Published in final edited form as:

Clin Trials. 2014 February ; 11(1): 38–48. doi:10.1177/1740774513500589.

A Bayesian Adaptive Phase I-II Clinical Trial for Evaluating Efficacy and Toxicity with Delayed Outcomes

Joseph S. Koopmeiners^{1,3} and Jaime Modiano^{2,3}

¹Division of Biostatistics, University of Minnesota, Minneapolis, Minnesota

²College of Veterinary Medicine, University of Minnesota, St. Paul, Minnesota

³Masonic Cancer Center, University of Minnesota, Minneapolis, MN 55455

Abstract

Background—In traditional phase I oncology trials, the safety of a new chemotherapeutic agent is tested in a dose escalation study to identify the maximum tolerated dose, which is defined as the highest dose with acceptable toxicity. An alternate approach is to jointly model toxicity and efficacy and allow dose finding to be directed by a pre-specified trade-off between efficacy and toxicity. With this goal in mind, several designs have been proposed to jointly model toxicity and efficacy in a phase I-II dose escalation study. A factor limiting the use of these designs is that toxicity and efficacy must be observed in a timely manner.

Purpose—One approach to overcoming this problem is to model toxicity and efficacy as time-to-event outcomes. This would allow new subjects to be enrolled before full information is available for previous subjects while incorporating partial information when adaptively assigning new subjects to a dose level.

Methods—We propose a phase I-II dose escalation study for evaluating toxicity and efficacy with delayed outcomes by jointly modeling toxicity and efficacy as time-to-event outcomes. We apply our proposed design to a phase I-II clinical trial of a novel targeted toxin for canine hemangiosarcoma.

Results—Our simulation results show that our design identifies the optimal dose at a similar rate to dose finding that treats toxicity and efficacy as binary outcomes but with a substantial savings in study duration.

Limitations—Our proposed design has acceptable operating characteristics and dramatically reduces the trial duration compared to a design that considers toxicity and efficacy as binary outcomes but comes at the cost of enrolling additional subjects when all dose levels are unacceptable.

Conclusions—We developed a novel phase I-II design that accounts for delayed outcomes by modeling toxicity and efficacy as time-to-event outcomes. Our design has similar operating characteristics to efficacy/toxicity trade-off designs that consider efficacy and toxicity as binary outcomes but with a dramatically shorter study duration.

Keywords

Bayesian adaptive design; Dose finding; Phase 1 clinical trial

1 Introduction

In traditional phase I oncology trials, the safety of a new chemotherapeutic agent is tested in a dose escalation study to identify the maximum tolerated dose (MTD), which is defined as the highest dose with acceptable toxicity. Phase I dose escalation studies can be broadly classified into two groups: rule-based designs and model based designs. Rule-based designs, of which the 3+3 design is the most common example [1], rely on a simple algorithm to guide dose escalation and identify the MTD. In contrast, model-based designs utilize a simple, typically parametric, model for the probability of toxicity, which guides dose escalation and identifies the MTD. The most common model-based design is the continual reassessment method (CRM) [2]. The CRM uses a one-parameter model for the probability of toxicity and assigns subjects to dose levels based on the current estimate of the maximum tolerated dose using the outcomes for all previous subjects. The CRM has since been adapted to a variety of scenarios to address its shortcomings and meet the needs of clinicians who run phase I oncology trials [3].

An alternate approach is to jointly model toxicity and efficacy and allow dose finding to be directed by a pre-specified trade-off between efficacy and toxicity. With this goal in mind, several phase I-II designs have been proposed to jointly model toxicity and efficacy in phase I-II dose escalation studies. Braun [4] uses a copula model to jointly model toxicity and efficacy. Thall and Cook [5] take a similar approach to jointly modeling efficacy and toxicity and introduce a contour for considering the trade-off between toxicity and efficacy. Zhang et al. [6] propose a trivariate CRM that models toxicity and efficacy as an ordered categorical variable with three levels: no toxicity or efficacy, efficacy without toxicity and toxicity with or without efficacy. A factor limiting the use of these designs is that toxicity and efficacy must be observed in a timely manner. This is particularly problematic for the efficacy outcome, which is often measured over a longer time-frame than the toxicity outcome. In this case, toxicity and efficacy must be modeled as time-to-event outcomes in order to avoid unnecessary delays in enrolling subjects due to the delay in observing outcomes for the previous cohorts

Consider the following motivating example. Researchers at the University of Minnesota School of Veterinary Medicine would like to complete a phase I-II clinical trial of a novel targeted toxin for canine hemangiosarcoma. Ideally, dose finding would be guided by a toxicity/efficacy trade-off because it is anticipated that there is a dose where further escalation will increase toxicity without increasing efficacy. Toxicity will be recorded and graded according to the Veterinary Cooperative Oncology Group (VCOG) criteria [7]. Adverse events are recorded on a 1-5 scale where severe, unacceptable adverse events are graded as 4 and lethal events are graded as 5. Dose limiting toxicities will be defined in this trial as any grade 4 or 5 toxicity in the first 28 days. Clinical response, measured by diagnostic imaging, is usually used in phase I-II designs that consider a toxicity/efficacy trade-off, but imaging of hemangiosarcoma is complicated due to the vascularity of the tumor

and the frequent occurrence of bleeding/clotting episodes that can add “mass” to a tumor, but that is not reflective of tumor growth. Instead, efficacy for this disease is defined as 6-month overall survival, which was chosen for comparison to the median overall survival of six months achievable with the standard of care [8]. This poses an obvious problem. A 6-month delay between cohorts is unacceptably long and would render the trial impractical. Modeling efficacy as a time-to-event outcome would allow additional cohorts to be enrolled during the 6-month follow-up and allow dose finding to be based on all information available at the time of enrollment.

Several approaches have been proposed for adapting the CRM to accommodate time-to-event outcomes for late-onset toxicities. Cheung and Chappell [9] introduce the time-to-event CRM (TITE-CRM), which models the probability of toxicity over time using a weight function and allows partial information to be incorporated into dose finding. Braun [10] takes a cure-rate model approach and shows that the cure-model likelihood is identical to the TITE-CRM likelihood derived by Cheung and Chappell [9]. Thall et al. [11] discuss a “look-ahead” approach and compare this to an approach where dose escalation decisions are based on all available data at subject enrollment. Finally, Bekele et al. [12] use predictive probabilities to control subject accrual in order to avoid excess toxicities in the event of late-onset toxicities.

In this manuscript, we present a phase I-II clinical trial for evaluating efficacy and toxicity with delayed outcomes. We model the time-to-toxicity with a cure-rate model, following the approach of Braun [10], and model the time-to-death with a mixture distribution that results in a similar parameterization to the cure-rate model for toxicity. A joint probability model for the time-to-toxicity and the time-to-death is developed using a copula model. Dose finding based on a trade-off between toxicity and efficacy is completed using the algorithm proposed Thall and Cook [5]. Our simulation results show that our design identifies the optimal dose at a similar rate to dose finding that treats toxicity and efficacy as binary endpoints but with a substantial savings in study duration.

The remainder of our manuscript proceeds as follows. In Section 2, we present a joint model for toxicity and efficacy with delayed outcomes and describe our dose-finding algorithm. In Section 3, we apply our proposed design to a phase I-II clinical trial of a novel targeted toxin for hemangiosarcoma and evaluate the operating characteristics of our design by simulation. Finally, we conclude with a brief discussion in Section 4.

2 Modeling Efficacy and Toxicity with Delayed Outcomes

2.1 Joint Probability Model

In this section, we propose a joint probability model for the time-to-toxicity and the time-to-death. To accomplish this, we specify univariate models for toxicity and death and derive a joint probability model using a copula approach. We begin by specifying a univariate model for toxicity

Let X_T be the time-to-toxicity and let H_T be the horizon for measuring toxicity. In our motivating example, H_T is 28 days and we are interested in estimating the probability of $X_T > H_T$. Following the approach of Braun [10], we assume a cure rate model for X_T ,

$$F_T(x_T|d) = \theta_T(d) F_{T,1}(x_T),$$

where $F_{T,1}(x_T) = 0$ for $x_T = 0$ and $F_{T,1}(x_T) = 1$ for $x_T = H_T$. In a standard cure model, $1 - \theta_T(d)$ is the fraction of the population that is cured. In our case, $\theta_T(d)$ is the probability of experiencing toxicity before H_T at dose level, d . This is our pre-determined parameter of interest and dose escalation will be, in part, based on estimates of $\theta_T(d)$. The relationship between dose and the probability of toxicity is often modeled using a simple, one-parameter model in early phase clinical trials but, in this case, we use a two-parameter logistic regression model,

$$\text{logit}(\theta_T(d)) = \beta_{0,T} + \beta_{1,T}(d - 1).$$

We note that subtracting one from the dose results in $\beta_{0,T}$ equaling the log odds of the probability of toxicity for dose one, which facilitates prior specification for $\beta_{0,T}$. At this point, we choose to leave $F_{T,1}(x_T)$ unspecified and will discuss potential options for parametric forms of $F_{T,1}(x_T)$ below.

Ideally, we would follow the same approach for modeling the time-to-death because the cure model results in a useful parameterization for dose finding. Unfortunately, the cure model is inappropriate for modeling survival in our case. Instead, we take a mixture approach, which results in a similar parameterization that easily facilitates dose finding. Let X_S be time-to-death and let H_S be the horizon for measuring survival (6-months in our motivating example). Our primary interest lies in estimating the probability of $X_S > H_S$. We assume the following mixture model for X_S ,

$$F_S(x_S|d) = \theta_S(d) F_{S,1}(x_S) + (1 - \theta_S(d)) F_{S,2}(x_S),$$

where $F_{S,1}(x_S) = 0$ for $x_S = 0$, $F_{S,1}(x_S) = 1$ for $x_S = H_S$ and $F_{S,2}(x_S) = 0$ for $x_S = H_S$ and $F_{S,2}(x_S) \rightarrow 1$ as $x_S \rightarrow \infty$. In this case, $\theta_S(d)$ is the probability of $X_S \leq H_S$ at dose level d . This distribution can be alternately expressed as a conditional distribution where $1_{\{X_S \leq H_S\}}$ is a Bernoulli random variable with success probability $\theta_S(d)$, X_S follows $F_{S,1}(x_S)$ if $X_S \leq H_S$ and X_S follows $F_{S,2}(x_S)$ if $X_S > H_S$.

The advantage of this mixture approach is that we can directly model the probability of $X_S > H_S$ and base dose finding on current estimates of $\theta_S(d)$. We also model $\theta_S(d)$ using a logistic regression model but this time include a quadratic term,

$$\text{logit}(1 - \theta_s(d)) = \beta_{0,s} + \beta_{1,s}(d - 1) + \beta_{2,s}(d - 1)^2.$$

We include a quadratic term for the probability of 6-month survival to account for the possibility that a dose exists such that further escalation would increase the probability of toxicity but would not result in a survival benefit. We note that our efficacy outcome, $X_S > H_S$, occurs when an event is not observed before H_S , which is why we specify our logistic regression model on $(1 - \theta_S(d))$. This implies that we can not observe our efficacy outcome until a subject has survived beyond H_S . This is in contrast to the typical phase I-II setting where efficacy is treatment response, in which case all information for efficacy is fully observed once response occurs. This does not impact our ability to model efficacy statistically but does result in longer trial durations.

We model the joint behavior of X_T and X_S using the Gumbel Copula [13]. Briefly, a copula model specifies a joint distribution on the unit square, from which a joint distribution of any two random variables can be derived using an inverse transformation. The joint CDF for the Gumbel Copula is,

$$F(x_T, x_S) = F_T(x_T)F_S(x_S) + \kappa F_T(x_T)(1 - F_T(x_T))F_S(x_S)(1 - F_S(x_S)),$$

where κ quantifies the correlation between X_T and X_S . κ takes values between -1 and 1, with $\kappa = 0$ implying independence, $\kappa > 0$ implying positive correlation and $\kappa < 0$ implying negative correlation. The joint PDF for the Gumbel Copula is

$$f(x_T, x_S) = f_T(x_T)f_S(x_S) + \kappa f_T(x_T)(1 - 2F_T(x_T))f_S(x_S)(1 - 2F_S(x_S)),$$

where $f_T(x_T)$ and $f_S(x_S)$ are the PDFs of $F_T(x_T)$ and $F_S(x_S)$, respectively.

Let $(x_{T,1}, x_{S,1}), (x_{T,2}, x_{S,2}), \dots, (x_{T,n}, x_{S,n})$ be pairs of time-to-toxicity and time-to-death, $y_1, y_2, \dots, y_n$ be the length of time each subject has been followed on the study and $d_1, d_2, \dots, d_n$ be the dose-level assigned to each subject. Subjects will be followed until H_S and subjects still alive at H_S will be considered censored. There are four potential likelihood contributions that depend on $x_{T,i}$, $x_{S,i}$ and y_i :

- Scenario 1: $x_{T,i} \leq y_i$ and $x_{S,i} \leq y_i$

$$L_{1,i} = \theta_T(d_i) f_{T,1}(x_{T,i}) \theta_S(d_i) f_{S,1}(x_{S,i}) (1 + \kappa (1 - 2\theta_T(d_i) F_{T,1}(x_{T,i})) (1 - 2\theta_S(d_i) F_{S,1}(x_{S,i})))$$

- Scenario 2: $x_{T,i} > y_i$ and $x_{S,i} \leq y_i$

$$L_{2,i} = (1 - \theta_T(d_i) F_{T,1}(y_i)) \theta_S(d_i) f_{S,1}(x_{S,i}) (1 - \kappa \theta_T(d_i) F_{T,1}(y_i) (1 - 2\theta_S(d_i) F_{S,1}(x_{S,i})))$$

- Scenario 3: $x_{T,i} \leq y_i$ and $x_{S,i} > y_i$

$$L_{3,i} = \theta_T(d_i) f_{T,i}(x_{T,i}) (1 - \theta_S(d_i) F_{S,1}(y_i)) (1 - \kappa (1 - 2\theta_T(d_i) F_{T,1}(x_{T,i})) \theta_S(d_i) F_{S,1}(y_i))$$

- Scenario 4: $x_{T,i} > y_i$ and $x_{S,i} > y_i$

$$L_{4,i} = (1 - \theta_T(d_i) F_{T,1}(y_i)) (1 - \theta_S(d_i) F_{S,i}(y_i)) (1 + \kappa \theta_T(d_i) F_{T,1}(y_i) \theta_S(d_i) F_{S,1}(y_i))$$

The full likelihood can be expressed as:

$$L(\vec{\beta} | \vec{x}_T, \vec{x}_S, \vec{y}) = \prod_{i=1}^n L_{1,i}^{1_{[x_{T,i} \leq y_i, x_{S,i} \leq y_i]}} L_{2,i}^{1_{[x_{T,i} > y_i, x_{S,i} \leq y_i]}} L_{3,i}^{1_{[x_{T,i} \leq y_i, x_{S,i} > y_i]}} L_{4,i}^{1_{[x_{T,i} > y_i, x_{S,i} > y_i]}}.$$

There are several important characteristics of this likelihood that are worth noting. First, by censoring subjects at time H_S , we need not specify the form of $F_{S,2}(x_S)$ as the likelihood does not depend on $F_{S,2}(x_S)$. As a result, the likelihood contribution for toxicity and the likelihood contribution for survival are the same. By specifying a mixture model for X_S , we are able to derive a model that is functionally equivalent to the cure-rate model specified for toxicity but without the unrealistic assumption that a fraction of the population has been cured and are not at-risk. Second, Braun [10] showed that, for the TITE-CRM, $\theta_T(d_i)$ and $f_T(x_{T,i})$ can be factored completely, which implies that the full conditional for $\vec{\beta}$ does not depend on $f_T(x_{T,i})$. That is no longer the case in our joint model unless $\kappa = 0$, implying independence between X_T and X_S . Finally, to this point, we have implicitly assumed that all subjects would survive past H_T and that all subjects would eventually be fully observed for the toxicity outcome. It is also possible that a subject could die before H_T . This is most likely to occur as a result of treatment, in which case the subject would have already experienced toxicity, but it is also possible that the subject could die from the disease or some other cause not related to treatment. In this case, we can simply censor the toxicity endpoint at the time of death and use the likelihood contribution for Scenario 2.

We must specify priors for all model parameters in order to complete a Bayesian analysis. For our motivating example, we assume normal priors for $\beta_{0,T}$ and $\beta_{0,S}$ ($N(-3, sd = 3)$ and $N(-1, 3)$, respectively), gamma(1/4, 1/4) priors for $\beta_{1,T}$ and $\beta_{1,S}$ (corresponding to a mean of 1 and a variance of 4) and a $N(0, 0.5)$ for $\beta_{2,E}$. Our prior distributions for $\beta_{0,T}$, $\beta_{0,S}$, $\beta_{1,T}$ and $\beta_{1,S}$ represent moderately informative priors. In all cases, the prior mean is set equal to our prior expectation for each parameter and the prior variance set large enough to provide support for all reasonable parameter values. Our prior for $\beta_{2,E}$ represents strong prior belief against the inclusion of a quadratic term but should allow enough flexibility to accommodate strong departures from linearity in the logistic regression for 6-month survival. Finally, we put a non-informative, uniform(-1, 1) prior on κ . We will provide simulation results evaluating the operating characteristics of our study using these prior distributions in Section 3. In general, the prior distributions can have a strong impact on the operating characteristics of a Phase I-II study and we recommend that investigators complete a thorough simulation study to evaluate the sensitivity of the operating characteristics to various prior specification before implementing our design.

To this point, we have not specified the form of $F_{T,1}(x_T)$ and $F_{S,1}(x_S)$. We will consider two parametric forms for $F_{T,1}(x_T)$ and $F_{S,1}(x_S)$ and compare their performance in a variety of scenarios through simulation. The first approach is to assume uniform distributions for $F_{T,1}(x_T)$ and $F_{S,1}(x_S)$: $X_T \sim \text{Unif}(0, H_T)$ and $X_S \sim \text{Unif}(0, H_S)$. The weight function investigated by Cheung and Chappell [9] results from assuming a uniform distribution for the time-to-toxicity. The second is to follow the approach of Braun [10] and assume *Beta*

$(a_T, 1)$ and *Beta* $(a_S, 1)$ distributions for $\frac{X_T}{H_T}$ and $\frac{X_S}{H_S}$, respectively. This should provide more flexibility than the uniform distribution, which is particularly important for the case of late-onset toxicity. We place an *Exponential* (1) prior on both a_T and a_S when fitting this model.

2.2 Dose Finding Algorithm

Several approaches to evaluating the trade-off between efficacy and toxicity have been proposed in the literature [4, 14, 5]. Braun [4] proposed a weighted Euclidean distance between the estimated probabilities of efficacy and toxicity and a set of target probabilities. Yin et al. [14] evaluate the trade-off between efficacy and toxicity using odds ratios. Finally, Thall and Cook [5] propose an efficacy/toxicity contour for evaluating the trade-off between efficacy and toxicity. The efficacy/toxicity contour proposed by Thall and Cook [5] is appealing in that it provides an intuitive framework for considering the trade-off between efficacy and toxicity and allows for great flexibility when designing a study. We will follow their approach in designing our study.

Let $\theta_{T,max}$ be the maximum acceptable probability of toxicity assuming all subjects survive past H_S and S,min be the minimum acceptable probability of survival assuming no toxicity. Thall and Cook [5] use a weighted L_p norm to evaluate the toxicity/efficacy trade-off for a dose, d_j ,

$$\delta_j = 1 - \left(\left(\frac{\theta_T(d_j)}{\theta_{T,max}} \right)^p + \left(\frac{1 - \theta_S(d_j)}{1 - \theta_{S,min}} \right)^p \right)^{1/p},$$

where p is determined by identifying a probability of toxicity and probability of survival combination, (θ_T^*, θ_S^*) , that is equally desirable to $(0, \theta_{S,min})$ and $(\theta_{T,max}, 1)$. p is determined by setting

$$\left(\left(\frac{\theta_T^*}{\theta_{T,max}} \right)^p + \left(\frac{1 - \theta_S^*}{1 - \theta_{S,min}} \right)^p \right)^{1/p} = 1$$

and solving for p . In our example, we set $\theta_{T,max} = 0.50$, $\theta_{S,min} = 0.55$ and $(\theta_T^*, \theta_S^*) = (0.40, 0.70)$, which results in $p = 2.27$.

In the context of our trial, we update the posterior whenever a new cohort is ready to be enrolled using all data available at that time. A dose level, d_j , is considered acceptable if the posterior probability of $\delta_j > 0$ exceeds a pre-specified threshold, π_1 ,

$$p(\delta_j > 0 | \vec{y}^s, \vec{y}^T, \vec{z}, \vec{d}) > \pi_1. \quad (1)$$

The study terminates for futility if no dose levels are considered acceptable. Otherwise, the acceptable dose level with maximum δ_j is considered optimal and the next cohort will be treated at that dose level under the restriction that untried dose levels may not be skipped when escalating. The trial continues until the maximum sample size has been reached and the dose with maximum δ_j at study completion is declared the optimal dose.

In conclusion, our proposed dose-finding algorithm is as follows:

1. Treat the first cohort of m patients at the lowest dose level.
2. Update the posterior distribution using all available data when the next cohort is to be enrolled.
3. Identify acceptable dose levels using criteria (1). The trial terminates for futility if no dose levels are acceptable.
4. Calculate δ_j for dose levels $j = 1, \dots, J$.
5. Treat the next cohort at the current estimate of the optimal dose under the restriction that untried dose levels may not be skipped when escalating.
6. The study continues until termination or the maximum sample size is reached. The acceptable dose with maximum δ_j at study completion is declared the optimal dose.

3 Application

We now return to our motivating example and discuss the application of our proposed design to a phase I-II clinical trial of a novel treatment for canine hemangiosarcoma. Hemangiosarcoma (HSA) is a rapidly fatal disease that can arise in dogs of any age. However, it is more often a disease of middle to old age dogs with alarmingly high predilection for some breeds. Unlike other tumors where there have been recent incremental gains in survival, the optimism voiced at the turn of this century regarding HSA treatments [15] now seems premature. More than 50% of dogs with HSA disease still die within 4 to 6 months of diagnosis, and clinical trials using new combinations of old drugs [16], and new drugs [17] have shown no benefit over the standard of care. Recent data suggest that resistance to conventional and targeted therapies may be due, at least in part, to the fact that the tumors include normal cells in an abnormal niche: the hemangiosarcoma cells seem to direct stromal constituents to undergo proliferation and establish a hypoxic and inflammatory tumor microenvironment that allows them to thrive [18]. We believe that attacking the mechanisms that orchestrate these interactions through targeted therapies will improve the outcomes for dogs with this disease.

Researchers at the University of Minnesota College of Veterinary Medicine would like to complete a Phase I-II clinical trial for canine hemangiosarcoma using a new targeted toxin that considers a trade-off between toxicity and efficacy during dose finding. Dose limiting toxicity is defined as any grade 4 or 5 toxicity within the first 28 days and efficacy will be evaluated by considering 6-month survival. The researchers would like to evaluate four dose levels and are limited to a maximum sample size of thirty dogs. In order to apply our proposed design we must specify the following parameters: $\theta_{T,max}$, $\theta_{S,min}$, (θ_T^*, θ_S^*) , π_1 and the cohort size. $\theta_{T,max}$, $\theta_{S,min}$ and (θ_T^*, θ_S^*) were set equal to 0.50, 0.55 and (0.40, 0.70), respectively, resulting in $p = 2.27$ as described in Section 2.2. The cohort size and π_1 were chosen by considering several values and evaluating the impact of these parameters on the operating characteristics of our study. We considered cohort sizes of 2 and 3 and π_1 equal to 0.05 and 0.10. At first glance, 0.05 and 0.10 may appear small for π_1 if π_1 represents the minimum probability that a dose level is acceptable for the trial to continue. It may be best to consider π_1 from the opposite perspective. A dose level is assumed acceptable and is considered unacceptable if the posterior probability that the dose has an unacceptable efficacy/toxicity trade-off is greater than $1 - \pi_1$. Therefore, we would only terminate the trial for futility if the probability that all doses are unacceptable exceeds 0.90 or 0.95.

We completed a small simulation study to evaluate the operating characteristics of our study. We considered several hypothetical scenarios in order to identify design parameters that result in good operating characteristics in a variety of settings. For each scenario, data were simulated assuming that either $F_{T,1}(x_T)$ and $F_{S,1}(x_S)$ were *Uniform* (0, H_T) and *Uniform*

(0, H_S), respectively, or that $\frac{X_T}{H_T}$ and $\frac{X_S}{H_S}$ follow a *Beta* (3, 1) distribution conditional on X_T or X_S , respectively. The latter is the more challenging case where the hazards for X_T and X_S increase over time. The waiting time between subjects was simulated from an exponential distribution with an average waiting time of one, two or four weeks. We anticipate an average waiting time of two weeks between dogs but consider average waiting times of one and four weeks to evaluate the impact of the waiting time on the operating characteristics of our study. It should be noted that the average waiting time between dogs is only relevant relative to the horizon for evaluating the two outcomes. For example, two weeks corresponds to half of the time required to evaluate the toxicity outcome but only 12.6% of the time required to evaluate the survival outcome. Therefore, the three average waiting times being considered differ substantially relative to the time needed to evaluate toxicity but vary only a small amount relative to the time needed to evaluate the survival outcome. In addition, we also considered the setting where toxicity and 6-month survival were treated as binary outcomes using the design proposed by Thall and Cook [5] in combination with a “look-ahead” method. In the “look-ahead” method, we determine if changing the unobserved outcomes would alter the dose assignment for the next cohort and enroll the next cohort immediately if the dose would be the same regardless of the unobserved outcomes. Otherwise, enrollment of the next cohort is delayed until all unobserved outcomes have been observed. The operating characteristics of the different designs were summarized by calculating the simulated probability of selecting each dose or stopping for futility, the average number of subjects treated at each dose level and the average study duration in weeks. 1,000 simulated studies were completed for each scenario.

Table 1 presents the five scenarios considered in our simulation study. For each scenario, we present the probability of DLT, the probability of 6-month survival and the corresponding δ for each dose level. In Scenarios 1 - 3, there is only one acceptable dose, which varies from dose 1 to dose 4, in Scenario 4, all doses are excessively toxic and in Scenario 5, all doses have acceptable toxicity but none have acceptable 6-month survival.

Table 2 presents the simulated operating characteristics of our study assuming $\pi_1 = 0.05$, cohorts of size 3, an average waiting time of 2 weeks and X_T and X_S simulated from uniform distributions conditional on experiencing toxicity or death before H_T or H_S , respectively. Treating efficacy and toxicity as time-to-event outcomes dramatically decreases the duration of the trial with only a moderate decrease in the probability of correctly identifying the correct dose and the average number of subjects treated at the optimal dose. This is true regardless if $F_{T,1}(x_T)$ and $F_{S,1}(x_S)$ are modeled as a uniform or beta distribution. In Scenario 1, the average study duration decreased from 175.8 months for the “look-ahead” method to approximately 80 months for the two time-to-event methods with no change in the probability of correctly identifying the optimal dose or the average number of subjects treated at the optimal dose. We observe a similar decrease in the study duration in Scenarios 2 and 3 with little change in the probability of correctly identifying the optimal dose but a modest decrease in the average number of subjects treated at the optimal dose. In Scenarios 4 and 5, there are no acceptable doses and the correct decision is to stop for futility. In Scenario 4, where all doses are over-toxic, the trial stops for futility almost 100% of the time regardless of method and in Scenario 5, where all doses are safe but do not have acceptable efficacy, the time-to-event methods stop for futility less often than the “look-ahead” method with a more dramatic decrease observed when we assume beta distributions for $F_{T,1}(x_T)$ and $F_{S,1}(x_S)$. The results for Scenarios 4 and 5 illustrate one limitation of the proposed method. Although the proposed design results in shorter trial duration and appropriately terminates for futility at an acceptable rate, the average number of subjects enrolled in the study is actually higher than for the “look-ahead” method because subjects are enrolled immediately upon availability. While we feel that this trade-off is acceptable for our current application, it may not be in all cases and researchers should carefully weigh this trade-off when designing future trials.

Table 3 presents similar results to Table 2 except that $\frac{X_T}{H_T}$ and $\frac{X_S}{H_S}$ are now simulated from *beta* (3,1) distributions conditional on experiencing toxicity or death before H_T or H_S , respectively. This represents the more challenging scenario where the hazards for X_T and X_S increase over time. We expect that specifying a more flexible beta distribution for $F_{T,1}(x_T)$ and $F_{S,1}(x_S)$ would result in better performance in this scenarios. Our results, though, show little difference between the two time-to-event methods and suggest that assuming a uniform distribution for $F_{T,1}(x_T)$ and $F_{S,1}(x_S)$ may provide adequate performance even when the model is misspecified.

Additional simulations to evaluate the impact of varying π_1 , the cohort size and the average waiting time between subjects were also completed. Increasing π_1 to 0.10 results in an increased probability of early termination for futility in Scenarios 4 and 5 but reduces the probability of correctly identifying the optimal dose in Scenarios 1 through 3 (results not

shown). Using cohorts of two resulted in a decrease in the probability of correctly identifying the optimal dose in Scenarios 1 through 3 and increased the probability of terminating for futility in all cases (results not shown). In general, though, using cohorts of size two instead of cohorts of size three resulted in only modest differences in the operating characteristics of our study. Finally, Table 4 illustrates that the average waiting time between subject enrollment has little impact on the probability of correctly identifying the optimal dose but the average number of subjects treated at the optimal dose increases as the waiting time between subjects increases. This is encouraging because it suggests that our study would still have acceptable operating characteristics if enrollment were twice as fast as expected, although, it is still possible that a dramatic increase in the rate of enrollment could negatively impact the operating characteristics of our study.

4 Discussion

In this manuscript, we propose a phase I-II clinical trial for evaluating efficacy and toxicity with delayed outcomes and discuss how our proposed design can be applied to a phase I clinical trial of a novel targeted toxin for canine hemangiosarcoma. We model toxicity using a cure-rate model and take a mixture approach to modeling overall survival. A joint model for the two outcomes was developed using the Gumbel copula. This model allows us to directly model our pre-determined parameters of interest and allows us to implement a dose finding algorithm that is identical to what would be used if efficacy and toxicity were binary outcomes. Our simulation results illustrate that our proposed design dramatically shortens study duration and identifies the optimal dose only slightly less often than a design that considers efficacy and toxicity as binary outcomes.

A limitation of existing phase I-II designs is that efficacy and toxicity must be observed in a timely manner. Our proposed design addresses this limitation by modeling efficacy and toxicity as time-to-event outcomes. This allows new subjects to be enrolled before the outcomes for previous cohorts have been fully observed and uses partial information to assign new subjects to a dose level. In our motivating example, there is a 6-month delay between enrollment and observation of the efficacy outcome but new subjects are expected to ready for enrollment every two weeks. Our proposed design allows this study to be completed in a little over a year and a half. For comparison, the study would take approximately four years to complete if we were forced to wait until full information is available for all subjects before enrolling new cohorts and it is unlikely that the trial would be run at all if that were the case.

Our proposed design includes a stopping rule for futility if at any point in the study there is substantial evidence that none of the dose levels are acceptable. We could also include a stopping rule that stops the study if it is clear that we have identified the optimal dose before reaching the maximum sample size. Stopping for efficacy is quite common in phase II and phase III clinical trials and results in a substantial reduction in sample size. The sample size of phase I clinical trials is often quite small (usually no more than 30 or 40) and it is unlikely that stopping early for efficacy would save more than a few patients. Furthermore, a more precise estimate of the probability of efficacy and toxicity for the new drug could be very helpful when design future phase II clinical trials. Considering the limited benefit of early

termination and the benefit of a more thorough understanding of the new drug, we chose not to include a stopping rule for efficacy.

Our proposed design has acceptable operating characteristics assuming an average waiting of one week between subjects but it is possible that a much faster enrollment rate could result in a decreased probability of correctly identifying the optimal dose. Polley [19] recently investigated several approaches to modifying the TITE CRM to accommodate fast accrual. These methods could easily be adapted to our setting to protect against fast accrual. Further work is needed to assure that implementing these methods does not negate the advantage of our proposed methods (i.e. that subjects can be enrolled as they become available) and to determine how these methods would be implemented in practice.

Acknowledgments

This work was supported by the University of Minnesota Cancer Center Support Grant [NIH P30 CA077598].

References

1. Storer, Barry E. Design and analysis of phase I clinical trials. *Biometrics*. 1989; 45(3):925–937. [PubMed: 2790129]
2. O’Quigley, John; Pepe, Margaret; Fisher, Lloyd. Continual reassessment method: A practical design for phase I clinical trials in cancer. *Biometrics*. 1990; 46(1):33–48. [PubMed: 2350571]
3. Goodman, Steven N.; Zahurak, Marianna L.; Piantadosi, Steven. Some practical improvements in the continual reassessment method for phase I studies. *Statistics in Medicine*. 1995; 14(11):1149–1161. [PubMed: 7667557]
4. Braun, Thomas M. The bivariate continual reassessment method: extending the CRM to phase I trials of two competing outcomes. *Controlled Clinical Trials*. 2002; 23(3):240–256. [PubMed: 12057877]
5. Thall, Peter F.; Cook, John D. Dose-finding based on efficacy/toxicity trade-offs. *Biometrics*. 2004; 60(3):684–693. [PubMed: 15339291]
6. Zhang, Wei; Sargent, Daniel J.; Mandrekar, Sumithra. An adaptive dose-finding design incorporating both toxicity and efficacy. *Statistics in Medicine*. 2006; 25(14):2365–2383. [PubMed: 16220478]
7. Vail, David M. Veterinary co-operative oncology group common terminology criteria for adverse events (vcog-ctcae) following chemotherapy or biological antineoplastic therapy in dogs and cats v1.0. *Veterinary and Comparative Oncology*. 2004; 2(4):195–213. [PubMed: 19379294]
8. Helfand, Stuart C. Canine Hemangiosarcoma: A Tumor of Contemporary Interest. *Cancer Therapy*. 2008; 6:457–462.
9. Cheung, Ying Kuen; Chappell, Rick. Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics*. 2000; 56(4):1177–1182. [PubMed: 11129476]
10. Braun, Thomas M. Generalizing the TITE-CRM to adapt for early- and late-onset toxicities. *Statistics in Medicine*. 2006; 25(12):2071–2083. [PubMed: 16217853]
11. Thall, Peter F.; Lee, J Jack; Tseng, Chi-Hong; Estey, Elihu H. Accrual strategies for phase I trials with delayed patient outcome. *Statistics in Medicine*. 1999; 18(10):1155–1169. [PubMed: 10363337]
12. Bekele, B Nebiyu; Ji, Yuan; Shen, Yu; Thall, Peter F. Monitoring late-onset toxicities in phase I trials using predicted risks. *Biostatistics*. 2008; 9(3):442–457. [PubMed: 18084008]
13. Murtaugh, Paul A.; Fisher, Lloyd D. Bivariate binary models of efficacy and toxicity in dose-ranging trials. *Communications in Statistics - Theory and Methods*. 1990; 19(6):2003–2020.
14. Yin, Guosheng; Li, Yisheng; Ji, Yuan. Bayesian dose-finding in phase I/II clinical trials using toxicity and efficacy odds ratios. *Biometrics*. 2006; 62(3):777–787. [PubMed: 16984320]

15. Clifford, Craig A.; Mackin, Andrew J.; Henry, Carolyn J. Treatment of canine hemangiosarcoma: 2000 and beyond. *Journal of Veterinary Internal Medicine*. 2000; 14(5):479–485. [PubMed: 11012108]
16. Lana, Susan; U'ren, Lance; Plaza, Susan; Elmslie, Robyn; Gustafson, Daniel; Morley, Paul; Dow, Steven. Continuous low-dose oral chemotherapy for adjuvant therapy of splenic hemangiosarcoma in dogs. *Journal of Veterinary Internal Medicine*. 2007; 21(4):764–769. [PubMed: 17708397]
17. Chon E, McCartan L, Kubicek LN, Vail DM. Safety evaluation of combination toceranib phosphate (Palladia) and piroxicam in tumour-bearing dogs (excluding mast cell tumours): a phase I dose-finding study. *Veterinary and Comparative Oncology*. 2011 pages no–no.
18. Tamburini BA, Phang TL, Fosmire SP, Scott MC, Trapp SC, Duckett MM, Robinson SR, Slansky JE, Sharkey LC, Cutter GR, Wojcieszyn JW, Bellgrau D, Gemmill RM, Hunter LE, Modiano JF. Gene expression profiling identifies inflammation and angiogenesis as distinguishing features of canine hemangiosarcoma. *BMC Cancer*. 2010; 10:619. [PubMed: 21062482]
19. Polley, Mei-Yin C. Practical modifications to the time-to-event continual reassessment method for phase I cancer trials with fast patient accrual and late-onset toxicities. *Statistics in Medicine*. 30(17):2130–2143. 2011. [PubMed: 21590790]

Table 1

Scenarios considered in our simulation studies. Presented for each scenario are the probability of DLT within 28 days, the probability of 6-month survival and the corresponding δ for each dose level.

Scenario 1				
	Dose 1	Dose 2	Dose 3	Dose 4
P(Toxicity)	0.05	0.12	0.27	0.50
P(Efficacy)	0.38	0.55	0.71	0.83
δ	-0.38	-0.22	0.19	-0.05
Scenario 2				
	Dose 1	Dose 2	Dose 3	Dose 4
P(Toxicity)	0.38	0.52	0.67	0.79
P(Efficacy)	0.77	0.82	0.86	0.89
δ	0.12	-0.09	-0.36	-0.59
Scenario 3				
	Dose 1	Dose 2	Dose 3	Dose 4
P(Toxicity)	0.02	0.07	0.15	0.31
P(Efficacy)	0.12	0.25	0.45	0.67
δ	-0.96	-0.67	-0.24	0.08
Scenario 4				
	Dose 1	Dose 2	Dose 3	Dose 4
P(Toxicity)	0.62	0.75	0.85	0.91
P(Efficacy)	0.27	0.40	0.55	0.69
δ	-0.96	-0.93	-0.91	-0.91
Scenario 5				
	Dose 1	Dose 2	Dose 3	Dose 4
P(Toxicity)	0.03	0.08	0.18	0.38
P(Efficacy)	0.18	0.25	0.33	0.43
δ	-0.82	-0.67	-0.51	-0.43

Table 2

Simulated selection probability and average number of subjects treated at each dose with $\pi_1 = 0.05$, a cohort size of 3, an average waiting time two weeks and X_7 and X_5 simulated from uniform distributions. 1000 simulations were completed for each scenario.

Scenario 1						
Method	Futility	Dose 1	Dose 2	Dose 3	Dose 4	Duration (weeks)
Uniform						
Selection Probability	0.04	0.04	0.19	0.50	0.24	80.6
Avg # Pts Treated		5.13	8.48	10.36	5.40	
Beta						
Selection Probability	0.01	0.04	0.20	0.51	0.24	81.2
Avg # Pts Treated		4.79	9.29	11.51	4.37	
Look Ahead						
Selection Probability	0.03	0.06	0.22	0.48	0.21	177.7
Avg # Pts Treated		5.11	9.26	10.22	5.01	
Scenario 2						
Method	Futility	Dose 1	Dose 2	Dose 3	Dose 4	Duration (weeks)
Uniform						
Selection Probability	0.13	0.70	0.16	0.02	0.0	76.8
Avg # Pts Treated		18.8	7.07	1.30	0.20	
Beta						
Selection Probability	0.08	0.76	0.14	0.02	0.0	78.9
Avg # Pts Treated		20.94	6.51	1.09	0.08	
Look Ahead						
Selection Probability	0.12	0.74	0.13	0.01	0.0	157.8
Avg # Pts Treated		22.22	4.78	0.50	0.03	
Scenario 3						
Method	Futility	Dose 1	Dose 2	Dose 3	Dose 4	Duration (weeks)
Uniform						
Selection Probability	0.16	0.0	0.01	0.06	0.77	76.9
Avg # Pts Treated		3.86	4.26	5.03	14.82	
Beta						
Selection Probability	0.04	0.0	0.02	0.05	0.88	81.0
Avg # Pts Treated		3.64	4.52	6.32	15.31	
Look Ahead						
Selection Probability	0.12	0.0	0.0	0.06	0.82	153.3
Avg # Pts Treated		3.30	3.50	4.66	16.86	
Scenario 4						
Method	Futility	Dose 1	Dose 2	Dose 3	Dose 4	Duration (weeks)

Method		Futility	Dose 1	Dose 2	Dose 3	Dose 4	
Uniform	Selection Probability	1.0	0.0	0.0	0.0	0.0	38.7
	Avg # Pts Treated		5.66	2.81	1.09	0.35	
Beta	Selection Probability	1.0	0.0	0.0	0.0	0.0	43.5
	Avg # Pts Treated		7.20	3.52	1.49	0.34	
Look Ahead	Selection Probability	1.0	0.0	0.0	0.0	0.0	58.8
	Avg # Pts Treated		3.99	2.40	1.19	0.38	

Scenario 5							
Method		Futility	Dose 1	Dose 2	Dose 3	Dose 4	Duration (weeks)
Uniform	Selection Probability	0.62	0.01	0.02	0.04	0.3	68
	Avg # Pts Treated		4.77	4.63	4.72	10.36	
Beta	Selection Probability	0.50	0.03	0.06	0.07	0.35	77.5
	Avg # Pts Treated		4.66	5.45	6.45	12.46	
Look Ahead	Selection Probability	0.69	0.0	0.02	0.04	0.25	130.2
	Avg # Pts Treated		3.93	4.24	4.22	9.86	

Table 3

Simulated selection probability and average number of subjects treated at each dose with $\pi_1 = 0.05$, a cohort size of 3, an average waiting time two weeks and X_7 and X_5 simulated from beta distributions. 1000 simulations were completed for each scenario.

Scenario 1						
Method	Futility	Dose 1	Dose 2	Dose 3	Dose 4	Duration (weeks)
Uniform	0.01	0.04	0.20	0.52	0.24	81.3
				10.81	3.40	
Beta	0.0	0.04	0.19	0.53	0.24	81.4
				11.72	2.97	
Look Ahead	0.03	0.04	0.24	0.47	0.22	191.1
				10.23	5.02	
Scenario 2						
Method	Futility	Dose 1	Dose 2	Dose 3	Dose 4	Duration (weeks)
Uniform	0.06	0.79	0.14	0.01	0.0	79.2
		21.86	6.28	0.67	0.03	
Beta	0.06	0.81	0.12	0.01	0.0	79.0
		22.15	5.92	0.75	0.01	
Look Ahead	0.11	0.76	0.12	0.01	0.0	166.4
		22.57	4.54	0.54	0.07	
Scenario 3						
Method	Futility	Dose 1	Dose 2	Dose 3	Dose 4	Duration (weeks)
Uniform	0.04	0.0	0.01	0.07	0.88	80.7
		3.69	5.98	7.03	13.17	
Beta	0.04	0.0	0.01	0.06	0.89	81.7
		3.66	5.44	7.70	13.15	
Look Ahead	0.12	0.0	0.01	0.05	0.82	174.2
		3.33	3.62	4.66	16.81	
Scenario 4						
Method	Futility	Dose 1	Dose 2	Dose 3	Dose 4	Duration (weeks)
Uniform	0.04	0.0	0.01	0.07	0.88	80.7
		3.69	5.98	7.03	13.17	
Beta	0.04	0.0	0.01	0.06	0.89	81.7
		3.66	5.44	7.70	13.15	
Look Ahead	0.12	0.0	0.01	0.05	0.82	174.2
		3.33	3.62	4.66	16.81	

Method		Futility	Dose 1	Dose 2	Dose 3	Dose 4	
Uniform	Selection Probability	0.99	0.0	0.0	0.01	0.01	49.3
	Avg # Pts Treated		9.47	3.75	1.28	0.28	
Beta	Selection Probability	1.0	0.0	0.0	0.0	0.0	50.6
	Avg # Pts Treated		10.11	3.95	1.28	0.28	
Look Ahead	Selection Probability	1.0	0.0	0.0	0.0	0.0	62.7
	Avg # Pts Treated		3.83	2.50	1.24	0.44	

Scenario 5							
Method		Futility	Dose 1	Dose 2	Dose 3	Dose 4	Duration (weeks)
Uniform	Selection Probability	0.42	0.02	0.05	0.07	0.43	78.8
	Avg # Pts Treated		4.37	6.51	6.88	11.31	
Beta	Selection Probability	0.42	0.03	0.06	0.06	0.43	80.1
	Avg # Pts Treated		4.12	6.15	8.07	11.33	
Look Ahead	Selection Probability	0.71	0.01	0.02	0.04	0.22	147.3
	Avg # Pts Treated		3.99	4.28	4.27	9.25	

Table 4

Simulated selection probability and average number of subjects treated at each dose when the average waiting time (WT) between subject enrollment is varied from one week to four week assuming $\pi_1 = 0.05$, cohorts of size 3 and X_T and X_S simulated from uniform distributions. 1000 simulations were completed for each scenario.

Scenario 1							
Method	Futility	Dose 1	Dose 2	Dose 3	Dose 4	Duration (weeks)	
1 Week WT	Selection Probability	0.05	0.02	0.15	0.54	0.24	51.8
	Avg # Pts Treated		5.04	8.23	10.17	5.92	
2 Weeks WT	Selection Probability	0.04	0.04	0.19	0.50	0.24	80.6
	Avg # Pts Treated		5.13	8.48	10.36	5.40	
4 Weeks WT	Selection Probability	0.02	0.05	0.22	0.48	0.22	137.0
	Avg # Pts Treated		5.17	8.93	10.69	4.82	
Scenario 2							
Method	Futility	Dose 1	Dose 2	Dose 3	Dose 4	Duration (weeks)	
1 Week WT	Selection Probability	0.14	0.71	0.12	0.02	0.0	50.1
	Avg # Pts Treated		17.7	7.36	1.90	0.40	
2 Weeks WT	Selection Probability	0.13	0.70	0.16	0.02	0	76.8
	Avg # Pts Treated		18.8	7.07	1.30	0.20	
4 Weeks WT	Selection Probability	0.10	0.77	0.11	0.02	0	130.0
	Avg # Pts Treated		21.52	5.57	0.72	0.05	
Scenario 3							
Method	Futility	Dose 1	Dose 2	Dose 3	Dose 4	Duration (weeks)	
1 Week WT	Selection Probability	0.19	0.0	0.0	0.06	0.75	50.3
	Avg # Pts Treated		4.32	4.31	5.43	14.04	
2 Weeks WT	Selection Probability	0.16	0.0	0.01	0.06	0.77	76.9
	Avg # Pts Treated		3.86	4.26	5.03	14.82	
4 Weeks WT	Selection Probability	0.15	0.0	0.01	0.05	0.79	130.5
	Avg # Pts Treated		3.53	3.74	4.73	15.96	

Scenario 4						
Method	Futility	Dose 1	Dose 2	Dose 3	Dose 4	Duration (weeks)
1 Week WT	Selection Probability Avg # Pts Treated	0.0 6.65	0.0 3.28	0.0 1.42	0.0 0.58	32.9
2 Weeks WT	Selection Probability Avg # Pts Treated	1.0 5.66	0.0 2.81	0.0 1.09	0.0 0.35	38.7
4 Weeks WT	Selection Probability Avg # Pts Treated	1.0 4.81	0.0 2.50	0.0 1.18	0.0 0.37	50.5
Scenario 5						
Method	Futility	Dose 1	Dose 2	Dose 3	Dose 4	Duration (weeks)
1 Week WT	Selection Probability Avg # Pts Treated	0.57 5.40	0.02 5.18	0.04 5.40	0.31 10.06	47.3
2 Weeks WT	Selection Probability Avg # Pts Treated	0.62 4.77	0.01 4.63	0.02 4.72	0.30 10.36	68.0
4 Weeks WT	Selection Probability Avg # Pts Treated	0.64 4.61	0.01 4.46	0.03 4.36	0.27 10.06	110.1