Published in final edited form as:

Stat Med. 2011 July 30; 30(17): 2081–2089. doi:10.1002/sim.4139.

Calibration of prior variance in the Bayesian Continual Reassessment Method

Shing M. Lee and Ying Kuen Cheung

Department of Biostatistics, Mailman School of Public Health, Columbia University, 722 West 168th Street, New York, New York 10032, U.S.A.

SUMMARY

The continual reassessment method (CRM) is an adaptive model-based design used to estimate the maximum tolerated dose in phase I clinical trials. Asymptotically, the method has been shown to select the correct dose given that certain conditions are satisfied. When sample size is small, specifying a reasonable model is important. While an algorithm has been proposed for the calibration of the initial guesses of the probabilities of toxicity, the calibration of the prior distribution of the parameter for the Bayesian CRM has not been addressed. In this paper, we introduce the concept of least informative prior variance for a normal prior distribution. We also propose two systematic approaches to jointly calibrate the prior variance and the initial guesses of the probability of toxicity at each dose. The proposed calibration approaches are compared with existing approaches in the context of two examples via simulations. The new approaches and the previously proposed methods yield very similar results since the latter used appropriate vague priors. However, the new approaches yield a smaller interval of toxicity probabilities in which a neighboring dose may be selected.

Keywords

Dose finding; indifference interval; least informative prior; phase I clinical trials

1. INTRODUCTION

Dose finding clinical trials are studies designed to determine the maximum tolerated dose (MTD) of a drug. The MTD is defined as the dose at which a specified percentage of the patients experience dose limiting toxicities (DLT). Various statistical methods for estimating the MTD have been proposed. The 3+3 design is the most popular in practice because it is easy to implement. However, it does not allow for the specification of a target probability of toxicity and only uses the toxicity information on the previous dose level for determining the dose assignment. The continual reassessment method (CRM; [1]) is a well-known model-based method that overcomes these issues and, at the same time, has good operating characteristics. The CRM is a sequential method by which patients are treated at the dose whose model-based probability of DLT is closest to a specified target probability of DLT. The main challenge when using the CRM is the model calibration. The method requires the specification of the functional form of dose toxicity model and the initial guesses of the probability of DLT at each dose. The Bayesian framework of the CRM also requires the specification of prior distribution of the model parameter.

Studies have examined the impact of model specification on the performance of the CRM. The papers by Chevret [3] and Paoletti and Kramar [4] show the importance of selecting appropriate model parameters via simulation studies and suggest using simulations to evaluate model specifications when designing a trial. In addition, approaches have been

proposed for calibrating the initial guesses. Yin and Yuan [5] propose using multiple parallel CRM models based on different initial guesses and estimating the posterior probabilities of toxicity using the Bayesian model averaging approach. Lee and Cheung [6] suggest a systematic approach to calibrate the initial guesses of the probabilities of DLT using the half-width of the indifference interval for a given functional form and prior distribution of the parameter. The indifference interval is an interval of DLT probability in which the CRM may select the neighboring dose levels, whose toxicity probability fall in this interval, instead of the MTD [7]. Moreover, they demonstrate that given properly selected initial guesses, different functional forms of the dose toxicity model can have comparable performance. The paper by Yin and Yuan does not propose ways to select the initial guesses. It proposes an approach to deal with the arbitrariness of the initial guesses provided by physicians. On the contrary, the paper by Lee and Cheung proposes a systematic approach to select initial guesses that will have reasonable operating characteristics.

While the selection of the initial guesses has been studied, the calibration of the prior distribution of the parameter for the Bayesian CRM has not been addressed in the literature. Chevret [3] examines the impact of the prior distribution on the performance of the method and concludes that the method is robust to the choice of prior as long as the prior is vague. In the context of dose selection, a vague prior should be defined in regards to the MTD instead of the model parameter. Thus, a large prior variance does not necessarily correspond to an uninformative prior in terms of the model-based MTD. For example, let the target probability of toxicity be 0.25 and the prior distribution of the parameter be normal with mean zero. Suppose that there are five dose levels with the initial guesses of the probabilities of DLT at each dose being 0.05, 0.12, 0.25, 0.40 and 0.55, respectively. Table 1 displays the distribution of the model-based MTD for various scenarios of prior variance and dose toxicity model. In this example, the distribution of the model-based MTD changes from unimodal to U-shaped as the prior variance increases. Thus, a prior with a large variance is not an uninformative prior. In addition, a vague prior for a given functional form may not be vague for another functional form (Table 1). Thus, it is important to select a reasonable prior variance given a specific functional form of the dose toxicity model.

In this paper, we introduce the concept of least informative prior variance for a normal prior. We also propose two systematic approaches to jointly specify the prior variance of the parameter and the initial guesses. These approaches use the concepts of the least informative prior variance and the half-width of the indifference interval [6]. The proposed calibration process simplifies the time consuming trial and error process that is currently required for designing the CRM. In addition, having a systematic approach for calibrating the prior variance makes the Bayesian CRM more accessible and easier to implement. We compare the proposed calibration approaches with the existing approach in the context of a bortezomib trial in patients with lymphoma [8] and the hypothetical example presented in Cheung and Chappell [9].

This paper is structured as follows. In Section 2, we introduce the notation for the CRM and review the current calibration method. In Section 3, we introduce the concept of the least informative prior variance and the new calibration approaches. In Section 4, we illustrate the application of the approaches in the context of the two motivating examples and compare the approaches to current techniques. Section 5 discusses the application of the approaches for designing a phase I trial. Further discussion is provided in Section 6.

2. MODEL CALIBRATION FOR THE CONTINUAL REASSESSMENT METHOD

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.

For the Bayesian framework, we assume that the prior distribution of β is normal with mean 0 and variance σ_β^2 . Given the model specification above and the toxicity data accrued up to the first n patients, the dose level recommended by the CRM is the dose with the model-based DLT probability closest to p_T , i.e.,

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

where $\hat{\beta}_n$ is the mean of the posterior distribution of β .

In the context of the CRM, the 'dose levels' ($d_1, d_2, \dots, d_K$) are not the actual doses administered, but rather they are obtained via backward substitution of the initial guesses of the DLT probabilities in the model, F . 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 . The dose toxicity model, $F(d; \beta)$, should be strictly increasing in d for all β , strictly monotone in the parameter β in the same direction for all d and flexible enough to yield a β such that $F(d; \beta) = p_T$ given any $p_T \in (0, 1)$. Common choices for the dose toxicity model are:

Empiric model:

$$F(d, \beta) = d^{\exp(\beta)} \quad (1)$$

One-parameter logistic model:

$$F(d, \beta) = \{1 + \exp(-a - \exp(\beta)d)\}^{-1} \quad (2)$$

where a is a fixed constant

Assume $\Theta = [b_1, b_{K+1}]$ is the parameter space (i.e. $\beta \in \Theta$) and $H_1 = [b_1, b_2]$, $H_j = (b_j, b_{j+1})$ for $j = 2, \dots, K-1$ and $H_K = (b_K, b_{K+1}]$ where b_j is the solution for $F(d_{j-1}; b_j) + F(d_j; b_j) = 2p_T$ for $j = 2, \dots, K$. To obtain a reasonable model specification Lee and Cheung [6] propose calibrating the initial guesses of the probabilities of DLT via the half width of the indifference interval for the MTD (l) which is defined as

$$\delta_l = \frac{F(d_{l+1}, b_{l+1}) - F(d_{l-1}, b_l)}{2}, l = 2, \dots, K-1. \quad (3)$$

By specifying a common half-width indifference interval for all dose levels, that is $\delta_l = \delta$, $d_1, \dots, d_K$ can be obtained recursively. Given a starting dose, v , and a target, p_T , d_v can be obtained via backward substitution, i.e. $p_T = F(d_v; \hat{\beta}_0)$. The remaining dose levels can be obtained by solving the following equations:

$$F(d_{v-1}, b_v) + F(d_v, b_v) = 2p_T \quad \text{and} \quad F(d_v, b_{v+1}) + F(d_{v+1}, b_{v+1}) = 2p_T \quad (4)$$

$$F(d_{v-1}, b_v) = p_T - \delta \quad \text{and} \quad F(d_{v+1}, b_{v+1}) = p_T + \delta \quad (5)$$

Selecting the initial guesses based on δ reduces the number of parameters from K to a single parameter. Thus, the calibration process is simplified. To calibrate the parameter δ , the

performance of CRM is evaluated based on a set of calibration scenarios such that the true probabilities of DLT follow the plateau configuration where $\mu_j = p_L$ for $j < l$, $\mu_j = p_U$ for $j > l$ and $\mu_j = p_T$ for $j = l$ where $l = 1, \dots, K$, $p_U = 2p_T/(1 + p_T)$ and $p_L = p_T/(2 - p_T)$. The δ with the highest average percentage of correct selection (PCS) across all scenarios is chosen and the corresponding dose levels are obtained via backward substitution.

3. LEAST INFORMATIVE PRIOR VARIANCE

3.1 Definition

Assuming the normal formulation for the prior distribution of the model parameter, the CRM requires the specification of $d_1, d_2, \dots, d_K, F$ and σ_β^2 . The parameter σ_β determines the distribution of the model-based MTD. Let v_β be the model-based MTD, then $P(v_\beta = j) = P(\beta \in B_j) = \Phi(b_{j+1}/\sigma_\beta) - \Phi(b_j/\sigma_\beta)$ for $j = 1, \dots, K$ where Φ is the cumulative distribution function of a standard normal. Depending on the value of σ_β , the distribution of v_β can be unimodal, uniform or U-shaped as illustrated in Table 1.

Ideally, an uninformative prior in terms of v_β is one such that the probability of a dose being equal to the model-based MTD is the same for all doses (i.e. $P(v_\beta = j) = 1/K$), that is, the distribution of v_β corresponds to the uniform distribution. Thus, we define σ_β^{LI} to be the σ_β such that the $SD(v_\beta) = \sqrt{(K^2 - 1)/12}$. For example, assuming five dose levels, the least informative normal prior is the one based on σ_β^{LI} such that $SD(v_\beta) = \sqrt{(5^2 - 1)/12} = 1.41$. For an empiric dose toxicity model, $\sigma_\beta^{LI} = 0.74$ and for a logistic with $a=3$, $\sigma_\beta^{LI} = 0.34$. Therefore, the least informative prior variance depends on the functional form. Moreover, given σ_β the distribution of the model-based MTD differs substantially depending on the functional form of the dose toxicity model. Consequently, it is necessary to calibrate σ_β given a specific functional form. A natural choice of σ_β is σ_β^{LI} since generally we have very little information at the start of a dose finding trial and we desire to minimize the effect of the prior in the estimation process.

3.2 Calibration

Instead of fixing σ_β like the algorithm proposed by Lee and Cheung [6], the proposed calibration algorithm calculates σ_β^{LI} for each value of δ given $F(d; \beta)$. It then selects the $(\delta, \sigma_\beta^{LI})$ that maximizes the average PCS across the calibration set specified by Lee and Cheung.

ALGORITHM 1—

1. Given the design parameters $p_T, K, v, F(d; \beta)$, iterate δ from 0.01 to $0.6p_T$ in a discrete domain with a grid width of 0.01.
2. For each δ value obtain the initial guesses of the probabilities of DLT via backward substitution, and calculate σ_β^{LI} .
3. Perform simulations using the CRM under each of the calibration scenarios of the plateau configuration where $\mu_j = p_L$ for $j < l$, $\mu_j = p_U$ for $j > l$ and $\mu_j = p_T$ for $j = l$ where $l = 1, \dots, K$, $p_U = 2p_T/(1 + p_T)$ and $p_L = p_T/(2 - p_T)$.
4. Average the PCS across all K scenarios of the calibration set.
5. Choose the $(\delta, \sigma_\beta^{LI})$ that maximizes the average PCS

To examine the adequacy of σ_β^{LI} , we also propose an calibration algorithm that iterates σ_β in a neighborhood σ_β^{LI} . The algorithm iterates δ and σ_β over a discrete two dimensional grid to find the (δ, σ_β) that maximizes the average PCS with respect to the calibration set specified by Lee and Cheung. This algorithm is bound to outperform Algorithm 1 since it iterates over a range of σ_β which includes σ_β^{LI} , however, it is much more computationally intensive.

ALGORITHM 2—

1. Given the design parameters $p_T, K, v, F(d; \beta)$, iterate δ from 0.01 to $0.6p_T$ in a discrete domain with a grid width of 0.01.
2. For each δ value obtain the initial guesses of the probabilities of DLT via backward substitution, and calculate σ_β^{LI} .
3. Iterate σ_β from $0.8\sigma_\beta^{LI}$ to $1.5\sigma_\beta^{LI}$ in a discrete domain with a grid width of 0.01 and perform simulations using the CRM under each of the calibration scenarios of the plateau configuration where $\mu_j = p_L$ for $j < l$, $\mu_j = p_U$ for $j > l$ and $\mu_j = p_T$ for $j = l$ where $l = 1, \dots, K$, $p_U = 2p_T/(1 + p_T)$ and $p_L = p_T/(2 - p_T)$.
4. Average the PCS across all K scenarios of the calibration set.
5. Choose the (δ, σ_β) that maximizes the average PCS

4. APPLICATION

The two proposed algorithms are applied to calibrate the motivating examples based on 2000 simulations for each calibration scenario. After obtaining the model specifications using the proposed algorithms, we perform 2000 simulations to compare the performance among the model specifications based on the original design (O), the algorithm proposed by Lee and Cheung (LC) and the two proposed algorithms (A1, A2). 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 in the case of the bortezomib trial or specified in the paper in the case of the hypothetical example presented in Cheung and Chappell [9]. The performance is evaluated based on the percentage of dose recommendation, the percentage treated at each dose, the average absolute difference between the true probability of toxicity of the dose selected and the target probability of toxicity, and the percentage of DLT observed during the trial. In addition, we examine the average percentage of correct selection as well as the range of the percentage of correct selection across the various scenarios. Ideally, we want the model specifications that yield a high average percentage of correct selection within a small range. The CRM does not allow dose skipping during escalation nor dose escalation immediately after a DLT is observed [10]. All simulations are performed in R using the 'dfcrm' package [11], [12].

4.1 Bortezomib Trial

The bortezomib study was a dose finding trial in patients with previously untreated diffuse large B cell or mantle cell non-Hodgkin's lymphoma [8]. The main objective of the trial was to determine the MTD of bortezomib when administered in combination with CHOP + Rituximab (CHOP-R). DLT was defined as a life threatening or disabling neurologic toxicity, a very low platelet count or a 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 dose levels of bortezomib with the third dose level being the starting dose. Dose escalation was conducted according to the CRM. The

initial guesses of the probabilities of DLT used in the original design of the trial were 0.05, 0.12, 0.25, 0.40 and 0.55, respectively. These were obtained through extensive simulations examining the operating characteristics under different scenarios of the validation set. The dose toxicity model was assumed to be empiric as defined in (1) with the prior distribution for β being normal with mean 0 and $\sigma_\beta = 1.16$. This was the normal prior that was previously used in the paper by O'Quigley and Shen [13]. Using the same prior distribution, the calibration approach by Lee and Cheung recommended a δ value of 0.10 with an average PCS of 0.482. The corresponding initial guesses were 0.01, 0.08, 0.25, 0.46 and 0.65, respectively [6].

To calibrate the trial using the proposed algorithms, we iterate δ from 0.01 to 0.15 on a discrete domain with a grid width of 0.01 and for each value of δ and we iterate σ_β between $0.8\sigma_\beta^{LL}$ and $1.5\sigma_\beta^{LL}$ on a discrete domain with a grid width of 0.01. As Algorithm 1 is encompassed in Algorithm 2 (i.e. the simulations corresponding to σ_β^{LL}), it is only necessary to perform Algorithm 2. The left panel of Figure 1 displays the average percentage of correct selection using Algorithms 1 and 2 for δ values between 0.01 and 0.15 given an empiric dose toxicity model. The average PCS using Algorithm 2 is 0.506 which corresponds to $\delta = 0.02$ and $\sigma_\beta = 0.28$, while the average PCS using Algorithm 1 is 0.505 which corresponds to $\delta = 0.06$ and $\sigma_\beta = 0.63$. The initial guesses corresponding to δ values of 0.02 and 0.06 are 0.17, 0.21, 0.25, 0.29, 0.33, and 0.06, 0.14, 0.25, 0.38, 0.50, respectively. The average PCS using Algorithm 1 is between 0.500 and 0.506 for δ values between 0.02 and 0.06. Thus, any δ in that range is reasonable. The figure also displays the choice of δ using the tables provided by Lee and Cheung [6] for comparison purposes.

We also assume a logistic model with an intercept of 3 as the dose toxicity model (i.e. equation (2) with $a=3$) and find the corresponding (δ, σ_β) pairs using Algorithms 1 and 2. The right panel of Figure 1 displays the average percentage of correct selection for both algorithms given δ values between 0.01 and 0.15. Again, the average PCS values using the two algorithms are very close. The average PCS are 0.503 and 0.505 for Algorithms 1 and 2, respectively. Both algorithms recommend a δ value of 0.05 which corresponds to initial guesses of 0.09, 0.16, 0.25, 0.36 and 0.46, respectively.

For each scenario of the validation set, Table 2 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 $|p(d^*) - p_T|$), and the percentage of patients with DLT (% DLT). The results are similar for the two algorithms using both the empiric and the logistic models. Even in comparison to existing methods, the average PCS across all scenarios ranges from 0.60 to 0.62 and the difference between the maximum PCS and minimum PCS ranged from 0.07 to 0.12. The performance of the empiric model and the logistic model with an intercept of 3 are also comparable.

4.2 Example in Cheung and Chappell

In the hypothetical example from Cheung and Chappell [9], the target probability of toxicity was 0.20. The study had 25 patients assigned to six dose levels with the third dose being the starting dose. The dose toxicity model was assumed to be empiric ($F(d; \beta) = d^\beta$) and the prior was exponential with rate 1. The initial guesses were 0.05, 0.10, 0.20, 0.30, 0.50, and 0.70. These were the same as those originally used in O'Quigley et al. [1].

For the calibration algorithms (LC, A1, A2), we assume that the dose toxicity model is empiric as defined in (1) and the prior is normal with a mean of 0 and variance of σ_β^2 . For the calibration approach of Lee and Cheung, which fixes the value of σ_β at 1.16, the optimal δ

value is 0.08 [6] with an average PCS of 0.46. This indifference interval corresponds to initial guesses of 0.01, 0.07, 0.20, 0.38, 0.56, and 0.71, respectively.

To calibrate the trial using the proposed algorithms, we iterate δ between 0.01 and 0.12 and for each value of δ we iterate σ_β between $0.8\sigma_\beta^{LI}$ and $1.5\sigma_\beta^{LI}$. Figure 2 displays the average percentage of correct selection using Algorithms 1 and 2 for δ values between 0.01 and 0.12.

The average PCS based on Algorithms 1 is 0.485 corresponding to $\delta = 0.05$ and $\sigma_\beta^{LI} = 0.69$, while the average PCS using Algorithm 2 is 0.488 which corresponds to $\delta = 0.04$ and $\sigma_\beta = 0.55$.

The validation set of true probabilities of DLT included five scenarios. For each scenario of the validation set, Table 3 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 $|p(d^*) - p_T|$), and the percentage of patients with DLT (% DLT). The performance of the original design is very unstable with an average PCS of 0.57 and the PCS ranging from 0.29 to 0.91 (i.e. range=0.62). This is also reflected in the average absolute difference between the true probability of the dose selected. The other methods (LC, A1, A2) have average PCS around 0.60 with the PCS ranging across scenarios between 0.40 and 0.46. The percentage of patients with DLT is similar for all methods.

5. GENERAL APPLICATIONS

In the two examples above, we observe that average PCS in the calibration set based on the optimal σ and σ_β^{LI} are very similar. In addition, we do not see a difference in performance using the validation set between the two algorithms. Thus, using σ_β^{LI} is reasonable and saves computation time. Table 4 displays the optimal $(\delta, \sigma_\beta^{LI})$ for the empiric model given target probabilities of toxicity, numbers of dose levels and sample sizes that may be encountered in practice. These are based on 2000 simulations under each scenario except for cases where the sample size (N) is deemed too small for the given number of dose levels (K) (Cheung YK, unpublished manuscript). 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 with the prior MTD being the highest dose level that is equal or less than the median dose.

6. DISCUSSION

The model specification task that is required when using the CRM to design a phase I trial is very challenging. It involves a lot of trial and error and an extensive number of simulations to determine the best model to use given an application setting. This deters many applied statisticians from using the CRM. The performance of the Bayesian CRM is volatile when the prior is strong. In those situations, the Bayesian CRM performs well under scenarios where the prior MTD is correctly specified, but performs poorly under scenarios when the prior MTD is misspecified. Given the lack of prior information when designing a dose finding trial, it is best to choose a vague prior. The prior should be uninformative in terms of the distribution of the model-based MTD instead of the model parameter. The definition of vague depends on the functional form of the dose toxicity model. This paper introduces the concept of least informative variance using a normal prior distribution and provides a new systematic and simple approach for selecting the set of initial guesses of the probabilities of DLT and least informative variance that yields a reasonably good performance. In addition, it provides reasonable model specifications for the most common scenarios encountered

assuming an empiric dose toxicity model. If other prior distributions are used, it is important to evaluate the distribution of the model-based MTD as part of the calibration process.

The approach yields results that are very similar to the previously proposed method of fixing the prior distribution of the parameter and the dose toxicity model to obtain the optimal initial guesses of the probability of DLT. Assuming the empiric and the logistic model with an intercept of 3, the method by Lee and Cheung fixes the variance at a larger value than the least informative variance. The similar performance of the Lee and Cheung method suggests that variances larger than the least informative variance can perform well as long as it is not too large to lead to incoherent performance and the initial guesses are well selected. In addition, the new algorithm yields a narrower interval of DLT probability in which a neighboring dose can be selected (i.e. δ) by adjusting the variance in the prior distribution of the parameter. Thus, asymptotically, the new algorithm is preferable. The performance of the algorithm is comparable regardless of the functional form of the dose toxicity model. Thus, both one parameter models can be used given that the dose levels and the prior distribution of the model parameter are well specified.

Acknowledgments

This work was supported by NIH grant R01 NS055809.

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. Shen L, O'Quigley J. Consistency of the continual reassessment method under model misspecification. *Biometrika*. 1996; 83:395–405.
3. Chevret S. The continual reassessment method in cancer phase I clinical trials: A simulation study. *Statistics in Medicine*. 1993; 12:1093–1108. [PubMed: 8210815]
4. Paoletti X, Kramar A. A comparison of model choices for the continual reassessment method in phase I cancer trials. *Statistics in Medicine*. 2009; 28:3012–3028. [PubMed: 19672839]
5. Yin G, Yuan Y. Bayesian model averaging continual reassessment method in phase I clinical trials. *Journal of the American Statistical Association*. 2009; 104:954–968.
6. Lee SM, Cheung YK. Model calibration in the continual reassessment method. *Clinical Trials*. 2009; 6:227–238. [PubMed: 19528132]
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. Leonard JP, Furman RR, Cheung YKK, et al. Phase I/II trial of bortezomib plus CHOP-Rituximab in diffuse large B cell (DLBCL) and mantle cell lymphoma (MCL): Phase I results. *Blood*. 2005; 106:147A–147A. 491 Part1.
9. Cheung YK, Chappell R. Sequential designs for phase I clinical trials with late-onset-toxicities. *Biometrics*. 2000; 56:1177–1182. [PubMed: 11129476]
10. Cheung YK. Coherence principles in dose-finding studies. *Biometrika*. 2005; 92:863–873.
11. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2008.
12. Cheung, YK. R package version 0.1-1. 2008. dfcrm: Dose-finding by the continual reassessment method. <http://www.columbia.edu/~yc632>
13. O'Quigley J, Shen L. Continual reassessment method: A likelihood approach. *Biometrics*. 1996; 52:673–684. [PubMed: 8672707]

Figure 1.

Figure 2.

Table 1
The distribution of the model-based MTD under various scenarios of prior variance (σ_β) and dose toxicity model

Dose	$\sigma_\beta = 0.20$	Empiric $\sigma_\beta = 0.74$	$\sigma_\beta = 1.16$	Logistic (a=1) $\sigma_\beta = 0.74$	Logistic (a=3) $\sigma_\beta = 0.74$
1	0.00	0.21	0.30	0.25	0.35
2	0.15	0.18	0.13	0.15	0.10
3	0.70	0.22	0.14	0.20	0.10
4	0.14	0.19	0.13	0.22	0.10
5	0.00	0.19	0.29	0.18	0.35

Table 2

Lymphoma Trial: Operating characteristics using Algorithm 1 (A1) and 2 (A2) with empiric (E) and logistic (L3) models. $p(d^*)$ is the true probability of the dose selected; O is original design; LC is the method by Lee and Cheung.

	% Recommendation					Average $ p(d^*) - p_l $	% DLT
	1	2	3	4	5		
Pr(DLT)	0.05	0.25	0.40	0.45	0.55		
O	13	56	25	5	1	0.077	29
LC	9	58	25	7	1	0.073	30
A1(E)	11	60	24	4	1	0.069	30
A2(E)	16	56	23	5	0	0.078	30
A1(L3)	13	58	24	4	1	0.073	30
A2(L3)	13	57	26	4	0	0.073	30
Pr(DLT)	0.05	0.05	0.25	0.45	0.55		
O	0	17	65	17	1	0.071	26
LC	0	15	67	17	1	0.067	26
A1(E)	0	15	67	17	1	0.067	25
A2(E)	0	18	64	16	1	0.073	25
A1(L3)	0	16	65	18	1	0.072	25
A2(L3)	0	15	68	17	1	0.066	25
Pr(DLT)	0.05	0.05	0.08	0.25	0.45		
O	0	1	22	61	16	0.071	23
LC	0	2	20	66	12	0.062	22
A1(E)	0	1	24	62	14	0.069	21
A2(E)	0	1	22	59	17	0.075	20
A1(L3)	0	1	22	62	15	0.069	21
A2(L3)	0	1	24	62	13	0.069	20
Pr(DLT)	0.05	0.05	0.08	0.12	0.25		
O	0	1	6	29	64	0.050	18
LC	0	1	6	36	57	0.059	17
A1(E)	0	0	6	37	57	0.059	16
A2(E)	0	0	6	31	62	0.052	15
A1(L3)	0	0	6	35	58	0.057	15

	% Recommendation						Average	% DLT
	1	2	3	4	5	56	$ p(d^{*s}) - p_T $	
A2(L3)	0	0	6	37	56	0.060		15

Table 3

Example from Cheung and Chappell 2000: Operating characteristics using Algorithm 1 (A1) and 2 (A2). $p(d^*)$ is the true probability of the dose selected. O is original design; LC is the method by Lee and Cheung.

	% Recommendation						Average $ p(d^*) - p_T $	% DLT
	1	2	3	4	5	6		
Pr(DLT)	0.05	0.10	0.20	0.30	0.50	0.70		
O	1	18	50	29	1	0	0.053	22
LC	1	20	53	25	1	0	0.049	22
A1	1	23	53	22	1	0	0.050	21
A2	1	24	50	23	1	0	0.052	20
Pr(DLT)	0.30	0.40	0.52	0.61	0.76	0.87		
O	91	8	1	0	0	0	0.110	35
LC	89	10	1	0	0	0	0.112	35
A1	93	7	0	0	0	0	0.107	34
A2	92	8	0	0	0	0	0.108	35
Pr(DLT)	0.05	0.06	0.08	0.11	0.19	0.34		
O	0	1	5	32	57	6	0.049	15
LC	0	2	8	29	49	12	0.060	16
A1	0	2	8	30	48	13	0.061	15
A2	0	1	7	32	46	14	0.062	15
Pr(DLT)	0.06	0.08	0.12	0.18	0.40	0.71		
O	0	4	22	59	14	0	0.064	19
LC	0	6	24	60	10	0	0.058	19
A1	0	6	27	56	12	0	0.063	18
A2	1	5	27	55	12	0	0.064	18
Pr(DLT)	0.00	0.00	0.03	0.05	0.11	0.22		
O	0	0	0	7	64	29	0.074	11
LC	0	0	0	8	43	49	0.061	13
A1	0	0	0	7	44	49	0.060	12
A2	0	0	0	7	43	50	0.059	12

Table 4

Optimal $(\delta, \sigma_{\beta}^{LL}(\delta))$ for the empiric model given the target probability of DLT (p_T), number of doses (K), and sample size (N)

K	N	Target Probability of DLT			
		0.10	0.20	0.25	0.33
4	20		(0.06, 0.61)	(0.06, 0.56)	(0.08, 0.71)
	25		(0.05, 0.51)	(0.05, 0.47)	(0.06, 0.53)
	30	(0.04, 0.59)	(0.05, 0.51)	(0.06, 0.56)	(0.06, 0.53)
	35	(0.04, 0.59)	(0.05, 0.51)	(0.05, 0.47)	(0.05, 0.44)
	40	(0.03, 0.43)	(0.05, 0.51)	(0.06, 0.56)	(0.08, 0.71)
5	20		(0.04, 0.45)	(0.04, 0.42)	(0.07, 0.70)
	25		(0.03, 0.34)	(0.05, 0.52)	(0.06, 0.60)
	30	(0.03, 0.48)	(0.04, 0.45)	(0.06, 0.63)	(0.04, 0.40)
	35	(0.03, 0.48)	(0.03, 0.34)	(0.06, 0.63)	(0.07, 0.70)
	40	(0.02, 0.32)	(0.04, 0.45)	(0.05, 0.52)	(0.06, 0.60)
6	20			(0.05, 0.64)	(0.06, 0.72)
	25		(0.05, 0.69)	(0.05, 0.64)	(0.05, 0.60)
	30		(0.04, 0.55)	(0.05, 0.64)	(0.06, 0.72)
	35	(0.03, 0.58)	(0.04, 0.55)	(0.05, 0.64)	(0.06, 0.72)
	40	(0.03, 0.58)	(0.04, 0.55)	(0.05, 0.64)	(0.06, 0.72)
7	20				
	25		(0.04, 0.61)	(0.05, 0.71)	(0.05, 0.67)
	30		(0.03, 0.46)	(0.04, 0.57)	(0.04, 0.54)
	35	(0.03, 0.65)	(0.03, 0.46)	(0.04, 0.57)	(0.04, 0.54)
	40	(0.02, 0.43)	(0.04, 0.61)	(0.04, 0.57)	(0.04, 0.54)