Published in final edited form as:

Clin Trials. 2021 June ; 18(3): 279–285. doi:10.1177/17407745211001521.

A Bayesian Dose-finding Design for Outcomes Evaluated with Uncertainty

Matthew J. Schipper^{1,2}, Ying Yuan³, Jeremy M.G. Taylor^{1,2}, Randall K. Ten Haken², Christina Tsien⁴, Theodore S. Lawrence²

¹Department of Biostatistics, University of Michigan, Ann Arbor, MI, United States of America;

²Department of Radiation Oncology, University of Michigan, Ann Arbor, MI, United States of America;

³Department of Biostatistics, The University of Texas M. D. Anderson Cancer Center, Houston, TX, United States of America;

⁴Department of Radiation Oncology, Washington University, St. Louis, MO, United States of America.

Abstract

Introduction: In some Phase I trial settings, there is uncertainty in assessing whether a given patient meets the criteria for dose limiting toxicity (DLT).

Methods: We present a design which accommodates DLT outcomes that are assessed with uncertainty for some patients. Our approach could be utilized in many available phase I trial designs, but we focus on the continual reassessment method (CRM) due to its popularity. We assume that for some patients, instead of the usual binary DLT outcome, we observe a physician assessed probability of DLT specific to a given patient. Data augmentation is used to estimate the posterior probabilities of DLT at each dose level based on both the fully observed and partially observed patient outcomes. A simulation study is used to assess the performance of the design relative to using the CRM on the true DLT outcomes (available in simulation setting only) and relative to simple thresholding approaches.

Results: Amongst the designs utilizing the partially observed outcomes, our proposed design has the best overall performance in terms of probability of selecting correct MTD and number of patients treated at the MTD.

Conclusions: Incorporating uncertainty in DLT assessment can improve the performance of the CRM design.

Keywords

Phase I; Maximum Tolerated Dose; Continual Reassessment Method (CRM); Data Augmentation

Introduction

The goal of phase I clinical trials is to estimate the maximum tolerated dose (MTD) of a new drug or treatment, defined as the dose associated with probability of dose limiting toxicity (DLT) closest to a specified target rate of toxicity. In the trial, successive cohorts of patients are adaptively assigned to a set of pre-specified doses so that the toxicity profile of the agent under investigation can be determined. Many statistical methods have been developed for single-agent phase I trial designs including model-based and algorithmic designs. One of the most popular model-based designs is the Continual Reassessment Method (CRM).¹ Numerous extensions to the CRM have been proposed, including those to handle right censored time-to-toxicity outcomes,² accommodate multiple toxicities or ordinal toxicities,^{3,4,5} based on multiple toxicity constraints,⁶ model bivariate competing outcomes,⁷ handle partial ordering of dose levels,⁸ and simultaneously determine a dose and schedule,⁹ amongst many others.

All of these designs assume that the toxicity endpoint is observed. In some instances however there is uncertainty in assessing whether a given patient meets the criteria for DLT.¹⁰ For example, in neuro-oncology radiation therapy trials, it can be difficult to differentiate radiation necrosis (a DLT) from tumor progression (not a DLT). In liver cancer, decrease in liver function may be a treatment related toxicity but may also occur in the absence of any treatment in patients with underlying liver dysfunction. Uncertainty can also arise in the context of multi-modality or combination therapy where it is often difficult to attribute a particular toxicity to a particular treatment component (e.g. radiation, chemo or new targeted agent). Iasonos and O'Quigley published a design for this setting in which clinicians are asked to classify toxicities as treatment related or not and to provide estimates of the conditional probability of DLT given they have classified the toxicity as treatment related.¹¹ They discuss the two possible types of classification errors and develop a design for the setting where there may be false positive findings (i.e. clinician calls a toxicity a DLT when it is not) but no errors in the other direction.

Hillman et al investigated attribution of Adverse Events (AEs) in the context of phase III trials and concluded that 'AE attribution is difficult to determine, unreliable, and of questionable value in interpreting AE data in randomized clinical trials'.¹² One approach to dealing with difficulty in attribution is to remove from the definition of DLT, the requirement that it be related to a particular treatment component. Then the target rate of DLT would be the overall rate of toxicity that is acceptable. In many instances however, protocols do not take this approach and prefer to define DLTs as those toxicities at least possibly related to the new treatment. One statistical reason for this is that the overall rate of toxicity will often be more variable than the rate of toxicity due to the investigational treatment. For example, suppose chemo-radiation is the standard treatment for a particular cancer and causes DLT in approximately 20% of patients. Then a phase I trial seeking to add a new agent to chemo-radiation could either define DLT as those toxicities caused by the new treatment and specify a target rate of this toxicity (e.g. 10%) or define DLT without regard to treatment attribution and target a rate of 30%. All else being equal, the estimated rate of DLT in the latter case is likely to be more variable due to larger binomial variance as well as uncertainty in estimating the 'background' rate of DLT associated with chemo-radiation alone.

In this paper we describe a Phase I design which accommodates DLT outcomes that are assessed with uncertainty for some patients. Our approach could be utilized in many available phase I trial designs, but we focus on the CRM due to its popularity and superior performance over algorithmic designs. We first describe the design and in particular how probabilities of DLT are estimated from partially observed data and then illustrate its performance in a simulation study.

Methods

Continual Reassessment Method

In the CRM design, it is assumed that dose-limiting toxicity (DLT) is recorded as a binary outcome y and the true rate of DLT monotonically increases with dose level. Each time a new patient is enrolled the probability of DLT at each dose level is estimated given the observed DLT outcomes from all previously treated patients in the trial. Based on the updated dose-DLT curve, a new cohort of patients is assigned to the dose with an estimated DLT probability closest to the pre-specified target.

Suppose that a set of J doses are under investigation for the new drug, and let $(p_1, \dots, p_J)$ be the pre-specified DLT probabilities (skeleton) at those doses, satisfying a monotonic dose-toxicity order $p_1 < \dots < p_J$. The first patient cohort is typically assigned to the first dose level. To direct the dose escalation for an incoming cohort, the CRM links the dose level d with the associated DLT probability π_d via a working dose-toxicity model, such as

$$\pi_d = p_d^{\exp(\alpha)}, \quad d = 1, \dots, J, \quad (1)$$

where α is an unknown parameter.

Suppose that n patients have entered the trial, and let y_i and d_i denote the DLT outcome and the received dose level for the i th subject, respectively. When $\mathbf{y} = \{y_i, i = 1, \dots, n\}$ is fully observed, the likelihood function is given by

$$f(\mathbf{y} | \alpha) = \prod_{i=1}^n \left\{ p_{d_i}^{\exp(\alpha)} \right\}^{y_i} \left\{ 1 - p_{d_i}^{\exp(\alpha)} \right\}^{1 - y_i}.$$

Let $f(\alpha)$ denote a prior distribution for the parameter α , for which we typically take a normal distribution $f(\alpha) = \mathcal{N}(0, 2)$ with mean 0 and variance 2. The posterior distribution of α is given by

$$f(\alpha | \mathbf{y}) \propto f(\alpha) f(\mathbf{y} | \alpha), \quad (2)$$

and the posterior means of the dose toxicity probabilities can be computed by

$$\hat{\pi}_d = \int p_d^{\exp(\alpha)} \frac{L(\mathbf{y} | \alpha) f(\alpha)}{\int L(\mathbf{y} | \alpha) f(\alpha) d\alpha} d\alpha. \quad (3)$$

After updating the posterior estimates of the DLT probabilities at all of the doses considered, the recommended dose level for the next cohort of patients is the one that has a DLT

probability closest to the target τ . The trial continues until the exhaustion of the total sample size, and then the dose with a posterior DLT probability closest to ϕ is selected as the MTD.

CRM when some outcomes are evaluated with uncertainty

Assume that at a given time point of clinical trial, for some treated patients, physicians can clearly identify whether these patients have experienced DLTs or not, and thus their values of y_i can be determined as 1 or 0 without ambiguity. However, for the other patients, due to various reasons described previously, physicians can not determine with certainty whether or not these patients experienced a DLT. Instead, physicians provide an estimate of how likely it is that each of these patients experienced DLTs. This estimate would be based on their clinical judgement. In other words, we assume that each patient has a true underlying DLT outcome (yes or no) but for some patients, we do not directly observe their DLT outcome y_i , but instead only have an estimate of $\phi_i = \Pr(y_i = 1)$, i.e., the probability of y_i being equal to 1.

The missing values of y_i (from patients whose DLTs cannot be identified without uncertainty) preclude direct implementation of the CRM. An intuitive approach to dealing with the unobserved DLT outcomes is to impute the missing data through data augmentation (DA) proposed by Tanner and Wong¹³ so that the standard complete-data method can be applied. The DA iterates between two steps: the imputation (I) step, in which the missing data are imputed, and the posterior (P) step, in which the posterior samples of unknown parameters are simulated based on imputed data.

Let $\phi = \{\phi_i\}$ and $\mathbf{y} = (\mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}})$, where $\mathbf{y}_{\text{obs}}$ and $\mathbf{y}_{\text{mis}}$ denote the observed and missing DLT data, respectively. The DA can be described as follows. At the I step, we “impute” the missing data $y_i \in \mathbf{y}_{\text{mis}}$ drawing samples from their full conditional distribution, given by

$$f(y_i | \mathbf{y}_{\text{obs}}, \alpha, \phi) = \text{Bernoulli}(\phi_i).$$

After filling in the missing data, we obtain the “imputed” complete data $\mathbf{y}$ and convert the missing data problem into a standard complete-data problem. Then, at the P step of the DA, given the imputed complete data $\mathbf{y}$, we sequentially sample the unknown model parameter from its full conditional distributions $f(\alpha | \mathbf{y})$ given by (2). We iterate the above imputation (I) step and posterior (P) step until the Markov chain converges. When the iterative procedure converges, the sequence of samples of α converges to its marginal posterior distribution, which can be used to make inference on α and direct dose finding.

If posterior means of the dose toxicity probabilities $\hat{\pi}_d$ are of interest, they can be calculated directly as follows:

$$\begin{aligned}\hat{\pi}_d &= \frac{\int \int p_d^{\exp(\alpha)} f(\alpha) f(y_{\text{obs}}, y_{\text{mis}} | \alpha) f(y_{\text{mis}} | \phi) dy_{\text{mis}} d\alpha}{\int \int f(\alpha) f(y_{\text{obs}}, y_{\text{mis}} | \alpha) f(y_{\text{mis}} | \phi) dy_{\text{mis}} d\alpha} \\ &= \frac{\int p_d^{\exp(\alpha)} e^{-\alpha^2/4} \prod_{i \in y_{\text{obs}}} \left\{ p_{d_i}^{\exp(\alpha)} \right\}^{y_i} \left\{ 1 - p_{d_i}^{\exp(\alpha)} \right\}^{1-y_i} \prod_{i \in y_{\text{mis}}} \left\{ p_{d_i}^{\exp(\alpha)} \phi_i + \left(1 - p_{d_i}^{\exp(\alpha)} \right) (1 - \phi_i) \right\} d\alpha}{\int e^{-\alpha^2/4} \prod_{i \in y_{\text{obs}}} \left\{ p_{d_i}^{\exp(\alpha)} \right\}^{y_i} \left\{ 1 - p_{d_i}^{\exp(\alpha)} \right\}^{1-y_i} \prod_{i \in y_{\text{mis}}} \left\{ p_{d_i}^{\exp(\alpha)} \phi_i + \left(1 - p_{d_i}^{\exp(\alpha)} \right) (1 - \phi_i) \right\} d\alpha}.\end{aligned}$$

, where the integration of the distribution of y_{mis} is just a sum over the two possible values (0,1). In the above approach, the likelihood contribution for a patient with missing y_i is

$$p_{d_i}^{\exp(\alpha)} \phi_i + \left(1 - p_{d_i}^{\exp(\alpha)} \right) (1 - \phi_i)$$

An alternate estimate can be obtained from (3) by replacing each missing y_i with its expected value $E[y_i | \phi_i] = \phi_i$

$$\begin{aligned}\hat{\pi}_d &= \frac{\int p_d^{\exp(\alpha)} e^{-\alpha^2/4} \prod_{i \in y_{\text{obs}}} \left\{ p_{d_i}^{\exp(\alpha)} \right\}^{y_i} \left\{ 1 - p_{d_i}^{\exp(\alpha)} \right\}^{1-y_i} \prod_{i \in y_{\text{mis}}} \left\{ p_{d_i}^{\exp(\alpha)} \right\}^{\phi_i} \left\{ \left(1 - p_{d_i}^{\exp(\alpha)} \right) \right\}^{1-\phi_i} d\alpha}{\int e^{-\alpha^2/4} \prod_{i \in y_{\text{obs}}} \left\{ p_{d_i}^{\exp(\alpha)} \right\}^{y_i} \left\{ 1 - p_{d_i}^{\exp(\alpha)} \right\}^{1-y_i} \prod_{i \in y_{\text{mis}}} \left\{ p_{d_i}^{\exp(\alpha)} \right\}^{\phi_i} \left\{ \left(1 - p_{d_i}^{\exp(\alpha)} \right) \right\}^{1-\phi_i} d\alpha}.\end{aligned}$$

Dose-finding Algorithm

Let τ denote the physician-specified DLT target, and assume that patients are treated in cohorts. We restrict dose escalation to one dose level between patients. The dose-finding algorithm for the proposed dose-finding design is described as follows.

1. Patients in the first cohort are treated at the second lowest dose level, following the common clinical practice of allowing a fall back dose under the starting dose.
2. At the current dose level j , based on the cumulated data, calculate the posterior means for the DLT probabilities, $\hat{\pi}_d (d = 1, \dots, J)$, and identify dose level j^* that has a DLT probability closest to τ , i.e., $j^* = \operatorname{argmin}_{d \in \{1, \dots, J\}} |\hat{\pi}_d - \tau|$.
  - If $j > j^*$, de-escalate the dose level to $j - 1$;
  - if $j < j^*$, escalate the dose level to $j + 1$ provided
  - otherwise, the dose stays at the same level as j for the next cohort of patients.
3. Once the maximum sample size is reached, the dose that has the DLT probability closest to τ is selected as the MTD.

Results

Simulation Study

We conducted simulation studies to evaluate the performance of the proposed design. The target rate was set to 0.30. Each trial was run until 30 patients were enrolled or the posterior mean probability of DLT at dose level 1 exceeded the target rate by .05 (and at least 6 patients had been treated). The first patient was assigned to dose level 2. Subsequent patients were assigned to the dose level with posterior mean $P(DLT)$ closest to the target rate but not exceeding the target by more than .05. The process used to generate the data for the simulation study is necessarily complex to reflect the hypothesized data generating mechanism. Let d_{ij} , y_{ij} , ϕ_{ij} and m_{ij} denote the dose level, true (possibly unobserved) DLT outcome, physician assessed probability of DLT and missingness indicator for the i^{th} patient on the j^{th} trial. Then we generate the observed outcome y_{ij}^{obs} as

$$y_{ij}^{obs} = I(M_{ij} = 0)y_{ij} + I(M_{ij} = 1)\phi_{ij} \quad (4)$$

where

1. $y_{ij} \sim \text{Bernoulli}(\pi_{d_{ij}})$
2. $m_{ij} | y_{ij} \sim \text{Bernoulli}(I(y_{ij} = 0)\mu_{m0} + I(y_{ij} = 1)\mu_{m1})$
3. $\phi_{ij} = \beta_{ij} + e_j$
  - a. $\beta_{ij} | y_{ij} \sim \text{beta}(I(y_{ij} = 0) * \mu_{\phi0} + I(y_{ij} = 1) * \mu_{\phi1}, \gamma)$
  - b. $e_j \sim N(0, \sigma^2)$

We set $\mu_{\phi0} = .3$, $\mu_{\phi1} = .7$, $\mu_{m1} = .5$, $\mu_{m0} = (\pi_{d1} * \mu_{m1} * (1 - \mu_{\phi1})) / (\mu_{\phi0} * (1 - \pi_{d1}))$, $\gamma = 3$ and $\sigma = .15$. Under this data generation scheme, $E[y_{ij}^{obs} | \text{trial } j] = \pi_{d_{ij}} + c$ where c is a function of e_j with $E[c] = 0$. We utilized a simple normal distribution for the trial level error term which could lead to values for ϕ outside of the (0,1) interval. This occurred infrequently but in the simulation study we truncated the generated values to ensure they were appropriately contained within (0,1). See appendix 1 for additional details on this data generating mechanism. In words, the physician estimates are right on average but may be biased within a given trial. For instance, a given trial may be conducted by a physician who tends to over estimate the probabilities of DLT (e.g. due to a conservative approach out of fear of harming patients). Or a given trial may be conducted by a physician who tends to under estimate the probabilities of DLT (e.g. because he or she ‘believes’ in the treatment being studied).

Simulation Designs

In addition to our proposed design, we also consider threshold approaches to handling this uncertainty in DLT outcomes. In the threshold approach, ϕ_{ij} is converted to a binary outcome by $y_{ij} = I(\phi_{ij} > \text{threshold})$. The analysis then proceeds as if y' were the observed data. We consider 3 threshold values; 0.33, 0.50 and 0.67. Finally, we also utilize the underlying true

outcome data y_{ij} , which is known only in the simulations setting, to conduct a standard CRM trial.

Other than the method for estimating the DLT probabilities across dose levels, all designs were identical. The target rate was set to .30. The prior estimates of the P(DLT) were the same probabilities used in scenario 1. Patients were enrolled in cohorts of 2 and no dose skipping was allowed. Then for each design we calculated how often each dose level is selected as the MTD under each simulation scenario. We also calculated an overall ‘average error’ index defined as the weighted average absolute difference between the probability of DLT at the selected MTD and the target rate, defined as

$$\text{average error} = \sum_{k=1}^K |\pi_k - \tau| * \text{probability of selecting dose } k \quad (5)$$

We note this is similar to the accuracy index described in¹⁴ but expressed as error so that smaller is better with the best possible value being 0 and the value directly interpretable as an average absolute difference in probabilities. In scenario 6 with no true MTD, we calculated the error as the average of the toxicity rate at the selected MTD or 0, for trials correctly recommending no dose level. Finally, we calculated the median number of patients treated at the MTD. This is an important metric both because of the desire to treat patients at the MTD and because more patients treated near the MTD allow for more precise preliminary estimates of efficacy such as response rate.¹⁵

Simulation Results

Table 2 gives the proportion of trials in which each dose level was selected as the MTD by each design under each simulation scenario. As expected, the standard CRM based on knowing the true DLT status for all patients, has the best performance in terms of the probability of correct MTD selection and the average error metric. Of the designs based on partially missing data, the proposed design has the best average performance, although in any given scenario, one of the thresholding designs may do better. For example, in scenario 1, the true MTD is at the highest dose level so thresholding at 0.67 approach does the best due to its bias to underestimate toxicity and thus select higher dose levels. Comparing the average error metric between designs, we see that the proposed design has the best (lowest) value in 2 of the first 4 simulation scenarios (2 and 3) and second best in the other 2 scenarios (1 and 4). Interpretation of the average error metric is more difficult in scenarios 5 and 6 because many trials stop early and recommend no dose level. In scenario 5, no design performs well with a majority of trials stopping early and declaring all doses too toxic. Performance would be improved if level 1 had a lower toxicity rate (e.g. .20 or .25). The performance of the designs with respect to the number of patients treated at the MTD are shown in Table 3. The relative performance of designs here largely mirrors that for the probability of selecting the correct MTD. Amongst the 4 designs based on partially missing data, the proposed method treated the largest number of patients at the MTD. As a check, using the full data, again resulted in the best performance.

We performed sensitivity analyses in which we varied the value of γ , σ , μ_{m1} , μ_{ϕ_0} or μ_{ϕ_1} . As expected, the differences between the methods decreased as μ_{m1} went to 0. Also, as σ increased, the performance of all methods which used the ϕ values, decreased. Otherwise, the results, and in particular, the relative rankings of designs, were not sensitive to choice of these parameters and are not shown.

Discussion

The proposed design allows for inclusion of information from clinicians on the uncertainty regarding DLT status for some patients. It will be of practical use only where this type of data can be reliably obtained from clinicians. One such class of examples, is the common setting in which the uncertainty in DLT status is due to the attribution of the observed adverse event to study treatment. In many ongoing trials in oncology at least, adverse events are scored according to Common Terminology Criteria for Adverse Events (CTCAE) criteria and clinicians score the attribution of the AE to the study treatment as unrelated, probably unrelated, possibly related, probably related or definitely related. To implement the proposed design, one could, in discussion with the clinicians, assign a numeric probability to each of these 5 levels (e.g. (.05, .2, .50, .8, .95)). Two of the authors have been or are currently members of our Cancer Center data safety monitoring committee. The clinicians on this committee routinely discuss attribution and whether the attribution provided by the treating clinician is reasonable. The observed attributions make full use of the 5 point scale. Our proposed design makes it possible for clinicians to avoid making the call as to unlikely vs possibly related in the setting of a Phase I trial where the former could make the event a non-DLT while the latter would make it a DLT. In other settings the source of the uncertainty is not the attribution but whether the observed event is in fact toxicity. For example in brain cancer treated with radiation therapy, this issue arises where it is difficult to distinguish toxicity (radiation necrosis) from pseudo-progression as both can result in similar patient symptoms and imaging. There have been recent reports of models that predict the probability of each of these events for individual patients (for example using radiomic features). For example, if the suspected region of radiation necrosis were biopsied, the true status of toxicity versus disease progression would be known. Given a dataset with this outcome and a paired image, a model could be developed providing an estimated probability of DLT. If such models were validated, the model estimated probabilities that a specific patient has radiation necrosis could be used in the proposed trial design.

Our proposed method performs well with average error values close to those obtained by using the true DLT status for all patients (something unknown outside of a simulation study). Perhaps the most striking results from the simulation study were the extent to which the selection of dose depended on the choice of threshold value. Using a threshold of 33% is conservative and, on average, results in selecting a dose level below the MTD. While this is to be expected, it has implications for the common DLT definition of toxicity including the criteria that it is at least possibly related to study treatment.

Our approach has both theoretical and practical differences from a recent publication on the same topic.¹¹ Iasonos and O'Quigley jointly model Y (true DLT outcome) and the clinicians

classification (Z in their paper), and link them through the error rate of the classification. In contrast, we directly model Y , and in the case that the value of Y cannot be ascertained, we ask clinicians to provide the likelihood that the toxicity is related to the drug, i.e., $\Pr(Y=1)$. There is also a practical difference in that Iasonos and O'Quigley require clinicians first to make a binary decision if a toxicity is related to the drug (i.e., the value of Z in their paper), and then ask the clinicians to provide an estimate of the conditional probability $P(Y=1 | Z=1)$. In contrast, our method directly asks the clinician how likely the toxicity is related to the drug, i.e., $P(Y=1)$, which simplifies the elicitation.

A simple competitor to our proposed design would be to threshold observed values at 0.50 and use the standard CRM design. While the differences are small, our proposed approach does better, both in terms of the probability of selecting the correct target dose and in terms of the median number of patients treated at the target dose. At the same time, we are unaware of trials that actually implemented such an approach. By allowing physicians to provide an estimate of how likely a patient is to have experienced DLT, rather than forcing them to provide a binary assessment, this design leverages a more informative outcome measure to improve performance.

Funding

This work was supported by the National Institutes of Health (grant number P30 CA046592).

1: Appendix

We here sketch the calculation of $E[Y_{ij}^{obs} | trial j]$. In what follows, we drop 'trial j' for notational convenience but compute the expectation conditional on trial j.

$$\begin{aligned} E[Y_{ij}^{obs} | trial j] &= E[E[Y_{ij}^{obs} | Y_{ij}]] \\ &= \pi_{d_{ij}} E[Y_{ij}^{obs} | Y_{ij} = 1] + (1 - \pi_{d_{ij}}) E[Y_{ij}^{obs} | Y_{ij} = 0] \end{aligned}$$

$$\begin{aligned} E[Y_{ij}^{obs} | Y_{ij} = 1] &= E[I(M_{ij} = 1)\phi_{ij} + I(M_{ij} = 0)Y_{ij} | Y_{ij} = 1] \\ &= \mu_{M_1}(\mu\phi_1 + e_j) + 1 - \mu_{M_1} \end{aligned}$$

$$\begin{aligned} E[Y_{ij}^{obs} | Y_{ij} = 0] &= E[I(M_{ij} = 1)\phi_{ij} + I(M_{ij} = 0)Y_{ij} | Y_{ij} = 0] \\ &= \mu_{M_0}(\mu\phi_0 + e_j) \end{aligned}$$

Plugging these expressions into the first equation, we obtain:

$$\begin{aligned}
 E[Y_{ij}^{obs} | trialj] &= \pi_{d_{ij}}(\mu_{M1}(\mu_{\phi_1} + e_j) + 1 - \mu_{M1}) + (1 - \pi_{d_{ij}})(\mu_{M0}(\mu_{\phi_0} + e_j)) \\
 &= \pi_{d_{ij}}\mu_{M1}\mu_{\phi_1} + \pi_{d_{ij}} - \pi_{d_{ij}}\mu_{M1} \\
 &\quad + \mu_{M0}\mu_{\phi_0}(1 - \pi_{d_{ij}}) \\
 &\quad + e_j(\pi_{d_{ij}}\mu_{M1} + \mu_{M0} - \pi_{d_{ij}}\mu_{M0}) \\
 &= \pi_{d_{ij}} + c
 \end{aligned}$$

where, $E[c] = 0$ with expectation calculated with respect to distribution of e_j .

References

1. O'Quigley J, Pepe M and Fisher L. Continual reassessment method: A practical design for Phase I clinical trials in cancer. *Biometrics* 1990; 46: 33–48. [PubMed: 2350571]
2. Cheung YK and Chappell R. Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics* 2000; 56: 1177–1182. [PubMed: 11129476]
3. Bekele BN, Thall PF. Dose finding based on multiple toxicities in a soft tissue sarcoma trial. *Journal of the American Statistical Association*. 2004; 99: 2636.
4. Ivanova A, Kim SH. Dose finding for continuous and ordinal outcomes with a monotone objective function: a unified approach. *Biometrics*. 2009; 65:247255.
5. Van Meter EM, Garrett-Mayer E and Bandyopadhyay D. Dose-finding clinical trial design for ordinal toxicity grades using the continuation ratio model: an extension of the continual reassessment method. *Clin Trials*. 2012; 9(3):303–13. [PubMed: 22547420]
6. Lee SM, Cheng B, and Cheung YK. Continual reassessment method with multiple toxicity constraints. *Biostatistics*. 2011; 12(2): 386398.
7. Braun TM. The bivariate continual reassessment method: Extending the CRM to phase I trials of two competing outcomes. *Controlled Clinical Trials*. 2002; 23 240–256. [PubMed: 12057877]
8. Wages NA, Conaway MR, O'Quigley J. Continual reassessment method for partial ordering. *Biometrics*. 2011;67(4): 1555–63. [PubMed: 21361888]
9. Zhang J and Braun T. A Phase I Bayesian Adaptive Design to Simultaneously Optimize Dose and Schedule Assignments Both Between and Within Patients. *J Am Stat Assoc*. 2013; 108(503): 892–901.
10. Zohar S and OQuigley J. Sensitivity of dose-finding studies to observation errors. *Contemp. Clin. Trials* 2009; 30, 523530.
11. Iasonos A and O'Quigley J. Phase I designs that allow for uncertainty in the attribution of adverse events. *J R Stat Soc Ser C Appl Stat*. 2017; 66(5): 10151030.
12. Hillman SL, Mandrekar SJ, Bot B, DeMatteo RP, Perez EA, Ballman KV, Nelson H, Buckner JC and Sargent DJ. Evaluation of the Value of Attribution in the Interpretation of Adverse Event Data: A North Central Cancer Treatment Group and American College of Surgeons Oncology Group Investigation *J Clin Oncology*. 2010; 28(18):3002–7.
13. Tanner MA and Wong WH. The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association*. 1987; 82: 528–540.
14. Cheung KY. Dose finding by the continual reassessment method. Chapman and Hall/CRC. 2011.
15. Boonstra PS, Shen J, Taylor JM, Braun TM, Griffith KA, Daignault S, Kalemkerian GP, Lawrence TS and Schipper MJ. A statistical evaluation of dose expansion cohorts in phase I clinical trials. *J Natl Cancer Inst*. 2015; 107(3).

Table 1:

Prior probabilities of toxicity

	Dose Level				
	1	2	3	4	5
Prior Values (skeleton)	0.02	0.05	0.10	0.20	0.3
Scenario 1	0.02	0.05	0.10	0.20	0.3
Scenario 2	0.05	0.10	0.20	0.30	0.45
Scenario 3	0.10	0.20	0.30	.45	0.60
Scenario 4	0.20	0.30	0.45	0.60	0.70
Scenario 5	0.30	0.45	0.60	0.70	0.80
Scenario 6	0.45	0.55	0.65	0.75	0.85

Table 2:

Simulation Results: Selected Dose Level and Average Error

Scenario	Method	Selected Dose Level (% of Trials)						Average Error
		Stop Early	1	2	3	4	5	
1	No Missing	0.0	0.0	0.0	1.7	29.3	69.0	3.3
1	Proposed Method	0.0	0.0	0.0	1.6	23.7	74.6	2.7
1	Threshold = 0.33	0.0	0.0	0.0	4.4	44.6	51.0	5.3
1	Threshold = 0.50	0.0	0.0	0.0	1.8	29.3	68.9	3.3
1	Threshold = 0.67	0.0	0.0	0.0	0.7	17.3	81.9	1.9
2	No Missing	0.2	2.4	21.7	21.7	58.1	17.6	5.4
2	Proposed Method	0.4	0.0	2.9	20.3	54.5	21.9	6.0
2	Threshold = 0.33	1.1	0.0	6.3	33.4	51.0	8.2	6.2
2	Threshold = 0.50	0.9	0.0	2.7	24.0	52.8	19.6	6.2
2	Threshold = 0.67	0.6	0.0	1.0	13.7	44.8	39.9	7.7
3	No Missing	3.8	1.6	24.7	46.6	22.7	0.6	7.5
3	Proposed Method	4.0	1.8	27.3	45.2	20.4	1.0	7.7
3	Threshold = 0.33	5.7	4.9	41.2	37.7	10.3	0.2	8.4
3	Threshold = 0.50	3.8	2.0	23.9	43.3	25.2	1.8	8.3
3	Threshold = 0.67	3.5	2.0	26.9	44.1	21.1	2.4	8.0
4	No Missing	18.4	18.0	48.3	14.1	1.2	0.0	9.8
4	Proposed Method	22.2	23.9	41.3	11.2	1.4	0.0	11.2
4	Threshold = 0.33	29.5	24.9	36.8	8.2	0.6	0.0	12.8
4	Threshold = 0.50	21.4	15.3	45.4	15.1	2.6	0.2	11.1
4	Threshold = 0.67	22.4	4.6	35.2	26.7	9.4	1.8	12.8
5	No Missing	54.1	30.3	14.8	0.80	0.0	0.0	18.7
5	Proposed Method	65.2	26.4	8.3	0.1	0.0	0.0	20.8
5	Threshold = 0.33	68.8	22.8	8.0	0.4	0.0	0.0	22.0
5	Threshold = 0.50	54.2	27.5	15.9	2.0	0.3	0.1	19.5
5	Threshold = 0.67	40.3	21.5	30.3	5.7	2.1	0.1	19.3
6	No Missing	89.4	9.6	1.0	0.0	0.0	0.0	4.9
6	Proposed Method	94.7	4.8	0.4	0.0	0.0	0.0	2.5
6	Threshold = 0.33	95.4	3.1	1.4	0.1	0.0	0.0	2.2
6	Threshold = 0.50	89.0	8.6	1.9	0.3	0.2	0.0	5.3
6	Threshold = 0.67	73.4	14.2	9.7	2.2	0.5	0.0	13.5

Table 3:

Median Number of Patients Treated at the MTD

	Scenario					
	1	2	3	4	5	6
No Missing	20	13	13	12	7	NA
Proposed Method	21	12	12	12	6	NA
Threshold = 0.33	14	10	11	11	5	NA
Threshold = 0.50	20	12	11	11	6	NA
Threshold = 0.67	23	10	11	9	5	NA