Submitted to *Service Science*
manuscript (Please, provide the manuscript number!)

Online Supplement to Optimal Stopping of Adaptive Dose-Finding Trials

Amir Ali Nasrollahzadeh
University of California, Davis, Department of Mechanical & Aerospace Engineering, Davis, CA, 95616, anasr@ucdavis.edu
Amin Khademi
Clemson University, Department of Industrial Engineering, Clemson, SC 29634, khademi@clemson.edu

---

1. Further Discussion on the Dose-Response Model

As discussed in Section 3, there are two main classic approaches to describe the dose-response curve function $f(z, \Theta)$: Functional forms and piecewise linear approximations. First, we take a closer look at the functional forms. For example, one may consider an $\mathrm{E}_{\max}$ model as $f(z, \Theta) = \theta_0 + \frac{\theta_1 z}{\theta_2 + z}$, where $(\theta_0, \theta_1, \theta_2)$ are unknown parameters to be estimated. There are a lot of potential choices for this function in addition to an $\mathrm{E}_{\max}$ model such as Hill's equation, Michaelis-Menten, first-order exponential, step functions, and so on (World Health Organization and others 2009). Choosing a functional form to describe the dose-response relationship dictates the shape of the underlying curve which exposes the model to misspecification error because information arising from observations might reveal that the true dose-response curve is of another shape. The above example of an $\mathrm{E}_{\max}$ model which is usually used to describe a sigmoid shaped dose-response curve is susceptible to misspecification error if response observations reveal that the underlying dose-response curve is in fact bell-shaped. Moreover, since we consider a normal distribution on error, thus the response, a Bayesian setup for functional forms suffers from nonconjugacy, which poses significant computational challenges as one has to use time-consuming Markov chain Monte Carlo approaches to generate samples from the posterior distribution. Also, the state space of a stochastic dynamic programming formulation involves the set of all probability distributions on the unknown parameters, making the analysis for decision making extremely challenging since the parameters' marginal distributions might not be known.

The second approach to estimating the dose-response curve is to use piecewise linear approximations to the curve, which addresses the challenges discussed above. In particular, because this method approximates the curve at each dose, it does not assume a predetermined functional form for the response upfront and it is therefore less likely to experience model misspecification errors. Moreover, as we next discuss, a proper choice for curve approximation will result in conjugacy, which significantly reduces computational efforts and simplifies analysis. More complicated models can also be considered to approximate the dose-response curve. For example, a second order normal dynamic linear model (NDLM) is also used to approximate the true dose-response curve, by which a patient's response for doses close enough to dose $z \in \mathcal{Z}$ is estimated by fitting a straight line using the expected response at dose z and a slope at that dose as model parameters. This is the approach proposed by Berry et al. (2002) and applied in dose-finding trials by Krams et al. (2003), Warner et al. (2015), Lenz et al. (2015) and Liu et al. (2017). The NDLM model is flexible in approximating any dose-response curve and allows for a linear correlation structure with random deviations. However, Nasrollahzadeh and Khademi (2018) showed that the first order NDLM is competitive with the second order approach in addition to being computationally more efficient in estimating different shaped dose-response curves. Furthermore, our proposed one-step look-ahead policy can be adapted to more complicated dose-response models such as NDLM, however, the diffusion approximation method relies on the proposed first-order model and its extension to more complicated models is not clear since df^*, the advantage over placebo, will not follow an Ito process any longer.

2. Proofs

Proposition 1. Consider the optimal stopping problem at time n when the system is in state s^n. Based on equation (12), the KG policy decides to continue the trial only if

$$\max\Big(0, -c_1' n_p + c_2 m_n \mathbb{E}[\mathbb{1}_{\{B^n\}} | \mathscr{F}^n]\Big) < \mathbb{E}^n\Big\{ -c_1 + \max_{a^{n+1}\in\{0,2\}} u(a^{n+1}, s^{n+1}, \mathscr{F}^{n+1}) \Big| \mathscr{F}^n \Big\},$$

where the left hand side of the inequality denotes the value of terminating or abandoning the trial at time n while the right hand side denotes the value of continuing the trial. To prove the proposition, it is enough to show that whenever the KG policy decides to continue the trial, the optimal policy also chooses continuation. The optimal policy decides to continue the trial if

$$\max\Big(0, -c_1' n_p + c_2 m_n \mathbb{E}[\mathbb{1}_{\{B^n\}} | \mathscr{F}^n]\Big) < \mathbb{E}^\tau \sup_{\tau \geq n+1} \Big\{ -c_1(\tau - n) + \mathbb{E}^n\big[\max_{a^\tau \in \{0,2\}} u(a^\tau, s^\tau, \mathscr{F}^\tau) \big| \mathscr{F}^n \big] \Big\},$$

where the value of termination or abandoning is equal to that of the KG policy, i.e., the left hand sides in both above inequalities are equal. Note that the supremum is taken over the set $\tau \geq n+1$ which contains $\tau = n+1$ and thus,

$$\begin{aligned}
\mathbb{E}^{\tau} \sup_{\tau \geq n+1} \Big\{ -c_1(\tau - n) + \mathbb{E}^n \big[\max_{a^\tau \in \{0,2\}} u(a^\tau, s^\tau, \mathscr{F}^\tau) \big| \mathscr{F}^n \big] \Big\} \\
&= \mathbb{E}^{\tau} \sup_{\tau \geq n+1} \Big\{ \mathbb{E}^n \big[-c_1(\tau - n) + \max_{a^\tau \in \{0,2\}} u(a^\tau, s^\tau, \mathscr{F}^\tau) \big| \mathscr{F}^n \big] \Big\} \\
&\geq \mathbb{E}^{n+1} \Big\{ \mathbb{E}^n \big[-c_1(n+1-n) + \max_{a^{n+1} \in \{0,2\}} u(a^{n+1}, s^{n+1}, \mathscr{F}^{n+1}) \big| \mathscr{F}^n \big] \Big\} \\
&= \mathbb{E}^n \Big\{ -c_1 + \max_{a^{n+1} \in \{0,2\}} u(a^{n+1}, s^{n+1}, \mathscr{F}^{n+1}) \Big| \mathscr{F}^n \Big\},
\end{aligned}$$

where the last equality is justified by the tower property of conditional expectation. Therefore, whenever the KG value of continuation is greater than abandoning or terminating the trial at time n, the optimal value of continuation is also greater and the optimal policy decides to continue the trial. Q.E.D.

Proposition 2. Assuming that $B_{ct}(x_t, t)$ is twice differentiable in the continuation set $\mathscr{C}$, $B_{ct}(x_{t+\Delta t}, t+\Delta t)$ in equation (18) may be written according to the Taylor series expansion by

$$\begin{aligned}
B_{ct}(x_t, t) = -c_1 \Delta t + \mathbb{E} \Big[B_{ct}(x_t, t) + \frac{\partial B_{ct}(x_t, t)}{\partial t} \Delta t + \frac{\partial B_{ct}(x_t, t)}{\partial x} \Delta x + \frac{1}{2} \frac{\partial^2 B_{ct}(x_t, t)}{\partial x^2} (\Delta x)^2 \\
+ \mathcal{O}(\partial t) \Big| \mathscr{F}^t_{ct} \Big],
\end{aligned}$$

where $\mathcal{O}(\partial t)$ denotes all the terms in the Taylor expansion with ∂t^2, $\partial t \partial x$, or higher degrees of differentiability in ∂t. Replacing Δx with $\theta\, dt + \sigma\, dW_n$, we have

$$\begin{aligned}
B_{ct}(x_t, t) = -c_1 \partial t + \mathbb{E} \Big[B_{ct}(x_t, t) + \frac{\partial B_{ct}(x_t, t)}{\partial t} \partial t + \frac{\partial B_{ct}(x_t, t)}{\partial x} (\theta\, \partial t + \sigma\, \partial W_t) \\
+ \frac{1}{2} \frac{\partial^2 B_{ct}(x_t, t)}{\partial x^2} (\theta\, \partial t + \sigma\, \partial W_t)^2 + \mathcal{O}(\partial t) \Big| \mathscr{F}^t_{ct} \Big].
\end{aligned}$$

Using Itô's lemma, and noting that ∂t^2 and $\partial t \partial W$ tend to zero faster than ∂W^2 when $\partial t \to 0$,

$$\begin{aligned}
B_{ct}(x_t, t) = -c_1 \partial t + \mathbb{E} \Big[B_{ct}(x_t, t) + \Big(\frac{\partial B_{ct}(x_t, t)}{\partial t} + \frac{\partial B_{ct}(x_t, t)}{\partial x} \theta + \frac{1}{2} \frac{\partial^2 B_{ct}(x_t, t)}{\partial x^2} \sigma^2 \Big) \partial t \\
+ \sigma \frac{\partial B_{ct}(x_t, t)}{\partial x} \partial W_t \Big| \mathscr{F}^t_{ct} \Big],
\end{aligned}$$

where ${\partial W_n}^2$ is substituted with ∂t. Noting that $\mathbb{E}[\partial W_t | \mathscr{F}^t_{ct}] = 0$ because W_t is a standard Brownian motion, and $\mathbb{E}[\theta | \mathscr{F}^t_{ct}] = \frac{x_t}{t}$ because the posterior belief about θ at time t is $\mathcal{N}(\frac{x_t}{t}, \frac{\sigma^2}{t})$,

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

where the terms $B_{ct}(x_t, t)$ cancel each other on both sides of the equality and the equation is divided by $\Delta t = \partial t$. The first boundary is derived by the definition of the continuation set, and the second boundary is a so-called smooth pasting condition. Q.E.D.

3. Details on Solution Algorithms

Simulation-based Gridding Algorithm. Algorithm 1 describes the simulation-based gridding method, which we run in an online fashion for each patient. To that end, construct a discretized grid over possible values of m and ν by selecting appropriate lower and upper bounds for them. We use the notation (m, ν, n) to refer to the grid constructed at decision epoch n.

At each period, dose allocation z^n is given and the state variable (μ^n, Σ^n) is known. Steps 3-14 describe the M forward simulations of the trial starting from decision epoch n to the end of trial N. In particular, for each $i \in \{1, \ldots, M\}$ and for any decision epoch $n \leq k \leq N$, dose z_i^k is determined by the fixed allocation policy. A future observation is simulated using the posterior predictive distribution. This observation is then used to update the state variable (μ_i^k, Σ_i^k). Using the updated state variable, a sample of T dose-response curves Θ are generated whereafter m_i^k and ν_i^k are evaluated using sample mean and sample variance, respectively. Steps 16-23 implement a similar procedure for the cells that are not populated already. Note that experiments $i \in \{1, \ldots, M\}$, or $i \in \{1, \ldots, M'\}$ are independent of each other and can run in parallel. We used the "*foreach*" package in the R programming language to parallelize the forward simulations and reduce the run time.

When the grid is fully populated, starting from step 25, we use backward induction to evaluate the optimal value function and thus the optimal decision for each cell in the grid. Starting from the last epoch N, the optimal value function is the maximum expected value of termination or abandonment for that cell since only these two decisions are available. The expected value of termination is easily evaluated because in each cell (m, ν) are known. Working backwards, in order to find the optimal value function and the optimal decision for each cell, we need to consider the expected value of continuation as well. This is achieved by tracking which experiments are visiting each grid cell in A_j^k at each decision epoch. Therefore, using equation (10), the utility of continuation in each cell j at decision epoch k is the average optimal value function of the cells at decision epoch $k+1$ which are visited by an experiment originating from cell j.

Algorithm 1 is run at each decision epoch n to evaluate the optimal value function and thus the optimal decision across the entire grid. Thereafter, at decision epoch n, true response y^n is observed, and using a similar procedure to steps $7-11$, (m_n, ν_n) are evaluated and the optimal decision is identified by finding the corresponding grid cell for (m_n, ν_n).

One-Step Look-Ahead Policy. In Algorithm 2, at each decision epoch n the state (μ^n, Σ^n) is given. To evaluate the utility of termination, a T-sized sample of dose-response curves Θ_t^n is generated. For each Θ_t^n, the target dose $z_t^{*,n}$ is evaluated via equation (1) and thus the advantage

over placebo, i.e., df^* can be calculated where $df^*_{n,t} = f(z^*_t, \Theta^n_t) - f(0, \Theta^n_t)$. Estimate m_n, and ν_n using a sample mean and sample variance. These values are later used in step 9 to evaluate the expected utility of termination. Then, n_p observations of y_* and y_0 are generated where $\hat{y}|\Theta, z \sim \mathcal{N}(\theta_z, \sigma^2)$ in order to estimate $\mathbb{E}[\mathbb{1}_{\{B^n\}}|\mathscr{F}^n]$ by Monte Carlo. The value of termination is then computed via equation (3).

Instead of simulating the entire trial to the last participant N to evaluate the value of continuation, a one-step look-ahead policy is implemented where at each decision epoch n, the next stage is assumed to be the last. Therefore, the value of continuation is computed by looking one step into the future. Starting from step 11 in Algorithm 2, the trial is simulated one-step into the future by generating future observations and updating the estimate of the dose-response curve Θ with respect to them. For each simulated observation, our belief about the dose-response curve is updated and the expected value of termination in the next stage is estimated. Taking a sample average over all these values results in an approximation of the expected utility of termination at decision epoch $n+1$. Since the expected value of abandonment is fixed to 0, one can approximate the value of continuation in equation (12) by taking the maximum over 0 and the approximated expected value of termination. The one-step look-ahead stopping rule is checked in step 18 of Algorithm 2.

Diffusion Approximation. The solution to the free boundary problem in Proposition 2 is evaluated using a trinomial tree discretization method. To that end, create a grid over (x_n, t_n) for all $0 \le n \le N$ by considering a rectangle $[\underline{x}, \overline{x}] \times [t_0, t_0 + N]$ where $\underline{x}$ and $\overline{x}$ are appropriately selected lower and upper bounds for x. These bounds are selected to be similar to the range considered in the simulation-based gridding algorithm. According to the trinomial tree method, the Brownian process defined in equation (17) in any given cell in the grid, i.e., (x_i, t_i), can move to one the following three different cells:

$$\begin{cases} (x_{i-1}, t_{i+1}) \text{ with probability } p_d, \\ (x_i, t_{i+1}) \quad \text{with probability } p_m, \\ (x_{i+1}, t_{i+1}) \text{ with probability } p_u, \end{cases}$$

where probabilities p_d, p_m, and p_u satisfy the following equations (see Ingber et al. 2001)

$$\begin{cases} p_u + p_m + p_d = 1, \\ \Delta x(p_u - p_d) = x\Delta t \\ (\Delta x)^2(p_u + p_d) = \tilde{\sigma}^2 \Delta t, \end{cases}$$

where Δx and Δt are carefully selected grid intervals, and $\tilde{\sigma} = \frac{2\sigma^2}{t} - \frac{2\sigma^2}{t+1}$ denotes the posterior variance at $t+1$. To choose appropriate grid intervals, set Δt such that $\frac{1}{\Delta t}$ is equal to an integer value. In Section 6, we considered $\Delta t = 0.05$. Following Arlotto et al. (2010), we also assume that $p_u = p_d$, thus $p_u = p_d = \frac{\tilde{\sigma}^2 \Delta t}{2(\Delta x)^2}$, and $p_m = 1 - 2p_d$. Therefore, the probabilities p_d or p_u are maximized when $t = t_0$. Since $p_u + p_d \leq 1$, $p_u \leq p_{max} \leq 0.5$, we have $\Delta x = \frac{\sigma\sqrt{2\Delta t}}{\sqrt{2t_0(t_0+1)p_{max}}}$. We assume $p_{max} = 0.495$. The backward solution to the grid is given by

$$B(x_i, t_i) = p_u B(x_{i+1}, t_{i+1}) + p_m B(x_i, t_{i+1}) + p_d B(x_{i-1}, t_{i+1}),$$

where at the top or bottom row of cells in the x axis, the cell values are extended in a linear fashion. Therefore, at $i = 0$, or $i = I$, we have

$$B(x_{I+1}, t) = 2B(x_I, t) - B(x_{I-1}, t),$$
$$B(x_{0-1}, t) = 2B(x_0, t) - B(x_1, t).$$

Note that in the last column of cells in t axis where $t = t_0 + N$, the cell values are calculated by the boundary condition $B(x_i, t_0 + N) = G(x_i, t_0 + N)$. After enumerating the entire grid by the values of $B(x_i, t_i)$, grid cells for which $B(x_i, t_i) = G(x_i, t_i)$ are recorded and their (x_i, t_i) values are extracted. To apply the smooth pasting condition in the free boundary problem, we apply a smoothing spline package in the R programming language to smooth the extracted (x_i, t_i) values; thus the lower and upper boundaries.

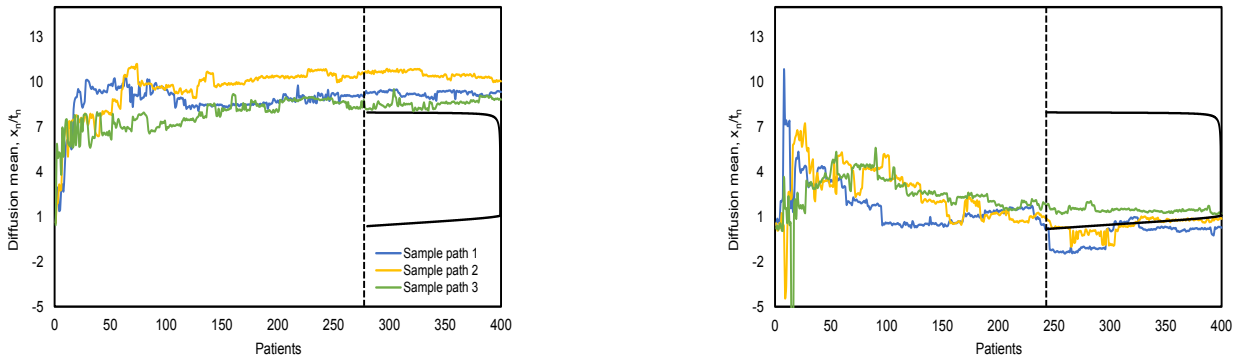
To find the optimal decision at each decision epoch, observe true observations y^n, update the values of m_n, ν_n, and n, then calculate x_n and t_n by using the following formulas already given in Section 5.3.

$$t_0 = \frac{2\sigma^2}{\nu_0^2},$$
$$t_n = t_0 + n,$$
$$x_n = m_n t_n.$$

If (x_n, t_n) correspond to a cell inside the area marked by lower and upper boundaries, the trial is still in the continuation set and the optimal decision is to continue. However, if the corresponding cell falls outside of the boundaries, the decision is to stop the trial, i.e., terminate or abandon.

4. Further Numerical Analysis

In Sections 4.2 and 4.1, we present additional numerical results including the probability of correct selection as another performance measure for the sigmoid dose-response curve and the



(a) Sigmoid curve (b) Flat curve

Figure 10 State variable paths

settings introduced in Section 6.1 of the main paper, and state variable paths for the diffusion approximation method presented in Section 6.3 of the main paper, respectively. Section 4.3 details some of the estimations necessary for reporting the results of application to real clinical trials. Sections 4.4 to 4.8 present a series of sensitivity analyses with respect to the variance of observation, shape of the underlying dose-response curve, the dose allocation procedure, the cost and benefit coefficients, and the discetization prameters of the gridding algorithm and the diffusion approximation method.

4.1. Probability of Correct Selection

One of the typical performance measures in the adaptive design of dose-finding (Phase II) clinical trials is the probability of correctly identifying the target dose. A "good" design results in a high probability of correct selection for the true target dose. However, when studying a flat dose-response curve, a correct dose does not exist. Therefore, we report the probability of correct selection for the sigmoid dose-response curve introduced in Section 6 of the main paper. The settings of the remaining simulations are similar to that of Section 6.1 of the main paper. Table 6 shows the probability of correctly identifying the target dose (PCS) for each solution algorithm. This probability is evaluated by generating samples of the posterior distributions of Θ at the time of stopping, identifying the ED_{95} dose, and comparing it with the true target dose. Recall that for this setting, the probability of correct decision is 1 for all the algorithms, that is, all of them correctly terminate the trial. However, the performance of the diffusion approximation method is better regarding the PCS because it stops a bit later.

4.2. State Variable Paths

Figure 10 shows a few state variable paths for sigmoid and flat dose-response curves for the sensitivity analysis in Section 6. In case of the sigmoid dose-response curve, all such paths would

Table 6 Probability of correctly identifying the right dose

	PCS
Simulation-based gridding	0.89
KG	0.85
Diffusion approximation	0.96

cross into the termination region before $\max_j \text{Var}[\theta_j|\mathscr{F}^n] \leq 4$. Thus, the average stopping time in Table 3 is reported to be equal to 280 because the diffusion algorithm decides on termination before further sampling. In case of the flat dose-response curve, a few paths cross into the abandonment region before $\max_j \text{Var}[\theta_j|\mathscr{F}^n] \leq 4$ is satisfied. However, a few paths remain in the continuation region a little longer and some do not cross into abandonment region at all which is why the average stopping time reported in Table 4 happens later and the probability of correct decision remains the same with respect to the results in Table 2.

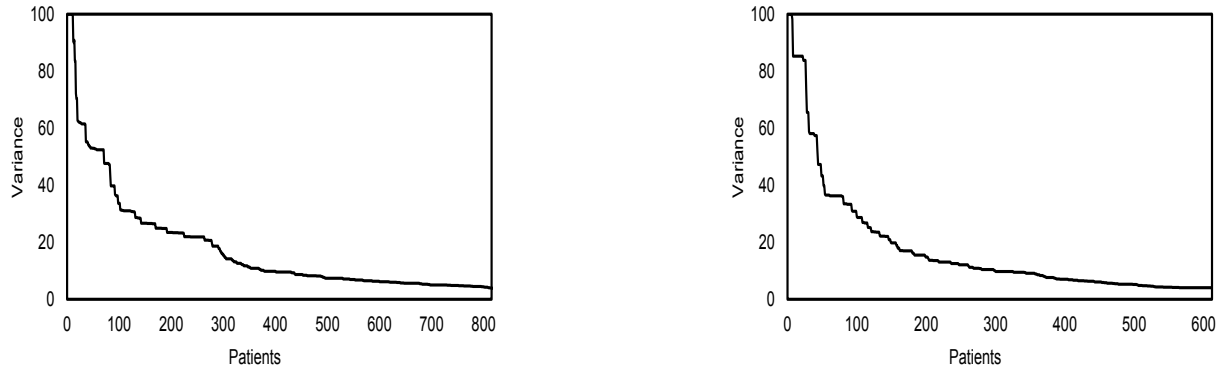
4.3. Estimate of c_2 for Application to Clinical Data

The incremental cost-effectiveness ratio of the genomic-directed chemotherapy was reported to be €5529. This price reflects the change in average total cost to a patient per QALY in the European Union if the proposed treatment was adopted. Note that Hall et al. (2011) used various monetary estimates based on different statistics of both the UK and the European Union. We consider this cost as the benefit of the developing company per unit advantage over placebo. However, to actually capture the benefit to the company, one needs to estimate the total benefit of the treatment in the market. To do that, we consider the female population of a European Union county (Germany, Schmid 2017) and the percentage of lymph node breast cancer incidence rate per 100,000 females (Norway, Hofvind et al. 2012), which enables us to estimate the total market value of a new treatment per QALY assuming that the stage II breast cancer incidence rate in Norway roughly estimates that of the German population. Therefore, assuming that all breast cancer patient diagnosed with lymph node involvement undergo the new treatment, the benefit will be approximately $63 million. In reporting the results in Table 6, we consider a very conservative c_2 equal to $1 million. A more accurate estimation of c_2 requires a detailed market analysis which is not in the scope of this study, however, we can refer the readers to an example of such a study in Grand View Research (2019).

4.4. Sensitivity to the Variance of Observations

The dose-response model in Section 3 includes a noise in response $\epsilon \sim \mathcal{N}(0, \sigma^2)$. In reporting the results in Section 6, we assume that σ^2 is constant and equal to 100 units throughout the

trial. However, this might not be the case in the real world, and thus we propose a sensitivity analysis with respect to the variance of observations, and show that our conclusions are robust. The following results are reported for $\sigma^2 = 100$. In particular, Figure 11 shows the maximum posterior variance at each decision epoch for both sigmoid and flat dose-response curves until $\max_j \text{Var}[\theta_j|\mathscr{F}^n] \leq 4$ is satisfied. Note that to satisfy this condition in case of the sigmoid dose-response curve, $n \geq 815$, where the maximum expected utility is $8,590,869$ achieved at $n = 825$. For the flat dose-response curve, $n \geq 612$ satisfies the condition and results in a maximum expected utility of $-589,975$ at $n = 621$. Tables 7 and 8 denote the performance measurements of the three solution algorithms with respect to the sigmoid and flat dose-response curves, respectively. The results for the sigmoid dose-response curve are compatible with those of Tables 3 and 4 in Section 6. However, in case of the flat dose-response curve, the probability of correct decision has improved significantly for the simulation-based gridding algorithm and the one-step look-ahead policy. Notice that by increasing the variance of observation, the allocation algorithm requires a larger number of samples such that the cost of sampling cancels out any incorrect identification of the improvement over placebo, and thus all algorithms abandon correctly.



(a) Sigmoid curve (b) Flat curve

Figure 11 Maximum posterior variance until $\max_j \text{Var}[\theta_j|\mathscr{F}^n] \leq 4$

Table 7 Sigmoid dose-response curve

	Estimated utility	True utility	Stopping time	PCD
Simulation-based gridding	8,555,127	7,848,623	830	1
KG	8,550,258	7,863,623	815	1
Diffusion approximation	8,550,258	7,863,623	815	1

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

Table 8 Flat dose-response curve

	Estimated utility	True utility	Stopping time	PCD
Simulation-based gridding	-627,207	-646,000	616	1
KG	-627,207	-643,000	613	1
Diffusion approximation	-792,184	-642,000	612	1

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

4.5. Sensitivity to the Shape of the Underlying Dose-Response Curve

In reporting the performance measures in Section 6 of the main paper, the underlying dose-response curves under investigation were sigmoid shaped (with a significant advantage over placebo) and flat (without a significant advantage over placebo). One can argue that the sigmoid shape is too specific since the response and dosage are positively correlated. Here, we present the performance of our proposed solutions to a non-monotonic curve which is very general and could capture real dose-response curves similar to Figure 1(c). Figure 12 shows the non-monotonic dose-response curve and the state of the learning process after 20 initial patients. Table 9 shows the estimated and true utilities, the stopping time, and the probability of correct decision for the simulation-based gridding algorithm and the proposed solutions. The results of this sensitivity analysis shows that our results and insights are robust with respect to the shape of the dose-response curve.

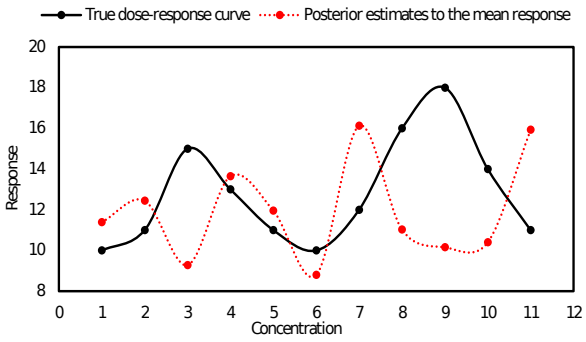


Figure 12 Posterior estimate to the dose-response curve after 20 patients

4.6. Sensitivity to the Dose Allocation Procedure

In an adaptive design of clinical trials, a decision maker is faced with two sets of decisions: (i) Whether to terminate, abandon, or continue the trial each time a new observation becomes available, and (ii) how to determine the next dose assignment if the first decision is to continue the trial. Recall that, we assume a decoupled stopping and allocation procedure, i.e., if the

Table 9 Non-monotonic dose-response curve

	Estimated utility	True utility	Stopping time	PCD
Simulation-based gridding	8,134,632	7,921,000	49	1
KG	7,711,279	7,945,000	25	1
Diffusion approximation	8,111,178	7,883,000	87	1

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

stopping decision is to continue the trial, the decision maker is given the next assignment by some known and fixed policy. This assumption enabled us to simplify model complexity and reduce computational efforts and allowed us to propose a diffusion approximation solution. In addition, this assumption is justified since most clinical trials are currently administered according to a fixed balanced randomization policy (Berry et al. 2010). Our numerical analysis reports the performance metrics of the standard and the proposed solutions using an efficient adaptive allocation approach proposed in Nasrollahzadeh and Khademi (2018). Here, we are assessing the performance of the proposed algorithms with respect to a static approach which assigns each dose according to balanced probabilities, and another adaptive approach which assigns each dose to maximize the probability of correctly identifying the target dose one step into the future (Berry et al. 2010, Chapter 4). Tables 10 and 11 show similar performance measures to Tables 1 and 2 of the main paper, respectively, for these two allocation procedures. The results of Tables 1 and 2 are also included in order to ease comparison and demonstrate the consistency of the results. We excluded the simulation-based gridding algorithm from these tables because of the extensive computational effort required for its simulation.

Table 10 Sigmoid dose-response curve

Allocation policy	Stopping policy	Estimated utility	True utility	Stopping time	PCD
Nasrollahzadeh and Khademi (2018)	KG	10,791,346	9,462,000	38	1
	Diffusion approx.	10,632,415	9,394,000	106	1
Balanced randomization	KG	10,664,504	9,476,000	24	1
	Diffusion approx.	10,495,772	9,361,000	139	1
Adaptive randomization of Berry et al. (2010)	KG	10,778,156	9,456,000	44	1
	Diffusion approx.	10,521,007	9,392,000	108	1

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

Table 11 Flat dose-response curve

Allocation policy	Stopping policy	Estimated utility	True utility	Stopping time	PCD
Nasrollahzadeh and Khademi (2018)	KG	1,840,765	-58,000	28	0.10
	Diffusion approx.	-3,910	-307,000	277	0.96
Balanced randomization	KG	2,881,252	-42,000	12	0.04
	Diffusion approx.	-17,314	-336,000	306	0.98
Adaptive randomization of Berry et al. (2010)	KG	1,817,181	-51,000	21	0.1
	Diffusion approx.	-8907	-319,000	289	0.96

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

4.7. Sensitivity to Monetary Parameters

Since the shape of the utility function relies on the monetary coefficients c_1, c_1', and c_2 and their respective ratios, we include a sensitivity analysis to test the performance of the simulation-based gridding algorithm, one-step look-ahead policy, and diffusion approximation method for different ratios. This is of particular interest because the termination and abandonment boundaries of the diffusion approximation method depend on the shape of the utility function. The numerical analyses in Section 6 are reported assuming that $c_1 = c_1' = 1000$ and $c_2 = 1,000,000$. These values are chosen to replicate the study by Müller et al. (2007) and are based on the assumption that the benefits of introducing a working treatment are orders of magnitude greater than the cost of sampling one patient, and that the cost of sampling a patient in Phase II and the confirmatory phase are roughly similar. Although, these assumptions might hold true on average (Battelle Technology Partnership Practice 2015), these ratios may end up being different for different trials. Therefore, we add a sensitivity analysis with respect to the monetary coefficients of the utility function. Because of the extensive computational effort required for the simulation-based gridding algorithm, we only report the performance measures of the one-step look-ahead policy and the diffusion approximation method.

Table 12 shows the performance measures for different cost/benefit ratios. We set $c_1 = 1000$. As the benefit coefficient reduces, and c_1 and c_1' costs increase, the uncertainty about the cost-effectiveness of the target dose increases. Therefore, the probability of correct decision (PCD) decreases because the solution algorithms cannot be sure if the marginal improvement detected in the advantage over placebo is beneficial at the given cost/benefit ratio. Notice that the diffusion approximation method performs better in terms of PCD in these scenarios.

Table 13 presents similar results for the flat curve. This time, as the cost/benefit ratio increases, the uncertainty about the flatness of the curve reduces, and thus the probability of correct

decision increases. Notice that when the cost/benefit ratio is at its highest, the performance of KG and diffusion approximation is essentially similar.

Table 12 Sigmoid dose-response curve

		KG			Diffusion approximation		
		$c_2 = 10^6$	$c_2 = 10^5$	$c_2 = 10^4$	$c_2 = 10^6$	$c_2 = 10^5$	$c_2 = 10^4$
Estimated utility	$c_1'/c_1 = 1$	10,791,346	1,178,392	77,836	10,632,415	1,078,426	83,029
	$c_1'/c_1 = 2$	10,761,346	1,148,392	58,836	10,602,415	1,048,426	-48,189
	$c_1'/c_1 = 5$	10,671,346	1,058,392	-3,163	10,512,415	950,426	-3,163
True utility	$c_1'/c_1 = 1$	9,462,000	885,000	43,300	9,394,000	835,000	22,300
	$c_1'/c_1 = 2$	9,432,000	855,000	20,300	9,364,000	805,000	-93,700
	$c_1'/c_1 = 5$	9,342,000	795,000	-55,700	9,274,000	723,000	-55,700
Stopping time	$c_1'/c_1 = 1$	38	38	22	106	88	43
	$c_1'/c_1 = 2$	38	38	15	106	88	129
	$c_1'/c_1 = 5$	38	38	1	106	80	1
PCD	$c_1'/c_1 = 1$	1	1	0.60	1	1	0.74
	$c_1'/c_1 = 2$	1	1	0.32	1	1	0.66
	$c_1'/c_1 = 5$	1	1	0	1	1	0

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

4.8. Sensitivity to Discretization Parameters

The simulation-based gridding algorithm and the diffusion approximation rely on discretization of the state space variables. In the simulation-based algorithm, the approximation is modified to depend on the true state variable, i.e., $s = (\mu, \Sigma)$, only through $\tilde{s} = (m, \nu)$. In constructing the grid, the range over m is considered to be $[0, 20]$ divided into 40 intervals. The range over ν is considered to be $[0, 10]$ which is divided into 20 intervals. Our numerical analysis shows that expanding the range does not contribute to the solution in a significant way. However, doubling the number of intervals in each axis produced better results shown here in Table 14. Because of the very expensive computation efforts, we could never discretize the grid as finely as the diffusion approximation method. Notice that in reporting the result of Table 14, the simulation-based gridding algorithm is improved by the condition described in Section 6.3.

In the case of diffusion approximation, the grid is constructed over another modified state variable $\hat{s} = (x, t)$. As described in details in Section 3, Δt is selected in such a way that $\frac{1}{\Delta t}$ is an

Table 13 Flat dose-response curve

		KG			Diffusion approximation		
		$c_2 = 10^6$	$c_2 = 10^5$	$c_2 = 10^4$	$c_2 = 10^6$	$c_2 = 10^5$	$c_2 = 10^4$
Estimated utility	$c_1'/c_1 = 1$	1,926,790	658,360	-4,987	-3,910	-14,482	-25,153
	$c_1'/c_1 = 2$	1,896,790	628,360	-55,153	-10,131	-27,468	-55,153
	$c_1'/c_1 = 5$	1,806,790	538,360	-145,153	-56,733	-95,231	-145,153
True utility	$c_1'/c_1 = 1$	-64,000	-64,000	-32,000	-307,000	-114,000	-31,000
	$c_1'/c_1 = 2$	-94,000	-94,000	-61,000	-317,000	-105,000	-61,000
	$c_1'/c_1 = 5$	-184,000	-184,000	-151,000	–361,000	-151,000	-151,000
Stopping time	$c_1'/c_1 = 1$	34	34	2	277	84	1
	$c_1'/c_1 = 2$	34	34	1	257	45	1
	$c_1'/c_1 = 5$	34	34	1	211	1	1
PCD	$c_1'/c_1 = 1$	0.10	0.30	0.85	0.96	1	1
	$c_1'/c_1 = 2$	0.10	0.30	0.90	1	1	1
	$c_1'/c_1 = 5$	0.10	0.30	0.98	1	1	1

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

integer. In reporting the results for the diffusion approximation method, we assume $\Delta t = 0.05$. The interval in the x-axis is calculated by the procedure described in Section 3 and is dependent on Δt. Our numerical analysis shows that refining the grid cell sizes by changing $\Delta t = 0.01$, or expanding the range over x-axis does not contribute to the solution significantly. Note that the range for the t-axis cannot be changed.

Table 14 Simulation-based gridding

Sigmoid dose-response curve				
	Expected utility	True utility	Stopping time	PCD
Grid interval=0.5	10,001,441	9,211,000	289	1
Grid interval=0.25	10,007,286	9,220,000	280	1
flat dose-response curve				
Grid interval=0.5	-11,199	-314,000	284	0.69
Grid interval=0.25	-8,907	-319,000	289	0.69

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

Algorithm 1 Simulation-based Gridding Algorithm

Input. The allocation scheme z^n and state (μ^n, Σ^n).
######## Forward simulations: Populate the grid with M experiments ########
for $i := 1$ **to** M **do**
  for $k := n$ **to** N **do**
    Using the fixed and given allocation scheme, evaluate dose z_i^k.
    Simulate future observation $\hat{y}_i^{k+1} | \mu_i^k, \Sigma_i^k, z_i^k \sim \mathcal{N}(\mu_{i,z}^k, \sigma^2 + \Sigma_{i,zz}^k)$.
    Update the state using $\hat{y}_i^{k+1}$ to obtain $(\mu_i^{k+1}, \Sigma_i^{k+1})$ by equation (2).
    Generate T samples of dose-response $\Theta_{i,t}^k \sim \mathcal{N}(\mu_i^{k+1}, \Sigma_i^{k+1})$.
    Estimate the target dose $z_{i,t}^{*,k}$ for each $\Theta_{i,t}^k$ using equation (1).
    Let $df_{i,t}^{*,k} = f(z_{i,t}^{*,k}, \Theta_{i,t}^k) - f(0, \Theta_{i,t}^k)$.
    Estimate m_i^k and ν_i^k using T-sample mean and T-sample variance.
    Record the trajectory of (m_i^k, ν_i^k) in the grid (m, ν, k) for experiment i.
  end for
end for
########## Forward simulations: Populate the empty cells ##########
for each empty cell j in the grid $(m, \nu, n : N)$ **do**
  Identify (m_j, ν_j, n_j).
  for $i := 1$ **to** M' **do**
    for $k := n_j$ **to** N **do**
      Repeat steps 5-12.
    end for
  end for
end for
################## Backward induction ##################
for $k := N$ **to** n **do**
  for each cell j in the grid (m, ν, k) **do**
    Determine A_j^k defined in Section 5.1.
    if $k = N$ **then**
      Evaluate the optimal approximated value function by equation (9).
    else
      Evaluate the approximated utility of continuation by equation (10).
      Evaluate the optimal approximated utility by equation (11), thus the optimal decision.
    end if
  end for
end for

Algorithm 2 One-Step Look-Ahead Policy

Input. State (μ^n, Σ^n) at the beginning of each decision epoch.
Generate T samples of $\Theta_t^n \sim \mathcal{N}(\mu^n, \Sigma^n)$.
Estimate the target dose $z_t^{*,n}$ using equation (1).
Let $df_{n,t}^* = f(z_t^*, \Theta_t^n) - f(0, \Theta_t^n)$.
Estimate m_n and ν_n using T-sample mean and T-sample variance.
Generate n_p observations of $\hat{y}_{*,t}^n - \hat{y}_{0,t}^n | \Theta_t^n$.
Check whether the event $B_t^n := \{\frac{\sqrt{n_p}(\bar{y}_{*,t} - \bar{y}_{0,t})}{\sqrt{2\sigma^2}} > q_\alpha\}$ holds true.
Estimate $\mathbb{E}[\mathbb{1}_{\{B^n\}} | \mathscr{F}^n]$ by taking a sample average over T samples.
Evaluate the value of termination, i.e., $a^n = 2$ using equation (3).
##################One-step forward simulation####################
for $t := 1$ **to** T **do**
    Simulate future observation $\hat{y}_t^{n+1} | \Theta_t^n, z^n \sim \mathcal{N}(\theta_{t,z^n}, \sigma^2)$.
    Update the state using $\hat{y}_t^{n+1}$ to obtain $(\mu_t^{n+1}, \Sigma_t^{n+1})$ by equation (2).
    Generate M posterior samples of $\Theta_{t,m}^{n+1} \sim \mathcal{N}(\mu_t^{n+1}, \Sigma_t^{n+1})$.
    Repeat steps 3-9 to evaluate the value of termination by taking sample average of M values.
end for
Evaluate $V_{a^n=1}^{\mathrm{KG}}$ via equation (12) by taking a sample average of T estimated termination values in the above "for" loop.
if $V_{a^n=1}^{\mathrm{KG}} > 0$ **then**
    The optimal decision is to continue, go to step 2, $n \leftarrow n+1$.
else
    Terminate or abandon the trial using the expected value of termination evaluated at step 9.
end if

References

Arlotto A, Gans N, Chick S (2010) Optimal employee retention when inferring unknown learning curves. *Simulation Conference (WSC), Proceedings of the 2010 Winter*, 1178–1188 (IEEE).

Battelle Technology Partnership Practice (2015) Biopharmaceutical industry-sponsored clinical trials: Impact on state economies. URL `http://phrma-docs.phrma.org/sites/default/files/pdf/biopharmaceutical-industry-sponsored-clinical-trials-impact-on-state-economies.pdf`.

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. *Case Studies in Bayesian Statistics*, 99–181 (Springer).

Berry SM, Carlin BP, Lee JJ, Muller P (2010) *Bayesian adaptive methods for clinical trials* (CRC press).

Grand View Research (2019) Breast cancer drugs market size, share & trends analysis report by type (hormonal receptors, mitotic inhibitors, HER2 inhibitors, anti-metabolites, CDK 4/6 inhibitors), by region, and segment forecasts, 2019 - 2025. URL `https://www.grandviewresearch.com/industry-analysis/breast-cancer-drugs-market`.

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

Hofvind S, Lee CI, Elmore JG (2012) Stage-specific breast cancer incidence rates among participants and non-participants of a population-based mammographic screening program. *Breast cancer research and treatment* 135(1):291–299.

Ingber L, Chen C, Mondescu RP, Muzzall D, Renedo M (2001) Probability tree algorithm for general diffusion processes. *Physical Review E* 64(5):056702.

Krams M, Lees KR, Hacke W, Grieve AP, Orgogozo JM, Ford GA, et al. (2003) Acute stroke therapy by inhibition of neutrophils (ASTIN) an adaptive dose-response study of UK-279, 276 in acute ischemic stroke. *Stroke* 34(11):2543–2548.

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

Liu F, Walters SJ, Julious SA (2017) Design considerations and analysis planning of a Phase 2: a proof of concept study in rheumatoid arthritis in the presence of possible non-monotonicity. *BMC Medical Research Methodology* 17(1):149.

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

Nasrollahzadeh AA, Khademi A (2018) Dynamic programming for response-adaptive dose-finding clinical trials. *Submitted for publication* .

Schmid D (2017) Population of europe 2016, by country and gender. URL `https://www.statista.com/statistics/611318/population-of-europe-by-country-and-gender`.

Warner P, Weir C, Hansen C, Douglas A, Madhra M, Hillier S, Saunders P, Iredale J, Semple S, Walker B, et al. (2015) Low-dose dexamethasone as a treatment for women with heavy menstrual bleeding: protocol for response-adaptive randomised placebo-controlled dose-finding parallel group trial (dexfem). *BMJ Open* 5(1):e006837.

World Health Organization and others (2009) Principles and methods for the risk assessment of chemicals in food .