Using Data Augmentation to Facilitate Conduct of Phase I-II Clinical Trials With Delayed Outcomes

Ick Hoon Jin, Suyu Liu, Peter F. Thall & Ying Yuan

Accepted author version posted online: 18 Feb 2014. Published online: 13 Jun 2014.

To cite this article: Ick Hoon Jin, Suyu Liu, Peter F. Thall & Ying Yuan (2014) Using Data Augmentation to Facilitate Conduct of Phase I-II Clinical Trials With Delayed Outcomes, Journal of the American Statistical Association, 109:506, 525-536, DOI: 10.1080/01621459.2014.881740

To link to this article: http://dx.doi.org/10.1080/01621459.2014.881740

Taylor & Francis makes every effort to ensure the accuracy of all the information (the “Content”) contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA

Using Data Augmentation to Facilitate Conduct of Phase I–II Clinical Trials With Delayed Outcomes

Ick Hoon Jin, Suyu Liu, Peter F. Thall, and Ying Yuan

A practical impediment in adaptive clinical trials is that outcomes must be observed soon enough to apply decision rules to choose treatments for new patients. For example, if outcomes take up to six weeks to evaluate and the accrual rate is one patient per week, on average three new patients will be accrued while waiting to evaluate the outcomes of the previous three patients. The question is how to treat the new patients. This logistical problem persists throughout the trial. Various ad hoc practical solutions are used, none entirely satisfactory. We focus on this problem in phase I–II clinical trials that use binary toxicity and efficacy, defined in terms of event times, to choose doses adaptively for successive cohorts. We propose a general approach to this problem that treats late-onset outcomes as missing data, uses data augmentation to impute missing outcomes from posterior predictive distributions computed from partial follow-up times and complete outcome data, and applies the design’s decision rules using the completed data. We illustrate the method with two cancer trials conducted using a phase I–II design based on efficacy–toxicity trade-offs, including a computer simulation study. Supplementary materials for this article are available online.

KEY WORDS: Bayesian adaptive clinical design; Data augmentation algorithm; Dose-finding; Missing data; Phase I–II clinical trial design; Piecewise exponential model.

1. INTRODUCTION

Phase I–II clinical trial designs combine conventional phase I and phase II trials by determining a dose of a new agent based on both toxicity and efficacy (Gooley et al. 1994; Thall and Russell 1998; O’Quigley, Hughes, and Fenton 2001; Braun 2002; Thall and Cook 2004; Bekele and Shen 2005; Zhang, Sargent, and Mandrekar 2006; Yin, Li, and Ji 2006; Yuan and Yin 2011a). Most commonly, a small phase I trial based on toxicity (Storer 1989; O’Quigley, Pepe, and Fisher 1990; Babb, Rogatko, and Zacks 1998; Conaway, Dunbar, and Peddada 2004) first is conducted to choose a putatively safe dose, the maximum tolerated dose (MTD), and a phase II trial then is conducted (Gehan 1969; Fleming 1982; Simon 1989; Thall and Simon 1994; Thall, Simon, and Estey 1995; Bryant and Day 1995), using the MTD, with efficacy the primary outcome. This conventional approach may lead to several problems. (1) Informal dose adjustments often are made in phase II if excessive toxicity is observed, which invalidates assumed properties of any efficacy-based design. (2) Operating characteristics of the entire phase I – phase II process are seldom computed. (3) Separate designs ignore the trade-off between efficacy and toxicity that often underlies therapeutic decision making. Since phase I–II designs address these problems explicitly, a natural question is why such designs are used infrequently for actual trials. While these are complex issues, the following logistical problem may play a prominent role.

In outcome-adaptive clinical trials, a major practical impediment arises if patient outcomes are not observed quickly enough to apply decision rules that choose treatments or doses for newly accrued patients. Because it is undesirable, and often impossible, to delay a new patient’s treatment while waiting for previous patients’ outcomes to be scored so that an adaptive statistical rule can be applied, outcome-adaptive rules may be at odds with clinical practice. In phase I–II, this problem arises if either toxicity or efficacy is not scored quickly, relative to the accrual rate. One solution is to turn away new patients and treat them off protocol. This may be less desirable than giving the experimental regimen, or impossible if no alternative treatment exists. Another solution is to give all new patients the dose or treatment that is optimal based on the most recent data. This may have very undesirable consequences if the most recent optimal dose later turns out to be overly toxic, and it is the main reason that dose-finding usually is done sequentially with small cohorts of 1, 2, or 3 patients. In Phase I–II, the opposite effect can occur if the most recent optimal dose is safe but inefficacious. A third solution when some patients’ outcomes have not yet been evaluated fully is to treat all new patients immediately, but use a dose one level below the design’s current optimal dose. In Phase I, the problem of dealing with delayed (“late onset”) toxicity was first addressed by Cheung and Chappell (2000), who introduced the time-to-event continual reassessment method (TiTE-CRM), and later by Braun (2006), Bekele et al. (2008), and Yuan and Yin (2011b). As an illustration, consider a phase I–II clinical trial of chemotherapy for acute leukemia, where both toxicity and efficacy events may occur at any time during a common 6-week evaluation period, doses are chosen adaptively for cohorts of size 3, and the accrual rate 1 patient per week. On average, the first cohort will be accrued in 3 weeks with all of their outcomes scored by week 9, when adaptive rules are applied using their data to choose the second cohort’s dose. Since one also can expect to accrue 6 new patients between weeks 3 and 9, the question is how to deal therapeutically with these new patients. A second example is an autologous stem cell transplantation trial for multiple myeloma (MM) where toxicity may occur at
any time during the first 30 days, but efficacy is evaluated only at 90 days post-transplant, and is defined as no detectable MM protein in the urine or blood serum and < 5% plasma cells in the bone marrow. If the accrual rate is 3 patients per month, one may expect to accrue 9 patients before efficacy is scored for any patients, so applying an adaptive rule to choose a dose for patients 4, 5, and 6, (cohort 2), using the data from patients 1, 2, and 3, (cohort 1), is not possible without delaying the second cohort’s therapy. The severity of this problem increases with accrual rate and persists throughout the trial.

In this article, we consider phase I–II designs where both efficacy (E) and toxicity (T) are characterized as binary variables, Y_E and Y_T, evaluated either during or at the end of specified time intervals, [0, U_E] for Y_E and [0, U_T] for Y_T. We refer to Y_E and Y_T as “delayed outcomes” because they are not observed immediately. Denote accrual rate by α. We quantify the severity of the problem using the logistical difficulty indexes \( \xi_j = U_j / \alpha \) for each outcome \( j = E, T \), and overall index \( \xi = \max[\xi_E, \xi_T] \). For example, if \( \alpha = 1 \) patient/month and \( U_E = U_T = 3 \) months, then \( \xi = 3 \). Doubling the accrual rate to \( \alpha = 2 \) gives the problematic value \( \xi = 6 \).

We propose a methodology that handles the problem of delayed outcomes in phase I–II by accounting for each patient’s follow-up time prior to evaluation of \( Y_E \) and \( Y_T \), and treating all \( Y_j \)’s that have not yet been observed as missing values. We use data augmentation (Tanner and Wong 1987) to impute each missing \( Y_j \) using partial follow-up times and complete outcome data. Combining observed and imputed \( (Y_E, Y_T) \) data for all patients who have been treated, we then apply the phase I–II design’s decision rules. Our approach of treating delayed outcomes as missing data is similar to that of Yuan and Yin (2011b), who deal with the phase I setting with \( Y_T \) but not \( Y_E \). Key differences are that the bivariate distribution and missingness patterns of \( (Y_E, Y_T) \) are much more complicated than those of \( Y_T \) alone. In addition, Yuan and Yin (2011b) used the EM algorithm under a frequentist framework to estimate toxicity probabilities, whereas we rely on predictive probabilities and imputation to obtain a completed dataset under a Bayesian formulation.

In Section 2, we propose a general approach to delayed outcomes in phase I–II trials. In Section 3, we present imputation and Bayesian data augmentation methods. Section 4 illustrates the method with two trials designed using an extended version of the efficacy-toxicity (Eff-Tox) trade-off based design of Thall and Cook (2004). In Section 5, we present simulations of the proposed method and comparison with alternative methods. We conclude with a discussion in Section 6.

### 2. OBSERVED OUTCOMES AND MISSING VALUES

#### 2.1 Data Structures

The data structure assumed by most phase I–II dose-finding method consists of assigned doses and two binary outcomes. We denote this by \( \mathbf{D}(Y) = \{d_{i,1}, Y_{1,E}, Y_{1,T}, \ldots, d_{i,n}, Y_{n,E}, Y_{n,T}\} \), where \( d_{i,n} \) is the dose given to the \( i \)th patient and interim sample size \( n = 1, \ldots, N_{\text{max}} \). To account for the fact that, at any interim decision, one or both of \( Y_E \) and \( Y_T \) may not yet be observed for a given patient, we denote the data including only the observed \( Y_j \)’s by \( \mathbf{D_{obs}}(Y) \) and the unobserved (“missing”) \( Y_j \)’s by \( \mathbf{D_{mis}}(Y) \). Our strategy will be to use \( \mathbf{D_{obs}}(Y) \) and partial follow-up time data to impute any missing \( Y_j \)’s in \( \mathbf{D_{mis}}(Y) \) and thus construct a completed version of \( \mathbf{D_{obs}}(Y) \) that has the form of \( \mathbf{D}(Y) \), and then apply the phase I–II method.

Let \( X_j \) denote time to outcome \( j = E, T \). Corresponding to the acute leukemia trial and the stem cell transplantation trial described previously, we distinguish between two data structures, depending on how \( Y_E \) is observed. In Case 1, \( Y_E \) is observed in real time, and \( Y_E = I(X_E \leq U_E) \). In Case 2, \( Y_E \) is evaluated only at \( U_E \), and there is no \( X_E \). In both cases, \( Y_T = I(X_T \leq U_T) \). To simplify exposition, we focus primarily on Case 1 since it is more complex, and later explain how to deal with Case 2. For trials where efficacy is evaluated periodically \( X_E \) is interval censored. We include this by smoothing the interval censored data, placing \( X_E \) at the midpoint or a randomly chosen value of the interval where efficacy was known to occur. We accommodate patient death during therapy by defining \( X_T \) as the time to either nonfatal toxicity or death. For efficacy, if the patient dies prior to \( U_E \) we define \( Y_E = 0 \), which is implied by defining \( X_E = \infty \). Denote \( U = \max \{U_T, U_E\} \) and \( V = \) follow-up time, where by design \( V \leq U \).

#### 2.2 Missingship Mechanism

Denote \( \pi_j(d, \theta) = \Pr(Y_j = 1|d, \theta) \) for \( j = E, T \), where \( \theta \) is the model parameter vector. In the following, we often will suppress \( d \) and \( \theta \) for brevity. The observed follow-up time of \( X_j \) during the evaluation interval \( [0, U_j] \) is \( X_j^o = V \wedge X_j \). We assume that \( V \) is independent of \( X_j \). The key to our method is that \( X_j^o \) provides useful information about \( Y_j \), because

\[
\Pr(Y_j = 1|X_j^o = V < X_j) = 1 - \frac{\Pr(U_j < X_j)}{\Pr(X_j^o = V < X_j)}
\]

must decrease as \( X_j^o \) increases from 0 to \( U_j \). This fact also underlies the TITE-CRM (Cheung and Chappell 2000). To account for missing values, we use \( V \) to extend the previous definitions of \( Y_E \) and \( Y_T \), as follows. We define

\[
Y_j = \begin{cases} 
\text{missing} & \text{if } X_j > V \text{ and } V < U_j \\
1 & \text{if } X_j \leq V \leq U_j \\
0 & \text{if } X_j > V = U_j 
\end{cases}
\]

That is, \( Y_j = \text{missing} \) if the patient has not yet experienced the event and has not been fully followed to \( U_j \), while \( Y_j \) is observed if the patient either has experienced the event \( (Y_j = 1) \) or has completed the defined follow-up time without the event \( (V = U_j \) and \( Y_j = 0) \). Denote the missingship indicators \( M_j = I(Y_j = \text{missing}) \) for \( j = E, T \), and \( M = (M_E, M_T) \).

The 4th patient’s data are \( D_4 = (d_{i,4}, Y_{1,E}, Y_{1,T}, X_{i,E}^o, X_{i,T}^o) \), and \( D_i \) determines \( M_i \). Our methodology uses the actual interim data from \( n \) patients, denoted by \( \mathbf{D} = (D_1, \ldots, D_n) \), to impute
missing \( Y_j \)'s and construct a completed binary dataset \( \mathcal{D}(Y) \) for implementing the phase I–II method.

### 2.3 Event Time Distributions

To construct flexible survival functions, we assume piecewise exponential marginals for \([X_j|d, Y_j = 1], j = E, T, \) by partitioning \([0, U_j]\) into \( K_j \) intervals, \([0, h_{j,1}], (h_{j,1}, h_{j,2}), \ldots, [h_{j,K_j-1}, h_{j,K_j}]\), and assuming hazard \( \lambda_{j,k} \) on \([h_{j,k-1}, h_{j,k})\). The marginal survival function for \( X_j \) is

\[
S_j(x|Y_j = 1, \lambda_j) = \exp \left\{ - \sum_{k=1}^{K_j} w_{j,k}(x) \lambda_{j,k} \right\}, \quad x > 0,
\]
denoting \( \lambda_j = (\lambda_{j,1}, \ldots, \lambda_{j,K_j}) \), and weights \( w_{j,k}(x) = h_{j,k} - h_{j,k-1} \) if \( x > h_{j,k} \), \( w_{j,k}(x) = x - h_{j,k-1} \) if \( x \in [h_{j,k-1}, h_{j,k}) \), and \( w_{j,k}(x) = 0 \) otherwise. Initially, we considered a more elaborate form of \( S_j \), in which each \( \lambda_{j,k} \) was replaced by \( \lambda_{j,k} \gamma_{j,k} \), where \( \gamma_j \) > 1, so that the event time distributions varied with dose in a proportional hazard model with piecewise exponential baseline hazard. However, we found that fixing \( \gamma_j = 1 \) did not change the method’s operating characteristics, because there is little or no information to estimate the \( \gamma_j \)'s. To determine whether a more parsimonious survival model might give a design with similar performance, we replaced the piecewise exponential with a Weibull. However, simulations across eight dose-outcome scenarios (Supplementary Table S9) showed that the Weibull gave larger probabilities of incorrectly stopping early, and either similar or lower probabilities of selecting desirable doses.

Our imputation method requires the joint conditional survival probabilities \( S(x_j, x_{T}|a, b) = \Pr(X_E > x_j, X_T > x_T|Y_E = a, Y_T = b) \) for \((a, b) = (0, 1), (1, 0), (1, 1)\). Assuming conditional independence, \( S_j(x_j|Y_E, Y_T) = S_j(x_j|Y) \) for \( j = E, T, \) implies that \( S(x_j, x_T|1, 0) \) is determined by the marginal of \( X_E \) and \( S(x_j, x_T|0, 1) \) is determined by the marginal of \( X_T \). Determining \( S(x_j, x_T|1, 1) \) requires accounting for association between \( X_E \) and \( X_T \). We do this by defining a joint distribution using the Clayton copula (Clayton 1978), given for \( \phi \geq 0 \) by

\[
S(x_j, x_T|1, 1) = \left( S_E(x_j|Y_E = 1)^{-1/\phi} + S_T(x_T|Y_T = 1)^{-1/\phi} - 1 \right)^{-\phi} = \left\{ \exp \left\{ \sum_{k=1}^{K_j} \lambda_{E,k} w_{E,k}/\phi \right\} + \exp \left\{ \sum_{k=1}^{K_T} \lambda_{T,k} w_{T,k}/\phi \right\} - 1 \right\}^{-\phi}.
\]

The likelihood of \( \mathcal{D} \) depends on \( Y_E, Y_T, \) and the censoring patterns of \( X_E \) and \( X_T \). Denote \( \lambda = (\lambda_E, \lambda_T), \xi = (\phi + 1)/\phi, \epsilon_j = I(X_j > X_j) \), the pdf of \( X_j \) by \( f_j(\cdot) \), and

\[
W(X_E^0, X_T^0) = S_E(X_E^0|Y_E = 1)^{-1/\phi} + S_T(X_T^0|Y_T = 1)^{-1/\phi} - 1.
\]

When \( Y_E = Y_T = 1 \), there are four possible likelihoods,

\[
L(\mathcal{D} | \lambda, \phi) = \begin{cases} 
\xi W(X_E^0, X_T^0)^{-\phi-2} \\
\times \prod_{j=E}^{T} f_j(X_j|Y_j = 1) \\
\times \left\{ \prod_{j=E}^{T} S_j(X_j^0|Y_j = 1) \right\}^\xi & \text{if } (\epsilon_x, \epsilon_y) = (1, 1) \\
W(X_E^0, X_T^0)^{-\phi-1} f_E(X_E^0|Y_E = 1) & \text{if } (\epsilon_x, \epsilon_y) = (1, 0) \\
W(X_E^0, X_T^0)^{-\phi-1} f_T(X_T^0|Y_T = 1) & \text{if } (\epsilon_x, \epsilon_y) = (0, 1) \\
W(X_E^0, X_T^0)^{-\phi} & \text{if } (\epsilon_x, \epsilon_y) = (0, 0)
\end{cases}
\]

For the cases \((Y_E, Y_T) = (1, 0) \) or \((0, 1)\), the likelihoods are given by

\[
L(\mathcal{D} | \lambda, \phi) = \left[ \frac{f_E(X_E^0|Y_E = 1)}{S_E(X_E^0|Y_E = 1)^{1-\epsilon_x}} \right]^{Y_E(1-Y_T)} \\
\times \left[ \frac{f_T(X_T^0|Y_T = 1)}{S_T(X_T^0|Y_T = 1)^{1-\epsilon_y}} \right]^{Y_T(1-Y_E)}
\]

Denoting the likelihood for \( n \) patients by \( L(\mathcal{D} | \lambda, \phi) = \prod_{i=1}^{n} L(\mathcal{D} | \lambda, \phi) \), the posterior is \( f(\lambda, \phi | \mathcal{D}) \propto f(\lambda) f(\phi) L(\mathcal{D} | \lambda, \phi) \), for priors \( f(\lambda) \) and \( f(\phi) \) of \( \lambda \) and \( \phi \).

### 3. IMPUTATION METHOD

Let \( \pi_{a,b}(d, \theta) = \Pr(Y_E = a, Y_T = b|d, \theta) \) for \( a, b \in \{0, 1\} \) denote the joint distribution of \([Y_E, Y_T|d]\), so \( \pi_E(d, \theta) = \pi_{1,1}(d, \theta) + \pi_{1,0}(d, \theta) \) and \( \pi_T(d, \theta) = \pi_{1,1}(d, \theta) + \pi_{0,1}(d, \theta) \). If no outcomes are missing the likelihood is the usual product

\[
L(\mathcal{D}(Y)|\theta) = \prod_{i=1}^{n} \prod_{a,b=0}^{1} \pi_{a,b}(d_i|\theta)^{I(Y_E=Y_T)^{a,b}}, \quad (1)
\]

with posterior \( f(\theta | \mathcal{D}(Y)) \propto L(\mathcal{D}(Y)|\theta) f(\theta) \), for prior \( f(\theta) \). Since \( L(\mathcal{D}(Y)|\theta) \) cannot be computed if some \( Y_j \)'s are missing, we obtain a completed version of the likelihood, of the form (1), by using Bayesian data augmentation (Tanner and Wong 1987). We iterate between an imputation step sampling missing \( Y_j \)'s from their full conditionals, and a step computing the posterior using the completed data. Importantly, the imputation depends on the dose-outcome model.

Missing \( Y_j \)'s are nonignorable because, at follow-up time \( V \), a patient who will not experience outcome \( j \) by \( U_j \) is more likely to have \( Y_j = \text{missing} (M_j = 1) \) than a patient for whom outcome \( j \) will occur (Yuan and Yin 2011b). By Bayes’ Law, \( \Pr(M_j = 1|Y_j = 0) > \Pr(M_j = 1|Y_j = 1) \) implies that

\[
\frac{\pi_j}{1 - \pi_j} > \frac{\Pr(Y_j = 1|M_j = 1)}{\Pr(Y_j = 0|M_j = 1)}.
\]

This says that the odds that \( Y_j = 1 \) decreases if \( Y_j \) is missing, so the missingness indicator \( M_j \) contains information about the future value of \( Y_j \).
A complication in carrying out the imputation is that there are three possible missingness patterns: (1) \( Y_E \) is missing and \( Y_T \) is observed, (2) \( Y_T \) is missing and \( Y_E \) is observed, and (3) both \( Y_E \) is missing and \( Y_T \) is missing. Since we impute missing values by sampling from their full conditional posteriors, these distributions must be specified for each missingness pattern. These posteriors are defined in terms of the following conditionals for \( X_E \) and \( X_T \), given each of the four possible future outcome pairs, which for brevity we denote by

\[
S_{ab} = \Pr(X_E > V, X_T > V | Y_E = a, Y_T = b), \quad a, b \in \{0, 1\}.
\]

In any case, \( S_{00} = 1 \) because \( \Pr(X_j > V | Y_j = 0) = 1 \) for \( 0 \leq V \leq U_j \), which also implies that \( S_{10} = \Pr(X_E > V | Y_E = 1, Y_T = 0) = 0 \) and \( S_{01} = \Pr(X_T > V | Y_E = 0, Y_T = 1) = 1 \). Thus, only \( S_{11} \) involves a joint distribution for \( (X_E, X_T) \) given \( Y_E = Y_T = 1 \).

To obtain consistent estimates when imputing the missing \( Y_j \)'s, we condition on the actual data \( D \). The following posterior conditional distributions are derived in the Appendix. When \( Y_E \) is missing and \( Y_T \) is observed, we impute \( Y_E \) from its conditional posterior

\[
\Pr(Y_E = 1 | D) = \left\{ \frac{\pi_{1,1}S_{10}}{\pi_{1,1}S_{10} + \pi_{1,0}} \right\} \frac{\pi_{1,0}S_{10}}{\pi_{1,0}S_{10} + \pi_{0,0}}^{1-Y_T}.
\]

When \( Y_T \) is missing and \( Y_E \) is observed, we impute \( Y_T \) from its conditional posterior

\[
\Pr(Y_T = 1 | D) = \left\{ \frac{\pi_{1,1}S_{01}}{\pi_{1,1}S_{01} + \pi_{1,0}} \right\} \frac{\pi_{0,1}S_{01}}{\pi_{0,1}S_{01} + \pi_{0,0}}^{1-Y_E}.
\]

When both \( Y_E \) and \( Y_T \) are missing, we impute \( (Y_E, Y_T) \) from the joint conditional posterior

\[
\Pr(Y_E = y, Y_T = z | D) = \frac{\pi_{y,z}S_{yz}}{\sum_{y=0}^{1} \sum_{z=0}^{1} \pi_{a,b}S_{ab}}, \quad \text{for} \quad y, z = 0, 1.
\]

At the posterior step using the completed data, we sample parameters from their full conditional posteriors in two steps: (1) sample \( \theta \) from \( f(\theta | D(Y)) \) and (2) sample \( \phi \) and \( \lambda_{j,k} \) for each \( k = 1, \ldots, K_j \) and \( j = (E, T) \) from \( f(\lambda, \phi | D) \). This is iterated until the Markov chain converges, with posteriors computed using adaptive rejection Metropolis sampling (Gilks, Best, and Tan 1995).

We now turn to Case 2, in which \( Y_E \) is evaluated only at \( U_E \). Analytically, Case 2 is much simpler than Case 1 because in Case 2 efficacy is \( Y_E \) with no random event time involved, and \( Y_E \) is missing completely at random (MCAR) at all \( V < U_E \). For the same reason, trial conduct in Case 2 is much harder logistically than in Case 1. This is because there is no \( X_E \) in Case 2 and hence no partial information about \( Y_E \) when \( V < U_E \). Inference for \( \pi_E \) relies entirely on observed \( Y_E \) values from previously treated patients, while \( Y_T \) may be imputed by exploiting the event time data \( (X_E, Y_E) \) using the marginal of \( X_T \).

Our proposed method can be applied to any phase I–II combination of probability model and decision rules based on \( [Y_E, Y_T | d] \) with delayed outcomes. To make things concrete, we apply the method to the Efficacy–Toxicity (EffTox) tradeoff method of Thall and Cook (2004). Case 1 is illustrated by the trial of chemotherapy for acute leukemia where \( U_T = U_E = 42 \) days, and Case 2 by the stem cell transplantation trial with \( U_T = 30 \) days and \( U_E = 90 \) days.

4. EFFTOX DESIGNS

4.1 The Design Assuming Complete Data

We first briefly review the EffTox phase I–II design (Thall and Cook 2004) assuming complete data \( D(Y) \). The model for \( [Y_E, Y_T | d] \) assumes marginal probabilities \( \pi_E = \logit^{-1}(\mu_E + \beta_E, d + \beta_E, d^2) \) and \( \pi_T = \logit^{-1}(\mu_T + \beta_T, d + \beta_T, d^2) \) and uses a Gumbel copula (Murtaugh and Fisher 1990) to obtain a joint distribution,

\[
\pi_{a,b} = (\pi_E)^{\psi}(1-\pi_E)^{1-\psi}(1-\pi_T)^{1-b} + (-1)^{\psi+b}\pi_E(1-\pi_E)\pi_T(1-\pi_T)\left(\frac{e^{\theta}-1}{e^{\theta}+1}\right),
\]

where \( \psi \) parameterizes association between \( Y_E \) and \( Y_T \). To prevent misestimation with small sample sizes, we require that \( \beta_E, 1 + 2\beta_E, d > 0 \) and \( \beta_T, 1 + 2\beta_T, d > 0 \), corresponding to agents where both \( \pi_T \) and \( \pi_E \) increase with dose.

The desirability of a dose is defined by first defining the desirability \( \delta(\pi_E, \pi_T) \) of each probability pair \( \pi = (\pi_E, \pi_T) \) in \([0, 1]^2\), with \( \delta(\pi_E, \pi_T) \) increasing in \( \pi_E \) and decreasing in \( \pi_T \). Denoting the posterior means \( \pi_{1,d} = E[\pi_{1,d}(\theta | D(Y))] \) for \( j = E, T \), the desirability of dose \( d \) is \( \delta(\pi_{E,d}, \pi_{T,d}) \), for use as a decision criterion. The function \( \delta \) may be obtained from a target efficacy–toxicity tradeoff contour, \( C \), in \([0, 1]^2 \) in several ways (see Thall, Cook, and Estey 2006). To avoid doses that are too toxic or inefficacious, two admissibility criteria are imposed. Given elicited fixed lower limit \( \pi_E \) on \( \pi_E \) and upper limit \( \pi_T \) on \( \pi_T \), a dose \( d \) is acceptable if

\[
Pr[\pi_{E,d} \geq \pi_E | D(Y)] > p_E \quad \text{and} \quad Pr[\pi_{T,d} < \pi_T | D(Y)] > p_T
\]

for prespecified cutoffs \( p_E \) and \( p_T \). The trial starts at a dose chosen by the physician and each new cohort is treated with the acceptable dose having largest \( \delta(d) \). An untried dose may not be skipped when escalating, and when escalating to an untried dose only the toxicity admissibility rule is imposed. If no dose is acceptable the trial is stopped with no dose selected.

We refer to the new version of the EffTox design that incorporates our proposed imputation methodology as the late onset (LO)-EffTox design. It differs from the EffTox design in one fundamental way, namely that event time data are exploited in a fundamental way, namely that event time data are exploited to compute posterior decision criteria using the Bayesian data augmentation methodology described in Section 3.

4.2 Prior Specification

In any model-based Bayesian adaptive clinical trial design, the prior must be sufficiently vague that the accumulating data dominates the posterior distribution, and thus the adaptive decisions. Thall and Cook (2004) provided a framework for establishing priors from elicited mean values of \( \pi_E(d, \theta) \) and \( \pi_T(d, \theta) \) that solves for prior hyperparameters using nonlinear least squares. We derive a prior based on the weakly informative prior for logistic regression proposed by Gelman et al. (2008). Following Gelman et al. (2008), for priors we
assume $\mu_T, \mu_E, \beta_T, 1, \beta_T, 2, \beta_E, 2 \sim \text{iid Cauchy}(0, 2.5)$, and then shift the six Cauchy prior location parameters from 0 to $\tilde{\mu} = (\tilde{\mu}_{\psi, 1}, \tilde{\mu}_{\psi, 2}, \tilde{\mu}_{\psi, 3}, \tilde{\mu}_{\psi, 4}, \tilde{\mu}_{\psi, 5}, \tilde{\mu}_{\psi, 6})$, to reflect prior opinion. To obtain $\tilde{\mu}$, similarly to Thall and Cook (2004), we first elicit the means $\tilde{m}_{E, d_r}$ and $\tilde{m}_{T, d_r}$ of $\pi_E(d_r, \theta)$ and $\pi_T(d_r, \theta)$ from the physician for each dose $d_r, r = 1, \ldots, R$. For each $j = E, T$, we use least squares to solve for $(\tilde{\mu}_{\psi, 1}, \tilde{\mu}_{\psi, 2}, \tilde{\mu}_{\psi, 3})$ by assuming $E(\log(\tilde{m}_{j, d_r})) = \tilde{\mu}_{\psi, 1} + \tilde{\mu}_{\psi, 2}d_r + \tilde{\mu}_{\psi, 3}d_r^2$.

To obtain a vague prior on $\lambda$, we assume $[X_j | Y_j = 1] \sim \text{Unif}(0, U_j)$, which implies that the hazard at the midpoint of the subinterval $[h_{j,k-1}, h_{j,k})$ of the partition $K_j = \{K_j = k + 0.5\}$. We assume that $\lambda_{j,k} \sim \text{Gam}(\lambda_{j,k}/C, 1/C)$, where $\text{Gam}(a, b)$ denotes the gamma distribution with mean $a/b$ and variance $(a/b)^2$. Thus, $\lambda_{j,k}$ has prior mean $\lambda_{j,k}$ and variance $C\lambda_{j,k}$, so $\psi$ is a tuning parameter that determines $\text{var}(\lambda_{j,k})$ and can be calibrated by simulation. In preliminary simulations, we found that $C = 2$ yields a reasonably vague prior and a design with good operating characteristics. Finally, we assume $\psi \sim \text{Normal}(0, 1)$ and $\phi \sim \text{Gam}(0.2, 0.2)$.

5. COMPUTER SIMULATIONS

5.1 Clinical Trial Designs

We simulated our proposed methodology to study its behavior when applied to the EffTox design for phase I-II trials in each of Cases 1 and 2. To assess robustness, we conducted a second set of simulations in which we varied several model, design, and simulation scenario parameters. In all simulations, we considered hypothetical phase I-II trials with five raw doses (2.5, 5.0, 7.5, 10.0, 12.5). In the model, we replaced each raw dose $d^r_s$ by the standardized dose $d_r = (0.5/s)(\log(d^r_s) - \log(d^r_0))$, where $s$ is the standard deviation of the centered log doses, so $d_1, \ldots, d_5$ are centered around 0 and have standard deviation 0.5. The trade-off contour, $C$, was determined by fitting a quadratic curve to the trade-off target probability pairs $(\pi_E(d_r, \theta), \pi_T(d_r, \theta))$. Applying the method described in Section 4.2, this gave location parameters $(\tilde{\mu}_{\psi, 1}, \tilde{\mu}_{\psi, 2}, \tilde{\mu}_{\psi, 3}) = ( \begin{pmatrix} -1.21 \ 0.96 \ 0.35 \end{pmatrix})$ and $(\tilde{\mu}_{\psi, 1}, \tilde{\mu}_{\psi, 2}, \tilde{\mu}_{\psi, 3}) = (\begin{pmatrix} -1.16 \ 1.39 \ 0.85 \end{pmatrix})$ for the shifted Cauchy priors. For the Gamma piecewise exponential event rate priors, we assumed $K = 6$. The formula in Section 4.2 gives $(\tilde{\lambda}_{E, 1}, \ldots, \tilde{\lambda}_{E, 6}) = (\begin{pmatrix} 0.182 \ 0.222 \ 0.286 \ 0.400 \ 0.667 \ 2.000 \end{pmatrix})$ in Case 1, and $(\tilde{\lambda}_{T, 1}, \ldots, \tilde{\lambda}_{T, 6}) = (0.364, 0.444, 0.571, 0.800, 1.333, 4.000)$ in Case 2. We used tuning parameter $C = 2$. Thus, for example, $\lambda_{E, 1} \sim \text{Gam}(0.093, 0.5)$ in Case 1.

5.2 Simulation Study Design

Each simulation scenario was specified in terms of assumed true efficacy and toxicity probabilities at each dose, $\pi_j(d_r)_{\text{true}}$, for $j = E, T$ and $r = 1, \ldots, 5$. We modeled association between $X_E$ and $X_T$ by assuming a Clayton copula (Clayton 1978) with $\phi = 1.0$. We considered eight scenarios, illustrated in Figure 2. In Scenario 8, no dose is admissible, since $d = 1.2$ are ineffectivc and $d = 3.45$ are too toxic. For the first simulation study, we generated $X_E, X_T$ from Weibull distributions. For each $j = E, T$, the Weibull scale and shape parameters were chosen so that (1) $\Pr(X_J \leq U)_{(j, r)}^{\text{true}} = \pi_j(d_r)_{\text{true}}$ and (2) $\pi_j(d_r)_{\text{true}} = \Pr(U_J / 2 \leq X J \leq U_J)_{(j, r)}^{\text{true}} = 0.50$, that is, 50% of the events occurred in the second half of the evaluation interval. Because $\pi_j(d_r)_{\text{true}}$ and $\pi_j(d_r)_{\text{true}}$ vary with $d_r$, in each scenario, the scale and shape parameters of the corresponding true Weibull distributions both vary with $d_r$. Each scenario was simulated 1000 times.

5.3 Simulation Results for Case 1

For Case 1, we compared the LO-EffTox design to three methods that are used in practice to deal with the late onset problem. The first method is the “One Level Down” rule. With

![Figure 1. Desirability contours: the target contour is represented by a solid line. Other contours on which all $(\pi_E, \pi_T)$ have the same desirability are shown as dashed lines.](image-url)
Figure 2. Illustration of scenarios: the solid lines with circles represent true probabilities of efficacy and the dashed lines with triangles represent true probabilities of toxicity. The horizontal lines represent $\bar{\pi}_T$ and $\bar{\pi}_E$. 

Scenario 1

Scenario 2

Scenario 3

Scenario 4

Scenario 5

Scenario 6

Scenario 7

Scenario 8
this method, if some of the patients treated at the current optimal dose, $d_{\text{opt}} = d$, have not yet been evaluated fully, that is, $Y_{i,E,d} = \text{missing}$ or $Y_{i,T,d} = \text{missing}$, then any new patient is treated at $d_{i-1}$. The second method is the “Look Ahead” rule (Thall et al. 1999) which says that, for each possible value $\bar{D}_{\text{mis}}(Y)$ that $D_{\text{mis}}(Y)$ may take on, use the completed data $D_{\text{obs}}(Y) \cup \bar{D}_{\text{mis}}(Y)$ to compute $d_{\text{opt}}$. If this dose is the same for all possible $\bar{D}_{\text{mis}}(Y)$, then use that dose to treat the next patient immediately. Otherwise, the only two options for new patients are to make them wait to be treated, which usually is impossible in practice, or to turn them away and treat them off protocol. The third method uses all complete cases, where both $Y_E$ and $Y_T$ are observed, to compute $d_{\text{opt}}$ and treat the next patient immediately.

We will use two summary criteria to evaluate each method’s performance and compare the three methods. Denote the true desirability of dose $d_i$ by $\delta_{r,\text{true}}^i$ and the true set of acceptable doses by $A_{\text{true}}$. The first criterion is the desirability-weighted selected percentage,

$$\delta = \frac{\sum_{r=1}^5 \delta_{r,\text{true}}^i \Pr(\text{select } d_i | d_i \in A_{\text{true}})}{\sum_{r=1}^5 \delta_{r,\text{true}}^i | d_i \in A_{\text{true}}},$$

which quantifies dose selection reliability and thus potential benefit for future patients. The second criterion is the ratio $N_E/N_T$ where $N_E$ and $N_T$ denote the number of patients who experienced efficacy and toxicity, respectively. This criterion quantifies benefit to the patients in the trial, hence may be considered an index of ethical desirability.

Table 1 gives the operating characteristics of the LO-EffTox design and three competing methods. The percentage of trials with no dose selected is denoted by “None,” with the numbers of patients turned away from the trial by the Look Ahead method given in parentheses. In general, LO-EffTox outperforms the One Level Down, Look Ahead rule, and Complete Case method. The One Level Down rule has much smaller correct selection rate and mean $N_E$. The Look Ahead rule design performs roughly equivalently to LO-EffTox in terms of dose selection percentages and $N_E/N_T$, but the trial durations under the Look Ahead rule are dramatically larger compared to LO-EffTox. This is because the Look Ahead rule turns many patients away, while LO-EffTox treats all patients. Compared to the Complete Case method, LO-EffTox has either similar or higher correct selection percentages and more patients treated at doses with higher desirability. In Scenario 8, where all doses are inefficacious or too toxic, LO-EffTox has by far the largest probability of correctly stopping early and selecting no dose.

Figure 3 illustrates the results in Table 1 in terms of $\delta$ plotted on the vertical axis and $N_E/N_T$ on the horizontal axis, under each of Scenarios 1–7. Scenario 8 is not included in Figure 3 since in this case no dose is acceptable, so $\delta$ is not relevant. Values in the upper right portion of the figure are more desirable, while values in the lower left are less desirable. Figure 3 shows that the One Level Down rule produces designs with very poor properties, in terms of both $\delta$ and $N_E/N_T$. These two criteria are roughly equivalent for LO-EffTox and the Look Ahebd version of EffTox for each of Scenarios 1–6. In Scenario 7, which has true desirability not monotone in dose, LO-EffTox has much greater $\delta$ and much smaller $N_E/N_T$ compared to the “Look Ahead” version of EffTox, so in this case there is no clear winner. However, since the Look Ahead rule turns away many patients and produces a very long trial, the apparent equivalence in terms of the two criteria in Figure 3 only tells part of the story. Compared to the Complete Case method, LO-EffTox has either similar or much larger $\delta$ values and similar $N_E/N_T$.

5.4 Simulation Results for Case 2

For Case 2, we compared the LO-EffTox design to the One Level Down, Look Ahead, and Complete Case rules, and also
Table 1. Simulation study comparing the four phase I–II designs in Case 1

<table>
<thead>
<tr>
<th>Method</th>
<th>Scenario 1</th>
<th>Scenario 2</th>
<th>Scenario 3</th>
<th>Scenario 4</th>
<th>Scenario 5</th>
<th>Scenario 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>None</td>
</tr>
<tr>
<td>LO-EffTox % selected</td>
<td>0.0 (0.05, 0.03)</td>
<td>0.0 (0.10, 0.05)</td>
<td>0.0 (0.20, 0.07)</td>
<td>0.0 (0.25, 0.08)</td>
<td>0.0 (0.35, 0.10)</td>
<td>0.0 (0.18, 0.20)</td>
</tr>
<tr>
<td>LO-EffTox # patients</td>
<td>6.15</td>
<td>7.50</td>
<td>18.73</td>
<td>27.25</td>
<td>38.49</td>
<td>54.00</td>
</tr>
<tr>
<td>One Down % selected</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>One Down # patients</td>
<td>6.15</td>
<td>7.50</td>
<td>18.73</td>
<td>27.25</td>
<td>38.49</td>
<td>54.00</td>
</tr>
<tr>
<td>Look Ahead % selected</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Look Ahead # patients</td>
<td>6.15</td>
<td>7.50</td>
<td>18.73</td>
<td>27.25</td>
<td>38.49</td>
<td>54.00</td>
</tr>
<tr>
<td>Complete Case % selected</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Complete Case # patients</td>
<td>6.15</td>
<td>7.50</td>
<td>18.73</td>
<td>27.25</td>
<td>38.49</td>
<td>54.00</td>
</tr>
<tr>
<td>N_e/N_T</td>
<td>11.8/3.8</td>
<td>5.7/2.4</td>
<td>3.1</td>
<td>16.7/12.5</td>
<td>13.1/11.9</td>
<td>16.2/8.9</td>
</tr>
<tr>
<td>Duration (weeks)</td>
<td>36.9</td>
<td>37.0</td>
<td>37.1</td>
<td>74.2</td>
<td>37.0</td>
<td>37.4</td>
</tr>
</tbody>
</table>

(Continued on next page)
Table 1. Simulation study comparing the four phase I–II designs in Case 1 (Continued)

<table>
<thead>
<tr>
<th>Method</th>
<th>% selected</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>None</th>
<th>$N_E/N_T$</th>
<th>Duration (weeks)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete Case</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td># patients</td>
<td>3.0%</td>
<td>3.0%</td>
<td>3.0%</td>
<td>3.0%</td>
<td>3.0%</td>
<td>3.0%</td>
<td>3.0%</td>
<td>3.0%</td>
<td>3.0%</td>
</tr>
<tr>
<td>Scenario 7</td>
<td>(0.02, 0.10)</td>
<td>(0.05, 0.25)</td>
<td>(0.30, 0.30)</td>
<td>(0.40, 0.55)</td>
<td>(0.50, 0.70)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LO-EffTox</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td># patients</td>
<td>3.0%</td>
<td>3.0%</td>
<td>3.0%</td>
<td>3.0%</td>
<td>3.0%</td>
<td>3.0%</td>
<td>3.0%</td>
<td>3.0%</td>
<td>3.0%</td>
</tr>
<tr>
<td>Scenario 8</td>
<td>(0.02, 0.10)</td>
<td>(0.05, 0.25)</td>
<td>(0.35, 0.55)</td>
<td>(0.40, 0.60)</td>
<td>(0.50, 0.70)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: The probability pairs given in parentheses for each dose level are $(\pi_T(d_r^{\text{ume}}), \pi_T(d_r^{\text{ume}}))$.

5.5 Sensitivity Analyses

To investigate robustness, we conducted additional simulations in which we varied each of $\pi_T(d_r^{\text{ume}})$, logistical difficulty index $\zeta$, $N_{\text{max}}$, the event time distributions, $K$ = number of sets in the piecewise exponential event time distribution partition, and the association parameter, $\phi$. Each sensitivity analysis was conducted for all scenarios, and the results are summarized in Supplementary Tables S2–S8. Table S2 shows that the design is not sensitive to changes in values of $\pi_T(d_r^{\text{ume}})$ over the range 0.10–0.90, illustrated in Figure 4(a). Tables S3 and S4 summarize sensitivity to the logistical difficulty index for the values $\zeta = 3.0–24.0$ in Case 1 and $\zeta = 9.0–54.0$ in Case 2. As $\zeta$ increases, the ratio $N_E/N_T$ and $\delta$ both decrease, illustrated in Figure 4(b). Table S5 shows, as expected, that the design’s performance improves with larger $N_{\text{max}}$, illustrated in Figure 4(c). Simulations with $X_E$ and $X_T$ generated from several combinations of the Weibull and Log-logistic (Table S6) showed that LO-EffTox is robust to the true event time distributions. Table S7 shows that the number of sets in the partition of the piecewise exponential has little effect on the method’s performance for $K = 6–12$. Varying the association parameter from $\phi = 0.1$ to 2.5 had almost no effect on performance (see Table S8).

6. DISCUSSION

We have proposed a general methodology to address the problem of late-onset outcomes in phase I–II clinical trials. The method treats unobserved binary outcomes as nonignorable missing data, uses data augmentation to impute the missing outcomes, and applies the design’s decision rules using the completed data. Simulations show that, in most cases, the proposed design performs better than alternative approaches to the late onset problem in phase I–II trials. Our results suggest that the...
Table 2. Simulation study comparing the five phase I–II designs in Case 2

<table>
<thead>
<tr>
<th>Method</th>
<th>Dose Level</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>None</td>
<td>N_I/N_T</td>
<td>Duration (days)</td>
<td></td>
</tr>
<tr>
<td>LO-EffTox</td>
<td></td>
<td>0.52</td>
<td>0.53</td>
<td>0.58</td>
<td>0.61</td>
<td>0.67</td>
<td></td>
<td></td>
<td>12.7/3.9</td>
<td>237</td>
</tr>
<tr>
<td></td>
<td>% selected</td>
<td>0.0</td>
<td>0.4</td>
<td>1.1</td>
<td>2.2</td>
<td>90.9</td>
<td>5.4</td>
<td></td>
<td>2.9/1.6</td>
<td>246</td>
</tr>
<tr>
<td></td>
<td># patients</td>
<td>3.58</td>
<td>4.25</td>
<td>5.99</td>
<td>4.91</td>
<td>27.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One Down</td>
<td></td>
<td>38.74</td>
<td>9.23</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
<td></td>
<td>13.3/4.0</td>
<td>399</td>
</tr>
<tr>
<td>Look Ahead</td>
<td></td>
<td>0.0</td>
<td>0.0</td>
<td>1.6</td>
<td>1.9</td>
<td>91.0</td>
<td>5.5</td>
<td></td>
<td>6.2/2.4</td>
<td>244</td>
</tr>
<tr>
<td>Complete Case</td>
<td></td>
<td>3.25</td>
<td>3.49</td>
<td>4.54</td>
<td>6.11</td>
<td>29.26</td>
<td>(49.2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>% selected</td>
<td>38.96</td>
<td>3.64</td>
<td>3.91</td>
<td>3.35</td>
<td>7.97</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TITE-CRM</td>
<td></td>
<td>3.22</td>
<td>3.23</td>
<td>3.18</td>
<td>3.40</td>
<td>34.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Continued on next page)

(Continued)
Table 2. Simulation study comparing the five phase I–II designs in Case 2 (Continued)

<table>
<thead>
<tr>
<th>Method</th>
<th>% selected</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>None</th>
<th>N_E/N_T</th>
<th>Duration (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TiTE-CRM</td>
<td>% selected</td>
<td>0.0</td>
<td>11.8</td>
<td>51.6</td>
<td>30.2</td>
<td>4.6</td>
<td>1.8</td>
<td>23.5/16.7</td>
<td>237</td>
</tr>
<tr>
<td># patients</td>
<td></td>
<td>4.47</td>
<td>7.42</td>
<td>14.68</td>
<td>13.57</td>
<td>7.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scenario 8</td>
<td>δtrue</td>
<td>0.43</td>
<td>0.34</td>
<td>0.29</td>
<td>0.28</td>
<td>0.27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.02, 0.10)</td>
<td></td>
<td>(0.05, 0.25)</td>
<td>(0.35, 0.55)</td>
<td>(0.40, 0.60)</td>
<td>(0.50, 0.70)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LO-EffTox</td>
<td>% selected</td>
<td>0.0</td>
<td>1.9</td>
<td>8.1</td>
<td>0.4</td>
<td>0.1</td>
<td>89.5</td>
<td>7.9/13.6</td>
<td>117</td>
</tr>
<tr>
<td># patients</td>
<td></td>
<td>3.36</td>
<td>7.21</td>
<td>13.99</td>
<td>4.18</td>
<td>1.79</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One Down</td>
<td>% selected</td>
<td>0.0</td>
<td>4.2</td>
<td>90.1</td>
<td>1.4</td>
<td>0.0</td>
<td>4.3</td>
<td>1.3/6.1</td>
<td>247</td>
</tr>
<tr>
<td># patients</td>
<td></td>
<td>38.62</td>
<td>9.31</td>
<td>0.05</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Look Ahead</td>
<td>% selected</td>
<td>0.0</td>
<td>1.6</td>
<td>18.8</td>
<td>1.1</td>
<td>0.1</td>
<td>78.4</td>
<td>8.2/13.4</td>
<td>367</td>
</tr>
<tr>
<td># patients</td>
<td></td>
<td>3.22</td>
<td>8.10</td>
<td>16.11</td>
<td>3.79</td>
<td>1.17</td>
<td>(53.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complete Case</td>
<td>% selected</td>
<td>0.0</td>
<td>8.2</td>
<td>33.5</td>
<td>8.7</td>
<td>3.9</td>
<td>45.7</td>
<td>5.0/10.6</td>
<td>204</td>
</tr>
<tr>
<td># patients</td>
<td></td>
<td>28.81</td>
<td>4.85</td>
<td>7.18</td>
<td>2.84</td>
<td>1.15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TiTE-CRM</td>
<td>% selected</td>
<td>5.8</td>
<td>58.7</td>
<td>31.7</td>
<td>0.9</td>
<td>0.0</td>
<td>2.9</td>
<td>9.1/17.5</td>
<td>236</td>
</tr>
<tr>
<td># patients</td>
<td></td>
<td>8.00</td>
<td>17.23</td>
<td>14.29</td>
<td>5.65</td>
<td>1.69</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: The probability pairs given in parentheses for each dose level are (π_E(d_i)|^{Dmax}, π_T(d_i)|^{Dmax}).

The general approach of imputing binary vectors $Y$ by using time-to-event variables used to define $Y$ may improve the logistics of any outcome-adaptive procedure based on the distribution of $Y$.

**APPENDIX A**

The probability of $Y_E = 1$ with the known $Y_T$ value is

$$\text{pr}(Y_E = 1|X_E > V, Y_T) = \frac{\sum_{a=0}^{b} \text{pr}(Y_E = a) \text{pr}(Y_T, X_E > V|Y_E = a) \text{pr}(X_E > V|Y_T, Y_E = a)}{\sum_{a=0}^{b} \text{pr}(Y_E = a) \text{pr}(X_E > V|Y_E = a)}$$

if $Y_T = 1$, then

$$= \frac{\pi_{1,1} \text{pr}(X_E > V|Y_E = 1) + \pi_{0,1}}{\pi_{1,1} + \pi_{0,1}}$$

where $S_{ab}$ and $\pi_{ab}$ are defined by Equations (2) and (3), respectively.

The probability of $Y_T = 1$ with the known $Y_E$ value is

$$\text{pr}(Y_T = 1|X_T > V, Y_E) = \frac{\sum_{b=0}^{b} \text{pr}(Y_T = b) \text{pr}(X_T > V|Y_T = b)}{\sum_{b=0}^{b} \text{pr}(Y_T = b) \text{pr}(X_T > V|Y_T = b)}$$

if $Y_E = 1$, then

$$= \frac{\pi_{1,0} \text{pr}(X_T > V|Y_E = 1) + \pi_{0,0}}{\pi_{1,0} + \pi_{0,0}}$$

Figure 4. Sensitivity analyses under Scenario 2. Solid lines represent the desirability weighted selection percentages $\delta$. Dashed lines represent the means of $N_E/N_T$. 
where $S_{0b}$ and $\pi_{a,b}$ are defined by Equations (2) and (3), respectively. When $Y_T$ and $Y_T$ are unknown, the probability of $(Y_E, Y_T)$ is

$$
\text{pr}(Y_E = 1, Y_T = 1|X_E > V, X_T > V) = (\text{pr}(X_E > V, X_T > V|Y_T = 1)\text{pr}(Y_E = 1, Y_T = 1)) \\
\times (\sum_{a=0}^{1} \sum_{b=0}^{1} \text{pr}(X_E > V, X_T > V|Y_E = a, Y_T = b))
$$

when $\pi_{a,b}$ and $\pi_{a,b}$ are defined by Equations (2) and (3), respectively, and $S_{0b} = 1$.

**SUPPLEMENTARY MATERIALS**

Table 1 in supplementary materials summarizes simulation results for the two remained scenarios. Table 2-9 in supplementary materials summarize the sensitivity analysis results conducted for all scenarios by varying each of $\pi_{j,late}(d)^{true}$, logistic difficulty index $\xi$, $N_{max}$, the event time distributions, $K$ = number of sets in the piecewise exponential event time distribution partition, and the association parameter, $\phi$, and survival functions for computing missing outcomes.

[Received December 2012. Revised October 2013.]

**REFERENCES**


