Published in final edited form as:

Clin Trials. 2009 June ; 6(3): 227–238. doi:10.1177/1740774509105076.

Model Calibration in the Continual Reassessment Method

Shing M. Lee and Ying Kuen Cheung

Department of Biostatistics, Mailman School of Public Health, Columbia University.

Abstract

Background—The continual reassessment method (CRM) is an adaptive model-based design used to estimate the maximum tolerated dose in dose finding clinical trials. A way to evaluate the sensitivity of a given CRM model including the functional form of the dose-toxicity curve, the prior distribution on the model parameter, and the initial guesses of toxicity probability at each dose is using indifference intervals. While the indifference interval technique provides a succinct summary of model sensitivity, there are infinitely many possible ways to specify the initial guesses of toxicity probability. In practice, these are generally specified by trial and error through extensive simulations.

Methods—By using indifference intervals, the initial guesses used in the CRM can be selected by specifying a range of acceptable toxicity probabilities in addition to the target probability of toxicity. An algorithm is proposed for obtaining the indifference interval that maximizes the average percentage of correct selection across a set of scenarios of true probabilities of toxicity and providing a systematic approach for selecting initial guesses in a much less time consuming manner than the trial and error method. The methods are compared in the context of two real CRM trials.

Results—For both trials, the initial guesses selected by the proposed algorithm had similar operating characteristics as measured by percentage of correct selection, average absolute difference between the true probability of the dose selected and the target probability of toxicity, percentage treated at each dose and overall percentage of toxicity compared to the initial guesses used during the conduct of the trials which were obtained by trial and error through a time consuming calibration process. The average percentage of correct selection for the scenarios considered were 61.5% and 62.0% in the lymphoma trial, and 62.9% and 64.0% in the stroke trial for the trial and error method versus the proposed approach.

Limitations—We only present detailed results for the empiric dose toxicity curve, although the proposed methods are applicable for other dose toxicity models such as the logistic.

Conclusions—The proposed method provides a fast and systematic approach for selecting initial guesses of probabilities of toxicity used in the CRM that are competitive to those obtained by trial and error through a time consuming process, thus, simplifying the model calibration process for the CRM.

Introduction

The estimation of the maximum tolerated dose (MTD) is considered the main objective of phase I clinical trials. The MTD is defined as the dose at which a specified proportion of the patients will experience dose limiting toxicity (DLT). Various methods for estimating the MTD have been proposed throughout the years among them the continual reassessment method (CRM; [1]). This method uses a single parameter dose toxicity model to describe the relationship between dose and probability of DLT. Using the observed toxicity data from each

patient, the model parameter is estimated and the dose associated with the DLT probability closest to a prespecified target is administered to the next patient. The CRM, being a Bayesian model-based method, requires the specification of (i) the functional form of dose-toxicity model, (ii) the prior distribution of its parameter, and (iii) the 'dose levels'. While Chevret [2] studies extensively the impacts of the functional form of the model and the prior distribution of its parameter on the performance of the method, the selection of the 'dose levels' has not been addressed in the literature. When using the CRM, the dose levels are not the doses administered, but rather they are obtained via backward substitution of the initial guesses of the DLT probabilities in the model. Ideally, these initial guesses correspond to the 'desirable range of toxicity probabilities for testing' [1]. However, in most circumstances that information is not available and selection of the initial guesses is performed in an ad hoc manner. In practice, they are generally chosen by trial and error based on extensive simulations to examine the operating characteristics of various sets of initial guesses. This process is complicated since there are as many initial guesses as the number of doses.

The challenges in the calibration of CRM are presented using two motivating examples. The first one is a dose finding trial in patients with previously untreated diffuse large B cell or mantle cell non-Hodgkin's lymphoma [3]. The main objective of the trial was to determine the MTD of VELCADE when administered in combination with CHOP + Rituximab (CHOP-R). DLT was defined as life threatening or disabling neurologic toxicity, very low platelet count or symptomatic non-neurologic or non-hematologic toxicity requiring intervention. The target probability of DLT was 0.25. Eighteen patients were treated for six 21-day cycles (126 days). The standard dose for CHOP-R was administered every 21 days. There were five doses of VELCADE with the third dose level being the starting dose. Dose escalation was conducted according to the time-to-event CRM (TITE-CRM; [4]). In this paper, however, we will focus on the dose-toxicity modeling issue and disregard the time-to-event component.

The second motivating example is a dose finding trial of short-term high dose lovastatin for the treatment of stroke within 24 hours of symptom onset [5]. The MTD was defined as the dose associated with a target probability of DLT of 0.10, where DLT was defined as liver or muscle toxicity based on changes in laboratory markers or symptoms. The treatment was administered to 33 patients for 30 days, with dose escalation for lovastatin occurring in the first 3 days and all patients receiving 20 mg/day for the following 27 days. The five doses of lovastatin that were evaluated up to 30 days after treatment were 1, 3, 6, 8 and 10 mg/kg. Dose escalation for the stroke trial was done in two stages. The first 3 patients were assigned the first dose level. If no DLT was observed, dose escalation was done as described in Table 1 per the coherence principle [10]. Once a DLT is observed, the trial would be switched to the TITE-CRM.

The second author served as the statistician on both trials. Initial guesses of the DLT probabilities were chosen based on model sensitivity [7] and simulated operating characteristics under a pre-specified set of dose-toxicity configurations. As there was no systematic way to choose the initial guesses for evaluation, many sets were tested resulting in a time-consuming ad hoc process. In both trials, the prior MTD was set at the third dose; thus the initial DLT rate at dose level 3 was set at the corresponding pre-specified target. Therefore, the search of the remaining initial guesses were done in a four-dimensional space. This could be a daunting task that deters applied statisticians from using the CRM. The motivation of this paper is to provide a systematic approach for choosing the initial guesses of the probabilities of DLT and consequently the dose levels. We propose selecting initial guesses used in the CRM by specifying a range of acceptable toxicity probabilities instead of a target probability of toxicity. An algorithm is provided for obtaining the range of acceptable toxicity probabilities by maximizing the average percentage of correct selection across a set of scenarios of true probabilities of toxicity. The algorithm provides a fast and systematic approach for selecting

near optimal initial guesses in a much less time consuming manner than the trial and error method. Thus, it simplifies the design calibration process, and enhances the use of the CRM.

Model specification in the CRM

Suppose that we are interested in estimating the dose level associated with a target DLT probability p_T . Let $d_1, d_2, \dots, d_K$ be the K test doses, and $F(d, \beta)$ be the dose toxicity model that is strictly increasing in the dose d for all β . (The model $F(\cdot, \cdot)$ also needs to satisfy some mild regularity conditions; see Appendix 1.) Common choices for the dose toxicity model are:

Empiric: $F(d, \beta) = d^\beta$

One-parameter logistic: $F(d, \beta) = \{1 + \exp(-a - \beta d)\}^{-1}$ where a is a fixed constant

As the CRM was originally proposed in a Bayesian framework, a prior distribution $\pi(\cdot)$ on the model parameter β is also assumed. Given the prior distribution and the data accrued up to the first n patients (i.e. the doses assigned to the patients and their corresponding toxicity outcomes), β can be estimated by the posterior mean (denoted as $\hat{\beta}_n$). The dose level recommended for the $(n+1)$ st patient is the dose with the model-based DLT probability closest to p_T , i.e.,

$$\arg \min_{d_j} |F(d_j; \hat{\beta}_n) - p_T|.$$

This process is continued until a pre-specified number of patients.

Here, we note an important distinction between the dose levels $d_1, d_2, \dots, d_K$ and the doses administered (e.g. 1 mg/kg, 3 mg/kg, etc.). In the context of the CRM, the dose level d_j is obtained by substituting the initial guess of the DLT probability p_j into the specified dose toxicity model. Precisely, d_j is defined such that $p_j = F(d_j; \hat{\beta}_0)$, for $j = 1, \dots, K$, where $\hat{\beta}_0$ denotes the prior mean of β . Thus, to utilize the CRM it is necessary to specify $p_1, \dots, p_K$ in addition to the functional form of F and the prior distribution π so that the doses $d_1, d_2, \dots, d_K$ can be determined.

Chevet suggests via simulations that the performance of the CRM as measured by the percentage of correct selection depends on the choice of F , with one-parameter logistic with $a = 3$ being superior; see Table II in [2]. This conclusion is based on an exponential prior $\pi(\beta) = \exp(-\beta)$ and a particular choice of the initial guesses of DLT probabilities. In what follows, we take a different approach by which the initial guesses are chosen by first fixing F and π , and demonstrate that by selecting the appropriate initial guesses, the empiric dose toxicity model can have comparable performance to the one-parameter logistic with $a = 3$.

Choosing dose levels based on indifference intervals

Assume that we have a dose finding trial with K doses ($d_1 \dots d_K$) and the target probability of DLT is p_T . Let $\Theta = [b_1, b_{K+1}]$ be the parameter space (i.e. $\beta \in \Theta$) and $H_1 = [b_1, b_2]$, $H_k = (b_k, b_{k+1})$ for $k = 2, \dots, K-1$ and $H_K = (b_K, b_{K+1}]$ where b_k is the solution for $F(d_{k-1}, b_k) + F(d_k, b_k) = 2p_T$ for $k = 2, \dots, K$. Shen and O'Quigley [8] showed that for large enough n the CRM will recommend the true MTD (l) with certainty, if $\beta_k \in H_l$ for all k , where β_k is defined such that $F(d_k, \beta_k) = \mu_k$ and μ_k is the true DLT rate associated with dose k . Cheung and Chappell [7] postulated that if the true dose toxicity function is steep around the MTD, for large enough n , the dose recommended by the CRM is the true MTD under the more relaxed conditions whereby $\beta_l \in H_l, \beta_k \in \cup_{i=k+1}^K H_i$ for $k = 1, \dots, l-1$, and $\beta_k \in \cup_{i=1}^{k-1} H_i$ for $k = l+1, \dots, K$.

While these consistency conditions are theoretically revealing, their practical use is limited because the conditions involve the unknown true DLT probabilities μ_k 's. Therefore, Cheung and Chappell suggested converting the intervals in the parameter space for β to intervals on the probability scale. Then, the indifference interval for a given correct dose level l was defined as an interval of DLT probabilities associated with the neighboring doses such that these neighboring doses may be selected instead of the true MTD (l). In notation, the indifference interval for the MTD (l) can be expressed as $(NA, F(d_{l+1}, b_{l+1}))$ for $l = 1$, $(F(d_{l-1}, b_l), F(d_{l+1}, b_{l+1}))$ for $l = 2, \dots, K-1$ and $(F(d_{l-1}, b_l), NA)$ for $l = K$.

For example, in the lymphoma trial p_T is 0.25 and the dose toxicity model is assumed to be empiric where $\beta \in [-5, 5]$. Assume that the initial guesses of the probabilities of DLT at each dose are 0.05, 0.12, 0.25, 0.40 and 0.55, respectively. Then the sets are $H_1 = [-5, -0.6)$, $H_2 = (-0.6, -0.2)$, $H_3 = (-0.2, 0.22)$, $H_4 = (0.22, 0.63)$ and $H_5 = (0.63, 5]$. If $l = 3$, the condition specified by Cheung and Chappell requires that $\mu_1 \in (0, 0.193)$, $\mu_2 \in (0, 0.176)$, $\mu_3 \in (0.179, 0.319)$, $\mu_4 \in (0.319, 1)$ and $\mu_5 \in (0.3254, 1)$ for the CRM to select the MTD for large enough n . If $\mu_2 \in (0.176, \mu_3)$, the CRM may select the second dose as the MTD, but the probability of DLT is close enough to the target probability of DLT that the investigator would be indifferent if the incorrect dose was selected. The same can be said if $\mu_4 \in (\mu_3, 0.319)$. Thus, the indifference interval in this case is $(0.176, 0.319)$. Indifference intervals can be calculated assuming that the MTD is each one of the doses and they are $(NA, 0.31)$, $(0.19, 0.32)$, $(0.18, 0.32)$, $(0.18, 0.32)$ and $(0.18, NA)$ for $l = 1, \dots, 5$. The overall indifference interval is then defined by taking the union of the indifference intervals for $l = 1, \dots, K$. Thus, for the lymphoma trial with the assumptions above, the overall indifference interval is $(0.18, 0.32)$. Thus, by only specifying p_T and dose levels or the initial guesses of the probabilities of DLT at each dose, indifference intervals provide a range in which the DLT probability of the recommended dose will fall, providing an approach to evaluate the sensitivity of different dose toxicity models.

In this article, instead of comparing different dose toxicity models using indifference intervals, we propose to calibrate the CRM model by first specifying the target length of an indifference interval. Specifying the indifference interval in turn facilitates the selection of the initial guesses of the probabilities of DLT due to the correspondence between indifference intervals and the initial guesses of the probabilities of DLT, as follows. Given the prior MTD (v), the total number of dose levels (K), the target probability of DLT (p_T) and the dose toxicity model ($F(d, \beta)$), d_v can be obtained from p_v which equals p_T using backward substitution since $p_v = F(d_v; \hat{\beta}_0)$ where $\hat{\beta}_0$ denotes the prior mean of β . If an indifference interval of length 2δ is desired, dose levels d_{v-1} and d_{v+1} can be obtained from the following equations:

$$F(d_{v-1}, b_v) + F(d_v, b_v) = 2p_T \text{ and } F(d_v, b_{v+1}) + F(d_{v+1}, b_{v+1}) = 2p_T \quad (1)$$

$$F(d_{v-1}, b_v) = p_T - \delta \text{ and } F(d_{v+1}, b_{v+1}) = p_T + \delta \quad (2)$$

Equations (1) are based on the definition of b_k for $k = 2 \dots K$. Equations (2) are based on the definition of indifference interval and its specified length. Thus, $d_{v-1} = F_{\beta}^{-1}(p_T - \delta)$ such that $F(d_v, \beta) = p_T + \delta$ and $d_{v+1} = F_{\beta}^{-1}(p_T + \delta)$ such that $F(d_v, \beta) = p_T - \delta$. These can be obtained using the equations above and they are unique given the regularity conditions for F . Keeping the indifference interval at each dose level the same as the overall indifference interval and using the same procedures above, dose levels $d_1, \dots, d_K$ can be iteratively obtained. From $d_1, \dots, d_K$ we can obtain $p_1, \dots, p_K$ using the dose toxicity model since $p_i = F(d_i; \hat{\beta}_0)$. Thus, by specifying the desired length (δ) of the indifference interval, the target probability of DLT (p_T), the number

of dose levels (K), the prior MTD (v) and the functional form of the dose toxicity model (F), we can obtain the initial guesses of DLT probabilities, without the need to search for good initial guesses (and hence dose levels) in a high dimensional space.

Additionally, it can be shown that given the regularity conditions specified on the dose toxicity model, as δ increases, the distance between p_i and p_T for $i \neq v$ also increases. However, the spread between each pair of doses does not necessarily increase.

THEOREM 1: Let $F(d, \beta)$ be the dose toxicity function such that it is strictly increasing in d for all β and strictly monotone in the parameter β in the same direction for all d , then $|p_{v+i}(\delta) - p_T|$ is strictly increasing in δ , where $p_{v+i}(\delta)$ is the initial guess of toxicity probability for dose $v + i$ given δ and i is an integer such that $1 - v \leq i \leq K - v$ and $i \neq 0$.

The proof for Theorem 1 is provided in Appendix 2. From Theorem 1, we can infer that as δ increases, the set of initial guesses of probabilities of toxicity will be steeper. Simulations were also performed to examine the effect of δ on the prior probability of choosing a dose, $P(\beta \in H_k)$. For example, in the lymphoma trial, if $\delta = 0.03$ then $P(\beta \in H_i)$ for $i = 1, \dots, 5$ are (0.42, 0.06, 0.06, 0.06, 0.40). If $\delta = 0.12$ then $P(\beta \in H_i)$ for $i = 1, \dots, 5$ are (0.19, 0.19, 0.25, 0.19, 0.17). Thus, small values of δ have higher prior probability of selecting extreme doses while large values of δ have higher prior probability of selecting the prior MTD.

Calibrating δ in finite-sample settings

For large sample sizes, the method will guarantee that the target probability of DLT will fall in the specified indifference interval, however, it is necessary to provide some guidelines for the selection of δ in the finite sample setting. We expect that for finite samples the optimal δ value that yields the highest percentage of correct dose selection (PCS) will depend on the other design parameters (p_T , K , v , F) as well as the sample size. We propose an algorithm to calibrate the CRM by selecting the δ that yields the highest PCS given a calibration set of true probabilities of DLT.

ALGORITHM

1. Let $\delta = 0.01$
2. Calculate the initial guesses of the probabilities of toxicity at each dose ($p_1, \dots, p_K$). If $F(d, \beta)$ is the empiric model,

$$p_{i+1} = \exp \left(\frac{\log(p_T + \delta) \log(p_i)}{\log(p_T - \delta)} \right) \text{ for } i = v, \dots, K - 1. \quad (3)$$

$$p_{i-1} = \exp \left(\frac{\log(p_T - \delta) \log(p_i)}{\log(p_T + \delta)} \right) \text{ for } i = 2, \dots, v. \quad (4)$$

If $F(d, \beta)$ is the logistic model,

$$\begin{aligned}
 d_{i+1} &= \frac{\left(\log\left(\frac{(p_T+\delta)}{1-(p_T+\delta)}\right)-a\right) d_i}{\log\left(\frac{(p_T-\delta)}{1-(p_T-\delta)}\right)-a} \text{ for } i=v, \dots, K-1 \\
 d_{i-1} &= \frac{\left(\log\left(\frac{(p_T-\delta)}{1-(p_T-\delta)}\right)-a\right) d_i}{\log\left(\frac{(p_T+\delta)}{1-(p_T+\delta)}\right)-a} \text{ for } i=2, \dots, v. \\
 p_i &= \frac{\exp(a+d_i)}{1+\exp(a+d_i)} \text{ for } i=1, \dots, K.
 \end{aligned}$$

3. Specify the calibration set and assume the K different scenarios of true probabilities of DLT follow the plateau configuration where $p_i = p_L$ for $i < l$, $p_i = p_U$ for $i > l$ and $p_i = p_T$ for $i = l$ where $l = 1, \dots, K$, $\psi = 2$, $p_U = p_T\psi/(1 - p_T(1 - \psi))$ and $p_L = p_T/(\psi + p_T(1 - \psi))$
4. Perform simulations using the CRM under each of the scenarios indicated in step 3 and obtain the corresponding PCS.
5. Average the PCS across all K scenarios of the calibration set.
6. Repeat steps 1–5 for δ values of 0.02 to the minimum of $p_T - 0.005$ and the upper limit of the desired indifference interval on a discrete domain with a grid width of 0.01.
7. Select the δ that yields the highest average PCS.
8. Calculate the initial guesses of the probabilities of toxicity at each dose ($p_1, \dots, p_K$) for the δ selected in step 7 and use these as the initial guesses of the probabilities of toxicity for the CRM.

The following steps are optional:

9. Repeat steps 1 through 8 with $\psi = 3$ and $\psi = 5$
10. For step 7, in addition to selecting the δ that yields the highest average PCS, select the range of δ that yields PCS within 1 percentage point of the highest average PCS for each value of ψ .
 - If the intersection of the ranges of δ for the various values of ψ is non empty, select the δ in the intersection that yields the highest average PCS across all values of ψ .
 - If the intersection of the ranges of δ for the various values of ψ is empty, select the δ values that yield the highest average PCS for each ψ and perform a sensitivity analysis using the validation set to select the optimal value of δ .
11. Calculate the initial guesses of the probabilities of toxicity at each dose ($p_1, \dots, p_K$) for the δ value or values selected in step 10 and use these as the initial guesses of the probabilities of toxicity for the CRM.

The plateau configuration was selected as the calibration set because it provides the most conservative scenario for the CRM. The performance of the CRM will improve for scenarios when the true probabilities of DLT are steeper. The size of the jump is specified using the odds ratio, ψ . The p_L and p_U associated with the most common target probability of DLT for $\psi = 2, 3, 5$ are listed in Table 2. This suggests that $\psi = 2$ yields the most relevant calibration set for practical applications with a range of p_T . It also yields the most conservative calibration set. PCS is expected to increase with increasing odds ratio since the CRM performs better when the true probabilities of DLT are steeper. Thus, in most cases it is not necessary to perform the optional part of the algorithm. The optional section is for scenarios when the true probabilities

of toxicities are believed to be steep. The algorithm along with the optional steps will be referred as “Extended Algorithm”.

General Applications

The proposed algorithm is used to obtain the δ that yields the highest average PCS for various scenarios of target probability of DLT, sample size, number of doses and prior MTD, given calibration sets using an odds ratio of 2. The target probabilities of DLT selected are 0.10, 0.20, 0.25 and 0.33. The sample sizes are 20, 25, 30, 35 and 40. The number of doses ranges from 4 to 7 and the prior MTD ranges from the lowest dose to the median dose. These were selected based on the most common scenarios encountered in practice. The dose toxicity model is assumed to be empiric with the prior distribution for β being normal with $\mu = 0$ and $\sigma^2 = 1.34$ in all cases. The model was chosen based on prior publications on the CRM [9]. We ran 2000 simulations for each scenario except for a sample size of 20 with 7 doses as on average there would be less than 3 patients per dose. The CRM did not allow dose skipping during escalation. The CRM also did not allow for dose escalation immediately after a DLT was observed [10]. However, that dose can be assigned after outcomes from more patients are obtained.

For each scenario, Table 3–Table 6 displays the δ that yields the highest average PCS as well as the range of δ for which the average PCS was within 1% of the optimal δ . Table 3, 4, 5, and 6 display the optimal δ corresponding to target probability of DLT of 0.10, 0.20, 0.25 and 0.33, respectively. It can be observed that the value of the optimal δ decreases as sample size and the number of doses increases. Additionally, the value of the optimal δ increases as a function of the prior MTD.

To illustrate the application of the tables to a particular study, we calibrate the hypothetical example presented in Chevret. The example had a sample size of 25 with a target probability of DLT of 0.25 ($p_T = 0.25$). It had six doses ($K=6$) with the third dose level being the prior MTD ($v = 3$). Given these parameters, we use Table 5 to find the optimal δ which in this case is 0.08. The initial guesses of the probabilities of DLT associated with a δ of 0.08 given the specifications above are 0.029, 0.109, 0.25, 0.42, 0.581, 0.712. These are obtained using equations (3) and (4).

The validation set of true probabilities of DLT included the six scenarios originally included in Chevret. For each scenario of the validation set, we obtain the distribution of the MTD based on 2000 simulations using the initial guesses of probabilities of DLT used by Chevret (i.e. 0.05, 0.10, 0.25, 0.35, 0.50, 0.70) along with a logistic with $a=3$ and $a=1$ as the dose toxicity model and then using the initial guess corresponding to a δ of 0.08 along with an empiric dose toxicity model. For each scenario of the validation set, Table 7 displays the percentage with which each dose level is recommended (% Recommendation), the average absolute difference between the true probability of the dose selected (d^*) and p_T (Average $|p(d^*) - p_T|$), the percentage of patients treated at each dose level (% Treated) and the percentage of patients with DLT (% DLT) using the initial guesses of probabilities of DLT used by Chevret along with a logistic with $a=3$ and $a=1$ as the dose toxicity model and the initial guesses using the Algorithm along with an empiric dose toxicity model. The proposed algorithm performs slightly better compared to the logistic $a=3$ when the MTD is in the middle and slightly worse when the MTD is at the extremes. However, the overall performance is similar, suggesting that the initial guesses obtained using the proposed algorithm with an empiric model is competitive to the logistic with $a=3$, which is the best performing model according to Chevret’s paper. The initial guesses of probabilities used by Chevret are not appropriate when the dose toxicity model is logistic with $a=1$ since the PCS is only 17% when the MTD is the highest dose. The average absolute difference between the true probability of the dose selected and p_T is between 0.05 and 0.07 in most cases with the exception of the logistic with $a=1$ when the MTD is the highest dose.

Applications

The Extended Algorithm is applied to calibrate the two motivating examples. The calibration is based on 2000 simulations. For both of the motivating examples, the dose toxicity model is assumed to be empiric ($F(d; \beta) = d^{\exp(\beta)}$) with the prior distribution for β being normal with $\mu = 0$ and $\sigma^2 = 1.34$. In addition, we performed 2000 simulations using the CRM to compare the initial guesses of probabilities of DLT obtained using the Extended Algorithm to those selected previously based on extensive simulations. The comparisons are done using a validation set of true probabilities of DLT. This is the set of true probabilities of DLT that were originally used when designing the trial. The CRM does not allow dose skipping during escalation. In addition, the CRM does not allow for dose escalation immediately after a DLT is observed.[10]

Application to the Lymphoma Trial

For the lymphoma trial the initial guesses of the probabilities of toxicity at each dose were 0.05, 0.12, 0.25, 0.40 and 0.55, respectively. These were selected based on extensive simulations evaluating the operating characteristics of various initial guesses of the probabilities of DLT based on 2000 simulations using the validation set. Table 8 shows the validation set of the true probabilities of DLT. The other design parameter values are: $p_T = 0.25$, $v = 3$ and $K = 5$ (see section 1). The starting dose is the third dose.

We calibrate the lymphoma trial using the Extended Algorithm. For each calibration set, we iterate the δ values between 0.01 and 0.24. Each δ value corresponds to a set of initial probabilities of DLT. For example, given a δ value of 0.05 which corresponds to an indifference interval of (0.20, 0.30), we can calculate the initial guesses of the probabilities of DLT using equations (3) and (4) as follow:

$$\begin{aligned} d_1 = p_1 &= \exp\left(\frac{\log(0.20)\log(0.157)}{\log(0.30)}\right) = 0.084 \\ d_2 = p_2 &= \exp\left(\frac{\log(0.20)\log(0.25)}{\log(0.30)}\right) = 0.157 \\ d_4 = p_4 &= \exp\left(\frac{\log(0.30)\log(0.25)}{\log(0.30)}\right) = 0.355 \\ d_5 = p_5 &= \exp\left(\frac{\log(0.30)\log(0.355)}{\log(0.20)}\right) = 0.460 \end{aligned}$$

Thus, the initial guesses of the probabilities of DLT at each dose are 0.084, 0.157, 0.250, 0.355, 0.460 respectively, and they yield an indifference interval of (0.20, 0.30).

The results are displayed in Figure 1. As expected, PCS increases with increasing odds ratio. However, the range of δ that yields high PCS varies very little depending on the odds ratio. For an odds ratio (ψ) of 2, the highest average PCS is 48.8% when δ is 0.09. The average PCS was within 1% for δ values between 0.06 and 0.10. For an odds ratio of 3, δ of 0.09 also yielded the highest average PCS of 64.6%. The range of δ is 0.09 to 0.13. The highest average PCS for an odds ratio of 5 was 79.0% which occurred when δ is 0.12. The range of δ is 0.10 to 0.14. Thus, across the various odds ratio a δ of 0.10 is the only δ value that gives an average PCS that is within 1% of the highest average PCS. The initial guesses of the probabilities of DLT at each dose for a δ of 0.10 are 0.011, 0.082, 0.25, 0.464, 0.654. We compare the performance of the initial guesses of the probabilities of DLT corresponding to a δ value of 0.10 to the initial guesses based on extensive simulations using the validation set. To evaluate the sensitivity of the performance to the δ selected based on different odds ratios, we also compared the above initial guesses to the initial guesses of the probabilities of DLT corresponding to a δ value of 0.12 which is the optimal for an odds ratio of 5. The initial guesses of the probabilities of DLT at each dose for a δ of 0.12 are 0.003, 0.058, 0.25, 0.509, 0.719. To evaluate the effect of δ on the operating characteristics, we also compared the above δ values to a δ value of 0.03.

The validation set of true probabilities of DLT included four scenarios. For each scenario of the validation set, Table 8 displays the percentage with which each dose level is recommended (% Recommendation), the average absolute difference between the true probability of the dose selected (d^*) and p_T (Average $|p(d^*) - p_T|$), the percentage of patients treated at each dose level (% Treated) and the percentage of patients with DLT (% DLT). The results are very similar with initial guesses from the extensive simulations doing slightly better when the MTD is at the extreme and slightly worse when the MTD is in the middle. The results for δ values of 0.10 and 0.12 were similar, suggesting that the performance is robust to the choice of odds ratio. By increasing the value of δ from 0.03 to 0.12, the PCS increased substantially when the MTD is in the middle and decreased substantially when the MTD is the highest dose. This is also demonstrated by the average difference between the true probability at the dose selected and p_T .

Application to the Stroke Trial

For the stroke trial, the initial guesses of the probabilities of DLT based on extensive simulations were 0.02, 0.06, 0.10, 0.18 and 0.30 for each dose respectively. These were selected similarly to the ones for the lymphoma trial. Table 9 shows the validation set for the stroke trial. The other design parameter values are: $p_T = 0.10$, $v = 3$ and $K = 5$. To calibrate the trial using the Extended Algorithm, we examine calibration sets within the same range of odds ratio (2, 3, 5). For each calibration set, we iterate the δ values between 0.01 and 0.095 since the target probability of DLT is 0.10. The results are displayed in Figure 2. It is visible in this example that for small odds ratios the PCS is low regardless of the choice of δ . Across all calibration sets, a δ of 0.04 yielded an average PCS was within 1% of the optimal δ . The initial guesses of the probabilities of DLT at each dose for a δ of 0.04 are 0.009, 0.037, 0.10, 0.20, 0.325.

We compare the performance of the initial guesses of the probabilities of DLT corresponding to a δ value of 0.04 to the initial guesses previously chosen using the validation set. The validation set of true probabilities of DLT for the stroke trial included five scenarios. For each scenario of the validation set, Table 9 displays the percentage with which each dose level is recommended (% Recommendation), the average absolute difference between the true probability of the dose selected (d^*) and p_T (Average $|p(d^*) - p_T|$), the percentage of patients treated at each dose level (% Treated) and the percentage of patients with DLT (% DLT) using extensive simulations and the Extended Algorithm. The results are similar with initial guesses from the extensive simulations doing slightly better when the MTD is at the extreme and slightly worse when the MTD is in the middle.

Discussion

To estimate the MTD using the CRM, it is necessary to specify the initial guesses of the probability of DLT at each dose. The literature on the CRM generally uses the same set of initial guesses of the probabilities of DLT that was used in the original paper published by O'Quigley et. al. [1] without any justification. The task of selecting the initial guesses can be daunting given the high dimensionality of the problem and the extensive number of simulations involved. We believe this deters many applied statisticians from using the CRM. The focus of this paper is to provide a systematic and simple approach for choosing the initial guesses of the probabilities of DLT. We provide the optimal δ for the most common scenarios encountered assuming an empiric dose toxicity model along with an example that illustrates the way to use the tables. By simplifying the selection of the initial guesses of the probabilities of DLT in practice, we hope to enhance the use of the CRM. All simulations in this paper can be easily performed using the `getprior` and the `crmsim` function in the R package 'dfcrm' [11], [12]. The `getprior` function calculates the initial guesses of the probabilities of DLT at each dose given

δ , p_T , K , v , and the dose toxicity model. The `crmsim` function performs the simulations by specifying the true probabilities of DLT, the initial guesses, p_T , the sample size, the number of simulations, and the dose toxicity model to be used. See Appendix 3 for the R program for the Chevret example. In addition to being simple, for both of the motivating examples and the hypothetical example in Chevret, our method yielded results that were very similar to those obtained using the ad hoc extensive simulations as well as the results published by Chevret.

The method proposed to select the initial guesses of the probabilities of DLT is based on the concept of indifference intervals. It can be argued that indifference interval is an asymptotic concept and sample sizes are small in dose finding trials. However, it has been demonstrated that poor asymptotic behavior implies poor finite sample operating characteristics. In cases when the investigator can specify the range of acceptable probabilities of toxicity, it is only necessary to calculate the initial guesses of the probability of toxicity at each dose based on that indifference interval and examine the operating characteristics given under scenarios of true probabilities of toxicity. In cases when the investigator can not specify a desired indifference interval, the proposed algorithm can be used to calibrate the CRM and to screen for model specifications. In addition, we propose an algorithm that chooses the optimal δ and provide the optimal δ for some common real life scenarios based on simulations examining the operating characteristics in the finite sample setting.

For all the examples and simulations, the dose toxicity model was assumed to be empiric as we generally use the empiric model in practice. The optimal δ will be different if the dose toxicity model is assumed to be logistic. These are not provided, but can be obtained using the approach described in the paper for the motivating examples. A logistic dose toxicity model can be used for the ‘`dfcrm`’ package.

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. [PubMed: 2350571]
2. Chevret S. The continual reassessment method in cancer phase I clinical trials: A simulation study. *Statistics in Medicine* 1993;12:1093–1108. [PubMed: 8210815]
3. Leonard JP, Furman RR, Cheung YKK, et al. Phase I/II trial of btezomib plus CHOP-Rituximab in diffuse large B cell (DLBCL) and mantle cell lymphoma (MCL): Phase I results. *Blood* 2005;106(11):147A–147A. 491 Part1.
4. Cheung YK, Chappell R. Sequential designs for phase I clinical trials with late-onset-toxicities. *Biometrics* 2000;56(4):1177–1182. [PubMed: 11129476]
5. Elkind MS, Sacco RL, MacArthur RB, et al. The Neuroprotection with Statin Therapy for Acute Recovery Trial (NeuSTART): an adaptive design phase I dose-escalation study of high-dose lovastatin in acute ischemic stroke. *Int J Stroke* 2008;3(3):210–218. [PubMed: 18705902]
6. Cheung YK. Coherence principles in dose-finding studies. *Biometrika* 2005;92(4):863–873.
7. Cheung YK, Chappell R. A simple technique to evaluate model sensitivity in the Continual Reassessment Method. *Biometrics* 2002;58:671–674. [PubMed: 12230003]
8. Shen L, O’Quigley J. Consistency of the continual reassessment method under model misspecification. *Biometrika* 1996;83:395–405.
9. O’Quigley J, Shen L. Continual reassessment method: A likelihood approach. *Biometrics* 1996;52:673–684. [PubMed: 8672707]
10. Cheung YK. Coherence principles in dose-finding studies. *Biometrika* 2005;92:863–873.
11. R Development Core Team. R Foundation for Statistical Computing. Vienna, Austria: 2008. R: A language and environment for statistical computing. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
12. Cheung, YK. `dfcrm`: Dose-finding by the continual reassessment method. 2008. R package version 0.1–1. <http://www.columbia.edu/~yc632>.

Appendix 1 Regularity conditions for dose toxicity model F

- F is strictly increasing in d for all β .
- F is strictly monotone in the parameter β in the same direction for all d .
- Given any $p_T \in (0,1)$, for each d , there exists β in the interior of Θ , the parameter space, such that $F(d; \beta) = p_T$.
- $F(d_k, \beta)$ is bounded away from 0 and 1 for all k and $\beta \in \Theta$.
- $F'(d, \beta) = \partial F(d, \beta) / \partial \beta$ is uniformly bounded in β .
- For each $0 < t < 1$ and each d , the function

$$t \frac{F'(d, \beta)}{F(d, \beta)} + (1-t) \frac{-F'(d, \beta)}{1 - F(d, \beta)}$$

is continuous and strictly monotone in β .

Appendix 2

THEOREM 1: Let $F(d, \beta)$ be the dose toxicity function such that it is strictly increasing in d for all β and strictly monotone in the parameter β in the same direction for all d , then $|p_{v+i}(\delta) - p_T|$ is strictly increasing in δ , where $p_{v+i}(\delta)$ is the initial guess of toxicity probability for dose $v+i$ given δ and i is an integer such that $1-v \leq i \leq K-v$ and $i \neq 0$.

Without loss of generality, assume that $F(d, \beta)$ is strictly decreasing in β for all d , $i > 0$ and $\delta_1 < \delta_2$. Thus $p_T < p_{v+i}$. It suffices to show that $p_{v+i}(\delta_1) < p_{v+i}(\delta_2)$ where $i = 1, \dots, K-v$. This can be proved by induction.

For $i = 1$, we need to show that $p_{v+1}(\delta_1) < p_{v+1}(\delta_2)$. Since $F(d, \beta)$ is decreasing in β ,

$F_{d_v}^{-1}(p_T - \delta_1) < F_{d_v}^{-1}(p_T - \delta_2)$. This implies that, $F_{\beta(\delta_1)}^{-1}(p_T + \delta_1) < F_{\beta(\delta_2)}^{-1}(p_T + \delta_2)$ where $F(d_v, \beta(\delta_k)) = p_T - \delta_k$ for $k = 1, 2$ since $F(d, \beta)$ is strictly increasing in d and $F^{-1}(p, \beta)$ is strictly increasing in β . Now, since $F(d, \beta)$ is strictly increasing in d , $F(d_{v+1}(\delta_1); 1) < F(d_{v+1}(\delta_2); 1)$ where $F(d_{v+1}(\delta_k), \beta(\delta_k)) = p_T + \delta_k$ for $k = 1, 2$. Thus, $p_{v+1}(\delta_1) < p_{v+1}(\delta_2)$.

Now, suppose that $p_{v+j}(\delta_1) < p_{v+j}(\delta_2)$, we want to show that $p_{v+j+1}(\delta_1) < p_{v+j+1}(\delta_2)$. Since $F(d, \beta)$ is strictly decreasing in β and $F^{-1}(p, \beta)$ is strictly increasing in β ,

$F_{d_{v+j}(\delta_1)}^{-1}(p_T - \delta_1) < F_{d_{v+j}(\delta_2)}^{-1}(p_T - \delta_2)$. This implies that, $F_{\beta(\delta_1)}^{-1}(p_T + \delta_1) < F_{\beta(\delta_2)}^{-1}(p_T + \delta_2)$ where $F(d_{v+j}(\delta_k), \beta(\delta_k)) = p_T - \delta_k$ since $F(d, \beta)$ is strictly increasing in d and $F^{-1}(p, \beta)$ is strictly increasing in β . Since $F(d, \beta)$ is strictly increasing in d , $F(d_{v+j+1}(\delta_1); 1) < F(d_{v+j+1}(\delta_2); 1)$ where $F(d_{v+j+1}(\delta_k), \beta(\delta_k)) = p_T + \delta_k$ for $k = 1, 2$. Thus, $p_{v+j+1}(\delta_1) < p_{v+j+1}(\delta_2)$.

Thus, by induction, $p_{v+i}(\delta_1) < p_{v+i}(\delta_2)$ where $i = 1, \dots, K-v$. Since $p_T < p_{v+i}(\delta_1) < p_{v+i}(\delta_2)$, $|p_{v+i}(\delta) - p_T|$ is strictly increasing in δ . The proof for $1-v \leq i \leq -1$ is similar and thus omitted.

Appendix 3 R program for Chevret example

```
nsim<-2000
```

```
n<-25
```

```
target<-0.25
```

```
K<-6  
nu<-3  
# using the delta from calibration  
bestdelta<-0.08  
prior<-getprior(bestdelta, target, nu, K)  
ptrue<-c(0.09, 0.16, 0.27, 0.38, 0.57, 0.75)  
new<-crmsim(ptrue, prior, target, n, nu, nsim=nsim, count=FALSE)
```

Figure 1.
Lymphoma Trial: Proportion of Correct Selection versus δ

Figure 2.
Stroke Trial: Proportion of Correct Selection versus δ

Table 1

Stroke Trial: Dose escalation plan before any DLT was observed

Cohort	Cohort Size	Lovastatin Dose (Day 1, 2) (mg/kg/day)	Lovastatin Dose (Day 3–30) (mg/kg/day)
1	3	1	20
2	3	3	20
3	3	6	20
4	3	6	20
5	3	8	20
6	3	8	20
7	3	8	20
8	3	10	20
9	3	10	20
10	3	10	20
11	3	10	20

Table 2

Lower (p_L) and upper (p_U) probabilities of the calibration sets for p_T of 0.10, 0.20, 0.25 and 0.33 where p_L and p_U are defined such that $(p_T/(1 - p_T))/(p_L/(1 - p_L)) = (p_U/(1 - p_U))/(p_T/(1 - p_T)) = \psi$.

p_T	ψ	p_L	p_U
0.10	2	0.05	0.18
	3	0.04	0.25
	5	0.02	0.36
0.20	2	0.11	0.33
	3	0.08	0.43
	5	0.05	0.55
0.25	2	0.14	0.40
	3	0.10	0.50
	5	0.06	0.63
0.33	2	0.20	0.50
	3	0.14	0.60
	5	0.09	0.71

Table 3

Optimal δ for OR=2 and $p_T = 0.10$ given number of doses (K), prior MTD (v) and sample size

K	v	Sample size				
		20	25	30	35	40
4	1	0.05	0.05	0.05(0.03–0.05)	0.04(0.03–0.05)	0.04(0.02–0.05)
	2	0.07(0.04–0.08)	0.07(0.05–0.08)	0.06(0.03–0.07)	0.05(0.03–0.07)	0.06(0.03–0.06)
5	1	0.05(0.04–0.05)	0.05(0.03–0.05)	0.04(0.03–0.05)	0.04(0.03–0.05)	0.03(0.02–0.05)
	2	0.05(0.04–0.07)	0.06(0.05–0.07)	0.05(0.03–0.06)	0.04(0.03–0.06)	0.05(0.03–0.06)
	3	0.07(0.05–0.08)	0.06(0.05–0.08)	0.06(0.03–0.08)	0.06(0.03–0.07)	0.03(0.03–0.06)
6	1	0.05(0.04–0.05)	0.04(0.03–0.05)	0.04(0.03–0.05)	0.04(0.03–0.04)	0.03(0.03–0.04)
	2	0.05(0.04–0.06)	0.05(0.04–0.06)	0.05(0.03–0.05)	0.04(0.04–0.05)	0.03(0.03–0.05)
7	3	0.05(0.05–0.07)	0.06(0.05–0.07)	0.05(0.04–0.07)	0.04(0.03–0.06)	0.05(0.03–0.06)
	1		0.04(0.03–0.05)	0.04(0.03–0.04)	0.03(0.03–0.04)	0.03(0.02–0.04)
2			0.04(0.03–0.05)	0.04(0.03–0.05)	0.04(0.03–0.04)	0.03(0.03–0.05)
	3		0.05(0.04–0.06)	0.05(0.03–0.06)	0.04(0.04–0.05)	0.03(0.03–0.05)
	4		0.06(0.05–0.07)	0.05(0.03–0.06)	0.04(0.03–0.06)	0.03(0.02–0.06)

Table 4
Optimal δ for OR=2 and $p_T = 0.20$ given number of doses (K), prior MTD (v) and sample size

K	v	Sample size				
		20	25	30	35	40
4	1	0.07(0.04–0.08)	0.05(0.04–0.08)	0.05(0.04–0.06)	0.04(0.03–0.05)	0.04(0.03–0.05)
	2	0.08(0.05–0.11)	0.08(0.04–0.08)	0.08(0.04–0.09)	0.08(0.05–0.08)	0.05(0.04–0.07)
5	1	0.06(0.03–0.08)	0.05(0.03–0.07)	0.05(0.03–0.06)	0.04(0.03–0.05)	0.04(0.03–0.04)
	2	0.08(0.05–0.10)	0.05(0.04–0.09)	0.07(0.04–0.08)	0.05(0.04–0.08)	0.05(0.04–0.07)
6	1	0.05(0.03–0.07)	0.04(0.03–0.07)	0.05(0.03–0.06)	0.04(0.03–0.05)	0.04(0.03–0.05)
	2	0.06(0.04–0.08)	0.05(0.04–0.08)	0.06(0.04–0.07)	0.06(0.03–0.07)	0.04(0.04–0.06)
7	1	0.08(0.05–0.10)	0.08(0.04–0.09)	0.07(0.05–0.08)	0.05(0.04–0.08)	0.04(0.04–0.07)
	2		0.04(0.03–0.06)	0.05(0.03–0.06)	0.03(0.03–0.05)	0.04(0.03–0.05)
	2		0.04(0.03–0.07)	0.06(0.03–0.07)	0.04(0.03–0.07)	0.04(0.04–0.05)
	3		0.07(0.04–0.08)	0.06(0.05–0.07)	0.06(0.04–0.08)	0.04(0.04–0.05)
	4		0.08(0.04–0.09)	0.06(0.04–0.08)	0.07(0.04–0.08)	0.04(0.04–0.07)

Table 5

Optimal δ for OR=2 and $p_T = 0.25$ given number of doses (K), prior MTD (v) and sample size

K	ν	Sample size				
		20	25	30	35	40
4	1	0.06(0.04–0.09)	0.07(0.04–0.08)	0.05(0.04–0.07)	0.04(0.03–0.07)	0.06(0.04–0.07)
	2	0.09(0.06–0.12)	0.07(0.05–0.10)	0.07(0.05–0.11)	0.07(0.05–0.08)	0.06(0.04–0.09)
5	1	0.06(0.03–0.08)	0.05(0.04–0.08)	0.04(0.04–0.07)	0.04(0.03–0.05)	0.04(0.03–0.05)
	2	0.08(0.05–0.11)	0.07(0.05–0.08)	0.07(0.05–0.09)	0.05(0.04–0.08)	0.06(0.04–0.07)
6	3	0.09(0.07–0.12)	0.07(0.05–0.10)	0.09(0.05–0.10)	0.07(0.05–0.08)	0.06(0.05–0.09)
	1	0.06(0.03–0.07)	0.04(0.03–0.07)	0.04(0.03–0.06)	0.03(0.03–0.04)	0.03(0.03–0.05)
7	2	0.07(0.04–0.08)	0.07(0.05–0.08)	0.07(0.04–0.07)	0.04(0.04–0.07)	0.04(0.04–0.07)
	3	0.08(0.05–0.10)	0.08(0.05–0.10)	0.07(0.05–0.09)	0.05(0.04–0.08)	0.06(0.05–0.07)
7	1		0.03(0.03–0.07)	0.04(0.03–0.05)	0.03(0.03–0.04)	0.03(0.03–0.05)
	2		0.07(0.04–0.07)	0.05(0.04–0.07)	0.04(0.03–0.06)	0.04(0.03–0.06)
8	3		0.07(0.05–0.08)	0.05(0.04–0.07)	0.05(0.04–0.07)	0.05(0.04–0.07)
	4		0.07(0.05–0.08)	0.07(0.05–0.09)	0.05(0.05–0.07)	0.06(0.04–0.07)

Table 6

Optimal δ for $OR=2$ and $p_T = 0.33$ given number of doses (K), prior MTD (v) and sample size

K	v	Sample size				
		20	25	30	35	40
4	1	0.06(0.05–0.08)	0.07(0.05–0.09)	0.05(0.04–0.08)	0.05(0.05–0.07)	0.06(0.05–0.07)
	2	0.10(0.07–0.12)	0.09(0.07–0.11)	0.06(0.05–0.11)	0.07(0.06–0.10)	0.07(0.05–0.09)
5	1	0.06(0.04–0.07)	0.06(0.04–0.08)	0.05(0.04–0.07)	0.05(0.04–0.06)	0.05(0.04–0.06)
	2	0.10(0.05–0.10)	0.09(0.06–0.10)	0.05(0.04–0.10)	0.07(0.05–0.09)	0.06(0.05–0.07)
6	1	0.05(0.03–0.07)	0.05(0.03–0.07)	0.04(0.04–0.05)	0.04(0.04–0.05)	0.04(0.04–0.05)
	2	0.07(0.04–0.10)	0.06(0.04–0.09)	0.05(0.04–0.08)	0.06(0.04–0.07)	0.05(0.04–0.06)
7	1	0.09(0.05–0.11)	0.07(0.06–0.10)	0.06(0.04–0.09)	0.07(0.05–0.09)	0.07(0.05–0.08)
	2		0.04(0.03–0.05)	0.04(0.03–0.05)	0.04(0.03–0.05)	0.04(0.03–0.05)
8	1		0.05(0.04–0.08)	0.05(0.04–0.06)	0.04(0.04–0.07)	0.05(0.03–0.06)
	2		0.07(0.04–0.09)	0.07(0.04–0.09)	0.06(0.05–0.07)	0.05(0.04–0.07)
9	1		0.07(0.06–0.09)	0.08(0.05–0.09)	0.07(0.05–0.09)	0.07(0.05–0.08)
	2					

Table 7

Operating characteristics of the initial guesses in Chevet with logistic $a=3$ (CL3) and logistic $a=1$ (CL1) models versus the initial guesses obtained using the Algorithm (A) with empiric model assuming $p_T = 0.25$ and the prior MTD is dose level 3. The initial guesses obtained using the Algorithm correspond to $\delta = 0.10$. $p(d^*)$ is the true probability of the dose selected.

	% Recommendation						Average $ p(d^*) - p_T $	% Treated						% DLT					
	1	2	3	4	5	6		1	2	3	4	5	6	1	2	3	4	5	6
P(DLT)	0.05	0.10	0.25	0.35	0.50	0.70		0.05	0.10	0.25	0.35	0.50	0.70						
CL3	0	18	53	25	4	0	0.062	7	20	38	23	10	2	26					
CL1	1	18	57	22	2	0	0.056	7	21	42	22	8	0	25					
A	0	18	56	24	2	0	0.056	4	20	44	24	7	1	26					
P(DLT)	0.09	0.16	0.27	0.38	0.57	0.75		0.09	0.16	0.27	0.38	0.57	0.75						
CL3	4	32	45	17	1	0	0.070	12	25	36	19	7	1	27					
CL1	5	31	48	16	0	0	0.063	14	26	37	18	6	0	25					
A	3	29	52	16	1	0	0.065	8	25	41	20	5	1	27					
P(DLT)	0.22	0.32	0.45	0.54	0.69	0.80		0.22	0.32	0.45	0.54	0.69	0.80						
CL3	57	35	7	1	0	0	0.059	51	27	15	5	2	0	31					
CL1	62	31	7	0	0	0	0.054	54	25	15	5	2	0	30					
A	51	40	8	1	0	0	0.062	42	33	18	5	2	0	32					
P(DLT)	0.30	0.40	0.52	0.61	0.76	0.87		0.30	0.40	0.52	0.61	0.76	0.87						
CL3	83	15	2	0	0	0	0.069	69	17	10	3	1	0	35					
CL1	87	12	1	0	0	0	0.064	71	15	10	3	1	0	35					
A	80	18	2	0	0	0	0.072	63	22	11	3	1	0	36					
P(DLT)	0.00	0.01	0.04	0.09	0.24	0.49		0.00	0.01	0.04	0.09	0.24	0.49						
CL3	0	0	1	15	74	11	0.060	1	1	8	20	51	18	23					
CL1	0	0	1	19	81	0	0.041	1	1	9	26	63	0	18					
A	0	0	1	20	71	9	0.063	0	1	9	27	51	12	21					
P(DLT)	0.00	0.00	0.03	0.05	0.06	0.22		0.00	0.00	0.03	0.05	0.06	0.22						
CL3	0	0	0	0	11	89	0.048	1	1	5	6	19	68	17					
CL1	0	0	0	1	82	17	0.163	0	1	6	8	80	5	7					
A	0	0	0	2	15	83	0.057	0	0	6	9	26	58	15					

Table 8

Lymphoma Trial: Operating characteristics of the extensive simulation, Extended Algorithm ($\delta=0.10$), the optimal δ for $\psi=5$ ($\delta=0.12$) and a δ value of 0.03 given $p_T=0.25$ and the prior MTD is dose level 3. $p(d^*)$ is the true probability of the dose selected.

		% Recommendation					Average $ p(d^*) - p_T $	% Treated					% DLT				
		1	2	3	4	5		1	2	3	4	5	1	2	3	4	5
Prob of DLT	0.05	0.25	0.40	0.40	0.45	0.55		0.05	0.25	0.40	0.45	0.55					
Simulation	13	56	25	5	1	0.077	0.077	19	37	28	11	5	29				
$\delta=0.10$	9	58	25	7	1	0.073	0.073	13	42	29	13	3	30				
$\delta=0.12$	8	59	27	6	0	0.069	0.069	10	43	32	13	2	31				
$\delta=0.03$	24	42	24	8	2	0.106	0.106	31	28	24	10	7	26				
Prob of DLT	0.05	0.05	0.25	0.45	0.55			0.05	0.05	0.25	0.45	0.55					
Simulation	0	17	65	17	1	0.071	0.071	5	21	48	20	7	26				
$\delta=0.10$	0	15	67	17	1	0.067	0.067	3	20	52	21	4	26				
$\delta=0.12$	0	12	72	15	1	0.059	0.059	0	18	58	21	2	26				
$\delta=0.03$	2	24	48	21	4	0.106	0.106	14	22	36	19	10	25				
Prob of DLT	0.05	0.05	0.08	0.25	0.45			0.05	0.05	0.08	0.25	0.45					
Simulation	0	1	22	61	16	0.071	0.071	2	5	29	44	21	23				
$\delta=0.10$	0	2	20	66	12	0.062	0.062	1	5	28	50	16	22				
$\delta=0.12$	0	1	22	67	10	0.057	0.057	0	3	31	54	12	21				
$\delta=0.03$	1	3	26	43	27	0.106	0.106	5	9	26	30	29	24				
Prob of DLT	0.05	0.05	0.08	0.12	0.25			0.05	0.05	0.08	0.12	0.25					
Simulation	0	1	6	29	64	0.050	0.050	1	3	15	28	52	18				
$\delta=0.10$	0	1	6	36	57	0.059	0.059	1	3	16	37	43	17				
$\delta=0.12$	0	1	8	39	52	0.066	0.066	0	2	18	43	37	16				
$\delta=0.03$	0	1	6	19	73	0.037	0.037	3	4	14	20	58	18				

Table 9

Stroke Trial: Operating characteristics of the extensive simulation versus the Extended Algorithm given $p_T = 0.10$. The initial guesses obtained using the Extended Algorithm correspond to $\delta = 0.04$. An initial design is used prior to observing the first DLT. $p(d^*)$ is the true probability of the dose selected.

	% Recommendation					Average $ p(d^*) - p_T $	% Treated					% DLT				
	1	2	3	4	5		1	2	3	4	5	1	2	3	4	5
Prob of DLT	0.10	0.25	0.30	0.35	0.40		0.10	0.25	0.30	0.35	0.40					
Simulation	88	11	1	0	0	0.019	76	17	6	1	0	14				
Algorithm 1	87	12	1	0	0	0.020	75	19	6	1	0	14				
Prob of DLT	0.04	0.10	0.25	0.30	0.35		0.04	0.10	0.25	0.30	0.35					
Simulation	32	54	13	1	0	0.041	39	38	19	3	0	11				
Algorithm 1	28	59	12	1	0	0.037	35	44	19	2	0	11				
Prob of DLT	0.01	0.04	0.10	0.25	0.30		0.01	0.04	0.10	0.25	0.30					
Simulation	2	30	55	12	1	0.040	15	27	41	16	1	10				
Algorithm 1	1	27	59	12	1	0.037	13	28	44	14	1	9				
Prob of DLT	0.01	0.01	0.04	0.10	0.25		0.01	0.01	0.04	0.10	0.25					
Simulation	0	3	26	53	18	0.045	11	13	31	36	9	7				
Algorithm 1	0	3	29	51	17	0.046	11	14	33	35	8	7				
Prob of DLT	0.01	0.01	0.01	0.04	0.10		0.01	0.01	0.01	0.04	0.10					
Simulation	0	1	5	30	65	0.023	11	10	21	34	24	4				
Algorithm 1	0	1	8	27	64	0.024	11	11	22	34	22	4				