Service Science

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

Adaptive Design of Personalized Dose-Finding Clinical Trials

Saeid Delshad, Amin Khademi

To cite this article:
Saeid Delshad, Amin Khademi (2022) Adaptive Design of Personalized Dose-Finding Clinical Trials. Service Science 14(4):273-291. https://doi.org/10.1287/serv.2022.0306

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

SERVICE SCIENCE
Vol. 14, No. 4, December 2022, pp. 273–291
ISSN 2164-3962 (print), ISSN 2164-3970 (online)

Adaptive Design of Personalized Dose-Finding Clinical Trials

Saeid Delshad,[a,*] Amin Khademi[a]

[a] Department of Industrial Engineering, Clemson University, Clemson, South Carolina 29631
*Corresponding author
Contact: sdelsha@clemson.edu, https://orcid.org/0000-0002-7676-8927 (SD); khademi@clemson.edu, https://orcid.org/0000-0002-5281-8715 (AK)

Received: July 5, 2021
Revised: February 5, 2022; May 28, 2022
Accepted: June 10, 2022
Published Online in Articles in Advance: July 21, 2022

https://doi.org/10.1287/serv.2022.0306

Abstract. A key and challenging step toward personalized/precision medicine is the ability to redesign dose-finding clinical trials. This work studies a problem of fully response-adaptive Bayesian design of phase II dose-finding clinical trials with patient information, where the decision maker seeks to identify the right dose for each patient type (often defined as an effective target dose for each group of patients) by minimizing the expected (over patient types) variance of the right dose. We formulate this problem by a stochastic dynamic program and exploit a few properties of this class of learning problems. Because the optimal solution is intractable, we propose an approximate policy by an adaptation of a one-step look-ahead framework. We show the optimality of the proposed policy for a setting with homogeneous patients and two doses and find its asymptotic rate of sampling. We adapt a number of commonly applied allocation policies in dose-finding clinical trials, such as posterior adaptive sampling, and test their performance against our proposed policy via extensive simulations with synthetic and real data. Our numerical analyses provide insights regarding the connection between the structure of the dose-response curve for each patient type and the performance of allocation policies. This paper provides a practical framework for the Food and Drug Administration and pharmaceutical companies to transition from the current phase II procedures to the era of personalized dose-finding clinical trials.

Funding: This research is supported by the National Science Foundation [Grant 1651912].
Supplemental Material: The online appendices are available at https://doi.org/10.1287/serv.2022.0306.

Keywords: adaptive dose-finding • clinical trials • precision medicine • one-step look-ahead policy

1. Introduction

1.1. Research Motivation

Personalized/precision medicine is widely believed to be the future of medicine and the most promising path toward higher quality of care (Bates 2010). Recently, a significant amount of research has been devoted to precision medicine highlighting the importance of considering patients' personal differences in treatment selection and finding the right therapeutic dose (Hayden 2015). That is all due to the growing evidence that as a result of disregarding personalization, many commonly prescribed treatment procedures and medications are ineffective or even potentially harmful to some patients (Schork 2015). In the broad literature of precision medicine, the goal of this study is to design adaptive trials to find dose levels in situations where the "right" dosage may depend on a set of already-identified patient characteristics.

The main motivation for this work is to address the growing need for designing innovative methods for personalized dose-finding trials. Peck (2021) defined precision dosing as the process of individually tailoring the dosage to have the greatest treatment benefit and the least health risk for different patient types. Maxfield and Zineh (2021, pp. 1505–1506) emphasized the necessity of precision dosing, which maximizes the balance between the benefits and potential risks at the level of individual patients. They highlighted the most important steps toward this goal by emphasizing on "optimizing dosing in patient subpopulations during drug development, understanding the determinants of response variability in patient populations, and enabling clinical practice for treatment decisions at the patient level." Peck (2018) urged practitioners to consider all relevant personal aspects of a patient in specifying optimal dose of a drug, emphasizing that the only way to achieve personalization is to redesign dose-finding trials. Using the advanced search tool on *clinicaltrials.gov*, as of May 2022, there are over 3,800 studies that are generally working on some personalized treatment or precision medicine, from which 200 trials are actively focused on dose-finding studies (U.S. National Library of Medicine—National Institutes of Health 2021).

The goal of these dose-finding trials is usually to identify an *effective target dose* corresponding to a certain level of drug efficacy. Motivated by toxicity and side effects concerns, the target dose, denoted by ED_L, is defined as the minimum dose that achieves at least

$L \times 100\%$ $(0 < L \leq 1)$ of the maximal response. In our numerical results, similar to Berry et al. (2002), we focus on identifying $ED_{0.95}$ as the target dose, defined as the minimum dose, which achieves 95% efficacy of the maximal-response dose. One key element to identify ED_L is to estimate the dose-response curve, which we will explain next.

1.2. Proposed Framework

Motivated by the abovementioned evidence that optimal dosing may depend on patient characteristics, this study relaxes the patient homogeneity assumption of Berry et al. (2002). In particular, we consider a setup where the dose-response curve may be a function of patient covariates. We assume that the set of patient covariates (types) is finite, discrete, low dimensional (in a sense that will be discussed later), prespecified, and observable. Moreover, we assume that all covariates are predictive (see Section 9 for discussions on predictive versus prognostic covariates). Because we consider patient covariates, the target doses for different patient types may be quite variant. For example, the problem instances in Figure 1 could represent the response behavior of two distinct patient types to a single drug for which their target dose is different. Figure 1(a) shows an instance of a dose-response curve in which $ED_{0.95}$ and the maximal-response dose coincide at dose 3, whereas Figure 1(b) represents another instance in which $ED_{0.95}$ occurs at dose 4; the maximal response is met at dose 5.

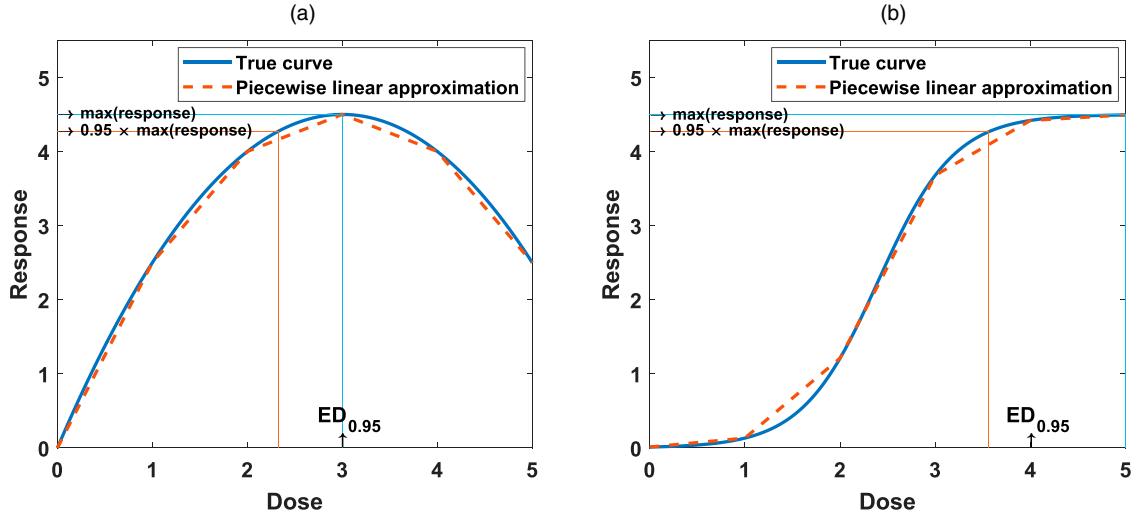
In fact, a key component of our setup is to construct a dose-response model in the presence of patient covariates. Such a model should (i) be flexible to incorporate a wide range of dose-response shapes, (ii) result in a low-dimensional dynamic programming formulation for analytical investigation, and (iii) incorporate the correlation between covariates and doses because, for instance, one may expect that assigning patients with related covariates to closer doses would result in similar responses. To that end, we propose a first-order normal dynamic linear model (NDLM) to approximate the response of each dose for a given patient type, where the mean response at each dose is a linear model. Although the NDLM is nonparametric in approximating the dose-response curve, our approach is semiparametric as we are considering linear models for the mean responses at each dose.

1.3. Model Properties and Distinction

First, the proposed model is flexible in characterizing a wide range of dose-response curves. Because for any given covariate, we use a first-order NDLM, which is a nonparametric model. In fact, for each pair of dose and patient type, we estimate a mean response and then by connecting the points of doses and corresponding mean responses, we construct a piecewise linear curve (presented by red dashed lines in Figure 1) as an approximation for the true dose-response curve (presented by the continuous blue curves in Figure 1) for each patient type. The motivation for such a construction is due to the results in Nasrollahzadeh and Khademi (2021), where they showed that the first-order NDLM is a competitive and robust approach to model dose-response curves.

Second, the dynamic programming formulation for the sequential allocation problem based on the proposed model is low dimensional. This is because by assuming a multivariate normal prior on the unknown model parameter Θ and normal responses, the posterior distribution on Θ will be multivariate normal. Therefore, the state space will be the vector of posterior mean and correlation matrix instead of the set of all probability distributions for the unknown parameter Θ. Although the state space is still multidimensional and continuous, the

Figure 1. Examples of Dose-Response Curves and Their Piecewise Linear Approximation in Dose-Finding Trials



dynamic programming formulation is amenable to analytical investigation.

Third, the proposed model is able to incorporate the correlation among doses and covariates. This is because we can construct the prior on $\Theta \sim \mathcal{N}(\mu^0, \Sigma^0)$ such that it specifies a given correlation. For example, if we initially believe that for a covariate x, the dose-response shape is bell shaped, we can initialize the segment in μ^0, which corresponds to x to have a bell-shaped trend. Or if we believe that closer doses have similar responses, we can construct Σ^0 such that closer doses have a higher correlation. For details regarding the choice of prior, see Sections 4 and 7.

Hence, the sequence of events in each decision epoch in our fully sequential setup is as follows. At the beginning of each epoch, a patient arrives to the trial. The covariates of the arrived patient is a realization of an independent and identically distributed multivariate random variable with a known distribution. A dose is allocated based on all the information available to the decision maker (DM) so far, and the patient response becomes available at the beginning of the next epoch. The allocation decisions are based on the objective of reducing the uncertainty regarding the target dose of each patient type. This objective is achieved by minimizing the expected (over patient covariates) variance of the personalized target dose at the end of the trial, which is often used in dose-finding studies. In fact, in early stage trials, the focus is on improving the accuracy of the estimates of the target dose because it will be used later in phase III clinical trials on a large patient population. An incorrect target dose choice may result in severe consequences in later stages: Low doses expose patients to nontherapeutic dosages of drugs and high doses expose patients to excessive toxicity, both resulting in failure for regulatory approval. See Section 4 for more on the choice of the objective function.

1.4. Paper Structure

We formulate this sequential sampling problem as a stochastic dynamic program (Section 3) by which we show a few natural properties of the learning problem (Section 6). Because the state space is multidimensional and continuous, the problem is not amenable to exact solutions. Instead, we apply the one-step look-ahead framework to our formulation in general (Section 5) and provide a closed-formed decision rule with two doses under the assumption that the belief about alternative doses are independent. In addition, in the impersonalized version of the problem with two doses (where we consider homogeneous patients), we provide some insights on the sampling behavior in the short term and show that the one-step look-ahead policy is optimal. We show that the asymptotic sampling behavior under the variance minimization in equivalent to that of knowledge gradient (KG) and optimal computing budget allocation (OCBA).

Finally, we propose four benchmarks each of which is an adaptation of available sampling policies in the literature of adaptive dose-finding clinical trials to our setting. In particular, we consider uniform randomization, greedy approach, allocation based on posterior sampling, and posterior predictive sampling. We implement our proposed policy and the benchmarks with synthetic (Section 7) and real (Section 8) data sets. In particular, we show the performance of the proposed policies under a variety of problem settings and performance measures. Our results show that the one-step look-ahead policy performs robustly. Our numerical result for the real case study of warfarin also sheds light on the connection between the structure of the dose-response curve for different patient types and the performance of tested policies. We also measure the value of considering the proposed personalized model with this data set. Section 9 discusses practical limitations. Furthermore, all the proofs, detailed algorithms, and additional numerical results are respectively provided in Online Appendices A, B, and C.

2. Related Literature

In terms of theory, there are two streams of literature related to our work: (i) contextual multiarmed bandits (CMBs) and (ii) response-adaptive design of clinical trials.

In CMB problems, a DM observes a context at each epoch and chooses an action that produces a reward that is a function of the action and the context observed plus, possibly, a random noise: See Agarwal et al. (2014) for details. There is a fundamental difference between the structure of our problem and that of standard CMB: Our objective is to minimize the expected (over covariates) variance of a target dose, which is a nonlinear function of unknown parameters, whereas the objective of standard CMB is to minimize the cumulative (or instant) regret. Goldenshluger and Zeevi (2013) proposed a greedy algorithm with occasional forced exploration for the contextual linear bandit problem. Bastani et al. (2021) built on this work and proposed a greedy-first algorithm. In a Bayesian setting and using the notion of Bayes regret, Russo and Van Roy (2014) investigated the regret of posterior sampling and its connection to UCB (upper confidence bound) algorithms.

Another class of related literature in stochastic learning is R&S in which the DM has a limited budget to explore the expected value of several alternatives and recommends the best at the end. In the Bayesian R&S (ranking and selection), a one-step look-ahead policy, called knowledge gradient, has shown success in several settings (Frazier et al. 2008, Ryzhov et al. 2012, Wang and Powell 2018). Ding et al. (2022) incorporated

covariates in a standard R&S and showed the consistency of KG. Negoescu et al. (2011) studied the problem of choosing the right chemical formula for a drug where they considered a number of attributes to represent the different chemical derivatives (alternatives) of a drug. They used the Free-Wilson model for the response of each alternative formula assuming that the existence of each attribute in the chemical formula of the drug has an additive contribution. Their special linear model considered binary attributes. Although our proposed model is similar to that model, there is a subtle difference between these models. Adapting the notation, for an alternative z, the mean response in their model is $\theta_1 x_1^z + \theta_2 x_2^z + \ldots + \theta_K x_K^z$, whereas that for our model is $\theta_1^z x_1 + \theta_2^z x_2 + \ldots + \theta_K^z x_K$ for a given covariate $x = (x_1, x_2, \ldots, x_K)$, that is, the covariates in their model determine the alternatives and the unknown parameters do not correspond to the alternatives, and that is because their covariates represent chains of atoms in a molecule specifying different alternative formulas. In our model though, given a covariate, the alternatives may have different mean responses as the unknown parameters depend on the alternatives. Furthermore, we develop a one-step look-ahead policy tailored to our setting and analyze its properties and show its optimality in a specific setting. In fact, our setup is structurally different, that is, we seek to minimize the expected variance of the target dose, whereas the objective in standard R&S is to maximize the posterior mean at the end of the trial.

Bayesian sequential design of clinical trials has received attention in operation management. For example, Ahuja and Birge (2016) studied the problem of multiple patient assignment in a two-armed trial. Williamson et al. (2017) used dynamic programming to propose a randomized allocation to maximize the number of patient successes and penalize it if the number of assignments to a dose is less than a threshold. Kotas and Ghate (2018) developed a response-guided dosing model to individualize treatment for cohorts of patients with evolving conditions over their treatment course. Incorporating patient covariates and the choice of objective function makes our setting different from these previous studies. Finally, Rojas-Cordova et al. (2020) reviewed the impact of applying sequential adaptive designs in clinical trials, highlighting the advantages and challenges of adaptive decision-making procedures.

In regard to sampling policies, adaptive design of dose-finding trials has received significant attention from statistics and biostatistics community. We review some relevant work where patient information is incorporated into the design of clinical trials. O'Quigley et al. (1999) designed a sampling process for phase I clinical trials where the goal is to find the maximum tolerable dose, where patients are divided into two groups. As for the adaptive allocation, their idea was to apply a version of the Continual Reassessment Method, which at each epoch selects the dose that its estimated probability of toxicity is the closest to a target probability. In fact, their allocation approach is greedy. Guo and Yuan (2017) designed a Bayesian personalized two-stage phase I/II clinical trial where at each epoch they first selected a set of important biomarkers and then used a utility function to represent dose desirability with respect to toxicity and efficacy. They applied adaptive randomization (AR) where the probability of selecting each dose was proportional to its posterior expected utility. Liu et al. (2018) expanded on the previous work by adding immune response to the utility function for a case of personalized cancer treatment. At each sampling epoch, they first found the admissible dose range and then used an AR allocation policy to randomize cohorts (one or more) of patients to each dose with respect to its posterior probability of being optimal, that is, having the highest posterior mean utility. Lin et al. (2019) also worked on adaptive phase I/II clinical trials; but in their case, they considered multiple subgroups of patients and delayed responses. Their allocation is hence different from the aforementioned literature as they assign patients from each subgroup to dose-schedule regimes. To make the allocation decision, they used a sequential procedure to first find an admissible dose regime for each subgroup and then allocated each patient to dose-schedule regimes based on an AR policy, where they assign each patient to each regime proportional to their posterior predictive probability of that regime being the best (having the maximum mean utility) within each subgroup.

3. Problem Formulation

We consider a fully sequential setup where at each epoch a patient arrives to the trial with a covariate vector generated from a known multivariate distribution; the DM assigns a treatment dosage after observing patient covariates; and the patient response is observed before the next epoch when a new patient arrives. Let $\mathcal{Z} := \{1, 2, \ldots, Z\}$ refer to the finite set of allowable doses (Z being the cardinality of $\mathcal{Z}$) for all patients. Note that each element of set $\mathcal{Z}$ may be the index referring to a specific dosage of an active drug. For example, dose 1 may refer to 5 mg of the drug, dose 2 may refer to 10 mg of the drug, and so on. In dose-finding trials, Z is usually between 2 and 10 (Berry et al. 2002).

Specifically, patients are heterogeneous in our setup, for example, they may have different genders, races, ages, genetic features, biomarkers, health conditions, and living environments. We assume that the DM is interested in identifying the target dose for each set of patient covariates. For example, in cancer clinical trials, the investigators are interested in identifying the right dose for patients with specific given biomarkers. We call those attributes that affect patient responses covariates. We consider a covariate vector of size K to represent each patient. In particular, let x_k

for $k = 1, 2, \ldots, K$ be the k^{th} covariate and $\mathcal{C}_k$ be the finite set of all values that covariate x_k can take. For example, if x_2, the second covariate, represents gender, then $\mathcal{C}_2 = \{0, 1\}$ where $x_2 = 0$ means that the second covariate represents the "male" and $x_2 = 1$ means that the second covariate represents the "female." Therefore, each patient is completely characterized by a $1 \times K$ vector $x = (x_1, x_2, \ldots, x_K)$. Let $\mathcal{X} := \{x = (x_1, x_2, \ldots, x_K) : x_k \in \mathcal{C}_k, k = 1, 2, \ldots, K\}$ be the finite set of all possible patient covariate vectors (patient types). We set $x_1 = 1$ aligned with the literature on linear models to create an intercept. We assume that in any epoch, the covariate vector x is realized with probability $\mathcal{P}_x$ (where $\sum_{x \in \mathcal{X}} \mathcal{P}_x = 1$) independent of everything else. Without loss of generality, we assume that for any $x \in \mathcal{X}$, we have $\mathcal{P}_x > 0$. Suppose that at epoch $n \in \{0, 1, \ldots, N-1\}$, a patient with the covariate vector $x^n = (x_1^n, x_2^n, \ldots, x_K^n)$ arrives where x^n is completely observable before making the sampling decision. We assume that $\mathcal{X}$ is low dimensional in the following sense: If we assume that x_ks are binary, in order to have stable estimates for linear models, in a frequentist setting, we need to have the number of covariates be less than or equal to the sample size and $(X^n)^\top X^n$ being invertible, where $X^n := [x^1, x^2, \ldots, x^n]^\top$ and $^\top$ denotes matrix transpose. One condition is that the distinct number of patient types is bounded by $\log_2(N-1)+1$ and the other $(X^n)^\top X^n$ being invertible. For example, in dose-finding trials with sampling budget $n = 100$, our setup can handle up to seven distinct patient types. Generally though, in the Bayesian setting, our algorithms always work as long as the initial prior covariance is positive definite.

Next, we describe the state space of the problem. To that end, we use a first-order normal dynamic linear model to characterize the dose-response curve. That is, for each dose z, we assume that the response of a patient with covariate vector x is normally distributed with unknown mean $\langle x, \Theta_z \rangle$ (because vector Θ_z is unknown) and a possibly dose-dependent known variance σ_z^2, where $\langle \cdot, \cdot \rangle$ denotes inner product. Specifically, in this setup, we assume that the mean response for dose z is a linear function of covariates with an unknown K-dimensional parameter column vector $\Theta_z = (\theta_{z,1}, \theta_{z,2}, \ldots, \theta_{z,K})^\top$. Therefore, $\Theta = (\Theta_1, \Theta_2, \ldots, \Theta_Z)^\top$ is the $ZK \times 1$ stacked parameter vector.

We follow a Bayesian setup in that we assume a multivariate normal prior distribution for Θ, which measures the DM's initial belief about the unknown parameter Θ. Denote μ^0 as the $ZK \times 1$ initial mean vector and Σ^0 as the $ZK \times ZK$ initial covariance matrix of Θ in the beginning of the experiment, that is, $\Theta \sim \mathcal{N}(\mu^0, \Sigma^0)$. Elements in Θ may be correlated, that is, Σ^0 may be nondiagonal. This structure captures correlation across the unknown parameters $\theta_{z,k}$, $\forall z, k$ corresponding to different covariates and alternative doses. This modeling feature is important because one may expect that the doses that are close to each other have close responses for a given patient type. Note that we consider continuous patient responses, and the normality assumption is quite standard in dose-finding studies with continuous response. In fact, the primary endpoint is assumed to be normal for a majority of clinical trials in the literature (Julious 2004); for example, see Liu et al. (2017).

Because the prior for Θ is multivariate normal and the patient response is also normally distributed, we have a conjugate model so that the posterior on Θ is also multivariate normal and the posterior parameters will be the state of the system. We next formalize the posterior and present a procedure for calculating the posterior given the patient covariate, dose assigned, and observed response. To that end, let $\delta_z^x = (0, 0, \ldots, x, \ldots, 0)$ be a $1 \times ZK$ sparse vector, where the $1 \times K$ dimensional row vector x is placed in the z^{th} component, and each zero is a $1 \times K$ row vector of zeros. Therefore, the mean response of a patient with covariate x assigned to dose z, which is $\langle x, \Theta_z \rangle$, can be represented by $\langle \delta_z^x, \Theta \rangle$.

Upon the arrival of the patient in the n^{th} epoch with covariate vector $x^n = x$, the DM assigns a dose $z^n = z$ to the patient. Let $y_z^{n+1} \in \mathbb{R}$ be the observed response. Recall that the response of the patient follows a normal distribution given by

$$\left(y_z^{n+1} \mid z, \Theta, x\right) \sim \mathcal{N}(\langle \delta_z^x, \Theta \rangle, \sigma_z^2), \qquad z \in \mathcal{Z}, n = 0, \ldots, N-1, \tag{1}$$

where σ_z^2 is the sampling variance of dose z.

We define $s^0 := (\mu^0, \Sigma^0)$ as our initial belief about Θ. Define $\mathcal{F}^n$ as the sigma-algebra generated by the set $\{\mu^0, \Sigma^0, x^0, z^0, y_{z^0}^1, \ldots, x^{n-1}, z^{n-1}, y_{z^{n-1}}^n\}$. Denote $\mathbb{P}_n(\cdot) := \mathbb{P}(\cdot \mid \mathcal{F}^n)$, $\mathbb{E}_n(\cdot) := \mathbb{E}(\cdot \mid \mathcal{F}^n)$, and $\mathrm{Cov}_n(\cdot) := \mathrm{Cov}(\cdot \mid \mathcal{F}^n)$ respectively as the probability, expectation, and covariance with respect to $\mathcal{F}^n$. Assuming a multivariate normal prior distribution about the parameter vector Θ and normally distributed responses, the posterior at epoch n will have a multivariate normal distribution $\Theta \mid \mathcal{F}^n \sim \mathcal{N}(\mu^n, \Sigma^n)$. Thus, we let $s^n = (\mu^n, \Sigma^n)$ be the state of the trial at epoch n, in which μ^n is the $ZK \times 1$ mean vector and Σ^n is the $ZK \times ZK$ covariance matrix of Θ at epoch n. Denote set $\mathcal{S}$ as the state space, that is,

$$s^n \in \mathcal{S} := \left\{(\mu, \Sigma) : \mu \in \mathbb{R}^{ZK}, \Sigma \in \mathbb{M}_+^{ZK \times ZK}\right\}, \tag{2}$$

where $\mathbb{M}_+^{ZK \times ZK}$ denotes the set of $ZK \times ZK$ positive semidefinite matrices.

Next, we describe how the parameters of the posterior distributions can be calculated. Let $z^n = z$ be the dose assigned to the patient at the n^{th} epoch with covariate vector x^n. Then we have the following formula to find the posterior distribution of Θ,

$$\begin{aligned} \Sigma^{n+1} &= \left[(\Sigma^n)^{-1} + \sigma_z^{-2} \delta_z^{x^n \top} \delta_z^{x^n}\right]^{-1}, \\ \mu^{n+1} &= \Sigma^{n+1}\left[(\Sigma^n)^{-1}\mu^n + \sigma_z^{-2} \delta_z^{x^n \top} y_z^{n+1}\right], \end{aligned} \tag{3}$$

for which we denote $\eta(\cdot)$ to be a state transition function, that is, $s^{n+1} = \eta(s^n, z, x^n, \omega_y^{n+1})$, where ω_y^{n+1} represents the randomness in the sample response y_z^{n+1}.

In this study, we consider a DM who is interested in identifying the target dose for each patient type as precisely as possible given a limited number of patients. Let $0 < L \le 1$ and denote ED_L^x as the smallest dose with at least $(L \times 100)\%$ effectiveness of the maximal response for a patient with covariate vector x. Given Θ

$$\mathrm{ED}_L^x := \min\left\{z \in \mathcal{Z} : \langle \delta_z^x, \Theta\rangle \ge L \times \langle \delta_{z_{\max}^x}^x, \Theta\rangle\right\}, \quad \forall x \in \mathcal{X}, \tag{4}$$

where $z_{\max}^x$ is the dose at which the maximal response is observed for the patient with covariate vector x. If $\langle \delta_{z_{\max}^x}^x, \Theta\rangle < 0$ for an $x \in \mathcal{X}$, we let ED_L^x be the smallest dose. This definition is motivated by considering toxicity and side effects, as higher concentrations of an active drug usually produce more severe side effects (Berry et al. 2002). Following a standard approach in optimal design of experiments for clinical trials, in order to increase the precision of the target dose for each patient type, we minimize the expected variance of ED_L^x over all covariates $x \in \mathcal{X}$, that is, patient allocation is derived by minimizing the expected (over covariates) variance of ED_L^x at the end of the trial.

Denote $\mathcal{F}^{n+}$ as the sigma-algebra generated by the set $\mathcal{F}^n \bigcup \{x^n\}$ as all the data available at epoch n. Let $A(s^n) := \mathcal{Z}$ denote the action space in any epoch n. Then we define a decision rule $d^n : \mathcal{S} \to \mathcal{Z}$ that is measurable with respect to the sigma-algebra $\mathcal{F}^{n+}$, which represents the history of states, actions (allocated doses), and covariate vectors observed until time n including the covariate vector of the patient who just arrived at time n. Also, let $\Pi := \{\pi = (d^0, \ldots, d^{N-1})\}$ be the set of measurable nonanticipative policies, where π is an element of Π. Now, let $g : (\Theta, x) \longmapsto z$ be the function that specifies ED_L^x for a patient with covariate vector x. Therefore, the expected variance of ED_L^x at the end of the trial given the initial state s^0 under any given policy π is $l_\pi^N(s^0) = \mathbb{E}_\pi\{\mathbb{E}_x[\mathrm{Var}(g(\Theta, x) \mid \mathcal{F}^N)] \mid s^0\}$, where $\mathbb{E}_\pi(\cdot)$ and $\mathbb{E}_x(\cdot)$ indicate the expectation taken with respect to the probability measure induced by policy π and covariate vector x, respectively. Thus, having an initial prior s^0, the DM solves for $B(s^0) = \inf_{\pi \in \Pi} l_\pi^N(s^0)$. Let $V^n(s^n)$ denote the value function at epoch n, which is a solution to the following Bellman equations

$$\begin{aligned} V^n(s^n, x^n) &= \min_{z^n \in \mathcal{Z}} \left\{\mathbb{E}\left\{V^{n+1}(s^{n+1}) \mid s^n, z^n, x^n\right\}\right\}, \ n = 0, 1, \ldots, N-1, \\ V^n(s^n) &= \mathbb{E}_x\{V^n(s^n, x)\}, \ n = 0, 1, \ldots, N-1, \\ V^N(s^N) &= \mathbb{E}_x\left\{\mathrm{Var}(g(\Theta, x) \mid s^N)\right\}, \end{aligned} \tag{5}$$

where $V^n(s^n, x^n)$ denotes the minimum value of being at state s^n at epoch n after a patient with covariate vector x^n is observed, and $V^0(s^0) = B(s^0)$.

Formulation (5) provides a framework for optimal allocation, but the DM has to recommend an ED_L^x for each covariate vector x at the end of trial. Our recommendation is to estimate ED_L^x by $\widehat{\mathrm{ED}}_L^x$ based on the posterior mean, that is,

$$\widehat{\mathrm{ED}}_L^x := \min\left\{z \in \mathcal{Z} : \langle \delta_z^x, \mu^n\rangle \ge L \times \max_{z \in \mathcal{Z}} \langle \delta_z^x, \mu^n\rangle\right\}, \quad \forall x \in \mathcal{X}, \tag{6}$$

which is aligned with recommendation decisions in R&S and is practical in dose-finding studies (Berry et al. 2002). If $\max_{z \in \mathcal{Z}} \langle \delta_z^x, \mu^n\rangle < 0$, we set $\widehat{\mathrm{ED}}_L^x$ to the minimum dose.

Remark 1. We consider ED_L^x in general for formulation and analyses. However, in Sections 7 and 8, the proposed policies are implemented for $\mathrm{ED}_{0.95}^x$ in all numerical experiments. One can set $0 < L \le 1$ to consider other effectiveness levels, for example, $\mathrm{ED}_{0.5}^x$ and ED_1^x.

4. Discussion on the Choice of the Objective and Prior

The literature on optimal allocation procedures in clinical trials is vast, and different communities have studied it from a variety of perspectives. Our study belongs to the literature on the optimal design of experiments applied for clinical trials. A main objective in optimal design of experiments is to allocate experimental resources to reduce the uncertainty about certain parameters of the statistical model, often achieved by optimizing different utility functions of the variance-covariance matrix of the estimated parameters (Wu and Hamada 2011). In particular, Wald (1943) introduced the minimization of the variance of parameters as a measure of efficiency of statistical investigations. Also, Fisher (1949) related the minimization of variance to maximizing Fisher information, declaring the calculation of variance as a basic way of measuring the amount of information that a random variable carries.

There is a significant literature that uses the optimal design of experiments to find optimal patient allocation in static or adaptive clinical trials in general and dose-finding trials in particular. Specifically, for dose-finding phase II trials, we refer the reader to Dette et al. (2008) from a frequentist and Berry et al. (2010) from a Bayesian viewpoint, where the objective is set to minimize the variance of the target dose. In addition, variance minimization designs are supported by regulatory agencies. For example, the U.S. Food and Drug Administration (FDA) (2019) suggested that the variance can be used to adjust sample sizes according to prespecified algorithms and to ensure the desired power level. Recall that given a sample size, reducing variance improves the power of the statistical tests. The FDA also suggested that interim estimates of

variance can be applied to adaptive sampling utilizing treatment assignment information. Also, the European Medicines Agency Committee for Medicinal Products for Human Use (2014) suggested the minimization of variance as an efficient statistical method for model-based design and analysis of phase II dose-finding studies under model uncertainty.

Furthermore, optimal design of experiment type objective functions is used in operations community. For example, Bhat et al. (2019) studied static and adaptive A-B optimal testing to maximize the precision (inverse of variance) of the treatment effect. Russo and Van Roy (2018) proposed a sampling approach for multiarmed bandit problems to minimize the uncertainty, measured by the entropy, regarding the arm with the highest mean. They called this method "information-directed" sampling; see also Delshad and Khademi (2020). One may interpret our approach as variance-directed sampling. Finally, Bertsimas et al. (2019) studied the problem of allocating patients to group treatments and showed how incorporating patient covariates into clinical trials can improve statistical power. They proposed a computationally efficient covariate-adaptive optimization method that decreased the duration and operating costs of clinical trials while protecting the results against experimental bias.

Next, we discuss the choice of prior for our setting, which is a delicate issue in general for any Bayesian framework. Ildstad et al. (2001) discussed the subjectivity of the prior in the Bayesian clinical trials and the concern of small sampling budgets in phase I and II clinical trials. Their suggested solution was to use data from other similar diseases and treatments that have been previously studied to create reasonable priors for the new drug/treatment. For example, they mentioned how for the problem of loss of bone mineral density during spaceflight, data from earlier spaceflights, and studies of osteoporosis in immobilized individuals could provide a strong basis for development of prior distributions. Berry et al. (2010) have also generally discussed the choice of prior for clinical trials. They described the prior distribution in the Bayesian setting as the reflection of information that includes the investigator's understanding of the biology of the disease and historical/preclinical results of related treatments and, therefore, the prior distributions are specific to the investigator and might not be accepted by anyone else. Guo and Yuan (2017, p. 512) stated that "for the purpose of early phase trial designs, a desirable prior should be sufficiently regularized (or informative) so that the design and model estimates are reasonably stable throughout the trial, while also being vague enough so that the accumulating data can rapidly dominate the prior as the trial proceeds." They generated random prior variances following a gamma distribution with a large variance to create weakly informative priors. Similarly, Liu et al. (2018) used a gamma distribution, with a mean that was suggested by specialized clinicians and a large variance to generate vague priors. They stated that they did not require any accurate prior estimates, as the accumulating data dominate the vague prior and guide dose transition within a few samples. Later in Section 7, we will explain how we create priors based on Guo and Yuan (2017) and conduct sensitivity analyses based on prescriptions proposed by Berry et al. (2010). For constructing priors of KG-type policies, Chick et al. (2021) created two types of priors: robust and tilted. They are based on the idea that having large values for prior mean encourages exploration as well as increasing the prior variance. Similarly, we ensure that the prior is vague enough.

5. Proposed Policy and Benchmarks

Finding an optimal solution of Equation (5) is intractable because the state space is continuous and multidimensional and, therefore, standard solution techniques do not apply. However, Equation (5) helps us with developing a one-step look-ahead policy where the DM assumes that the next epoch is the last one. Our proposed policy, as we call it the Dose-finding One-step Look-ahead (DOL) policy, is different from the standard KG in two major ways: (i) The objective here is to minimize the expected variance of ED_L^x over all x, where the standard KG's objective is to maximize the expected final response. (ii) Our setup involves covariates, and the response model is a linear function of patients' covariates. Given the fact that DOL is stationary, its decision at each epoch only depends on the state of the trial. At any epoch n with the state $s^n = s$ and patient covariate vector $x^n = x'$, DOL minimizes $\mathbb{E}\{V^n(\eta(s,z,x',\omega_y)) - V^n(s)\}$ over all doses $z \in \mathcal{Z}$. Therefore, DOL allocates dose $z^n_{DOL} : (\mathcal{S},\mathcal{X}) \to \mathcal{Z}$ based on

$$z^n_{DOL}(s,x') \in \arg\min_{z\in\mathcal{Z}} \mathbb{E}[\mathbb{E}_x\{\mathrm{Var}(g(\Theta,x))\}\,|s^n = s, z^n = z, x^n = x'], \tag{7}$$

where ties are broken randomly. To that end, we use Monte Carlo simulation in which at each epoch, given the state and the patient visited, we create a pool of sample responses by allocating to every dose $z \in \mathcal{Z}$. Next, given each of these sample responses, we find the next state. Then, we use another Monte Carlo inside each replication of the first one to estimate the variance of ED_L^x for each $x \in \mathcal{X}$ given sampling each dose. Finally, using the given probabilities $\mathcal{P}_x$, we find the overall expectation of the variance of ED_L^x over all $x \in \mathcal{X}$. Algorithm 1 in Online Appendix B.1 presents the details of implementing DOL.

Next, we describe four benchmark policies used in our study. The first benchmark is the Uniform Allocation (UA) policy in which for each patient the dose for assignment is chosen randomly, where each dose has an equal chance of being chosen. The UA policy is a traditional approach for patient allocation in clinical trials. The second benchmark is the Greedy Allocation (GA), which has been used for patient allocation in dose-finding trials (O'Quigley et al. 1999), and we adapt it to our setting. In particular, at epoch n when a patient arrives with covariate vector x^n, the posterior mean response for each dose is calculated, that is, $\langle x^n, \mu_z^n \rangle$, $\forall z \in \mathcal{Z}$, by which ED_L is identified and assigned to the patient.

The third benchmark is perhaps the most common sampling technique used in Bayesian dose-finding clinical trials. Given a utility function, Posterior Adaptive Sampling (PAS) randomly selects a dose according to the probability it is optimal. It is also referred to as Thomson sampling or probability matching. For instance, Guo and Yuan (2017) applied adaptive randomization where the probability of selecting each dose was proportional to its posterior expected utility. Assuming that our utility function is the variance of the target dose at the end of the trial, we modify PAS to allocate at each epoch a patient with covariate vector x to each dose with respect to the probability that it is ED_L^x. This adaptive sampling approach can be seen as an adaptation of posterior sampling applied for the best arm identification problem. Hence, it involves two main steps: (i) sampling the model and (ii) selecting the action. Because our setup is similar to best arm identification where the best arm (with the maximal response) is replaced by the target dose, we sample from the model but the action selection is to choose the target dose. Formally, at any epoch n with the state s, our PAS allocates the patient with covariate vector x to each dose $z \in \mathcal{Z}$ with probability $\text{w}_z^n := \mathbb{P}_n\{z = \text{ED}_L^x\}$. Algorithm 3 in Online Appendix B.2 elaborates the sampling procedure following PAS. However, note that PAS may fail in general: See Russo et al. (2018, section 8.2) for a discussion.

The fourth benchmark adds more exploration to PAS by allowing sampling from posterior predictive. We call it Posterior Predictive Adaptive Sampling (PPAS), and it only differs from PAS in terms of sampling probabilities w_z^n, where for PPAS we allocate based on the posterior predictive probability of each dose being the target dose. This sampling policy is also used in dose-finding literature (Lin et al. 2019), and we adapt it to our setting. To implement PPAS in our setting, at epoch n, given the patient with covariate vector x^n, we sample $\tilde{y}_z^n$ from the posterior predictive distribution of $\langle \delta_z^{x^n}, \Theta \rangle$ for each dose $z \in \mathcal{Z}$, that is, we sample $\tilde{y}_z^n$ from $\mathcal{N}(\langle \delta_z^{x^n}, \mu^n \rangle, \delta_z^{x^n \top} \Sigma^n \delta_z^{x^n} + \sigma_z^2)$, $\forall z \in \mathcal{Z}$ and then allocate to ED_L^x found in the sampled vector $\tilde{y}^n = [\tilde{y}_z^n : z \in \mathcal{Z}]$. Algorithm 5 in Online Appendix B.3 elaborates the sampling procedure following PPAS.

6. Model and Proposed Policy Analysis

In this section, we show a few natural properties of the learning problem for the case with independent beliefs about the alternative doses. In Section 3, we formulate the personalized dose-finding problem in a general correlated setting, where all the parameters of Θ corresponding to different doses and covariates may be correlated. In this section though, although we still let the parameters corresponding to different covariates be correlated for each dose, we assume that the belief about the doses are independent from each other. To that end, we denote a set of Z many $K \times K$ posterior covariance matrices Σ_z^n, $\forall z \in \mathcal{Z}$, which represent the correlation among covariates for each alternative dose. In other words, Σ^n introduced in Section 3 becomes a block diagonal matrix with Σ_z^n s on the diagonal.

In this setting, we still have a standard linear model for the observations, that is, $y_z^{n+1} = \langle x^n, \Theta_z \rangle + \epsilon_z^{n+1}$, $\forall z \in \mathcal{Z}$, $n \in \{0, 1, \ldots, N-1\}$. However, after taking a sample at epoch n from a specific dose z, we only update our belief about its posterior mean μ_z^n and covariance matrix Σ_z^n , whereas the belief about any other doses will not change. Specifically, for the dose $z^n = z$ sampled at epoch n for the patient with covariate vector x^n, we have

$$\begin{aligned} \Sigma_z^{n+1} &= \left[(\Sigma_z^n)^{-1} + \sigma_z^{-2} x^{n\top} x^n \right]^{-1}, \\ \mu_z^{n+1} &= \Sigma_z^{n+1} \left[(\Sigma_z^n)^{-1} \mu_z^n + \sigma_z^{-2} x^{n\top} y_z^{n+1} \right], \end{aligned} \tag{8}$$

whereas for any other dose $z' \neq z$, we have $\mu_{z'}^{n+1} = \mu_{z'}^n$ and $\Sigma_{z'}^{n+1} = \Sigma_{z'}^n$.

Now define $Q^n(s, z, x') := \mathbb{E}\{V^{n+1}(\eta(s^n, z^n, x^n, \omega_y^{n+1})) \mid s^n = s, z^n = z, x^n = x'\}$ as the "cost" of assigning dose z to a patient with covariate vector x' when the trial is in state s. The next result states that the optimal policy always prefers to measure an alternative dose rather than to measure nothing at all.

Proposition 1. *We have $Q^n(s, z, x') \leq V^{n+1}(s)$ for all $s \in \mathcal{S}$, $x' \in \mathcal{X}$, $z \in \mathcal{Z}$, $n \in \{0, 1, \ldots, N-1\}$.*

As a result of Proposition 1, we have the following corollaries. The first one states that sampling from a dose with a known true response will not change the expected variance of ED_L^x, and the second one implies that the extra measurement always reduces the value function. The function $V^{n+1}(s)$ can be interpreted as the value of no measurement in state $s^n = s$.

Corollary 1. *Let $z, z' \in \mathcal{Z}$ be two different doses; if dose z is known almost surely, that is, $\Sigma_z^n = 0$ for some $n < N$, then $Q^n(s^n, z, x^n) = V^{n+1}(s^n) \geq Q^n(s^n, z', x^n)$.*

Corollary 2. *We have $V^n(s) \leq V^{n+1}(s)$ for all states $s \in \mathcal{S}$.*

Next, we focus on a setting with two alternatives to derive novel insights about optimality and the asymptotic sampling behavior of DOL in the variance

minimization setting. First, in Proposition 2, we show that DOL has a closed-form allocation rule. This result facilitates the computation of DOL and bypasses the need to carry out Monte Carlo simulations.

To that end, given $\mathcal{F}^{n+}$ (which includes x^n) if we sample from dose $z^n = z$, the posterior predictive distribution is $y_z^{n+1} \sim \mathcal{N}(\langle x^n, \mu_z^n \rangle, \sigma_z^2 + x^n \Sigma_z^n (x^n)^\top)$, that is, defining a standard normal random variable $\mathfrak{Z}^{n+1}$, we can write $y_z^{n+1} = \langle x^n, \mu_z^n \rangle + \sqrt{\sigma_z^2 + x^n \Sigma_z^n (x^n)^\top}\, \mathfrak{Z}^{n+1}$. One can show that $\mu_z^{n+1} = \mu_z^n + \left[\sigma_z^{-2} \Sigma_z^{n+1} (x^n)^\top \sqrt{\sigma_z^2 + x^n \Sigma_z^n (x^n)^\top}\right] \mathfrak{Z}^{n+1}$ is a stochastic process with Gaussian increments. Hence, we have $\mu_z^{n+1} \mid (\mathcal{F}^{n+}, z^n = z) \sim \mathcal{N}(\mu_z^n, \tilde{\Sigma}_z^n)$, where

$$\tilde{\Sigma}_z^n = \left[\sigma_z^{-2} + \sigma_z^{-4} x^n \Sigma_z^n (x^n)^\top\right] \Sigma_z^{n+1} (x^n)^\top x^n \Sigma_z^{n+1}, \quad \forall z \in \mathcal{Z}, \tag{9}$$

is the change in the covariance of dose z assuming that we sample from it for a patient with covariate vector x^n (see Online Appendix A.4 for derivations).

Proposition 2. *Seeking the personalized ED_L^x in a set of two doses, $\mathcal{Z} = \{z_1, z_2\}$ with $z_1 < z_2$, the DOL's allocation rule is given by*

$$z_{DOL}^n(s^n, x^n) = \begin{cases} z_1 & \text{if } \sum\limits_{x \in \mathcal{X}} \mathcal{P}_x \Phi(\mathcal{A}_{z_1}^x)[1 - \Phi(\mathcal{A}_{z_1}^x)] \\ & \quad < \sum\limits_{x \in \mathcal{X}} \mathcal{P}_x \Phi(\mathcal{A}_{z_2}^x)[1 - \Phi(\mathcal{A}_{z_2}^x)], \\ z_2 & \text{if } \sum\limits_{x \in \mathcal{X}} \mathcal{P}_x \Phi(\mathcal{A}_{z_1}^x)[1 - \Phi(\mathcal{A}_{z_1}^x)] \\ & \quad > \sum\limits_{x \in \mathcal{X}} \mathcal{P}_x \Phi(\mathcal{A}_{z_2}^x)[1 - \Phi(\mathcal{A}_{z_2}^x)], \end{cases}$$

where $\mathcal{A}_{z_1}^x = \frac{L x \mu_{z_2}^n - x \mu_{z_1}^n}{\sqrt{x \Sigma_{z_1}^{n+1} x^\top + L^2 x \Sigma_{z_2}^n x^\top}}$ and $\mathcal{A}_{z_2}^x = \frac{L x \mu_{z_2}^n - x \mu_{z_1}^n}{\sqrt{x \Sigma_{z_1}^n x^\top + L^2 x \Sigma_{z_2}^{n+1} x^\top}}$, and $\Phi(\cdot)$ is the standard normal cumulative distribution function. In the case of equality, DOL randomizes between the two doses.

As can be seen from Proposition 2, the DOL policy is structurally unique. To make this more clear, the next result provides an intuition about how DOL samples from two doses if we focus on the *impersonalized* dose-finding clinical trials where the patients are homogeneous. For the impersonalized setting, suppose that we have normally distributed measurements for any dose $z \in \mathcal{Z}$ with the unknown mean parameter θ_z, and θ_z given filtration $\mathcal{F}^n$ is normally distributed with posterior mean μ_z^n and posterior variance Σ_z^n, that is, $(y_z^{n+1} \mid z, \theta_z) \sim \mathcal{N}(\theta_z, \sigma_z^2)$ and $\theta_z \mid \mathcal{F}^n \sim \mathcal{N}(\mu_z^n, \Sigma_z^n)$ for all z and n. Note that in this setting, $\tilde{\Sigma}_z^n = \Sigma_z^n - [(\Sigma_z^n)^{-1} + \sigma_z^{-2}]^{-1}$ is the change in the variance of dose z assuming that we sample from it at epoch n (see proposition 3.1 of Frazier et al. (2008) and Online Appendix A.5).

Proposition 3. *For the impersonalized dose-finding problem seeking ED_L in a set of two doses, $\mathcal{Z} = \{z_1, z_2\}$ with $z_1 < z_2$, the DOL's allocation rule is given by*

$$z_{DOL}^n(s^n) = \begin{cases} z_1 & \text{if } \tilde{\Sigma}_{z_1}^n > L^2 \tilde{\Sigma}_{z_2}^n, \\ z_2 & \text{if } \tilde{\Sigma}_{z_1}^n < L^2 \tilde{\Sigma}_{z_2}^n, \end{cases}$$

and, in case of equality, DOL randomizes between the two doses.

Proposition 3 provides a novel insight about the sampling behavior of the one-step look-ahead policy in the variance minimization setting. In particular, Proposition 3 reveals that DOL seeks to balance the weighted posterior variance of the *next epoch* between two alternatives. A special case is $L = 1$ when the DM seeks to find the dose with the maximal response (ED_1). In this case, it can be seen from the allocation rule in Proposition 3 that DOL allocates sampling to the dose that reduces the posterior variance the most. The next result characterizes the optimal policy in the impersonalized setting with two doses. In fact, it reveals that DOL is optimal in this setting with respect to our value function.

Proposition 4. *For the impersonalized dose-finding problem seeking ED_L among two doses, DOL is optimal.*

As can be seen, the structure of the optimal policy that seeks to balance the weighted posterior variances among alternatives is unique to our setting. The next result goes beyond identifying the optimal allocation rule and sheds light on the long-run sampling rate of the optimal policy in this specific setting.

Corollary 3. *For the impersonalized dose-finding problem seeking ED_L in a set of two doses, $\mathcal{Z} = \{z_1, z_2\}$ with $z_1 < z_2$, the optimal asymptotic sampling rate is in proportion to sampling standard deviations, that is, $\frac{n_{z_1}}{n_{z_2}} \longrightarrow \frac{1}{L^2}\left(\frac{\sigma_{z_1}}{\sigma_{z_2}}\right)$ as the total number of samples grows to infinity.*

Based on Corollary 3, if we look for the dose with the maximal response (ED_1), the asymptotic sampling rate will be $\frac{n_{z_1}}{n_{z_2}} \longrightarrow \frac{\sigma_{z_1}}{\sigma_{z_2}}$ for the impersonalized DOL with two doses. This is insightful because Ryzhov (2016) showed that, for two alternatives, the asymptotic sampling rate of OCBA and KG coincide with $\frac{n_{z_1}}{n_{z_2}} \longrightarrow \frac{\sigma_{z_1}}{\sigma_{z_2}}$. Corollary 3 shows that for two alternatives, while the proposed DOL policy is different from OCBA and KG in regard to the allocation rule, its asymptotic sampling behavior is similar.

7. Numerical Analysis

In this section, we present the results of fully Bayesian numerical experiments carried out through Monte Carlo simulations of five different policies including DOL and four benchmark policies, for example, PAS, PPAS, GA, and UA. We first explain the general setup, for example, the prior and the Monte Carlo parameters, as well as the performance measures used to

compare different allocation policies in our experiments. We then set up the base experiment of this section for which we conduct several sensitivity analyses.

7.1. Prior Setup and Performance Measures

For all the experiments, we assume a $ZK \times ZK$ initial prior covariance matrix Σ^0 found by an additive approach similar to Guo and Yuan (2017) that represents a logical relation among doses and covariates and at the same time guarantees a weakly informative prior, which lets the posterior converge by a few samples. Specifically, we assume that each element of the initial covariance matrix $\Sigma^0_{zk,z'k'}$, defined as the initial covariance between unknown parameters $\theta_{z,k}$ and $\theta_{z',k'}$, is found by summing three components: (i) *Base factor* $\gamma_b = b$ is fixed over all the matrix elements and should be large enough to create a weakly informative covariance matrix. The base factor b has to be chosen based on the scale of sampling responses; (ii) *Dose correlation effect* $\gamma_d^{zz'} = b \times e^{-a(z-z')^2}$ provides larger values for doses that are closer to each other. In fact, the factor a relates the magnitude of correlation to the distance between the pair of doses z and z', that is, the larger the value of a, the less correlated the response of doses z and z' will be. In our experiments, because we have 10 doses, we set $a = 0.1$, which results in negligible correlation between doses 1 and 10. (iii) *Covariate correlation effect* $\gamma_c^{kk'} = (\mathbb{1}\{S_x^{kk'} \geq 0\} - \mathbb{1}\{S_x^{kk'} < 0\}) \times b \times e^{|S_x^{kk'}|-1}$ is where $-1 \leq S_x^{kk'} \leq 1$ denotes a measure of similarity between the elements k and k' of the covariate vector x and $\mathbb{1}\{\cdot\}$ is the indicator function. For instance, in the case that covariates k and k' are binary and we have some retrospective data, this measure of similarity could be the Pearson correlation coefficient between the responses observed for two groups of patients, one with covariate $x_k = 1$ and the other with covariate $x_{k'} = 1$. If such data are not available, one may use measures of distance between covariate vectors for $S_x^{kk'}$. Hence, assuming $\Sigma^0_{zk,z'k'} = \gamma_b + \gamma_c^{kk'} + \gamma_d^{zz'}$, our prior covariances are in proportion to the base factor b and all the main diagonal elements of Σ^0 will be equal to $3b$. For simplicity, we set $S_x^{kk'} = 0.5$ and $S_x^{kk} = 1$ and also choose $b = 2$ for all the experiments in this section.

In regard to the prior means, Berry et al. (2010) introduced the *community of several priors* where the investigator considers mainly three different priors: (i) *Enthusiastic*: known as the best case prior or the prior that the investigator finds the most likely, for example, assuming a prior which makes an increasing dose-response relation in phase II clinical trials; (ii) *Skeptical*: known as the worst case prior or the prior that the investigator finds the least likely, for example, assuming a prior which makes a decreasing dose-response relation in phase II clinical trials; (iii) *Reference (flat/noninformative)*: known as the prior that attempts to express no particular opinion about the treatment's merit, for example, assuming the same level of efficacy for the alternative doses in phase II clinical trials.

To compare the performance of different policies, three different performance measures are considered. The first measure is the probability of correct selection (PCS) at each epoch defined as the posterior probability of selection recommendation (found by Equation (6) at that epoch) being equal to true $\text{ED}^x_{0.95}$ for each $x \in \mathcal{X}$ over all replications. Second, the expected opportunity cost (EOC) represents the absolute difference, on average, between the response of the true $\text{ED}^x_{0.95}$ and the response of the $\text{ED}^x_{0.95}$ recommended by the posterior at each epoch for $x \in \mathcal{X}$ over all replications. The last performance measure is the expected variance (EVar) of the target dose $\text{ED}^x_{0.95}$ averaged over all patient types given the posterior belief at each epoch, which corresponds to the value function of our formulation. In this section, we only present the EVar plots and leave the results obtained for PCS, EOC, and the average allocation (to each alternative dose) to Online Appendix C.

For all the experiments in this section, we create a pool of 10,000 random problem instances (true dose-response curves). In order to create the problem instances in each replication of our simulation, each time we sample the true parameter vector for replication r of the simulation from the prior, that is, $\Theta^{(r)} \sim \mathcal{N}(\mu^0, \Sigma^0)$ for $r = 1, 2, \ldots, 10{,}000$. We use the randomly generated problem instances to control the simulation and to find the value of performance measures in each epoch on average. Note that a simulation with 10,000 replications results in the precision of ± 0.01, ± 0.02, and ± 0.04 (with 95% confidence) for PCS, EOC, and EVar plots, respectively. We analyze the performance of each policy for a budget of $n = 100$ allocations (number of patients). Also, in all the experiments, the sampling procedure begins following each policy without any initialization at the beginning of the trial, that is, we implement each policy over the entire sampling budget making sure we use similar random seeds and patient types when simulating different policies.

As for the implementation of DOL policy, setting parameters $M = 100$ and $C = 100$ in Algorithm 1 of Online Appendix B.1, and using an Intel Core i9 processor with 12 cores, 24 threads, and 16.5 MB cache, MATLAB takes almost 70 seconds to allocate 100 doses to the patients following DOL.

7.2. Base Experiment

In the base numerical experiment, we consider the set of alternative doses $\mathcal{Z} = \{1, 2, \ldots, Z\}$, where $Z = 10$ is the number of alternative doses. Also, we consider two patient types represented by a set of covariate vectors $\mathcal{X} = \{(1,0), (1,1)\}$, where there is an equal probability of visiting any patient type at each epoch, that is,

$\mathcal{P}_{(1,0)} = \mathcal{P}_{(1,1)} = 0.5$. We set equal sampling/measurement standard deviations of $\sigma_z = 3$ over all doses $z \in \mathcal{Z}$.

As for the prior in the base experiment, we assume an increasing (namely, enthusiastic) prior mean vector μ^0. Specifically, we assume a $ZK \times 1$ initial prior mean vector $\mu^0 = [\mu_1^0, \mu_2^0, \ldots, \mu_Z^0]^\top$ with $\mu_z^0 = [0.2z - 0.01z^2, 0]^\top$, $\forall z \in \mathcal{Z}$.

Figure 2(a) presents the EVar results for each allocation policy. The additional PCS, EOC, and average allocation plots for the base experiment are presented in Figure 6 of Online Appendix C.1. As can be seen from these results, the PCS and EOC plots are consistent with the EVar results, that is, DOL outperforms other benchmarks in terms of PCS and EOC as well as EVar. Next, we change the parameters and the underlying setup of the base experiment to see how the results obtained for the base experiment are sensitive to different setup values to gain further numerical insight about DOL.

7.3. Sensitivity Analyses for the Base Experiment

We implement a variety of sensitivity analyses for the base experiment. The first test measures the sensitivity of results with respect to the prior from a classical Bayesian perspective. Specifically, we follow Berry et al. (2010) and repeat the base experiment with flat (reference/noninformative) and also skeptical priors. For the flat prior, we set the $ZK \times 1$ initial prior mean vector $\mu^0 = [\mu_1^0, \mu_2^0, \ldots, \mu_Z^0]^\top$ with $\mu_z^0 = [0.5, 0.5]^\top$, $\forall z \in \mathcal{Z}$. For the skeptical prior, we assume a decreasing initial dose-response belief by setting $\mu_z^0 = [0.2(Z - z + 1) - 0.01(Z - z + 1)^2, 0]^\top$, $\forall z \in \mathcal{Z}$. The plots for EVar for the flat and skeptical priors are presented in Figure 2, (b) and (c), respectively. Additional results for these experiments are presented in Figures 7 and 8 of Online Appendix C.2. We can see from these results that in both cases, DOL still performs competitively in terms of different performance measures, but the performance of PAS is closer to DOL compared with the results we have for the experiment with the enthusiastic prior.

We run another sensitivity analysis with respect to the prior, but this time we explore the impact of a misspecified prior. Specifically, we consider a DM who is assuming a flat (reference/noninformative) prior mean response generated by $\mu^0 = [\mu_1^0, \ldots, \mu_Z^0]^\top$ with $\mu_z^0 = [0.5, 0.5]^\top$, $\forall z \in \mathcal{Z}$, in an environment where the instances are generated based on the enthusiastic prior means (increasing initial dose-response beliefs). In other words, we use the same problem instances (true dose-response curves) that we generate for the base experiment to control this experiment but we run the trial each time starting from a flat prior. In Figure 2(d) and Figure 9 of Online Appendix C.3, we see a drop in performance, especially at the beginning of the trial because of this misspecified prior.

The next sensitivity analysis is concerned with sampling/measurement variances. First, we reduce the sampling standard deviations to $\sigma_z = 1$ over all doses $z \in \mathcal{Z}$. Second, we increase the sampling standard deviations to $\sigma_z = 5$ over all doses $z \in \mathcal{Z}$. Figure 2, (e) and (f) plus Figures 10 and 11 of Online Appendix C.4 show their results. Note that the one-step reduction in the variance of the target dose, calculated for DOL allocation decision, depends on the value of sampling variances across alternatives. We can see that in both cases, DOL is still competitive in terms of all performance measures and reaches PCS of close to one and EOC of close to zero overall when we reduce the sampling error. We can also see that decreasing/increasing the sampling variances will consistently result in higher/lower overall performance for all policies. However, in these cases, where we have notably high or low sampling errors, DOL is not significantly better than posterior adaptive sampling methods that are computationally easier to implement.

In another sensitivity analysis, we explore the consequences of considering an independent belief in our policies among doses (our analytical setup) where the environment is correlated (our modeling setup). To that end, we repeat the first experiment in this section for DOL, PAS, and PPAS by implementing a version of these policies that considers independent beliefs about the alternative doses and follows the updates mentioned in Section 6. Although we implement Algorithms 1, 3, and 5 of Online Appendix B for the base experiment, here we implement Algorithms 2, 4, and 6 of Online Appendix B to see how much disregarding the dependency between alternative doses affects the overall performance. Figure 2(g) and Figure 12 in Appendix C.5 show the results. In terms of performance, we see a drop in the performance of all three policies. Looking closer at the PCS and EVar plots, we observe that the performance of these policies do not improve throughout the trial as much as they did in the base experiment. This is to some extent due to different allocations compared with the base experiment and the fact that the belief about the alternative doses stays independent throughout the trial. We note that although for large sampling budgets the cross correlations wash off, in trials with a limited number of samples, ignoring the correlations clearly impacts the performance of sampling policies.

Additionally, we run a sensitivity analysis to see the impact of having more patient types. Specifically, we add two binary covariates with $K = 4$, corresponding to eight different patient types with uniform distribution. Results in Figure 2(h) show that DOL still outperforms the best benchmark for slightly higher-dimensional covariates.

Moreover, we run two experiments, one with a total of $Z = 5$ and the other with $Z = 15$ alternative doses,

Figure 2. Expected Variance of the Personalized Target Dose Averaged Over Both Patient Types in Each Sensitivity Analysis

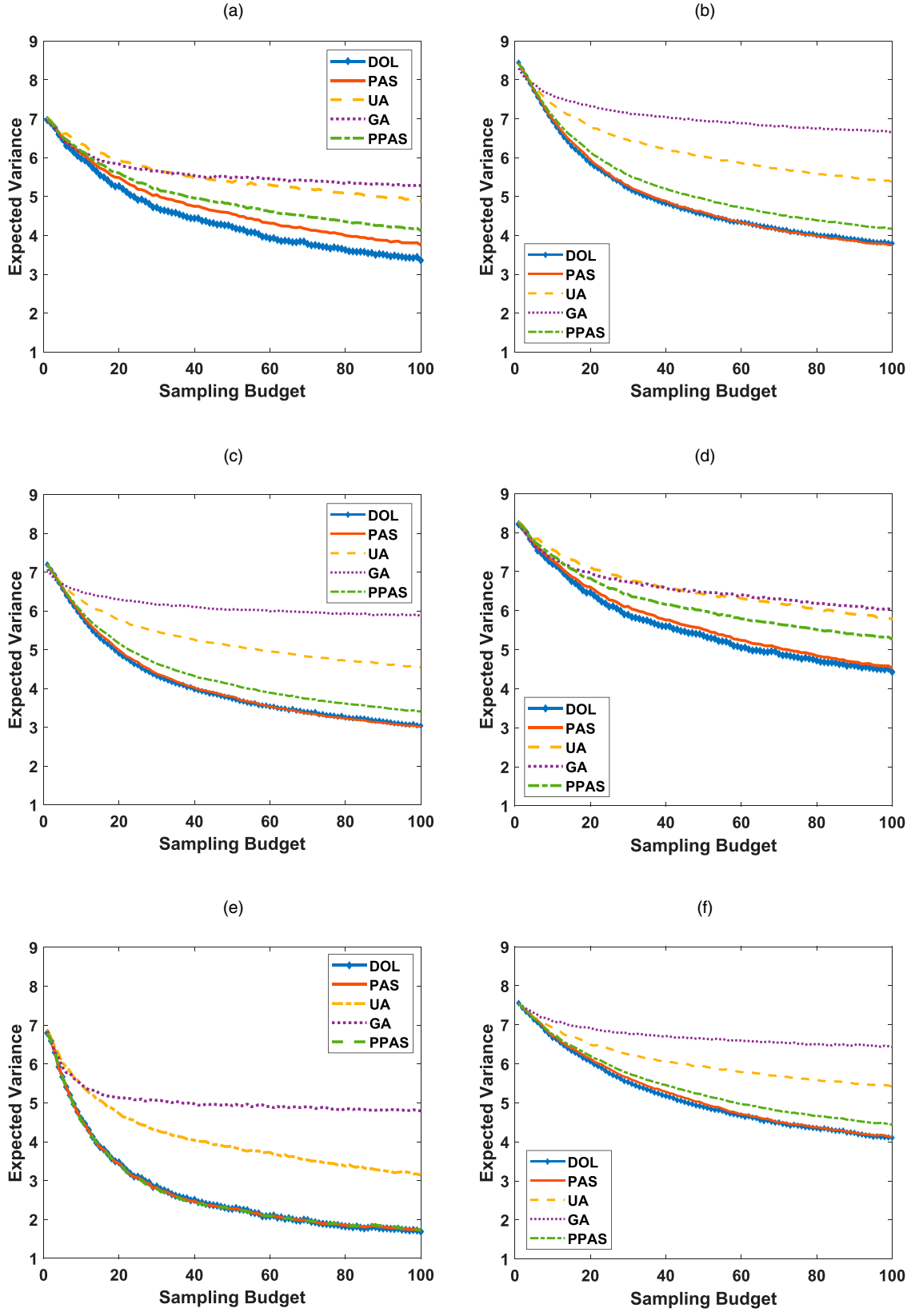
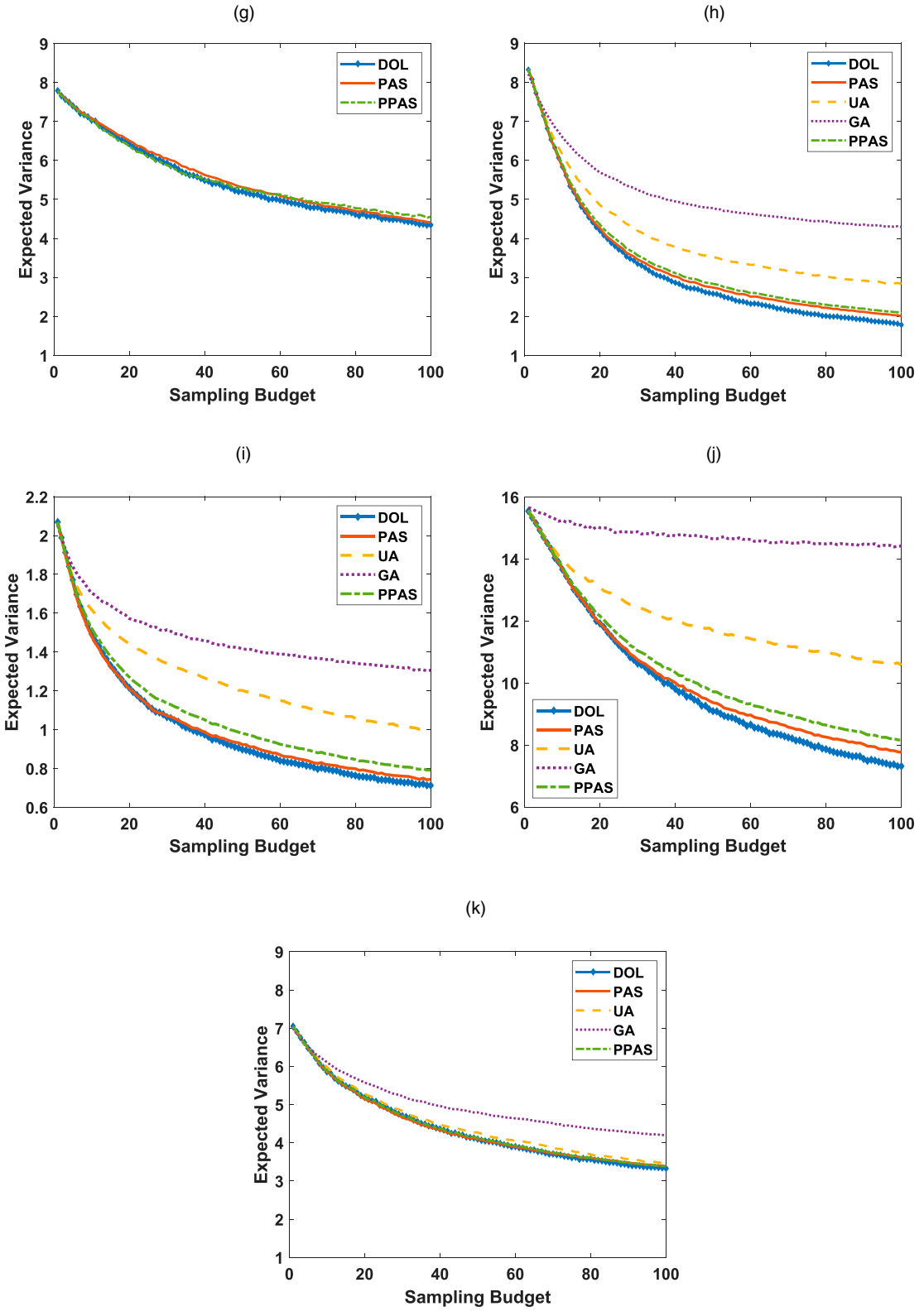


Figure 2. (Continued)



to see the impact of the number of doses on overall performance of DOL and other benchmarks. We see from Figure 2, (i) and (j) that decreasing/increasing the number of doses decreases/increases the overall expected variance of the target doses. The order of allocation policies in terms of performance has not changed, that is, DOL still performs slightly better than other benchmarks in both experiments.

Furthermore, we run a sensitivity analysis where we change the probability of observing each patient type. Specifically, we assume that $\mathcal{P}_{(1,0)} = 0.1$ and $\mathcal{P}_{(1,1)} = 0.9$, that is, we expect nine times more patients with covariate vector (1,1) than (1,0) in our simulation. In another words, patient type (1, 0) is rare. Figure 2(k) shows that the performance of all policies except for GA become similar for both patient types. The performance of DOL is similar to the base experiment, but the performance of adaptive posterior sampling methods improve compared with the base experiment.

Lastly, we conduct a sensitivity analysis with respect to the delay in observing the patient responses after allocating doses to patients. To that end, we assume that the delay τ is a discrete number measured in the unit of sampling epochs, that is, the patient response of any allocation will be available τ many epochs later. Then, the allocation decision at epoch $n \geq \tau$ is made given all the information available at epoch $n - \tau$, that is, the decision rule d^n is a function of data $\mathcal{F}^{n-\tau}$. In this case, the sampling decisions for the first τ samples are made by just considering the initial prior. In practice, for the first τ samples, we may use other algorithms for batch assignment of patients to treatment, which can potentially improve the performance of DOL at the end of the trial. We repeat the base experiment above for delays of $\tau = 0, 5, 10, \ldots, 40$ and plot the performance measures at the end of the trial ($N = 100$) for each τ. Results in Online Appendix C.6 show that the performance generally drops by increasing τ and DOL still outperforms other benchmarks considering different delays. Delay significantly impacted the GA policy, mainly because GA is inherently very exploitative and therefore is more sensitive to the initial belief.

8. Application to a Case Study with Real Data

In this section, we analyze a real dose-finding case study for an anticoagulant drug called warfarin. The goal is to investigate the use of personalized models for this problem. In the following, we first introduce the data set, then explain the setup of the main experiment, and, finally, present the results of three sensitivity analyses to gain more insight about effective allocations.

8.1. Warfarin Data Set

The warfarin data set used in this experiment is made available by Pharmacogenomics Knowledgebase (2021b). It includes multiple patient covariates and prescribed therapeutic doses of warfarin and their prothrombin time in international normalized ratio (PT-INR) scale for 5,700 patients from nine countries. PT-INR is measured as a quick response to warfarin. There are several genetic, clinical, demographic, and environmental covariates that may affect the response of a patient to this drug.

In fact, White (2010) carried out a statistical analysis of the PT-INR response in a diverse sample of patients and examined specific clinical factors that can affect the pharmacokinetics and pharmacodynamics of warfarin. Although she found several demographic (e.g., age and race) and genetic features that have a meaningful impact, she introduced the Vitamin K epOxide Reductase Complex Subunit 1 (VKORC1) genotype as the key factor with the highest influence on warfarin PT-INR response. Pharmacogenomics Knowledgebase (2021a) also has provided research results regarding clinical annotation of the VKROC1 genotype for warfarin PT-INR response. There are three different VKORC1 alleles denoted by AA, AG, and GG. Pharmacogenomics Knowledgebase (2021a) emphasized that patients with the GG allele may require a lower dose of warfarin as compared with patients with AG and AA alleles. For simplicity of presentation, we consider the VKORC1 genotype as the single factor determining the patient types. Excluding the data records with missing VKORC1-1639 attribute, we use the data from 2,997 patients to test the quality of sequential learning under the proposed policies. We denote the target dose of patients with AA, AG, and GG alleles with $\text{ED}^{AA}_{0.95}$, $\text{ED}^{AG}_{0.95}$, and $\text{ED}^{GG}_{0.95}$, respectively. Retrospective data analysis suggests that the true $\text{ED}^{GG}_{0.95}$ is clearly lower than the true $\text{ED}^{AA}_{0.95}$ and $\text{ED}^{AG}_{0.95}$, whereas there is not a significant difference between the true $\text{ED}^{AA}_{0.95}$ and $\text{ED}^{AG}_{0.95}$. Because the range of therapeutic dosages in the data set belongs to $[5, 105]$ mg/week, we discretize the dose range to belong to $D = \{10, 20, \ldots, 100\}$, representing the dose intervals between bin edges of $5, 15, 25, \ldots, 95, 105$ mg/week.

8.2. Main Experiment Setup

For simplicity of presentation, we assign indexes to the dose values in D and represent the set of allowable doses with $\mathcal{Z} = \{1, 2, \ldots, 10\}$. Figure 3(e) represents the average response of each patient type to each dose (given the entire data set). We can see from this plot that the true $\text{ED}^{GG}_{0.95}$ falls in the third dose range (between 25 and 35 mg/week), whereas the true $\text{ED}^{AA}_{0.95}$ and $\text{ED}^{AG}_{0.95}$ both fall into the fifth dose range (between 45 and 55 mg/week).

Figure 3. Average Performance Measures and Dose-Response Relations in the Real Case Study of Warfarin

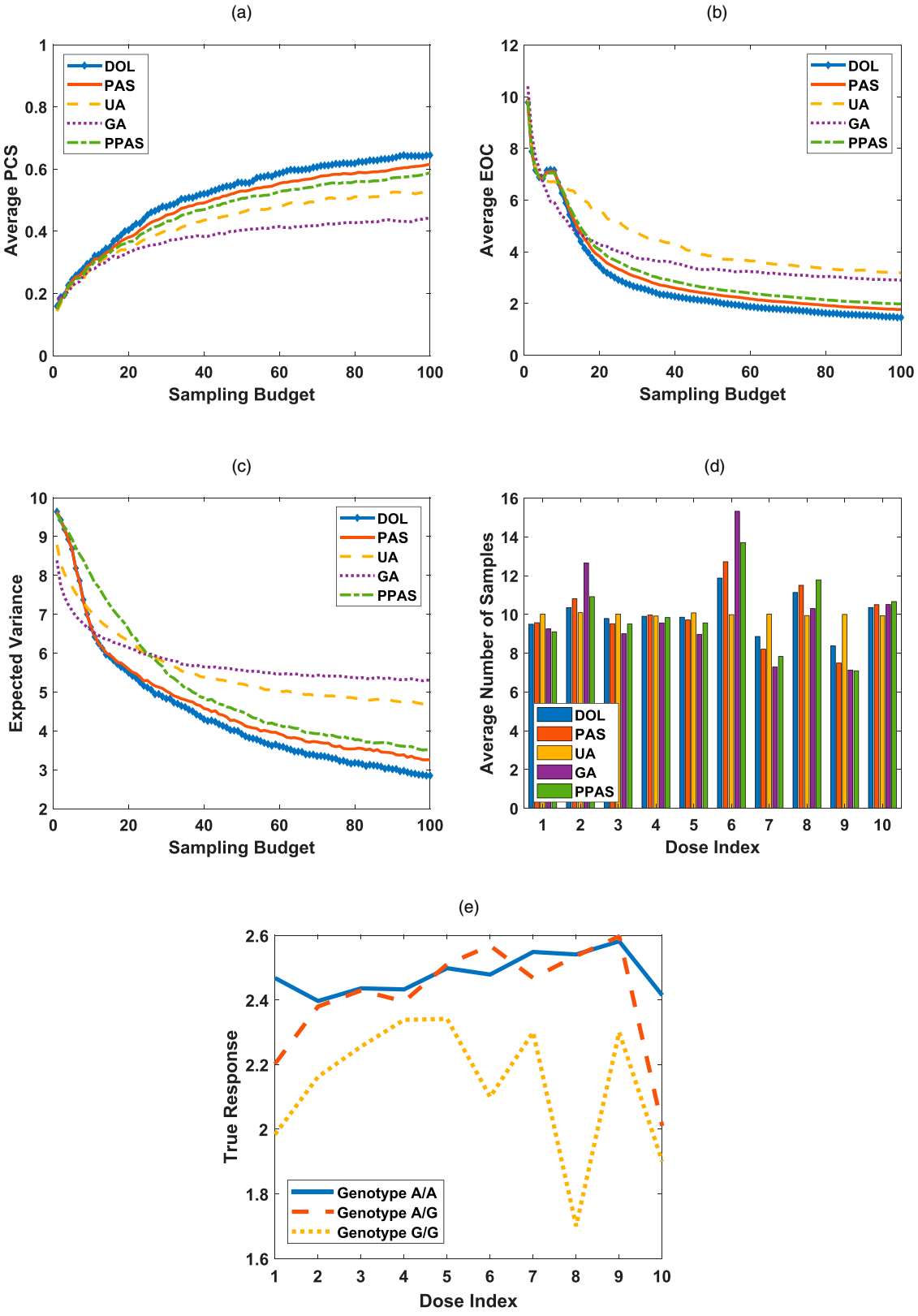


Table 1. Sampling Variances Calculated from the Data Set for Each Alternative Dose

Dose index	1	2	3	4	5	6	7	8	9	10
Sampling variance	0.240	0.179	0.145	0.165	0.156	0.093	0.306	0.091	0.065	0.012

For this numerical experiment, we run a simulation with 10,000 replications each starting from a fixed initial prior and randomly generated problem instances. We consider the sampling budget of $n = 100$ allocations in each replication for which we use a random permutation of patients in the data set. We note that the number of patients from three alleles are almost equal in the data set (995 patients with AA allele, 1,029 patients with AG allele, and 973 patients with GG allele). Hence, the probability of visiting each patient type at each epoch is almost equal (around one-third). In order to represent these three patients types, we use two binary variables ($K = 3$) with the set of covariate vectors $\mathcal{X} = \{(1,1,0),(1,1,1),(1,0,1)\}$ (respectively to represent patients with AA, AG, and GG alleles). We assume a $ZK \times 1$ initial flat (noninformative) prior mean vector of all zeros and a $ZK \times ZK$ initial prior covariance matrix with $\Sigma^0_{zk,z'k'} = \gamma_b + \gamma_c^{kk'} + \gamma_d^{zz'}$. Specifically, the PT-INR response is often a number between one and four; therefore, we set $b = 4$ to ensure a vague prior.

Here, we assume that the underlying setting is controlled by a fixed true mean response found by averaging over responses of each patient type assigned to each dose in our data set. In fact, this is an appealing feature of our experiment because it allows for *model misspecification*. In addition to the true mean response of each patient type to each dose, we can use the data set to calculate the sampling variance of each dose, that is, σ_z^2, $\forall z \in \mathcal{Z}$. Table 1 shows these values for the set of alternative doses $\mathcal{Z}$. We generate random responses in each replication from this fixed truth given these sampling variances. This makes our simulation consistent with the data set.

Similar to Section 7, we begin the sampling according to each policy without any initialization and compare the results of different policies at each epoch using the same performance measures. In terms of the average PCS, EOC, and the EVar of the target dose over all patient types, represented in Figure 3,(a)–(c), DOL slightly outperforms other policies. Figure 3(d) presents the average sampling allocations for this experiment. We can see from this plot that although GA allocates the largest number of patients to true target doses, it performs poorly with regard to all performance measures, especially EVar of the target dose. This is because the target doses (and their variance) are functions of the entire dose-response curve and GA is too exploitative, which causes it to perform even worse than UA in terms of PCS and EVar. On the other hand, we can see that DOL is robust in its trade-off between exploitation and exploration.

8.3. Additional Analyses of the Real Case Study

In order to gain more insight into the learning process, we present some additional plots to see how DOL and the benchmarks perform in *identifying the target dose for each patient type separately*. Figures 14 and 15 in Online Appendix C.7 show the results of implementing these policies for each patient type, in which the PCS, EOC, and EVar results are consistent, and that DOL is quite competitive in learning the personalized target dose for each patient type in terms of all performance measures. However, learning the target dose does not occur with the same pace for patients carrying different alleles. In particular, PCS is lower and the variance of the target dose is higher for the AA allele compared with AG and GG alleles. The main reason is that the true mean responses over all doses in patients with the AA allele are closer together compared with patients with AG and GG alleles, that is, the true dose-response curve for patients with the AA allele is almost flat, as presented in Figure 3(e). Therefore, it becomes more difficult to distinguish the corresponding target dose with the same level of sampling budget. Recall that the number of patients observed from each patient type is almost equal, and this difference in performance is due to the shape of the dose-response curve for each patient type.

Moreover, we design a setup to measure the *value of considering personalization* in dose-finding trials. To that end, we consider a DM who uses an impersonalized model for adaptive sampling in an environment in which dose-response curves depend on patient types. In fact, we use the true dose-response curves presented in the Figure 3(e) to control the simulation, and the patients randomly enter the trial similar to the previous experiment. However, we use all the responses, produced from all patient types, to update a single belief about the dose-response curve. That is, we consider a $Z \times 1$ vector Θ to model a single dose-response curve. We also assume a similar flat prior mean of all zeros and covariance structure similar to the experiment described above where $\gamma_c^{kk'} = 0$. Our results in Figure 16 of Online Appendix C.8 show a significant drop in performance, on average, over all three patient types. In particular, for $n = 100$, the DM who is running DOL with an impersonalized model will have a 20% and 45% drop in PCS and EVar performance, respectively. This is simply because in reality, we have different personalized target doses (dose 5 for patients with AA and AG alleles and dose 3 for patients with GG allele); however, the DM who assumes homogeneous patients estimates only a single

target dose for the entire population, which results in a poor performance.

Finally, we carry out another test to see the *importance of considering correlations among covariates*. Specifically, in this setup each group of patients with a particular allele is independent from others in that we learn each personalized target dose without considering the responses found for other patient types. We have three separate dose-response curves in parallel and simulate 10,000 different permutations of patients, where the belief about the alternative doses are correlated for each covariate. We use a learning model that does not allow information sharing among different patient types. In fact, we run three separate trials with $Z \times 1$ unknown parameters $\Theta^{(AA)}$, $\Theta^{(AG)}$, and $\Theta^{(GG)}$, simulating the results assuming a $Z \times 1$ initial flat prior mean vector of all zeros and a $Z \times Z$ initial prior covariance matrix of $\Sigma^0_{z,z'} = \gamma_b + \gamma_d^{zz'}$ for each trial. The results, presented in Figure 17 of Online Appendix C.9, show a significant drop in performance with regard to PCS and EVar, that is, information sharing across covariates is significant specially when we have a small sampling budget.

9. Practical Challenges and Limitations

There are a variety of practical issues in personalized dose-finding clinical trials that our model does not address. First of all, we assume that the response of a patient becomes available before the next epoch. Although this assumption is common in the literature (see, e.g., Guo and Yuan 2017), there are other studies that include delayed responses. For example, Lin et al. (2019) studied a different type of dose-finding trials where they considered delayed toxicity and efficacy outcomes. Although our model does not include the delay in observation, we present its effect on the performance of the proposed DOL in a sensitivity analysis presented in Online Appendix C.5.

Also, although assuming a continuous response is a common assumption in dose-finding trials, in some trials, the patient response may be binary (dose being effective or ineffective). In that case, the personalized target dose could be defined as the smallest dose with at least L times the maximum *probability* of observing an effective response over all doses for the patient with covariate vector x; a natural modification to our model is to consider a probit regression model instead of Equation (1), where y_z^x follows a Bernoulli distribution with the success probability of $p_z^x = \Phi(\langle \delta_z^x, \Theta \rangle)$ (Albert and Chib 1993). However, the probit regression model does not have any conjugate prior and the posterior samples should be generated by the Gibbs sampling approach introduced by Albert and Chib (1993). This makes the computation of DOL for adaptive sampling more time demanding. Regardless, our proposed procedure is still applicable with more computational efforts to create posterior samples.

Additionally, our proposed model is designed for low-dimensional covariate spaces. However, the set of covariate space could be large, whereas the number of patients in dose-finding trials is limited. One may use linear models, designed for high-dimensional data, instead of Equation (1). For example, Bastani and Bayati (2020) studied a contextual linear bandit setup with high-dimensional covariates for which they proposed an efficient algorithm based on the LASSO estimator. In addition, some of covariates could be continuous, such as blood pressure or viral load. One standard approach to make such covariates discrete/categorical is to use cut-off thresholds (European Network for Health Technology Assessment 2013).

Furthermore, we assume that significant covariates, which have the highest impact on the outcome, are prespecified, that is, they are known to the experimenter before starting the trial. This is often the output of clinical feature selection studies, where they specify which, say, biomarkers are important for toxicity/efficacy outcomes. In fact, several current personalized clinical trials are conducted to explore the impacts of a specific *given biomarker*: See Janiaud et al. (2019) for a list of oncology trials conducted for specific biomarker(s). However, as Berry et al. (2012) emphasized, gradually identifying the set of biomarkers in a precision clinical trial is crucial if the impactful biomarkers are not known upfront.

Another point to consider is that in practice, we may have two different types of covariates: (i) predictive covariates that interact with the treatment/dosage, for example, the presence of certain proteins in tumor biopsy in the case of cancer, (ii) prognostic covariates that do not interact with the treatment/dosage but may be informative of the outcome. We are assuming that all the covariates are predictive. However, in order to separate the effect of these two types of covariates, one needs to modify the response model to account for an additional number of unknown parameters that do not correspond to any doses/treatments. Using such a model, Alban et al. (2021) showed that misidentifying a predictive covariate and assuming it to be prognostic can be detrimental in terms of expected regret, yet the performance of an all-predictive model (our approach) turned out to be very close to the performance of a more comprehensive predictive-prognostic model. That is, the effect of misspecifying a prognostic covariate and taking it to be predictive in the model did not much influence the performance, mostly because the model will eventually learn to take proper values for the parameters that in fact do not relate to any dose/treatment.

Finally, we assume that the covariates of patients are observable upon arrival. However, in some cases, collecting the covariate data is expensive, might take

time (the laboratory results to come back), or may raise patient safety issues. Also, in other cases, the full covariate vector may not become known, either because of missing data or the delay in the process of collecting patient information. In such circumstances, one approach is to apply longitudinal models for biomarkers; see Berry (2012) and the references therein. Furthermore, we assume that the sampling continues until the last patient is allocated to a treatment. However, a trial may be stopped because there is significant evidence that the drug is efficacious or terminated if there is significant evidence that the drug is not efficacious or toxic. Nasrollahzadeh and Khademi (2020) studied the optimal stopping of a dose-finding clinical trial.

10. Conclusion

This study modeled a personalized response-adaptive dose-finding problem, where the DM seeks to find a target dose among a set of allowable doses given the personal characteristics of each patient. Setting different target doses for different patient types can lead to more successful phase III clinical trials, as patients assigned to their personal target dose are expected to show more promising responses. We considered a setting where the belief about alternative doses and similar covariates are correlated. We formulated this problem as a stochastic dynamic program and proposed DOL, which is a one-step look-ahead policy applied to our objective function of minimizing the expected variance of the personalized target dose over all covariate vectors. We also modified three other allocation policies, for example, PAS, PPAS, and GA, and used them as the benchmark. Using the model, we showed some properties of the learning problem and analyzed a few properties of the proposed DOL policy. For example, we derived a closed-form formula for DOL with two doses and showed that, in the impersonalized version of this problem, DOL is optimal. We also analyzed the asymptotic sampling behavior of DOL in this setting.

Finally, we implemented our proposed policies in various settings with synthetic data and conducted a variety of sensitivity analyses (with the results in the Online Appendix C). The results provided some insight into the behavior of each policy and revealed the connections between the proposed policies and the choice of the objective function. We also illustrated the results of applying our policies for a real case study of warfarin. We showed the merits of applying personalization and using the proposed DOL in terms of standard performance measures. Overall, considering the limited number of patients in dose-finding clinical trials and patients' variation, the proposed model and policy show promise for adaptive allocation of patients in dose-finding trials.

In terms of performance, the proposed DOL seems quite robust in a variety of settings. However, in some cases with notably large or small sampling variance, and in settings where the target doses are close to each other, the performance of some of the benchmarks are close to DOL; therefore, spending the extra computational effort for implementing DOL could be difficult to justify. In summary, DOL appears to be the best option in personalized dose-finding trials with moderate sampling variances and sparse target doses.

Acknowledgments

The authors thank the editors and two anonymous reviewers for their constructive comments.

References

Agarwal A, Hsu D, Kale S, Langford J, Li L, Schapire R (2014) Taming the monster: A fast and simple algorithm for contextual bandits. *Proc. 31st Internat. Conf. Machine Learn.* 32(2):1638–1646.

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

Alban A, Chick SE, Zoumpoulis S (2021) A value of information approach to designing sequential clinical trials for precision medicine. Accessed November 30, 2021, https://cattendee.abstractsonline.com/meeting/10390/presentation/6533.

Albert JH, Chib S (1993) Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.* 88(422):669–679.

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

Bastani H, Bayati M, Khosravi K (2021) Mostly exploration-free algorithms for contextual bandits. *Management Sci.* 67(3): 1329–1349.

Bates S (2010) Progress toward personalized medicine. *Drug Discovery Today* 15(3-4):115–120.

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

Berry DA, Herbst RS, Rubin EH (2012) Reports from the 2010 clinical and translational cancer research think tank meeting: Design strategies for personalized therapy trials. *Clinical Cancer Res.* 18(3):638–644.

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. Carlin B, Carriquiry A, Gatsonis C, Gelman A, Kass RE, Verdinelli I, West M, eds. *Case Studies in Bayesian Statistics*, Lecture Notes in Statistics, vol. 162 (Springer, New York), 99–181.

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

Bertsimas D, Korolko N, Weinstein AM (2019) Covariate-adaptive optimization in online clinical trials. *Oper. Res.* 67(4):1150–1161.

Bhat N, Farias VF, Moallemi CC, Sinha D (2019) Near optimal A-B testing. *Management Sci.* 66(10):4477–4495.

Chick SE, Gans N, Yapar O (2021) Bayesian sequential learning for clinical trials of multiple correlated medical interventions. *Management Sci.*, ePub ahead of print November 5, https://doi.org/10.1287/mnsc.2021.4137.

Delshad S, Khademi A (2020) Information theory for ranking and selection. *Naval Res. Logist.* 67(4):239–253.

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

Ding L, Hong LJ, Shen H, Zhang X (2022) Knowledge gradient for selection with covariates: Consistency and computation. *Naval Res. Logist.* 69(3):496–507.

European Medicines Agency Committee for Medicinal Products for Human Use (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. Report No. EMA/CHMP/SAWP/757052/2013, European Medicines Agency. https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/qualification-opinion-mcp-mod-efficient-statistical-methodology-model-based-design-analysis-phase-ii_en.pdf.

European Network for Health Technology Assessment (2013) *Endpoints Used for Relative Effectiveness Assessment of Pharmaceuticals: Health-Related Quality of Life and Utility Measures.* Report, EUnetHTA, Diemen, Netherlands.

Fisher RA (1949) *The Design of Experiments*, 5th ed. (Oliver and Boyd, London).

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

Goldenshluger A, Zeevi A (2013) A linear response bandit problem. *Stochastic Systems* 3(1):230–261.

Guo B, Yuan Y (2017) Bayesian phase I/II biomarker-based dose finding for precision medicine with molecularly targeted agents. *J. Amer. Statist. Assoc.* 112(518):508–520.

Hayden E (2015) California unveils 'precision-medicine' project. *Nature*, https://doi.org/10.1038/nature.2015.17324.

Ildstad ST, Evans CH Jr, eds. (2001) *Small Clinical Trials: Issues and Challenges* (National Academies Press, Washington, DC).

Janiaud P, Serghiou S, Ioannidis JP (2019) New clinical trial designs in the era of precision medicine: An overview of definitions, strengths, weaknesses, and current use in oncology. *Cancer Treatment Rev.* 73:20–30.

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.

Lin R, Thall PF, Yuan Y (2019) An adaptive trial design to optimize dose-schedule regimes with delayed outcomes. *Biometrics* 76(1):304–315.

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. Methodology* 17(1):1–14.

Liu S, Guo B, Yuan Y (2018) A Bayesian phase I/II trial design for immunotherapy. *J. Amer. Statist. Assoc.* 113(523):1016–1027.

Maxfield K, Zineh I (2021) Precision dosing: A clinical and public health imperative. *J. Amer. Medical Assoc.* 325(15):1505–1506.

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

Nasrollahzadeh AA, Khademi A (2022) Dynamic programming for response-adaptive dose-finding clinical trials. *INFORMS J. Comput.* 34(2):1176–1190.

Negoescu DM, Frazier PI, Powell WB (2011) The knowledge-gradient algorithm for sequencing experiments in drug discovery. *INFORMS J. Comput.* 23(3):346–363.

O'Quigley J, Shen LZ, Gamst A (1999) Two-sample continual reassessment method. *J. Biopharmaceutical Statist.* 9(1):17–44.

Peck RW (2018) Precision medicine is not just genomics: The right dose for every patient. *Annual Rev. Pharmacology Toxicology* 58:105–122.

Peck RW (2021) Precision dosing: An industry perspective. *Clinical Pharmacology Therapeutics* 109(1):47–50.

Pharmacogenomics Knowledgebase (2021a) Clinical annotation for rs61742245 (VKORC1); Warfarin; (level 2a dosage). Accessed November 30, 2021, https://www.pharmgkb.org/gene/PA133787052/clinicalAnnotation/1183703748.

Pharmacogenomics Knowledgebase (2021b) Pharmacogenetics and pharmacogenomics knowledge base downloads. Accessed November 30, 2021, https://www.pharmgkb.org/downloads.

Rojas-Cordova AC, Bish EK, Hosseinichimeh N (2020) Decision-making in sequential adaptive clinical trials, with implications for drug misclassification and resource allocation. Smith A, ed. *Women in Industrial and Systems Engineering* (Springer, Cham, Switzerland), 321–345.

Russo D, Van Roy B (2014) Learning to optimize via posterior sampling. *Math. Oper. Res.* 39(4):1221–1243.

Russo D, Van Roy B (2018) Learning to optimize via information–directed sampling. *Oper. Res.* 66(1):230–252.

Russo DJ, Van Roy B, Kazerouni A, Osband I, Wen Z (2018) A tutorial on Thompson sampling. *Foundations Trends Machine Learn.* 11(1):1–96.

Ryzhov IO (2016) On the convergence rates of expected improvement methods. *Oper. Res.* 64(6):1515–1528.

Ryzhov IO, Powell WB, Frazier PI (2012) The knowledge gradient algorithm for a general class of online learning problems. *Oper. Res.* 60(1):180–195.

Schork NJ (2015) Personalized medicine: Time for one-person trials. *Nature* 520(7549):609–611.

U.S. Food and Drug Administration (FDA) (2019) Adaptive designs for clinical trials of drugs and biologics guidance for industry. Report No. FDA-2018-D-3124, Center for Biologics Evaluation and Research & Center for Drug Evaluation and Research. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adaptive-design-clinical-trials-drugs-and-biologics-guidance-industry.

U.S. National Library of Medicine—National Institutes of Health (2021) Clinicaltrials.gov advanced search. Accessed November 30, 2021, https://www.clinicaltrials.gov/ct2/search/advanced.

Wald A (1943) On the efficient design of statistical investigations. *Ann. Math. Statist.* 14(2):134–140.

Wang Y, Powell WB (2018) Finite-time analysis for the knowledge-gradient policy. *SIAM J. Control Optim.* 56(2):1105–1129.

White PJ (2010) Patient factors that influence warfarin dose response. *J. Pharmacy Practice* 23(3):194–204.

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

Wu CJ, Hamada MS (2011) *Experiments: Planning, Analysis, and Optimization*, vol. 552 (John Wiley & Sons, Hoboken, NJ).