Bayesian Analysis (2013) **8**, Number 3, pp. 703–722

A Bayesian Dose-finding Design for Drug Combination Trials with Delayed Toxicities

Suyu Liu * and Jing Ning †

Abstract. We propose a Bayesian adaptive dose-finding design for drug combination trials with delayed toxicity. We model the dose-toxicity relationship using the Finney model, a model widely used in drug-drug interaction studies. The intuitive interpretations of the Finney model facilitate incorporating the available prior dose-toxicity information from single-agent trials into combination trials through prior elicitation. We treat unobserved delayed toxicity outcomes as missing data and handle them using Bayesian data augmentation. We conduct extensive simulation studies to examine the operating characteristics of the proposed method under various practical scenarios. Results show that the proposed design is safe and able to select the target dose combinations with high probabilities.

Keywords: Adaptive design, Late-onset toxicity, Combining drugs, Missing data, Maximum tolerated dose, Phase I trial

1 Introduction

The goal of phase I drug combination clinical trials is to find the maximum tolerated dose (MTD) combinations of multiple agents that will then be further examined for synergistic treatment effects in subsequent phase II and III trials. Numerous designs have been proposed for phase I drug combination clinical trials, including those by Thall et al. (2003); Conaway et al. (2004); Wang and Ivanova (2005); Yuan and Yin (2008); Yin and Yuan (2009b); Yin and Yuan (2009a); Braun and Wang (2010); and Wages et al. (2011), among others. A fundamental assumption underlying these adaptive dose-finding methods is that the toxicity outcome can be observed quickly, such that, by the time of the next dose assignment, complete information on toxicity will be available for the currently treated patients. In other words, the follow-up for all enrolled patients must be completed before a new patient can be enrolled.

However, in practice, toxicities are not always immediately observable and sometimes require a relatively long follow-up time for assessment. For example, in radiotherapy trials, dose-limiting toxicities (DLTs) often occur long after the treatment is finished (Coia et al. 1995; Cooper et al. 1995). Muler et al. (2004) reported a phase I trial of combined cisplatin and gemcitabine to treat patients with pancreatic cancer. In that trial, patients were required to be followed for 9 weeks to fully assess toxicity (e.g., grade 4 thrombocytopenia and neutropenia), but the patient accrual rate was as fast

*Department of Biostatistics, The University of Texas MD Anderson Cancer Center, TX, syliu@mdanserson.org

†Department of Biostatistics, The University of Texas MD Anderson Cancer Center, TX, jning@mdanserson.org

 DOI:10.1214/13-BA839

as one per week. Consequently, by the time of dose assignment for a new patient, some patients under treatment had not completed follow-up and might experience toxicity later during the follow-up period (i.e., late-onset toxicity).

The issue of late-onset toxicity is of particular importance in the emerging era of novel molecularly targeted agents. A recent review paper in the *Journal of Clinical Oncology* found that among a total of 445 patients in 36 trials involving molecularly targeted agents, 57% of the grade 3 and 4 toxicities were late-onset (Postel-Vinay et al. 2011).

In the context of single-agent clinical trials, a handful of methods have been proposed to address the issue of late-onset toxicity. Cheung and Chappell (2000) proposed the time-to-event continual reassessment method, in which subjects who have not experienced toxicity at a given time point are weighted by the proportions of their follow-up times with respect to the entire evaluation period, and subjects who have experienced toxicity are given a full weight of 1. Mauguen et al. (2011) extended this weighting approach to the escalation with overdose control design. From a different perspective, Yuan and Yin (2011) treated late-onset toxicities as a missing data problem and proposed an expectation-maximization (EM) algorithm to handle the unobserved toxicity outcome. However, because these designs focus on single-agent trials and are based on the assumption that toxicity monotonically increases with the dose, they cannot be directly applied to drug combination trials, for which the dose-toxicity relationship is only partially ordered in the two-dimensional drug combination space (Conaway et al. 2004).

In this article, we propose a Bayesian dose-finding design for drug combination trials with late-onset toxicities. We model the two-dimensional dose-toxicity surface using the Finney model (Finney 1971), which is widely used in drug-drug interaction studies (Greco et al. 1995). The intuitive interpretations of the Finney model greatly facilitate our ability to understand the model and elicit priors for the model parameters. We treat the late-onset toxicity as missing data and handle these missing data using the Bayesian data augmentation approach. By "imputing" the missing toxicity outcome, our design not only efficiently incorporates the partial information from patients who have not completed follow-up but also supports real-time dose assignment for newly accrued patients.

The rest of the paper is organized as follows. In Section 2, we present the Finney model and propose the Bayesian data augmentation approach to handle the missing data caused by late-onset toxicity. In Section 3, we outline the estimation procedure; we describe the dose-finding algorithm for the proposed design in Section 4. In Section 5, we present extensive simulation studies to examine the operating characteristics of the new design. We conclude with a brief discussion of our findings in Section 6.

2 Method

2.1 Dose-toxicity Model

Let a_j be the jth dose for agent A, $a_1 < \ldots < a_J$, b_k be the kth dose for agent B, $b_1 < \ldots < b_K$, and (j, k) represent the combination of dose a_j and dose b_k. We assume that the toxicity probability of combination (j, k), denoted as p_{jk}, follows the Finney model of the form

$$\text{logit}(p_{jk}) = \beta_0 + \beta_1 \log(a_j + \rho b_k + \gamma(a_j \rho b_k)^{1/2}), \tag{1}$$

where β_0, β_1, γ, and ρ (> 0) are unknown parameters. We require the regression slope $\beta_1 > 0$ to ensure that the probability of toxicity increases with the dose. One important advantage of the Finney model is its intuitive interpretations for drug combination studies. Specifically, ρ measures the relative potency of agent B versus agent A to induce toxicity, that is, the amount of agent A required to produce the same (toxicity) effect as a unit of agent B. A value of $\rho > 1$ indicates that, given the same dose, agent B is more likely to cause toxicity than agent A. The drug-drug interaction between A and B is characterized by the synergy-antagonism parameter γ, with $\gamma = 0$ corresponding to the Loewe additivity effect (Loewe and Muischnek 1926), $\gamma > 0$ to synergy, and $\gamma < 0$ to antagonism.

Another important advantage of the Finney model is that, unlike the dose-toxicity models used in most existing designs, which have been devised mainly for statistical convenience, the Finney model has been extensively studied and validated by real-world data (Greco et al. 1995; Govindarajulu 2001). Thus, the Finney model is expected to provide a good approximation of the true dose-toxicity relationship. In addition, when each agent is administered individually (i.e., setting the dose of the other agent to zero), the Finney model becomes the standard logistic model:

$$\text{logit}(p_{j0}) = \beta_0 + \beta_1 \log(a_j), \tag{2}$$

$$\text{logit}(p_{0k}) = \beta_0 + \beta_1 \log(\rho b_k), \tag{3}$$

where p_{j0} denotes the toxicity probability of dose level j for agent A as a single agent and p_{0k} denotes the toxicity probability of dose level k for agent B as a single agent.

Suppose that n patients have entered the trial. For the ith subject, let y_i and (j_i, k_i) denote the binary toxicity outcome and the received dose combination, respectively. When the toxicity outcomes $\boldsymbol{y} = \{y_i, i = 1, \ldots, n\}$ are fully observed, the *complete-data* likelihood function is given by

$$L(\boldsymbol{y}|\boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{\exp\{y_i\beta_0 + y_i\beta_1 \log(a_{j_i} + \rho b_{k_i} + \gamma(a_{j_i}\rho b_{k_i})^{1/2})\}}{1 + \exp\{\beta_0 + \beta_1 \log(a_{j_i} + \rho b_{k_i} + \gamma(a_{j_i}\rho b_{k_i})^{1/2})\}},$$

and the posterior distribution of $\boldsymbol{\theta} = (\beta_0, \beta_1, \gamma, \rho)$ is

$$f(\boldsymbol{\theta}|\boldsymbol{y}) \propto f(\boldsymbol{\theta}) L(\boldsymbol{y}|\boldsymbol{\theta}), \tag{4}$$

where $f(\boldsymbol{\theta})$ is the prior distribution of $\boldsymbol{\theta}$. Unfortunately, when toxicity is of late onset, this complete-data likelihood function is not available because some values of $\boldsymbol{y}$ are missing, as described below.

2.2 Accommodating Late-onset Toxicity

In a phase I dose-finding trial, we accrue patients sequentially and follow them for a fixed period of time $(0, T)$ to assess their toxicity outcome y_i. In contrast to the common misconception that late-onset toxicity is that which requires a long assessment time T, whether or not the toxicity is of late onset actually depends on the relative length of the assessment time and the inter-arrival time of the patients, (i.e., the ratio of the assessment time and the inter-arrival time, denoted as the A/I ratio). In other words, late-onset toxicity is a relative concept. If the assessment time T is not longer than the inter-arrival time of the patients (i.e., the A/I ratio ≤ 1), the toxicity is not of late onset, even when T is very large. This is because by the time of dose assignment for a newly arrived patient, all the previously treated patients would have finished the assessment period, and their toxicity outcomes would be fully observed. By contrast, if T is longer than the patient inter-arrival time (i.e., the A/I ratio > 1), for example in the case of a fast accrual, the toxicity may be of late onset even when T is short. This is because at the moment of dose assignment for a new cohort of patients, some previously enrolled patients would not have completed their assessment period. Even though these patients would not have experienced toxicity at that point in time, they might yet experience toxicity during the remaining period of follow-up.

A consequence of late-onset toxicity is that some patients' toxicity outcomes y_i may be subject to missingness. For subject i, let t_i denote the time to toxicity, and let s_i $(s_i \leq T)$ denote the actual follow-up time at the moment of dose assignment for a newly arrived cohort. If subject i will not experience toxicity (i.e., $y_i = 0$), we set $t_i = \infty$. Then, y_i is missing when $t_i > s_i$ and $s_i < T$, and is observed otherwise. That is, the toxicity outcome is missing for patients who have not yet experienced toxicity $(t_i > s_i)$ and have not been fully followed to T $(s_i < T)$. In the case that y_i is observed, $y_i = 1$ if $t_i \leq s_i$ and 0 if $t_i > s_i = T$.

A natural approach to handling the unobserved toxicity outcome is to "impute" and "fill in" the missing data so that standard complete-data methods can be applied. Under the Bayesian paradigm, this can be achieved using data augmentation (Tanner and Wong 1987). The data augmentation process consists of two iterative steps: the imputation (I) step, in which the missing data are imputed, and the posterior (P) step, in which the posterior samples of unknown parameters are simulated based on imputed data. Specifically, at the I step, we impute the missing data by drawing samples from their posterior predictive distribution

$$f(y_i | t_i > s_i, \boldsymbol{\theta}) = \text{Bernoulli}(\pi_i)$$

with

$$\begin{aligned}\pi_i &= \text{pr}(y_i = 1 | t_i > s_i, \boldsymbol{\theta}) \\ &= \frac{\text{pr}(y_i = 1|\boldsymbol{\theta})\text{pr}(t_i > s_i | y_i = 1)}{\text{pr}(y_i = 0|\boldsymbol{\theta})\text{pr}(t_i > s_i|y_i = 0) + \text{pr}(y_i = 1|\boldsymbol{\theta})\text{pr}(t_i > s_i|y_i = 1)} \\ &= \frac{p_{j_i k_i}\text{pr}(t_i > s_i|y_i = 1)}{1 - p_{j_i k_i} + p_{j_i k_i}\text{pr}(t_i > s_i|y_i = 1)}, \end{aligned} \tag{5}$$

where $p_{j_i k_i}$ is given in model (1). Because π_i involves the unknown survival probability $\text{pr}(t_i > s_i|y_i = 1)$, in order to impute the missing value of y_i, we need to model the time to toxicity t_i for patients who will experience toxicity (i.e., $y_i = 1$). For this purpose, we specify a flexible piecewise exponential model for the time to toxicity for patients who will experience toxicity.

Specifically, we partition the assessment period $[0, T]$ into L disjoint intervals $[0, u_1)$, $[u_1, u_2)$, …, $[u_{L-1}, u_L \equiv T)$ and assume a constant hazard λ_l in the lth interval. We set the hazard at time T as infinity to ensure that the survival function drops to 0 at the end of the follow-up. Typically, it is adequate to set the number of partitions equal to the average number of patients/cohorts who have not finished their follow-up (i.e., $L =$ the A/I ratio -1) because that is the number of survival rates required to be evaluated. Define the observed time $x_i = \min(s_i, t_i)$ and $\delta_{il} = 1$ if the ith subject experiences toxicity in the lth interval and $\delta_{il} = 0$ otherwise. Let $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_L)$. Conditional on $\boldsymbol{y}$, the likelihood function of $\boldsymbol{\lambda}$ is given by

$$L(\boldsymbol{x}|\boldsymbol{\lambda}, \boldsymbol{y}) = \prod_{i=1}^{n}\prod_{l=1}^{L}(\lambda_l)^{\delta_{il}}\exp\{-y_i\lambda_l d_{il}\},$$

where $d_{il} = u_l - u_{l-1}$ if $x_i > u_l$; $d_{il} = x_i - u_{l-1}$ if $x_i \in [u_{l-1}, u_l)$ and $d_{il} = 0$ otherwise. To pool information across different doses and obtain more reliable estimates, we here assume that the time-to-toxicity distribution is invariant to the dose level, conditioning on the patient experiencing toxicity ($y_i = 1$). The sensitivity analysis in Section 5.2 shows that our method is not sensitive to the violation of this assumption.

Under this piecewise exponential model, the Bernoulli probability π_i in equation (5) is given by

$$\pi_i = \frac{p_{j_i k_i}\exp(-\sum_{l=1}^{L}\lambda_l d_{il})}{1 - p_{j_i k_i} + p_{j_i k_i}\exp(-\sum_{l=1}^{L}\lambda_l d_{il})}, \tag{6}$$

which can be used to draw the missing data for the I step. After completing the I step, we turn to the P step to draw the posterior samples of unknown parameters conditional on the imputed data. In this step, the standard Markov chain Monte Carlo method for complete data can be used because the missing data have been imputed in the I step. The details of implementing the data augmentation are provided in Section 3.

2.3 Specification of Priors

The intuitive interpretations of the Finney model greatly facilitate the prior specification for unknown parameters. Noting that in practice the relative (toxicity) potency of agent B versus agent A is most likely within a 10-fold difference, we assigned the relative potency parameter ρ a lognormal prior $LN(0.81, 0.81)$, which covers the range (0.1, 10) about 95% of the time. In addition, because the prior mode of ρ is 1, this prior naturally centers around the neutral opinion that agent A and B have the same potency. For the synergy-antagonism parameter γ, we chose a normal prior $N(3.3, 4)$ based on two considerations. (1) For toxicity, antagonism is much rarer than synergy. Under this prior, the prior probability of antagonism is about 5%. (2) A two–standard-deviation change from the prior mean yields a range of (-0.7, 7.3), which covers the values of γ typically encountered in practice.

The prior specification for the intercept β_0 and slope β_1 requires more consideration. In practice, it is almost invariably true that before two agents are combined, each of them has been thoroughly investigated individually. Therefore, the investigator has sufficient prior information regarding the toxicity profile of each of the agents as a single agent, i.e., the values (or estimates) of the p_{j0}'s and p_{0k}'s, denoted as $\hat{p}_{j0}$ and $\hat{p}_{0k}$. We incorporate this prior information into the design through specifying appropriate priors for β_0 and β_1 and ascribing "effective" dose values to a_j and b_k. That is, the values of a_j and b_k that we use to fit the model will not be the actual clinical dose values but instead transformed values that are consistent with the single-agent toxicity profiles. The approach of using the effective dose has been used in model-based phase I trial designs, such as the continual reassessment method (CRM), to improve the estimation stability and operating characteristics of designs (Paoletti and Kramar 2009; Yin and Yuan 2009a).

Without a loss of generality, we set $a_1 = 1$ and $a_J = 10$ to fix the scale of the dose for agent A. Then, given the prior estimates of the single-agent toxicity probabilities $\hat{p}_{j0}$ and $\hat{p}_{0k}$, we calculate the effective dose as follows: based on the single-agent dose-toxicity model (2), we first back-solve the values of β_0 and β_1

$$\hat{\beta}_0 = \text{logit}(\hat{p}_{10}) \tag{7}$$

$$\hat{\beta}_1 = \log\left\{\frac{\hat{p}_{J0}/(1-\hat{p}_{J0})}{\hat{p}_{10}/(1-\hat{p}_{10})}\right\}/\log(10). \tag{8}$$

Then we calculate the effective dose values of agent A and B as

$$a_j = \exp[\{\text{logit}(\hat{p}_{j0}) - \hat{\beta}_0\}/\hat{\beta}_1], \tag{9}$$

$$b_k = \exp[\{\text{logit}(\hat{p}_{0k}) - \hat{\beta}_0\}/\hat{\beta}_1], \tag{10}$$

for $j = 2, \ldots, J-1$ and $k = 1, \ldots, K$ (by setting $\rho = 1$). By doing this, we ensure that when $\beta_0 = \hat{\beta}_0$ and $\beta_1 = \hat{\beta}_1$, the single-agent toxicity probabilities for a_j and b_k match the prior estimates $\hat{p}_{j0}$ and $\hat{p}_{0k}$. Subsequently, we center the prior distributions of β_0 and β_1 at $\hat{\beta}_0$ and $\hat{\beta}_1$, respectively. We assign β_0 a normal prior $N(\hat{\beta}_0, \sigma_0^2\hat{\beta}_0^2)$ and set $\sigma_0 = 2$ so that the prior standard deviation is twice the prior mean. Theoretically, this

prior may run into the problem of the variance being 0 if $\hat{\beta}_0 = 0$, however that problem almost never occurs in practice. This is because $\hat{\beta}_0 = 0$ means $\hat{p}_{10} = 0.5$, and the latter is extremely unlikely as the toxicity rate of the lowest dose of agent A is almost always well below 0.5. We assign β_1 a gamma prior with mean $\hat{\beta}_1$ and scale parameter σ_1. We set $\sigma_1 = 4$ so that the prior variance is four times as large as the prior mean.

For the piecewise exponential time-to-toxicity model, we assign each component of $\boldsymbol{\lambda}$ an independent gamma prior distribution. As a neutral prior opinion between toxicity occurring late versus early during the follow-up, we assume that *a priori* toxicity occurs uniformly throughout the follow-up period $(0, T)$. Under this assumption, the hazard at the middle of the lth partition is $\tilde{\lambda}_l = L/\{T(L - l + 0.5)\}$. Thus, we assign λ_l a gamma prior distribution with mean $\tilde{\lambda}_l$ and variance $D\tilde{\lambda}_l$. The simulation study shows that $D = 2$ represents a reasonably vague prior and generally yields good operating characteristics. When prior knowledge is available regarding the shape of the hazard for the time to toxicity, the hyperparameters of the gamma prior can be tuned to match the prior information.

3 Estimation

We now describe the estimation procedure for the proposed models based on Bayesian data augmentation. Let $\boldsymbol{y} = (\boldsymbol{y}_{\text{obs}}, \boldsymbol{y}_{\text{mis}})$, where $\boldsymbol{y}_{\text{obs}}$ and $\boldsymbol{y}_{\text{mis}}$ denote the observed and missing toxicity data, respectively; and let $\mathcal{D}_{\text{obs}} = (\boldsymbol{y}_{\text{obs}}, \boldsymbol{r})$ denote the observed data with $\boldsymbol{r} = \{r_i, i = 1, \ldots, n\}$ where r_i is the missing data indicator. If data are missing, $r_i = 1$, otherwise $r_i = 0$. Note that in our case, as described by Yuan and Yin (2011), the missing data induced by the late-onset toxicity are nonignorable because y_i is more likely to be missing when $y_i = 0$ than when $y_i = 1$; that is, the probability of y_i being missing depends on the value of y_i. Therefore, the observed data used for inference include not only the observed toxicity outcomes $\boldsymbol{y}_{\text{obs}}$ but also the missing data indicators $\boldsymbol{r}$. Inference that ignores $\boldsymbol{r}$ would lead to biased estimates.

Our estimation procedure iterates between the I step and the P step until the Markov chain converges. At the I step, we impute the missing data by drawing samples from their full conditional distribution, $f(y_i|\mathcal{D}_{\text{obs}}, \boldsymbol{\theta}, \boldsymbol{\lambda})$, i.e.,

$$\text{Bernoulli}\left(\frac{\exp\{\beta_0 + \beta_1 \log(a_{j_i} + \rho b_{k_i} + \gamma(a_{j_i}\rho b_{k_i})^{1/2}) - \sum_{l=1}^{L} \lambda_l d_{il}\}}{1 + \exp\{\beta_0 + \beta_1 \log(a_{j_i} + \rho b_{k_i} + \gamma(a_{j_i}\rho b_{k_i})^{1/2}) - \sum_{l=1}^{L} \lambda_l d_{il}\}}\right).$$

Then, at the P step, given the imputed data $\boldsymbol{y}$, we sequentially sample the unknown model parameters from their full conditional distributions, as follows:

(i) Sample each component of $\boldsymbol{\theta}$ sequentially from their full conditional distribution, which is proportional to $f(\boldsymbol{\theta}|\boldsymbol{y})$ given by equation (4).

(ii) Sample $\lambda_l, l = 1, \ldots, L$, from

$$f(\lambda_l|\boldsymbol{y}) = Ga\Big(\tilde{\lambda}_l/D + \sum_{i=1}^{n} \delta_{il}, 1/D + \sum_{i=1}^{n} y_i d_{il}\Big).$$

where $Ga(a, b)$ denotes a gamma distribution with the shape parameter a and the rate parameter b.

We iteratively draw a sequence of samples of the missing data and model parameters through the I step and P step. When the algorithm converges, the sequence of samples of $\boldsymbol{\theta}$ converges to its marginal posterior distribution, which can be used to make inferences about the toxicity probabilities and direct dose finding.

4 Trial Design and Conduct

As is common to model-based clinical trial designs, the model-based dose-finding algorithm is difficult to apply at the beginning of the trial, because very limited information is available and the posterior estimates of the probabilities of toxicity for dose combinations are highly variable. When toxicity is of late onset, this issue is of even greater concern because it is possible that none of the toxicity outcomes will be observed during the early stage of the trial due to the missing data. To overcome this difficulty, we use a start-up rule to obtain some preliminary toxicity data before switching to the model-based dose-finding algorithm. Specifically, we begin the start-up phase by treating the first cohort of patients at the lowest dose combination (a_1, b_1) and then escalate the dose along the diagonal of the dose combination matrix from (a_1, b_1) to (a_2, b_2) and so on until the first toxicity occurs. If the dose matrix is not square, say $J > K$, we first escalate the dose along the diagonal from (a_1, b_1) to (a_2, b_2) and so on until we reach (a_K, b_K); thereafter, we escalate the dose by holding the dose level of B at K and increasing the dose level of A from (a_K, b_K) to (a_{K+1}, b_K) and so on until we reach the highest dose combination (a_J, b_K). During this start-up phase, we require that each of the treated patients is fully monitored for toxicity assessment before a new patient is enrolled. Upon observing one toxicity outcome, the start-up phase is completed. Note that although we escalate the dose along the diagonal, because the start-up rule is very conservative (i.e., terminates after the first toxicity occurs), further ensuring patient safety is not of great concern. The maximum number of patients experiencing toxicity in the start-up phase will not be greater than that of one cohort, which in practice is typically no more than 3 patients, making our method even safer than the standard 3+3 design. The advantage of escalating along the diagonal is that it accumulates the dose-toxicity information for two agents simultaneously. This approach has been used previously by Thall et al. (2003) and Houede et al. (2010).

After the start-up phase is completed, we switch to the following model-based dose-finding algorithm. Let ϕ be the physician-specified target toxicity limit, and c_e and c_d be the fixed probability cutoffs for dose escalation and de-escalation, respectively. The values of c_e and c_d can be calibrated through simulation studies to obtain desirable operating characteristics. To guard against dramatic changes in dose and to improve safety, at this stage we restrict dose escalation or de-escalation of each agent to no more than one dose level and prohibit dose escalation of two agents at the same time (i.e., dose escalation along the diagonal). The model-based dose-finding algorithm is described as follows:

i. If at the current dose combination (j, k),

$$\Pr(p_{jk} < \phi | \mathcal{D}_{\text{obs}}) > c_e,$$

the dose will be escalated to one of the adjacent dose combination $\{(j+1, k), (j, k+1), (j+1, k-1), (j-1, k+1)\}$, at which the posterior mean of the toxicity probability is higher than p_{jk} and closest to ϕ.

ii. If at the current dose combination (j, k),

$$\Pr(p_{jk} > \phi | \mathcal{D}_{\text{obs}}) > c_d,$$

the dose will be de-escalated to one of the adjacent dose combinations $\{(j-1, k), (j, k-1), (j+1, k-1), (j-1, k+1), (j-1, k-1)\}$, at which the posterior mean of the toxicity probability is lower than p_{jk} and closest to ϕ.

iii. Otherwise, the next cohort of patients will continue to be treated at the current dose combination (i.e., the current dose combination is retained).

iv. Once the maximum sample size is reached, the dose combination that has a posterior mean of the toxicity probability closest to ϕ is selected as the MTD combination.

For patient safety, we impose the following stopping rule in our algorithm: if $\text{pr}(p_{11} > \phi | \mathcal{D}_{\text{obs}}) > 0.8$, then the trial is terminated and no MTD is selected. In practice, additional stopping rules can be used to allow early selection of the MTD. For example, we can terminate the trial early and select the MTD when a certain number of consecutive cohorts are treated at the same dose.

In addition to finding the MTD combination, we may also be interested in identifying a set of safe doses as the candidates to be further investigated in phase II trials for efficacy. We define the set of safe doses as the doses that satisfy the safety requirement $\text{pr}(p_{jk} < \phi | \mathcal{D}) > c_s$, where c_s is a prespecified safety threshold.

5 Numerical Examples

5.1 Simulation Study

We examined the operating characteristics of the proposed design by simulating all 8 toxicity scenarios previously considered by Yin and Yuan (2009b), in which each agent had four dose levels (see Table 1). The target toxicity probability was $\phi = 0.3$, and the maximum sample size was 20 cohorts of 3 patients each. We assumed that the follow-up time for assessing toxicity was 3 months and that the patient accrual followed a Poisson process with the rate of three patients per month (i.e., $A/I = 3$).

We simulated times to toxicity based on a Weibull distribution. At each dose level, the scale and shape parameters of the Weibull distribution were chosen such that (1) the cumulative distribution function at time T, the end of the follow-up, is the toxicity

Table 1: Eight toxicity scenarios for a two-agent combination trial with a target probability of toxicity of 0.3. The target maximum tolerated dose combinations are shown in boldface.

		Drug A							
	Dose Level	1	2	3	4	1	2	3	4
		Scenario 1				Scenario 2			
	4	0.50	0.55	0.60	0.70	0.48	0.52	0.55	0.58
	3	0.15	**0.30**	0.50	0.60	0.42	0.45	0.50	0.52
	2	0.10	0.12	**0.30**	0.45	**0.30**	0.40	0.48	0.50
	1	0.06	0.08	0.10	0.15	0.15	**0.30**	0.40	0.45
		Scenario 3				Scenario 4			
	4	**0.30**	0.50	0.55	0.60	0.50	0.55	0.60	0.70
	3	0.12	**0.30**	0.50	0.55	**0.30**	0.50	0.55	0.60
	2	0.10	0.15	**0.30**	0.45	0.12	**0.30**	0.50	0.55
	1	0.08	0.12	0.16	0.18	0.10	0.15	**0.30**	0.45
Drug B		Scenario 5				Scenario 6			
	4	0.20	**0.30**	0.45	0.50	**0.30**	0.50	0.60	0.70
	3	0.16	0.18	**0.30**	0.45	0.15	**0.30**	0.52	0.60
	2	0.14	0.16	0.20	**0.30**	0.10	0.20	**0.30**	0.55
	1	0.08	0.13	0.16	0.18	0.08	0.14	0.19	**0.30**
		Scenario 7				Scenario 8			
	4	0.16	0.18	0.20	**0.30**	0.70	0.75	0.80	0.85
	3	0.13	0.16	0.18	0.20	0.60	0.65	0.70	0.80
	2	0.12	0.14	0.16	0.18	0.55	0.60	0.65	0.70
	1	0.10	0.12	0.14	0.16	0.50	0.55	0.60	0.65

probability of that dose and (2) among all the toxicities that occur in $(0, T)$, half of them occur in $(T/2, T)$, the latter half of the follow-up period. The first condition ensures that the toxicity probability of each dose matches that given in Table 1, and the second condition ensures that the toxicity event is of late onset. Because the toxicity probability varies across the dose levels, the scale and shape parameters of the Weibull distribution need to be different for different dose levels.

We used $L = 2$ partitions to construct the piecewise exponential time-to-toxicity model. We took the single-agent toxicity probabilities $(p_{10}, \ldots, p_{40})$ =(0.06, 0.12, 0.20, 0.30) and $(p_{01}, \ldots, p_{04})$ =(0.04, 0.08, 0.16, 0.30) and set $c_e = 0.8$ and $c_d = 0.45$ to direct dose escalation and de-escalation. We compared the proposed design with the latent-table design proposed by Yin and Yuan (2009b). Like most existing combination trial designs, the latent-table design requires fully observed data for dose assignment. Therefore, when implementing that design, we suspended accrual until all of the toxicity outcomes in the trial were completely observed prior to the next dose assignment, although it may not be feasible in practice. For the purpose of comparison, we also implemented

a complete-data version of the proposed design, which suspended accrual and made dose assignment based on completely observed data as the latent-table design. This complete-data design provides an optimal upper bound and benchmark "limit" of the proposed design. Under each scenario, we simulated 1,000 trials.

Table 2 displays the selection percentage, the average percentage of patients treated at each dose combination under the latent-table and proposed designs. For ease of comparison, these simulation results are summarized in Table 3, including the total selection percentage of the target MTD combinations, the selection percentage of overly toxic doses (i.e, doses with toxicity probabilities larger than the target ϕ), the percentage of patients treated at the MTD combinations, the percentage of patients treated at overly toxic doses, and the average trial duration. Note that when the toxicity is of late onset, it is particularly important to examine the selection percentage of overly toxic doses and the percentage of patients treated at the overly toxic doses because the danger of late-onset toxicity is that patients who have not experienced toxicity at a given time during follow-up may still experience toxicity during the late phase of follow-up.

In particular, scenario 1 had two MTD combinations in the two-dimensional space. The proposed design outperformed the latent-table design with a 27.7% higher selection percentage of the target MTD combinations and 9.3% more patients allocated to the MTD combinations. The proposed design was also safer. It selected the overly toxic doses 21.2% less frequently and treated 6.0% fewer patients at the overly toxic doses compared to the latent-table design. More notably, because our proposed design allows for continual accrual, the trial duration using this design was less than half of that using the latent-table design, shortening the trial from 56.9 months to 26.8 months. The proposed design tended to assign more patients to the toxic doses (i.e., the doses above the MTD) located on the diagonal than the latent-table design. This was because the two designs used different start-up rules: the proposed design escalated the dose along the diagonal, whereas the latent-table design escalated the dose along the edges of the dose matrix. This should not be a concern because the total number of patients assigned the overly toxic doses (not limited to the diagonal) was smaller under the proposed design than the latent-table design. Overall, the performance of the proposed design was close to its optimal bound (i.e., the complete-data design), suggesting that the proposed design was able to efficiently handle the missing toxicity outcome.

Scenario 2 also had two MTD combinations, but at different locations. In this case, the rate of selecting the MTD combinations by the proposed design was 9.3% higher than that of the latent-table design, but the latter allocated 7.4% more patients to the MTD combinations. In terms of safety, the two designs are comparable, with similar percentages of patients treated at the overly toxic doses and similar selection percentages for overly toxic doses. For scenarios 3 through 5, three MTD combinations were selected, but they were located at different positions. The proposed design outperformed the latent-table design under these scenarios. Compared to the latent-table design, the proposed design was better at selecting target MTD combinations and avoiding overly toxic doses. The percentages of patients treated with the target MTD combinations and overly toxic doses were often comparable between the two designs, but the proposed design cut the trial duration from about 56 months to 26 months. Scenario 6 had four

MTD combinations, and scenario 7 had only one. The selection percentage of the MTD combinations under the proposed design was comparable to that of the latent-table design in scenario 6 and was slightly higher in scenario 7. Scenario 8 was designed to examine whether the proposed method would terminate the trial early if all the dose combinations were excessively toxic. Both the proposed and latent-table designs stopped the trials before large numbers of patients were treated at overly toxic doses.

Table 4 displays the performance of the proposed designs in identifying the set of safe doses with the safety threshold $c_s = 1 - c_d = 0.55$. We can see that the proposed design was able to identify the true "safe set" with high probabilities similar to those of the complete-data design.

5.2 Sensitivity Analysis

We conducted four types of sensitivity analysis by manipulating four simulation parameters. Specifically, we examined the performance of the proposed design when (1) the true distribution of the time to toxicity was a log-logistic distribution; (2) the distribution of the time to toxicity is heterogeneous across doses by controlling $(10j + 10k)\%$ of events occurred in $(0, T/2)$ for combination (j, k), that is, higher doses result in earlier toxicities; (3) the number of partitions used for the piecewise exponential time-to-toxicity model was 6; (4) the prior distributions of β_0, β_1 and λ were more vague by setting $\sigma_0 = 4$, $\sigma_1 = 16$ and $D = 4$; and (5) the A/I ratio was 5 (under a faster accrual of one cohort per 0.6 month). Note that the A/I ratio controls the amount of missing data in the trials. When the A/I ratio is high, the new cohort arrives rapidly and the trial requires more frequent decision making about dose assignment. Consequently, at each moment of decision making, the cohorts that have already entered the trial are followed for only a short period of time, resulting in a high percentage of missing toxicity outcomes. In other words, a higher A/I ratio (e.g., 5) makes it more difficult to find the correct dose.

Table 5 shows the results of the sensitivity analysis under the 8 toxicity scenarios listed in Table 1. The results under the first three conditions are similar to those reported in Table 3, suggesting that the proposed design was robust with respect to the distribution of the time to toxicity, the number of partitions (used for the piecewise exponential model), and the prior specification. When the A/I ratio was 5, we observed that the selection percentage of the target MTD combinations and the percentage of patients treated at the MTD combinations were slightly lower than those when the A/I ratio was 3 (see Table 3). This was expected, because a high A/I ratio of 5 induced more missing data and thus less information was available to make the decision about dose assignment. Nevertheless, in general the performance of the proposed design was similar under the two A/I ratios. Because the accrual was faster under the A/I ratio of 5, the corresponding average trial duration was shorter than that under the A/I ratio of 3.

6 Concluding Remarks

We have proposed a Bayesian adaptive dose-finding design for combination trials with delayed toxicity outcomes. We model the dose-toxicity relationship using the Finney model. The advantage of using the Finney model is that the parameters in the model have intuitive interpretations, which greatly facilitates incorporating the available dose-toxicity information from single-agent trials through prior elicitation. In addition, the Finney model has been validated by empirical studies, and thus is expected to improve the goodness of fit of the model and the performance of the trial design. To accommodate delayed toxicity outcomes, we treat unobserved delayed toxicity outcomes as missing data and use Bayesian data augmentation to handle the resulting missing data. We employ a flexible piecewise exponential model to capture the partial information from the patients whose outcomes have not been observed due to incomplete follow-up. We have also conducted extensive simulation studies to examine the operating characteristics of the proposed method under various practical scenarios. The proposed design allows for continuous accrual and shortens the trial duration without sacrificing the MTD selection percentage or patient safety.

The proposed design focuses on finding the MTD, but it can be extended to a phase I/II design that simultaneously considers toxicity and efficacy. Because efficacy often requires a longer assessment time than toxicity, it is more likely that at the moment of dose assignment for a newly accrued patient, the efficacy outcome for patients who have entered the trial has not been fully assessed. Similar to the way our design accounted for late-onset toxicity, the proposed methodology based on Bayesian data augmentation can be used to handle late-onset efficacy.

References

Braun, T. M. and Wang, S. (2010). "A hierarchical Bayesian design for phase I trials of novel combinations of cancer therapeutic agents." *Biometrics*, 66: 805–812. 703

Cheung, K. Y. and Chappell, R. (2000). "Sequential designs for phase I clinical trials with late-onset toxicities." *Biometrics*, 56: 1177–1182. 704

Coia, L. R., Myerson, R., and Tepper, J. E. (1995). "Late effects of radiation therapy on the gastrointestinal tract." *International Journal of Radiation Oncology, Biology, Physics*, 31: 1213–1236. 703

Conaway, M. R., Dunbar, S., and Peddada, S. D. (2004). "Designs for single- or multiple-agent phase I trials." *Biometrics*, 60: 661–669. 703, 704

Cooper, J. S., Fu, K., Marks, J., and Silverman, S. (1995). "Late effects of radiation therapy in the head and neck region." *International Journal of Radiation Oncology, Biology, Physics*, 31: 1141–1164. 703

Finney, D. (1971). *Probit Analysis*. Cambridge: Cambridge University Press. 704

Govindarajulu, Z. (2001). *Statistical Techniques in Bioassay*. Basel: Karger., 2 edition. 705

Greco, W. R., Bravo, G., and Parsons, J. C. (1995). "The search for synergy: a critical review from a response surface perspective." *Pharmacological Reviews*, 47: 331–385. 704, 705

Houede, N., Thall, P. F., Nguyen, H., Paoletti, X., and Kramar, A. (2010). "Utility-based optimization of combination therapy using ordinal toxicity and efficacy in phase I/II trials." *Biometrics*, 66: 532–540. 710

Loewe, S. and Muischnek, H. (1926). "Effect of combinations: mathematical basis of the problem." *Naunyn-Schmiedebergs Archiv für experimentelle Pathologie und Pharmakologie*, 114: 313–326. 705

Mauguen, A., Le Deley, M. C., and Zohar, S. (2011). "Dose-finding approach for dose escalation with overdose control considering incomplete observations." *Statistics in Medicine*, 30: 1584–1594. 704

Muler, J. H., McGinn, C. J., Normolle, D., Lawrence, T., Brown, D., Hejna, G., and Zalupski, M. M. (2004). "Phase I trial using a time-to-event continual reassessment strategy for dose escalation of cisplatin combined with gemcitabine and radiation therapy in pancreatic cancer." *Journal of Clinical Oncology*, 22: 238–243. 703

Paoletti, X. and Kramar, A. (2009). "A comparison of model choices for the continual reassessment method in phase I cancer trials." *Statistics in Medicine*, 28: 3012–3028. 708

Postel-Vinay, S., Gomez-Roca, C., Molife, L. R., Anghan, B., Levy, A., Judson, I., Bono, J. D., Soria, J., Kaye, S., and Paoletti, X. (2011). "Phase I trials of molecularly targeted agents: should we pay more attention to late toxicities?" *Journal of Clinical Oncology*, 29: 1728–1735. 704

Tanner, M. A. and Wong, W. H. (1987). "The calculation of posterior distributions by data augmentation (with discussion)." *Journal of the American Statistical Association*, 82: 528–550. 706

Thall, P. F., Millikan, R. E., Müller, P., and Lee, S.-J. (2003). "Dose-finding with two agents in phase I oncology trials." *Biometrics*, 59: 487–496. 703, 710

Wages, N. A., Conaway, M. R., and O'Quigley, J. (2011). "Continual Reassessment Method for Partial Ordering." *Biometrics*, 67: 1555–1563. 703

Wang, K. and Ivanova, A. (2005). "Two-Dimensional Dose Finding in Discrete Dose Space." *Biometrics*, 61: 217–222. 703

Yin, G. and Yuan, Y. (2009a). "Bayesian dose finding in oncology for drug combinations by copula regression." *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 58: 211–224. 703, 708

— (2009b). "A latent contingency table approach to dose finding for combinations of two agents." *Biometrics*, 27: 866–875. 703, 711, 712

Yuan, Y. and Yin, G. (2008). "Sequential continual reassessment method for two-dimensional dose finding." *Statistics in Medicine*, 27: 5664–5678. 703

— (2011). "Robust EM Continual Reassessment Method in Oncology Dose Finding." *Journal of the American Statistical Association*, 106: 818–831. 704

Acknowledgments

We would like to thank the two referees and the Associate Editor for many insightful suggestions which immensely strengthened the work.

Table 2: The selection percentage (Sel) and average percentage (Pts) of patients treated at each dose combination for the simulation study. Values in boldface correspond to the target maximum tolerated dose combinations.

%	Latent-table				Complete-data				Proposed			
					Scenario 1							
Sel	13.4	1.7	0.1	0.0	9.8	0.7	0.0	0.0	5.8	1.2	0.1	0.0
	8.9	**24.5**	6.5	0.5	9.6	**41.5**	8.2	0.0	9.0	**38.7**	7.4	0.1
	0.1	1.9	**16.8**	14.6	0.0	1.4	**24.7**	1.5	0.1	2.1	**30.3**	1.0
	0.0	0.1	1.2	9.8	0.0	0.0	0.7	1.1	0.0	0.0	1.0	2.5
Pts	9.8	3.0	1.1	0.5	5.2	3.3	0.9	1.1	4.2	4.1	1.9	1.8
	11.4	**13.2**	4.9	3.0	13.6	**25.9**	8.9	0.6	11.2	**17.8**	10.4	1.7
	6.3	3.7	**9.3**	9.0	2.2	14.5	**12.7**	0.9	2.2	16.4	**14.1**	1.3
	5.6	4.7	5.3	9.1	6.4	0.6	2.3	1.0	7.7	0.7	3.3	1.4
					Scenario 2							
Sel	0.5	0.1	0.0	0.0	0.1	0.1	0.0	0.0	0.3	0.3	0.0	0.0
	7.9	1.0	0.2	0.0	2.9	2.3	0.0	0.0	4.7	0.6	0.1	0.0
	27.1	8.1	1.2	0.1	**23.6**	14.1	1.1	0.0	**25.1**	14.8	0.6	0.0
	4.7	**22.7**	7.3	0.6	6.6	**37.1**	4.1	0.0	7.0	**34.0**	4.7	0.1
Pts	2.0	0.4	0.0	0.0	0.9	0.7	0.3	0.3	1.1	1.0	0.4	0.4
	9.9	1.8	0.2	0.0	5.1	3.5	1.6	0.1	4.6	3.8	2.4	0.3
	25.1	7.9	1.6	0.4	**16.2**	15.7	1.0	0.0	**15.5**	16.8	1.5	0.1
	19.2	**21.2**	8.7	1.6	25.4	**26.1**	3.1	0.1	25.7	**23.3**	2.9	0.1
					Scenario 3							
Sel	**20.0**	9.4	0.8	0.1	**15.5**	5.1	0.5	0.1	**11.0**	4.6	0.6	0.2
	3.9	**19.9**	6.2	1.2	6.3	**37.5**	7.8	0.1	7.3	**32.3**	7.3	0.0
	0.1	3.8	**13.1**	9.9	0.0	2.5	**19.4**	1.0	0.5	5.1	**24.3**	1.2
	0.0	0.0	2.6	8.6	0.0	0.1	1.8	1.1	0.0	0.1	1.6	2.3
Pts	**14.6**	7.8	2.6	1.1	**8.9**	5.7	1.7	1.2	**4.9**	5.8	2.6	2.0
	9.0	**11.8**	4.6	2.3	10.8	**24.1**	8.1	0.6	9.4	**16.2**	9.4	1.2
	6.2	3.5	**7.4**	6.3	2.5	14.4	**9.6**	0.8	3.6	16.7	**11.8**	1.0
	5.8	4.8	5.1	7.1	7.1	1.1	2.5	0.8	9.1	1.4	3.7	1.2
					Scenario 4							
Sel	2.5	0.3	0.0	0.0	0.8	0.0	0.0	0.0	1.5	0.1	0.0	0.0
	27.3	7.3	0.2	0.0	**22.2**	7.0	0.1	0.0	**19.8**	6.5	0.2	0.0
	5.5	**22.9**	6.8	0.2	5.8	**45.9**	4.1	0.1	7.8	**45.4**	4.8	0.0
	0.0	5.6	**17.5**	2.6	0.1	1.8	**9.5**	0.2	0.1	1.5	**9.3**	0.0
Pts	6.2	1.2	0.2	0.0	2.7	0.9	0.4	0.3	3.2	1.8	0.8	0.7
	18.7	5.7	0.7	0.2	**13.4**	6.7	2.9	0.2	**11.0**	6.9	3.9	0.5
	12.5	**15.3**	4.8	1.2	13.3	**32.1**	2.9	0.1	12.7	**29.5**	3.9	0.1
	6.8	9.3	**12.6**	4.6	9.2	7.8	**6.9**	0.2	11.7	7.2	**6.2**	0.1

Table 2, continued.

%	Latent-table				Complete-data				Proposed			
					Scenario 5							
Sel	4.9	**19.1**	13.2	2.8	9.4	**24.8**	14.0	2.9	8.5	**23.2**	11.3	2.8
	1.0	5.9	**16.2**	11.3	1.3	8.8	**20.5**	3.6	1.7	8.2	**25.4**	5.6
	0.2	0.9	4.7	**15.4**	0.5	1.6	3.8	**6.3**	0.2	1.2	2.7	**7.0**
	0.0	0.0	0.4	3.1	0.0	0.1	0.4	0.8	0.0	0.0	0.1	1.0
Pts	7.6	**10.6**	7.1	3.5	5.8	**15.1**	8.1	6.5	5.6	**11.8**	7.7	8.3
	6.4	5.7	**9.7**	6.2	3.7	12.4	**14.7**	4.9	3.9	11.7	**14.4**	4.9
	6.7	1.8	4.2	**8.5**	2.4	9.4	3.3	**3.2**	2.9	9.6	4.0	**3.5**
	6.4	5.3	4.6	5.7	7.2	1.4	0.7	1.1	8.9	1.2	0.7	1.2
					Scenario 6							
Sel	**19.9**	8.8	0.1	0.0	**15.5**	5.4	0.4	0.0	**11.4**	3.9	0.4	0.0
	5.6	**21.5**	4.6	0.2	8.2	**34.7**	4.9	0.0	10.2	**29.0**	6.1	0.0
	0.3	5.3	**13.3**	3.7	0.2	8.1	**16.6**	0.5	0.5	9.7	**21.1**	0.3
	0.0	0.3	5.7	**10.3**	0.0	0.1	3.1	**1.1**	0.1	0.6	2.2	**2.4**
Pts	**14.4**	7.6	2.0	0.4	**8.7**	5.4	1.4	0.8	**5.2**	5.3	2.4	1.3
	9.9	**12.4**	3.9	1.1	11.9	**21.1**	6.7	0.4	10.3	**15.8**	7.7	0.9
	6.7	5.1	**8.2**	3.9	3.6	17.3	**9.3**	0.3	4.5	19.5	**10.4**	0.5
	5.9	5.1	6.4	**7.1**	7.3	2.0	3.3	**0.6**	9.0	2.4	4.0	**0.9**
					Scenario 7							
Sel	0.8	3.6	13.6	**60.5**	1.5	4.0	7.9	**72.7**	1.7	3.6	9.3	**73.3**
	0.1	1.0	3.6	11.9	0.6	0.7	1.0	7.4	0.1	0.7	1.1	5.7
	0.1	0.4	1.3	1.9	0.0	0.5	0.2	0.6	0.1	0.6	0.3	0.6
	0.0	0.1	0.2	0.2	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.1
Pts	6.9	6.5	10.4	**24.8**	3.1	6.1	8.6	**44.0**	2.7	6.0	8.2	**42.3**
	5.8	2.1	3.2	8.7	3.0	4.0	5.2	6.9	2.7	4.0	5.0	6.9
	6.4	0.9	1.4	4.1	2.8	6.3	0.4	0.6	2.6	7.4	0.4	0.7
	6.2	4.6	3.9	4.1	7.9	0.9	0.2	0.1	9.9	0.8	0.2	0.1
					Scenario 8							
Sel	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.2	0.0	0.0	0.0	0.6	0.0	0.0	0.0	0.7	0.0	0.0	0.0
Pts	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0
	3.5	0.0	0.0	0.0	0.5	0.1	0.3	0.0	0.0	0.3	0.4	0.0
	29.6	0.0	0.0	0.0	3.2	6.8	0.1	0.0	1.8	6.2	0.2	0.0
	60.0	6.1	0.9	0.0	83.3	5.3	0.1	0.0	86.0	4.9	0.1	0.0

Table 3: Summary of the simulation results.

	Scenario							
Design	1	2	3	4	5	6	7	8
	Selection percentage of MTD combinations							
Latent-table	41.3	49.8	53.0	67.7	50.7	65.0	60.5	0.0
Complete-data	66.2	60.7	72.4	77.6	51.6	67.9	72.7	0.0
Proposed	69.0	59.1	67.6	74.5	55.6	63.9	73.3	0.0
	Selection percentage of overly toxic doses							
Latent-table	36.8	27.0	27.6	19.9	27.3	17.4	0.0	0.2
Complete-data	20.2	24.7	14.6	12.3	20.5	11.2	0.0	0.6
Proposed	15.6	26.2	13.9	13.1	19.7	10.7	0.0	0.7
	% of patients treated at MTD combinations							
Latent-table	22.5	46.3	33.8	46.6	28.8	42.1	24.8	0.0
Complete-data	38.6	42.4	42.6	52.4	33.0	39.6	44.0	0.0
Proposed	31.8	38.9	32.9	46.6	29.7	32.3	42.3	0.0
	% of patients treated at overly toxic doses							
Latent-table	31.3	34.5	24.7	24.8	16.8	18.9	0.0	100
Complete-data	20.8	32.4	18.1	17.3	19.5	15.0	0.0	100
Proposed	25.3	35.5	22.0	21.8	20.8	18.1	0.0	100
	Trial duration (months)							
Latent-table	56.9	49.5	56.4	56.2	56.5	56.4	56.5	11.5
Complete-data	56.8	50.1	56.6	56.1	56.3	56.5	56.5	12.4
Proposed	26.8	24.1	26.3	25.4	26.6	26.0	26.4	6.5

Table 4: The selection percentage of each combination as a safe dose under the proposed methods based on 1,000 simulated trials. The boxes indicate the true set of safe doses.

Scenario	Complete-data				Proposed			
1	17.4	3.5	0.9	0.0	15.7	3.9	0.9	0.6
	96.5	66.8	10.4	0.9	95.4	62.6	8.3	1.1
	99.2	99.2	78.4	8.1	99.3	99.1	74.9	7.3
	99.2	99.2	99.0	52.8	99.3	99.3	99.2	44.7
2	0.7	0.2	0.0	0.0	1.4	0.6	0.5	0.2
	6.5	2.5	0.6	0.0	6.8	2.1	1.3	0.5
	44.5	16.7	4.3	0.4	48.1	18.3	3.4	1.1
	91.7	72.0	21.2	2.9	91.6	75.0	24.3	2.3
3	30.2	9.4	2.3	0.6	25.5	9.4	3.5	1.4
	95.4	71.7	19.2	2.9	92.2	64.8	16.6	3.8
	98.7	98.4	78.5	13.9	98.3	97.2	72.2	12.7
	98.8	98.8	98.2	55.3	98.4	98.3	97.3	48.9
4	1.5	0.6	0.4	0.20	2.5	1.1	0.5	0.2
	33.9	7.4	0.9	0.5	32.3	8.1	2.2	0.5
	95.9	72.6	13.9	0.8	95.4	70.6	13.0	1.8
	97.6	97.4	75.5	8.0	97.0	96.7	71.4	7.4
5	81.2	56.5	19.5	5.8	82.7	56.0	20.1	6.9
	96.8	93.0	71.5	17.9	98.1	94.9	72.9	19.7
	98.7	98.1	95.2	58.6	98.8	98.5	96.8	58.1
	98.8	98.8	98.0	91.7	98.9	98.9	98.5	91.2
6	27.5	9.9	3.3	0.8	24.0	7.7	3.1	1.6
	89.4	63.5	16.3	3.6	86.2	56.1	13.2	3.1
	98.7	96.8	71.0	13.9	97.6	95.8	66.4	10.1
	98.8	98.8	97.1	50.3	97.9	97.9	94.9	43.9
7	95.7	94.7	90.1	70.4	95.8	94.2	91.4	71.8
	97.1	96.8	95.5	88.9	96.9	96.6	95.4	90.4
	97.5	97.5	96.9	95.1	97.2	97.1	96.7	94.7
	97.5	97.5	97.4	96.9	97.2	97.2	97.1	96.5
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.0	0.0	0.06	0.0	0.0	0.0	0.0

Table 5: Summary of the results for the sensitivity analysis.

Condition	Scenario							
	1	2	3	4	5	6	7	8
	Selection percentage of MTD combinations							
Log-logistic	65.2	61.3	65.7	77.4	55.5	63.4	72.3	0.0
Heterogeneous	66.0	58.4	65.9	75.3	53.6	61.8	72.9	0.0
6 partitions	67.1	56.3	67.6	76.3	52.1	64.9	71.6	0.0
Vague priors	67.8	58.8	69.5	81.4	53.4	65.2	67.4	0.0
A/I ratio=5	65.2	56.1	63.2	73.4	50.1	59.7	66.0	0.0
	Selection percentage of overly toxic doses							
Log-logistic	17.3	25.0	17.8	10.2	18.3	8.7	0.0	1.1
Heterogeneous	17.6	26.1	15.2	11.5	20.5	10.8	0.0	0.7
6 partitions	15.8	25.6	14.8	12.1	18.8	10.3	0.0	0.3
Vague priors	14.3	25.1	14.2	9.4	20.2	9.9	0.0	0.3
A/I ratio=5	15.1	28.0	15.4	13.9	18.3	10.6	0.0	0.6
	% of patients treated at MTD combinations							
Log-logistic	32.1	37.3	34.6	45.7	29.4	31.0	42.9	0.0
Heterogeneous	33.5	38.9	35.2	47.3	31.4	34.0	40.9	0.0
6 partitions	33.3	38.0	34.0	46.7	29.2	31.8	39.1	0.0
Vague priors	32.7	39.5	34.8	48.5	30.0	33.1	39.3	0.0
A/I ratio=5	26.9	36.0	29.7	42.5	24.8	25.8	36.6	0.0
	% of patients treated at overly toxic doses							
Log-logistic	24.3	35.9	22.9	21.0	20.9	18.7	0.0	100
Heterogeneous	24.3	35.2	21.4	20.2	18.9	17.5	0.0	100
6 partitions	23.2	33.8	21.6	19.9	19.1	17.6	0.0	100
Vague priors	22.8	34.6	21.8	18.7	20.2	17.9	0.0	100
A/I ratio=5	26.7	35.0	24.5	22.0	21.5	20.5	0.0	100
	Trial duration (months)							
Log-logistic	26.9	23.9	26.2	25.3	26.6	25.9	26.3	6.7
Heterogeneous	26.9	24.0	26.2	25.3	26.4	26.0	26.4	6.8
6 partitions	26.7	23.8	26.3	25.5	26.5	26.1	26.4	6.5
Vague priors	26.8	24.1	26.2	25.6	26.3	26.1	26.0	6.4
A/I ratio=5	20.1	17.5	19.6	18.7	19.9	19.4	19.9	5.5