Generalizing the TITE-CRM to adapt for early- and late-onset toxicities

Thomas M. Braun[*],[†]

Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, U.S.A.

SUMMARY

Due to the staggered entry of subjects in phase I trials, some subjects will only be partially through the study when others are ready to be enrolled. Nonetheless, many phase I designs focus solely upon whether or not subjects experience toxicity, thereby determining the maximum tolerated dose (MTD) with a binomial likelihood using data from fully observed subjects. The time-to-event continual reassessment method (TITE-CRM) was the first attempt to incorporate information from partially observed subjects by using a weighted binomial likelihood in which the weights are based upon the actual toxicity time distribution. Unfortunately, it is difficult to accurately estimate the toxicity time distribution because only a small proportion of enrolled subjects will experience toxicity. Creators of the TITE-CRM propose the simple alternative of weighting subjects by the proportion of time observed, as well as two adaptive weights to adjust for late-onset toxicities. As a alternative to these approaches, we suggest assuming the toxicity times, as a proportion of the total time under observation, have a Beta distribution with parameters 1.0 and θ; we also allow θ to vary by dose. The value of θ allows us to reflect the occurrence of early- or late-onset toxicities without correctly specifying the actual distribution of toxicity times. Through this model, we do not necessarily expect to improve identification of the MTD, but rather hope to reduce the exposure of subjects to overly toxic doses. Through simulation, we examine how well our model identifies the MTD and allocates dose assignments in three scenarios investigated by previous publications. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: dose escalation trial; phase I trial; CRM; mixture model; cure model; Bayesian statistics

1. INTRODUCTION

The continual reassessment method (CRM) [1] was a significant contribution to the statistical methodology of phase I trials and was followed by later modifications [2]. The CRM assumes that the probability of toxicity increases monotonically with dose, d, via a function $p(d; \beta)$. Most applications of the CRM apply Bayesian methods and assign a prior distribution to β,

Copyright © 2005 John Wiley & Sons, Ltd.

although maximum likelihood methods have been proposed [3]. Each enrolled subject is followed for a finite maximum amount of time defined by the interval in which toxicity may occur. Based upon the toxicity experience of enrolled subjects and the prior placed upon β, the CRM computes a posterior distribution for β.

The posterior mean of β is then used to estimate the probability of toxicity seen at each dose. The next subject is subsequently entered into the trial at the dose whose estimated probability of toxicity is closest to a desired threshold. This algorithm is repeated as each subject enters the trial, and when the pre-specified stopping criteria have been met, data from all the subjects are used to define the maximum tolerated dose (MTD) as the dose whose toxicity rate is closest to the optimal threshold rate.

Although competing phase I designs have been proposed [4–6] those designs share a common limitation with the CRM: censored subjects, i.e. subjects without toxicity who have not reached the end of their observation period, do not contribute to estimation of the MTD until they are fully observed. When partially observed subjects exist, the authors of some designs suggest that the newest subject be enrolled on the current estimate of the MTD instead of delaying enrolment altogether [1, 4]. However, such an allocation scheme is still inefficient because it completely ignores the partial information collected on censored subjects.

The time-to-event CRM (TITE-CRM) was developed to incorporate censored subjects when estimating the MTD [7]. In its simplest form, the TITE-CRM weights each subject without toxicity by the proportion of time each subject has been observed and fully weights each subject with toxicity. The weights are applied to the computation of the posterior mean of β that determines the updated estimate of the MTD. As a result, each subject can be entered into the study as soon as each is eligible, thereby shortening the duration of the entire trial and efficiently using patient information throughout the study.

Selecting the appropriate weight for each subject is not trivial because the actual weights depend upon the distribution of toxicity times in those who experience toxicity. For example, subjects with very little follow-up should receive less weight if toxicities are expected to occur near the end of the observation period rather than near the beginning of the observation period. Furthermore, it is reasonable to assume that the average toxicity time varies by dose, i.e. higher doses are associated with earlier toxicity times. Thus, the appropriate weights should be a function of dose and correctly reflect when toxicities are expected to occur during the study.

Unfortunately, estimation of the toxicity time distribution is difficult because only a small proportion of enrolled subjects will experience toxicity, and a non-parametric estimate of the toxicity time distribution will be infeasible. As an alternative, we suggest using a parametric model and assume the toxicity times, as a proportion of the total time under observation, have a Beta distribution with parameters 1.0 and θ, allowing θ to be a function of dose d. Although this parametric model is unlikely to reflect the true distribution of toxicity times, the model is flexible enough to capture the pattern of early- or late-onset toxicities for a number of distributions. Through this model, we expect the TITE-CRM to assign more subjects at the correct MTD and away from doses above the MTD, while still estimating the MTD as well as when using the simple weighting scheme mentioned earlier.

We present a justification for our model in Section 2. In Section 3, we compare the performance of our proposed design to the performance of the TITE-CRM with other weighting schemes in terms of correctly identifying the MTD and dose allocation during the study. We present summarizing comments in Section 4.

2. NOTATION

2.1. Dose and probability of toxicity

We have a phase I study designed to determine which of J doses is the MTD. Each of the doses is represented by a numeric value $X_j, j = 1, 2, \ldots, J$. We will enrol a maximum of N subjects, and we denote $X_{[i]}, i = 1, 2, \ldots, N$, as the dose assigned to subject i. We observe each subject for a fixed period of time T, during which we measure the binary outcome

$$Y_i = \begin{cases} 0 & \text{no toxicity by time } T \\ 1 & \text{toxicity by time } T \end{cases}$$

which occurs with probability $p(X_{[i]}; \beta)$, a function describing the association of dose and toxicity. For example, if we believe dose and toxicity follow a logistic model, we could use

$$p(X_{[i]}, \beta) = \frac{\exp(-a + \beta X_{[i]})}{1 + \exp(-a + \beta X_{[i]})}$$

where a is a known constant selected so that the probability of toxicity at $X_{[i]} = 0$ is sufficiently low. Some investigators may instead prefer the assumption that dose effects are additive on a logarithmic scale, i.e. $p(X_{[i]}; \beta) = X_{[i]}^{\beta}$. A third possibility is $p(X_{[i]}; \beta) = \{(1 + \tanh X_{[i]})/2\}^{\beta}$ originally suggested by O'Quigley *et al.* [1]. Note in all three examples of $p(X_{[i]}; \beta)$, the value given to each dose X_j may need to be re-scaled to appropriately fit the chosen model for $p(X_{[i]}; \beta)$.

2.2. Dose and time to toxicity

For those subjects experiencing toxicity during the study, we assume that the time to toxicity U_i follows a conditional distribution $F_c(u|Y_i = 1; \lambda)$. As a result, U_i has the marginal distribution

$$F_m(u; \beta, \lambda | X_{[i]}) = p(X_{[i]}; \beta) F_c(u|Y_i = 1; \lambda)$$

and marginal survivor function

$$S_m(u; \beta, \lambda | X_{[i]}) = [1 - p(X_{[i]}; \beta)] + p(X_{[i]}; \beta) S_c(u|Y_i = 1; \lambda)$$

which is a mixture of $S_c(u|Y_i = 1; \theta_i)$, the survivor function for subjects who experience toxicity, and unity, the survivor function for subjects who never experience toxicity. This mixture model is commonly known as a cure model because we assume there is a proportion of subjects who will never experience toxicity and are considered to be 'cured'; the value $[1 - p(X_{[i]}; \beta)]$ is typically known as the cured fraction [8].

It has been shown that in situations where the likelihood function for the parameters (in our case β and λ) is quite flat, there will be a degree of non-identifiability between the parameters [9]. Furthermore, a relatively small proportion of subjects in a phase I trial will experience toxicity, and it will be difficult to pinpoint the actual distribution for toxicity times without strong prior information. As a result, non-identifiability between β and λ will impede estimation of the MTD.

2.3. Proposed model

One solution to the non-identifiability of β and λ is to assume that all subjects who experience toxicity do so uniformly over $[0, T]$ regardless of dose, i.e.

$$F_{\mathrm{c}}(u|Y_i = 1) = \begin{cases} u/T & 0 \leqslant u \leqslant T \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

as presented by Cheung and Chappell [7]. However, they also demonstrate that if toxicities are likely to occur near the end of the interval $[0, T]$, the TITE-CRM with uniform weights tends to be liberal, meaning dose escalation occurs too quickly, being based on partial information gathered before the toxicities are observed. Cheung and Chappell proposed an alternate function to equation (1) that adaptively adjusts for the number of observed toxicities. If toxicities occur near the end of the observation period, the data from partially observed subjects are weighted less than they would be with uniform weights. However, this adaptive function pools information from all subjects regardless of dose and does not allow for the distribution of toxicity times to vary by dose.

   As an alternative to Cheung and Chappell's adaptive weight function, we propose that the uniform distribution in equation (1) be generalized to a Beta distribution, i.e. $F_{\mathrm{c}}(u|Y_i = 1, \theta_i) = [u/T]^{\theta_i}$, $\theta_i > 0$, with uniform weights occurring when $\theta_i = 1$. We pool information across subjects by adopting the model $\theta_i = Z_{[i]}^{\lambda}$, $-\infty < \lambda < \infty$, where $Z_1, Z_2, \ldots, Z_J$ are transformed values of $X_1, X_2, \ldots, X_J$ constrained to the interval $[0, 1]$, and $Z_{[i]}$ is the transformed value corresponding to the dose assigned to subject i. If we model the probability of toxicity as $p(X_{[i]}; \beta) = X_{[i]}^{\beta}$, then the dose values will already be confined to $[0, 1]$ and no additional transformation is needed. If a different toxicity probability model were used, the dose values could simply be divided by a constant $C_x > X_J$. The selection of C_x should be motivated by the desire to keep Z_1 sufficiently above 0, keep Z_J sufficiently below 1, and maintain sufficient variability among the transformed values.

   By placing a prior $\mathcal{N}(0, \sigma^2)$ distribution on λ, with σ^2 known, we allow the distribution of toxicity times to vary by dose, although $\theta_i \approx 1$ on average for all doses, coinciding with equation (1). Non-zero values of $\lambda(\theta_i \neq 1)$ describe the degree to which the toxicity time distribution deviates from a uniform distribution. If $\lambda > 0$, the parameter $\theta_i < 1$ and $F_{\mathrm{c}}(u|Y_i = 1, \theta_i)$ is a right-skewed Beta distribution, indicating that toxicities occur relatively early in the interval $[0, T]$. In contrast, if $\lambda < 0$, the parameter $\theta_i > 1$ and $F_{\mathrm{c}}(u|Y_i = 1, \theta_i)$ is left-skewed, indicating that toxicities occur relatively late in the interval $[0, T]$. Although these Beta distributions are unlikely to exactly describe the toxicity time distributions, our model allows the study to continually adapt to early- and late-onset toxicities without forcing the *a priori* assumption that one is more likely than the other.

2.4. Likelihood for β and λ

When investigators wish to enrol subject $k + 1$ in the study, the prior information from all k currently enrolled subjects is used to estimate the MTD, which is assigned to the new subject. Each currently enrolled subject will have been followed for an amount $T_i \leqslant T$ at which subject i has either been observed with toxicity or is censored without toxicity. If we

denote

$$\delta_i = \begin{cases} 0 & \text{censored without toxicity at } T_i \\ 1 & \text{observed with toxicity at } T_i \end{cases}$$

subject i has a likelihood contribution of

$$\mathscr{L}_i(\beta, \lambda | X_{[i]}, T_i, \delta_i) = [p(X_{[i]}; \beta)(T_i/T)^{\theta_i}]^{\delta_i}[1 - p(X_{[i]}; \beta)(T_i/T)^{\theta_i}]^{1-\delta_i} \tag{2}$$

where $\theta_i = Z_{[i]}^\lambda$ and $Z_{[i]}$ is the transformed value of $X_{[i]}$. We compute $\tilde{\beta}$, the posterior mean of β by assuming λ has a prior $\mathscr{N}(0, \sigma^2)$ distribution and β has a general prior distribution $g(\beta)$. For each dose j, we compute $\tilde{p}_j = p(X_j; \tilde{\beta})$, the estimated probability of toxicity for a subject receiving dose j. We define the MTD as that dose for which $d(\tilde{p}_j, p^*)$ is a minimum, where p^* is the desired rate of toxicity and $d(x, y)$ is a function used to measure the distance between $\tilde{p}_j$ and p^*. However, other criterion for the MTD are possible [10].

Note that the likelihood contribution for subjects with toxicity ($\delta_i = 1$) can be factored directly into separate functions of β and λ. As a result, the weight placed on subjects with toxicity is irrelevant for the computation of $\tilde{\beta}$, and the value of $\tilde{\beta}$ impacts the computation of $\tilde{\lambda}$ only through the subjects without toxicity. Furthermore, Theorem 1 of Cheung and Chappell's paper regarding the consistency of the TITE-CRM applies to our model as well. Although based upon a maximum likelihood approach, the theorem states that, under certain regularity conditions, the probability of correctly identifying the MTD converges to unity as the sample size increases to infinity as long as the weights do not depend upon β. But, as we have argued, our proposed weights were not designed to improve estimation of the MTD. Instead, our weights were designed to improve the sequential assignment of doses by assigning more subjects to the MTD and fewer subjects to toxic doses than other proposed weighting schemes.

3. NUMERICAL EXAMPLES

3.1. Background

In the following section, we use Monte Carlo simulations to examine the performance of our model in terms of: (1) how often the correct MTD is identified; (2) the number of subjects assigned to each dose; and (3) the pattern of dose assignment in an average study. The results from our model, which we denote TITE-B, are compared to two applications of the TITE-CRM: (1) using uniform weights, which we denote TITE-U; and (2) using adaptive weights corresponding to equation (3) of Cheung and Chappell, which we denote TITE-A.

We have six doses under consideration in the trial, and we wish to identify the dose with a probability of toxicity closest to 0.20. Our prior estimates of the probabilities of toxicity for subjects receiving each dose are 0.05, 0.10, 0.20, 0.30, 0.50, and 0.70, respectively. A total of 30 subjects are enrolled in the trial, and subjects are simulated to be eligible for entry at a rate of one subject every 0.5 months. Each subject is followed for a maximum of 6 months. To promote patient safety, all the designs begin enrollment at the lowest dose. Furthermore, all three designs forbid escalation to a dose until all doses below it have been assigned to at least one subject. For example, if an algorithm recommends escalating from dose 2 to dose 4,

Table I. Actual toxicity probabilities; boldface indicates MTD.

Scenario	Dose					
	1	2	3	4	5	6
1	0.05	0.10	**0.20**	0.30	0.50	0.70
2	0.06	0.08	0.12	**0.18**	0.40	0.71
3	0.00	0.00	0.03	0.05	0.11	**0.22**

the study will only permit escalation to dose 3 if no subjects had yet been assigned to dose 3. Dose de-escalation is unconstrained.

The three scenarios of actual toxicity probabilities are displayed in Table I and are a subset of those investigated previously [1, 7]. In scenario 1, the probabilities of toxicity are identical to their prior estimates; thus, dose 3 is the MTD. In scenario 2, the true probabilities vary from their prior values and no dose has an exact probability of toxicity equal to 0.20; dose 4 is the MTD. In scenario 3, all six doses are safe, and the last dose is closest to the true MTD. We also ran simulations under a setting where all six doses were toxic (scenario 2 of Reference [7]), but results for that setting are not presented because all three weighting schemes performed equally well.

The actual toxicity times for subjects who experience toxicity are simulated with three different models. The first model simulates a uniform random variable in the interval [0,6] for all subjects who experience toxicity. Note that in each scenario, times to toxicity have the same uniform distribution for all six doses. The second model simulates times to toxicity using a Weibull distribution with shape parameter fixed at 4 and scale parameter selected for each dose such that the cumulative distribution function at 6 months equals the probability of toxicity for each dose. The third model simulates times to toxicity using a Pareto distribution with a mode of 10 days and scale parameter selected for each dose such that the cumulative distribution function at 6 months equals the probability of toxicity for each dose. The Weibull distribution will simulate settings with late-onset toxicities, while the Pareto distribution will simulate settings with early-onset toxicities.

We use the power model $p(X_j; \beta) = X_j^{\beta}$ to describe the relationship of dose and toxicity, and we assume that β has a prior exponential distribution with mean 1.0. Based upon the power model evaluated at $\beta = 1$, the re-scaled values assigned to each dose are $X_1 = 0.05$, $X_2 = 0.10$, $X_3 = 0.20$, $X_4 = 0.30$, $X_5 = 0.50$, and $X_6 = 0.70$. In the TITE-B model, we set σ^2, the prior variance for λ, at 0.5. Therefore, we assume that 99 per cent of plausible values for λ are in the range $[-1.8, 1.8]$, which we found contained sufficient variability in λ but still allowed for stable estimates of the MTD. Further discussion regarding the selection of σ^2 is relegated to Section 3.2.4.

3.2. Results

3.2.1. Identifying the MTD. Table II displays how well each of the three approaches correctly identifies the MTD in each of the three scenarios. In the first scenario, there are three doses above the MTD and the prior toxicity probabilities match the actual toxicity probabilities. All three models correctly identify the MTD in about 50 per cent of simulations, regardless of the underlying toxicity time distribution. With late-onset toxicities, the TITE-U

Table II. Summary of the abilities of three methods to identify the MTD.

Scenario	Dist'n	Model	Number of simulations MTD identified		
			Below MTD	At MTD	Above MTD
1	W	TITE-B	305	479	216
		TITE-U	241	511	248
		TITE-A	314	478	208
	U	TITE-B	266	541	193
		TITE-U	279	509	212
		TITE-A	265	510	225
	P	TITE-B	283	535	182
		TITE-U	341	484	175
		TITE-A	264	514	222
2	W	TITE-B	340	576	84
		TITE-U	277	641	82
		TITE-A	350	565	85
	U	TITE-B	387	544	69
		TITE-U	391	542	67
		TITE-A	373	558	69
	P	TITE-B	425	502	73
		TITE-U	479	476	45
		TITE-A	362	563	75
3	W	TITE-B	758	242	n/a
		TITE-U	812	188	n/a
		TITE-A	812	188	n/a
	U	TITE-B	811	189	n/a
		TITE-U	851	149	n/a
		TITE-A	845	155	n/a
	P	TITE-B	821	179	n/a
		TITE-U	860	140	n/a
		TITE-A	843	157	n/a

Note: Time-to-toxicity distribution: $W = $ Weibull; $U = $ uniform; $P = $ Pareto. Results are based upon 1000 simulations.

model tends to identify the MTD at toxic doses more often than the other two models. With early-onset toxicities, the TITE-A model tends to identify the MTD at toxic doses more often than the other two models. With uniform toxicities, there is relatively little difference between the three models.

In the second scenario, there are two doses above the MTD, and the prior toxicity probabilities do not equal the actual toxicity probabilities. All three approaches correctly identify the MTD slightly more frequently than they do in the first scenario. With late-onset toxicities, the TITE-U model tends to outperform the other two models, and with early-onset toxicities,

the TITE-A model tends to outperform the other two models, although by assigning more subjects to toxic doses than TITE-U does (see Figure 1). As in scenario 1, all three models perform similarly with uniform toxicities. One might believe that the slightly weakened ability of the TITE-B model in this scenario might be explained by the lower toxicity probabilities and the relative dearth of toxicities from which to determine $\tilde{\lambda}$. We postpone discussion of this conjecture to Section 3.2.4. In the third scenario, all six doses are either at or below the MTD, and the TITE-B model is more likely than the other models to correctly identify the MTD at dose 6, even though the rate of toxicity is extremely low.

3.2.2. Number of subjects assigned. Figures 1 and 2 correspond to scenarios 2 and 3, respectively, and display the average numbers of subjects assigned to each dose across all 1000 simulations for the three approaches. A plot of the results for scenario 1 is omitted, as they mimic the results for scenario 2. In Figure 1, we hope to see a majority of subjects assigned to the first four doses, and in particular, the fourth dose. With late-onset toxicities, we find that the TITE-B model assigns more subjects than the others to the fourth dose and that the TITE-B model assigns fewer subjects than the others to the last two toxic doses. These findings concur with published results that the TITE-CRM with uniform weights can expose subjects to highly toxic doses in the presence of late-onset toxicities [7]. The TITE-B model continues to perform better than the other models with uniform and early-onset toxicities, although in less magnitude than with late-onset toxicities.
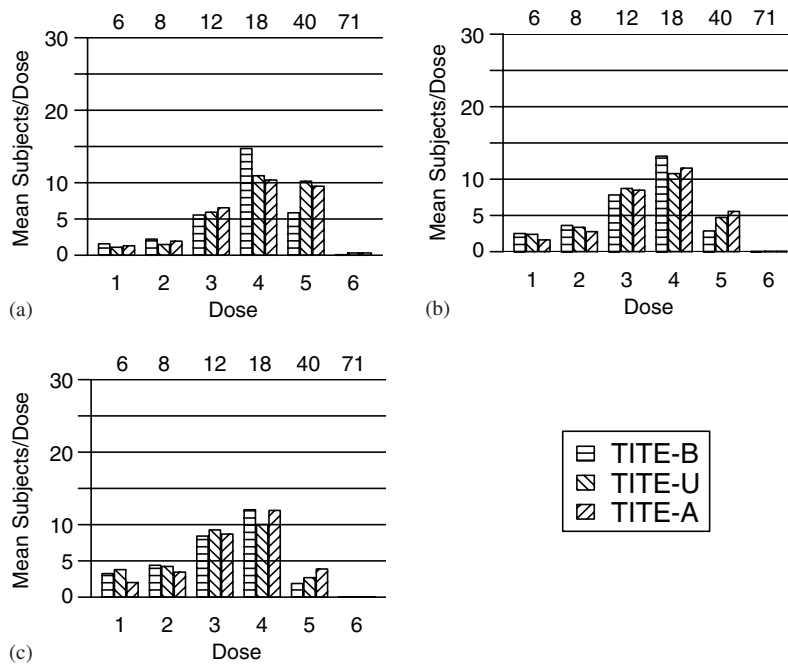

Figure 1. Number of subjects assigned by TITE-B, TITE-U and TITE-A when the fourth of six doses is the MTD (scenario 2). Each bar represents the average number of subjects assigned to each dose over 1000 simulations. Top horizontal axis represents the toxicity probabilities ($\times 100$) of each dose. Toxicity time distribution: (a) = Weibull; (b) = Uniform; and (c) = Pareto.
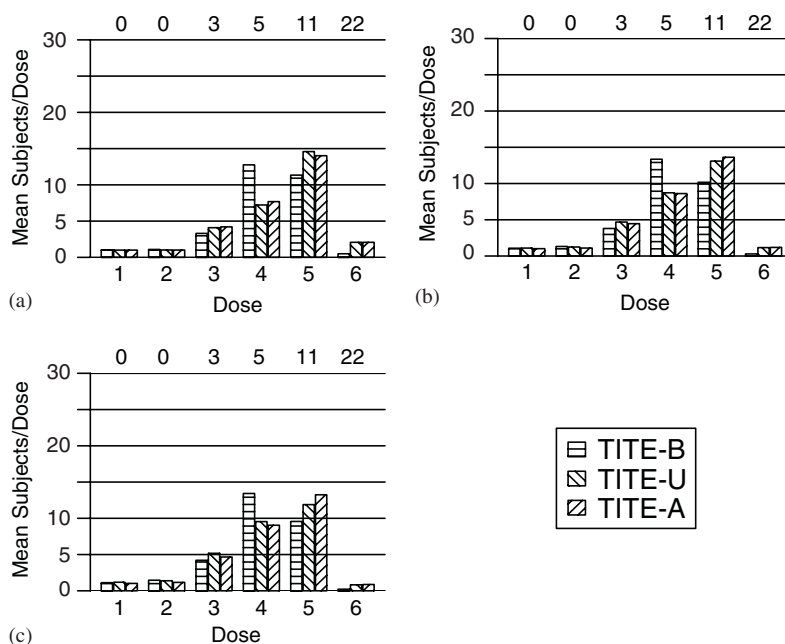
Figure 2. Number of subjects assigned by TITE-B, TITE-U and TITE-A when all six doses are at or below the MTD (scenario 3). Each bar represents the average number of subjects assigned to each dose over 1000 simulations. Top horizontal axis represents the toxicity probabilities ($\times 100$) of each dose. Toxicity time distribution: (a) = Weibull; (b) = Uniform; and (c) = Pareto.

In Figure 2, all six doses are at or below the MTD, and all four approaches assign few subjects to the sixth dose. Moreover, the TITE-U and TITE-A models tend to assign more subjects to the fifth dose, while the TITE-B model assigns more subjects to the fourth dose, exemplifying a slightly conservative nature of the TITE-B model in this scenario. Nonetheless, from Table II, we saw that the TITE-B model was more likely than the other models to correctly identify the MTD. Thus, our results suggest there is a trade-off between estimation of the MTD and allocation of subjects to the MTD when the final dose is the MTD and the toxicity probabilities are relatively low. This fact is further exemplified in Figure 4, which we discuss next and shows the average pattern of dose assignment of the three designs in scenario 3.

3.2.3. Pattern of dose assignment. Figures 3 and 4 correspond to scenarios 1 and 3, respectively, and display the mean dose assigned to each of the 30 subjects across all 1000 simulations of the three approaches. A plot of the results for scenario 2 is omitted, as it was similar to that for scenario 1. Figures 3 and 4 emphasize the consistency of the TITE-CRM, as both show dose assignment converging to the true MTD. However, Figure 4 does indicate that a sample size of 30 subjects may not be sufficient for scenario 3, as all the models tend to remain below the MTD at the last subject.

Nonetheless, as we have emphasized, our focus is not on the final result of the study, but rather on the path taken to reach that final result. Figure 3(a) demonstrates how the TITE-B
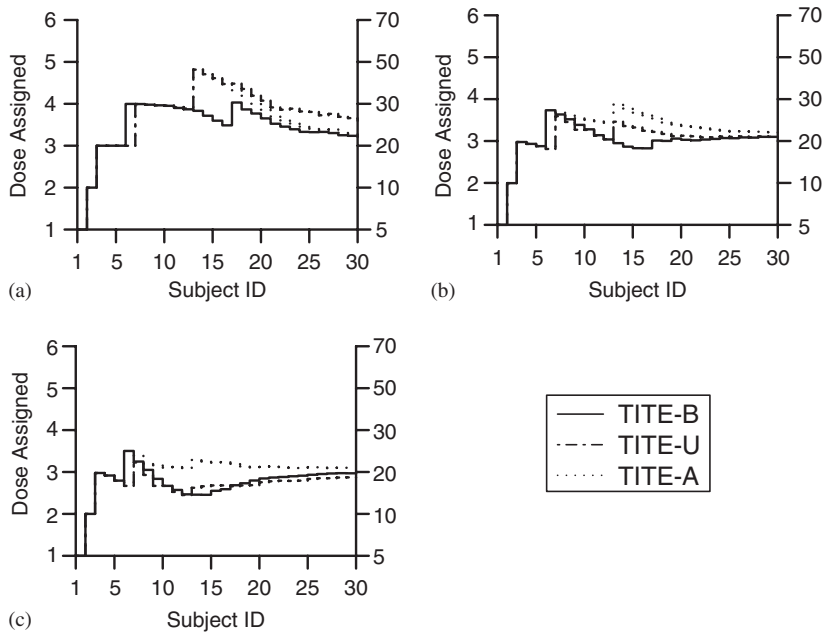
Figure 3. Pattern of dose assignment by TITE-B, TITE-U and TITE-A when the third of six doses is the MTD (scenario 1). Each point represents the mean dose assigned to each subject over 1000 simulations. Right vertical axis represents the toxicity probabilities ($\times 100$) of each dose. Toxicity time distribution: (a) = Weibull; (b) = Uniform; and (c) = Pareto.

model weights tend to limit dose escalation above the MTD with late-onset toxicities better than the other models. Although all three models have escalated beyond the MTD to the fourth dose after 7 or 8 subjects, the TITE-B model begins de-escalating near the thirteenth subject. In contrast, the TITE-U and TITE-A models actually escalate at this point to the fifth dose.

Figure 3(a) also demonstrates how the TITE-A model responds to late-onset toxicities sooner than the TITE-U. By the end of the study, the TITE-A model has de-escalated faster than the TITE-U model and has reached the same point as the TITE-B model; the TITE-U model remains noticeably above the others at the end of the study. The TITE-B model continues to limit dose escalation better than the other models with uniform toxicities; with early-onset toxicities (Figure 3(c)), the TITE-B weights and TITE-U weights lead to similar results, while the adaptive weights of the TITE-A model are less responsive.

Figure 4 concurs with the findings reached earlier with Figure 2: by estimating the additional parameter λ to reflect the nature of the toxicity times, the TITE-B model tends to be more conservative than the other models. Although all three models are similar in the early part of the study, the TITE-B model tends to remain at the fourth dose longer than the other models. As a result, the TITE-A and TITE-U models escalate sooner than the TITE-B model, which is acceptable in this scenario as all six doses are at or below the MTD.

3.2.4. Reflecting toxicity time distribution. For each simulation in each scenario, we computed $\tilde{\lambda}$, the posterior mean for λ, as well as the bounds of the 95 per cent posterior credible
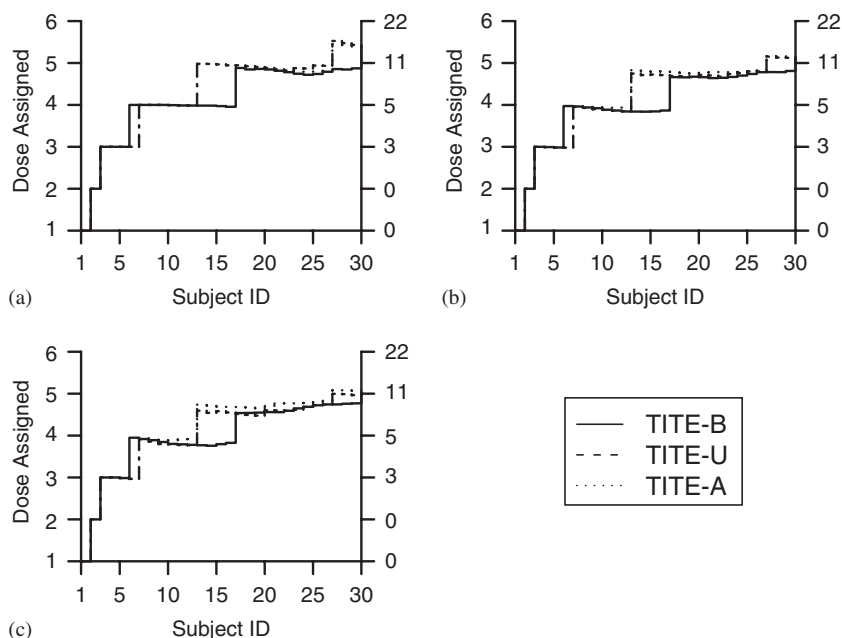
Figure 4. Pattern of dose assignment by TITE-B, TITE-U and TITE-A when all six doses are at or below the MTD (scenario 3). Each point represents the mean dose assigned to each subject over 1000 simulations. Right vertical axis represents the toxicity probabilities ($\times 100$) of each dose. Toxicity time distribution: (a) = Weibull; (b) = Uniform; and (c) = Pareto.

interval for λ. Those values help us to assess the ability of our model to properly reflect the toxicity time distribution. Specifically, with late-onset (early-onset) toxicities, we would expect $\tilde{\lambda}$ to be negative (positive), with a 95 per cent credible interval that is skewed toward negative (positive) values. With uniform toxicities, we expect $\tilde{\lambda} \approx 0$, with a 95 per cent credible interval nearly symmetric about zero.

In scenario 1, the average values of $\tilde{\lambda}$ were -0.82, -0.01, and 0.24, respectively, with Weibull, uniform, and Pareto toxicity times, and the average bounds on the corresponding 95 per cent credible intervals were $(-1.37, -0.20)$, $(-0.50, 0.57)$, $(-0.23, 0.81)$, respectively. Thus, the TITE-B model does a good job of reflecting the appropriate pattern of toxicities. However, it does least well with early-onset toxicities, as the 95 per cent credible interval does overlap with zero, and although the interval is skewed toward positive values, it is also wide enough to encompass a significant range of negative values.

In scenario 2, we found similar average values of $\tilde{\lambda}$ to those in scenario 1. However, we also found that the 95 per cent posterior credible intervals tended to be much wider than those in scenario 1. This added variation in $\tilde{\lambda}$ is likely due to the lower toxicity probabilities in scenario 2 as compared to scenario 1 (see Table I), and with fewer toxicities observed, computation of $\tilde{\lambda}$ is more difficult.

This conjecture is further supported by the results of scenario 3, which had even lower probabilities of toxicity. In scenario 3, the average values of $\tilde{\lambda}$ were -0.43, 0.00, and 0.19,

respectively, with Weibull, uniform, and Pareto toxicity times, and the average bounds on the corresponding 95 per cent credible intervals were $(-1.62, 0.77)$, $(-1.07, 1.09)$, $(-0.83, 1.22)$, respectively. Thus, $\tilde{\lambda}$ tended to remain closer to its prior mean of 0, and the resulting confidence intervals were much wider.

As we stated earlier, 30 subjects may be an insufficient sample size for the third scenario or we may need to refine the prior variance of λ. More specifically, for a given prior variance of λ, estimating λ requires the observation of a 'sufficient' number of toxicities. We can derive a benchmark for 'sufficient' from the TITE-B results displayed in Figures 1 and 2 by multiplying the probabilities of toxicity for each dose by the average number of subjects assigned to each dose. By doing so, we compute an average number of expected toxicities and the degree of 'information' supplied for λ. Using this approach, we found that there were 4–5 expected toxicities in scenarios 1 and 2, depending upon the distribution of toxicity times, and this value dropped to 2 expected toxicities in scenario 3. Therefore, one might conservatively propose a doubling of the sample size in scenario 3 to have the same information as that in scenarios 1 and 2.

We also performed a sensitivity analysis of the influence of σ^2, the prior variance of λ, on the performance of the TITE-B algorithm in scenario 3 with a sample of 30 subjects. With the increased value $\sigma^2 = 1$, the average values of $\tilde{\lambda}$ over 1000 simulations were -0.62, -0.01, and 0.25, respectively, with Weibull, uniform, and Pareto toxicity times, and the average bounds on the corresponding 95 per cent credible intervals were $(-2.06, 0.92)$, $(-1.31, 1.38)$, $(-0.99, 1.58)$, respectively. Thus, by increasing the prior variance, we place less emphasis on the prior and the posterior values of λ are pulled further from 0. The consequence of this result is that dose escalation is less aggressive and, in our simulations, we found that an average of 2 fewer subjects were assigned to doses 5 and 6 (the MTD) and were instead assigned to dose 4.

In contrast, a decreased value $\sigma^2 = 0.25$ lead to average values of $\tilde{\lambda}$ over 1000 simulations of -0.27, 0.00, and 0.13, respectively, with Weibull, uniform, and Pareto toxicity times, and the average bounds on the corresponding 95 per cent credible intervals were $(-1.20, 0.64)$, $(-0.85, 0.83)$, $(-0.69, 0.92)$, respectively. Thus, by decreasing the prior variance, we place more emphasis on the prior and the posterior values of λ are closer to 0. The consequence of this result is that dose escalation is more aggressive and, in our simulations, we found that an average of 2 fewer subjects were assigned to dose 4 and were instead assigned to doses 5 and 6.

4. CONCLUDING REMARKS

We have developed a model for the TITE-CRM that maintains the consistency properties acknowledged by its creators, yet is able to adaptively adjust for late- and early-onset toxicities while the study is running. Although requiring an additional parameter, our model worked as well as the other TITE models in a reasonable sample size of 30 subjects and permits the weights to vary by dose. As noted by a reviewer, the adaptive weights of Cheung and Chappell (TITE-A) are a non-parametric estimate of the toxicity time distribution that requires the estimate of a quantile for each observed toxicity. As our approach requires the estimate of a single parameter rather than multiple quantiles, the TITE-B model may provide a more

efficient estimate of the toxicity time distribution than the TITE-A model. However, we did find indications that the TITE-B model may require additional subjects in some settings to reach its full potential.

If investigators are willing to enrol additional subjects, our model is also flexible enough to allow the modelling of toxicity times on other patient characteristics, such as prior therapy. For example, our model $\theta_i = Z_{[i]}^{\lambda}$ can be re-written as the log-linear model $\log(\theta_i) = \lambda \log(Z_{[i]}) + \gamma \mathbf{W}_i$, where $\mathbf{W}_i$ is a vector of covariates for subject i and γ is the vector of corresponding regression parameters.

We assumed a prior normal distribution for λ that essentially made no assumption about whether early- or late-onset toxicities were more likely; investigators could easily incorporate other prior distributions to reflect other prior beliefs. Moreover, traditional Bayesian approaches (for use with large sample sizes) may promote the use of 'non-informative' priors, but the small sample sizes of most phase I trials require some degree of 'informativeness' in the prior distributions used. In our design, 'informativeness' is quantified by the value of the hyper-parameter σ^2. As implied by our results in Section 3.2.4, investigators should evaluate several values for σ^2 before running the actual study, as there will certainly be a trade-off between the value of σ^2, the number of subjects enrolled, and the number of doses that will be used in the final study design.

REFERENCES

1. O'Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase I clinical trials in cancer. *Biometrics* 1990; **46**:33–48.
2. Goodman S, Zahurak M, Piantadosi S. Some practical improvements in the continual reassessment method for phase I studies. *Statistics in Medicine* 1995; **14**:1149–1161.
3. O'Quigley J, Shen LZ. Continual reassessment method: a likelihood approach. *Biometrics* 1996; **52**:673–684.
4. Babb J, Rogatko A, Zacks S. Cancer phase I clinical trials: efficient dose escalation with overdose control. *Statistics in Medicine* 1998; **17**:1103–1120.
5. Gasparini M, Eisele J. A curve-free method for phase I clinical trials. *Biometrics* 2000; **56**:609–615.
6. Leung DHY, Wang YG. Isotonic designs for phase I trials. *Controlled Clinical Trials* 2001; **22**:126–138.
7. Cheung Y, Chappell R. Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics* 2000; **56**:1177–1182.
8. Berkson J, Gage R. Survival curves for cancer patients following treatment. *Journal of the American Statistical Association* 1952; **47**:510–515.
9. Farewell V. The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* 1982; **38**:1041–1046.
10. Thall PF, Russell KT. A strategy for dose-finding and safety monitoring based on efficacy and adverse outcomes in phase I/II clinical trials. *Biometrics* 1998; **54**:251–264.