*University of Michigan School of Public Health*

The University of Michigan Department of Biostatistics Working Paper Series

*Year* 2008 *Paper* 80

# Parametric Non-Mixture Cure Models for Schedule-Finding of Therapeutic Agents

Thomas M. Braun* Changying A. Liu†

*University of Michigan, tombraun@umich.edu

†University of Michigan, liuchang@umich.edu

# Parametric Non-Mixture Cure Models for Schedule-Finding of Therapeutic Agents

Thomas M. Braun and Changying A. Liu

**Abstract**

We propose a Phase I clinical trial design that seeks to determine the cumulative safety of a series of administrations of a fixed dose of an investigational agent. In contrast to traditional Phase I trials that are designed to solely find the maximum tolerated dose (MTD) of the agent, our design instead identifies a maximum tolerated schedule (MTS) that includes an MTD as well as a vector of recommended administration times. Our model is based upon a non-mixture cure model that constrains the probability of toxicity for all subjects to monotonically increase with both dose and the number of administrations received. We assume a specific parametric hazard function for each administration and compute the total hazard of toxicity for a schedule as a sum of individual administration hazards. Throughout a variety of settings motivated by an actual study in allogeneic bone marrow transplant recipients, we demonstrate that our approach has excellent operating characteristics and performs as well as the only other currently published design for schedule-finding studies. We also present arguments for the preference of our non-mixture cure model over the existing model.

# Parametric Non-Mixture Cure Models for Schedule-Finding of Therapeutic Agents

**Changying A. Liu**

**and**

**Thomas M. Braun**

University of Michigan

Department of Biostatistics

Ann Arbor, MI 48109

November 29, 2007

SUMMARY

We propose a Phase I clinical trial design that seeks to determine the cumulative safety of a series of administrations of a fixed dose of an investigational agent. In contrast to traditional Phase I trials that are designed to solely find the maximum tolerated dose (MTD) of the agent, our design instead identifies a maximum tolerated schedule (MTS) that includes an MTD as well as a vector of recommended administration times. Our model is based upon a non-mixture cure model that constrains the probability of toxicity for all subjects to monotonically increase with both dose and the number of administrations received. We assume a specific parametric hazard function for each administration and compute the total hazard of toxicity for a schedule as a sum of individual administration hazards. Throughout a variety of settings motivated by an actual study in allogeneic bone marrow

transplant recipients, we demonstrate that our approach has excellent operating characteristics and performs as well as the only other currently published design for schedule-finding studies. We also present arguments for the preference of our non-mixture cure model over the existing model.

KEY WORDS: Phase I trial, dose-finding study, adaptive design, Bayesian statistics, Weibull distribution

## 1. Introduction

Conventional Phase I clinical trials have been designed for the sole purpose of identifying a maximum tolerated dose (MTD) based upon a single administration. However, in many clinical settings, the agent will be given repeatedly over a sequence of administrations and patients will be followed to assess the cumulative safety of the agent. As a result, treatment is two dimensional, consisting not only of the dose given at each administration, but also the timing of each administration.

Braun et al. (2003) applied the Time-to-Event Continual Reassessment Method (TITE-CRM) of Cheung and Chappell (2000) by attempting to model each schedule of administrations as a "dose." Due to the limitations of this approach, Braun et al. (2005) presented the first design specifically created to determine a maximum tolerated schedule (MTS) instead of a traditional MTD. In their approach, the authors chose to model separately the hazards of toxicity for the individual administrations in a schedule. It is reasonable to assume that the hazard of toxicity for many cytotoxic agents increases steadily after administration, reaches a peak, and then begins to decrease as the agent is cleared from the patient. For simplicity,

Braun et al. (2005) adopted a piecewise linear model

$$h(u \mid \boldsymbol{\theta}) = \begin{cases} \theta_2 \, \frac{u}{\theta_1} & 0 \leq u \leq \theta_1 \\ \theta_2 \, \frac{\theta_3 - u}{\theta_3 - \theta_1} & \theta_1 < u \leq \theta_3 \\ 0 & u > \theta_3 \text{ or } u < 0 \end{cases} \tag{1}$$

in which $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$, with $\theta_1$ the time at which $h(u \mid \boldsymbol{\theta})$ reaches its maximum, $\theta_2$ the maximum hazard, and $\theta_3$ the time when the hazard vanishes to zero. Using this hazard for each administration, the total hazard for a series of administrations was modeled simply as the sum of the hazards of each administration. Given the timing of each administration, the total hazard implies a cumulative hazard and cumulative probability for toxicity at any point in time. Thus, the outcome of interest in this design is the time to toxicity, rather than a binary indicator of toxicity ever occurring. Subject accrual, Bayesian estimation, and outcome-adaptive decision-making are done in a sequential fashion as in classical Phase I trial designs.

Even though Braun et al. (2005) demonstrated the triangle hazard model performed well in the scenarios they investigated, there are several aspects of this hazard model that could be improved. First, none of the parameters in $\boldsymbol{\theta}$ have interpretations that relate directly to the overall probability of toxicity. Second, the hazard function in Equation (1) is not smooth due to the abrupt change in slope at $\theta_1$, i.e $\theta_1$ is a change-point for the hazard, and maximum likelihood estimation of $\theta_1$ is difficult (Liu, 2007). Third, the hazard is truncated at $\theta_3$, which conflicts with the usual approach in time-to-event modeling of assuming an infinite support of toxicity times for the hazard. Finding a maximum likelihood estimate (MLE) of $\theta_3$ is also computationally challenging as truncation parameters do not share the regularity characteristics common to most parameters. Fourth, the model is

inflexible for the inclusion of subject-level or administration-level covariates, i.e. if the dose were to vary with each administration. Although the triangular hazard was generalized to incorporate varying doses, the approach used required a different value of $\boldsymbol{\theta}$ for each dose, forcing the total number of parameters to increase rapidly as more doses are studied (Braun et al., 2007). Although one could conceive of a regression model for each parameter in $\boldsymbol{\theta}$ as a function of subject-level covariates, the direct relationship of each covariate to the probability of toxicity would be difficult or impossible to discern.

As an alternative to the method of Braun et al. (2005), we propose a parametric non-mixture cure model for determining the MTS. Through this approach, we model the cure fraction and marginal probability of toxicity directly as a function of the number of administrations. Most importantly, this model includes parameters whose values can be directly interpreted in relationship to the probability of toxicity. We continue with the approach of Braun et al. (2005) and model the total hazard of toxicity as a sum of single administration hazards. However, due to the limitations described earlier, we instead choose to model the hazard for each administration to be proportional to a standard two-parameter Weibull density function. Given the flexibility of the shape of this density, our model can accommodate a variety of hazard functions beyond the "up-and-down" triangular hazard if so desired. In Section 2, we describe basic notation and develop our parametric cure model. Section 3 contains a method for developing prior distributions for our parameters, and Section 4 contains a specific outline of trial conduct using our design. Section 5 describes the performance of our algorithm and compares it to the approach of Braun et al. (2005). Section 6 contains concluding remarks.

## 2. Notation and Cure Model Development

### 2.1 *Notation*

We have $K$ treatment schedules, $\boldsymbol{s}^{(1)}, \boldsymbol{s}^{2}, \ldots, \boldsymbol{s}^{(K)}$ under investigation in a trial. Schedule $k$ has a total of $m_k$ administrations and $\boldsymbol{s}^{(k)}$ is a vector of planned administration times, i.e. $\boldsymbol{s}^{(k)}$=$(s_1, s_2, ..., s_{m_k})$. Furthermore, $\boldsymbol{s}^{(k)}$ is nested in $\boldsymbol{s}^{(k+1)}$ for each $k = 1, ..., K-1$, so that the duration of a treatment schedule increases with $k$ and $m_1 < m_2 < ... < m_K$. For example, if the first schedule was comprised of three weekly administrations after enrollment and the second schedule was a three-week extension of the first schedule, $\boldsymbol{s}^{(1)} = \{0, 7, 14\}$ and $\boldsymbol{s}^{(2)} = \{0, 7, 14, 21, 28, 35\} = \{\boldsymbol{s}^{(1)}, \boldsymbol{s}^{(1)} + 21\}$. Let $\omega$ denote the fixed maximum length of follow-up for each subject, defined by the medical investigators as a clinically meaningful endpoint that is also late enough to accommodate the longest schedule, $\boldsymbol{s}^{(K)}$. For example, in toxicity studies of possible treatments for acute graft-versus host disease (aGVHD) in allogeneic bone marrow transplant patients, $\omega$ is often set at $100$ days, as this is the the duration of time required for aGVHD to develop. A fixed target probability $p_\omega$ is elicited from investigators and is defined as the target cumulative probability of toxicity by time $\omega$.

A maximum of $N$ subjects will be enrolled in the study, each observed to the earlier of toxicity or completing $\omega$ days of follow-up without toxicity. We assume that entry of the first subject coincides with day $t_1^* = 0$ of the trial, while $t_j^*$ denotes the duration of the study when each successive subject $j = 2, 3, \ldots, N$ is enrolled. The shortest schedule is assigned to the first subject, and the schedule assigned to each successive subject $j$ is determined through an interim analysis at time $t_j^*$ using the data from previously enrolled subjects $i = 1, 2, \ldots, j-1$. We let $T_i$ denote the actual, possibly unobserved, time after enrollment for subject $i$.

At interim evaluation time $t_j^*$, we let $U_i$ denote the length of follow-up for subject $i, i < j$, i.e. $U_i = \min(T_i, t_j^* - t_i^*)$, and $C_i$ is the indicator of whether or not $U_i$ corresponds to a toxicity i.e.

$$C_i = \begin{cases} 1 & ; \quad T_i = U_i, \\ 0 & ; \quad T_i \neq U_i. \end{cases}$$

We emphasize that the study is designed ideally so that each subject will receive a series of administrations at specific, fixed administration times as defined by $\boldsymbol{s}^{(1)}, ..., \boldsymbol{s}^{(K)}$. However, in practice, subjects may deviate from their assigned schedule after they begin treatment for a host of reasons, including clinician error or delay of treatment for medical reasons unrelated to toxicity, such as infection. Therefore, we let $\boldsymbol{s}_i = \{s_{i,1}, ..., s_{i,m_i}\}$ denote the actual times after enrollment at which subject $i$ receives an administration, where $s_{i,1}$ coincides with enrollment and $m_i$ is the number of administrations received. Although $m_k$ administrations are planned for subjects assigned to schedule $\boldsymbol{s}^{(k)}$, it may be the case that $m_i < m_k$ either due to administrative censoring or because subject $i$ had toxicity at time $s_{i,m_i}$ and thus received no further administrations.

### 2.2 *Model Development*

Consider a parametric probability density function (pdf) $f(u \mid \boldsymbol{\phi})$ consisting of parameters $\boldsymbol{\phi}$, with cumulative distribution function $F(u \mid \boldsymbol{\phi})$. We then model the hazard of toxicity for each administration using a scaled form of the density $h(u \mid \boldsymbol{\phi}, \theta) = \theta f(u \mid \boldsymbol{\phi})$, where $\theta > 0$ is an additional parameter quantifying the proportionality of $h(\cdot)$ to $f(\cdot)$. As a result, each administration has a survival function equal to

$$S(u \mid \boldsymbol{\phi}, \theta) = \exp\{-\theta F(u \mid \boldsymbol{\phi})\} \tag{2}$$

that is improper in the sense that $S(\infty) = \exp\{-\theta\} > 0$. Such a model is known as a non-mixture cure model with cure rate $\exp\{-\theta\}$; see Chen et al. (1999) and Tsodikov et al. (2003) for detailed descriptions. In the context of Phase I trials, the probability of toxicity is $1 - \exp\{-\theta\}$.

We emphasize that the concept of a cure fraction is appealing in Phase I trials, as we expect a proportion of subjects to never experience toxicity, even with infinite follow-up. Unlike our proposed model, the cure fraction was not directly parameterized in Braun et al. (2005) and was instead implied through the finite duration of the hazard $\theta_3$ in Equation (1). Another benefit of our approach over that in Braun et al. (2005) is that the single-administration hazard $h(u \mid \boldsymbol{\phi}, \theta)$ can be quite general, provided that it is biologically plausible and is sufficiently tractable for parameter estimation. Most importantly, the hazard is not restricted to have an up-and-down pattern nor have finite support like that in Braun et al. (2005).

Thus, we recommend the hazard be based upon a two-parameter Weibull density

$$h(u \mid \boldsymbol{\phi}, \theta) \;=\; \theta \exp(-\gamma)\alpha u^{\alpha-1} \exp[-u^{\alpha} \exp(-\gamma)] \tag{3}$$

in which $\boldsymbol{\phi} = (\alpha, \gamma)$, $\alpha > 0$ and $-\infty < \gamma < \infty$. This function implies a non-montonic hazard when $\alpha \geq 2$, and also allows for a monotonically decreasing hazard when $0 \leq \alpha < 2$, which may be suitable when the agent is given as a bolus with immediate exposure to maximum toxicity that then decreases over time. In our application, we choose to specifically constrain $\alpha \geq 2$ so that the resulting hazard function has our desired non-monotonic pattern. Figure 1 displays an example of such a hazard for $\alpha = 2, \gamma = 4$, and $\theta = 0.3$ and demonstrates that although an infinite support exists for the time to toxicity, the hazard basically

vanishes 20 days after administration.

[Figure 1 about here.]

We assume that the form of $h(\cdot)$ is unchanged among administrations and that all administrations have an additive cumulative effect on toxicity. Based upon these assumptions, we generate the total and cumulative hazards of toxicity at a general study time $t^*$ for a subject treated with schedule $s^{(k)}$ and follow-up $U$ as

$$h(U \mid \boldsymbol{\phi}, \theta, \boldsymbol{s}^{(k)}) = \theta(m_k) \sum_{\ell=1}^{m_k} f(U - s_\ell \mid \boldsymbol{\phi})/m_k \quad (4)$$

$$H(U \mid \boldsymbol{\phi}, \theta, \boldsymbol{s}^{(k)}) = \theta(m_k) \sum_{\ell=1}^{m_k} F(U - s_\ell \mid \boldsymbol{\phi})/m_k, \quad (5)$$

and corresponding survival function

$$S(U \mid \boldsymbol{\phi}, \theta, \boldsymbol{s}^{(k)}) = \exp\left[-\theta(m_k) \sum_{\ell=1}^{m_k} F(U - s_\ell \mid \boldsymbol{\phi})/m_k\right]. \quad (6)$$

In Equations (4)-(6), the parameter $\theta$ is now generalized to be a function of the number of administrations, which is necessary so that the limiting cumulative probability of toxicity changes with the number of administrations and is equal to $1 - \exp\{-\theta(m_k)\}$. We have also expressed Equations (4) and (5) as an average of hazards among administrations; if scaling by $m_k$ were not done, then we would have $S_k(\infty) = \exp\{-m_k\theta(m_k)\}$. Instead, scaling by $m_k$ maintains $\exp\{-\theta(m_k)\}$ as the cure rate for schedule $k$.

In order to force the cumulative probability of toxicity to increase with the number of administrations, we adopt the regression model $\log\{\theta(m_k)\} = \beta_0 + \beta_1 \log(m_k)$, $\beta_1 \geq 0$, leading to a cure fraction equal to $\exp\{-\exp[\beta_0 + \beta_1 \log(m_k)]\}$. This model also lends to a simple interpretation of $\beta_0$ as quantifying a single adminstration's impact on the cure fraction.

Note that our model assumes that the limiting cumulative probability of toxicity, or conversely the cure fraction, is the same for two subjects receiving the same number of administrations, regardless of when they receive those administrations. The administration times only impact the rate at which the limit is reached. Other models could certainly be adopted that allow the cure fraction to vary by the administration times. Most importantly, our model is much more flexible than that of Braun et al. (2005) and allows for the inclusion of additional covariates both in the cure fraction, as well as the hazard function. For example, if dose were to vary among the subjects, we could directly incorporate dose effects either in the cure fraction or the hazard function using standard regression models.

For the interim analysis at $t^*$, subject $i$ contributes three pieces of data: (1) $U_i$, the length of follow-up, (2) $C_i$, an indicator of toxicity, and (3) $\boldsymbol{s}_i$, a vector of $m_i$ administration times. We let $\boldsymbol{D}_i^* = \{U_i, C_i, \boldsymbol{s}_i\}$, so that the total data collected on all $n^*$ subjects at $t^*$ is $\boldsymbol{D}^* = \{\boldsymbol{D}_1^*, \boldsymbol{D}_2^*, \ldots, \boldsymbol{D}_{n^*}^*\}$. Using this data, the interim likelihood for our parameters is

$$\mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\beta} \mid \boldsymbol{D}^*) \;=\; \prod_{i=1}^{n^*} \{h(U_i \mid \boldsymbol{\phi}, \beta_0, \beta_1, \boldsymbol{s}_i)\}^{C_i} S(U_i \mid \boldsymbol{\phi}, \beta_0, \beta_1, \boldsymbol{s}_i) \qquad (7)$$

where $\boldsymbol{\beta} = \{\beta_0, \beta_1\}$ and the survivor and hazard functions are defined in Equations (4) and (6), respectively.

## 3. Developing Priors for $\boldsymbol{\phi}$ and $\boldsymbol{\beta}$

Since the sample sizes of most Phase I trials are small and very little information is available at the beginning of a trial, we choose to use sequential Bayesian methods to find updated estimates of $\boldsymbol{\phi}$ and $\boldsymbol{\beta}$. Once we define prior distributions for our parameters, we use Markov chain Monte Carlo (MCMC) techniques to sample from the posterior distributions of $\boldsymbol{\phi}$ and $\boldsymbol{\beta}$. As there are no constraints on $\gamma$ and

$\beta_0$, we assume both are normally distributed with respective means $\mu_\gamma$ and $\mu_{\beta_0}$ and respective variances $\sigma^2_\gamma$ and $\sigma^2_{\beta_0}$. To reflect our need for $\alpha \geq 2$ so that we have a non-monotonic hazard, we assume $\alpha$ is equal in distribution to $Z_\alpha + 2$, where $Z_\alpha$ has a Gamma distribution with mean $(\mu_\alpha - 2)$ and variance $\sigma^2_\alpha$. Note that there is no need to shift the distribution of $\alpha$ if one desires a monotonically decreasing hazard. To satisfy our constraint $\beta_1 > 0$, we assume that $\beta_1$ has a Gamma distribution with mean $\mu_{\beta_1}$ and variance $\sigma^2_{\beta_1}$.

To develop mean hyperparameter values for the cure fraction parameters $\boldsymbol{\beta}$, we ask the investigators to specify an *a priori* value, $P_k$, for the cumulative probability of toxicity for schedule $k, k = 1, 2, \ldots, K$. Based upon the simple linear regression model $E\{\log[-\log(1-P_k)]\} = b_0 + b_1 \log(m_k)$, we use ordinary least squares to find estimates of $b_0$ and $b_1$. We let $\mu_{\beta_0}$ and $\mu_{\beta_1}$ equal the estimates of $b_0$ and $b_1$, respectively. With regard to the hazard shape parameters $\boldsymbol{\phi}$, we ask the investigators to specify an *a priori* value for the limiting cumulative probability of toxicity for a single administration. We denote this value $Q_0$ and note that $Q_0$ must be less than the value of $P_1$ elicited earlier. We also ask investigators to select two time points $t_1$ and $t_2$ and supply *a priori* values $Q_1$ and $Q_2$ for the cumulative probabilities of toxicity at $t_1$ and $t_2$, respectively, for a single administration. Based upon Equation (2), we first derive the value $\theta^* = -\log(1 - Q_0)$. Plugging $\theta^*$ and Equation (3) into Equation (2), some algebra gives us two equations in terms of two parameters $a$ and $g$:

$$\begin{aligned} \log\{-\log[1 + \log(1 - Q_1)/\theta^*]\} &= a \log(t_1) - g \\ \log\{-\log[1 + \log(1 - Q_2)/\theta^*]\} &= a \log(t_2) - g \end{aligned}$$

If we let $\hat{a}$ and $\hat{g}$ denote the respective solutions to $a$ and $g$ in the above equations,

we set $\mu_\alpha = \max\{2.01, \hat{a}\}$ and $\mu_\gamma = \hat{g}$.

We do not derive variance hyperparameter values from elicited information; we choose rather to treat them as tuning parameters that will determine how informative the priors will be. We run exhaustive simulations using a handful of subjects and a variety of variance hyperparameter values until "suitable" values are determined. Specifically, we consider two extreme scenarios: (1) the prior means indicate that the longest schedule is tolerable, yet the first few subjects experience toxicity after very few administrations, and (2) the prior means indicate that only the lowest schedule is tolerable, yet the first few subjects fail to experience toxicity after several administrations. Our approach is to first set the prior variance of each parameter to one-tenth the magnitude of the corresponding prior mean and run small simulation studies under settings (1) and (2) and continually modify the variances until we find values that allow the data to override the prior means in both settings. We also draw 10,000 samples from each of the priors and plug those values into Equation (6) to generate 10,000 prior estimates of the probability of toxicity by $\omega$ for each schedule $k$. We then plot histograms of these values for each schedule $k$ to visually inspect that the distributions have means tending to increase with $m_k$ and with a modest amount of variability around each of the means. Figure 2 is an example of such a plot in the context of the numerical examples of Section 5.

## 4. Trial Conduct

Once suitable prior distributions are selected, we enroll the first subject on the lowest schedule, $\boldsymbol{s}^{(1)}$. With each additional subject $j = 2, 3, \ldots N$, the following procedure is followed:

(1) For each subject $i = 1, 2, \ldots, j-1$, identify $\boldsymbol{s}_i$, the number and timing

of administrations, $U_i$, the length of follow-up, and $C_i$, an indicator for the occurrence of toxicity;

(2) Insert the data into the likelihood in Equation (**??**) and combine it with the prior to sample from the posterior distributions of $\boldsymbol{\phi} = (\alpha, \gamma)$ and $\boldsymbol{\beta} = (\beta_0, \beta_1)$ using MCMC;

(3) Plug the posterior means $\widehat{\boldsymbol{\phi}}$ and $\widehat{\boldsymbol{\beta}}$ into Equation (6) and compute for each schedule $k$, $\widehat{p}_k = 1 - S_k(\omega \mid \widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\beta}}, \boldsymbol{s}^{(k)})$, the estimated probability of toxicity by $\omega$;

(4) Compare all $\widehat{p}_k$ to $p_\omega$, the target probability of toxicity by $\omega$. Then, determine the schedule $k$ with $\widehat{p}_k$ closest to $p_\omega$ and denote the schedule as $\tilde{k}$;

(5) Let $k_{(j-1)}$ denote the schedule assigned to subject $j-1$; assign subject $j$ to schedule $k^* = \min([k_{(j-1)} + 1], \tilde{k})$;

(6) Stop treating any subjects $i = 1, 2, \cdots j-1$ who are still receiving administrations and have received administrations beyond those included in $k^*$;

(7) Reassign schedule $k^*$ to subjects $i = 1, 2, \cdots j-1$ who are still receiving administrations and are assigned to a schedule other than $k^*$.

Once all $N$ subjects have been enrolled and fully followed for a maximum of $\omega$ days, the MTS is defined as the schedule satisfying trial conduct rule (4) using the data of all $N$ subjects.

We now highlight some important aspects in the trial conduct. First, all future planned treatment for a subject is stopped once a toxicity occurs. Second, conduct rule (4) uses a criterion, as a function of treatment schedule, that is identical to

the CRM criterion as a function of dose (O'Quigley et al., 1990). If one were concerned with further limiting the rate of escalation, an alternative criterion proposed in Braun et al. (2005) could be used that reflects the point estimate $\widehat{p}_k$, as well as the percentage of its corresponding posterior distribution lying above $p_\omega$. One could further modify conduct rule (4) to terminate the study if all $\widehat{p}_k$ were greater than $p_\omega$, subject to a minimum accrual requirement. Third, conduct rule (5) forbids non-incremental schedule escalation, so that each subject can be assigned to at most the next-longest schedule beyond that of the preceding subject. However, we do not put any constraint on schedule de-escalation, i.e. we enroll the next subject on the recommended schedule if it is shorter than the schedule assigned to the preceding subject. Fourth, conduct rules (6) and (7) are implemented to promote patient safety while increasing the likelihood that each patient is assigned to the actual MTS. Note that this feature of schedule reassignment was not discussed in the approach of Braun et al. (2005), although such an approach could be easily implemented.

## 5. Numerical Studies

We compared the performance of our algorithm with that of Braun et al. (2005) in a variety of settings via simulations programmed in SAS (SAS Institute Inc., Cary, NC, USA). We adopted the motivating example of Braun et al. (2005), a study in which investigators wished to study $K = 6$ schedules corresponding to 2, 4, 6, 8, 10, 12 weeks of therapy. Each week of therapy consisted of three consecutive daily doses followed by four consecutive days of rest so that schedule $k$ consisted of $m_k = 6k$ administrations. The maximum period to monitor toxicity was specified to be $\omega = 100$ days. Our goal was to determine how long a subject could be treated while maintaining the cumulative probability of toxicity by $\omega$ to

be as close as possible to the threshold value, $p_\omega = 0.40$.

We studied the design with a maximum sample size of $N = 30$ patients, which is feasible in Phase I trials and one that we have also found is sufficient to determine the MTS with reasonable accuracy. In each simulation, subject interarrival times were assumed to be uniformly distributed within 12 to 16 days. The first subject was assigned to the shortest schedule, with subsequent schedule assignments based upon the optimal schedule determined from the methods described in Section 4. At each interim analysis, a single chain of 5000 observations, after a burn-in of 1000 observations was drawn from the posterior distribution of each parameter.

We performed simulations in a series of eight scenarios; the actual probabilities of toxicity within 100 days for each schedule in all scenarios are shown in Table 1. The first six scenarios correspond to settings in which schedule $s^{(j)}$ was optimal for scenario $j, j = 1, 2, \ldots 6$ and there is relatively small differential change in the toxicity probabilities as the schedules increase. The final two scenarios correspond to settings in which there is a large jump in toxicity probabilities near the MTS; in scenario 7, the true MTS is schedule 3, while in scenario 8 the MTS lay between schedules 3 and 4. The proximity of toxicity probabilities for neighboring schedules impacts the ability of any algorithm to identify the target schedule. Therefore, we quantify the difficulty in identifying the optimal schedule in each scenario with a measure $\Delta_p$, which is the average absolute distance of $p_\omega$ from the toxicity probabilities. The value of $\Delta_p$ for each scenario is shown in the final column of Table 1; smaller values indicate greater difficulty in locating the MTS. Thus, we predict that the MTS is easier to identify in scenarios 1,2,7, and 8 and harder to identify in scenarios 3-6. In all simulations, toxicity times were

not simulated under our assumed model; instead, toxicity times were simulated to occur uniformly over the interval $[10 + 14(j-1), 10 + 14j]$ under schedule $s^{(j)}$. As a result, not only can we directly compare the two approaches, we can also examine the performance of our algorithm under model misspecification for the toxicity times.

[Table 1 about here.]

With regard to the prior distributions for $\boldsymbol{\phi}$ and $\boldsymbol{\beta}$, the investigators supplied the values $P_1 = 0.09$, $P_2 = 0.17$, $P_3 = 0.23$, $P_4 = 0.29$, $P_5 = 0.35$, $P_6 = 0.40$. Thus, they believed the longest schedule, $\boldsymbol{s}^{(6)}$, was optimal, a belief that led to a misspecified prior for all but the sixth scenario. The investigators also believed that one administration had a limiting cumulative probability $Q_0 = P_1/6$ (one-sixth of the shortest schedule), with corresponding cumulative probabilities of toxicity $Q_1 = Q_0/4$ and $Q_2 = Q_0/2$ at times $t_1 = 6$ days and $t_2 = 9$ days, respectively. From these elicited values, we used the methods described in Section 3 to estimate the mean hyperparameter values $\mu_\alpha = 2.2$, $\mu_\gamma = 0.1$, $\mu_{\beta_0} = -4.3$ and $\mu_{\beta_1} = 1.1$. Through a detailed sensitivity analysis, we identified variance hyperparameter values $\sigma_\alpha = 0.50, \sigma_\gamma = 0.20, \sigma_{\beta_0} = 1.2$ and $\sigma_{\beta_1} = 0.3$ that allowed for adequate performance of our algorithm. Figure 2 displays histograms of 10,000 draws from the resulting priors of $F(100|\boldsymbol{\phi}, \boldsymbol{\theta}, s^{(j)})$ for $j = 1, \ldots, 6$. As expected, the prior for schedule 6 is centered most closely to the target $p_\omega = 0.40$ and the centers of the distributions increase with the number of administrations. The prior distributions used for the triangular model of Braun et al. (2005) were derived as described in their manuscript, assuming the longest schedule was optimal and one administration had a hazard of 18 days and a peak at $2 \pm 2$ days.

[Figure 2 about here.]

The results in Table 2 demonstrate that both approaches do an excellent job of identifying the MTS even when their corresponding assumed models do not reflect the actual toxicity times. However, this result is not unexpected, as the MTS selected at the end of the study is impacted strongly by the overall rate of toxicity and less so by the actual times of toxicity. The times of toxicity instead influence the schedule assigned to each subject during the study and influence the overall percentage of subjects assigned to a neighborhood of the MTS. And with regard to this latter metric, we see that our approach does a slightly better job in some scenarios of assigning subjects to schedules close to the MTS. For example, our approach assigned an average of 55% of subjects within a neighborhood of the MTS in scenario 2, as compared to 43% with the approach of Braun et al. (2005); the corresponding percentages in scenario 6 are 60% and 52%, respectively.

However, no clear trend exists as to which approach does a better job of assigning subjects during a study. A worthy area for future research would be to formally assess whether one approach is more likely than the other to respond to the actual toxicity times and adjust patient assignments accordingly. Regardless, it is very encouraging that both approaches perform well in scenarios 7 and 8 when there is a drastic increase in the risk of toxicity at schedules beyond the true MTS, even when toxicities have a very late time of onset.

[Table 2 about here.]

## 6. Conclusion

In this paper, we have proposed a non-mixture cure model to identify an optimal schedule among a fixed number of possible nested treatment schedules. Via

simulation, we have demonstrated the excellent operating characteristics of our algorithm when the assumed model is misspecified, as well as when the prior is incorrectly specified. By adopting a cure model framework, we have created a very flexible design that can be used in a variety of settings and allows for the adjustment of patient-level characteristics as sample size permits. We want to emphasize that we did not seek to improve the results of Braun et al. (2005), but rather to develop a more flexible and appealing model for the cumulative hazard of toxicity. Based upon our arguments in Section 1 and the simulation results in Section 5, we feel our approach is an extremely useful contribution to the design of schedule-finding studies.

We have begun to extend our algorithm by including dose as a covariate in our model for the cure fraction and plan to further examine this algorithm to that proposed in Braun et al. (2007). Our approach would assume proportional hazards among doses, which may seem overly restrictive to some readers. However, it is recognized in dose-finding studies that the specific parametric model used is of mild consequence, as the goal of the design is only to identify the optimal dose rather than estimate the probability of toxicity across a continuum of doses. A similar argument can be made for the assumption of proportional hazards, as we suspect that such an assumption will still lead to correct identification of optimal dose and schedule combinations with fewer parameters, although not accurately estimating the probability of toxicity among all dose and schedule combinations. Nonetheless, we could allow the parameters $\alpha$ and $\gamma$ to vary by dose if proportional hazards was not a reasonable assumption. We can also easily generalize our model to allow for hazards that change with each administration by modeling the hazard parameters as a function of administration number or the time between

administrations.

Our model can be used for the design of any clinical trial in which investigators wish to measure the impact of multiple administrations on a binary outcome. Thus, our algorithm could be used in a Phase II study seeking to determine how many administrations are necessary for a desired rate of efficacy, or in a Phase III study comparing two different schedules or doses of the same agent or two different agents in a large sample of (randomized) subjects. Furthermore, if our methods were applied to a large cohort of subjects like that in a Phase III trial, we could model the single-administration hazard non-parameterically with standard techniques rather than forcing a parametric pattern on the event times.

We could also extend our models to allow for optimal treatment schedule finding with combinations of two agents where both agents have a multiple treatment schedule. In this scenario, our outcome remains the time to toxicity; however, the non-mixture cure model would incorporate main effects of both agents into the cure fraction, as well as a term for any possible interaction between the agents. The more challenging aspect of this design is how to incorporate both agents into the time to toxicity hazard, as the two agents will likely differ in both the number of administrations, as well as the times of administration. Nonetheless, once we have a reasonable model, the Bayesian estimation procedures developed in this paper could be used in the design for evaluating the combination therapies.

## REFERENCES

Braun, T. M., Levine, J. E. and Ferrara, J. (2003). Determining a maximum cumulative dose: Dose reassignment within the tite-crm. *Controlled Clinical Trials* **24**, 669–681.

Braun, T. M., Thall, P. F., Nguyen, H. and de Lima, M. (2007). Simultaneously optimizing dose and schedule of a new cytotoxic agent. *Clinical Trials* **4**, 113–124.

Braun, T. M., Yuan, Z. and Thall, P. F. (2005). Determining a maximum tolerated schedule of a cytotoxic agent. *Biometrics* **61**, 335–343.

Chen, M. H., Ibrahim, J. G. and Sinha, D. (1999). A new bayesian model for survival data with a surviving fraction. *Journal of American Statistical Association* **94**, 909–919.

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

Liu, C. A. (2007). Parametric models for optimal treatment schedule finding in adaptive early-phase clinical trials. *PhD Thesis, University of Michigan* .

O'Quigley, J., Pepe, M. and Fisher, L. (1990). Continual reassessment method: A practial design for Phase I clinical trials in cancer. *Biometrics* **46**, 33–48.

Tsodikov, A. D., Ibrahim, J. G. and Yakovlev, A. Y. (2003). Estimating cure rates from survival data: an alternative to two-component mixture models. *Journal of American Statistical Association* **98**, 1063–1078.

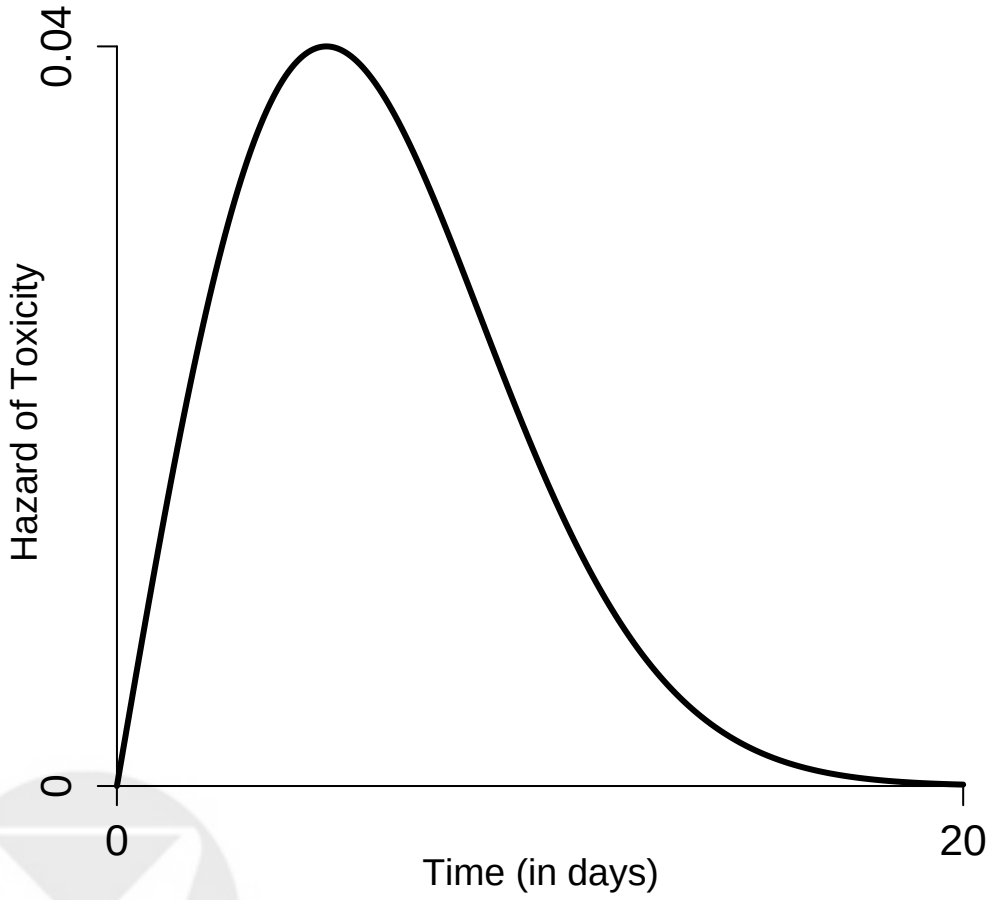


**Figure 1.** Visual representation of single administration hazard based upon Equation (3) with $\alpha = 2, \gamma = 4$, and $\theta = 0.3$.

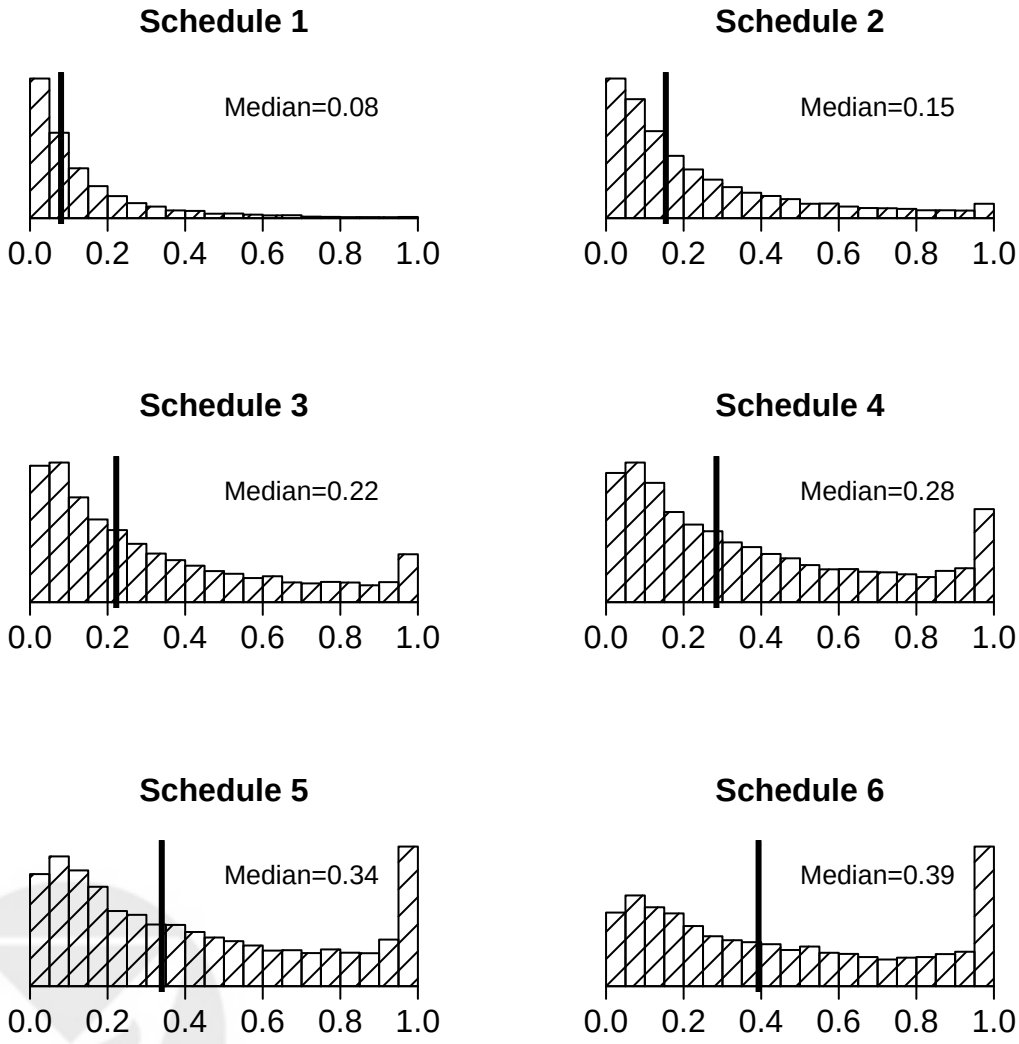


**Figure 2.** Empirical prior distributions for cumulative probability of toxicity by day 100 for each schedule. Solid vertical line represents median.

**Table 1**
*Numeric description of scenarios 1-8*

| | Probability of DLT by $\omega$ | | | | | | |
|---|---|---|---|---|---|---|---|
| Scenario | 1 | 2 | 3 | 4 | 5 | 6 | $\Delta_p$ |
| 1 | 0.40 | 0.64 | 0.79 | 0.87 | 0.92 | 0.95 | 0.36 |
| 2 | 0.23 | 0.40 | 0.54 | 0.64 | 0.72 | 0.79 | 0.21 |
| 3 | 0.16 | 0.29 | 0.40 | 0.49 | 0.57 | 0.64 | 0.14 |
| 4 | 0.12 | 0.23 | 0.32 | 0.40 | 0.47 | 0.54 | 0.12 |
| 5 | 0.10 | 0.19 | 0.26 | 0.34 | 0.40 | 0.46 | 0.13 |
| 6 | 0.08 | 0.16 | 0.23 | 0.29 | 0.35 | 0.40 | 0.15 |
| 7 | 0.10 | 0.20 | 0.30 | 0.60 | 0.70 | 0.80 | 0.25 |
| 8 | 0.20 | 0.30 | 0.40 | 0.70 | 0.80 | 0.90 | 0.25 |

**Table 2**

*Comparison of proposed design to that of Braun,et al (2005) with an incorrectly specified model. For each scenario, each entry is the percentage of simulations in which each schedule was chosen as optimal, with the average percentage of patients assigned to each schedule in parentheses. Line A corresponds to the results using the proposed model; Line B corresponds to the results using Braun,et al (2005). Boldfaced values correspond to schedules within a 10 point neighborhood of* $p_\omega = 0.40$

| | | Schedule (number of weeks) | | | | | |
|---|---|---|---|---|---|---|---|
| Scenario | Method | 1 (2) | 2 (4) | 3 (6) | 4 (8) | 5 (10) | 6 (12) |
| 1 | (A) | **90 (60)** | 9 (26) | 1 (9) | 0 (3) | 0 (2) | 0 (0) |
| | (B) | **88 (58)** | 12 (33) | 0 (6) | 0 (2) | 0 (1) | 0 (0) |
| 2 | (A) | 29 (33) | **61 (55)** | 8 (10) | 2 (1) | 0 (1) | 0 (0) |
| | (B) | 25 (20) | **60 (43)** | 15 (24) | 0 (8) | 0 (3) | 0 (2) |
| 3 | (A) | 9 ( 7) | **25 (25)** | **41 (24)** | **19 (22)** | 6 (14) | 0 (8) |
| | (B) | 8 (14) | **29 (23)** | **45 (32)** | **16 (18)** | 2 ( 9) | 0 (3) |
| 4 | (A) | 4 (10) | 20 (18) | **27 (22)** | **32 (26)** | **14 (13)** | 3 (9) |
| | (B) | 3 ( 9) | 18 (17) | **23 (18)** | **34 (28)** | **15 (16)** | 7 (11) |
| 5 | (A) | 0 (3) | 4 ( 4) | 19 (23) | **26 (30)** | **34 (25)** | **17 (15)** |
| | (B) | 0 (2) | 2 (14) | 22 (22) | **27 (20)** | **33 (28)** | **16 (14)** |
| 6 | (A) | 0 (2) | 0 ( 4) | 11 ( 9) | 22 (25) | **27 (28)** | **40 (32)** |
| | (B) | 0 (3) | 0 (11) | 12 (16) | 20 (18) | **25 (22)** | **43 (30)** |
| 7 | (A) | 10 ( 9) | 25 (25) | **37 (28)** | 20 (15) | 6 (11) | 2 (10) |
| | (B) | 4 (11) | 37 (27) | **41 (36)** | 13 (13) | 4 ( 7) | 1 ( 6) |
| 8 | (A) | 2 (2) | **27 (25)** | **41 (30)** | 19 (24) | 9 (11) | 2 (6) |
| | (B) | 6 (5) | **30 (24)** | **42 (33)** | 13 (19) | 8 (11) | 1 (8) |