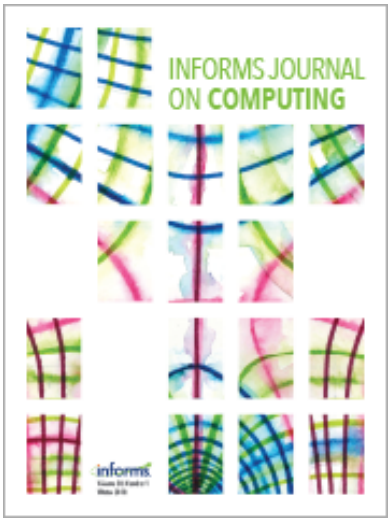


INFORMS Journal on Computing

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

Dynamic Programming for Response-Adaptive Dose-Finding Clinical Trials

Amir Ali Nasrollahzadeh, Amin Khademi

To cite this article:
Amir Ali Nasrollahzadeh, Amin Khademi (2021) Dynamic Programming for Response-Adaptive Dose-Finding Clinical Trials. INFORMS Journal on Computing

Published online in Articles in Advance 21 Oct 2021

. https://doi.org/10.1287/ijoc.2021.1082

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

INFORMS JOURNAL ON COMPUTING
Articles in Advance, pp. 1–15
ISSN 1091-9856 (print), ISSN 1526-5528 (online)

Dynamic Programming for Response-Adaptive Dose-Finding Clinical Trials

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

[a] Department of Industrial Engineering, Clemson University, Clemson, South Carolina 29634
Contact: snasrol@clemson.edu, https://orcid.org/0000-0002-8023-4629 (AAN); khademi@clemson.edu, https://orcid.org/0000-0002-5281-8715 (AK)

Received: July 5, 2019
Revised: October 7, 2020; February 21, 2021
Accepted: February 24, 2021
Published Online in Articles in Advance: October 21, 2021

https://doi.org/10.1287/ijoc.2021.1082

Abstract. Identifying the right dose is one of the most important decisions in drug development. Adaptive designs are promoted to conduct dose-finding clinical trials as they are more efficient and ethical compared with static designs. However, current techniques in response-adaptive designs for dose allocation are complex and need significant computational effort, which is a major impediment for implementation in practice. This study proposes a Bayesian nonparametric framework for estimating the dose-response curve, which uses a piecewise linear approximation to the curve by consecutively connecting the expected mean response at each dose. Our extensive numerical results reveal that a first-order Bayesian nonparametric model with a known correlation structure in prior for the expected mean response performs competitively when compared with the standard approach and other more complex models in terms of several relevant metrics and enjoys computational efficiency. Furthermore, structural properties for the optimal learning problem, which seeks to minimize the variance of the target dose, are established under this simple model.

Summary of Contribution: In this work, we propose a methodology to derive efficient patient allocation rules in response-adaptive dose-finding clinical trials, where computational issues are the main concern. We show that our methodologies are competitive with the state-of-the-art methodology in terms of solution quality, are significantly more computationally efficient, and are more robust in terms of the shape of the dose-response curve, among other parameter changes. This research fits in "the intersection of computing and operations research" as it adapts operations research techniques to produce computationally attractive solutions to patient allocation problems in dose-finding clinical trials.

History: Accepted by Paul Brooks, Area Editor for Applications in Biology, Medicine, and Healthcare.
Funding: This work was supported by the Division of Civil, Mechanical, and Manufacturing Innovation [Grant 651912]. This research is funded by the National Science Foundation [Grant 1651912].
Supplemental Material: The online supplement is available at https://doi.org/10.1287/ijoc.2021.1082.

Keywords: adaptive clinical trials • dose-finding studies • dynamic learning • knowledge gradient

1. Introduction

Clinical trials are medical studies in which participants are assigned to one or more treatments to evaluate their effects on health-related outcomes. The objective is usually to determine whether new treatments are safe and effective by measuring certain responses in trial participants (National Institutes of Health 2014). The U.S. Food and Drug Administration (FDA) classifies the approval procedure of a medical product into four phases. Phase I studies a small group of volunteers with the disease/condition for several months to identify a safe dosage range and potential side effects. Phase II increases the number of participants up to several hundred, and extends the length of study up to two years. These studies are not large enough to establish the efficacy of the treatment with certainty; however, they provide safety evaluations and allow researchers to refine their methods for the next phases. In Phase III, 300–3,000 volunteers are studied for a period of one to four years in order to confirm the drug's efficacy and to monitor its adverse reactions, particularly its long-term and rare side effects. Phase IV is carried out once the medical product has been approved by the FDA and involves several thousand volunteers and postmarket safety monitoring (U.S. Food and Drug Administration 2017). The average cost of inventing, developing, and introducing a new drug to market has exponentially increased in recent years and has surpassed $2.6 billion (Tufts 2014). The biggest drivers of this rise are expensive clinical trials whose costs can reach $300–$600 million for large trials (Griffin et al. 2010). Phase II clinical trials constitute about 18% of pharmaceutical companies research and development expenditures while their probability of success remains almost half of that in Phase I (Hay et al. 2014). Identifying the "right" dose in Phase II is a critical step in drug

development partly because of high attrition rates in Phase III (the most costly phase), which may be due to inadequate dose selection, that is, doses that are too low to achieve a desired benefit (futility because of insignificant positive effect) or doses that are too high and result in adverse reactions (exposure to unnecessary risk) (Snapinn et al. 2006).

In a standard (static) clinical trial, patients are randomly assigned to predetermined doses such that the number of patients allocated to each dose is roughly equal. Such a design may be inefficient. For example, if the slope of a dose-response curve is observed at a dose range not anticipated, equal assignment of patients to other dose ranges may lead to inefficient use of resources. These allocations may expose patients to toxic or ineffective doses, which raises ethical concerns. In addition, observations in the trial may indicate a larger variability in response to a particular dose; thus, fixed sample sizes cannot compensate for the unanticipated variability (Berry et al. 2002).

A better strategy is adaptive design, which accommodates modifications to the trial as information accrues while the trial is still in progress. For example, the experimenters may increase the number of doses under consideration in the study, drop some doses from analysis, and change the patient randomization procedure to avoid large sample sizes at doses where the shape of the dose-response curve is reasonably well estimated by the available data. Thus, adaptive designs tend to generally reduce length, total sample size, and costs of trials without compromising their integrity. Furthermore, such designs have an ethical motivation as they randomize patients to doses currently thought to be the best with greater probability (Rosenberger 1996).

In adaptive designs, dose allocation decisions are based on the current estimate of the dose-response curve. This estimate may be obtained via a parametric or nonparametric adaptive design. A standard parametric approach assumes a particular shape for the dose-response curve upfront and estimates its parameters as data accrues; see Model 1. A standard Bayesian nonparametric approach uses a second-order piecewise linear model where the curve is approximated by a pair of expected response and slope at each dose. This modeling approach is called the second-order normal dynamic linear model (NDLM), and we refer to it as Model 2. The main motivation for Model 2 is its modeling flexibility because, unlike parametric models, it can be used for estimating any dose-response curve. That is, it is immune to model misspecification, which may cause severe consequences for patients. Conversely, although parametric models have a fewer number of parameters to estimate, they are susceptible to model misspecification.

In this study, we propose an alternative Bayesian nonparametric framework for estimating the dose-response curve. In particular, our proposal approximates the curve using only the expected response at each dose (θ_j for dose j) and consecutively connects (j, θ_j) pairs to create a piecewise linear approximation. We propose three models based on different prior structure on $\Theta = (\theta_1, \ldots, \theta_J)$ (assuming J doses), each with different design purposes, assuming a normally distributed patient response. Specifically, Model 3 assumes a normal prior on Θ with an unknown mean and known correlation structure, which enjoys conjugacy. Model 4 imposes a hierarchical normal-Wishart prior on Θ, which allows for unknown correlation structure, but it loses conjugacy in our fully sequential setting. Model 5 considers a hierarchical normal-Wishart prior on Θ, where the unknown observation variance is proportional to the correlation on Θ. Although this proportionality may be limiting, Model 5 enjoys conjugacy approximately.

Although a dynamic programming (DP) formulation can be written for all of the dose-response approximation models, using Model 3 results in a formulation that is amenable to analytical study. In particular, we study structural properties of the produced optimal learning problem. A unique feature of the learning problem lies in its objective, which is the minimization of the variance of the target dose. We apply the one-step look-ahead framework into this class of learning problems and show that it learns the true target dose when the number of patients becomes large. This is in stark contrast with the standard Model 2 where the same one-step look-ahead framework shows inconsistent behavior in our numerical results. Our extensive numerical study, which includes real data from a clinical trial, shows that Model 3 is competitive in performance and significantly reduces the computational time and efforts, which are major impediments in applying Bayesian adaptive designs in practice.

Section 2 reviews different streams of literature related to our work. Section 3 describes the dose-response relationship and introduces different parametric and nonparametric approaches to its modeling. In Section 4, we present a DP formulation of the adaptive design of a Phase II clinical trial under Model 3. Section 5 derives a couple of structural properties and discusses an algorithmic way to find high-quality solutions. Section 6 discusses numerical results of our experiments with respect to various performance metrics and presents a case study using real-world clinical trial data. Section 7 briefly explores practical considerations and limitations of the proposed approaches. Finally, Section 8 concludes the paper. A table of notation, proofs of the analytical results, DP formulation for different dose-response approximation methods, and additional numerical analyses are presented in the online supplement (OS).

2. Related Works

There are three streams of literature related to this study: literature on (i) optimal design of dose-finding trials in which finding a target dose is the objective, (ii) adaptive design of dose-finding trials in which sampling policies are adapted to observed responses while the trial is in progress, and (iii) dynamic/optimal learning which focuses on deriving adaptive policies.

2.1. Optimal Design of Dose-Finding Trials

In this line of literature, researchers have investigated efficient designs for estimating a target dose by, for example, minimizing its asymptotic variance under a particular dose-response model. For example, Dette et al. (2008) proposed various optimal designs estimating the dose-response curve but did not consider response-adaptivity; thus, the designs are unable to modify dose range, sample size, or allocation scheme while the trial is in progress. These designs are also dependent on prespecified dose-response models (i.e., parametric), which are susceptible to model misspecification. In contrast, we study a range of parametric and nonparametric models to estimate the dose-response curve; our policies are adaptive to data.

2.2. Response-Adaptive Clinical Trials

In response-adaptive designs, patient allocation, dose range, and sample size are subject to modification when a new response is observed. Multiarmed bandit framework and Bayesian decision theory have been two of the most active lines of literature in response-adaptive designs. In the multiarmed bandit approach, a decision maker (DM) selects a treatment based on observed information to maximize an expected (cumulative) reward. For example, Press (2009) developed response-adaptive two-armed bandits for sequential experiments, such as clinical trials where information acquired during the trial was used to modify the allocation scheme and sampling size. Such designs were applicable only when two treatments are considered and their responses are binary (success or failure). Villar and Rosenberger (2018) extended the two-armed design to multiarmed bandits capable of comparing multiple treatments with continuous responses. However, the bandit structure is designed to identify the maximum reward when compared with a control treatment; thus, the policies derived are applicable for Phase III where the goal is to test the benefits of new treatments versus a control treatment. For more details on benefits and challenges of applying multiarmed bandits in clinical trials, see Villar et al. (2015).

Here, our focus is on response-adaptive dose-finding clinical trials where Bayesian decision theory is utilized to design response-adaptive sequential sampling policies to identify a target dose. For example, Berry et al. (2002) and Müller et al. (2006) used a second-order NDLM to formulate an adaptive dose allocation scheme (see details in OS 3). Weir et al. (2007) compared the standard Markov chain Monte Carlo (MCMC) simulation of NDLMs to estimate the dose-response curve with that of an importance sampling method. Furthermore, Krams et al. (2003) employed a similar approach in Acute Stroke Therapy by Inhibition of Neutrophils (ASTIN) clinical trial and used a fully Bayesian analysis for patient randomization, which was approved by the regulatory authorities. Lenz et al. (2015) implemented this approach in assessing the efficacy of ABT-089 in Alzheimer disease. Similar to response-adaptive designs that employ NDLM to approximate the unknown dose-response curve, our proposed dose-response modeling choices also use a piecewise linear approximation. However, in two of our three proposed models, evaluating posterior distribution parameters eliminates the required time-consuming MCMC simulation. We also show that the NDLM approach in approximating the unknown dose-response relationship may fail to identify the target dose (OS 7) in some cases because of its sensitivity to some prespecified parameters.

2.3. Dynamic Learning

Next, we briefly review the dynamic learning literature related to our work. For a comprehensive review of optimal learning, see Powell and Ryzhov (2012) and references therein.

Ranking and selection is a class of learning problems in which a risk-neutral DM seeks to find the best population, in terms of its expected value given a fixed budget to learn the unknown true distribution of the population. Gupta and Miescke (1996) introduced a special case of the one-step look-ahead policy for offline versions of ranking and selection problems where the algorithm chooses its future measurements by optimizing the one-step expected value function with respect to the posterior distributions. Frazier et al. (2008) adapted this technique to ranking and selection problems by assuming an independent multivariate normal prior. The authors referred to this technique as "knowledge gradient" (KG). Frazier et al. (2009) further developed this method to accommodate correlated normal prior beliefs. Furthermore, Wang et al. (2016) provided a KG policy for multiarmed bandits with binary responses; Parizi and Ghate (2016) implemented a KG algorithm in lot-sizing for gamma and Dirichlet priors. The objective in these studies is to maximize the expected total reward, where the reward function is equivalent to the posterior mean of the unknown parameter. In contrast, our approach minimizes the variance of the target dose, a nonlinear

function of the unknown parameter describing the dose-response curve.

There are several recent studies employing dynamic learning into clinical trials. Kotas and Ghate (2016, 2018) formulated a dynamic programming to optimal dose finding and approximated the Bellman equation by (semistochastic) certainty equivalent control techniques. Ahuja and Birge (2016) and Chick et al. (2017) considered adaptive two-armed bandits and implemented dynamic learning techniques to identify the most efficacious treatment in a variety of settings. However, as elaborated earlier, the structure of such designs are appropriate for reward-maximizing frameworks or deriving probability of correct selection in two-armed settings.

3. Approximate Models to the Dose-Response Relationship

A dose-response curve identifies the relationship between treatment dose of a drug and the patient response usually measured by a numerical score. Figure 1 shows three typical dose-response relationships. For example, the increase in fractional excretion of sodium as the response to the amount of loop diuretic dose prescribed is shown for heart failure patients in Figure 1(a). Let $f(z,\Theta)$ denote the true mean response at dose z parameterized by vector Θ. We assume that the patient response as a function of the prescribed dose z is given by

$$y = f(z,\Theta) + \epsilon, \tag{1}$$

where $\epsilon \sim \mathcal{N}(0,\sigma^2)$ denotes the noise in observation of patient responses (Berry et al. 2002).

3.1. Target Dose

The ultimate goal of a dose-finding study (Phase II clinical trial) is to find the target dose. This is typically achieved based on the approximation model to the unknown dose-response curve, that is, $f(z,\Theta)$. Existing literature usually focuses on three definitions for the target dose: (1) minimum effective dose defined as the smallest dose producing a particularly relevant response; (2) maximum tolerable dose defined as the highest dose producing a desired response without unacceptable toxicity; (3) ED_{95} defined as the smallest dose at which 95% of the maximal response is achieved. In this study, we focus on estimating the ED_{95} formally represented as

$$ED_{95} = \min_z \{z \in \mathcal{Z} : f(z,\Theta) \geq 0.95 f(z_{max},\Theta)\}, \tag{2}$$

for a given Θ, where z_{max} is a dose at which maximal response is observed. If $f(z_{max},\Theta) < 0$, we let ED_{95} be the initial dose. Note that ED_{95} is also motivated by toxicity concerns: Higher responses usually correspond to higher doses, which are more likely to cause adverse effects. Therefore, minimizing the dosage level that achieves a desired response is a predominant choice. For more discussion on this issue, see Berry et al. (2002). Nonetheless, our proposed approach can be used for other target dose definitions as well.

The next step in identifying the target dose is to choose a model for $f(z,\Theta)$, which in turn determines how patient responses are used to update its parameters, that is, Θ. As mentioned earlier, there are two main modeling approaches, which we will discuss next.

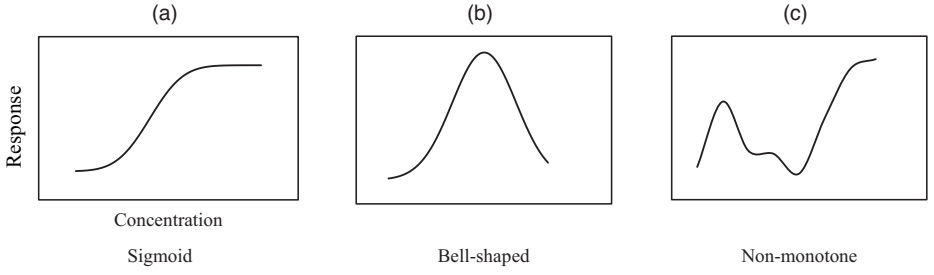
3.2. Parametric Models

In some Phase II clinical trials, a reliable guess for the shape of the dose-response curve may be available to practitioners. In such a case, functional forms that assume a fixed form for $f(z,\Theta)$ upfront and estimate the parameter vector Θ by observing patient responses may be appropriate. Hill's equation, the Michaelis-Menten model, and E_{max} models are typical examples of functional forms that assume a sigmoid shape for the underlying dose-response curve as in Figure 1(a). Consider the following representation of the Hill's equation (Gadagkar and Call 2015)

$$f(z,\Theta) = \theta_1 + \frac{\theta_2 - \theta_1}{1 + \left(\frac{\theta_3}{z}\right)^{\theta_4}},$$

where $\Theta = (\theta_1,\theta_2,\theta_3,\theta_4)$. Here, θ_1 denotes the expected response at placebo; θ_2 denotes the expected response for an infinite dosage; θ_3 may be interpreted as ED_{50}; and θ_4 is the slope at the steepest part of the

Figure 1. Typical Dose-Response Curves



Note. (a) Sigmoid, (b) bell-shaped, (c) nonmonotone.

sigmoid curve. Given a prior on Θ (and σ if it is unknown), the observation setup in Equation (1), that is, $y|\Theta,\sigma,z \sim \mathcal{N}(f(z,\Theta),\sigma^2)$, does not result in a conjugate setup and, thus, the need for MCMC simulation to sample from the posterior of Θ given a new observation y. There are two main approaches to infer about parameters Θ and σ: a single-level method that assumes independent uniform distributions over these parameters (Johnstone et al. 2016) and a hierarchical approach in which the prior over σ is inverse-gamma and θ_2, θ_3, and θ_4 are distributed according to log-normal distributions, the parameters of which are in turn distributed according to multiple normal and half-Cauchy distributions (Hennessey et al. 2010). Because hierarchical models result in overparameterization, which defeats the main purpose of functional forms, and Johnstone et al. (2016) reported little sensitivity to prior distributions, we choose to implement the single-level priors. Therefore, the choice of the dose-response relationship is given by

$$y|z,\Theta \sim \mathcal{N}(f(z,\Theta),\sigma^2), \quad f(z,\Theta) = \theta_1 + \frac{\theta_2 - \theta_1}{1 + (\frac{\theta_3}{z})^{\theta_4}},$$
$$\theta_3 \sim \mathcal{U}(\underline{\theta}_3, \overline{\theta}_3), \quad \theta_4 \sim \mathcal{U}(\underline{\theta}_4, \overline{\theta}_4),$$

where $\mathcal{U}(\cdot,\cdot)$ denotes a uniform distribution. Hyperparameters $\underline{\theta}$ and $\bar{\theta}$ are used to denote the range of θ_3 and θ_4. In our setup, θ_1 and θ_2 are assumed to be predetermined after an initial Phase I study. One main drawback of parametric models is that if the true shape of the dose-response curve is significantly different from the assumed one, the parameter estimation may be poor resulting in poor patient assignment, which may have severe consequences.

3.3. Nonparametric Models

In early stage clinical trials, the dose-response relationship is usually not known in advance and nonparametric methods are proposed to represent the unknown dose-response curve in order to avoid model misspecification. Piecewise linear approximations where a series of straight lines represent different parts of the unknown curve are a predominant approach. We represent the standard model (Model 2), propose three alternative models, and discuss their pros and cons. In the following nonparametric models, $\mathcal{Z} := \{Z_j : j = 1,\dots,J\}$ refers to the set of allowable doses, where Z_1 may denote the placebo. Hereafter, we use index j to refer to dose Z_j in order to ease notation.

3.4. Standard Second-Order NDLM

One may approximate a dose-response curve at doses close enough to dose $j \in \mathcal{Z}$ by fitting a line passing through the expected response at dose j and a slope at that dose. Therefore, for doses z close enough to dose j, the patient response becomes the straight line $\theta_j + (z-j)\delta_j$, where $\Theta = (\theta_1,\dots,\theta_J)$ is a vector of expected responses and $\boldsymbol{\delta} = (\delta_1,\dots,\delta_J)$ is a vector of slopes at each dose. Linear extrapolation for consecutive doses in set $\mathcal{Z}$ results in the relationship $\theta_j = \theta_{j-1} + \delta_{j-1}$ assuming doses are spaced equally and within one unit of each other. This piecewise linear structure is used in a second-order NDLM, which allows for random normal deviations from this structure in the following form

$$y|z,\Theta,\boldsymbol{\delta} \sim \mathcal{N}(f(z,\Theta,\boldsymbol{\delta}),\sigma^2),$$
$$\begin{pmatrix}\theta_j\\ \delta_j\end{pmatrix} = \begin{pmatrix}\theta_{j-1}+\delta_{j-1}\\ \delta_{j-1}\end{pmatrix} + \begin{pmatrix}\nu_j\\ \omega_j\end{pmatrix}, \quad \nu_j \sim \mathcal{N}(0,\mathcal{V}_j), \quad \omega_j \sim \mathcal{N}(0,\mathcal{W}_j), \tag{Model 2}$$

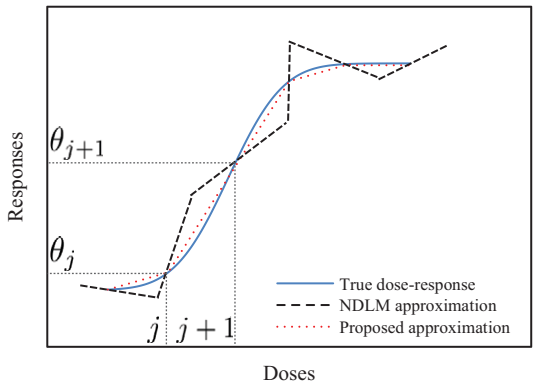
where $\mathcal{V}_j$ and $\mathcal{W}_j$ are known error terms and are set equal to each other for simplification in our implementation. Assuming a multivariate normal prior on the tuple $\mathcal{W}_j$ results in a posterior in proportion to a multivariate normal distribution under which (θ_j,δ_j) follows a bivariate normal whose parameters are given by a set of recursive equations (West and Harrison 1997); see OS 3. Therefore, sampling from the joint posterior distribution is challenging and requires special algorithms, such as "forward filtering backward sampling" (Frühwirth-Schnatter 1994, p. 183), or MCMC simulations, which are not computationally efficient. In comparison with Model 1, the NDLM approach is flexible in approximating any shape for the dose-response curve and accommodates nontrivial correlation structure between different doses.

3.5. Proposed Dose-Response Curve Approximations

We propose first-order but possibly hierarchical piecewise linear approximations to the dose-response curve. Similar to Model 2, the relationship between the assigned dose and the expected response is identified by $f(z=j,\Theta) = \theta_j$, where $\Theta = (\theta_1,\dots,\theta_J)$ denotes the vector of expected responses at dose j. The straight line approximating a part of the dose-response curve is constructed by connecting points (j,θ_j) such that the slope between consecutive doses becomes $\theta_{j+1} - \theta_j$. Figure 2 shows the second-order NDLM and the proposed piecewise linear approximation. In fact, instead of tracking two parameters, (θ_j,δ_j), per dose in the second-order NDLM approach, we only keep track of one parameter, θ_j, at each dose. Next, we propose three models based on the first-order approximation of the dose-response curve.

Setting a multivariate normal prior on the belief about the dose-response curve, that is, Θ, results in conjugacy of prior and likelihood distribution for a single observation assuming that the observation

Figure 2. (Color online) Piecewise Linear Approximation to Dose-Response Curve



variance is known. Calculating the posterior distribution in this setup is instantaneous. The model is given by

$$y|z,\Theta \sim \mathcal{N}(f(z,\Theta),\sigma^2),\quad \Theta \sim \mathcal{N}(\mu^0,\Sigma^0), \qquad \text{(Model 3)}$$

where μ^0, a $J\times 1$ vector, and Σ^0, a $J\times J$ positive semi-definite matrix, are both known. Assuming a known correlation structure, Σ^0, between different doses may be interpreted as losing some flexibility in allowing for unknown correlation structures when compared with the second-order NDLM. Therefore, we secondly propose a hierarchical normal-Wishart prior on Θ, which allows for unknown and nontrivial correlation structures according to the following model

$$y|z,\Theta \sim \mathcal{N}(f(z,\Theta),\sigma^2),\quad \Theta|\Sigma \sim \mathcal{N}(\mu^0,\frac{1}{q^0}\Sigma),\quad \Sigma^{-1}\sim \mathcal{W}(b^0,\beta^0), \qquad \text{(Model 4)}$$

where $\mathcal{W}(\cdot)$ denotes a Wishart distribution, μ^0 is a known $J\times 1$ vector, q^0 and b^0 are known scalars, and β^0 denotes a $J\times J$ known positive definite matrix. In Model 4, a single observation in a purely sequential setting does not result in conjugacy; thus, the posterior calculation requires time-consuming MCMC simulations.

In summary, assuming an unknown true underlying dose-response curve, a set of admissible doses, and multiple response observations, Model 1 assumes a sigmoid shape for the dose-response curve, whereas posterior calculation requires MCMC simulations. However, Model 1 only keeps track of two variables in our implementation. The correlation structure between responses of different doses is built in the functional form, and the model is susceptible to misspecification when the true dose-response curve is not sigmoid. Model 2 is flexible in that its piecewise linear structure is capable of approximating any dose-response shape while allowing a nontrivial correlation relationship. Posterior estimations of Model 2 require an implementation of a forward filtering backward sampling method, which can be time consuming, even though it might be faster than the MCMC simulation. However, the trade-off is that Model 2 updates $2J$ variables as opposed to two variables in Model 1. A first-order piecewise linear approximation to the dose-response curve reduces this number to J variables, one variable per dose. Model 3 assumes a known correlation structure between the mean responses of different doses, which is its limitation. However, the resulting conjugacy simplifies posterior calculation compared with both Model 1 and Model 2. Finally, Model 4 relaxes the known correlation structure of Model 3 using a hierarchical normal-Wishart prior setup. However, it does not seem to enjoy conjugacy for single sequential observations and thus requires MCMC simulations for posterior estimation.

In presenting Models 1–4, the observation variance is assumed to be known and similar for all doses. However, a simple modification to Equation (1) will allow for possibly different (known) observation variances for each dose in Models 2–4, that is, $y_j = f(z=j,\Theta)+\epsilon_j$, where $\epsilon_j \sim \mathcal{N}(0,\sigma_j^2)$. We next propose a model that extends Model 3 for an unknown observation variance

$$y|z,\Theta \sim \mathcal{N}(f(z,\Theta),\Sigma_{zz}),\quad \Theta|\Sigma \sim \mathcal{N}(\mu^0,\frac{1}{q^0}\Sigma),\quad \Sigma^{-1}\sim \mathcal{W}(b^0,\beta^0), \qquad \text{(Model 5)}$$

which enjoys conjugacy for an observation vector of size J. Also, an approximation to the posterior distribution of the hyperparameters in Model 5 is developed for a single observation setting, which reduces the posterior calculation computational time significantly. However, a drawback of Model 5 is that the variance of observation is proportional to that of Θ, which might not be the case in practice. In reporting the results of this study, our main focus is on presenting a comparison between Model 1 and Model 4. Numerical experiments for Model 5 are presented in OS 7.7.

4. Dynamic Programming Formulation Under Model 3

Assume that in the course of the trial, a total of N homogeneous patients are sequentially assigned to different doses and their responses are observed before the assignment of the next patient. The DM starts with a prior belief of the unknown dose-response curve and chooses assignment doses to ultimately reduce the uncertainty about ED_{95}. Once a patient observation becomes available, the current belief of the underlying dose-response curve is updated and the DM chooses the dose assignment for the next patient. For the underlying dose-response model, we choose

Model 3 because its conjugate property results in a formulation that is amenable to analytical investigation. The DP formulation assuming dose-response Models 1, 2, 4, and 5 are presented in OS 2, 3, 4, and 5, respectively.

Given Model 3, we have

$$y_j^{n+1} = \theta_j + \epsilon_j^{n+1}, \qquad j = 1,\dots,J, \quad n = 0\dots,N-1, \tag{3}$$

where y_j^{n+1} denotes the response of patient $n+1$ assigned to dose j and $\epsilon_j^{n+1} \sim \mathcal{N}(0,\sigma^2)$. Let z^n denote the dose assigned to patient $n+1$. Note that conditional on Θ and z^n, the sampled observation $\hat{y}^{n+1}$ has a normal distribution $(\hat{y}^{n+1}|\Theta,z^n) \sim \mathcal{N}(\theta_{z^n},\sigma^2)$. Recall that in Model 3, we assume a multivariate normal prior on the belief about Θ, that is,

$$\Theta \sim \mathcal{N}(\mu^0,\Sigma^0), \tag{4}$$

where μ^0, a $J\times 1$ vector, can be thought of as the initial belief of the mean response at each dose informed by Phase I, whereas Σ^0, a $J\times J$ positive semidefinite matrix, can reflect the DM's belief about the initial correlation between different doses. Define filtration $\mathcal{F}^n$ as the sigma-algebra generated by prior information, sampling doses, and corresponding responses by decision epoch n, that is, $\mathcal{F}^n$ is a sigma-algebra generated by $\mu^0,\Sigma^0,z^0,\hat{y}^1,z^1,\hat{y}^2,z^2,\dots,z^{n-1},\hat{y}^n$. Notice that z^0 denotes the assignment dose to the first patient before observing any response.

4.1. State and Action Space

Set decision epochs at times when a dose is assigned to a patient where $n=0,\dots,N-1$. No decision is made at time N. The state of the system at decision epoch n is the current belief on the dose-response curve, which is captured by the posterior distribution on Θ given $\mathcal{F}^n$. Defining $\mu^n=\mathbb{E}[\Theta|\mathcal{F}^n]$ and $\Sigma^n=\mathrm{Cov}[\Theta|\mathcal{F}^n]$ in addition to the conjugacy of normal likelihoods and priors will allow us to summarize the state of the system at decision epoch n by the tuple (μ^n,Σ^n). Therefore, the state space can be written in the following form:

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

where $\mathbb{S}_+^J$ denotes the set of $J\times J$ positive semidefinite matrices. An action is described by the dose assigned to patient $n+1$, that is, z^n. Therefore, the action space is equivalent to the set of admissible doses $\mathcal{Z}$.

4.2. Transitions

Our prior belief on Θ is a multivariate normal distribution. In addition, sample observations $\hat{y}^{n+1}$ are normally distributed. Therefore, the posterior distribution on Θ, which is specified by μ^{n+1} and Σ^{n+1} is also a multivariate normal distribution. The relationship between the prior and posterior is characterized by state s^n, action z^n and the random response $\hat{y}^{n+1}$. Assuming that the covariance matrix Σ^n is nonsingular for now, μ^{n+1} and Σ^{n+1} can be written as

$$\begin{aligned}\mu^{n+1} &= \Sigma^{n+1}((\Sigma^n)^{-1}\mu^n + (\sigma^2)^{-1}\hat{y}^{n+1}e_{z^n}),\\ \Sigma^{n+1} &= ((\Sigma^n)^{-1} + e_{z^n}(\sigma^2)^{-1}e_{z^n}^\top)^{-1},\end{aligned} \tag{5}$$

where e_{z^n} is a J-vector of zeroes and a single one at j^{th} index assuming $z^n=j$. This formulation only holds when Σ^n is positive-definite and invertible; however, notice that $e_{z^n}(\sigma^2)^{-1}e_{z^n}^\top$ only changes one element of matrix $(\Sigma^n)^{-1}$. Using the Sherman-Morrison formula to adjust the inverse of a matrix when only one element has changed, formulation (5) can be written in such a way that Σ^n is positive semidefinite and no longer needs to be invertible, that is,

$$\begin{aligned}\mu^{n+1} &= \mu^n + \frac{\hat{y}^{n+1}-\mu_j^n}{\sigma^2+\Sigma_{jj}^n}\Sigma^n e_j^\top,\\ \Sigma^{n+1} &= \Sigma^n - \frac{\Sigma^n e_j e_j^\top \Sigma^n}{\sigma^2+\Sigma_{jj}^n},\end{aligned} \tag{6}$$

where it is assumed that $z^n=j$. Define $\tilde{\sigma}$ as a vector-valued function $\tilde{\sigma}(\Sigma,z^n=j) := \frac{\Sigma e_j}{\sqrt{\sigma^2+\Sigma_{jj}}}$, and note that $\mathrm{Var}[\hat{y}^{n+1}-\mu^n|\mathcal{F}^n] = \mathrm{Var}[\theta_{z^n}+\epsilon^{n+1}|\mathcal{F}^n] = \sigma^2+\Sigma_{jj}^n$. Define random variable $X^{n+1} := \frac{(\hat{y}^{n+1}-\mu^n)}{\sqrt{\mathrm{Var}[\hat{y}^{n+1}-\mu^n|\mathcal{F}^n]}}$ by which formulation (6) is equivalent to

$$\begin{aligned}\mu^{n+1} &= \mu^n + \tilde{\sigma}(\Sigma^n,z^n)X^{n+1},\\ \Sigma^{n+1} &= \Sigma^n - \tilde{\sigma}(\Sigma^n,z^n)\tilde{\sigma}^\top(\Sigma^n,z^n),\end{aligned} \tag{7}$$

where random variable X^{n+1} is standard normal when conditioned on $\mathcal{F}^n$.

4.3. Objective Function

In response-adaptive dose-finding trials, identifying the target dose, for example, ED_{95}, is considered amongst the ultimate goals. The DM must choose a sequence of dose assignments such that learning the target dose is achieved quickly and accurately. A common practice in dose-finding trials is to minimize the variance of the target dose (see, e.g., Krams et al. 2003, Berger and Wong 2009, Lenz et al. 2015, Holm Hansen et al. 2017). In statistics literature, such an approach is called D-optimal design, which is widely used in optimal design of experiments (O'Quigley et al. 2017, chapter 13). The reason for this choice of objective is that reducing the variance of a random variable corresponds to better estimation and confidence, which is the primary factor in experimental design (Chow and Pong 2016, chapter 10). More importantly, this approach in experiment design has support from regulatory authorities in both the United States and Europe.

The European Medicines Agency (2014) qualifies variance minimization for model-based design of uncertain Phase II dose-finding studies. In addition, the U.S. Food and Drug Administration (2018) notes that a pragmatic approach to adaptive randomization may be based on the minimization of test statistics to improve efficiency of the design.

Therefore, in order to learn the target dose, we set the expected cost at the end of the trial to be $\mathrm{Var}(\mathrm{ED}_{95}|s^N)$ where $s^N = (\mu^N, \Sigma^N)$ under Model 3. At each time, the allocation dose z^n is allowed to depend on samples by time n, that is, z^n is $\mathcal{F}^n$-measurable. Define $\Pi := \{(z^0, \ldots, z^{N-1}) : z^n \in \mathcal{F}^n\}$ to be the set of measurement policies where $\pi = (z^0, \ldots, z^{N-1})$ is an element in Π. Let $l_\pi(s^0)$ denote the expected variance of ED_{95} at the end of the trial when the initial prior on Θ is $\mathcal{N}(\mu^0, \Sigma^0)$. Choosing a policy that minimizes the expected cost is achieved by solving $V(s^0) = \inf_{\pi\in\Pi} l_\pi(s^0)$, where $l_\pi(s^0) = \mathbb{E}^\pi\{\mathrm{Var}^N[\mathrm{ED}_{95}]|s^0 = (\mu^0, \Sigma^0)\}$, $\mathbb{E}^\pi\{\cdot\}$ indicates an expectation taken with respect to probability distribution imposed by a fixed measurement policy π and $\mathrm{Var}^N(\cdot)$ is the variance with respect to $\mathcal{F}^N$. Defining a sequence of value functions at each decision epoch $n \le N-1$ as $V^n(s^n)$, the optimal value function is the solution to the Bellman equation

$$V^n(s^n) = \min_{z^n} \left\{\mathbb{E}\left\{V^{n+1}(s^{n+1})\,|\,s^n, z^n\right\}\right\},$$
$$V^N(s^N) = \mathrm{Var}\left(\mathrm{ED}_{95}\,|\,s^N\right). \tag{8}$$

One may also consider to maximize the probability of correctly identifying ED_{95} at the end of the trial as an objective. We consider minimizing the variance of ED_{95} at the end of the trial as our main objective and report the probability of correct selection as another performance measure. We also implement an adaptive randomization allocation as a benchmark in our numerical analysis in Section 6, which is motivated by the probability of correct selection.

5. Formulation Analysis and Proposed Approximate Solution

Under Model 3, we show a few structural properties of our learning problem. In this section, these properties are presented under Model 3; however, some can be extended for Model 5 as well. Define $\mathcal{Q}^n(s,z) := \mathbb{E}[V^{n+1}(\eta(s^n, z^n, X^{n+1}))|s^n = s, z^n = z]$ for any $s \in \mathcal{S}$ as a function measuring the value of assigning dose $z^n = z$ to patient $n+1$ when the trial is in state s^n. Note that $\eta(\cdot)$ is a transition function by which the next state is determined via Equation (7), that is, $s^{n+1} = \eta(s^n, z^n, X^{n+1})$. Denote $V^{n+1}(s^n)$ as the value of making no measurements while in state s^n. The following theorem states that the optimal policy always prefers to make a measurement. In other words, measuring any dose will result in a better estimate of ED_{95}.

Theorem 1. *The optimal policy always prefers to measure an alternative dose rather than to measure nothing at all, that is, $Q^n(s,z) \le V^{n+1}(s)$ for every $s \in \mathcal{S}$, $0 \le n < N$ and $z \in \{1, \ldots, J\}$.*

Theorem 1 shows that any extra measurement would be beneficial (not worse) to the value function at time n. Proof of Theorem 1 is given in OS 6. The following corollary suggests that the extra measurement should be made according to the optimal policy. The proof is presented in OS 6.

Corollary 1. *For all states $s \in \mathcal{S}$, $V^n(s) \le V^{n+1}(s)$.*

Corollary 2 states that there is no value in measuring a dose that is already known (its variance is zero). Therefore, the optimal policy avoids measuring a dose that does not provide information on the curve. Proof of Corollary 2 is presented in OS 6.

Corollary 2. *Let i and j denote any two doses where $i \ne j$, $n < N$, and $s = (\mu, \Sigma)$. If $\Sigma_{jj} = 0$, then $Q^n(s,i) \le Q^n(s,j)$.*

5.1. Approximate Solution

Solving formulation (8) to optimality is impractical because of the state space continuity. We propose a one-step look-ahead framework (KG) in which the DM assumes that the next decision epoch will be the last and allocates a dose to the next patient in order to minimize the expected value of a single period decision process. This setup will remain similar in general for all the dose-response approximation models in Section 3; see OS 2, 3, 4, and 5 for details. Note that $\mathrm{Var}^n(\mathrm{ED}_{95})$ is the value we would receive if we were to stop the trial at decision epoch n. The KG policy chooses a decision to minimize the expected variance of the target dose with respect to the posterior state variables, that is, maximize the information gained. Define the KG policy π^{KG} for every $s \in \mathcal{S}$ formally according to

$$\mathcal{X}^{\pi^{KG}}(s) \in \arg\min_z \mathbb{E}_n\Big\{\mathrm{Var}^{n+1}(\mathrm{ED}_{95}) - \mathrm{Var}^n(\mathrm{ED}_{95})\,\Big|\,s^n = s, z^n = z\Big\} \text{ for every } n < N, \tag{9}$$

where $\mathcal{X}^{\pi^{KG}}(s^n)$ is a decision function that returns the dose selected in state s^n under the KG policy π^{KG}, that is, $\mathcal{X}^{\pi^{KG}}(s^n) := z^n$. In order to compute the KG policy, one needs to evaluate

$$\min_z \mathbb{E}_n\{\mathrm{Var}^{n+1}(\mathrm{ED}_{95})\,|\,s^n = s, z^n = z\}$$

for every $s^n \in \mathcal{S}$ at each decision epoch n because $\mathrm{Var}^n(\mathrm{ED}_{95})$ is constant with respect to $\mathcal{F}^n$. To that end, we apply Algorithm 1 at each decision epoch. Algorithm 1 describes a measurement policy where the next dose assignment is selected in such a way to

minimize the variance of ED_{95}. In particular, given the input information consisted of the state variable $s^n = (\mu^n, \Sigma^n)$ and the set of admissible doses $\mathcal{Z}$ at decision epoch n, Algorithm 1 returns the single period optimal assignment dose. Note that after each patient has been assigned a dose at the end of the algorithm, its response is observed before the next assignment; the observation is used to update the approximate posterior parameters according to Equation (7). Moreover, $\tilde{y}, \tilde{\Theta}, \hat{\mu}, \hat{\Sigma}, \hat{\Theta}$, and ED_{95} are temporary and are discarded at the end of each decision epoch in Algorithm 1.

Algorithm 1 (Knowledge Gradient Dose Selection Policy Under Model 3)

Input: s^n, M, T; Output: z^n.
for each dose $z \in \mathcal{Z}$ **do**
-Generate M samples of $\tilde{\Theta}$ from $\mathcal{N}(\mu^n, \Sigma^n)$.
for each sampled $\tilde{\Theta}_{(m=1:M)}$ **do**
-Simulate future observation $\tilde{y}_{zm} \sim \mathcal{N}(\tilde{\theta}_{zm}, \sigma^2)$.
-Using $\tilde{y}_{zm}$, update (μ^n, Σ^n) by formulation (7) to obtain $(\hat{\mu}^n_{zm}, \hat{\Sigma}^n_{zm})$.
-Generate T posterior samples of $\hat{\Theta}$ by sampling from $\mathcal{N}(\hat{\mu}^n_{zm}, \hat{\Sigma}^n_{zm})$
for each sampled $\hat{\Theta}_{(t=1:T)}$ **do**
-Find $g(\hat{\Theta}_{(t)}) = \min_z\{z \in \mathcal{Z} : f(z, \hat{\Theta}_{(t)}) \geq 0.95 f(z_{max}, \hat{\Theta}_{(t)})\}$
end for
-Estimate the observed variance $U_{zm} = \mathrm{Var}[g(\hat{\Theta})| \mathcal{F}^n \cup (z, \tilde{y}_{zm})]$ using sample variance. {$\mathcal{F}^n \cup \sigma(z, \tilde{y}_{zm})$ denotes a sigma-algebra generated by $\mu^0, \Sigma^0, z^0, y^1, \ldots, z^{n-1}, y^n, z, \tilde{y}_{zm}$.}
end for
-Calculate expected variance U_z for each dose by taking a Monte Carlo sample average $\sum_m \frac{U_{zm}}{M}$.
end for
-Select the dose z^* that minimizes U_z.

5.2. Consistency

A measurement policy is "consistent" if it is able to learn the truth perfectly in the limit. In a response-adaptive dose-finding study, learning the true value of ED_{95} is achieved only if the true underlying dose-response relationship is known in the limit. Consistency of some sampling policies has been studied in the literature. For example, reinforcement learning algorithms that force the measuring policy to explore all alternatives infinitely many times are consistent. However, in most response-adaptive designs, it is difficult to ensure that all alternatives are sampled frequently often to prove consistency. Frazier et al. (2008, 2009) derived the consistency conditions for a class of KG policies, with independent and correlated normal prior beliefs in ranking and selection problems. Furthermore, Frazier and Powell (2011) provided a general set of sufficient conditions for consistency for a broad class of sequential sampling policies. These methods do not directly apply to our setup because our objective function is minimizing the variance of ED_{95}. Proof of the following theorem is presented in detail in OS 6.

Theorem 2. *Under Model 3, and assuming independent responses, the KG policy for response-adaptive dose-finding clinical trials is consistent.*

6. Numerical Results

This section presents the results of our numerical experiments assuming that the underlying true but unknown dose-response curve is of sigmoid shape. Similar results for other dose-response curves are presented in OS 7.1. One may interpret our numerical setup as applying a Bayesian policy in frequentist settings, where there is a true unknown dose-response curve.

6.1. Benchmark Policy

The policy described by Equation (9) in Section 4 assigns a dose in order to reduce the uncertainty about ED_{95} at the end of the trial, that is, reduce the estimate of its variance at the end of the trial. In reporting the results, this policy is considered to be the primary dose allocation method implemented in all of our numerical experiments; we call it "minVar." For a benchmark policy, we implement another randomization policy with a different objective. The adaptive randomization policy described by Carlin et al. (2010), chapter 4, attempts to maximize the probability of correctly identifying the target dose at the end of the trial in order to reduce the uncertainty about it. In fact, the adaptive randomization policy, hereafter "maxProb," applies the one-step look-ahead framework for the following Bellman equation

$$V^n(s^n) = \max_{z^n}\{\mathbb{E}\{V^{n+1}(s^{n+1})|s^n, z^n\}\}, \quad n = 0, 1, \ldots, N-1$$
$$V^N(s^N) = \max_j\{\mathbb{P}(ED_{95} = j|s^N)\}.$$

The maxProb policy is implemented in a similar fashion to Algorithm 1, that is, the same one-step look-ahead framework is used to find the assignment dose. However, instead of evaluating the variance of ED_{95}, we estimate $\mathbb{P}(ED_{95} = j|s^n)$ for assigning each dose j by Monte Carlo.

6.2. Simulation Initialization

Our underlying true dose-response curve is sigmoid and is made up of 11 doses, which are placed equidistant. Typically, this number in Phase II trials is between 4 and 12 doses (Berry et al. 2002). Aligned with the literature (e.g., ASTIN trial of Krams et al. 2003), the first dose is considered as a placebo with its

response marking the baseline score for the treatment in the trial. A simulation model is developed to assess the performance of the dose-response approximation Models 1–4 under two dose allocation policies minVar and maxProb. At each decision epoch in the simulation, a patient arrives at the trial and is given a dose. The dose assignment is dependent on two factors: (i) the posterior belief about the dose-response curve under each of the four dose-response approximation models, and (ii) the choice of the adaptive policy. The patient response is then generated from the true dose-response curve and is added to the data. This new observation is used to update the posterior belief about the dose-response curve according to the dose-response approximation model. Algorithm 1 describes this dose assignment procedure for Model 3 and minVar policy in details. Similar algorithms implementing the same logic for Models 1, 2, and 4 are presented in OS 2, 3, and 4, respectively.

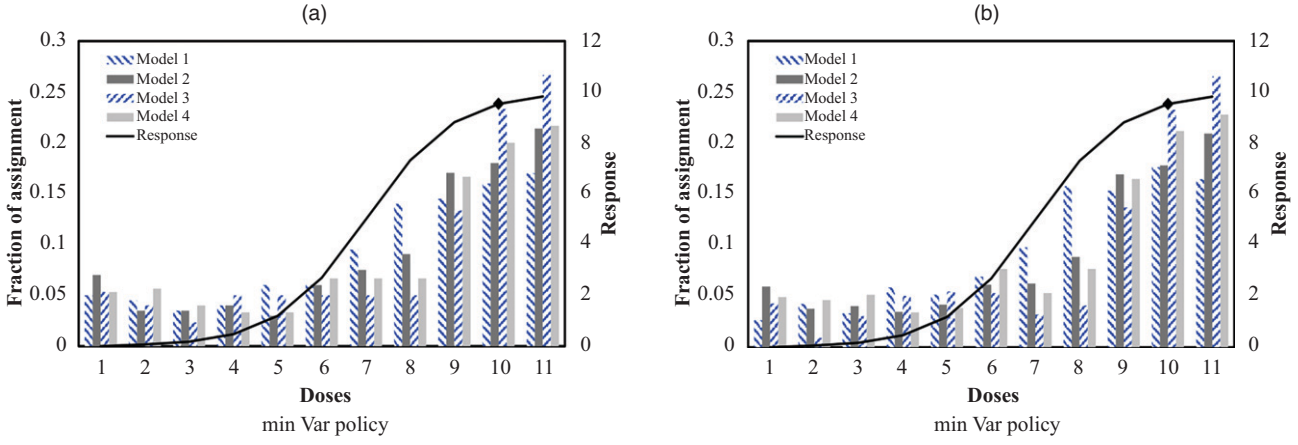
In simulating each dose-response approximation model and adaptive policy, sample size parameters M and T (in Algorithm 1) are set to 500 and 1,000, and a sequence of 300 patients is used in reporting the results. A thinning factor of five is considered for random variable generation in each run where every fifth randomly generated number was used in the simulation to avoid serial correlation in the sequence of random numbers. In reporting each performance measure, 30 simulations with a different sequence of random numbers are considered. Confidence intervals in reporting the performance metrics are calculated with respect to these 30 simulations for a significance level of 0.05. The simulations are coded in R and are run on an Intel core i7 3.7 GHz processor with 16 GB of RAM. The codes that include the input data (choice of hyperparameter initial values) are available in the authors GitHub repository at https://github.com/AmirAli-N/DynamicProgramming-DoseFinding.

Each of the dose-response approximation models introduced in Section 3 have their own set of hyperparameters to be initialized at the start of the simulation. Various choices of priors for these hyperparameters exist in the literature and for some, such as Wishart distribution, the choice of prior is itself an active area of research which is beyond the scope of our analysis. Therefore, we use the clinical trial context and assume that preliminary results are available from Phase I. Assuming that a Phase I clinical trial is run before Phase II and the information of 33 patients, that is, three per each dose, is available, we initialize the hyperparameters of Models 1–4. Here, we only include hyperparameter initialization of Model 3 where μ^0 is the sample average of Phase I responses and Σ^0 is the sample covariance matrix. See OS 2, 3, and 4 for details about the choice of hyperparameters for Models 1, 2, and 4, respectively. The variance of observation is fixed to 10 units for all doses.

6.3. Results

We first present the fraction of patients assigned to each dose under different dose-response approximation models and allocation rules. In particular, Figure 3(a) and (b) show the patient assignment pattern for Models 1–4 under minVar and maxProb adaptive policies, respectively. The horizontal axis shows the dose indices, the left-hand side vertical axis denotes the proportion of patients assigned to a dose, while the solid line represents the true dose-response curve, with response scores on the right-hand side vertical axis. Figure 3 shows that Model 3 and Model 4 outperform other dose-response approximation models in terms of the fraction of assignments to the true ED_{95}. In fact, Model 3

Figure 3. (Color online) Patient Assignments to Dose-Response Curves



Notes. (a) minVar policy, (b) maxProb policy. The diamond identifies the true ED_{95}.

Table 1. Number of Assignments to the True ED_{95} Dose

Allocation policy	Model 1	Model 2	Model 3	Model 4
minVar	48	54	70	60
maxProb	53	54	69	64

Notes. Total number of patients is 300. Fractional numbers are rounded up.

which is the simplest approximation has the highest fraction of patients assigned to ED_{95}. Also, the results are consistent with respect to the dose allocation policies, that is, for both minVar and maxProb we observe a similar trend. Table 1 also confirms this result by reporting the number of patients assigned to the true ED_{95}.

In Table 2, two measures of uncertainty about the true ED_{95} at the end of the trial are reported. Variance of the target dose at the end of trial for either minVar or maxProb policies is given by Σ_{jj} where $j = ED_{95}$. The probability of correct selection shows the probability of correctly identifying the true ED_{95} at the end of the trial for each dose-response approximation model and allocation policy. To evaluate the probability of correct selection after the responses of all 300 patients were observed, the final posterior evaluation of the state variable is used to generate 1000 dose-response curves. For each curve, ED_{95} is evaluated and the fraction of curves whose ED_{95} matched the true ED_{95} is reported. The confidence intervals reported in Table 2 are calculated with respect to 1,000-sample average. These results also confirm the insight from Figure 3 and Table 1. The minVar policy achieves a slightly better variance of the target dose and the maxProb policy achieves a slightly better probability of correct selection.

Although type I/II error analyses are constructed for frequentist settings, some translation of them may be applicable in Bayesian settings as well (Carlin et al. 2010). In evaluating type I/II errors, we assume that the null hypothesis denotes a flat dose-response curve, and the alternate hypothesis represent a curve with at least one dose showing a clinically meaningful advantage over placebo. This is also the approach taken by Berry et al. (2002) and Smith et al. (2006) in their Bayesian implementation. A clinical meaningful ED_{95} is identified only when $\mathbb{P}(|\theta_{ED_{95}} - \theta_1| \geq \Delta) \geq p$, where p and Δ are design parameters. These parameters may be determined by expert opinion or can be thought of as parameters to control for in order to achieve reasonable type I/II error. Note that type I/II errors in this case are evaluated with respect to two underlying dose-response curve: a flat curve to evaluate type I error and the sigmoid curve in Figure 3 to assess the power of the hypothesis test. We follow Yin et al. (2012) in setting up a framework to control for a difference level Δ that achieves reasonable type I/II error rates. Table 3 shows type I error and power rates for $\Delta = 1$ and $p = 0.8$ with respect to the minVar allocation policy. For more details on calculating type I/II errors and selection of (Δ, p), see OS 7.5. These results suggest that with careful selection of (Δ, p), adaptive Bayesian approaches such as those presented in this study can achieve acceptable levels of type I/II error rates which is a regulatory necessity for drug approval.

6.4. Computational Time

A motivating factor for this study is the development of computationally efficient dose assignment frameworks in adaptive dose-finding trials. As mentioned earlier in Section 2, the difficulty in implementation and significant computational requirements are major impediments in applying the NDLM framework (Model 2) in practice. For example, Holm Hansen et al. (2017) reported that an implementation of Model 2 in their case took approximately two months of processing time for 10,000 MCMC runs to evaluate posterior updates and 100 observation simulations per dose to decide on the next dose assignment. In all of our numerical experiments that need MCMC simulation, the number of simulated observation is $M = 1{,}000$ and that of the MCMC runs is 2,500 to evaluate posterior distributions. Table 4 shows that Model 3 is significantly more efficient because it does not require the time-consuming MCMC simulation. Moreover, Model 2 is slightly better than Model 1 and Model 4 because it uses the forward filtering backward sampling method. Note that the results in Table 4 are reported for simulating a single decision epoch along 30 sample paths for the minVar dose allocation policy.

6.5. Additional Numerical Results and Sensitivity Analysis

OS 7.1 and 7.2 present similar results to Table 2 and Figure 3 for other dose-response curves, which include a bell-shaped curve, a nonmonotonic curve, and

Table 2. Uncertainty About the True ED_{95} at the End of Trial

	Probability of correct selection (%)				Variance of the target dose			
Allocation policy	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
minVar	80	83.2	84.5	83.9	0.79	0.65	0.38	0.60
maxProb	82.0	85.3	88.0	84.9	0.80	0.65	0.45	0.65

Note. Probabilities and variances are within ± 0.5 and ± 0.1 for a 95% confidence interval, respectively.

Table 3. Type I Error and Power Rates

	Model 1	Model 2	Model 3	Model 4
Type I (flat)	0.05	0.04	0.02	0.04
Power (sigmoid)	87.4	90.2	95.1	88.0

Note. Results are reported within ±0.02 for a 95% confidence interval.

a flat one. OS 7.3 presents figures demonstrating the trend in estimation of both allocation policies' objective functions, that is, the variance of the target dose and probability of correct selection, as the trial progresses. In OS 7.4, we assess the quality of solution with respect to another performance metric that attempts to measure the L^2 distance between the estimated dose-response curve and the true underlying dose-response curve over the trial's progression. Finally, in OS 7.8, we show that Model 2 is very sensitive to the parameter choice to the extent that it may produce nonconsistent results under the minVar policy.

6.6. A Clinical Trial Case Study

In this section, we repeat our numerical experiments with a new underlying dose-response curve from a real clinical trial by Eli Lilly and Company (2018) on the effects of LY2951742 in participants with mild to moderate osteoarthritis knee pain where four doses at 5 mg, 50 mg, 120 mg, and 300 mg were allocated to patients in the following fashion: $\frac{1}{3}$ were assigned to placebo and the other four doses were equally assigned to participants. We use the credible intervals reported in their study to retrospectively estimate θ_z, the true dose-response curve, and σ_z^2, the true observation variance for dose z. Therefore, the true underlying dose-response curve in this experiment is estimated by the results of Eli Lilly and Company (2018), and we do not mean that it is the "truth." Figure 4 shows the assignment pattern of LY2951742 for placebo and the four doses described. For a sense of completion, we also report the actual patient assignment pattern of the case study. In addition, Table 5 reports the number of assignments to the true ED_{95}, the expected variance of ED_{95} at the end of the trial, and the probability of correct selection. In this dose-response curve, the difference between the highest and the lowest response is small; however, Model 3 still outperforms all other dose-response approximation models in terms of the number of patients assigned to the true ED_{95}. We did not consider Model 1 because its sigmoid functional form is far from the estimate of the dose-response curve depicted in Figure 4. Because the doses in this case study are not equidistant, Model 2 requires slight modifications for implementation; the details of which are presented in OS 7.6.

Table 4. Computational Time for a Single Dose Assignment and 30 Samples

	Model 1	Model 2	Model 3	Model 4
Computational time (in hours)	2.9	2.8	0.16	2.9

7. Practical Considerations and Limiting Assumptions

Our results show that the first-order dose-response approximation Model 3 is computationally appealing for adaptive dose assignment decisions while being competitive or even superior to the standard approach in several metrics. However, these results are based on some assumptions that may not hold true in a real-world clinical trial. We discuss some important issues in this section.

7.1. Known Observation Variance

This limiting assumption is embedded in Equation (1). However, it may be partly justified because in most Phase II clinical trials, some estimates of the observation variance may be available from Phase I. In addition, we perform a sensitivity analysis on the observation variance and show that the results remain consistent in terms of the assignment pattern and variance of the target dose; see OS 7.7. Furthermore, we propose Model 5, which considers an unknown observation variance and enjoys properties that result in efficient computation of its posterior distribution. We compare Model 5 with Model 2 and Model 4, which can be easily extended to unknown variance case, in terms of the assignment pattern, probability of correct selection, and variance of the target dose: see OS 7.7 for more details.

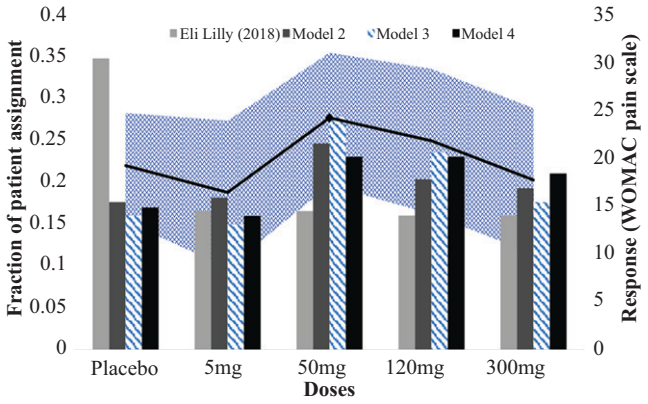
7.2. Batch Assignment

In some clinical trials, it might be necessary or desirable to assign multiple doses to multiple patients at the same time. This extension can be accommodated within our DP framework regardless of the dose-response approximation model. One heuristic solution is that instead of a one-step look-ahead policy, the approximate solution for multiple assignments at decision epoch n involves simulating the algorithms, for example, Algorithm 1, several steps into the future to select multiple doses. However, developing a sophisticated policy for this setting is beyond the scope of this study because of the unique feature of the objective.

7.3. Delays or Interim Responses

We assume that the response of a patient becomes available before the arrival of the next patient. However, the response may become available with delay. Although this feature does not impact the dose-response model approximation, it poses a significant

Figure 4. (Color online) Patient Assignments to the Dose-Response Curve



Notes. The diamond identifies the true ED_{95}. The hashed area represents the credible intervals reported in the trial, and the solid black line shows our estimate of the true dose-response curve.

challenge for updating and constructing the DP framework, especially when the delays may change over time. In addition, although the response of a patient may be available after a delay, an interim response (surrogates) might be observable. This is typically handled by a longitudinal model where, given the interim observation, a full observation is estimated and is used to update the model. Furthermore, interim analysis may be conducted to stop the trial because enough evidence is gathered that the drug is not effective. For optimal stopping decisions for adaptive dose-finding trials, see Nasrollahzadeh and Khademi (2020).

7.4. Patient Covariates

Recall that we assume a pool of homogeneous patients. However, in some dose-finding trials, patient characteristics, such as age, gender, and preexisting conditions, might be taken into account in both dose allocation and dose-response approximation. Addressing this issue requires introducing a model to capture the dependency between covariates and response, which we leave for future studies.

7.5. Dosing Scenario

Some dose-finding Phase II trials are looking for a dosing scenario instead of a single target dose whereby each patient is treated with a particular dosing scenario, for example, increasing or decreasing in dosage during the course of treatment. The structure of these trials is completely different from the one we consider in this study.

7.6. Normality of Response

The efficient solutions produced by Model 3 rely on the conjugacy of the likelihood function and prior, which only holds true when the response follows a normal distribution. Although this is a reasonable assumption for Phase II clinical trials with continuous responses, relaxing it would take away the observed numerical efficiency.

8. Conclusion

In this study, we developed a framework to identify the target dose in sequential dose-finding clinical trials that is capable of incorporating different dose-response approximation models. In fact, we proposed several dose-response approximation models and compared the implications of using each of them via simulation in terms of dose assignment patterns, probability of correctly identifying the true target dose, and variance of the target dose at the end of the trial. These dose-response approximation models have different levels of flexibility, design purposes, and limitation. We presented a formal DP formulation for the sequential dose assignment problem and derived analytical results for our novel learning problem under Model 3.

Table 5. End of Trial Results

Outcomes	Model 2	Model 3	Model 4
Probability of correct selection (%)	81.63	88.34	80.99
Variance of the target dose	0.87	0.51	1.35
Number of assignments to the true ED_{95}	46	52	43

Notes. Total number of patients is 187. The number of assignments are rounded up, and the rest of results are reported within a ± 0.5 for a 95% confidence interval. The number of assignments to the true ED_{95} in the trial was 31, and the variance of ED_{95} was 7.68 at the end of the trial.

Our results show that it might be beneficial to lose some flexibility in nonparametric models in order to gain efficiency in terms of the computational time required for selecting a dose assignment. In particular, our results suggest that assuming a known correlation structure between different doses (which might be reasonable because of preliminary Phase I data) results in significant reduction in the required time for dose assignment without compromising solution quality.

Acknowledgments

The authors thank the department editor, associate editor, and four anonymous reviewers for their helpful comments.

References

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

Berger MP, Wong WK (2009) *An Introduction to Optimal Designs for Social and Biomedical Research*, vol. 83 (John Wiley & Sons, New York).

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

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

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

Chow SC, Pong A (2016) *Handbook of Adaptive Designs in Pharmaceutical and Clinical Development* (CRC Press, Boca Raton, FL).

Dette H, Bretz F, Pepelyshev A, Pinheiro J (2008) Optimal designs for dose-finding studies. *J. Amer. Statist. Assoc.* 103(483): 1225–1237.

Eli Lilly and Company (2018) A study of LY2951742 in participants with mild to moderate osteoarthritis knee pain. Accessed October 1, 2020, https://clinicaltrials.gov/ct2/show/results/NCT02192190.

Frazier PI, Powell WB (2011) Consistency of sequential Bayesian sampling policies. *SIAM J. Control Optim.* 49(2):712–731.

Frazier PI, Powell WB, Dayanik S (2008) A knowledge-gradient policy for sequential information collection. *SIAM J. Control Optim.* 47(5):2410–2439.

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

Frühwirth-Schnatter S (1994) Data augmentation and dynamic linear models. *J. Time Ser. Anal.* 15(2):183–202.

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

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

Gupta SS, Miescke KJ (1996) Bayesian look ahead one-stage sampling allocations for selection of the best population. *J. Statist. Planning Inference* 54(2):229–244.

Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J (2014) Clinical development success rates for investigational drugs. *Nature Biotechnology* 32(1):40–51.

Hennessey VG, Rosner GL, Bast RC Jr, Chen MY (2010) A Bayesian approach to dose–response assessment and synergy and its application to in vitro dose–response studies. *Biometrics* 66(4):1275–1283.

Holm Hansen C, Warner P, Parker RA, Walker BR, Critchley HO, Weir CJ (2017) Development of a Bayesian response-adaptive trial design for the Dexamethasone for excessive menstruation study. *Statist. Methods Medical Res.* 26(6):2681–2699.

Johnstone RH, Bardenet R, Gavaghan DJ, Mirams GR (2016) Hierarchical Bayesian inference for ion channel screening dose-response data. *Wellcome Open Res.* 1–6.

Kotas J, Ghate A (2016) Response-guided dosing for rheumatoid arthritis. *IIE Trans. Healthcare Systems Engrg.* 6(1):1–21.

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, (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.

Lenz RA, Pritchett YL, Berry SM, Llano DA, Han S, Berry DA, Sadowsky CH, Abi-Saab WM, Saltarelli MD (2015) Adaptive dose-finding phase 2 trial evaluating the safety and efficacy of ABT-089 in mild to moderate Alzheimer disease. *Alzheimer Disease Associated Disorders* 29(3):192–199.

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

Nasrollahzadeh AA, Khademi A (2020) Optimal stopping of adaptive dose-finding trials. *Service Sci.* 12(2-3):80–99.

National Institutes of Health (2014) Notice of revised NIH definition of clinical trial. Accessed October 1, 2020, http://grants.nih.gov/grants/guide/notice-files/NOT-OD-15-015.html.

O'Quigley J, Iasonos A, Bornkamp B (2017) *Handbook of Methods for Designing, Monitoring, and Analyzing Dose-Finding Trials* (CRC Press, Boca Raton, FL).

Parizi MS, Ghate A (2016) Lot-sizing in sequential auctions while learning bid and demand distributions. *2016 Winter Simulation Conf. (WSC)* (IEEE, Piscataway, NJ), 895–906.

Powell WB, Ryzhov IO (2012) *Optimal Learning*, vol. 841 (John Wiley & Sons, New York).

Press WH (2009) Bandit solutions provide unified ethical models for randomized clinical trials and comparative effectiveness research. *Proc. Natl. Acad. Sci. USA.* 106(52):22387–22392.

Rosenberger WF (1996) New directions in adaptive designs. *Statist. Sci.* 11(2):137–149.

Smith MK, Jones I, Morris MF, Grieve AP, Tan K (2006) Implementation of a Bayesian adaptive design in a proof of concept study. *Pharmaceutical. Statist.* 5(1):39–50.

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

The European Medicines Agency (2014) Qualification opinion of MCP-Mod as an efficient statistical methodology for model-based design and analysis of Phase II dose finding studies under model uncertainty. Accessed October 1, 2020, https://www.ema.europa.eu/en/committees/committee-medicinal-products-human-use-chmp.

Tufts (2014) Cost to develop and win marketing approval for a new drug is $2.6 billion. Accessed October 1, 2020, http://csdd.tufts.edu/news/complete_story/pr_tufts_csdd_2014_cost_study.

U.S. Food and Drug Administration (2017) The drug development process: Clinical research. Accessed October 1, 2020, https://www.fda.gov/ForPatients/Approvals/Drugs/ucm405622.htm.

U.S. Food and Drug Administration (2018) Adaptive designs for clinical trials of drugs and biologics: Draft guidance for industry. Accessed October 1, 2020, https://www.fda.gov/downloads/drugs/guidances/ucm201790.pdf.

Villar SS, Bowden J, Wason J (2015) Multi-armed bandit models for the optimal design of clinical trials: Benefits and challenges. *Statist. Sci.* 30(2):199–215.

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

Wang Y, Wang C, Powell W (2016) The knowledge gradient for sequential decision making with stochastic binary feedbacks. Lawrence N, Reid M, eds. *Internat. Conf. Machine Learn.* (PMLR, New York), 1138–1147.

Weir CJ, Spiegelhalter DJ, Grieve AP (2007) Flexible design and efficient implementation of adaptive dose-finding studies. *J. Biopharmaceutical Statist.* 17(6):1033–1050.

West MJ, Harrison PJ (1997) *Bayesian Forecasting and Dynamic Models* (Springer-Verlag, New York).

Yin G, Chen N, Jack Lee J (2012) Phase II trial design with Bayesian adaptive randomization and predictive probability. *J. Roy. Statist. Soc. Ser. C: Appl. Statist.* 61(2):219–235.