Phase II) and provide a reliable assurance for efficacious results. Early futility detection in Phase II reduces the cost and risk of exposure to futile treatments, whereas an assurance for efficacious results may prevent the high attrition rate in Phase III. Therefore, formulating an optimal stopping problem capable of making optimal stopping decisions for efficacy or futility while adapting to different treatments and patient assignment procedures has the potential to impact the drug development process in practice. Compared with classic static designs of clinical trials, adaptive designs generally can reduce the cost of conducting a clinical trial because of their ability to modify the design while the trial is still in progress. This enables an adaptive clinical trial to abandon an ineffective treatment early in the process (Berry et al. 2002), and thus the optimal stopping of adaptive clinical trials is naturally motivated.

We formulate the optimal stopping of an adaptive dose-finding Phase II trial as a finite-horizon stochastic dynamic program (DP), where at each intermediate decision epoch, the decision maker (DM) may abandon the trial as a result of futility, continue the trial to collect more evidence about the dose-response curve, or terminate the trial for efficacy and move to the confirmatory phase. The dose-finding trial is assumed to investigate the efficacy of multiple doses (≥ 2) with normally distributed patient responses with an unknown mean and a known observation variance. Considering that the main goal of a Phase II trial is to identify a target dose for which the efficacy should be tested versus a standard treatment or placebo in a large patient population in Phase III, a key feature in our formulation is the incorporation of an approximation of the probability of success at the end of a confirmatory phase (Phase III) when deciding to terminate the trial. Before further discussion on this key feature, note that we decouple the dose allocation procedure (i.e., select the assignment dose for the next patient when the decision is to continue the trial) from the stopping problem and fix it to some given policy. This choice will potentially result in easier implementation in practice because most clinical trials are still administered with a balanced randomized allocation policy (similar to the setting studied in Chick et al. (2017) and our case study, Hall et al. (2011)). In fact, we include a randomized allocation policy in addition to other adaptive benchmarks in our sensitivity analyses to demonstrate the performance of the proposed stopping rules with respect to different dose allocation policies.

In order to approximate the probability of success in Phase III, it is natural for the DM to consider the power of the hypothesis test $H_0 : df^* \leq 0$ versus $H_1 : df^* > 0$, where df^* denotes the expected response improvement over a placebo or standard treatment (assuming higher responses are favorable) (Müller et al. 2006). By approximating the probability of success in the next phase, and considering the cost of sampling patients and monetary benefits of identifying a significant improvement over the placebo, a utility function is constructed capable of incorporating costs and benefits as well as the quality of the stopping decisions, thus fulfilling our motivation. Moreover, adaptive designs of clinical trials may address some ethical issues. An adaptive approach considering early abandonment is ethically motivated because it may prevent the DM from assigning patients to ineffective or toxic doses in early stages. Also, one can argue that considering a term for approximating the probability of success and ensuring a level of certainty for treatments that clear Phase II to be tested for a larger population in Phase III is also ethically motivated. However, ethical debates on the design of clinical trials are not settled; see, for example, Berry et al. (2004) and Bothwell and Kesselheim (2017) for an extensive literature that spans a couple of decades.

Two main challenges are involved in the definition of df^* in a Bayesian setting: (i) the target dose is a random variable at the beginning of each decision epoch given the history of the states, actions, and observations; and (ii) the expected response of any dose (including the target dose) is also a random variable at the beginning of each decision period given the said history. These conditions introduce difficulties in evaluating the distribution of df^*. Addressing such challenges requires a proper dose-response model and a stochastic DP setup, which are discussed in Sections 3 and 4, respectively. The resulting stochastic DP formulation, however, suffers heavily from the curse of dimensionality because the state space is multidimensional and unbounded.

For this problem, Brockwell and Kadane (2003) proposed an approximation procedure, which was partially applied to the optimal stopping of a fully Bayesian dose-finding trial (Berry et al. 2002). The approximation is based on discretizing the state space over a grid, using forward simulation until the last decision epoch to create sample paths, and utilizing backward induction to estimate the value function in each cell of the grid at each time period. This method is computationally extremely time consuming, and Berry et al. (2002) stated that applying this method at each decision epoch in a fully adaptive design is “impractical.” Moreover, Grieve and Krams (2005) noted computational difficulties of this method as the main reason for considering posterior estimates instead of forward simulation and backward induction as a basis for the stopping rule in the Acute Stroke Therapy by Inhibition of Neutrophils (ASTIN) trial. Therefore, the main goals of this study are (i) developing computationally efficient algorithms for the optimal stopping problem of a dose-finding trial and (ii) testing the performance of the proposed policies with available benchmarks via simulation in settings where the correct stopping decision is abandonment or termination.

We propose two solution methods for this problem. The first one adapts the one-step look-ahead framework, in which the DM assumes that the next decision epoch is the last one. This approach is computationally much less demanding than the benchmark method because it only requires one-step forward simulations. However, the induced stopping time by this method may happen earlier than the optimal stopping time (Proposition 1). In the second proposed method, we consider a two-armed bandit version of the problem with one unknown arm (the target dose) and one known arm (placebo), where the posterior of df^* (i.e., the advantage of the target dose over placebo) is normally distributed as a result of our normality assumption on patients' responses. Therefore, in a continuous sampling regime, a scaled mean of df^* follows an Itô process, which enables us to formulate a continuous-time Bellman equation for the continuous-time optimal stopping counterpart. By using Itô's lemma, we show that the optimal value function to the continuous-time Bellman equation satisfies a partial differential diffusion-advection equation with boundary conditions (Proposition 2). In addition, the solution to the partial differential equation depends only on the utility function (objective function of the DM), which can be found up front, and identifies a continuation region over the mean response (vertical axis) and time (horizontal axis), which is easy to understand and implement. We show how the two-armed bandit results can be used to make effective decisions in the setting of multiple doses and adopt the method for the original stopping problem where the patients arrive in discrete decision epochs. This method is also computationally appealing because it bypasses forward simulations, which are computationally time consuming to find the decision regions.

In summary, our formulation is a methodological exercise in exploring the potential advantages of the diffusion approximation method in correctly terminating, abandoning, or continuing a clinical trial. We test the performance of the two proposed methods along with the benchmark developed by Berry et al. (2002) via simulation. We also conduct a case study where data from a real clinical trial (Hall et al. 2011) are used to investigate the performance of the proposed solutions in a real-world setting. In addition to the monetary value of a stopping decision, which is the primary objective function, we report the probability of correct decision at the stopping time for each method. We test the results on two settings: one where there is a significant difference between the average response of the target dose and placebo (the ultimate decision is termination) and one where the said difference is negligible (the ultimate decision is abandonment). Our simulation results shed light on the behavior and performance of each method and reveal an important insight on the performance of these methods for implementation in practice. Although the adaptive optimal stopping problem is motivated because classic solutions require complex simulations and extensive computational efforts, our results show that there is a key distinction in performance with respect to treatments that do not produce significant advantageous responses over the placebo. In fact, we show that the diffusion approximation is particularly effective in correctly abandoning ineffective treatments, whereas the one-step look-ahead policy and the simulation-based gridding fail to correctly abandon the trial in our setting. This result is also emphasized in our case study where the original trial did not conclude with a clear termination or abandonment result and suggested further investigation to establish the efficacy of the underlying treatment. In our retrospective analysis of the trial, the diffusion approximation method also decided to continue the trial because positive evidence suggested efficacy, but they were not significant enough to justify a termination decision. Because abandoning the trial as a result of futility of the treatment is an important factor in reducing the cost and health risks associated with clinical trials, our results are of important potential practical value in designing adaptive Phase II clinical trials.

Appendices in the online supplement include detailed discussions of the dose-response approximation model, proofs, details of the proposed algorithms, additional numerical results, and sensitivity analyses.

2. Related Works

Optimal stopping is an important decision making problem and is studied in different communities because of its vast applications. The optimal stopping of a clinical trial has also received significant attention because of its importance. For an overview of advancements in optimal stopping of clinical trials, see a survey by Hee et al. (2016) from a Bayesian perspective and Jennison and Turnbull (1999) from a frequentist perspective, as well as references therein. In addition, Jitlal et al. (2012) and Deichmann et al. (2016) reviewed hundreds of clinical trials considering early stopping rules as a result of futility, effectiveness, or safety concerns across different phases. Stallard et al. (2001) reviewed different stopping rules specifically for Phase II clinical trials.

Here, we only focus on Bayesian decision-theoretic methods developed for optimal stopping of dose-finding trials. One main method used in Bayesian decision theoretic designs is based on forward simulation of the trial to the end and using backward induction to estimate the value function over a grid to ultimately evaluate the stopping region. Details of this methodology are presented in Brockwell and Kadane (2003), and it is adapted in Berry et al. (2002) for a fully adaptive trial. However, this method is computationally extremely demanding, and we propose two more efficient algorithms to solve this problem. Our simulation results compare the performance of the proposed solutions and the method developed by Berry et al. (2002) with respect to estimated and true utility, stopping time, and the probability of correctly deciding whether to terminate or abandon the trial.

Our first proposal is to adapt the one-step look-ahead approach to the optimal stopping of dose-finding trials. This approach is used for sequential sampling in Bayesian settings, and Frazier and Powell (2008) applied this method for the optimal stopping of a ranking and selection problem. However, the optimal stopping of a dose-finding trial is different from a ranking and selection setup because the DM has to consider the effects of the termination decision on the next confirmatory phase of the drug development process. This is accommodated for by a hypothesis test in which the significance of the advantage of the target dose over placebo, calculated by subtracting the placebo response from that of the (random) target dose, is tested. Rojas-Cordova and Bish (2018) analyzed an adaptive sequential allocation and termination of the trial and investigated the trade-off between establishing the efficacy early on and more sampling to increase the accuracy of the estimation. However, their setup is different in objective and is focused on Phase III trials with binary responses.

Our second approach is inspired by a work presented in Chernoff (1961), where a diffusion approximation is used to test whether the mean of a normal distribution is positive. Such a method is used in the optimal stopping of a Phase III clinical trial (Chick et al. 2017) and in the optimal stopping of a clinical trial with correlated treatments (Chick et al. 2018). However, the structure of our problem is different from previous studies because the DM has to consider the power of a hypothesis test. Moreover, the heuristic to extend the diffusion results to multiple doses settings is particularly tailored to our setting.

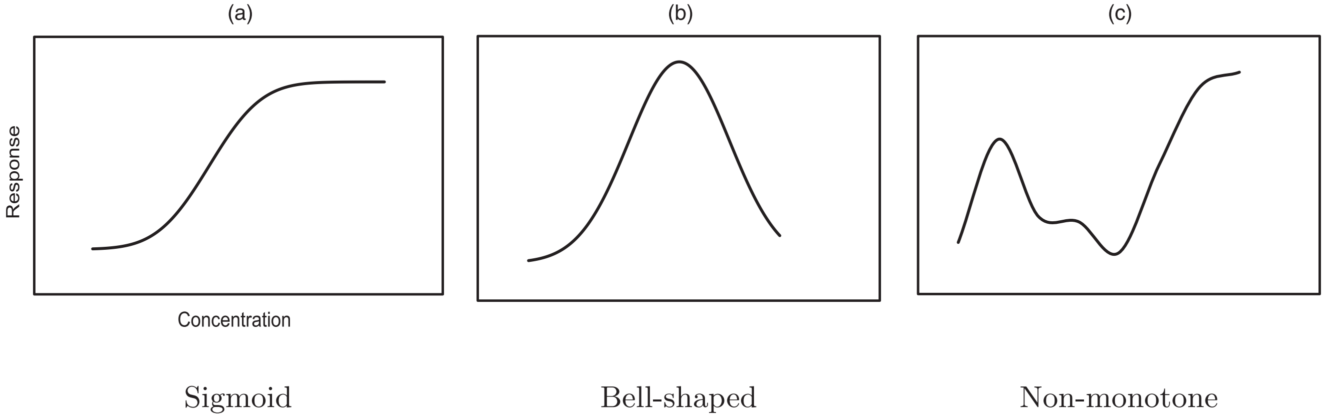
3. Dose-Response Model

The relationship between the treatment dose of a drug and its induced response (e.g., change in a measurable medical outcome) is essential in dose-finding studies and is usually described by a curve or function referred to as a dose-response curve. For example, Figure 1 presents three typical dose-response curves, where the sigmoid shape in Figure 1(a) is one of the most recurring dose-response relationships in theory and practice (Gadagkar and Call 2015).

There are two main classic approaches to estimate the dose-response curve. The first approach considers a functional form up front (parametric), dictating the shape of the underlying dose-response curve, and seeks to estimate the parameters by adapting the model to available observations (e.g., Kotas and Ghate 2018). Another approach is to use a piecewise linear approximation (nonparametric) to the curve over a discrete set of available doses (Berry et al. 2002). Here, we present a first-order Bayesian nonparametric piecewise linear approximation to the curve, and we refer the reader to Online Supplement Section 1 for further discussion on the choice of the dose-response model, the pros and cons of the parametric and nonparametric models with respect to misspecification error, and their computational and potential practical implications.

Denote the numerical score of a patient's response by y and the prescribed dose by $z \in \mathcal{Z}$, where $\mathcal{Z} := \{Z_j : j = 1, \ldots, J\}$ is the set of all admissible doses. Let $\Theta = (\theta_1, \ldots, \theta_J)'$ denote a J-sized column vector of unknown parameters, where θ_j refers to the mean response to dose Z_j (index j is used in lieu of Z_j throughout this work). In particular, we assume that at any given dose j, the response is given by $y_j = \theta_j + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \sigma^2)$ (see, e.g., Berry et al. (2002)), and σ^2 refers to the uncertainty in observing a patient's response. By this construction, the response of patients at dose j is normally distributed with an unknown mean θ_j and a known variance σ^2. We can interpret that such construction approximates the dose-response curve by fitting a piecewise linear function connecting consecutive θ_js. We consider a Bayesian setup where the DM has a multivariate normal belief about Θ and updates her belief by each patient's response.

Figure 1. Typical Dose-Response Curves



Although some early Phase II studies investigate binary or categorical responses to describe the success or failure of a treatment, finding the most promising dose usually involves a continuous response (Berry et al. 2010). In fact, there are cases when categorical responses are produced by considering a priori cutoff thresholds for underlying continuous responses (European Network for Health Technology Assessment 2013). Patients' responses such as improvement in life-years, progression free survival, blood pressure, temperature, weight, Hamilton depression score, time to event (e.g., time to heal), and various scoring systems are some examples of continuous responses (see, e.g., Biswas and Bhattacharya (2016)). In constructing this dose-response model, we assume that the error is normally distributed with a known variance. This is a standard assumption, as the primary endpoint (response) in the majority of trials with continuous response is assumed to follow a normal distribution (Julious 2004). See, for example, Krams et al. (2003) for implementation of such a setup in practice. In addition, our case study also considers a continuous outcome, as it reports the probability of achieving a five-year quality-adjusted life-years (QALYs) for standard and treatment arms. Furthermore, considering a known variance is usually justified for Phase II because an estimate of the variance is already available from Phase I results (Spiegelhalter et al. 2004).

3.1. Target Dose

The ultimate goal of a Phase II study is to identify a target dose. We focus on the efficacy of the target dose in Phase II of dose-finding clinical trials in order to define a utility function for the stopping decision, and thus ED_{95}, the smallest dose achieving 95% of the maximal response, is considered as the target dose. We formally define ED_{95} as

$$\text{ED}_{95} := \min_{z}\{z \in \mathcal{Z} : f(z, \Theta) \geq 0.95 f(z_{\max}, \Theta)\}, \quad (1)$$

where $z_{\max}$ denotes the dose with maximal response. The target dose is different from the assignment dose given to the patients throughout the trial by a fixed allocation policy. The responses of patients to these assignment doses are used to update the mean response of different doses. The motivation for ED_{95} is that the highest response may correspond to high dosages (toxic doses), which may induce undesired adverse side effects.

4. Problem Formulation

In this section, we present a stochastic DP formulation for the response-adaptive optimal stopping of dose-finding Phase II clinical trials. At each decision epoch, using the information accrued so far, a DM decides whether to (i) abandon the trial because of significant evidence of the inefficacy of the treatment, (ii) continue the trial to collect more information if there is insignificant evidence of the efficacy of the treatment with an expectation of improvement, or (iii) terminate the trial for efficacy and move to a confirmatory study when efficacy is verified by testing the treatment for a large population.

The optimal stopping problem is accompanied by an allocation procedure that upon continuation decision identifies the next dose assignment. In this work, we assume that the allocation decision is made according to some predetermined rule. In particular, we use a one-step look-ahead policy introduced in Nasrollahzadeh and Khademi (2018) for dose assignment. We test the effects of other dose assignment procedures on the performance of our proposed solutions in Online Supplement Section 4.6. The objective of the optimal stopping problem is expressed in terms of monetary values. This objective, which is known as net present value, is appropriate in stopping problems where sampling costs/rewards are financial measures (Brealey et al. 2012).

Recall that Θ represents a vector of unknown expected responses corresponding to doses in set $\mathcal{Z}$ where the resulting dose-response function is approximated by connecting consecutive θ_j's. Let n denote decision epochs, let N be the total number of (potential) homogeneous patients in the trial, and let $y^{n+1} = \theta_{z^n} + \epsilon^{n+1}$ be the observed response of patient $n+1$ after assignment to dose z^n, where $(y^{n+1}|\Theta, z^n) \sim \mathcal{N}(\theta_{z^n}, \sigma^2)$. We assume that the response of a patient is observed before the next decision epoch. Define $\mathscr{F}^n$ as the σ-algebra generated by $z^0, y^1, z^1, y^2, \ldots, z^{n-1}, y^n$. Note that z^0 is the assignment dose before observing any response, z^n is given by the fixed allocation policy, and τ represents some stopping time at which $\mathscr{F}^\tau$ describes the accrued information gathered by sampling τ patients. We use y and $\hat{y}$ to denote true and simulated observations, respectively.

4.1. State Space

Decision epochs are set at the times when a patient becomes available. We assume a possibly correlated multivariate normal prior on our belief about Θ; that is, $\Theta \sim \mathcal{N}(\mu^0, \Sigma^0)$. Observations y form a normal likelihood distribution resulting in a Bayesian conjugate setup where posterior distributions on Θ are also multivariate normal. Define $\mu^n := \mathbb{E}[\Theta|\mathscr{F}^n]$ and $\Sigma^n := \text{Cov}[\Theta|\mathscr{F}^n]$ as posterior moments of the belief about Θ. At decision epoch n, the DM decides on abandoning, continuing, or terminating only based on the current estimate of the dose-response curve, which is summarized by the posterior distribution on parameter Θ given historical information $\mathscr{F}^n$ (i.e., $\mathbb{P}(\Theta|\mathscr{F}^n)$). This posterior can be completely described by the state variable $s^n = (\mu^n, \Sigma^n)$. Thus, the state space $\mathcal{S}$ is defined as

$$s^n \in \mathcal{S} := \left\{ (\mu, \Sigma) : \mu \in \mathbb{R}^J, \Sigma \in \mathbb{S}_+^J \right\} \cup \nabla,$$

where $\mathbb{S}_+^J$ denotes the set of $J \times J$ positive semidefinite matrices, and ∇ denotes an absorbing state showing the end of the decision-making process.

4.2. Action Space

At each decision epoch, if enough evidence (in the form of current estimate of the dose-response) has emerged to suggest that an effective target dose is identified, and sampling more patients will not improve the estimate by a significant margin considering the cost of sampling, the DM may decide to "terminate" the trial and switch to a confirmatory phase where the target dose is further tested to confirm its effectiveness. On the contrary, the DM might learn that the current estimate of the dose-response curve shows no signs of effectiveness (e.g., a flat dose-response curve), and sampling more patients will only increase trial costs, and thus the DM may "abandon" the trial. However, if the current estimate of the dose-response curve suggests that an effective target dose may be identified, and continuing the trial with more sampling potentially may lead to a significant improvement of the estimate and utility, then the DM may "continue" the trial by allocating a dose to the next patient, observe the response, and update the current estimate of the dose-response curve. Recall that the allocation scheme is assumed to be given and independent of the optimal stopping problem. Thus, define $a^n(s) \in \{0, 1, 2\}$ as the decision variable when in state s, where 0 shows that the decision is to abandon the trial, 1 shows the continuation of the trial, and 2 shows that the trial is terminated. Thus, the action space is described by $A(s) := \{a^n(s) \in \{0, 1, 2\}, \forall n \leq N\}$, where at stopping time $n = \tau$, or $n = N$, $a^n \in \{0, 2\}$. For $s = \nabla$, set $A(s) := \emptyset$.

4.3. Transitions

Terminating or abandoning the trial at decision epoch n determines the stopping time as $\tau = n$, and the system transits to state ∇, where no more sampling is allowed and the current estimate of the dose-response curve remains unchanged. However, if the decision is to continue the trial, a dose is selected according to an allocation policy, and its observed response will be used in order to update the current estimate of the dose-response curve (i.e., transit to a new state). The new state $s^{n+1} = (\mu^{n+1}, \Sigma^{n+1})$ is described by

$$\begin{aligned} \mu^{n+1} &= \mu^n + \tilde{\sigma}(\Sigma^n, j) X^{n+1}, \\ \Sigma^{n+1} &= \Sigma^n - \tilde{\sigma}(\Sigma^n, j) \tilde{\sigma}'(\Sigma^n, j), \end{aligned} \tag{2}$$

where j denotes the allocated dose, $\tilde{\sigma}(\Sigma^n, j) := \frac{\Sigma^n e_j}{\sqrt{(\sigma^2 + \Sigma_{jj}^n)}}$, e_j is a J-vector of 0′s and a single 1 at the jth index, and $X^{n+1} := \frac{y^{n+1} - \mu^n}{\sqrt{(\sigma^2 + \Sigma_{jj}^n)}}$ follows a standard normal distribution conditioned on $\mathscr{F}^n$.

4.4. Objective Function

We consider maximizing a monetary equivalent of benefits acquired as a result of early termination or abandonment of the trial versus costs incurred by continuing the trial with more sampling. If the decision is to abandon the trial (i.e., $a^n = 0$), then no immediate reward or cost is incurred. If the decision is to continue the trial (i.e., $a^n = 1$), then only a sampling cost $c_1 > 0$ is paid. When the decision is to terminate the trial (i.e., $a^n = 2$), the immediate reward consists of the monetary value of the advantage over a placebo, if such an advantage is significant, minus the setup/sampling cost in the confirmatory phase. Define utility function $u(a^n, s^n, \mathscr{F}^n)$ as the expected immediate benefit (reward − cost) incurred when deciding on action a^n in state s^n given information $\mathscr{F}^n$ by

$$u(a^n, s^n, \mathscr{F}^n) = \begin{cases} 0 & \text{if } a^n = 0, \\ -c_1 & \text{if } a^n = 1, \\ -c_1' n_p + c_2 m_n \mathbb{E}\left[\mathbb{1}_{\{B^n\}} | \mathscr{F}^n\right] & \text{if } a^n = 2, \end{cases} \tag{3}$$

where $c_1' n_p$ is the cost of sampling n_p patients in the confirmatory phase ($c_1' > 0$), and $c_2 > 0$ is the payoff per unit advantage of the target dose over the placebo. Considering economic health outcomes as primary or secondary objective is well established in the literature and practice. For example, refer to Flight et al. (2019) for a review of clinical trials with economic health outcomes. A typical outcome in clinical trials is QALYs gained, which may have a corresponding monetary value that can be used to estimate the benefit per unit improvement over the placebo (see, e.g., Hall et al. (2011)). Let $m_n = \mathbb{E}[df^* | \mathscr{F}^n]$ denote the expected advantage over the placebo where $df^* = \theta_{z^*} - \theta_0$, let z^* be the (random) target dose, and let θ_0 be the known and fixed response of the placebo. Because z^* is random with respect to $\mathscr{F}^n$, θ_{z^*} denotes the posterior expected response at dose z^*, and thus, df^* identifies the posterior advantage over the placebo. Furthermore, the indicator function $\mathbb{1}_{\{B^n\}}$ determines the significance of the advantage over the placebo by considering the event B^n in which the null hypothesis is rejected when comparing $H_0 : df^* \leq 0$ with $H_1 : df^* > 0$. In particular,

$$B^n := \left\{ \frac{\sqrt{n_p}(\bar{y}_* - \bar{y}_0)}{\sqrt{2\sigma^2}} > q_\alpha \right\}, \tag{4}$$

where $\bar{y}_*$ and $\bar{y}_0$ denote n_p-sample average responses of the estimated target dose and placebo at decision epoch n, respectively; and q_α denotes the $(1-\alpha)$ quantile of normal distribution with α being the significance level for the hypothesis. Hypothesis tests do not directly transfer to Bayesian settings. However, an approximation may be developed to evaluate a "Bayesian significance" where we wish to evaluate the predictive probability of obtaining significant results when testing the null hypothesis versus an alternative hypothesis; see Spiegelhalter et al. (2004), chapter 6.5.3. A similar setup is also implemented by Müller et al. (2006), and we use it in our setting as well. Moreover, Lewis et al. (2013) and Liu et al. (2017) are examples of real clinical trials that implement similar frequentist setups to determine stopping times in Bayesian frameworks.

The expectation $\mathbb{E}[\mathbb{1}_{\{B^n\}}|\mathscr{F}^n]$ can be estimated with an arbitrary accuracy by Monte Carlo as follows: create a sample from Θ and calculate the target dose, z^*, for the said sample by Equation (1); create n_p samples from $\mathcal{N}(\theta_{z^*},\sigma^2)$ and $\mathcal{N}(\theta_0,\sigma^2)$; calculate $\bar{y}_*$ and $\bar{y}_0$ and identify whether the event B^n occurs; and continue this process for enough samples and take a sample average to estimate $\mathbb{E}[\mathbb{1}_{\{B^n\}}|\mathscr{F}^n]$. The utility function defined here is tailored to Phase II clinical trials and is designed to capture the important trade-offs in this decision-making process. For example, if the trial stops early because of futility, the treatment has no merit, which is reflected in its reward of 0. The cost of sampling patients up to that point is captured by the cost associated with the previous continuation decisions. If positive evidence for the advantage over the placebo emerges, the DM has to make sure that the evidence is significant and then consider the cost of setting up a confirmatory phase in addition to the cost of sampling patients up to that point.

Given that a decision to abandon or terminate the trial has been made at stopping time $n=\tau$, the optimal expected utility is given by

$$\begin{aligned} G(s^\tau) &= \max_{a^\tau\in\{0,2\}} u(a^\tau, s^\tau, \mathscr{F}^\tau) \\ &= \max\left\{0, -c_1' n_p + c_2 m_\tau \mathbb{E}\left[\mathbb{1}_{\{B^\tau\}}|\mathscr{F}^\tau\right]\right\}. \end{aligned} \tag{5}$$

Therefore, for every $n<\tau$, the decision has to be $a^n=1$, and a sampling cost c_1 is paid as the expected immediate utility (i.e., $u(a^n=1, s^n, \mathscr{F}^n)=-c_1$). Let $l_\pi(s^0)$ denote the expected utility at stopping time τ, given historical information $\mathscr{F}^\tau$ under policy π when the initial prior on the belief about Θ is $s^0=(\mu^0,\Sigma^0)$; that is,

$$l_\pi(s^0) = \mathbb{E}^\pi\left\{-c_1\tau + \max_{\pi(a^\tau)\in\{0,2\}} u(\pi(a^\tau), s^\tau, \mathscr{F}^\tau)\Big|s^0\right\}, \quad \forall\pi\in\Pi, \tag{6}$$

where Π is the set of all nonanticipative admissible policies, and the DM selects a policy $\pi\in\Pi$ such that $V(s^0)=\sup_{\pi\in\Pi} l_\pi(s^0)$. Therefore, the optimal value function is the solution to the following optimality equations:

$$\begin{aligned} V(s^n) &= \mathbb{E}^\tau \sup_{\tau\geq n+1}\{-c_1(\tau-n)+\mathbb{E}[V(s^\tau)|\mathscr{F}^n]\}, \\ V(s^\tau) &= G(s^\tau), \qquad \forall s\in\mathcal{S}. \end{aligned} \tag{7}$$

The state space defined on Θ is unbounded, and thus standard stochastic DP techniques are computationally intractable. We describe and implement the solution developed by Berry et al. (2002) and propose two different alternative techniques to solve the optimal stopping problem.

5. Approximate Solutions

In this section, we explain three approximate methods. Berry et al. (2002) used a simulation-based gridding algorithm discussed in Section 5.1 to evaluate stopping times. This approach is computationally extremely expensive in implementation and gives rise to "static terminators" where for a few sample dose-response curves, a large number of trials are simulated forward in time to compute their average expected utility over a discretized grid by backward induction. These approximations are used statically to evaluate stopping times for "similar" dose-response curves. Section 5.2 proposes a one-step look-ahead policy to find stopping times, and Section 5.3 proposes a diffusion approximation method, which are computationally more efficient.

5.1. Simulation-Based Gridding Approximation

Brockwell and Kadane (2003) and Müller et al. (2007) proposed an approximation method for the problem defined in Section 4 where the state space is discretized by a grid over which a sufficient number of experiments are run to estimate the final-stage value function. The key idea is that in each cell of this grid—say, cell j—the value of termination and abandonment can be evaluated easily. The value of continuation is the sample average of all the cells that are visited in the next decision epoch by the experiments currently visiting cell j. We present the details of the approach for completeness and to clarify the differences between the assumptions used in this approach with those in our setup.

To construct the grid, Berry et al. (2002) assumed a normal prior on the advantage over placebo (i.e., $df^*\sim\mathcal{N}(m_0,v_0^2)$). Let (m_n,v_n) denote the posterior mean and standard deviation of the advantage over placebo at ED$_{95}$ at time n; that is, $m_n=\mathbb{E}[df^*|\mathscr{F}^n]$ and $v_n^2=\text{Var}[df^*|\mathscr{F}^n]$. Construct a bivariate grid over possible values of m and v, carefully considering their upper and lower bounds as follows. Given the allocation scheme and thus the allocation dose z^n, simulate trials $i = 1, \dots, M$ by generating observations $\hat{y}_i^{(n+1:N)}$ and update the current estimate of Θ by calculating $\mu_i^{(n+1:N)}$ and $\Sigma_i^{(n+1:N)}$. In order to estimate m_n and ν_n^2, after the current Θ is evaluated, simulate samples from Θ, identify the target dose for each sample, and calculate the posterior mean and standard deviation of df^* through sample mean and variance estimation. Record the trajectory of each trial (i.e., the sequence of (m_n, ν_n)) over the bivariate grid for (m, ν). For example, Figure 2(a) shows 30 trial trajectories of m_n on a simplified univariate grid (only m_n versus n) for $N = 10$. It is likely that some of the grid cells remain empty; that is, no simulated trials resulted in m and ν values corresponding to that cell, which affects the approximation quality. To fix that, consider a particular (m_n, ν_n) corresponding to those cells as priors and simulate a number of trials starting from those cells. Thus, the entire grid is populated.

Remark 1. We assume a correlated multivariate normal prior on Θ; that is, $\Theta \sim \mathcal{N}(\mu_0, \Sigma_0)$. However, z^*=ED_{95} is random with respect to $\mathscr{F}^n$, and thus, θ_{z^*} is not normally distributed with respect to $\mathscr{F}^n$. In the simulation-based gridding algorithm, the actual unknown distribution of θ_{z^*} is approximated by a normal distribution in the literature. However, we do not make such an assumption in our proposed solutions.

To evaluate the optimal decision in each cell, start from the last decision epoch N, when the continuation decision is not possible, and the value function can be computed by Equation (5). Denote by A_j^n the subset of indices $i \in \{1, \dots, M\}$ whose trajectories terminate in the jth cell (which corresponds to an (m, ν) pair) in the grid (m_n, ν_n, n). For the last decision epoch N, this is demonstrated by darker trajectories that end up in a specific cell in Figure 2(a). The termination utility function in the jth cell is evaluated by taking a sample average of the value functions corresponding to trial simulations whose trajectories terminated in that grid cell; that is,

$$\hat{U}_j^N(a^N = 2) \approx \frac{1}{|A_j^N|} \sum_{i \in A_j^N} u_j^N(a^N = 2, \hat{s}_i^N), \tag{8}$$

where $\hat{U}_j^N(a^N = 2)$ is the approximated utility function at decision epoch N in the grid cell j when the decision is to terminate the trial, $|\cdot|$ denotes set cardinality, and the utility function $u_j^N(a^N, \hat{s}_i^N)$ is known for $a^N \in \{0, 2\}$ and $\hat{s}_i^N = (m_j^N, \nu_j^N)$ for all $i \in A_j^N$, where m_j^N and ν_j^N correspond to the jth cell values for m and ν, respectively. Therefore, the expected utility of termination at the last decision epoch N is given by

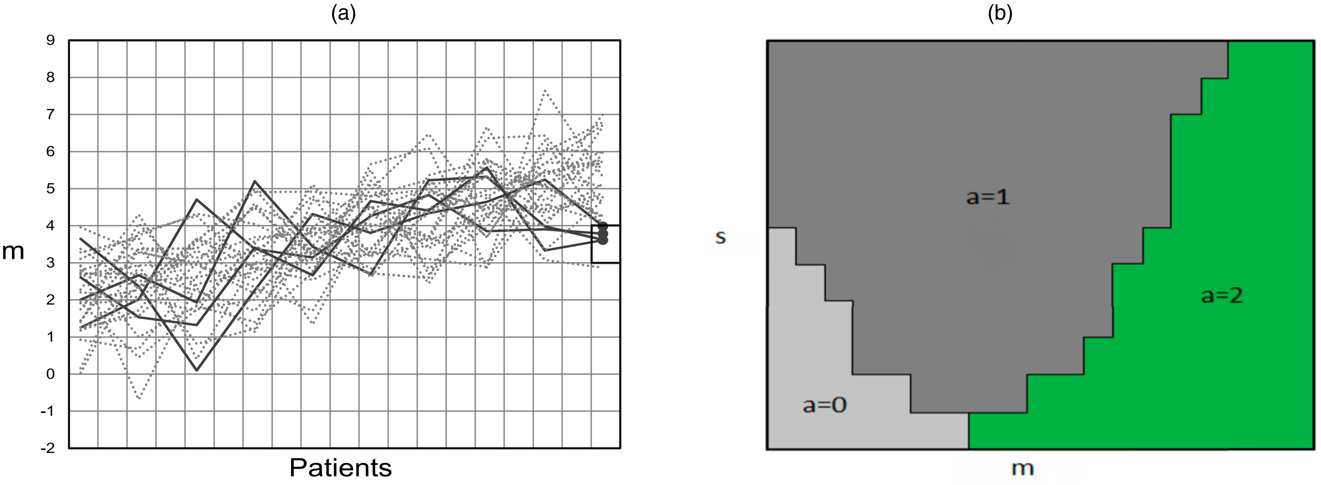
$$u_j^N(a^N = 2, \hat{s}_i^N) = -c_1' n_p + c_2 m_j^N \mathbb{E}[\mathbb{1}_{\{B^N\}} | \mathscr{F}^N]$$
$$\forall i \in A_j^N,$$

where $B^N := \{\frac{\sqrt{n_p}(\bar{y}_* - \bar{y}_0)}{\sqrt{(2\sigma^2 + (\nu_j^N)^2)}} > q_\alpha\}$, with $2\sigma^2 + (\nu_j^N)^2$ denoting the posterior predictive variance of $\bar{y}_* - \bar{y}_0$. Thus, the approximated value function in each cell of the grid at decision epoch N is

$$\hat{V}_j^{*,N} = \max\left\{0, \hat{U}_j^N(a^N = 2)\right\}, \tag{9}$$

where if $\hat{V}_j^{*,N} = 0$, the optimal decision is to abandon the trial in the jth grid cell (i.e., $a_j^{*,N} = 0$). Otherwise, the optimal decision is to terminate the trial, $a_j^{*,N} = 2$.

Figure 2. Gridding Approximation



An example of trajectories of simulated experiments on a grid over (m, n)

Optimal decisions on a grid over (m, ν) at decision epoch n

Working backward, the utility function in the jth cell for $n < N$ when the decision is to continue the trial is given by

$$\hat{U}_j^n(a^n = 1) \approx \frac{1}{|A_j^n|}\sum_{i \in A_j^n} \hat{V}_{j(i)}^{*,n+1}, \tag{10}$$

where $j(i)$ denotes a cell that trajectory i visits at decision epoch $n + 1$. Therefore, the approximated value function in each cell of the grid at decision epoch $n < N$ is

$$\hat{V}_j^{*,n} = \max\Big\{0, \hat{U}_j^n(a^n = 1), \hat{U}_j^n(a^n = 2)\Big\}. \tag{11}$$

Enumerating the entire grid backward until decision epoch n identifies the optimal decision and value function for each cell. Figure 2(b) shows a hypothetical example of optimal decisions on the grid (m, ν) at a particular decision epoch n. Algorithm 1 in Online Supplement Section 3 describes the method in more detail. Berry et al. (2002) applied this approach under a set of "typical dose-response curves" where the approximate value function for each grid cell was computed by taking the average of expected utilities under the same set of dose-response curves. Therefore, when a true observation from a dose-response curve investigated in the trial becomes available at decision epoch n, an (m_n, ν_n)-tuple is evaluated, and depending on which grid cell it falls into, the optimal decision is identified. This approach may be problematic particularly when the unknown dose-response curve does not closely resemble those in the typical set. Furthermore, when the shape of the dose-response curve is unknown and a response-adaptive dynamic allocation scheme is trying to learn it, the resulting response-adaptive optimal stopping problem becomes computationally demanding. In Sections 5.2 and 5.3, we propose alternative methods that are significantly more efficient and can be used in a fully sequential setting.

5.2. A One-Step Look-Ahead Policy

Frazier and Powell (2008) proposed a kind of one-step look-ahead policy (knowledge gradient) to optimal stopping of ranking and selection problems by assuming that the experiment has to terminate at the next decision epoch. We adapt such a framework into the optimal stopping of a dose-finding trial with unique challenges. In particular, we consider three actions at each decision epoch—abandonment, continuation, and termination—whereas in most standard ranking and selection problems, only continuation and termination decisions are available. Furthermore, our utility function consists of $\mathbb{E}[\mathbb{1}_{\{B^n\}}|\ \mathscr{F}^n]$, which is emanated from evaluating the significance of the advantage over the placebo via a hypothesis test, when the decision is to terminate the trial.

To quantify the value gained in continuing the trial, define $V_{a^n=1}^{\mathrm{KG}}(s)$ as a function that measures the difference between terminating or abandoning the trial at time n and continuing the trial, incurring the cost of sampling, and terminating or abandoning the trial at time $n + 1$; that is,

$$V_{a^n=1}^{\mathrm{KG}}(s^n) = \mathbb{E}\Big\{-c_1 + \max_{a^{n+1}\in\{0,2\}} u\big(a^{n+1}, s^{n+1}, \mathscr{F}^{n+1}\big)\big|\mathscr{F}^n\Big\} - \max_{a^n\in\{0,2\}} u(a^n, s^n, \mathscr{F}^n), \tag{12}$$

where the knowledge gradient policy π^{KG}, hereafter referred to as the KG policy, decides to continue the trial (i.e., $a^{\pi^{KG}}(s^n) = 1$), when $V_{a^n=1}^{\mathrm{KG}}(s^n) > 0$. In the case that $V_{a^n=1}^{\mathrm{KG}}(s^n) \le 0$, the optimal decision is identified by $a^{\pi^{\mathrm{KG}}}(s^n) \in \arg\max_{a^n\in\{0,2\}} u(a^n, s^n, \mathscr{F}^n)$. Note that $a^{\pi^{\mathrm{KG}}}(s)$ is a function returning the optimal decision selected when in state s^n under the KG policy π^{KG}. In order to evaluate $V_{a^n=1}^{\mathrm{KG}}(s^n)$, one needs to estimate both the current expected utility function $u(a^n, s^n, \mathscr{F}^n)$ and the one-step utility function $u(a^{n+1}, s^{n+1}, \mathscr{F}^{n+1})$ by taking a sample average (Monte Carlo). Algorithm 2 in Online Supplement Section 3 presents the details of this procedure. This approach replaces multistep forward simulations of the trial from decision epoch n to N by one-step forward simulations, which significantly reduces the complexity and computational time of the algorithm.

The following result bounds the optimal decision from below, and it shows that the KG policy may stop sooner than the optimal policy; that is, whenever the KG policy decides to continue the trial, the optimal decision is also the continuation of the trial. This proposition motivates a sensitivity analysis with respect to the history of the trial. We later show that stopping sooner than the optimal policy may result in a low probability of correct decision in certain situations. This is also showcased in our case study where both the simulation-based gridding and the KG policy result in unacceptable probabilities of correct decision.

Proposition 1. *The optimal stopping time τ is bounded below by the KG stopping time τ^{KG}; that is, $\tau^{KG} \le \tau$.*

5.3. Diffusion Approximation

Although the complexity and computational time of the knowledge gradient method is significantly better than the simulation-based gridding method, both require forward simulations to approximate the optimal solution to the value functions in (7). Instead, we propose a method that assumes a prior belief about the actual benefit of the target dose over the placebo and approximates its increments over time by a continuous-time Wiener process, which enables us to construct the optimal stopping boundaries up front. This framework is inspired by Chernoff (1961), where a diffusion approximation is used to sequentially test whether the drift of a Wiener process is positive. Our approach also approximates the stopping time of sequential normal means (i.e., the advantage over placebo) by solving a continuous-time Bellman equation. To that end, we first consider a setting where there is a single unknown dose versus a known placebo. Then, we design a heuristic that uses the said boundaries for decision making in multiple doses settings.

5.3.1. A Single Dose with Unknown Mean Response vs. a Placebo. For now, assume that the trial involves a placebo with known mean response and a single dose with unknown mean response. Without loss of generality, assume that $y_0 \sim \mathcal{N}(0, \sigma^2)$ and $y_* \sim \mathcal{N}(\theta, \sigma^2)$, where θ is unknown and a prior $\theta \sim \mathcal{N}(m_0, \nu_0^2)$ is given. Set $t_0 = \frac{\sigma^2}{\nu_0^2}$. For a single dose, the advantage over placebo is given by $df^* = \theta - 0 = \theta$; see Section 4. Therefore, at each time period, a sample from the dose with unknown mean response is observed, and the posterior of θ and, therefore, df^*, becomes $df^*| \mathscr{F}^n \sim \mathcal{N}(m_n, \frac{\sigma^2}{t_n})$, where

$$\begin{aligned} t_n &= t_0 + n, \\ m_n &= \frac{t_0}{t_n} m_0 + \frac{\sum_{i=1}^n \hat{y}_*^i}{t_n}. \end{aligned} \tag{13}$$

In this setting, df^* naturally follows a normal distribution. Recall that in the utility calculation there is an expectation to calculate, which by this construction has a closed form. In particular,

$$\begin{aligned} \mathbb{E}[\mathbb{1}_{\{B^n\}}|\mathscr{F}^n] &= \mathbb{P}\left\{\frac{\sqrt{n_p}(\bar{y}_* - \bar{y}_0)}{\sqrt{\left(2\sigma^2 + \frac{\sigma^2}{t_n}\right)}} > q_\alpha|\mathscr{F}^n\right\} \\ &= 1 - \Phi(Q_\alpha(m_n, t_n)), \end{aligned}$$

where $Q_\alpha(m_n, t_n) = q_\alpha - \frac{m_n\sqrt{n_p}}{\sqrt{(2\sigma^2 + \frac{\sigma^2}{t_n})}}$, $2\sigma^2 + \frac{\sigma^2}{t_n}$ is the posterior predictive variance of $\bar{y}_* - \bar{y}_0$, and $\Phi(\cdot)$ denotes a normal cumulative distribution function.

Redefine the state variable $\hat{s} = (m_n, t_n)$, and using $\hat{s}^0 = (m_0, t_0)$, let $\tilde{l}_\pi(\hat{s}^0)$ denote the expected utility at stopping time τ under policy $\pi \in \Pi$ when the prior is parameterized by (m_0, t_0); that is,

$$\begin{aligned} \tilde{l}_\pi(\hat{s}^0) = \mathbb{E}^\pi[&-c_1\tau + \max\{0, -c_1' n_p \\ &+ c_2 m_\tau(1 - \Phi(Q_\alpha(m_\tau, t_\tau)))\}|\hat{s}^0], \end{aligned} \tag{14}$$

where the DM selects a policy $\pi \in \Pi$ such that $V^*(\hat{s}^0) = \sup_{\pi\in\Pi} \tilde{l}_\pi(\hat{s}^0)$. Define $x_0 = m_0 t_0$ and $x_n = x_0 + \sum_{i=1}^n \hat{y}_*^i$, where $m_n = \frac{x_n}{t_n}$. Using these definitions, the state variable can be rewritten as $\hat{s}^n = (x_n, t_n)$. Let $G(x_\tau, t_\tau)$ denote the expected utility at the stopping time given by

$$G(x_\tau, t_\tau) = \max\left\{0, -c_1' n_p + c_2 \frac{x_\tau}{t_\tau}\left(1 - \Phi\left(Q_\alpha\left(\frac{x_\tau}{t_\tau}, t_\tau\right)\right)\right)\right\}. \tag{15}$$

Because the utility functions are uniformly bounded for any state and action, and the action space is finite, there exits a Markovian and deterministic optimal policy. The optimal policy to $V^*(m_0, t_0) = \sup_{\pi\in\Pi} \tilde{l}_\pi(m_0, t_0)$ is the solution to the following Bellman equation:

$$\begin{aligned} &B(x_n, t_n) = \max\{G(x_n, t_n), -c_1 + \mathbb{E}[B(x_{n+1}, t_{n+1})|x_n, t_n]\}, \\ &B(x_\tau, t_\tau) = G(x_\tau, t_\tau), \end{aligned} \tag{16}$$

where $t_{n+1} = t_n + 1$, and $x_{n+1} = x_n + \hat{y}_*^{n+1}$.

Optimality equation (16) has a continuous state space, and thus it is computationally intractable to solve. Therefore, in order to approximate the solution to (16), suppose that patients' responses are observed continuously rather than at discrete decision epochs t_n. This assumption is necessary for a rigorous development of the method and is not required for implementation in practice. In fact, we use stopping boundaries developed in a continuous regime for the stopping decisions of trials where patients arrive in discrete time epochs. One can think of the continuous equivalent of the cumulative sum $x_n = x_0 + \sum_{i=1}^n \hat{y}_*^i$ as a Brownian motion with unknown drift θ and variance σ^2 per unit time, which satisfies the following stochastic differential equation:

$$dx_t = \theta\, dt + \sigma\, dW_t, \tag{17}$$

where x_t is an extension of x_n to continuous real values for real-valued t, and W_t is a standard Brownian motion. Extend the definition of filtration $\mathscr{F}^n$ to be the natural σ-algebra generated by the process $\{x_t\}_{t\in[t_0,t_n]}$ (i.e., $\mathscr{F}_{ct}^{t\in[t_0,t_n]}$). Therefore, the continuous-time approximation of the Bellman equation in (16) is given by

$$\begin{aligned} &B_{ct}(x_t, t) = \max\{G(x_t, t), -c_1\Delta t + \mathbb{E}[B(x_{t+\Delta t}, t + \Delta t)|\mathscr{F}_{ct}^t]\}, \\ &B_{ct}(x_\tau, \tau) = G(x_\tau, \tau). \end{aligned} \tag{18}$$

The following proposition shows that $B_{ct}(x_t, t)$ is the solution to a free boundary problem with a partial differential diffusion-advection equation and two boundary conditions.

Proposition 2. *The term $B_{ct}(x_t, t)$ is the solution to the following partial differential equation in the continuation set $\mathscr{C} := \{(x_t, t) : -c_1\Delta t + \mathbb{E}[B_{ct}\ (x_{t+\Delta t}, t+\Delta t)|\mathscr{F}^t_{ct}] > G(x_t, t)\}$:*

$$0 = -c_1 + \frac{\partial B_{ct}(x_t,t)}{\partial t} + \frac{\partial B_{ct}(x_t,t)}{\partial x}\frac{x_t}{t} + \frac{1}{2}\frac{\partial^2 B_{ct}(x_t,t)}{\partial x^2}\sigma^2, \qquad (19)$$

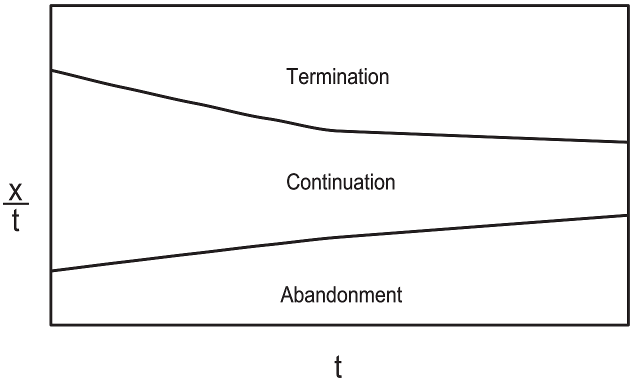
where $B_{ct}(x_t, t) = G(x_t, t)$ outside of the set $\mathscr{C}$. The free boundary $\partial\mathscr{C}$ is given by

$$\begin{aligned} B_{ct}(x_t,t) &= G(x_t,t) && \text{on } \partial\mathscr{C}, \\ \frac{\partial B_{ct}(x_t,t)}{\partial x} &= \frac{\partial G(x_t,t)}{\partial x} && \text{on } \partial\mathscr{C}. \end{aligned} \qquad (20)$$

Boundaries to the continuation set $\mathscr{C}$ can be found without any trial simulation, which significantly reduces the complexity and computational effort required to obtain optimal stopping times. The solution algorithm to the free boundary problem is described in Online Supplement Section 3. We use trinomial tree discretization method to solve the partial differential problem of Proposition 2. Figure 3 demonstrates an example of the solution to the free boundary problem. The approximated optimal decision is identified by determining the region of the state variable after each new observation.

5.3.2. Multiple Doses with Unknown Mean Responses. In the previous subsection, we construct the continuation boundaries where there is only a single dose with an unknown mean response. However, the original problem consists of multiple doses for which the mean response is unknown. Therefore, the target dose $z^* = \text{ED}_{95}$ is random, and each continuation decision may yield a different target dose with respect to the sample path. This results in an unknown distribution for df^* when multiple doses are considered. In fact, if there was only a single dose with an unknown mean response, the allocation dose for continuation decisions and the target dose were similar, and the posterior advantage over the placebo, df^*, was distributed according to a normal distribution. However, in the multiple doses setting, the allocation dose for continuation decisions may estimate a different target dose, which results in df^* not to enjoy conjugacy with respect to the patient's response. In the literature, a variety of heuristic approaches have been proposed to extend the results of a single alternative case to multiple alternative settings. See, for example, Chick and Frazier (2012) and Chick et al. (2018). However, those approaches depend on the assumption that the reward is equivalent to the maximum expected reward of simulating the arm with the highest mean, whereas in our formulation, the arm with the highest mean does not necessarily yield the maximum expected utility. Thus, they are not applicable in our setting. Therefore, we propose the following heuristic to extend the single dose setting to the multiple doses case.

Figure 3. An Example of Boundaries of the Continuation Set



Recall that for each dose Z_j, there is a prior on the expected response θ_j. The diffusion approximation boundaries only depend on the prior and the shape of the utility function. Therefore, for each dose j, we construct the continuation boundaries up front. The idea is that, at each decision epoch, we estimate $m^*_n = \frac{x^*_n}{t_n}$ of the target dose z^* and make decisions by considering the optimal region corresponding to dose z^* to check whether m^*_n falls into termination, abandonment, or continuation regions. To that end, at each decision epoch, we create a sample from the posterior on Θ, and for each sample, we use Equation (1) to find the target dose. Then, we take a sample average to estimate the target dose, and because the said sample average may not be in $\mathcal{Z}$, we round it to the closest dose. Given the estimate of target dose z^*, we simply have $m^*_n = \mathbb{E}\{\theta_{z^*}|\mathscr{F}^n\}$. The decision is found by referring to the optimal decision region corresponding to dose z^* and checking whether m^*_n belongs to an abandonment, continuation, or termination zone at t_n. Assuming a continuation decision at time n, a patient is assigned to a dose according to the allocation scheme. Its response y^{n+1} is observed and is used to update the estimate of the dose-response curve (i.e., Θ). Then, this process continues until the stopping time or all patients are tested.

6. Numerical Results

In this section, we present implementation results of the simulation-based gridding algorithm, one-step look-ahead policy, and diffusion approximation for a variety of settings. Because the performance of these solution methods may differ depending on the adaptive dose allocation scheme, we assume that the dose allocation algorithm is given by that in Nasrollahzadeh and Khademi (2018) for all of the solution algorithms. We include a sensitivity analysis in Online Supplement Section 4.6 with different allocation policies to investigate the robustness of our approach. To that end, we implement a typical balanced randomization allocation policy that assigns patients to each dose with equal probabilities and is described as the design norm in many areas of medical research. An example of such an allocation policy is in our case study where the trial had a balanced randomization patient assignment. We also implement an adaptive randomization policy rooted in Thompson sampling presented in Berry et al. (2010), chapter 4. To assess the quality of solution methods with respect to termination, abandonment, and continuation decisions, two different types of dose-response curves are tested:

i. A sigmoid curve with a significant advantage over the placebo, one of the most recurring dose responses in practice (e.g., Gadagkar and Call 2015). This curve is used to test the performance of different stopping rules with respect to continuation and termination decisions. For this curve, the optimal decision at stopping is to terminate the trial for efficacy.

ii. A flat dose-response curve, which is used to assess the quality of different algorithms when the correct decision at stopping is to abandon the trial for futility. For further analysis on the shape of the underlying dose-response curve, see Online Supplement Section 4.5, where we discuss the performance of the proposed methods with respect to a bimodal nonmonotonic dose-response curve.

Recall that the problem is modeled as a Bayesian Markov decision process, and naturally, the policies are optimal when assessed according to a fully Bayesian setup (i.e., problem instances). In other words, true dose-response curves must be generated randomly from the same prior, and the performance must be measured with respect to the expectation under the particular prior. However, because of computational difficulties in generating results for the simulation-based gridding approach, we assess the performance of these approximation methods with respect to two dose-response curves (a frequentist setting). Assessing different algorithms with respect to a specific configuration is not unprecedented particularly in clinical trials; see, for example, Berry et al. (2002) and Krams et al. (2003). A sigmoid and a flat curve are considered to highlight the performance of these algorithms when facing favorable and unfavorable cases.

6.1. Simulation Initialization

A typical number of doses under investigation in Phase II of clinical trials is between 4 and 12 (e.g., Berry et al. 2002). We consider 11 doses including a placebo. The first dose is considered to be a placebo, and its known and fixed mean response marks the baseline score for any particular treatment. At each decision epoch, if the decision is to continue the trial, a dose must be allocated to the next patient. We use a one-step look-ahead policy to optimally select a dose that minimizes the one-step posterior variance of the target dose ED_{95}. Thereafter, the patient's response is generated from the true distribution and is used to update the posterior estimate of the dose-response curve. Aligned with the literature, we assume that the stopping algorithm is applied only after observing the responses of a certain number of patients (e.g., 20) have already been through the trial. The total number of patients volunteered for the trial is assumed to be 400. We assume that the observation variance is known and is fixed at 10 units. A sensitivity analysis is conducted on this assumption in Online Supplement Section 4.4, where we increase the observation variance to test the performance of the proposed policies when observations are less informative. The significance level is considered to be 1% across all experiments.

We assume that the sampling cost $c_1 = 1{,}000$, sampling cost in confirmatory phase $c_1' = 1{,}000$, and reward per unit advantage over the placebo $c_2 = 1{,}000{,}000$. This is to replicate the original settings of the study by Berry et al. (2002). It is also a reasonable initialization for our case study where c_1 and c_2 are in the same range. We also conduct a sensitivity analysis in Online Supplement Section 4.7, where we investigate the sensitivity of the proposed solutions to the ratio of cost/benefit for different configurations of c_1, c_1', and c_2. The prior (μ_0, Σ_0) is set according to $\mu_0 = (0, \ldots, 0)$, and Σ_0 is initiated by a Gaussian covariance function where $\text{Cov}(\theta_i, \theta_j) = \beta \exp\{-\gamma(i-j)^2\}$, where β is usually estimated by $\text{Var}(\theta_i)$. The Gaussian structure of the covariance function allows for less correlation when doses are farther apart. To keep the symmetry of the covariance matrix, β is chosen to be equal to $\frac{\text{Var}(\theta_i) + \text{Var}(\theta_j)}{2} = 100$, and γ, the lengthscale factor, is set to 0.01 for both sigmoid and flat curves. This is to ensure that initial values of the expected responses carry little prior information about the shape of the dose-response curve. A thinning factor of 5 is used in generating random variables where every fifth random variable created is used to avoid serial correlation in the sequence of random numbers. In reporting the results, 30 simulations with different sequences of random numbers are considered. The simulation is coded in the R programming language and is run on an Intel core i7 3.7 GHz processor with 16 GB of RAM.

In case of the simulation-based gridding algorithm, recall that the advantage over the placebo (i.e., df^*), in the literature, is assumed to be normally distributed according to $\mathcal{N}(m, \nu^2)$. The prior values for m_0 and ν_0 are set equal to 0 and 10 to ensure that the prior carries little information about the belief on df^*. In constructing the grid over m and ν, we consider the range of m to be 20 units (i.e., $[0, 20]$) and the range of ν to be 10 units (i.e., $[0, 10]$). The grid is divided into 40 and 20 intervals in the m and ν axes, respectively. Initially, to populate the grid, $M = 1{,}000$ experiments are run, and their (m, ν) trajectories are recorded over the grid. Afterward, from each empty cell in the grid, $M' = 10$ more simulations are initiated, and their trajectories are recorded. To implement the algorithm in an online fashion, we parallelize forward simulations to speed up the computation. For more details, we refer readers to Online Supplement Section 3.

For the diffusion approximation method, the prior values for m_0 and ν_0 are chosen to replicate those of the simulation-based gridding algorithm. We also assume a similar range for m as in the gridding algorithm (i.e., $m \in [0, 20]$). The discretization in diffusion approximation is different from the grid construction in the simulation-based gridding algorithm. Here, the grid is constructed over values of x and t. Since 20 patients have already been through the trial, t is considered to be in $[20 + t_0, 400 + t_0]$, where $t_0 = \frac{\sigma^2}{\nu_0^2}$. The details to calculate both axis intervals are given in Online Supplement Section 3. We later do a sensitivity analysis on the grid size of (m, ν) and (m, t) for both simulation-based gridding and diffusion approximation policies to investigate whether finer grids improve the performance significantly: see Online Supplement Section 4.8.

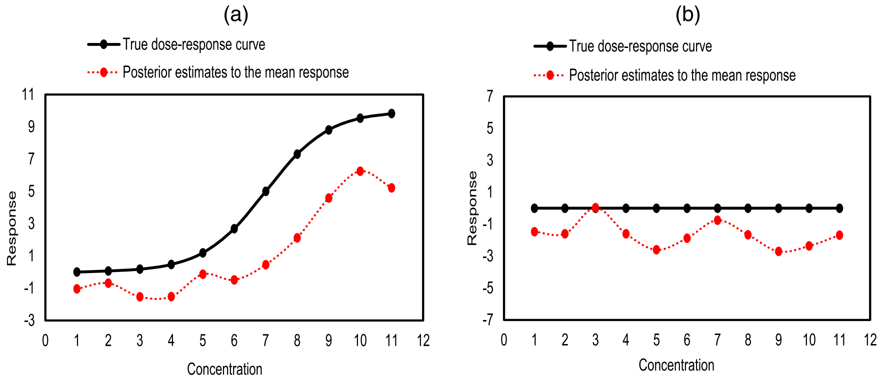
Because the simulation of the trial for all three methods is the same, we report the computational time required to find the stopping decision for each method. At each decision epoch, the gridding algorithm runs a forward simulation and uses backward induction which takes 1 hour on average. In this method the computational time in early stages when there are many patients to consider is considerably longer than the later stages when fewer patients are left. At each time period, the one-step look-ahead policy takes about 30 seconds to find the decision. The diffusion approximation creates the stopping regions up front and for a given dose allocation and its response, finding the stopping decision is instantaneous. These results confirm that the proposed methods are much less demanding than the standard method.

6.2. Results

Figures 4 and 5 show the state of the dose-response estimation after assigning 20 patients. In particular, Figure 4, (a) and (b), shows the posterior estimates to the dose-response curve where each point on the piecewise linear dotted line is the sample average of 30 posterior estimates of μ in $\Theta \sim \mathcal{N}(\mu, \Sigma)$ after observing 20 patients. Furthermore, Figure 5, (a) and (b), shows the maximum posterior variance where each point denotes the sample average of 30 posterior estimates of maximum Σ_{jj}, $j = 1, \ldots, J$, for each patient number.

Tables 1 and 2 show the estimated utility at stopping time, the true utility at stopping, the average stopping time, and the probability of correct decision (PCD) for the three stopping rules with respect to sigmoid and flat dose-response curves, respectively. In reporting the true utility at stopping, we assume that the true underlying dose-response curve is known and thus its ED$_{95}$ and the true benefit over the placebo. Therefore, the probability of success at the end of the confirmatory phase by the term $\mathbb{E}[\mathbb{1}_{\{B^n\}}|\mathscr{F}^n]$ reduces to the probability $1 - \Phi(q_\alpha - \frac{df^*\sqrt{n_p}}{\sqrt{2\sigma^2}})$. What differentiates the policies is the cost of patient sampling, which depends on stopping times corresponding to each policy. In case of detecting a significant advantage

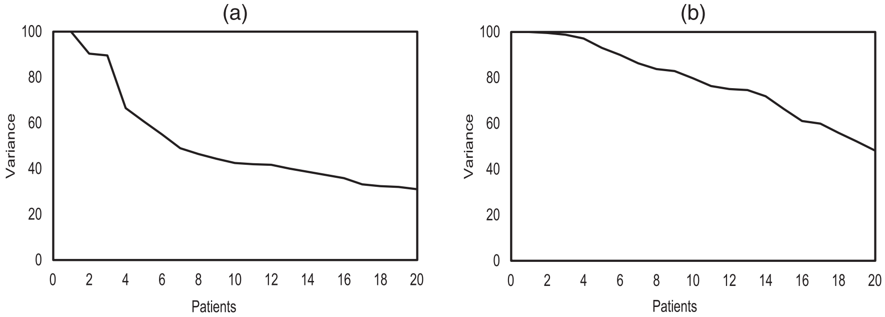
Figure 4. Posterior Estimates to the Dose-Response Curve After 20 Patients



Sigmoid curve

Flat curve

Figure 5. Maximum Posterior Variance



Sigmoid curve

Flat curve

over the placebo, the correct decision is to terminate the trial. If the true dose-response curve is flat, abandoning the trial is considered the correct decision. We also include a discussion on the probability of correctly identifying the target dose as a performance measure in Online Supplement Section 4.1. Notice that all three algorithms correctly terminate the sequential sampling process when the dose-response curve is sigmoid with a significant advantage over the placebo. The KG policy stops sooner, but the estimated utility at the stopping time is higher for the simulation-based gridding algorithm in spite of observing more patients. This is because upon stopping it has a slightly higher estimation of the advantage over the placebo and the ratio of the benefit per unit advantage over the placebo is much higher than the sampling cost. Recall that in our setting, this ratio is $\frac{c_2=1{,}000{,}000}{c_1=1{,}000}$, and thus detecting even marginal improvements in the estimate of the advantage over the placebo will overwhelm the sampling cost. The simulation-based gridding algorithm is uniquely effective in detecting these marginal improvements given high-quality prior knowledge because it simulates a large number of trials to the end of the horizon and can trade off sampling patients for beneficial marginal improvements in the estimate of the target dose mean response. The KG policy performs weaker in this regard because the decision is made with respect to the advantage over the placebo one step into the future and thus may fail to identify similar trade-off opportunities several steps in the future. The diffusion approximation method achieves lower estimated utility and stops later. In particular, Figure 6(a) demonstrates a few simulated state variable paths crossing into the termination region from the continuation region. The average stopping time and expected utilities reported in both Tables 1 and 2 are the average over 30 sample paths. The reported probability of correct decision only considers the stopping decisions and is independent of the target dose selection in case of detecting a significance.

When sampling from a flat dose-response curve, the gridding algorithm and the KG policy incorrectly terminate the trial most of the times. In the case of the KG policy, as soon as the next step expected utility is estimated to be less than the current one, the policy stops sampling. If the policy overestimates the expected response of the target dose, the current estimated utility may become positive and thus the incorrect decision to terminate the trial instead of abandoning. Table 2 shows that the diffusion approximation algorithm correctly abandons 96% of times when the dose-response curve is flat, although the average abandonment time comes significantly further in the trial. Also, notice the difference between estimated and true utilities of the simulation-based gridding and KG policy. Both methods incorrectly estimate a monetary benefit, whereas the underlying dose-response curve will only result in sampling costs. Figure 6(b) shows a few simulated state variable paths of the diffusion approximation method crossing into the abandonment region. Furthermore, the estimated utility at the stopping time for the diffusion approximation algorithm, although lower than the simulation-based gridding algorithm and the KG policy, is closer to the

Table 1. Sigmoid Dose-Response Curve

	Estimated utility	True utility	Stopping time	PCD
Simulation-based gridding	11,076,870	9,420,000	80	1
KG	10,791,346	9,462,000	38	1
Diffusion approximation	10,632,415	9,394,000	106	1

Note. Stopping times are reported in terms of number of patients going through the trial before a stopping decision is made.

Table 2. Flat Dose-Response Curve

	Estimated utility	True utility	Stopping time	PCD
Simulation-based gridding	1,926,790	−64,000	34	0.10
KG	1,840,765	−58,000	28	0.10
Diffusion approximation	−3,910	−307,000	277	0.96

Note. Stopping times are reported in terms of number of patients going through the trial before a stopping decision is made.

true expected utility for the flat dose-response curve. Therefore, one might conclude that the gridding and KG algorithms do not produce reliable estimates when the true dose-response curve is flat. Results show that in this setting, the standard method may produce significantly poor solutions, which may have severe consequences in terms of costs of the next phase and the health of future patients; see Rojas-Cordova and Hosseinichimeh (2018) for a discussion on consequences of misspecification errors in adaptive clinical trials.

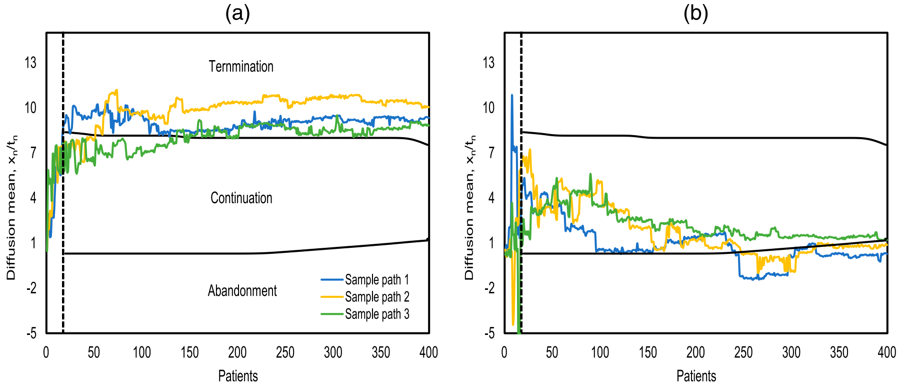
One approach to address such a shortcoming of the gridding and KG policies is to start considering stopping decisions only if enough evidence is gathered regarding the dose-response curve. This evidence may be interpreted as the accuracy of the dose-response estimation (i.e., the diagonal of the covariance matrix Σ in state variable s^n). Next, we address such an extension.

6.3. The Effect of History $\mathscr{F}^n$

Motivated by our results, we propose applying the stopping rule only after a certain number of patients' responses have already been observed. As more patients' responses are added to the history, the accuracy of the estimation about Θ increases. This is because sampling dose j results in lowering Σ_{jj}, which in turn is a measure of uncertainty about the dose-response estimation at dose j. Therefore, considering a bound on $\max_j \text{Var}[\theta_j|\mathscr{F}^n]$ ensures a minimum level of accuracy about the dose-response curve estimation. This modification does not contribute to the complexity of the stopping rule because $\text{Var}[\theta_j|\mathscr{F}^n] = \Sigma_{jj}$ is already available to the DM as a part of the state space. We propose the following heuristic: at each decision epoch for any solution method, if the decision is to continue, continue; if the decision is to stop, check whether $\max_j \text{Var}[\theta_j|\mathscr{F}^n] \leq \bar{V}$ is satisfied; if it is satisfied, follow the decision; otherwise, continue.

One can tune $\bar{V}$ to change the amount of evidence gathered before the stopping decisions are applied. We consider $\bar{V} = 4$ units in presenting the results. Figure 7 shows the state of dose-response estimation in terms of posterior estimate to the dose-response curve when $\max_j \text{Var}[\theta_j|\mathscr{F}^n] \leq 4$ for the first time. Figure 8 shows the maximum posterior variance from the start of the trial until $\max_j \text{Var}[\theta_j|\mathscr{F}^n] \leq 4$ for the first time. In case of the sigmoid dose-response curve, $\max_j \text{Var}[\theta_j|\mathscr{F}^n] \leq 4$ when $n \geq 280$, and $\sup_{\pi\in\Pi} l_\pi(s^0) = 10{,}011,765$, which is achieved at patient 286. For a flat dose-response curve, $\max_j \text{Var}[\theta_j|\mathscr{F}^n] \leq 4$ when $n \geq 243$, and $\sup_{\pi\in\Pi} l_\pi(s^0) = -312$ is achieved at patient 296. Similar to Section 6, Tables 3 and 4 show the performance measures for the three stopping rules with respect to the sigmoid and flat dose-response curve, respectively. See Online Supplement Section 4.2 for state variable paths of the diffusion approximation.

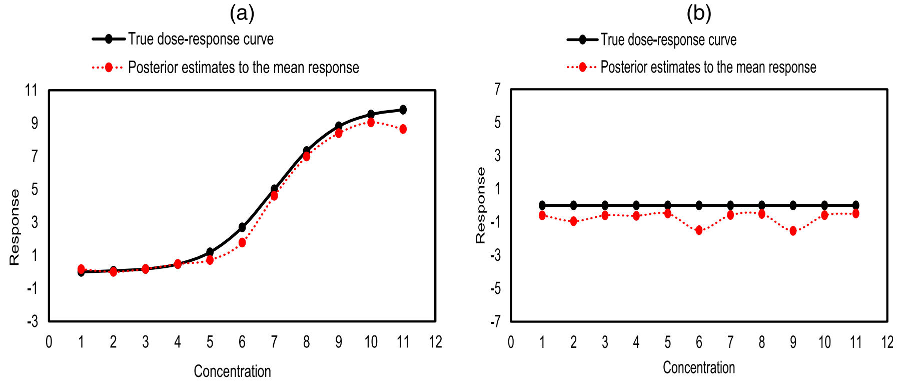
Figure 6. State Variable Paths



Sigmoid curve

Flat curve

Figure 7. Posterior Estimates to the Dose-Response Curve When $\max_j \mathrm{Var}[\theta_j|\mathscr{F}^n] \le 4$



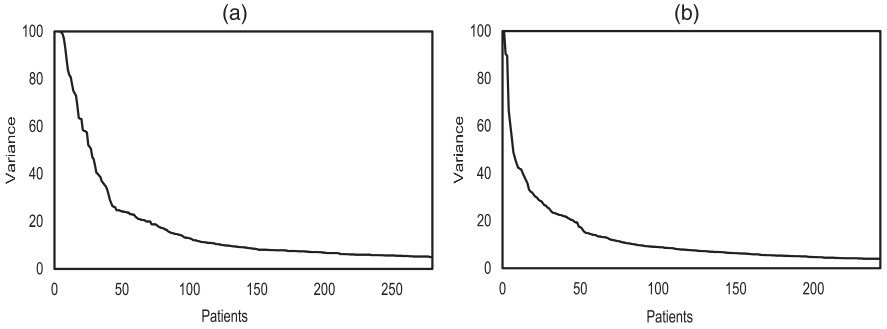
Sigmoid curve

Flat curve

6.4. Sensitivity Analyses

To assess the robustness of the proposed solutions with respect to model/simulation parameters, dose-response model specifications, and different allocation procedures, we conduct further sensitivity analyses on the variance of observations, shape of the underlying dose-response curve, dose allocation procedure, monetary parameters, and discretization parameters. In particular, in Online Supplement Section 4.4, the variance of observations is increased by a factor of 10 to add uncertainty to the information present in each observation. Our results demonstrate that the proposed policies are robust with respect to the variance of observation in that their relative performance in terms of the estimated utility, stopping time, and probability of correct decision persist for the sigmoid dose-response curve. However, as expected, the number of patients required to satisfy the condition of Section 6.3 increases, which improves the performance of the simulation-based gridding and KG policies because the cost of sampling more patients may cancel out any incorrect identification of the advantage over the placebo. In Online Supplement Section 4.5, the performance of the proposed solutions is investigated for a bimodal curve as a representative of a more general dose-response relationship. The results suggest that the proposed policies are robust with respect to the shape of the underlying dose-response curve. Online Supplement Section 4.6 compares the performance of the proposed policies with respect to the underlying dose allocation procedure. Balanced randomization and another adaptive randomization (developed by Berry et al. (2010)) are also investigated, which show similar results to Section 6.2. Note that in Section 6.1, we assume that $c_1 = c_1' = \$1{,}000$ and $c_2 = \$1{,}000{,}000$. Such a cost/benefit ratio may not hold up for every trial, and thus Online Supplement Section 4.7 increases the $\frac{c_1'}{c_1}$ by factors of 2 and 5 and decreases c_2 by factors of 10 and 100. The effects of all these combinations are reported on the performance of KG and the diffusion approximation. The results show that diffusion approximation outperforms KG in every case. Finally, the descretization parameters for simulation-based gridding and diffusion approximation methods are changed to generate finer grids. However, our results show that the improvement in both cases is marginal.

Figure 8. Maximum Posterior Variance Until $\max_j \mathrm{Var}[\theta_j|\mathscr{F}^n] \le 4$



Sigmoid curve

Flat curve

Table 3. Sigmoid Dose-Response Curve

	Estimated utility	True utility	Stopping time	PCD
Simulation-based gridding	10,001,441	9,211,000	289	1
KG	9,957,969	9,218,000	282	1
Diffusion approximation	9,916,273	9,220,000	280	1

Note. Stopping times are reported in terms of number of patients going through the trial before a stopping decision is made.

6.5. Case Study: Application to Clinical Data

In order to show the performance of our policies with respect to real data from clinical trials, we simulate the proposed policies for a treatment studied in Hall et al. (2011) on the effects of genomic test-directed chemotherapy for early-stage lymph node–positive breast cancer. This treatment was studied for its potential on saving people from unnecessary chemotherapy and reducing costs. The trial investigated two treatment methods: standard care, which means chemotherapy for all patients, and genomic test-directed chemotherapy. For participants in the standard care arm, the probability of a five-year disease-free survival is 0.79 with a standard error of 0.028. The genomic test-directed chemotherapy achieved a 0.88 probability of a five-year disease-free survival on average with a standard error of 0.038. A total of 300 patients were recruited for the preliminary trial and were allocated to each arm according to a balanced randomization policy. Compared with the standard care, the genomic test-directed chemotherapy had an incremental cost of €860 ($1,125). Hall et al. (2011) estimated that, on average, the genomic test-directed chemotherapy had 0.16 years in QALYs gained with an incremental cost-effective ratio of €5,529 per year per patient. A conservative estimate of the market value per QALY is presented in Online Supplement Section 4.3. The significance level is assumed to be 5%. Table 5 summarizes the trials specifications and parameters for this simulation. Other parameters not mentioned in Table 5 are initialized according to Section 6.1.

Table 6 shows the estimated utility, the true utility, the stopping time, and the PCD for the gridding algorithm, the KG policy, and the diffusion approximation method considering the underlying dose-response curve of Hall et al. (2011). Aligned with the actual trial, we use a randomized allocation rule for assigning patients to treatments upon the continuation decision. The results are consistent with our numerical experiments and demonstrate further that the proposed diffusion approximation performance is more reliable. In fact, the standard method produces inaccurate estimates for the utility and terminates the trial for efficacy. However, Hall et al. (2011, p.57) concluded that “there is substantial uncertainty regarding the cost-effectiveness of Oncotype DX–directed chemotherapy” at the end of the trial and stated further research has to be done to collect more information about the cost effectiveness of the treatment. In particular, we consider the correct decision to be abandonment or continuation at the end of the trial in our simulations. Specifically, our results show that the majority of state variable paths in the diffusion approximation remain in the continuation region at the end of the trial with 300 patients. Notice that the results in Table 6 agree with the previous results of Tables 2 and 3 where the simulation-based gridding algorithm performs best with respect to the estimated utility, whereas KG achieves the highest true utility. However, both methods are susceptible to incorrectly detect a significant response, and their probability of correct decision is much lower than the diffusion approximation method. Also note that in Tables 2 and 3, simulation-based gridding and KG produce the lowest probabilities of correct decision and overestimate the utility significantly.

7. Conclusions

In this work, we studied the optimal stopping problem of an adaptive dose-finding clinical trial capable of terminating the trial for efficacy or abandoning it as a result of futility. We implemented a simulation-based gridding solution method and compared it with two

Table 4. Flat Dose-Response Curve

	Estimated utility	True utility	Stopping time	PCD
Simulation-based gridding	−11,199	−314,000	284	0.69
KG	−12,889	−274,000	244	0.64
Diffusion approximation	−12,571	−334,000	304	0.96

Note. Stopping times are reported in terms of number of patients going through the trial before a stopping decision is made.

Table 5. Hall et al. (2011) Trial Simulation Parameters

Parameters	Value
Dose-response curve	Standard arm with 0.79 probability of 5-year increase in QALYs Treatment arm with 0.88 probability of 5-year increase in QALYs
Observation variance	Standard arm: 0.12 Treatment arm: 0.22
$c_1 = c_1'$	\$1,125
c_2	\$1,000,000

proposed methods in terms of solution quality and computational effort. Our first proposed method assumes that the next decision epoch is the last one (KG) and produces stopping decisions accordingly. Our second proposal considers a two-dose continuous version of the sampling and stopping problem and creates an Itô process for the state transition by which solving the continuous Bellman equation coincides with solving a partial differential advection-diffusion equation. We proposed a heuristic approach to extend the algorithm to multiple doses setting.

Our results show that if in the true dose-response curve the target dose has a significant advantage over the placebo, all three methods make a right decision in terminating the trial for efficacy; the KG policy stops sooner, followed by the simulation-based gridding, although the diffusion approximation requires more sampling epochs. Therefore, the estimate of the utility for the standard approach and the policy KG are higher than that of the diffusion approximation. However, if in the true dose-response curve the target dose does not have a significant advantage over the placebo, the gridding and KG methods perform extremely poorly in terms of the probability of correct decision and estimating the utility. In particular, these two methods decide on termination 90% of the time on average, although the correct decision is abandonment; that is, the error probability is 0.9 for these methods, which may have significant adverse consequences and is unacceptable for regulatory approvals. In fact, these two methods stop too early and significantly overestimate the benefits upon termination. By stark contrast, the diffusion approximation method produced abandonment decisions in 96% of the times in this setting, resulting in only 4% error, which shows that the diffusion method stops the trial much later, when it has enough evidence for making decisions.

Our results suggest that applying the standard method in a fully adaptive setting from early on, where a DM can stop or terminate the trial at each decision epoch, may have severe consequences when the correct decision is to abandon. Motivated by such observations, we proposed a modified stopping rule, where the stopping decision is activated only if the maximum posterior variance about the mean response Θ falls below a threshold. In fact, $\max_j \mathrm{Var}[\theta_j|\mathscr{F}]$ is a metric that measures the uncertainty about the whole dose-response curve and is available to DMs at each decision epoch because it is a part of the state variable in the stopping problem. Our results show that using a constrained method significantly improves the performance of the simulation-based gridding algorithm and the KG policy.

Therefore, for recommendation purposes, the diffusion approximation method proved to be more robust with respect to different shapes of the underlying dose-response curve. In fact, based on the Food and Drug Administration (2018) estimates, only 33% of treatments clear Phase II of clinical trials, which shows that in the majority of trials, the underlying dose-response curve does not include a dose with a significant advantage over placebo. Therefore, our results suggest that the diffusion approximation method is potentially more accurate with respect to different scenarios. However, if strong evidence is available that a significant advantage over placebo exists, our results suggest that the KG policy

Table 6. Application to Clinical Trial

	Estimated utility	True utility	Stopping time	PCD
Simulation-based gridding	64,406	−120,600	62	0.2
KG	47,269	−92,475	53	0.2
Diffusion approximation	−290,634	−370,350	300	0.90

Note. Results are reported in terms of number of patients going through the trial before a stopping decision is made.

potentially stops sooner with respect to different dose-response curves and dose assignment procedures, and it is computationally less demanding than the standard gridding method. We also note that a potential extension of our methodology is to consider seamless Phase II/III trials where the probability of success at the end of Phase III can be integrated naturally in the utility function of the stopping problem.

Acknowledgments

We thank Pinar Keskinocak, associate editor, and three anonymous reviewers for their helpful comments.

References

Berry DA (2004) Bayesian statistics and the efficiency and ethics of clinical trials. *Statist. Sci.* 19(1):175–187.

Berry SM, Carlin BP, Lee JJ, Muller P (2010) *Bayesian Adaptive Methods for Clinical Trials* (CRC Press, Boca Raton, FL).

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*, Vol. 5 (Springer, New York), 99–181.

Biswas A, Bhattacharya R (2016) Response-adaptive designs for continuous treatment responses in phase III clinical trials: A review. *Statist. Methods Medical Res.* 25(1):81–100.

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. Biopharm. Statist.* 17(6):965–995.

Bothwell LE, Kesselheim AS (2017) The real-world ethics of adaptive-design clinical trials. *Hastings Center Rep.* 47(6):27–37.

Brealey RA, Myers SC, Allen F, Mohanty P (2012) *Principles of Corporate Finance*, 10th ed., Special Indian ed. (Tata McGraw-Hill Education, Dehli, India).

Brockwell AE, Kadane JB (2003) A gridding method for Bayesian sequential decision problems. *J. Comput. Graph. Statist.* 12(3):566–584.

Chernoff H (1961) Sequential tests for the mean of a normal distribution. Neyman J, ed. *Proc. Fourth Berkeley Sympos. Math. Statist. Probab.*, Vol. 1 (University of California Press, Berkeley), 79–91.

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

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

Chick SE, Gans N, Yapar O (2018) *Bayesian sequential learning for clinical trials of multiple correlated medical interventions*. INSEAD Working Paper 2020/40/TOM/ACGRE, INSEAD, Fontainebleau, France.

Deichmann RE, Krousel-Wood M, Breault J (2016) Bioethics in practice: Considerations for stopping a clinical trial early. *Ochsner J.* 16(3):197–198.

European Network for Health Technology Assessment (2013) Endpoints used for relative effectiveness assessment of pharmaceuticals: Clinical endpoints. Report, EUnetHTA, Diemen, Netherlands.

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.

Food and Drug Administration (2018) The drug development process—Step 3: Clinical research. Accessed May 1, 2019, https://www.fda.gov/patients/drug-development-process/step-3-clinical-research.

Frazier P, Powell WB (2008) The knowledge-gradient stopping rule for ranking and selection. Mason SJ, Hill RR, Mönch L, Rose O, Jefferson T, Fowler JW, eds. *Proc. 2008 Winter Simulation Conf.*, (IEEE, Washington, DC), 305–312.

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

Grieve AP, Krams M (2005) ASTIN: A Bayesian adaptive dose–response trial in acute stroke. *Clinical Trials* 2(4):340–351.

Griffin R, Lebovitz Y, English R (2010) *Transforming Clinical Research in the United States: Challenges and Opportunities: Workshop Summary* (National Academies Press, Washington, DC).

Grignolo A, Pretorius S (2016) Phase III trial failures: Costly, but preventable. *Appl. Clinical Trials* 25(8/9):36–42.

Hall PS, McCabe C, Stein RC, Cameron D (2011) Economic evaluation of genomic test–Directed chemotherapy for early-stage lymph node–positive breast cancer. *J. Natl. Cancer Inst.* 104(1):56–66.

Hee SW, Hamborg T, Day S, Madan J, Miller F, Posch M, Zohar S, Stallard N (2016) Decision-theoretic designs for small trials and pilot studies: A review. *Statist. Methods Medical Res.* 25(3): 1022–1038.

Jennison C, Turnbull BW (1999) *Group Sequential Methods with Applications to Clinical Trials* (Chapman and Hall/CRC, Boca Raton, FL).

Jitlal M, Khan I, Lee S, Hackshaw A (2012) Stopping clinical trials early for futility: Retrospective analysis of several randomised clinical studies. *British J. Cancer* 107(6):910–917.

Julious SA (2004) Sample sizes for clinical trials with normal data. *Statist. Medicine* 23(12):1921–1986.

Kotas J, Ghate A (2018) Bayesian learning of dose-response parameters from a cohort under response-guided dosing. *Eur. J. Oper. Res.* 265(1):328–343.

Krams M, Lees KR, Hacke W, Grieve AP, Orgogozo JM, Ford GA, ASTIN Study Investigators (2003) Acute Stroke Therapy by Inhibition of Neutrophils (ASTIN): An adaptive dose-response study of UK-279,276 in acute ischemic stroke. *Stroke* 34(11): 2543–2548.

Lewis RJ, Viele K, Broglio K, Berry SM, Jones AE (2013) An adaptive, phase II, dose-finding clinical trial design to evaluate L-carnitine in the treatment of septic shock based on efficacy and predictive probability of subsequent phase III success. *Critical Care Medicine* 41(7):1674–1678.

Liu F, Walters SJ, Julious SA (2017) Design considerations and analysis planning of a Phase 2a proof of concept study in rheumatoid arthritis in the presence of possible non-monotonicity. *BMC Medical Res. Methodol.* 17:149.

Müller P, Berry DA, Grieve AP, Krams M (2006) A Bayesian decision-theoretic dose-finding trial. *Decision Anal.* 3(4):197–207.

Müller P, Berry DA, Grieve AP, Smith M, Krams M (2007) Simulation-based sequential Bayesian design. *J. Statist. Planning Inference* 137(10):3140–3150.

Nasrollahzadeh AA, Khademi A (2018) Dynamic programming for response-tive dose-finding clinical trials. Working paper, Clemson University, Clemson, SC.

Rojas-Cordova A, Bish EK (2018) Optimal patient enrollment in sequential adaptive clinical trials with binary response. Working paper, Southern Methodist University, Dallas.

Rojas-Cordova AC, Hosseinichimeh N (2018) Trial termination and drug misclassification in sequential adaptive clinical trials. *Service Sci.* 10(3):354–377.

Roy A (2012) Stifling new cures: The true cost of lengthy clinical drug trials. Report, Manhattan Institute for Policy Research, New York.

Sacks LV, Shamsuddin HH, Yasinskaya YI, Bouri K, Lanthier ML, Sherman RE (2014) Scientific and regulatory reasons for delay and denial of FDA approval of initial applications for new drugs, 2000–2012. *J. Amer. Medical Assoc.* 311(4):378–384.

Snapinn S, Chen MG, Jiang Q, Koutsoukos T (2006) Assessment of futility in clinical trials. *Pharmaceutical Statist.* 5(4):273–281.

Spiegelhalter DJ, Abrams KR, Myles JP (2004) *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*, Vol. 13 (John Wiley & Sons, Chichester, UK).

Stallard N, Whitehead J, Todd S, Whitehead A (2001) Stopping rules for Phase II studies. *British J. Clinical Pharmacol.* 51(6):523–529.

Tufts Center for the Study of Drug Development (2014) Cost to develop and win marketing approval for a new drug is $2.6 billion. Press release (November 18), Tufts Center for the Study of Drug Development, Boston. http://csdd.tufts.edu/news/complete_story/pr_tufts_csdd_2014_cost_study.