Research Article

Statistics in Medicine

Received 2 November 2011, Accepted 25 August 2012 Published online 17 September 2012 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.5621

Adaptive prior variance calibration in the Bayesian continual reassessment method

Jin Zhang,*† Thomas M. Braun and Jeremy M. G. Taylor

The use of the continual reassessment method (CRM) and other model-based approaches to design Phase I clinical trials has increased owing to the ability of the CRM to identify the maximum tolerated dose better than the 3 + 3 method. However, the CRM can be sensitive to the variance selected for the prior distribution of the model parameter, especially when a small number of patients are enrolled. Although methods have emerged to adaptively select skeletons and to calibrate the prior variance only at the beginning of a trial, there has not been any approach developed to adaptively calibrate the prior variance throughout a trial. We propose three systematic approaches to adaptively calibrate the prior variance during a trial and compare them via simulation with methods proposed to calibrate the variance at the beginning of a trial. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: adaptive design; Bayes factor; dose-finding study; dose-escalation study

1. Introduction

Phase I clinical trials are studies of human subjects aimed at estimating the maximum tolerated dose (MTD), with the sample size typically in the range of 20–40 subjects. The MTD is the dose at which the probability of having a dose-limiting toxicity (DLT) is near a predefined target $0 < \eta < 1$. Because the dose identified as the MTD will be further investigated for efficacy in Phase II trials, it is important to obtain an accurate estimate for the MTD. Owing to the severity of most DLTs, patient safety dictates that the study begins at low doses and escalates doses as patients are accrued so that exposure of patients to doses above the MTD is minimized. However, escalation of doses should also occur as quickly as possible as lower doses are also expected to be ineffective for treating or preventing recurrence of cancer.

A vast amount of methodology exists for the design of Phase I trials. The 3 + 3 design is the standard algorithmic design using cohorts of three patients. Although algorithmic designs are simple to understand and implement, their resulting MTD estimates have large bias and variance. Also, many subjects are likely to be treated at doses below the MTD [1, 2].

A preferred design would incorporate a parametric model for the association of dose and probability of DLT. One popular model-based method is the continual reassessment method (CRM), which provides the MTD estimate from a fixed set of dose levels by using a one-parameter model for the dose–toxicity relationship. The parameter estimate is updated every time a new subject or cohort completes its follow-up by using either a Bayesian approach proposed by O'Quigley *et al.* [3] or maximum likelihood methods proposed by O'Quigley and Shen [4].

In the Bayesian CRM, one must determine *a priori* DLT rates for each dose, referred to as a skeleton, and the first subject is assigned to the dose whose skeleton value is closest to η. Faries [5], Korn *et al.* [6], and Moller [7] proposed modifications to the original CRM to promote patient safety and slow dose escalation. Specifically, the modified CRM suggests that the first patient be assigned to the lowest dose, regardless of the skeleton, and that skipping of doses during dose escalation should not be allowed.

Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.
**Correspondence to: Jin Zhang, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, U.S.A.*
†*E-mail: zhjin@umich.edu*

Numerous extensions to the CRM have been published since the original CRM manuscript, including the time-to-event CRM (TITE-CRM) of Cheung and Chappell [8] to account for incomplete follow-up of patients and the later generalization of Braun [9] to adapt the TITE-CRM for early-onset and late-onset DLTs. Yin and Yuan [10] proposed the Bayesian model averaging CRM to allow for the incorporation of multiple skeletons, and Yuan and Yin [11] developed a hybrid design to combine rule-based methods and the CRM. Lee and Cheung [12, 13] suggested a systematic but computationally intensive approach to calibrate the skeleton through the use of indifference intervals.

Lee and Cheung [13] also proposed methods to determine the value of the variance given to the prior distribution of the parameter in the dose–toxicity model at the onset of the trial. In general Bayesian applications, a large (i.e., vague) prior variance usually connotes a less influential prior distribution, and Chevret [14] suggested the use of a vague prior variance with the Bayesian CRM, although the specific definition of vagueness is controversial. Lee and Cheung [13] proposed a least informative prior variance, defined as the value of the prior variance that results in all doses being *a priori* equally likely of being the MTD. The value of the least informative prior variance tends to be much smaller than what is traditionally considered a vague prior variance.

It is also not appreciated that the level of vagueness of the prior variance is dependent upon the values selected for the skeleton. The aggressive behavior of the CRM in the case studies of Moller [7] and Neuenschwander *et al.* [15] can be entirely explained by the dependence between the prior variance and the skeleton, so that the prior variance used in each study was too small for the chosen skeleton. As a specific example, O'Quigley *et al.* [3] suggested the use of a standard exponential distribution as a vague prior distribution, but the prior variance may still be too small in some settings. Consider two skeletons for five dose levels: the original skeleton used by O'Quigley *et al.* [3] and a skeleton developed using the methods of Lee and Cheung [13]. Both skeletons specify the third as the MTD. The target probability is 0.20, the true MTD is dose 6, the maximum number of enrolled patients is 25, and the dose–response model is the hyperbolic model defined by O'Quigley *et al.* [3]. From the results presented in Scenario 1 in Table I, we see a notable difference in the dose selected as the MTD under these two skeletons even though both use the same prior distribution for the parameter. The prior variance works well with the second skeleton but may be too small for the first skeleton. If we increase the prior variance by using a Gamma distribution with mean 1 and variance 4, the first skeleton now gives results comparable with the second skeleton. Hence, the vagueness of a prior variance heavily depends on the skeleton used.

For a specific skeleton used in a trial, the choice of the prior variance also depends on the relative location of the true MTD and the MTD defined by the skeleton. If the MTD specified by the skeleton is close to the true MTD, a small prior variance could help find the correct MTD more efficiently. However, if the skeleton does not match the truth well, then a larger prior variance is needed to help find the MTD. In Table I, for the same Skeleton A, we find that a larger prior variance works better when the true MTD is dose 6 and that a smaller prior variance works better when the true MTD is dose 4. For both scenarios, the *a priori* MTD is dose 3. Again, the vagueness of a prior variance depends on the skeleton used.

Table I. A simulation study comparing the impact of the prior variance and skeleton on the ability to identify the MTD for the traditional continual reassessment method with a fixed prior variance.

Scenario	Skeleton	Prior		1	2	3	4	5	6
1			DLT rates:	0.00	0.00	0.03	0.05	0.11	0.22
	A	*Exp*(1)		0	0	0	6	65	**29**
	B	*Exp*(1)		0	0	0	4	36	**60**
	A	*Gamma*(1/4, 4)		0	0	0	3	36	**61**
2			DLT rates:	0.02	0.05	0.10	0.20	0.30	0.50
	A	*Exp*(1)		0	2	20	**56**	23	0
	A	*Gamma*(1/4, 4)		0	2	18	**47**	31	1
	B	*Exp*(1)		0	2	22	**50**	23	2
	B	*Gamma*(1/4, 4)		0	2	20	**49**	26	3

Numbers 1–6 in the first row stand for doses 1 to 6. Other numbers in the table stand for the proportion of simulations that select each dose as the maximum tolerated dose (MTD). Skeleton A denotes the original skeleton used by O'Quigley *et al.*: {0.05, 0.10, **0.20**, 0.30, 0.50, 0.70}. Skeleton B denotes the skeleton used by Lee and Cheung: {0.05, 0.11, **0.20**, 0.31, 0.42, 0.53}. Numbers in bold indicate which dose is the MTD. DLT, dose-limiting toxicity.

To achieve similar performance with the use of Skeleton A and standard exponential distribution as the prior, one might need to further reduce the prior variance used for Skeleton B. It seems that the CRM using a constant prior variance could perform well in specific scenarios but might not perform well in other scenarios, no matter which value is selected for the constant prior variance.

The aforementioned motivating example indicates that the traditional methods to calibrate the prior variance may not work well in many scenarios, because the traditional approaches try to find a prior variance based on the skeleton at the onset of a trial and keep it constant during a trial but fail to take into account the relative location of the true MTD and *a priori* MTD for the specific scenario. However, the accumulating data during a trial might provide us information regarding the relative distance between the truth and the skeleton. Hence, we consider the prior variance as a tuning parameter that should be adaptively calibrated during the entire study to determine whether or not the variance chosen at the beginning of the study should be modified. In this paper, we introduce three systematic approaches for adaptively calibrating the prior variance throughout a Phase I trial. In Section 2, we review the CRM and the work of Lee and Cheung [13]. In Section 3, we present the details for our three variance calibration approaches. In Section 4, we apply our methods to two hypothetical settings and compare operating characteristics with current approaches. In Section 5, we conclude with some discussion.

2. Existing methods

2.1. Continual reassessment method

Under the assumption that the probability of DLT increases monotonically with dose, the CRM procedure could update the dose–response relationship as new observations are available through the trial. Patients are assigned to the dose whose estimated DLT rate is closest to the target probability η, subject to possible restrictions. Let J denote the number of doses examined, and let N denote the number of subjects enrolled by the end of the trial. For each dose j, $j = 1, \ldots, J$, there is a skeleton value p_j, denoting the *a priori* DLT rate for dose j. The response y_i of patient i is binary: $y_i = 1$ if there is DLT, or $y_i = 0$ if there is no DLT, $i = 1, \ldots, N$. The CRM uses a one-parameter model given by $\pi_i = \psi(x_i; \beta)$, where β is some unknown parameter, ψ is a monotonic function with the range [0, 1], and x_i denotes the rescaled value of the assigned dose for subject i. In this paper, we consider two commonly used models: (1) a logistic model with intercept 3 given by $\psi(x_i; \beta) = 1/\{1 + \exp[-3 - \exp(\beta)x_i]\}$ and (2) a power model given by $\psi(x_i; \beta) = x_i^{\exp(\beta)}$. In both models, we place a normal prior on β, with mean zero and variance σ^2.

The rescaling of doses attempts to mirror the investigators' prior assumptions and provides a good fit over the skeleton probabilities for the dose levels under the study [3]. Specifically, x_i can take one of the rescaled values x_j^* that are determined from the equations

$$p_j = \int \psi\left(x_j^*; \beta\right) g(\beta) \mathrm{d}\beta, \qquad j = 1, \ldots, J,$$

where $g(\beta)$ is the prior distribution for β. In practice, this computation is replaced with the simplified formula

$$x_j^* = \psi^{-1}_{\beta = E(\beta)}(p_j). \tag{1}$$

Let $\mathbf{Y}_n = \{y_1, \ldots, y_n\}$ denote the observed DLT responses for subjects $1, \ldots, n$, $1 \leqslant n \leqslant N$, after subject n has completed follow-up for DLT. Then the likelihood function for $\mathbf{Y}_n$ is given by

$$L(\mathbf{Y}_n|\beta) = \prod_{i=1}^{n} \{\psi(x_i; \beta)\}^{y_i} \{1 - \psi(x_i; \beta)\}^{1-y_i}.$$

By Bayes' theorem, the posterior mean of the DLT rate at dose d_j, given the observed data, is given by

$$\tilde{\pi}_j = E\left(\psi\left(\beta; x_i = x_j^*\right)|\mathbf{Y}_n\right) = \int \psi\left(\beta; x_i = x_j^*\right) \frac{L(\mathbf{Y}_n|\beta)g(\beta)}{\int L(\mathbf{Y}_n|\beta)g(\beta)\mathrm{d}\beta} \mathrm{d}\beta.$$

In practice, the plug-in estimator, $\psi\left(x_j^*; \tilde{\beta}\right)$ where $\tilde{\beta} = E(\beta|\mathbf{Y}_n)$, is commonly used to simplify the calculation for $\tilde{\pi}_j$. On the basis of the updated posterior DLT rates $\tilde{\pi}_j$, $j = 1, \ldots, J$, the recommended

dose for the next patient is chosen as the one with a DLT rate closest to the target η. So the next subject is assigned to dose level j such that

$$j = \underset{j\in(1,\ldots,J)}{\arg\min} \ |\tilde{\pi}_j - \eta|. \tag{2}$$

The CRM usually does not allow dose skipping during dose escalation. The trial either progresses until the total number of subjects N is reached or is terminated if a certain stopping rule is satisfied. The MTD is determined at the end of the trial by simply selecting dose j according to (2) based upon $\mathbf{Y}_N$.

To address the ethical concern of overdosing subjects, many authors have developed stopping rules for dose-finding studies that halt a study if all doses under study are too toxic, including Korn *et al.* [6], O'Quigley [16], and O'Quigley and Reiner [17]. In our simulations presented in Section 4, we used a variant of the stopping rule proposed by Thall and Russell [18], in which the trial is stopped and no dose is selected as the MTD once the posterior probability that the DLT rate of the lowest dose is higher than the target probability is larger than a pre-specified value.

2.2. *Least informative prior variance of Lee and Cheung* [13]

We first briefly review the concept of indifference intervals proposed by Cheung and Chapell [19] in the context of the CRM. The parameter space of β can be divided into J intervals: $I_1 = [b_l, b_1)$, $I_j = (b_j, b_{j+1})$ for $j = 1, \ldots, J-2$, and $I_J = (b_{J-1}, b_u)$, where $b_1, \ldots, b_{J-1}$ are solved from

$$\psi\left(x_j^*; b_j\right) + \psi\left(x_{j+1}^*; b_j\right) = 2\eta, \quad \text{for} \quad j = 1, \ldots, J-1.$$

It is obvious that the CRM would assign dose j to the next subject if and only if the estimate $\tilde{\beta}$ falls in the interval I_j, $j = 1, \ldots, J$. Although $\beta \in (-\infty, \infty)$, finite values for b_l and b_u are used in practice to avoid computational difficulty.

The least informative prior variance, denoted as σ_{LI}^2, is the prior variance that results in β being equally likely of belonging to any of the J intervals, that is, all doses being *a priori* equally likely of being the MTD. These J probabilities can be regarded as being from a discrete uniform distribution, although it is usually not possible to make them exactly equal. Instead, Lee and Cheung [13] defined σ_{LI}^2 as the prior variance such that the variance of the J probabilities matches $(J^2-1)/12$, the variance of a discrete uniform distribution. Although σ_{LI}^2 is uninformative in terms of the prior model-based MTD distribution, the value of σ_{LI}^2 is usually not large with respect to what is usually considered to be an uninformative variance.

For example, in the setting where there are five dose levels, the skeleton is $\{0.05, 0.10, 0.20, 0.35, 0.50\}$, the target η is 0.20, and a logistic model with intercept 3 is used; the resulting five intervals of β in which doses 1 to 5 are the MTD are $I_1 = (-\infty, -0.23)$, $I_2 = (-0.23, -0.08)$, $I_3 = (-0.08, 0.10)$, $I_4 = (0.10, 0.29)$, and $I_5 = (0.29, \infty)$, respectively. The least informative prior variance σ_{LI}^2 is 0.32^2, which would usually be regarded as an informative prior variance in general Bayesian applications.

3. Methods for adaptive variance calibration

3.1. *Defining a large prior variance* σ_{HI}^2

When the MTD defined by the skeleton is not the first or last dose, a prior variance larger than σ_{LI}^2 would result in a U-shaped distribution of the *a priori* model-based MTD [13]. As a result, dose 1 and J would be more likely to be selected as the MTD. Hence, σ_{LI}^2 could perform poorly when the MTD is the lowest or highest dose and the MTD defined by the skeleton lies elsewhere, at least when no stopping rule is used. Therefore, we further define a larger prior variance, σ_{HI}^2, as the prior variance that satisfies $Pr(\beta \in I_1 \cup I_J) = 0.8$, producing a U-shaped distribution for the model-based MTD. Presumably, σ_{HI}^2 could perform well when σ_{LI}^2 performs poorly. A value other than 0.80 can certainly be used to determine the value of σ_{LI}^2 . However, values larger than 0.80 will place more mass in the tails of the MTD distribution and may be too aggressive in situations when the MTD is not the highest dose. Conversely, values smaller than 0.80 will place less mass in the tails of the MTD distribution and will lessen the ability to find the MTD when it is the highest dose. We found that 0.80 was a good compromise between these two situations.

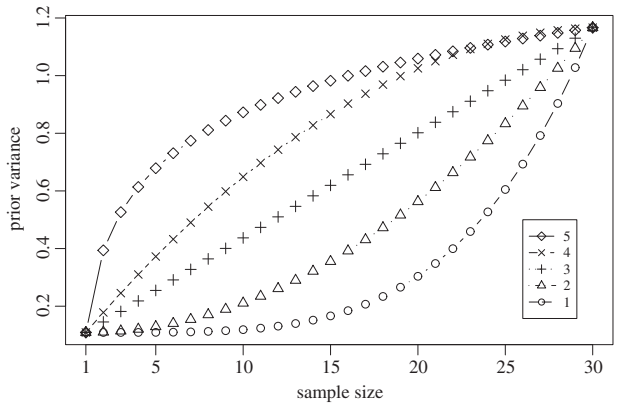

Figure 1. CRM-VC1: the prior variance increases with the sample size in five different patterns.

3.2. CRM-VC1: increasing the prior variance with the sample size

We denote CRM-VC1 as our first approach to adaptively calibrate the prior variance in the CRM. Because the sample size is small early in a trial, it may be appropriate to use σ_{LI}^2 at the beginning of a trial so that each dose is *a priori* equally likely to be selected as the MTD. However, it would not be desirable for the prior to dominate the data [20], especially when the MTD is the lowest or highest dose. Hence, a sufficiently large prior may be preferred later in a trial. One natural approach is to start the prior variance at σ_{LI}^2 and increase it to σ_{HI}^2 at a rate based upon n, the number of currently enrolled patients. We have selected five different functions explaining how the prior variance increases with n so that different rates of change could be captured:

1. $\sigma_n^2 = \sigma_{LI}^2 + \left(\sigma_{HI}^2 - \sigma_{LI}^2\right)(n-1)^4/(N-1)^4$
2. $\sigma_n^2 = \sigma_{LI}^2 + \left(\sigma_{HI}^2 - \sigma_{LI}^2\right)(n-1)^2/(N-1)^2$
3. $\sigma_n^2 = \sigma_{LI}^2 + \left(\sigma_{HI}^2 - \sigma_{LI}^2\right)(n-1)/(N-1)$
4. $\sigma_n^2 = \sigma_{LI}^2 + \left(\sigma_{HI}^2 - \sigma_{LI}^2\right)\log(2n-1)/\log(2N-1)$
5. $\sigma_n^2 = \sigma_{LI}^2 + 2N\left(\sigma_{HI}^2 - \sigma_{LI}^2\right)(n-1)/(N^2-1) - \left(\sigma_{HI}^2 - \sigma_{LI}^2\right)(n-1)^2/(N^2-1)$.

Figure 1 displays the five different patterns when $N = 30$, $\sigma_{LI} = 0.33$, and $\sigma_{HI} = 1.08$, a setting we will further explore in our simulations. These five functions represent typical variance–sample size relationships: (a) the prior variance increases slowly at first and then quickly reaches σ_{HI}^2; (b) the prior variance increases with n with a constant rate; and (c) the prior variance increases quickly at first and slowly reaches σ_{HI}^2.

3.3. CRM-VC2: a hypothesis testing approach

If we start with a certain prior variance in the CRM, it would be ideal if the accumulating data could help determine whether the current prior variance should change. If the skeleton specifies the correct MTD, then the prior variance should be small, and the prior information is incorporated to enhance estimation of the MTD. Otherwise, it is preferable to change the prior variance if the data indicate that the skeleton has misidentified the MTD. This is the motivation for CRM-VC2.

A trial based on CRM-VC2 starts with the prior variance σ_{LI}^2. When the data favor the hypothesis that the MTD is the highest dose but the MTD defined by the skeleton lies elsewhere, CRM-VC2 increases the prior variance to σ_{HI}^2, because a large prior variance would increase the probability of selecting the tail dose levels owing to the U-shaped distribution. We do not increase the prior variance if the MTD is dose 1, because the use of a stopping rule makes it unnecessary. However, when the MTD defined by the skeleton coincides with the highest dose, the prior variance determined by CRM-VC2 would remain at σ_{LI}^2 because increasing the prior variance is no longer helpful when the prior information is correct.

The decision to switch from σ_{LI}^2 to σ_{HI}^2 involves a hypothesis testing approach similar to what Yuan and Yin [11] proposed for their hybrid design. We propose three hypotheses: $H_1 : \beta \in I_1$, $H_2 : \beta \in I_2 \cup I_3 \ldots \cup I_{J-1}$, and $H_3 : \beta \in I_J$. We also propose two reasonable bounds b_l and b_u for β to avoid technical difficulties, that is, $\beta \in [b_l, b_u]$ rather than $(-\infty, \infty)$. Specifically, b_l satisfies

$\psi(x_1, b_l) = \eta + 0.05$, and b_u satisfies $\psi(x_J, b_u) = \eta - 0.05$. Although it is guaranteed that b_l is smaller than b_u for our model parameterization, such a result may not hold true for all models, in which case one would switch b_l with b_u. Via simulation, we also examined using bounds defined by $\eta \pm 0.025$ and $\eta \pm 0.10$ and found little change in the operating characteristics when using $\eta \pm 0.05$ (results not shown). Actually, when the true β falls outside $[b_l, b_u]$, the true DLT rates for all the J doses would be far away from the target η, implying that the doses examined would be either too toxic or overly safe. A trial would hence either be terminated by a stopping rule or quickly find the highest dose as the MTD.

To be objective, we assign a uniform prior distribution under each hypothesis: $\beta|H_1 \sim \text{Unif}[b_l, b_1)$, $\beta|H_2 \sim \text{Unif}[b_1, b_J)$, and $\beta|H_3 \sim \text{Unif}[b_J, b_u]$. The marginal distribution of $\mathbf{Y}_n$ under H_1 is then given by

$$p(\mathbf{Y}_n|H_1) = \int_{b_l}^{b_1} \prod_{i=1}^{n} \{\psi(x_i;\beta)\}^{y_i} \{1 - \psi(x_i;\beta)\}^{1-y_i} \frac{1}{b_1 - b_l} \, \mathrm{d}\beta.$$

Similarly, we can compute $p(\mathbf{Y}_n|H_2)$ and $p(\mathbf{Y}_n|H_3)$. The posterior probability of H_k, $k = 1, 2, 3$, is given by

$$p(H_k|\mathbf{Y}_n) = \frac{p(H_k)p(\mathbf{Y}_n|H_k)}{p(H_1)p(\mathbf{Y}_n|H_1) + p(H_2)p(\mathbf{Y}_n|H_2) + p(H_3)p(\mathbf{Y}_n|H_3)}.$$

If we let $BF_{hk} = p(\mathbf{Y}_n|H_h)/P(\mathbf{Y}_n|H_k)$, $h = 1, 2, 3$, denote the Bayes factor for comparing H_h and H_k, then

$$p(H_k|\mathbf{Y}_n) = \frac{p(H_k)}{p(H_1)BF_{1k} + P(H_2)BF_{2k} + P(H_3)BF_{3k}}.$$

We specify $p(H_1) = P(H_2) = P(H_3) = 1/3$ and use Jeffreys' rule that $\log_{10}(BF_{kk'}) > 1/2$ indicates substantial evidence in favor of H_k against $H_{k'}$ [11, 21]. This rule translates to the criterion that if $p(H_3|\mathbf{Y}_n) > 0.61$, then there is substantial evidence that $\beta \in I_J$. Once such evidence exists, the prior variance would increase to σ^2_{HI}; otherwise, the prior variance stays at σ^2_{LI}.

3.4. CRM-VC3: adaptively changing skeletons

Instead of changing the prior variance during a trial to make the MTD more likely to be selected, CRM-VC3, our third approach to calibrate the prior variance, is to modify the skeleton adaptively but keep the prior variance constant. Consequently, the intervals $I_1, \ldots, I_J$ would also change, because the intervals $I_1, \ldots, I_J$ only depend on the skeleton and the model used. If we can properly adjust these intervals, then more mass of the prior distribution could be placed over the interval that results in selecting the correct MTD.

For CRM-VC3, a trial starts with the prior variance σ^2_{LI}, and once the new estimates for π_j are obtained, the dose values would be rescaled again and used as the new dose values. Let $\tilde{\beta}_n$ denote the posterior mean of β after n subjects have finished follow-up for DLT, and let $x^*_{j,0} = x^*_j$. The updated skeleton p^n_j is set equal to the current DLT probabilities based on $\tilde{\beta}_{(n)}$, that is,

$$p^n_j = \psi\left(x^*_{j,(n-1)}; \tilde{\beta}_{(n)}\right), \tag{3}$$

and similar to (1), the updated rescaled dose value for dose j is

$$x^*_{j,n} = \psi^{-1}_{\beta = E(\beta)}(p^n_j). \tag{4}$$

Skeletons and rescaled dose values are updated according to (3) and (4) during a trial. All other facets of the design, including the model and prior distribution, remain the same. The resulting intervals $I_1 \ldots I_J$ would change adaptively with the updating of $\tilde{\beta}$. Consider the setting where there are five dose levels, the skeleton is $\{0.05, 0.10, 0.20, 0.35, 0.50\}$, the target η is 0.20, and the model used is a logistic model with intercept 3. Figure 2 shows that we could assign more mass to tail areas adaptively if $\tilde{\beta}$ falls in I_1 or I_J. After the first subject is observed, if $\tilde{\beta} = 0$, then the new skeleton will be the same

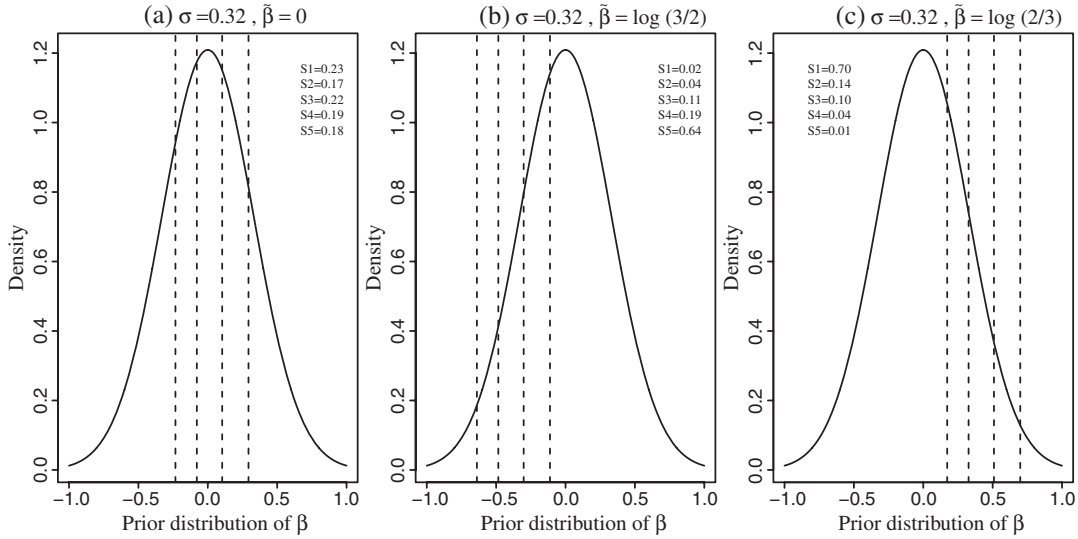

Figure 2. Areas under the Normal prior density curve of β with mean zero and variance $\sigma_{LI}^2 = 0.32^2$ for intervals I_1 to I_5: $S1$ to $S5$. They are also the prior probabilities for selecting each of the five doses as the maximum tolerated dose. Rescaling the doses sequentially could change the area under the curve for each interval. The four vertical lines stand for the boundaries for the five intervals.

with the original skeleton, indicating that the prior information is close to the truth. As a result, the J intervals and the resulting areas under the prior density curve for the J intervals do not change because we are using σ_{LI}^2 as the prior variance (Figure 2(a)). This is reasonable because the data suggest that the MTD does not lie in the tail. If $\tilde{\beta} = \log(3/2)$, suggesting that the MTD is dose 5, then the prior density will place more mass in I_5 after rescaling the dose values with the area under $I_5 = 0.64$ (Figure 2(b)). If $\tilde{\beta} = \log(2/3)$, suggesting that the MTD is dose 1, then the prior density places more mass on I_1 after rescaling the dose values with the area under $I_1 = 0.70$ (Figure 2(c)). As more subjects enter the study, $\tilde{\beta}$ becomes more accurate, and the resulting updated skeleton is driven by the data, avoiding the effect that a misspecified skeleton would have in the conventional CRM. One may be concerned that the updated skeleton values may be unstable early in the study when little data exist, and restrict the use of CRM-VC3 after a minimum sample size has been accrued. However, restricting any skipping of doses during escalation will alleviate any possible instability. Furthermore, we examined the mean dose assigned to the first 10 subjects in the settings presented in Section 4 (results not shown) and found that using CRM-VC3 was no more or less stable than the other methods.

4. Simulation results

4.1. Rules used in simulations

We used a cohort size of one subject in our study. Like most dose-finding studies, we restrict dose escalation to be no more than one dose above the assignment of the most recent subject. However, we do not impose any restriction on dose de-escalation. Also, the first subject is always assigned to the lowest dose. However, in the simulations of the hypothetical trial of Lee and Cheung [13], we assign the third dose to the first subject to make our results comparable with theirs.

The prior variance (CRM-VC1 and CRM-VC2) or skeleton (CRM-VC3) will be updated after a new subject finishes follow-up for DLT. For patient safety, we will stop the trial if at least two out of the first three patients experience DLT or if $Pr(\pi_1 > \eta|\mathbf{Y}_n) > 0.9$ after four or more patients have been enrolled. To reduce the sensitivity to the prior variance and instability due to small sample size, our stopping rule mimics the 3 + 3 method for the first three subjects and then switches to the use of the posterior probability that the DLT rate for the lowest dose is above the target. We found that the threshold of 0.9 works well in simulations but could be adjusted, depending upon how great the need for early stopping is. We performed 2000 simulations in each scenario; all simulations were performed in the statistical package `R` [22], the code for which is available upon request.

4.2. A hypothetical trial

In our hypothetical clinical trial of $N = 30$ subjects with five dose levels and the target DLT rate $\eta = 0.20$, we used the logistic model with intercept 3. The prior distribution for β was normal with mean 0 and variance σ_n^2. We considered two commonly used skeletons that specify the *a priori* MTDs to be the middle dose and the highest dose. Specifically, Skeleton 1 is $\{0.05, 0.10, 0.20, 0.35, 0.5\}$ ($\sigma_{LI} = 0.32; \sigma_{HI} = 1.04$), and Skeleton 2 is $\{0.01, 0.04, 0.07, 0.11, 0.20\}$ ($\sigma_{LI} = 0.35; \sigma_{HI} = 0.68$).

Table II. Simulation study comparing CRM-VC1, CRM-VC2, and CRM-VC3 with the traditional CRM under Skeleton 1: $\{0.05, 0.10, 0.20, 0.35, 0.5\}$.

			Percentage of simulations selected as MTD						Mean number of subjects assigned				
Scenario			None	1	2	3	4	5	1	2	3	4	5
1	Pr(DLT)			**0.20**	0.30	0.35	0.45	0.50					
	CRM	σ_{LI}^2	17	**44**	29	9	1	0	11	9	5	1	0
		σ_{HI}^2	32	**36**	23	8	1	0	11	6	3	1	1
	CRM-VC1	1	23	**47**	21	7	1	0	12	8	4	2	0
		2	26	**43**	24	9	1	0	11	9	5	1	0
	CRM-VC2		18	**44**	28	9	1	0	10	8	6	2	0
	CRM-VC3		34	**36**	20	8	1	0	10	6	4	1	1
2	Pr(DLT)			0.10	**0.20**	0.35	0.45	0.50					
	CRM	σ_{LI}^2	4	15	**62**	16	1	0	6	14	7	2	0
		σ_{HI}^2	6	18	**56**	19	2	0	8	12	6	2	1
	CRM-VC1	1	4	22	**57**	22	2	0	7	13	7	2	0
		2	4	20	**56**	19	1	0	7	13	7	2	0
	CRM-VC2		4	15	**61**	19	1	0	6	14	7	2	0
	CRM-VC3		9	18	**55**	17	2	0	8	12	6	2	1
3	Pr(DLT)			0.05	0.10	**0.20**	0.35	0.45					
	CRM	σ_{LI}^2	0	1	19	**66**	13	0	2	7	15	5	0
		σ_{HI}^2	2	1	19	**59**	18	1	3	7	12	6	2
	CRM-VC1	1	0	1	23	**61**	14	1	2	7	14	6	1
		2	2	1	19	**60**	18	1	2	7	13	6	1
	CRM-VC2		0	1	20	**65**	14	1	2	7	15	5	1
	CRM-VC3		2	1	20	**61**	15	1	3	7	12	6	2
4	Pr(DLT)			0.02	0.05	0.10	**0.20**	0.35					
	CRM	σ_{LI}^2	0	0	1	28	**61**	10	1	2	10	14	3
		σ_{HI}^2	0	0	1	23	**60**	16	2	3	7	12	7
	CRM-VC1	1	0	1	1	25	**58**	16	1	2	9	14	4
		2	0	0	1	21	**61**	16	1	2	8	13	5
	CRM-VC2		0	0	1	27	**56**	16	1	2	9	12	5
	CRM-VC3		0	0	1	23	**60**	15	2	3	7	12	7
5	Pr(DLT)			0.01	0.04	0.07	0.11	**0.20**					
	CRM	σ_{LI}^2	0	0	0	7	37	**56**	1	2	5	12	11
		σ_{HI}^2	0	0	0	4	27	**69**	1	2	3	7	16
	CRM-VC1	1	0	0	0	6	28	**66**	1	2	4	11	12
		2	0	0	0	4	27	**68**	1	2	4	9	14
	CRM-VC2		0	0	0	6	26	**68**	1	2	4	8	14
	CRM-VC3		0	0	0	5	25	**70**	1	2	3	7	17

'None' denotes the proportion of trials that stop early. Numbers 1–5 in the first row stand for doses 1 to 5. Numbers in bold indicate which dose is the maximum tolerated dose (MTD). CRM, continual reassessment method; DLT, dose-limiting toxicity.

We examined the performances of CRM-VC1, CRM-VC2, and CRM-VC3 in five different scenarios under both Skeletons 1 and 2 based on the percentage of simulations selected as the MTD and the average number of patients assigned to each dose. The true MTDs are doses 1 to 5 for scenarios 1 to 5, respectively. We also performed the traditional CRM using the fixed prior variances σ^2_{LI} and σ^2_{HI} in the same five scenarios for comparison.

We summarize the performances under each scenario in the final 11 columns of Tables II and III for Skeletons 1 and 2, respectively. The first five of the 11 columns display the percentage of simulations in

Table III. Simulation study comparing CRM-VC1, CRM-VC2, and CRM-VC3 with the traditional CRM under Skeleton 2: {0.01, 0.04, 0.07, 0.11, 0.20}.

			Percentage of simulations selected as MTD						Mean number of subjects assigned				
Scenario			None	1	2	3	4	5	1	2	3	4	5
1	Pr(DLT)			**0.20**	0.30	0.35	0.45	0.50					
	CRM	σ^2_{LI}	18	**41**	30	8	1	0	9	9	4	2	1
		σ^2_{HI}	26	**44**	24	5	1	0	13	7	3	2	1
	CRM-VC1	1	21	**46**	26	6	1	0	10	8	4	2	1
		2	22	**45**	26	8	2	0	10	9	4	2	1
	CRM-VC2		18	**41**	30	8	1	0	9	9	4	2	1
	CRM-VC3		31	**43**	20	5	1	0	12	6	3	1	1
2	Pr(DLT)			0.10	**0.20**	0.35	0.45	0.50					
	CRM	σ^2_{LI}	6	16	**60**	17	2	0	5	13	7	3	1
		σ^2_{HI}	5	20	**57**	14	2	0	8	12	5	2	1
	CRM-VC1	1	6	20	**57**	16	2	0	5	13	6	3	1
		2	6	16	**57**	17	3	1	6	13	6	3	1
	CRM-VC2		6	16	**60**	17	2	0	5	13	7	3	1
	CRM-VC3		9	21	**55**	14	2	0	9	11	5	2	1
3	Pr(DLT)			0.05	0.10	**0.20**	0.35	0.45					
	CRM	σ^2_{LI}	1	1	22	**54**	21	2	2	7	11	7	3
		σ^2_{HI}	1	2	25	**49**	22	2	3	8	10	6	3
	CRM-VC1	1	1	1	29	**49**	20	2	2	8	11	7	3
		2	2	1	26	**49**	20	2	2	8	11	7	3
	CRM-VC2		1	1	22	**54**	21	2	2	7	11	7	3
	CRM-VC3		2	2	29	**50**	17	1	3	8	10	6	3
4	Pr(DLT)			0.02	0.05	0.10	**0.20**	0.35					
	CRM	σ^2_{LI}	0	0	3	22	**58**	18	1	2	6	12	8
		σ^2_{HI}	0	0	2	21	**55**	22	2	3	6	11	9
	CRM-VC1	1	0	0	3	22	**58**	18	1	2	6	12	8
		2	0	0	2	23	**54**	21	1	2	6	12	9
	CRM-VC2		1	0	3	22	**58**	18	1	2	6	12	8
	CRM-VC3		0	0	3	24	**55**	18	2	3	7	11	8
5	Pr(DLT)			0.01	0.04	0.07	0.11	**0.20**					
	CRM	σ^2_{LI}	0	0	0	2	23	**75**	1	1	2	7	19
		σ^2_{HI}	0	0	0	2	23	**75**	1	2	2	6	18
	CRM-VC1	1	0	0	0	3	23	**73**	1	1	2	7	18
		2	0	0	0	2	22	**75**	1	1	2	6	19
	CRM-VC2		0	0	0	2	23	**75**	1	1	2	7	19
	CRM-VC3		0	0	1	4	23	**73**	1	2	3	6	18

'None' denotes the proportion of trials that stop early. Numbers 1–5 in the first row stand for doses 1 to 5. Numbers in bold indicate which dose is the maximum tolerated dose (MTD). CRM, continual reassessment method; DLT, dose-limiting toxicity.

which each dose was identified as the MTD at the end of the study, and the last five columns display the average number of patients assigned to each dose. For each scenario, we list the true toxicity probabilities in the first row, the results for the CRM using the fixed prior variance σ_{LI}^2 and σ_{HI}^2 in rows 2 and 3, the results obtained by CRM-VC1 using functions (1) and (2) in Section 3.2 in rows 4 and 5, the results obtained by CRM-VC2 in row 6, and the results obtained by CRM-VC3 in row 7. We do not present the results for CRM-VC1 using functions (3) to (5) as they did no better or worse than functions (1) and (2). We note that in general, we have found that all five functions perform similarly in terms of finding the correct MTD, indicating that there is no real need to choose among them in application.

In Table II, we note that the traditional CRM is sensitive to the value of the prior variance. Using the prior variance σ_{LI}^2 (a similar approach with Method A1 in [13]) performs better than using σ_{HI}^2 in Scenarios 2, 3, and 4, where the true MTD is close to the MTD defined by the skeleton. But using σ_{LI}^2 performs poorly relative to σ_{HI}^2 in Scenario 5, where the true MTD is the highest dose but the MTD defined by the skeleton is dose 3. Overall, the CRM using σ_{HI}^2 is more robust than using σ_{LI}^2 in finding the MTD except when the true MTD is at or close to the skeleton MTD. We also see that using a large prior variance could produce more unnecessary early stopping. For example, in Scenario 1, using σ_{HI}^2 results in 44% early stopping compared with 36% when using σ_{LI}^2. This is why the prior variance does not increase to σ_{HI}^2 in CRM-VC2 when there is evidence that the true MTD is dose 1.

When the true MTD is similar to that specified by the skeleton, CRM-VC1 gives comparable results with the traditional CRM using σ_{LI}^2 but performs much better when the true MTD is the highest dose. Compared with the traditional CRM using the prior variance σ_{HI}^2, CRM-VC1 performs slightly better in Scenarios 1, 2, and 3 and has comparable performance in Scenarios 4 and 5. CRM-VC2 performs consistently well overall, even though it performs slightly worse than other competing methods in Scenario 4. CRM-VC2 performs as well as the CRM using the prior variance σ_{LI}^2 in Scenarios 1, 2, and 3, but CRM-VC2 performs much better in Scenario 5, where the true MTD is dose 5. Compared with the traditional CRM using σ_{HI}^2, CRM-VC2 also demonstrates a better ability in identifying the MTD: 61% versus 56% in Scenario 2 and 65% versus 59% in Scenario 3. CRM-VC3 performs similarly with the traditional CRM using the prior variance σ_{HI}^2.

We also see that the design giving a higher probability of selecting the MTD also assigns more patients to the correct dose. Hence, similar results are obtained if we compare mean dose assignments among the methods examined. Overall, CRM-VC2 performs best among all the methods examined across the five scenarios.

It is also common in practice that the MTD determined by the skeleton is the highest dose. Skeleton 2 is one such skeleton. When using Skeleton 2 while keeping other facets of the design unchanged, we notice that the results presented in Table III are quite similar among the five scenarios for most of the methods examined. However, CRM-VC2 and the traditional CRM using prior variance σ_{LI}^2 slightly outperform other methods across the five scenarios.

4.3. A hypothetical trial by Lee and Cheung [13]

In this setting, there are $N = 25$ subjects and six dose levels, and the target DLT rate is $\eta = 0.20$. The first patient is assigned to dose 3, and the model used is the power model. To make our results comparable with those of Lee and Cheung, we used the skeleton $\{0.05, 0.11, 0.20, 0.31, 0.42, 0.53\}$ $(\sigma_{LI} = 0.68; \sigma_{HI} = 2.45)$ used with Method A1 of Lee and Cheung, which we denote as LC-A1. Table IV shows results obtained from the traditional CRM, LC-A1, CRM-VC1, CRM-VC2, and CRM-VC3 in the similar scenarios examined by Lee and Cheung. Note that LC-A1 is equivalent to the traditional CRM using the prior variance σ_{LI} in this example because the skeleton is the same for both methods.

In Scenario 1, all the methods work similarly. The traditional CRM using prior variance σ_{HI}^2 results in more trials being terminated than using σ_{LI}^2. In Scenarios 2 and 3, LC-A1, CRM-VC2, and the traditional CRM using σ_{LI}^2 perform slightly better than other approaches; all three approaches correctly identify the MTD in approximately 55% of simulations in contrast to 49–53% for the other approaches. However, LC-A1 and the traditional CRM using σ_{LI}^2 perform poorly relative to CRM-VC2 in Scenario 5 where the MTD is the highest dose.

As seen earlier, the traditional CRM using σ_{HI}^2 does not perform as well as σ_{LI}^2 in Scenarios 2 and 3 where the true MTD is close to the MTD defined by the skeleton, even though the difference is not large. CRM-VC1 and CRM-VC2 perform as well as or better than LC-A1 in Scenarios 2,3 and 4 but performs better than LC-A1 in Scenario 5. CRM-VC3 performs similarly with the traditional CRM using σ_{HI}^2.

Table IV. Simulation study comparing CRM-VC1, CRM-VC2, and CRM-VC3 with the method LC-A1 and the traditional CRM in scenarios examined by Lee and Cheung.

			Percentage of simulations selected as MTD							Mean number of subjects assigned					
Scenario			None	1	2	3	4	5	6	1	2	3	4	5	6
1	Pr(DLT)			**0.20**	0.35	0.50	0.61	0.76	0.87						
	CRM	σ^2_{LI}	37	**44**	18	1	0	0	0	8	6	3	1	0	0
		σ^2_{HI}	43	**43**	14	1	0	0	0	11	4	4	2	1	0
	LC-A1		37	**44**	18	1	0	0	0	8	6	3	1	0	0
	CRM-VC1	1	44	**43**	13	1	0	0	0	8	5	3	1	0	0
		2	45	**42**	14	0	0	0	0	8	5	3	1	0	0
	CRM-VC2		39	**44**	16	1	0	0	0	8	6	3	0	0	0
	CRM-VC3		41	**44**	14	1	0	0	0	10	4	3	1	0	0
2	Pr(DLT)			0.05	0.10	**0.20**	0.30	0.50	0.70						
	CRM	σ^2_{LI}	5	1	21	**51**	22	1	0	1	5	11	6	1	0
		σ^2_{HI}	4	3	22	**49**	22	1	0	3	5	9	5	2	1
	LC-A1		5	1	21	**51**	22	1	0	1	5	11	6	1	0
	CRM-VC1	1	4	2	22	**47**	22	2	0	1	5	11	6	1	0
		2	5	1	22	**48**	22	2	0	1	6	10	6	1	0
	CRM-VC2		4	1	20	**52**	21	2	0	1	5	11	5	1	0
	CRM-VC3		4	2	22	**48**	22	1	0	2	5	10	5	2	1
3	Pr(DLT)			0.06	0.08	0.12	**0.18**	0.40	0.71						
	CRM	σ^2_{LI}	3	0	4	27	**55**	11	0	0	2	8	10	3	0
		σ^2_{HI}	2	1	7	27	**49**	14	0	2	2	7	9	4	1
	LC-A1		3	0	4	27	**55**	11	0	0	2	8	10	3	0
	CRM-VC1	1	3	1	5	23	**53**	16	0	0	2	8	10	3	0
		2	3	0	5	25	**52**	15	0	1	2	8	10	4	0
	CRM-VC2		3	0	5	27	**54**	12	0	0	2	9	10	3	1
	CRM-VC3		2	1	6	26	**52**	14	0	1	2	8	9	4	1

Table IV. *Comtinued.*

Scenario			Percentage of simulations selected as MTD							Mean number of subjects assigned					
			None	1	2	3	4	5	6	1	2	3	4	5	6
4	Pr(DLT)			0.05	0.06	0.08	0.11	**0.19**	0.34						
	CRM	σ^2_{LI}	1	0	1	8	33	**46**	11	0	1	5	8	8	3
		σ^2_{HI}	1	0	2	6	23	**47**	20	1	1	3	6	8	6
	LC-A1		1	0	1	8	33	**46**	11	0	1	5	8	8	3
	CRM-VC1	1	2	0	1	7	23	**45**	22	0	1	5	7	8	4
		2	1	0	1	7	25	**45**	20	0	1	5	7	8	5
	CRM-VC2		2	0	1	8	26	**43**	21	0	1	5	7	6	5
	CRM-VC3		1	0	2	7	23	**46**	21	1	1	4	5	7	6
5	Pr(DLT)			0	0	0.03	0.05	0.11	**0.22**						
	CRM	σ^2_{LI}	0	0	0	0	7	43	**50**	0	0	3	5	9	8
		σ^2_{HI}	0	0	0	0	4	30	**65**	0	0	2	3	7	13
	LC-A1		0	0	0	0	7	43	**50**	0	0	3	5	9	8
	CRM-VC1	1	0	0	0	0	4	29	**66**	0	0	3	4	8	10
		2	0	0	0	0	4	31	**64**	0	0	3	4	7	11
	CRM-VC2		0	0	0	0	6	29	**65**	0	0	3	5	6	12
	CRM-VC3		0	0	0	0	4	28	**68**	0	0	3	3	6	13

'None' denotes the proportion of trials that stop early. Numbers 1–6 in the first row denote doses 1 to 6. Numbers in bold indicate which dose is the maximum tolerated dose (MTD). CRM, continual reassessment method; DLT, dose-limiting toxicity; LC-A1, Method A1 of Lee and Cheung.

We also see that the design giving a higher probability of selecting the MTD also assigns more patients to the correct dose. Hence, similar results are obtained if we compare mean dose assignments among the methods examined. Overall, in this setting, CRM-VC2 and CRM-VC3 seem to perform best among all the methods examined in terms of the ability of identifying the MTD across the five scenarios.

5. Discussion

In the present paper, we relax the assumption of a fixed prior variance in the traditional CRM and propose three systematic approaches to adaptively calibrate the prior variance continually throughout the trial. Our approaches have the ability to perform better than the traditional CRM using a constant prior variance as well as methods that calibrate the prior variance only at the beginning of the trial.

Although Lee and Cheung [13] suggested the use of σ^2_{LI} after first calibrating the skeleton at the beginning of the trial, this approach does not perform well when the true MTD is far away from the MTD defined by the skeleton. Although Lee and Cheung proposed an alternate, computationally intensive design, which we refer to as LC-A2, we found that LC-A2 generally offers no improvement to the results of LC-A1. Our approaches, however, are able to improve upon the results of LC-A1 in scenarios where the MTD is the highest dose, without sacrificing the performance much in other scenarios, and they are less computationally expensive than LC-A2. However, as seen in our simulation results, our methods might be more aggressive than the CRM using σ^2_{LI} in certain scenarios. Nonetheless, CRM-VC2 and CRM-VC3 can be modified to be less conservative by simply changing some of the thresholds used in those methods. For example, we could increase the threshold of 0.61 for the posterior probability of H_3 in CRM-VC2 to a larger value, in which case the performance would be similar to that of Lee and Cheung. However, we note that there is no one value for the threshold that will work best in all scenarios, and we feel that our threshold is a good choice in most scenarios.

One reviewer questioned whether or not our designs are coherent in the sense that dose escalation is possible when the most recent patient experiences a DLT [23]. In the scenarios of Section 4.3, escalation after a DLT never occurred, and in the scenarios of Section 4.2, the dose escalation never occurred after an observed DLT in more than 3% of simulations. Thus, although there is no guarantee of coherence of our designs in all settings, any deviation from coherence is quite small, and patient safety is not compromised.

Our approaches could be extended to accommodate a wider range of applications. CRM-VC1 could be easily applied to more complex studies, including finding the most successful dose (MSD) or the most tolerated schedule, once we determine σ^2_{LI} and σ^2_{HI}. CRM-VC2 and CRM-VC3 rely on the skeleton and hence could be naturally extended to a study where a skeleton is specified and the dose values are rescaled, for example, modeling the toxicity in the study of finding the MSD. For a model with more than two parameters, CRM-VC2 and CRM-VC3 are still applicable even though it may be hard to find the indifference regions in high-dimensional parameter spaces.

Acknowledgements

This research was supported by the National Institute of Health (grant no. 5R01CA148713). We thank the editor, the associate editor, and two referees for their constructive and thoughtful reviews that help the manuscript significantly.

References

1. Storer B. Consistency of continual reassessment method under model misspecification. *Biometrics* 1989; **45**:925–937. DOI: 10.2307/2531693.
2. Rosenberger W, Haines L. Competing designs for Phase I clinical trials: a review. *Statistics in Medicine* 2009; **28**:3012–3028. DOI: 10.1002/sim.1229.
3. 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. DOI: 10.2307/2531628.
4. O'Quigley J, Shen L. Continual reassessment method: a likelihood approach. *Biometrics* 1996; **52**:673–684. DOI: 10.2307/2532905.
5. Faries D. Practical modifications of the continual reassessment method for Phase I cancer clinical trials. *Journal of Biopharmaceutical Statistics* 1994; **4**:147–164. DOI: 10.1080/10543409408835079.
6. Korn EL, Midthune D, Chen TT, Rubinstein LV, Christain MC, Simon R. A comparison of two phase I trial designs. *Statistics in Medicine* 1994; **13**:1799–1806. DOI: 10.1002/sim.4780131802.

7. Moller S. An extension of the continual reassessment methods using a preliminary up-and-down design in a dose-finding study in cancer patients, in order to investigate a greater range of doses. *Statistics in Medicine* 1995; **14**:911–22. DOI: 10.1002/sim.4780140909.
8. Cheung YK, Chappell R. Sequential designs for Phase I clinical trials with late-onset toxicity. *Biometrics* 2000; **56**:1177–1182. DOI: 10.1111/j.0006-341X.2000.01177.x.
9. Braun T. Generalizing the TITE-CRM to adapt for early- and late-onset toxicities. *Statistics in Medicine* 2005; **20**:153–169. DOI: 10.1002/sim.2337.
10. Yin G, Yuan Y. Bayesian model averaging continual reassessment method in Phase I clinical trials. *Journal of the American Statistical Association* 2009; **104**:954–972. DOI: 10.1198/jasa.2009.ap08425.
11. Yuan Y, Yin G. Bayesian hybrid dose-finding design in Phase I oncology clinical trials. *Statistics in Medicine* 2011; **30**:2098–2108. DOI: 10.1002/sim.4164.
12. Lee SM, Cheung YK. Model calibration in the continual reassessment method. *Clinical Trials* 2009; **6**:227–238. DOI: 10.1177/1740774509105076.
13. Lee SM, Cheung YK. Calibration of prior variance in the Bayesian continual reassessment method. *Statistics in Medicine* 2011; **30**:2081–2089. DOI: 10.1002/sim.4139.
14. Chevret S. The continual reassessment method in cancer Phase I clinical trials: a simulation study. *Statistics in Medicine* 1993; **30**:1093–1108.
15. Neuenschwander B, Branson M, Gsponer T. Critical aspects of the Bayesian approach to phase I cancer trials. *Statistics in Medicine* 2008; **27**:2420–2439. DOI: 10.1002/sim.3230.
16. O'Quigley J. Estimating the probability of toxicity at the recommended dose following a Phase I clinical trial in cancer. *Biometrics* 1992; **48**:853–862. DOI: 10.2307/2532350.
17. O'Quigley J, Reiner E. A stopping rule for the continual reassessment method. *Biometrika* 1998; **85**:741–748. DOI: 10.1093/biomet/85.3.741.
18. Thall P, Russell K. 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. DOI: 10.2307/2534012.
19. Cheung YK, Chappell R. A simple technique to evaluate model sensitivity in the continual reassessment method. *Biometrics* 2002; **58**:671–674. DOI: 10.1111/j.0006-341X.2002.00671.x.
20. Iasonos A, O'Quigley J. Continual reassessment and related designs in dose-finding studies. *Statistics in Medicine* 2011; **30**:2057–2061. DOI: 10.1002/sim.4215.
21. Jeffreys H. *Theory of Probability*, Vol. 3. Oxford University Press: London, 1961.
22. R Development Core Team. *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing. Vienna: Austria, 2008.
23. Cheung YK. Coherence principles in dose-finding studies. *Biometrika* 2005; **92**:863–873. DOI: 10.1093/biomet/92.4.863.