A Bayesian Adaptive Phase I/II Clinical Trial Design with Late-onset Competing Risk Outcomes

Yifei Zhang[1], Sha Cao[1,2], Chi Zhang[2,3], Ick Hoon Jin[4] and Yong Zang[1*,2]

1. Department of Biostatistics, Indiana University, USA

2. Center of Computational Biology and Bioinformatics, Indiana University, USA

3. Department of Medical Genetics, Indiana University, USA

4. Department of Applied Statistics, Yonsei University, Republic of Korea

*For correspondence: Yong Zang, Department of Biostatistics, Indiana University, 410 W 10th Street, Indianapolis, IN, 46075, US. Email: zangy@iu.edu

Abstract

Early-phase dose-finding clinical trials are often subject to the issue of late-onset outcomes. In phase I/II clinical trials, the issue becomes more intractable because toxicity and efficacy can be competing risk outcomes such that the occurrence of the first outcome will terminate the other one. In this paper, we propose a novel Bayesian adaptive phase I/II clinical trial design to address the issue of late-onset competing risk outcomes. We use the continuation-ratio (CR) model to characterize the trinomial response outcomes and the cause-specific hazard rate method to model the competing-risk survival outcomes. We treat the late-onset outcomes as missing data and develop a Bayesian data augmentation method to impute the missing data from the observations. We also propose an adaptive dose-finding algorithm to allocate patients and identify the optimal biological dose (OBD)

during the trial. Simulation studies show that the proposed design yields desirable operating characteristics.

KEY WORDS: Adaptive design; Competing risk; Late-onset; Phase I/II clinical trial; Targeted therapy

1   Introduction

The purpose of a traditional phase I dose-finding clinical trial is to identify the maximum tolerated dose (MTD) based on the toxicity outcome (Storer, 1989; O'Quigley et al., 1990). An underlying assumption for this type of trial is that both efficacy and toxicity increase monotonically with the dose level. However, recently developed novel molecularly targeted agents (MTAs) have challenged this paradigm, as the monotonic assumption may not always hold. MTAs are developed to modulate specific aberrant pathways in cancer cells, of which the dose-efficacy curve may not follow a monotonic pattern (Korn, 2004; Parulekar and Eisenhauer, 2004; Tourneau et al., 2009; LoRusso et al., 2010). Indeed, many studies have revealed that the efficacy of MTAs often increases initially with the dose and then plateaus (Ellis, 2003; Morgan et al., 2003). As a result, finding the MTD is no longer a desirable goal for a dose-finding trial for MTA, because a dose below the MTD may be sufficient to achieve maximum efficacy with substantially lower toxicity.

Therefore, a more reasonable goal of a dose-finding trial for MTA is to find the optimal biological dose (OBD), which yields the most desirable risk-benefit trade-off by jointly monitoring the efficacy and toxicity outcomes. This type of trial is typically referred to as the phase I/II clinical trial. A lot of phase I/II clinical trial designs have been proposed in the

2

literature (Thall and Russell, 1998; O'Quigley et al., 2001; Braun, 2002; Thall and Cook, 2004; Yin et al., 2006; Zang etl al., 2017; Muenz et al., 2019). All these designs treat toxicity and efficacy as categorical outcomes and assume that they can be observed immediately. That means the response outcomes for all the patients in the trial are observable when a new cohort of patients has entered the trial. Unfortunately, this assumption is often violated in practice, and late-onset outcomes are prevalent in practical trial implementation. In the presence of late-onset, part of the response outcomes may be missing at the interim analysis of the trial due to the lack of enough follow-up. Ignoring these missing data can result in a bias in estimating the response rate.

To address the issue of late-onset outcomes in dose-finding clinical trials, many statistical methods have been proposed (Cheung and Chappell, 2000; Braun, 2006; Yuan and Yin, 2011b; Liu et al., 2013; Jin et al., 2014). Most of the methods are applicable for the phase I clinical trial only with a focus on the toxicity outcome, and the remaining one for the phase I/II clinical trial models toxicity and efficacy outcomes separately and uses a copula model to build the joint distribution (Jin et al., 2014). However, in practice, the toxicity and efficacy endpoints are often competing risks, and thus cannot be modeled separately. Any drug used in the oncology trial is known to have a "double-edged sword" effect: on the one hand, it prevents the tumor progression; on the other hand, it induces severe toxicity on normal cells. Consequently, patients in the trial can experience different cause-specific adverse events. In a clinical trial, if a patient experiences either tumor progression or dose-limiting toxicity (DLT), he/she may receive a second-line treatment off the protocol for ethical consideration. Since only the adverse event that happens first is observable, competing risks issues arise.

The competing risk outcomes are quite common in clinical trials. A recent study has reviewed forty clinical trials with survival outcomes that were published in high-impact

general medical journals and found that thirty-one of them were potentially susceptible to competing risks (Austin and Fine, 2017). However, the issue of competing risk is not appropriately addressed for the majority of these studies. Instead, the standard methods such as the Kaplan-Meier survival curve and Cox regression were still used for statistical analysis, which can result in an overestimate of the incidence rate in the presence of competing risk (Wolbers et al., 2014). In phase I/II trial the situation gets worse because a promising drug may be rejected for toxicity, and patients may be assigned to a less efficacious dose due to the overestimate of incidence rate. Hence, an adaptive design to address the late-onset competing risk outcomes for phase I/II clinical trials is needed.

Our study is motivated by a phase I/II clinical trial being conducted at Indiana University Melvin and Bren Simon Cancer Center. The purpose of this trial is to find the OBD for a PI3K inhibitor combined with chemotherapy. Sixty patients with advanced solid tumors were enrolled and treated with PI3K inhibitor at five dose levels ranging from 200 mg to 800 mg, combined with a fixed dose of chemotherapy. Dose-limiting toxicity (DLT) and tumor progression are co-primary endpoints for this trial, and they are not immediately observable after the treatment. Instead, the DLT event and the tumor progression event may occur at any time point during a three-month follow-up. For every two months, a cohort of three patients will be enrolled in the trial for dose allocation before all the toxicity and efficacy outcomes become available for previous patients in the trial. In other words, some of the patients may still need to be followed when a new cohort of patients come in. Moreover, if any patient in the trial has experienced either DLT or tumor progression, he/she will be treated by an active second-line treatment off the protocol. In this motiving trial the DLT and tumor progression are competing risk outcomes, and how to design an adaptive phase I/II clinical trial that could handle late-onset competing risk outcomes is the challenge of

this trial.

In this paper, we develop a Bayesian adaptive phase I/II clinical trial design accounting for the late-onset competing risk endpoints. We use the continuation-ratio (CR) method (McCullagh and Nelder, 1989; Zhang et al., 2006) to model the trinomial response outcomes and the cause-specific hazard rate method to model the competing-risk survival outcomes. We treat the late-onset outcomes as missing data and develop a Bayesian data augmentation method to impute the missing data from the observation. We also propose an adaptive dose-finding algorithm to allocate patients and identify the OBD adaptively during the trial. The performance of the proposed design is evaluated through comprehensive simulation studies.

2   Methodology

2.1   Probability model

For majority of the dose-finding studies the outcomes are typically binary and the binomial distribution is used to fit the data. However, the response outcomes in the competing risk problem have three possible levels denoted as "no adverse event", "tumor progression occurring first" and "DLT occurring first". The binomial distribution is inappropriate for the three levels outcomes and we use the trinomial distribution to model the data. We first assume that each patient's response outcome is immediately observable and then extend the model to handle the late-onset issue. Let $y_i = (y_{i0}, y_{i1}, y_{i2})$ be the trinomial response outcome with $y_{i0} + y_{i1} + y_{i2} = 1$ for the ith patient in the trial $(i = 1, \cdots, n)$. We define $y_{i0} = 1$ if no adverse event occurs during the follow-up, $y_{i1} = 1$ if tumor progression occurs first and $y_{i2} = 1$ if DLT occurs first. Let $\pi_i = (\pi_{i0}, \pi_{i1}, \pi_{i2})$ be the associated cell probability and d_i be the dose level assigned to the ith patient $(d_i = 1, \cdots, L)$. We propose to use the

CR model (McCullagh and Nelder, 1989; Zhang et al., 2006) to characterize π_i as:

$$
\begin{aligned}
\pi_{i0} &= \frac{exp[\alpha_1 + \beta_1(d_i - 1)]}{\{1 + exp[\alpha_1 + \beta_1(d_i - 1)]\}\{1 + exp[\alpha_2 + \beta_2(d_i - 1)]\}} \\
\pi_{i1} &= \frac{1}{\{1 + exp[\alpha_1 + \beta_1(d_i - 1)]\}\{1 + exp[\alpha_2 + \beta_2(d_i - 1)]\}} \\
\pi_{i2} &= \frac{exp[\alpha_2 + \beta_2(d_i - 1)]}{1 + exp[\alpha_2 + \beta_2(d_i - 1)]},
\end{aligned}
\tag{1}
$$

where α_1, α_2, β_1 and β_2 are unknown parameters with $\beta_1 > 0$ and $\beta_2 > 0$. Based on model (1), π_{i1} monotonically decreases with dose level d_i and π_{i2} monotonically increases with d_i. It is reasonable to assume that the DLT rate will increase with the dose level. However, the tumor progression rate may decrease first and then, starting from a certain dose level, reaches a plateau due to the biological mechanism of the MTA (Ellis, 2003; Morgan et al., 2003). However, as π_{i1} is the probability of tumor progression occurring first under the competing risk framework, it is still fair to assume a monotonic decreasing pattern for π_{i1} because the monotonically increasing DLT rate will make DLT taking the precedence of tumor progression.

If y_i is quickly observable, we can use model (1) to guide the dose-finding. Unfortunately, model (1) is inappropriate in the presence of late-onset outcome. To see this, let T be the fixed total follow-up time for each patient and τ be the inter-arrival time, which means a new cohort of patients will enter the trial within each length τ time interval. Then the late-onset issue arises if $\tau < T$ because part of the existing patients have not been completely followed when a new cohort of patients come in, and adverse events that may not have so far occurred are still likely to occur in the remaining follow-up time.

Theoretically speaking, the late-onset outcomes are different with the missing data be-

cause these outcomes will eventually become observable when the patients have been com-
pletely followed. However, from the prospective of interim analysis the late-onset outcomes
can be naturally framed as missing data. Let t_i be the time that the first adverse event oc-
curs, v_i be the actual follow-up time at any interim stage, and $m(v_i)$ be the missing indicator
for whether a patient has been followed up to time v_i, we have:

$$m(v_i) = \begin{cases} 1 & \text{if } t_i > v_i \text{ and } v_i < T \\ 0 & \text{if } t_i \leq v_i \text{ or } v_i = T. \end{cases} \tag{2}$$

In other words, the response outcome is missing, and the late-onset issue arises if the
patient has not been fully followed and no adverse event has been observed by time v_i. In
Figure 1, we provide a hypothetical trial example to illustrate the missing mechanism. Based
on Figure 1, y_1 is missing when the second patient enters the trial because the first patient
has not been fully followed, and there is no adverse event occurring for that patient yet.
When the third patient enters the trial, y_2 should be treated as missing for a similar reason,
but not y_1, as the first patient has already experienced tumor progression. Finally, when the
last patient comes in, y_3 is not missing because a DLT has been observed for the patient. y_2
is not missing either because the second patient has been fully followed.

Ideally, to avoid missing data, the trial can be temporarily suspended to allow all enrolled
patients to be fully evaluated before a new dose assignment. However, in practice, such
suspension is often unlikely because it may result in an unfeasibly long trial, wastes resources,
and also causes tremendous administrative burden. Also, it is ethnically undesirable to
delay a new patient's treatment while waiting for previous patients' outcomes. Another
solution is discarding the missing data and using only those patients with complete response
outcomes. Unfortunately, in the following theorem, we prove that the missing due to late-

onset competing risk outcome is non-ignorable, which means the complete-data-only strategy can cause the estimate of π_i to be biased.

Theorem 1. The missing data caused by late-onset competing risk outcome are non-ignorable with the following inequality

$$\frac{P(y_{i1} + y_{i2} = 1 | m(v_i) = 0)}{P(y_{i1} + y_{i2} = 0 | m(v_i) = 0)} > \frac{P(y_{i1} + y_{i2} = 1 | m(v_i) = 1)}{P(y_{i1} + y_{i2} = 0 | m(v_i) = 1)}.$$

The proof of Theorem 1 is provided in the Appendix.

As it is unrealistic to suspend the trial and the missing data is non-ignorable, more sophisticated statistical model is needed to address the challenge. In what follows, we provide a Bayesian data augmentation approach (Tanner and Wong, 1987) to impute the missing data and derive the posterior distribution. We first of all use a piecewise exponential model to characterize the cause-specific hazard for the competing risk survival outcome for patients who would experience adverse event during the total follow-up time interval $[0, T]$. Specifically, we partition $[0, T]$ into K disjoint intervals $[0, h_1), [h_1, h_2), \cdots, [h_{K-1}, h_K \equiv T]$ and assume two constant cause-specific hazards λ_{1k} and λ_{2k} for tumor progression and DLT respectively. Let $x_i = \min(v_i, t_i)$ be the observation time and define $\delta_{ijk} = 1$ if the ith patient experiences event j first in the kth interval and $\delta_{ijk} = 0$ otherwise with $j = 1$ representing tumor progression and $j = 2$ representing DLT. Then, if all the data are observable, the overall likelihood function for $y = (y_1, \cdots, y_n)$ and $x = (x_1, \cdots, x_n)$ given all the parameters $\theta = (\alpha_1, \alpha_2, \beta_1, \beta_2, \lambda_{11}, \cdots, \lambda_{1K}, \lambda_{21}, \cdots, \lambda_{2K})$ can be written as

$$L(x, y|\theta) = \prod_{i=1}^{n} \left\{ \pi_{i0}^{y_{i0}} \left[\pi_{i1}^{y_{i1}} \pi_{i2}^{y_{i2}} \left(\prod_{k=1}^{K} \lambda_{1k}^{\delta_{i1k}} \lambda_{2k}^{\delta_{i2k}} \exp[-(y_{i1} + y_{i2})\lambda_{1k}s_{ik}]\exp[-(y_{i1} + y_{i2})\lambda_{2k}s_{ik}] \right) \right] \right\},$$

where $s_{ik} = h_k - h_{k-1}$ if $x_i > h_k$; $s_{ik} = x_i - h_{k-1}$ if $x_i \in [h_{k-1}, h_k)$ and $s_{ik} = 0$ otherwise. In order to handle missing data, without loss of generality, let us assume that data for the first n_1 patients are complete, and data for the remaining $n_2 = n - n_1$ patients are missing. Then, the complete likelihood function can be written as

$$L_c(x, y|\theta) = \left\{ \prod_{i=1}^{n_1} \pi_{i0}^{y_{i0}} \left[\pi_{i1}^{y_{i1}} \pi_{i2}^{y_{i2}} \prod_{k=1}^{K} \lambda_{1k}^{\delta_{i1k}} \lambda_{2k}^{\delta_{i2k}} \exp\{-(y_{i1} + y_{i2})\lambda_{1k}s_{ik}\}\exp\{-(y_{i1} + y_{i2})\lambda_{2k}s_{ik}\} \right] \right\}$$
$$\cdot \left\{ \prod_{i=n_1+1}^{n} \pi_{i0}^{I(y_{i0}=1)} \left[\exp\left\{ -\sum_{k=1}^{K}(\lambda_{1k} + \lambda_{2k})s_{ik} \right\} \pi_{i1} \right]^{I(y_{i1}=1)} \left[\exp\left\{ -\sum_{k=1}^{K}(\lambda_{1k} + \lambda_{2k})s_{ik} \right\} \pi_{i2} \right]^{I(y_{i2}=1)} \right\}.$$

Under the Bayesian framework, we assign α_1 and α_2 an independent uninformative uniform prior U$(-5, 5)$, and β_1 and β_2 an independent uninformative uniform prior U$(0, 5)$. Then, the prior distributions for π_i can be determined correspondingly through formula (1). In Table 1 we provide the prior mean and standard deviation for π_i with 5 dose levels. Based on the large standard deviation of π_i (about 0.2 to 0.4) in Table 1 we claim that the prior for π_i is uninformative. Following Liu et al. (Liu et al., 2013), we assign λ_{1k} and λ_{2k} an independent vague gamma prior Ga$(\tilde{\lambda}_k/2, 1/2)$ with $\tilde{\lambda}_k = K/\{T(K - k + 0.5)\}$. The underlying assumption for this gamma prior is that the adverse event occurs uniformly throughout the follow-up time $(0, T)$. The gamma prior is a conjugate prior for λ_{1k} and λ_{2k}, which will dramatically facilitate the MCMC procedure.

The data augmentation algorithm consists of an imputation step and a posterior sampling step. In the imputation step, we impute the missing data by their posterior samples drawing from the full conditional distributions. In the posterior sampling step, with the imputed missing data, we sequentially draw posterior samples of the unknown parameters θ from their full conditional distributions using the Metropolis-within-Gibbs-sampler algorithm. The data augmentation algorithm can be summarized as follows.

1. Impute any missing y_i from the multinomial distribution $\text{Mul}(1, p_{i0}, p_{i1}, p_{i2})$ with

$$
\begin{aligned}
p_{i0} &= \frac{\pi_{i0}}{\pi_{i0} + (\pi_{i1} + \pi_{i2})\exp\{-\sum_{k=1}^{K}(\lambda_{1k} + \lambda_{2k})s_{ik}\}} \\
p_{i1} &= \frac{\pi_{i1}\exp\{-\sum_{k=1}^{K}(\lambda_{1k} + \lambda_{2k})s_{ik}\}}{\pi_{i0} + (\pi_{i1} + \pi_{i2})\exp\{-\sum_{k=1}^{K}(\lambda_{1k} + \lambda_{2k})s_{ik}\}} \\
p_{i2} &= \frac{\pi_{i2}\exp\{-\sum_{k=1}^{K}(\lambda_{1k} + \lambda_{2k})s_{ik}\}}{\pi_{i0} + (\pi_{i1} + \pi_{i2})\exp\{-\sum_{k=1}^{K}(\lambda_{1k} + \lambda_{2k})s_{ik}\}}.
\end{aligned}
$$

2. Given all the imputed y_i, sequentially sample α_1, α_2, β_1 and β_2 from their full conditional distributions based on the CR model (1).

3. Given all the imputed y_i, sample λ_{jk} $(j = 1, 2; k = 1, \cdots, K)$ from the conjugate posterior distribution $\text{Ga}\left(\frac{\tilde{\lambda}_k}{2} + \sum_{i=1}^{n} \delta_{ijk}, \frac{1}{2} + \sum_{i=1}^{n}(y_{i1} + y_{i2})s_{ik}\right)$.

The data augmentation algorithm continuously repeats steps 1 to 3 until the Markov chain converges. The posterior samples of θ will be used to measure the efficacy/toxicity profile of the MTA at each dose level and further direct the dose-finding process, which will be illustrated in Section 2.2.

2.2   Dose-finding algorithm

We propose a phase I/II dose-finding algorithm to handle the late-onset competing risk outcomes based on the aforementioned probability model. The goal of this dose-finding algorithm is to: (1) evaluate the toxicity/efficacy profile of the drug at each dose level and; (2) select the optimal biological dose (OBD) and treat the patients with the identified OBD. Considering the trade-off between the competing risk outcomes, we propose to use a utility function to measure the desirability of the drug at each dose level.

To construct the utility function we need to elicit from physicians three utility values ω_0, ω_1 and ω_2 representing three events $y_{i0} = 1$, $y_{i1} = 1$ and $y_{i2} = 1$. Two requirements for the utility values are $\omega_0 > \omega_1$ and $\omega_0 > \omega_2$ because no adverse event happening is the most desirable case. Constructing the utilities requires close collaboration between clinicians and statisticians, and should be customized for each trial to reflect the clinical needs and practice best. In our experience, the process of elicitation of utility values is quite natural. It can be done by simply explaining what the utilities represent to the clinicians during the decision process and ask them to specify all values after fixing the ones for the worst and best outcome. Specifically, the utility values can be elicited in the following steps.

1. Specify a utility value $\omega_0 = 100$ for the most desirable event $y_{i0} = 1$.

2. Determine which adverse event is the most undesirable event. If tumor progression is the worst one, specify $\omega_1 = 0$. Otherwise, specify $\omega_2 = 0$.

3. Elicit the utility value for the remaining adverse event by using the two boundaries as reference. This utility values must be located between 0 and 100.

After eliciting the utility values, the true utility function for dose level $d = l$ can be

constructed as

$$U_{\text{true}}(l|\alpha_1, \alpha_2, \beta_1, \beta_2) = \sum_{j=0}^{2} \omega_j \pi_{ij}(d_i = l|\alpha_1, \alpha_2, \beta_1, \beta_2)$$

The true utility function relies on unknown parameters, which can estimated based on the observed data. During the trial, given the interim data $\mathcal{D}_n$, we can calculate the posterior mean utility as

$$U_n(l|\mathcal{D}_n) = \int U_{\text{true}}(l|\alpha_1, \alpha_2, \beta_1, \beta_2) f(\alpha_1, \alpha_2, \beta_1, \beta_2|\mathcal{D}_n) d\alpha_1 d\alpha_2 d\beta_1 d\beta_2,$$

where the posterior distribution $f(\alpha_1, \alpha_2, \beta_1, \beta_2|\mathcal{D}_n)$ can be approximated based on the afore-mentioned data augmentation algorithm.

In addition to the utility function, we also need an admissible set to conduct dose-finding. The purpose of the admissible set is to protect patients from being treated at overly-toxic or less-efficacious doses. With the interim data $\mathcal{D}_n$ the admissible set is constructed as

$$\mathcal{A} = \left\{ l : \mathrm{P}(\pi_{i1}(l|\alpha_1, \alpha_2, \beta_1, \beta_2) > \phi_1|\mathcal{D}_n) < c_1 \cap \mathrm{P}(\pi_{i2}(l|\alpha_1, \alpha_2, \beta_1, \beta_2) > \phi_2|\mathcal{D}_n) < c_2; l = 1, \cdots, L \right\},$$

where ϕ_1 and ϕ_2 are the highest acceptable adverse event rates and c_1 and c_2 are the cut-offs, which can be calibrated through simulation studies.

The OBD is defined as the dose yielding the highest posterior mean utility within the admissible set. To find the OBD, we propose the following dose-finding algorithm.

1. The first cohort of patients are treated at the lowest dose level or another physician-specified dose level.

2. At the current dose level, update the posterior distribution of θ.

3. Use the updated posterior distribution of θ to construct the admissible set and identify the OBD. If the admissible set is empty, early terminate the trial and conclude that no dose can be selected as the OBD.

4. Treat the next cohort of patients at the identified OBD.

5. Repeat steps 1-4 until the maximum sample size is reached. Dose skipping is not allowed when a dose-escalation occurs.

Once the trial ends, we select the dose with the highest posterior mean utility within the admissible set as the final OBD. As the proposed design uses the data augmentation method to calculate the posterior distributions of the unknown parameters, we refer it as the DA design.

3 Numerical Studies

We conducted comprehensive simulation studies to investigate the performance of the proposed design. We considered five dose levels and used the Weibull distribution to generate the cause-specific hazards at each dose level. The Weibull distributions were specified in a way that 70% of the adverse events will occur within the inter-arrival time. We specified a tumor progression upper-bound $\phi_1 = 0.6$ with a cut-off value $c_1 = 0.95$, a DLT upper-bound of $\phi_2 = 0.25$ with a cut-off value $c_2 = 0.95$. We assumed a maximum sample size of 60 in cohorts of size 3. We also specified a total follow-up time $T = 3$ months and an inter-arrival time $\tau = 2$ months. Also, we equally partitioned the follow-up $[0, T]$ into $K = 3$ intervals. We specified the utility values as $\omega_0 = 100$, $\omega_1 = 25$ and $\omega_2 = 0$. We compared the proposed design with two conventional designs. The first one makes the decision solely based on the

observed data only (O'Quigley et al., 1990), and we refer it as the ODO design. The second one treats patients one level down the current OBD if there are missing data, and we refer it as the OLD design. The same CR model and dose-finding algorithm are used for these two designs.

Table 2 shows the simulation results based on 5,000 replicates, including the dose selection probability, the average percentage of patients treated at each dose level, the average numbers of patients that experienced tumor progression first (N_{TP}) and DLT first (N_{DLT}), the average sample size (N) and the average duration of the trial. In the online supporting information, we also report the standard deviation of the total sample size N across 5,000 simulated trials. The probability pairs in parentheses are the probability of tumor progression occurring first and the probability of DLT occurring first at each dose level. The specific cause-specific hazard functions used to generate the data in Table 2 are depicted in Figure S1 of the online supporting information for this paper.

Scenarios 1 and 2 represent the circumstances of no OBD existing. In Scenario 1, the probability of tumor progression occurring first is at least 0.8, which is higher than the upper-bound of 0.6. The proposed DA design has the highest percentage of 94.6% to select no dose as OBD at the end of the trial, followed by 93.1% for the ODO design and 75% for the OLD design. In scenario 2, the probability of DLT occurring first is at least 0.4, whereas the upper-bound is 0.25. All the proposed designs yield similar percentages around 85% to report no OBD at the end of the trial. Besides, although the maximum sample size is 60, the average sample size for all the designs in scenarios 1 and 2 are substantially lower than 60, indicating the effectiveness of the proposed admissible set and early stopping rules.

In scenario 3 the OBD is at dose level 1 with the highest utility value of 60. The DA design yields a percentage of 59.1% to correctly identify dose level 1 as the OBD, which is about 7%

higher than that for the ODO design. On the other hand, due to the intrinsic mechanism of the OLD design, it yields the highest OBD selection percentage of 69.1%. Also, the OLD design assigns 45.8 patients to the OBD on average, which is 17.3 and 19.3 higher than that for the DA and ODO designs, respectively. In scenario 4 the OBD is at dose level 2. The DA design is the best design under this scenario, which yields the highest OBD selection percentage of 56.1%, followed by 51.5% for the ODO design and 46.2% for the OLD design. All the designs have comparable average number of patients treated at the OBD. In scenario 5 where dose 5 is the true OBD, the performance of the DA design is overwhelming. In terms of OBD selection percentage, the DA design is 10% higher than the ODO design and 20% higher than the OLD design. In terms of patient allocation, the DA design allocates about 5 and 23 more patients at the OBD than the ODO and OLD designs on average. The results for scenarios 6-8 are similar. In scenario 9, the probability of happening tumor progression first π_{i1} increases first. It then decreases, and dose 3 is the true OBD, which violates the parametric model assumption we have made in formula (1). The DA design still selects dose 3 with the highest percentage, which is slightly better than the ODO and OLD designs. Finally, the last scenario represents the situation of multiple OBDs because both dose 4 and dose 5 yield the highest utility value of 57.5. When there are multiple OBDs existing, the DA design still outperforms the other designs in both OBD selection and patient allocation. In summary, based on Table 2, the DA design outperforms the other conventional designs in most of the scenarios and should be recommended in practice.

We also conduct simulation studies to compare the proposed DA design with a Full design where $\tau = T = 3$ months. Noticing that when $\tau = T$, there is no missing data, and the data augmentation process is not required for this design. Scenarios 3 to 10 in Table 2 are used to generate the data. Figures S2 to S4 in the online supporting information summarizes the

simulation results, including the OBD selection percentage, number of patients treated at the OBD, and the trial durations. The results indicate that the Full design in general yields slightly better performances than the DA design in OBD selection and patient allocation at the price of the substantially prolonged trial duration.

We conducted additional simulation studies to investigate the robustness of the proposed DA design. We first change the data generating function for cause-specific hazards from Weibull distribution to uniform distribution. The parameter for the uniform distribution is specified such that it will yield the same π_i and similar missing rate as the Weibull distribution. Table 3 summarizes the results with DA_w indicating the one based on the Weibull distribution and DA_u indicating the one based on the uniform distribution. Based on Table 3, DA_w and DA_u report highly similar results meaning that proposed DA design is robust against data generating function. Also, in Table 4 we conducted a sensitivity analysis for the DA design with different piecewise exponential distributions to model the cause-specific hazards. In particular, we use DA_3 and DA_6 to denote the designs which partition the follow-up interval $[0, T]$ into 3 pieces and 6 pieces respectively. Based on Table 4, DA_3 and DA_6 have comparable performances, and DA_3 is slightly better for the last scenario where the OBD is located at dose 5. Hence, we claim that our choice of $K = 3$ is favorable. In Table 5 we investigate the sensitivity of the proposed design in terms of the prior distributions. We change the prior distributions of α_1 and α_2 from uniform distribution $U(-5, 5)$ to normal distribution $N(0, 5^2)$ and the prior distributions of β_1 and β_2 from uniform distribution $U(0, 5)$ to gamma distribution Gamma$(1, 0.2)$. The results show that the proposed DA design is not vulnerable to the prior specification. Lastly, in Table 6, we generate the data in such a way that 30% of the adverse event will occur within the inter-arrival time. The results are similar to those in Table 2, where 70% of the adverse event will occur within the

inter-arrival time. Hence, based on Table 6 we conclude that the performance of the DA design is robust against the missing rate of the adverse events.

4   Conclusion Remark

We have proposed a Bayesian adaptive design to tackle the issue of competing risk late-onset outcomes in phase I/II clinical trial. We use the cause-specific hazard to characterize the competing risk outcomes and develop a statistically rigorous Bayesian data augmentation algorithm to impute the missing data caused by late-onset outcomes, based on the accumulating response outcomes during each interim analysis of the trial. We also add early-stopping rules in the design, such that the phase I/II trial can be terminated earlier due to toxicity or futility, which enhances the ethics of the trial. We conduct comprehensive simulation studies to investigate the performance of the proposed design under different scenarios. The results show that the proposed design yields desirable operating characteristics and outperforms all the conventional designs used in practice. The R code to implement the proposed phase I/II clinical trial design is available on the online supporting information.

Besides competing risk outcomes, the toxicity and efficacy may be considered as semi-competing risk outcomes in some clinical trials (Murray et al., 2017). That is, although both toxicity and efficacy are still of primary interest, a subject will be treated off the protocol only if a specific adverse event has been observed for the subject. The proposed design cannot handle the semi-competing risk scenario, and a new design is required to address this problem. Also, in this paper, we assume population homogeneity for all the subjects in the trial. However, an increased understanding of population heterogeneity of cancer has already brought us to the era of personalized medicine, providing the clinicians unbeatable opportu-

nity to select individually tailored treatment taking into account each subject's variability. A lot of personalized clinical trial designs have been proposed (Piantadosi and Liu, 1996; Babb and Rogatko, 2001; O'Quigley and Paoletti, 2003; Yuan and Chappell, 2004; Ivanova and Wang, 2006; Thall et al., 2008), but none has investigated the issues of late-onset and competing risk. Therefore, it is of interest to extend the proposed design to integrate the personalized information into the trial.

Acknowledgments

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this paper.

References

Austin PC, Fine JP. (2017). Accounting for competing risks in randomized controlled trials: a review and recommendations for improvement. *Statistics in Medicine* **36**, 1203-1209.

Babb JS, Rogatko A. (2001). Patient specific dosing in a phase I cancer trial. *Statistics in Medicine* **20**, 2079-2090.

Braun TM. (2002). The bivariate continual reassessment method: Extending the CRM to phase I trials of two competing outcomes. *Controlled Clinical Trials* **23**, 240-256.

Braun TM. (2006). Generalizing the TiTE-CRM to adapt for early- and late-onset toxicities. *Statistics in Medicine* **25**, 2071-2083.

Cheung YK, Chappell R. (2000). Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics* **56**, 1177-1182.

Ellis LM. (2003). Antiangiogenic therapy: more promise and, yet again, more questions. *Journal of Clinical Oncology* **21**, 3897-3899.

Ivanova A, Wang K. (2006). Bivariate isotonic design for dose-finding with ordered groups. *Biometrics* **25**, 2018-2026.

Jin IH, Liu S, Thall PF, Yuan Y. (2014). Using data augmentation to facilitate conduct of phase I-II clinical trials with delayed outcomes. *Journal of Statistical Association* **109**, 525-536.

Korn EL. (2004). Nontoxicity endpoints in phase I trial designs for targeted, non-cytotoxic agents. *Journal of the National Cancer Institution* **96**, 977-978.

Liu S, Yin G, Yuan Y. (2013). Bayesian data augmentation dose finding with continual reassessment method and delayed toxicity. *Annals of Applied Statistics* **4**, 2138-2156.

LoRusso PM, Boerner SA, Seymour, L. (2010). An overview of the optimal planning, design, and conduct of phase I studies of new therapeutics. *Clinical Cancer Research* **16**, 1710-1718.

McCullagh P, Nelder JA. (1989). Generalized Linear Models (2nd edn). *Chapman & Hall: New York*, 160-164.

Morgan B, Thomas AL, Drevs J, Hennig J, Buchert M, Jivan A, et al. (2003). Dynamic contrast-enhanced magnetic resonance imaging as a biomarker for the pharmacological response of PTK787/ZK 222584, an inhibitor of the vascular endothelial growth factor receptor tyrosine kinases, in patients with advanced colorectal cancer and liver metastases: results from two phase I studies. *Journal of Clinical Oncology* **21**, 3955-3964.

Muenz DG, Taylor JMG, Braun TM. (2019). Phase I/II trial design for biological agents using conditional auto-regressive models for toxicity and efficacy. *Journal of the Royal Statistical Society-Series C*, in press.

Murray T, Thall P, Yuan Y, McAvoy S, Gomez D. (2017). Robust treatment comparison based on utilities of semi-competing risks in non-small-cell lung cancer. *Journal of the American Statistical Association*, **112**: 11-23.

O'Quigley J, Pepe M, Fisher L. (1990). Continual reassessment method: a practical design for phase I clinical trials in cancer. *Biometrics*, **46**: 33-48.

O'Quigley J, Hughes MD, Fenton T. (2001). Dose-finding designs for HIV studies. *Biometrics*, **57**: 1018-1029.

O'Quigley J, Paoletti X. (2003). Continual reassessment method for ordered groups. *Biometrics*, **59**: 430-440.

Parulekar WR, Eisenhauer EA. (2004). Phase I trial design for solid tumor studies of targeted, non-cytotoxic agents: Theory and practice. *Journal of the National Cancer Institution*, **96**: 990-997.

Piantadosi S, Liu G. (1996). Improved designs for dose-escalation studies using pharmacokinetic measurements. *Statistics in Medicine*, **15**: 1605-1618.

Storer BE. Design and analysis of phase I clinical trials. (1989). *Biometrics*, **45**: 925-937.

Tanner MA, Wong WH. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, **82**: 528-550.

Thall PF, Russell KE. (1998). A strategy for dose-finding and safety monitoring based on efficacy and adverse outcomes in phase I/II clinical trials. *Biometrics*, **54**: 251-264.

Thall P, Cook J. (2004). Dose-finding based on efficacy-toxicity trade-offs. *Biometrics*, **60**: 684-693.

Thall P, Nguyen H, Estey E. (2008). Patient-specific dose finding based on bivariate outcomes and covariates. *Biometrics*, **64**: 1126-1136.

Tourneau CL, Lee, JJ, Siu LL. (2009). Dose escalation methods in phase I cancer clinical trials. *Journal of the National Cancer Institution*, **101**: 708-720.

Wolbers M, Koller MT, Stel VS, Schaer B, Jager KJ, Leffondre K, Heinze G. (2014). Competing risks analyses: objectives and approaches. *European Heart Journal*, **35**: 2936-2941.

Yin G, Li Y, Ji Y. (2006). Bayesian dose-finding in phase I/II clinical trials using toxicity and efficacy odds ratios. *Biometrics*, **62**: 777-787.

Yuan Z, Chappell R. (2004). Isotonic designs for phase I cancer clinical trials with multiple risk groups. *Clinical Trials*, **1**: 499-508.

Yuan Y, Yin G. (2011). Robust EM continual reassessment method in oncology dose finding. *Journal of American Statistical Association*, **106**: 818-831.

Zang Y, Lee JJ. (2017). A robust two-stage design identifying the optimal biological dose for phase I/II clinical trials. *Statistics in Medicine*, **36**: 27-42.

Zhang W, Sargent DJ, Mandrekar S. (2006). An adaptive dose-finding design incorporating both toxicity and efficacy. *Statistics in Medicine*, **25**: 2365-2383.

Supporting Information

R codes to implement the proposed DA design, web tables and figures referenced in Section 3 are available with this paper at the Biometrics website on Wiley Online Library.

Appendix: Proof of Theorem 1

For the ith patient who does not experience any adverse events after the entire follow-up, the probability that his/her response outcome will be missing at time v_i can be expressed as

$$
\begin{aligned}
P(m(v_i) = 1|y_{i1} + y_{i2} = 0) &= P(t_i > v_i, v_i < T|y_{i1} + y_{i2} = 0) \\
&= P(v_i < T|y_{i1} + y_{i2} = 0)P(t_i > v_i|v_i < T, y_{i1} + y_{i2} = 0) \\
&= P(v_i < T|t_i > T)P(t_i > v_i|v_i < T, t_i > T) \\
&= P(v_i < T).
\end{aligned}
$$

Similarly, for the ith patient who will experiences an adverse event, the probability that his/her response outcome will be missing at time v_i can be expressed as

$$
P(m(v_i) = 1|y_{i1} + y_{i2} = 1) = P(v_i < T)P(t_i > v_i|v_i < T, t_i < T) < P(m(v_i) = 1|y_{i1} + y_{i2} = 0).
$$

Alone the same line, we have $P(m(v_i) = 0|y_{i1} + y_{i2} = 0) < P(m(v_i) = 0|y_{i1} + y_{i2} = 1)$.

Under Bayes rule, it is easy to see

$$\frac{P(y_{i1} + y_{i2} = 1|m(v_i) = 0)}{P(y_{i1} + y_{i2} = 0|m(v_i) = 0)} = \frac{P(m(v_i) = 0|y_{i1} + y_{i2} = 1)P(y_{i1} + y_{i2} = 1)}{P(m(v_i) = 0|y_{i1} + y_{i2} = 0)P(y_{i1} + y_{i2} = 0)}$$

$$\frac{P(y_{i1} + y_{i2} = 1|m(v_i) = 1)}{P(y_{i1} + y_{i2} = 0|m(v_i) = 1)} = \frac{P(m(v_i) = 1|y_{i1} + y_{i2} = 1)P(y_{i1} + y_{i2} = 1)}{P(m(v_i) = 1|y_{i1} + y_{i2} = 0)P(y_{i1} + y_{i2} = 0)}.$$

Therefore, we have

$$\frac{P(y_{i1} + y_{i2} = 1|m_i = 0)}{P(y_{i1} + y_{i2} = 0|m_i = 0)} > \frac{P(y_{i1} + y_{i2} = 1)}{P(y_{i1} + y_{i2} = 0)} > \frac{P(y_{i1} + y_{i2} = 1|m_i = 1)}{P(y_{i1} + y_{i2} = 0|m_i = 1)},$$

and based on this inequality, the missing data caused by late-onset competing risk outcome are non-ignorable.

Figure 1: Illustration of the missing mechanism. The total follow-up time is $T = 3$ and the inter-arrival time is $\tau = 2$. The black triangle indicates a tumor progression event and the black circle indicates a DLT event. y_1 is missing when the second patient enters the trial because the first patient has not been fully followed, and there is no adverse event occurring for that patient yet. When the third patient enters the trial, y_2 should be treated as missing for a similar reason, but not y_1, as the first patient has already experienced tumor progression. When the last patient comes in, y_3 is not missing because a DLT has been observed for the patient. y_2 is not missing either because the second patient has been fully followed. This figure appears in color in the electronic version of this article, and any mention of color refers to that version.
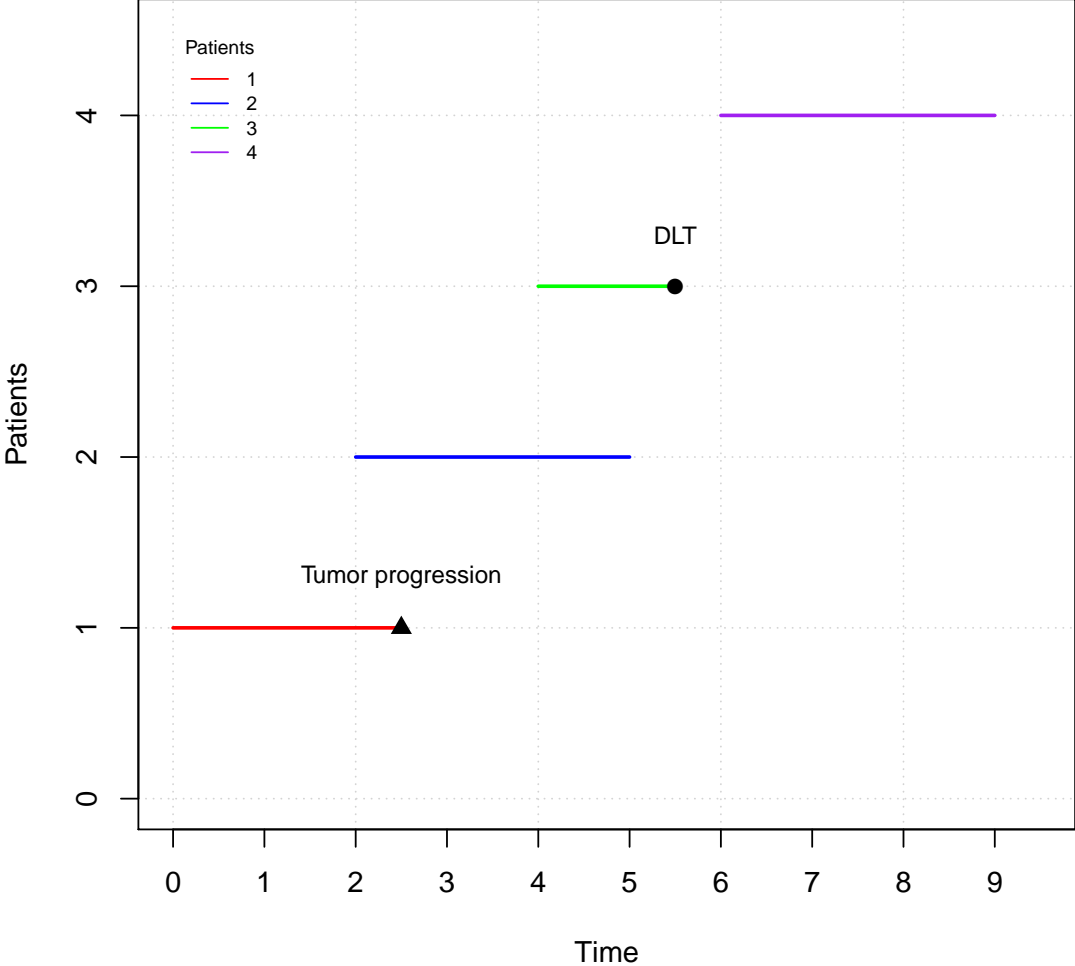
Table 1: The prior mean and standard deviation (s.d.) of $\pi_i = (\pi_{i0}, \pi_{i1}, \pi_{i2})$ with 5 dose levels. π_{i0} is the probability of no adverse event occurring; π_{i1} is the probability of tumor progression occurring first; π_{i2} is the probability of DLT occurring first.

Dose level	π_{i0}		π_{i1}		π_{i2}	
	mean	s.d.	mean	s.d.	mean	s.d.
1	0.253	0.316	0.253	0.315	0.494	0.389
2	0.199	0.282	0.071	0.168	0.730	0.332
3	0.123	0.244	0.020	0.092	0.857	0.269
4	0.086	0.216	0.009	0.064	0.904	0.229
5	0.066	0.193	0.005	0.049	0.929	0.202

Table 2: Operating characteristics of the DA, ODO and OLD designs based on 5,000 replicates. DA indicates the proposed data augmentation design; ODO indicates the observable data only design; OLD indicates the one level down design. The probability pairs in parentheses are the probability of occurring tumor progression first and the probability of occurring DLT first for each dose level. The percentage of trials with no dose selected is denoted by "None". N_{TP} is the number of patients experiencing tumor progression first, N_{DLT} is the number of patients experiencing DLT first and N is the total number of patients. The numbers in bold font indicate the OBD selection rates and patients allocation under the true OBD.

Design		Dose Level					None	N_{TP}/N_{DLT}	N	Duration (month)
		1	2	3	4	5				
	Scenario 1	(0.95,0.02)	(0.93,0.04)	(0.9,0.06)	(0.85,0.08)	(0.8,0.12)				
	Utility	26.75	26.25	26.5	28.25	28				
DA	% selected	0	0	0	0	5.4	**94.6**	27.9/2.9	32.8	22.1
	#patients	3.9	3.1	3.4	5.0	17.5				
ODO	% selected	0	0	0	0.4	6.5	**93.1**	27.2/2.9	32.2	21.7
	#patients	3.1	3.2	3.5	5.7	16.7				
OLD	% selected	0	0.2	0	2.3	22.5	**75.0**	33.8/3.1	39.3	26.5
	#patients	3.5	3.8	6.0	18.1	8.0				
	Scenario 2	(0.4,0.4)	(0.3,0.5)	(0.25,0.6)	(0.2,0.7)	(0.1,0.8)				
	Utility	30	27.5	21.25	15	12.5				
DA	% selected	15	0	0	0	0	**85**	10.3/11.5	27.0	18.5
	#patients	22.7	3.0	0.9	0.3	0.2				
ODO	% selected	15.1	0.6	0	0	0	**84.3**	10.5/11.4	27.2	18.6
	#patients	22.2	3.9	0.8	0.2	0.1				
OLD	% selected	14.3	0.7	0	0	0	**85**	10.5/11.2	26.8	18.4
	#patients	23.4	2.5	0.7	0.2	0.1				
	Scenario 3	(0.4,0.1)	(0.38,0.22)	(0.3,0.35)	(0.25,0.5)	(0.2,0.6)				
	Utility	60	49.5	42.5	31.25	25				
DA	% selected	**59.1**	32.9	5.5	0.2	0.1	2.2	21.6/11.5	58.8	40.0
	#patients	**28.5**	20.7	6.0	2.2	1.3				
ODO	% selected	**52.4**	37.9	6.0	0.7	0	3.0	21.9/11.5	58.4	39.8
	#patients	**26.2**	22.4	6.7	2.2	1.0				
OLD	% selected	**69.1**	21.7	5.4	0.9	0	2.9	22.7/8.4	58.4	39.8
	#patients	**45.8**	7.1	3.1	1.6	0.7				
	Scenario 4	(0.75,0.05)	(0.55,0.08)	(0.52,0.23)	(0.45,0.35)	(0.4,0.45)				
	Utility	38.75	50.75	38	31.25	25				
DA	% selected	5.6	**56.1**	28.1	7.8	1.7	0.7	32.4/11.0	59.8	40.6
	#patients	7.0	**23.3**	17.0	7.9	4.6				
ODO	% selected	6.4	**51.5**	29.9	7.5	2.4	2.3	31.9/11.0	59.3	40.2
	#patients	5.5	**23.0**	18.9	8.0	3.8				
OLD	% selected	5.8	**46.2**	39.5	5.1	2.4	1.0	36.0/7.8	59.5	40.4
	#patients	20.9	**21.9**	9.1	5.1	2.5				
	Scenario 5	(0.65,0.1)	(0.58,0.12)	(0.5,0.15)	(0.35,0.18)	(0.2,0.2)				
	Utility	41.25	44.5	47.5	55.75	65				
DA	% selected	2.7	6.9	8.5	15.7	**64.2**	2.0	21.7/9.8	59.1	40.3
	#patients	6.9	6.5	7.8	11.9	**26.0**				
ODO	% selected	3.8	9.7	11.3	18.8	**53.6**	2.8	22.4/9.7	58.5	39.8
	#patients	6.3	8.5	9.7	12.8	**21.3**				
OLD	% selected	5.4	9.1	15.0	22.2	**44.4**	3.9	27.9/8.5	58.0	39.5
	#patients	11.6	10.3	13.3	19.6	**3.2**				
	Scenario 6	(0.65,0.05)	(0.55,0.08)	(0.4,0.1)	(0.35,0.3)	(0.3,0.4)				
	Utility	46.25	50.75	60	43.75	37.5				
DA	% selected	2.2	18.2	**55.9**	20.6	3.0	0.1	25.9/9.8	59.9	40.8
	#patients	6.1	11.2	**23.7**	13.2	5.7				
ODO	% selected	3.0	21.3	**51.8**	18.3	4.8	0.8	25.8/9.5	59.5	40.5
	#patients	5.2	12.8	**24.1**	12.1	5.4				
OLD	% selected	5.0	22.8	**41.1**	25.8	4.6	0.7	30.3/6.9	59.6	40.6
	#patients	14.7	20.3	**15.3**	6.6	2.7				

26

Design		\multicolumn{5}{c}{Dose Level}					Duration (month)
		1	2	3	4	5	None	N_{TP}/N_{DLT}	N	
	Scenario 7	(0.65,0.1)	(0.58,0.15)	(0.5,0.18)	(0.2,0.2)	(0.15,0.4)				
	Utility	41.25	41.5	44.5	65	48.75				
DA	% selected	4.7	10.7	17.5	**49.1**	16.3	1.7	21.9/12.4	59.3	40.4
	#patients	8.4	8.1	12.6	**20.4**	9.8				
ODO	% selected	6.6	11.7	16.2	**44.1**	17.7	3.7	22.2/11.8	58.2	39.6
	#patients	7.7	10.4	12.3	**19.3**	8.5				
OLD	% selected	8.8	14.3	18.9	**29.1**	25.5	3.4	28.2/9.7	58.3	39.6
	#patients	15.2	13.2	15.7	**11.8**	2.4				
	Scenario 8	(0.7,0.08)	(0.65,0.1)	(0.5,0.13)	(0.4,0.18)	(0.25,0.2)				
	Utility	39.5	41.25	49.5	52	61.25				
DA	% selected	1.0	5.2	10.8	13.6	**68.0**	1.4	23.5/9.7	59.3	40.4
	#patients	5.7	5.8	8.5	11.6	**27.8**				
ODO	% selected	0.7	4.2	13.4	17.1	**63.5**	1.1	23.9/9.7	59.4	40.5
	#patients	4.6	6.1	10.0	13.0	**25.7**				
OLD	% selected	1.8	7.5	14.4	27.5	**46.2**	2.6	29.7/8.2	58.7	40.0
	#patients	8.9	10.0	14.7	21.5	**3.7**				
	Scenario 9	(0.65,0.05)	(0.5,0.1)	(0.3,0.15)	(0.4,0.2)	(0.6,0.22)				
	Utility	46.25	52.5	62.5	50	33				
DA	% selected	3.9	26.5	**41.1**	15.5	12.2	0.8	27.2/9.1	59.7	40.6
	#patients	6.4	12.5	**18.2**	11.6	11.0				
ODO	% selected	4.4	28.9	**39.7**	13.7	11.1	2.2	26.1/8.9	59.0	40.2
	#patients	5.5	14.0	**18.7**	11.5	9.4				
OLD	% selected	4.6	23.5	**38.7**	21.9	10.5	0.8	28.5/7.2	59.5	40.5
	#patients	15.4	18.3	**13.4**	9.6	2.8				
	Scenario 10	(0.7,0.1)	(0.6,0.15)	(0.5,0.18)	(0.3,0.2)	(0.3,0.2)				
	Utility	37.5	40	44.5	57.5	57.5				
DA	% selected	3.1	8.5	9.5	**15.0**	**61.1**	2.8	24.4/10.6	58.8	40.1
	#patients	7.1	7.7	7.8	**11.8**	**24.6**				
ODO	% selected	2.5	10.9	10.5	**18.3**	**54.9**	2.9	24.4/10.6	58.7	39.9
	#patients	6.0	9.3	9.1	**12.0**	**22.2**				
OLD	% selected	5.8	11.6	13.2	**21.2**	**43.3**	4.9	28.1/9.6	57.9	39.4
	#patients	13.7	9.9	12.7	**18.6**	**3.0**				

Table 3: Sensitivity analysis for the DA (data augmentation) design with different data generating functions. Weibull distribution is used to generate the cause-specific hazards for DA_w and uniform distribution is used for DA_u.

Design		Dose Level					None	N_{TP}/N_{DLT}	N	Duration (month)
		1	2	3	4	5				
		(0.95,0.02)	(0.93,0.04)	(0.9,0.06)	(0.85,0.08)	(0.8,0.12)				
	Utility	26.75	26.25	26.5	28.25	28				
DA_w	% selected	0	0	0	0	5.4	**94.6**	27.9/2.9	32.8	22.1
	#patients	3.9	3.1	3.4	5.0	17.5				
DA_u	% selected	0	0	0	0.3	7.3	**92.4**	28.7/3.0	33.8	22.7
	#patients	4.1	3.2	3.4	5.2	17.9				
		(0.4,0.1)	(0.38,0.22)	(0.3,0.35)	(0.25,0.5)	(0.2,0.6)				
	Utility	60	49.5	42.5	31.25	25				
DA_w	% selected	**59.1**	32.9	5.5	0.2	0.1	2.2	21.6/11.5	58.8	40.0
	#patients	**28.5**	20.7	6.0	2.2	1.3				
DA_u	% selected	**59.4**	32.0	6.7	0.3	0.1	1.5	22.1/11.6	59.2	40.3
	#patients	**28.8**	20.1	6.6	2.4	1.3				
		(0.65,0.1)	(0.58,0.15)	(0.5,0.18)	(0.2,0.2)	(0.15,0.4)				
	Utility	41.25	41.5	44.5	65	48.75				
DA_w	% selected	4.7	10.7	17.5	**49.1**	16.3	1.7	21.9/12.4	59.3	40.4
	#patients	8.4	8.1	12.6	**20.4**	9.8				
DA_u	% selected	5.5	9.3	16.5	**50.7**	16.4	1.6	21.8/12.4	59.2	40.3
	#patients	8.4	8.1	11.7	**21.1**	9.9				
		(0.65,0.1)	(0.58,0.12)	(0.5,0.15)	(0.35,0.18)	(0.2,0.2)				
	Utility	41.25	44.5	47.5	55.75	65				
DA_w	% selected	2.7	6.9	8.5	15.7	**64.2**	2.0	21.7/9.8	59.1	40.3
	#patients	6.9	6.5	7.8	11.9	**26.0**		21.7/9.8	59.1	40.3
DA_u	% selected	2.2	8.2	11.2	14.5	**62.3**	1.6	22.0/9.9	59.1	40.3
	#patients	6.7	7.6	8.0	11.8	**25.0**		22.0/9.9	59.1	40.3

Table 4: Sensitivity analysis for the DA (data augmentation) design with different piecewise exponential distributions to model the cause-specific hazards. The follow-up interval $[0, T]$ is partitioned into 3 pieces for the DA_3 design and 6 pieces for the DA_6 design.

Design		\multicolumn{5}{c}{Dose Level}			None	N_{TP}/N_{DLT}	N	Duration (month)
		1	2	3	4	5	None	N_{TP}/N_{DLT}	N	(month)
		(0.95,0.02)	(0.93,0.04)	(0.9,0.06)	(0.85,0.08)	(0.8,0.12)				
	Utility	26.75	26.25	26.5	28.25	28				
DA_3	% selected	0	0	0	0	5.4	**94.6**	27.9/2.9	32.8	22.1
	#patients	3.9	3.1	3.4	5.0	17.5				
DA_6	% selected	0	0	0	0	6.6	**93.4**	28.3/3.0	33.5	22.6
	#patients	4.0	3.1	3.4	5.0	18.1				
		(0.4,0.1)	(0.38,0.22)	(0.3,0.35)	(0.25,0.5)	(0.2,0.6)				
	Utility	60	49.5	42.5	31.25	25				
DA_3	% selected	**59.1**	32.9	5.5	0.2	0.1	2.2	21.6/11.5	58.8	40.0
	#patients	**28.5**	20.7	6.0	2.2	1.3				
DA_6	% selected	**57.7**	33.7	6.1	0.3	0.1	2.1	22.0/11.6	58.8	40.1
	#patients	**28.6**	20.2	6.4	2.4	1.3				
		(0.65,0.1)	(0.58,0.15)	(0.5,0.18)	(0.2,0.2)	(0.15,0.4)				
	Utility	41.25	41.5	44.5	65	48.75				
DA_3	% selected	4.7	10.7	17.5	**49.1**	16.3	1.7	21.9/12.4	59.3	40.4
	#patients	8.4	8.1	12.6	**20.4**	9.8				
DA_6	% selected	5.8	10.2	14.4	**50.1**	18.6	0.9	21.7/12.4	59.6	40.6
	#patients	8.5	8.2	11.4	**21.2**	10.3				
		(0.65,0.1)	(0.58,0.12)	(0.5,0.15)	(0.35,0.18)	(0.2,0.2)				
	Utility	41.25	44.5	47.5	55.75	65				
DA_3	% selected	2.7	6.9	8.5	15.7	**64.2**	2.0	21.7/9.8	59.1	40.3
	#patients	6.9	6.5	7.8	11.9	**26.0**				
DA_6	% selected	2.7	7.2	11.8	17.0	**60.0**	1.3	22.3/10.0	59.3	40.4
	#patients	6.7	7.1	9.0	12.5	**24.0**				

Table 5: Sensitivity analysis for the DA (data augmentation) design with different prior distributions. Uniform distributions are used as the prior distributions for DA_1, and normal and gamma distributions are used as the prior distributions for DA_2.

Design		1	2	3	4	5	None	N_{TP}/N_{DLT}	N	Duration (month)
		(0.95,0.02)	(0.93,0.04)	(0.9,0.06)	(0.85,0.08)	(0.8,0.12)				
	Utility	26.75	26.25	26.5	28.25	28				
DA_1	% selected	0	0	0	0	5.4	**94.6**	27.9/2.9	32.8	22.1
	#patients	3.9	3.1	3.4	5.0	17.5				
DA_2	% selected	0	0	0	0.1	5.1	**94.8**	26.4/2.7	31.0	20.9
	#patients	3.9	3.2	3.3	5.0	15.6				
		(0.4,0.1)	(0.38,0.22)	(0.3,0.35)	(0.25,0.5)	(0.2,0.6)				
	Utility	60	49.5	42.5	31.25	25				
DA_1	% selected	**59.1**	32.9	5.5	0.2	0.1	2.2	21.6/11.5	58.8	40.0
	#patients	**28.5**	20.7	6.0	2.2	1.3				
DA_2	% selected	**61.8**	28.8	7.6	0.7	0.2	0.9	21.9/11.9	59.5	40.5
	#patients	**29.2**	18.9	6.9	2.8	1.6				
		(0.65,0.1)	(0.58,0.15)	(0.5,0.18)	(0.2,0.2)	(0.15,0.4)				
	Utility	41.25	41.5	44.5	65	48.75				
DA_1	% selected	4.7	10.7	17.5	**49.1**	16.3	1.7	21.9/12.4	59.3	40.4
	#patients	8.4	8.1	12.6	**20.4**	9.8				
DA_2	% selected	4.5	10.1	17.9	**48.7**	17.7	1.1	21.5/12.6	59.4	40.5
	#patients	7.7	8.0	12.2	**21.1**	10.4				
		(0.65,0.1)	(0.58,0.12)	(0.5,0.15)	(0.35,0.18)	(0.2,0.2)				
	Utility	41.25	44.5	47.5	55.75	65				
DA_1	% selected	2.7	6.9	8.5	15.7	**64.2**	2.0	21.7/9.8	59.1	40.3
	#patients	6.9	6.5	7.8	11.9	**26.0**				
DA_2	% selected	3.0	6.8	8.7	18.0	**62.7**	0.8	22.0/10.1	59.6	40.6
	#patients	6.8	6.4	8.6	13.0	**24.8**				

Table 6: Sensitivity analysis with different cause-specific hazards rate within the inter-arrival time. 30% adverse events occurs within the inter-arrival time. DA indicates the proposed data augmentation design; ODO indicates the observable data only design; OLD indicates the one level down design.

Design		1	2	3	4	5	None	N_{TP}/N_{DLT}	N	Duration (month)
		(0.95,0.02)	(0.93,0.04)	(0.9,0.06)	(0.85,0.08)	(0.8,0.12)				
	Utility	26.75	26.25	26.5	28.25	28				
DA	% selected	0	0	0	0.1	8.6	**91.3**	30.9/3.2	36.4	24.9
	#patients	5.5	3.2	3.5	5.4	18.9				
ODO	% selected	0	0	0	0.2	8.9	**90.9**	28.8/3.1	34.2	23.5
	#patients	3.2	3.3	3.7	5.8	18.2				
OLD	% selected	0	2.6	1.2	4.3	31.4	**60.5**	38.6/3.1	44.3	30.2
	#patients	4.7	4.1	7.2	24.9	3.3				
		(0.4,0.4)	(0.3,0.5)	(0.25,0.6)	(0.2,0.7)	(0.1,0.8)				
	Utility	30	27.5	21.25	15	12.5				
DA	% selected	16.8	0.8	0	0	0	**82.4**	11.2/12.6	29.8	20.6
	#patients	24.6	3.5	1.2	0.4	0.1				
ODO	% selected	17.4	0.3	0	0	0	**82.3**	10.9/12.8	29.6	20.5
	#patients	22.8	4.9	1.3	0.5	0.1				
OLD	% selected	15.4	0.7	0	0	0	**83.9**	11.0/12.1	28.9	20.0
	#patients	24.7	2.8	0.9	0.4	0.1				
		(0.4,0.1)	(0.38,0.22)	(0.3,0.35)	(0.25,0.5)	(0.2,0.6)				
	Utility	60	49.5	42.5	31.25	25				
DA	% selected	**60.1**	31.7	6.8	0.7	0	0.7	22.2/11.7	59.7	40.7
	#patients	**29.7**	19.2	6.6	2.6	1.6				
ODO	% selected	**56.3**	31.8	7.4	0.4	0.1	4.0	21.2/11.7	57.8	39.5
	#patients	**26.5**	20.2	7.3	2.5	1.2				
OLD	% selected	**68.2**	23.3	5.5	0.7	0	2.3	22.4/8.8	58.7	40.1
	#patients	**45.8**	6.9	3.0	1.9	1.1				
		(0.75,0.05)	(0.55,0.08)	(0.52,0.23)	(0.45,0.35)	(0.4,0.45)				
	Utility	38.75	50.75	38	31.25	25				
DA	% selected	6.0	**55.1**	29.2	7.7	1.5	0.5	32.4/11.2	59.8	40.7
	#patients	7.8	**21.4**	18.2	8.0	4.3				
ODO	% selected	7.0	**53.4**	30.0	6.0	2.2	1.4	31.4/11.1	59.2	40.4
	#patients	5.6	**22.7**	18.8	7.9	4.2				
OLD	% selected	6.6	**46.6**	38.0	5.0	3.0	0.8	36.4/7.3	59.6	40.6
	#patients	24.3	**20.6**	7.8	4.8	2.1				