Published in final edited form as:

J Biopharm Stat. 2017 ; 27(3): 495–506. doi:10.1080/10543406.2017.1290650.

Selection of the initial design for the Two-Stage Continual Reassessment Method

Xiaoyu Jia[1,*], **Anastasia Ivanova**[2,**], and **Shing M. Lee**[3,***]

[1]Department of Population Health Science and Policy, Institute for Healthcare Delivery Science, Icahn School of Medicine at Mount Sinai, New York, New York, United States

[2]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, U.S.A

[3]Department of Biostatistics, Columbia University, 722 West 168th Street, New York, New York 10032, U.S.A

Summary

The continual reassessment method (CRM) was proposed in a Bayesian framework whereby the first patient is assigned to the prior guess of the maximum tolerated dose which is usually not the lowest dose level. This assignment may lead to safety concerns in practice because physicians usually prefer not to skip lower dose levels before escalating to the higher dose levels. The two-stage CRM was proposed to address such concern whereby model based dose escalation is preceded by a pre-specified escalating sequence starting from the lowest dose level. While a theoretical framework to build the two-stage CRM has been proposed, the selection of the initial dose escalating sequence, generally referred to as the initial design, remains arbitrary, either by specifing cohorts of three patients or by trial and error through extensive simulations. Motivated by a currently ongoing oncology dose finding study for which physicians stated their desire to start from the lowest dose even though the maximum tolerated dose was thought to be one of the higher dose levels, we proposed a systematic approach for selecting the initial design for the two-stage CRM. The initial design obtained using the proposed algorithm yields better operating characteristics compared to using a cohort of three initial design with a calibrated CRM. The proposed algorithm simplifies and provides a systematic approach for the selection of initial design for the two-stage CRM. Moreover, initial designs to be used as reference for planning a two-stage CRM are provided.

Keywords

Adaptive design; Phase I study; Dose finding; Indifference intervals; Initial design; Likelihood CRM

*xiaoyu.jia@mssm.edu
**aivanova@bios.unc.edu
***sml2114@columbia.edu

1. Introduction

The objective of dose-finding studies in oncology is to identify the maximum tolerated dose (MTD), defined as the dose level with toxicity probability closest to a pre-specified target value. These studies are sequential and assign doses based on outcomes of previously enrolled patients either using rule-based algorithms or model-based designs. While rule-based designs, for example the 3+3 algorithm (Storer, 1989), are easy to implement, model-based designs such as the continual reassessment method (CRM) have been shown to have better performance, and assign fewer patients at suboptimal doses (O'Quigley, Pepe and Fisher, 1990). The CRM allows for a specified sample size and, therefore, the MTD can be estimated with desired precision. Moreover, specifying the cohort size irrespective of the target value permits more frequent evaluation of doses. However, it also requires the specification of model parameters at the start of the trial. The statistical and computational complexity, and the lack of guidance regarding the specification of these parameters may contribute to the slow uptake of the method in practice (Braun, 2014). Model calibration requires experience with these designs and even for experienced statisticians it can be a time consuming process. Calibration algorithms that simplify this process have been proposed to select the model parameters for the Bayesian CRM (Lee and Cheung, 2009, 2011).

The CRM, as originally proposed, assigns the first patient at the best guess of the MTD, which usually is not the lowest dose level. This may be viewed as overly aggressive by clinicians whose preference is generally to start the trial at the lowest dose level to minimize risk (Ivy et al., 2010), and have sufficient number of patients assigned to each of the lower dose levels before escalating to higher dose levels. To address this concern, the two-stage CRM was proposed (O'Quigley and Shen, 1996). In the first stage, doses are assigned based on a pre-specified dose escalating sequence starting from the lowest dose level. Once the first toxicity is observed, the second stage is initiated whereby doses are assigned based on the CRM. The selection of the initial dose escalating sequence for the two-stage CRM, generally referred to as the initial design, has not been addressed in the literature. The choice of initial design is important because an inappropriately chosen one can lead to incoherent dose assignment in a two-stage CRM (Cheung, 2005), that is, the dose may escalate despite a dose limiting toxicity (DLT) being observed in the previous patient. Coherence is important because in practice most clinicians would not be comfortable escalating the dose following the occurrence of a DLT. Thus, up and down designs were proposed following the coherence principle. While the one stage Bayesian CRM is coherent (Cheung, 2005), coherence for the two-stage CRM is not guaranteed and over-conservative initial designs can be incoherent depending on the number of dose levels and the target toxicity value (Jia, Lee and Cheung, 2014). However, the literature on the selection of the initial design is sparse. The selection of initial designs remains largely arbitrary without rigorous justification, and specifics are lacking on how to determine such an initial design for a given clinical scenario (Storer, 1989). In the original proposal of the two-stage CRM, the initial design was based on a cohort size of three (O'Quigley and Shen, 1996), that is, if no DLT is observed after three patients, the dose is escalated. In other cases, faster escalation schemes were employed such as one patient on each of the first two dose levels and two patients each on the 3rd and 4th dose levels before escalating to the highest dose level (O'Quigley and Paoletti, 2003;

Ivanova and Wang, 2006). Using these initial designs, coherence can be forced by not allowing dose escalation after a DLT is observed, or the CRM in the second stage can be calibrated by trial and error to ensure coherence. These approaches are either ad hoc or require extensive simulations.

This research was motivated by an oncology dose finding study which is currently being conducted at Columbia University Medical Center (clinicaltrials.gov identifier NTC 02202772). The study objective is to determine the MTD, defined as the dose level associated with a target toxicity probability of 25%, for the combination chemotherapy of Gemcitabine, Cabazitaxel and Cisplatin in patients with bacillus calmette guerin (BCG) refractory or recurrent non-muscle invasive bladder cancer via intravesical administration. The dosage of Gemcitabine is fixed at 2g for all dose levels. Two dosages of Cabazitaxel and three dosages of Cisplatin are of interest, yielding a total of 5 dose levels (Table 1). The study sample size is 24 patients. A DLT is defined as any grade 3 or 4 systemic toxicity or any grade 3 or 4 hematuria, dysuria, urinary retention, urinary frequency/urgency, or bladder spasms according to the National Cancer Institute Common Terminology Criteria for Adverse Events (CTCAE) version 4.0 (NCI, 2009). During the study planning stage, the investigators explicitly expressed their desire of starting the trial from the lowest dose level and testing each dose level of the combinations of Gemcitabine and Cabazitaxel with at least one patient due to safety concern, before introducing Cisplatin. While they thought the MTD would most likely involve the combination of all three drugs, they did not want to start the trial with the combination of all three and skip the lower dose levels with the combination of two.

To address the concern of the investigators, the two-stage CRM was proposed, as it not only addresses the physician's concern, but also performs similar to the original CRM and outperforms the 3+3 design in selecting the correct dose as the MTD (O'Quigley and Shen, 1996). However, starting at the lowest dose level using the two-stage CRM also opened up new design issues because it required the pre-specification of the initial escalating sequence, in addition to the model parameters for the CRM.

This paper proposes a systematic approach for selecting an initial design for the two-stage CRM based on the theoretical framework established using the coherence principle (Jia, Lee and Cheung, 2014). In Section 2, we review the model specification of two-stage CRM. The systematic algorithm is proposed in Section 3, followed by the general application results and practical recommendations on design parameters under various clinical scenarios in Section 4. In Section 5, we illustrate the proposed algorithm in the context of the oncology dose finding study, and compare the results to those using initial designs with cohorts of 1 and 3 paired with model calibration using the approach proposed for the Bayesian single-stage CRM.

2. Model specification of two-stage CRM

Suppose we are interested in estimating the dose level associated with a target DLT probability p using a two-stage CRM, where $d_1, d_2, \ldots, d_K$ are the K doses of interest. It is necessary to specify an initial dose escalating sequence for the first stage, also referred to as

initial design, and the CRM model for the second stage. The initial dose sequence can be represented using cohort sizes $D_0 = \{m_{01}, m_{02}, \ldots, m_{0K}\}$, where m_{0k} indicates the number of patients assigned to dose level k. The total number of patients assigned to the first $K-1$ dose levels, $\sum_{k=1}^{K-1} m_{0k}$, is referred to as the conservatism index which measures the escalation speed of the initial design (Cheung, 2011). A large index indicates a more conservative initial design with slow escalation speed, and a large number of patients being assigned to lower dose levels.

For the second stage, the CRM requires the specification of a dose toxicity function ($F(d, \beta)$), the initial estimates of the dose toxicity probabilities at each dose $\{p_0 = \{p_{01}, p_{02}, \ldots, p_{0K}\}$), also referred to as the skeleton, and the prior distribution of the model parameter ($\pi(\beta)$) if the Bayesian framework is used. Once heterogeneity in toxicity outcomes is observed, the trial switches to the second stage which uses the CRM to update the estimate of the dose-toxicity model using the accrued toxicity information, and adaptively assigns incoming patients to the dose with an estimated toxicity probability closest to p. Once the planned sample size is reached, the dose with estimated toxicity probability closest to p is selected as the MTD.

2.1 Model Calibration

For the original CRM, given the dose toxicity function, $F(d, \beta)$, an algorithm based on indifference intervals was proposed to select the skeleton $\{p_{01}, p_{02}, \ldots, p_{0K}\}$. While for large sample sizes, the indifference interval specifies a range of acceptable toxicity probabilities, such that the CRM would converge to the dose levels with toxicity probabilities falling within a symmetric interval of length 2δ around the target toxicity probability p (Cheung and Chappell, 2002), in the finite sample setting the optimal δ can be chosen based the algorithm proposed by Lee and Cheung (2009). Starting with the prior guess of the MTD location, v_0, and setting it equal to p, $p_{0,v_0} = \mathrm{F}(\mathrm{d}_{v_0}, \beta_0) = p$ where β_0 denotes the prior mean of β, the initial estimates of each dose level can be obtained using the following iterative algorithm

$$F^{-1}(p_{0,k+1})F^{-1}(p-\delta)=F^{-1}(p_{0k})F^{-1}(p+\delta), k=1,\ldots,k-1 \quad (1)$$

The two-stage design allows using likelihood estimation in addition to Bayesian methods to guide the dose escalation. Using the likelihood approach and obtaining the initial estimates of dose toxicity probabilities $\{p_{01}, p_{02}, \ldots, p_{0K}\}$ using algorithm (1), the initial value β_0 is irrelevant to the design performance (Cheung, 2011) and the estimated model parameter is invariant to the choice of the prior guess of MTD (v_0) (Jia, Lee and Cheung, 2014). For example, suppose $p = 0.25$, $K = 5$, and $\delta = 0.05$. Assuming an empiric dose toxicity function $F(d_k, \beta)=d_k^{\beta}$, the initial estimates of dose toxicity probabilities are $p_0 = \{0.08, 0.16, 0.25, 0.35, 0.46\}$ when the prior guess of the MTD is dose level 3, and $p_0 = \{0.16, 0.25, 0.35, 0.46, 0.56\}$ when the prior guess is dose level 2. The two skeletons have the same spread, but different prior guesses of the MTD. Given the same dose assignments and corresponding toxicity outcomes, the two models yield identical likelihood estimates and dose assignments for the next patient. Therefore, when using the likelihood approach, the prior guess of the

MTD, ν_0, can be arbitrarily set at any dose level upon the activation of the CRM without affecting the design performance.

2.2 Coherence

Given the dose-toxicity function F and the associated initial estimate of dose toxicity probabilities p_0 constructed using the indifference interval approach (1), there exists a unique most conservative and coherent initial design for the two-stage CRM (Jia, Lee and Cheung, 2014), which can be determined through constructing a series of dose escalating sequences with increasing conservatism by adding one patient at a time from dose level $K - 1$ to 1 and starting over again from dose level $K - 1$ every time once reaching dose level 1. Each intermediate sequence would be checked for coherence before moving on to the next sequence. The iterative process continues until adding one more patient renders the dose escalating sequence incoherent. The coherent sequence immediately before is the unique most conservative and coherent initial design. This initial design is invariant to the choice of the prior guess of the MTD when using maximum likelihood estimation.

Assuming a total sample size of 15 in the aforementioned example, the unique most conservative and coherent initial design would be $D_0 = \{2, 2, 2, 2, 7\}$, regardless of which of the two skeletons ($p_0 = \{0.08, 0.16, 0.25, 0.35, 0.46\}$ or $p_0 = \{0.16, 0.25, 0.35, 0.46, 0.56\}$) is used in the CRM model. This intial design has a conservatism index of $\sum_{k=1}^{K-1} m_{0k}=8$, In this case, a sequence of initial designs $D_{01} = \{1, 1, 1, 1, 11\}$, $D_{02} = \{1, 1, 1, 2, 10\}$, $D_{03} = \{1, 1, 2, 2, 9\}$, $D_{04} = \{1, 2, 2, 2, 8\}$,and $D_{05} = \{2, 2, 2, 2, 7\}$ are examined and proven to be coherent given the CRM model in the second stage. However, the next sequence in line $D_{06} = \{2, 2, 2, 3, 6\}$ is over-conservative and leads to incoherence. Therefore, the previous sequence, D_{05}, is the most conservative and coherent initial design for this example.

2.3 Pruning

Certain number of patients, m_{0K}, can be reserved for the highest dose level if no toxicity occurs during the entire trial to gain reassurance before selecting the highest dosel level as the MTD. In the case that the conservatism index of the most conservative and coherent initial design exceeds such constraint, and fewer than desired patients are reserved for the highest dose, the initial design can be pruned down by iteratively removing one patient at a time in reverse fashion from dose level 1 to $K - 1$ until the constraint is satisfied, resulting in a less conservative and still coherent initial design.

Using the example above, if we specify 8 patients being reserved on the highest dose level, the most conservative and coherent initial design $D_0 = \{2, 2, 2, 2, 7\}$ needs to be pruned down to “re-allocate” patients from the lower dose levels to the highest dose level. Following the pruning algorithm, we start from dose level 1 and re-allocate one patient at a time to dose level 5 starting at the lowest dose level, until the constraint is met. Since re-allocating one patient is sufficient in this particular example, a pruned initial design of $D_0 = \{1, 2, 2, 2, 8\}$ would be used as the initial design.

3. Selection of the initial design

Given a target DLT probability p, with K dose levels, sample size N, and dose toxicity function F, below is the algorithm to select an initial design which yields the maximum average probability of correct selection (PCS) across a calibration set of K toxicity scenarios of true probabilities of DLT. Given the existence of a unique most conservative and coherent initial design with the initial estimates of dose toxicity probabilities, p_0, obtained using the indifference interval approach, the inital design can also be selected by specifying δ. However, to allow for a pre-specified number of patients being reserved at the highest dose level m_{0K}, an additional tuning parameter λ is included in this algorithm. Specifically, let $m_{0K} = \lambda/p$. For example, with a target toxicity probability $p = 0.10$ and $\lambda = 1$, 10 patients will be researved for the highest dose level, such that approximately one toxicity is expected to occur if dose level K is the MTD. Larger λ values force faster escalation through the lower dose levels and may be more appropriate for larger values of p. The number of patients reserved at the highest dose level, m_{0K}, determines whether pruning is necessary.

Algorithm

1. Given a target DLT probability p, with K dose levels, sample size N, and dose toxicity function $F(d, \beta)$, vary the indifference interval width, δ, in a range between 0.01 to $0.6 \times p$ on a grid width of 0.01. The range can be smaller for large values of p.

   For each δ, repeat steps 2-6 below.

2. Determine the skeleton p_0 based on δ using equation (1), with the prior guess of the MTD, ν_0, set at 1.

3. Given $F(d, \beta)$ and p_0, obtain the most conservative coherent initial design D_0 as explained in the Coherence section.

4. (Optional) If the number of patients assigned to the highest dose level is less than that desired, prune the most conservative and coherent initial design as explained in the Pruning section.

5. Specify the toxicity scenarios for calibration set by assuming K different scenarios of true probabilities of DLT, with each scenario selecting one of the K dose levels as the MTD. The scenarios are specified following the plateau configuration where $p_i = p_L$ for $i < l$, $p_i = p_U$ for $i > l$ and $p_i = p$ for $i = l$ where $l = 1, \ldots, K$, odds ratio $\psi = 2$, $p_U = p\psi/(1-p(1-\psi))$ and $p_L = p/(\psi + p(1-\psi))$. Other values of ψ can be investigated depending on the application.

6. Perform simulations using the two-stage likelihood CRM as specified under each of the K scenarios indicated in step 5, and obtain the corresponding PCS for each scenario. Average the PCS across all K scenarios of the calibration set.

Once the average PCS across all K scenarios for each δ value is obtained, select the δ and the correspoding initial design D_0 and skeleton p_0, that yields the highest average PCS across the scenarios. These are the initial design and skeleton to be used for the two-stage likelihood CRM.

4. General Application

The proposed algorithm is applied to various scenarios of target probability of toxicity, sample size, and number of doses. The target toxicity probabilities considered are 0.10, 0.20, 0.25 and 0.33. The sample sizes are 25, 30, 35, and 40. The number of dose level ranges from 4 to 7. These are selected based on common scenarios encountered in practice. Two different values for the tuning parameter $\lambda = 1$ and 2 are included for each scenario. The dose-toxicity model is assumed to be empiric and the maximum likelihood estimation method is used to estimate the model parameter. The toxicity scenarios used for calibration are based on an odds ratio of 2. Two thousand simulations were performed for all scenarios. Simulations were not performed for $p = 0.10$, with a sample size of 25, seven dose levels, and $\lambda = 2$, because on average there would be less than one patient assigned on the lower dose levels. No dose skipping is allowed during dose escalation. For each scenario defined by N, K and λ, the optimal δ value and the corresponding (pruned) initial design that yields the highest average PCS are displayed (Table 2), along with the δ values and D_0's that yield average PCS within 1% of the peak value (Appendix A). The optimal δ ranges between 0.02 and 0.04 for $p = 0.10$, between 0.03 and 0.04 for $p = 0.20$, between 0.04 and 0.06 for $p = 0.25$, and between 0.05 and 0.07 for $p = 0.33$. The optimal δ value increases as the target toxicity probability increases, and decreases as the number of doses increases. The initial design also depends on other parameters and tends to escalate faster with higher target toxicity probability and higher number of doses. When target toxicity probability is low, such as 10%, or the number of doses are small, the skeleton is flatter and slower escalation is permitted without causing incoherent dose assignment. However, as the target toxicity probability increases, such as 25% and 33%, or number of doses increases, faster escalation is required to reach the higher dose levels earlier, and the initial designs allow at most one or two patients to be assigned on each dose level before escalating. Under these scenarios, using cohorts of three as the initial design may lead to incoherence given the specified initial estimates of toxicity probability. In addition, smaller sample sizes and larger λ values generally require more pruning on the most conservative and coherent initial designs, thus leading to more aggressive dose escalation through lower dose levels.

These general results can be used as a reference to guide the design of dose-finding studies using the two-stage CRM.

5. Application to the Oncology Trial

As described in the introduction, the objective of the oncology dose finding study is to identify the dose level closest to a target toxicity rate of $p = 0.25$, given $K = 5$ dose levels and a sample size of 24. The best guess of the MTD is dose level 3. The investigator's primary concern was the safety of the lower dose levels. Thus, they wanted to start at the lowest dose level and only designs starting at the lowest dose level were considered. Moreover, we specified a safety rule whereby the trial would stop early if consecutive toxicities occurred on the first two enrolled patients. In this case, the MTD would be below dose level 1.

Since both the 3+3 design and the two-stage CRM can address the primary concern, we evaluated both against the toxicity profiles provided by the investigators and the results are included in Table 3. For the two-stage CRM, we took four different approaches to select the initial design and the design parameters. In the first approach, we used the algorithm specified in Section 3 to selected the initial design. The calibration toxicity profiles were the plateau scenarios with an odds ratio of 2. We iterated the δ value between 0.01 and 0.15, with each δ value corresponding to a different set of initial probabilities of DLT and most conservative and coherent initial design. For each δ value, 2000 simulations were performed and the average PCS was calculated across the five calibration scenarios. A δ of 0.04 was selected as it yielded the highest average PCS. The corresponding initial guesses of the probabilities of DLT are $p_0 = \{0.11, 0.17, 0.25, 0.33, 0.42\}$, and the associated most conservative and coherent initial design is $D_0 = \{1, 1, 2, 2, 18\}$. While we performed the simulations to obtain the optimal δ value and corresponding initial design, we note here that we could have obtained these using Table 2 and simply looking up the values for $p = 0.25$, $K = 5$ and a sample size of 25 which suggest using a δ value of 0.04 with the initial design of $D_0 = \{1, 1, 2, 2, 19\}$. Given that we have a sample size of 24, we could remove a patient from the highest dose level and use $D_0 = \{1, 1, 2, 2, 18\}$ without affecting coherence.

In the second approach for selecting the initial design, we performed a tailored calibration by incorporating the safety rule in each simulated trial during the calibration and using Bayesian methods instead of maximum likelihood estimation to estimate the model parameter. The optimal δ value and initial design were selected based directly on simulations conducted using the toxicity profiles provided by the investigators in Table 3, instead of using the calibration scenarios. For each δ value ranging between 0.01 and 0.15, the initial guesses of dose toxicity probabilities were obtained along with the (pruned) most conservative and coherent initial designs. Given the selected initial design and skeleton, 2000 simulations were conducted. A δ value of 0.05 ($p_0 = \{0.08, 0.16, 0.25, 0.35, 0.46\}$) was selected as it yielded the highest average PCS, and the corresponding initial design was $D_0 = \{2, 2, 2, 2, 16\}$.

In the third and fourth approaches for selecting the initial design, we used initial designs with cohort sizes of 3 and 1 separately, and obtained the skeleton using the optimal δ value from the Bayesian CRM (Lee and Cheung, 2009), given $K = 5$, $p = 0.25$ and the third dose level being the prior guess of the MTD. The optimal δ is 0.07 with corresponding initial guesses of probability of DLT, $p_0 = \{0.04, 0.12, 0.25, 0.40, 0.54\}$, for which the cohort of 3 and 1 initial designs are coherent.

Table 3 presents simulation results using the toxicity profiles provided by the investigators, for the 3+3 design and the four approaches to select the initial design and calibrate the two-stage CRM mentioned above. For each approach, we performed 2000 simulations using a sample size of 24 patients and an empiric function $(F(d_k;\beta)=d_k^{exp(\beta)})$ = with the least informative prior distribution $\beta \sim N(0, 0.55)$ in the CRM model (Lee and Cheung, 2011). In contrast, the 3+3 design has a built-in stopping rule, and yielded from 9 to 19 patients on average across the five scenarios. Since the sample size for the CRM can be set to any number of patients depending on resources and patient availability it can provide much

better estimation of the MTD compared to the 3+3 design. The average PCS using the 3+3 design is 34.8%, whereas the proposed algortihm yields an average PCS of 54.8%. In addition, the performance of the two-stage CRM calibrated using the general algorithm presented in section 3 is comparable to the two-stage CRM calibrated using the tailored algorithm specific for this study, with average PCS of 54.8% and 54.4%, respectively. With a more conservative initial design than using a cohort of 1, and coupled with a flatter skeleton, the proposed algorithm achieves similar average PCS (54.8% versus 54.4%), by performing better when the true MTD is at the extremes, but worse when the true MTD is one of the middle dose levels. With a steeper skeleton and a more conservative initial design, the cohort of 3 intial design yields a lower average PCS of 52.0% and performs substantially worse when the true MTD is the highest dose level 51% versus 69%.

As expected, the average DLT proportion was lower for more conservative initial designs using the two-stage CRM. However, it was within 0.05 for an initial design using a cohort size of 1 versus a cohort size of 3, suggesting that a less conservative initial design does not result in a significantly higher DLT proportion. For scenario 1, the 3+3 design, in fact, has the highest proportion of DLTs among all methods considered, due to its different stopping rule.

6. Discussion

In this paper, we have proposed a systematic algorithm to select the initial design for the two-stage CRM. The general algorithm can be utilized to facilitate the selection of the initial design and the design parameters for the two-stage CRM design in practice. The design parameters included in Table 2 can be used as reference and starting point when planning a two-stage CRM trial. Additional design parameters can be evaluated easily using the dfcrm package in R (dfcrm, 2013, R, 2008) and the sample code provided in Appendix B. In situations when the true MTD is thought to be at the extremes and the cohort of 1 initial design is considered too aggressive, our algorithm provides an initial design that is more conservative and yields good operating characteristics.

The literature on the two-stage CRM has not been consistent on the specification of the initial design. While physicians generally prefer slow dose escalation to guarantee enough information on lower doses before escalation, the idea of escalating fast through lower doses by testing only one patient on each dose if no DLT occurs has been proposed previously (O'Quigley and Conaway, 2010). While slow dose escalation is conservative, it can lead to incoherence and poor performance especially with a large number of dose levels. On the other hand, fast dose escalation in the initial design, such as one patient at each dose, always guarantees coherence, but can be too aggressive for clinicians. Our method can identify a more conservative initial design with similar perfomance compared to a cohort of 1 initial design, when possible. Our results do not contradict such aggressive recommendation and suggest that fast escalation is necessary to satisfy the coherence principle especially when the number of doses is large and the target toxicity rate is high. However, under other settings, slower escalation can be allowed without breaching the coherence principle and maintaining similar average proportion of correct selection compared to the cohort of one initial design. Slower escalation is desirable particularly when the clinicians prefer to test

sufficient number of patients on the lower doses to gain reassurance on the safety of the experimental drug at lower doses before proceeding to higher dose levels.

Here we note that although the algorithm presented in this paper is developed for two-stage CRM using likelihood estimation, it can be applied even if Bayesian estimation is adopted in the second stage as illustrated in the oncology dose finding study example. We also like to note that our recommendation is one of many choices for coherent design. As illustrated in our oncology trial example, a more conservative and still coherent initial design may exist when a steeper skeleton using the indifference interval approach is chosen. Howeover, it will have lower average PCS given the calibration scenarios, given that our algorithm selects the initial design and skeleton with the highest PCS. Moreover, a more conservative and still coherent initial design can also be found using trial and error for a skeleton specified not using the indifference interval approach (O'Quigley and Shen, 1996). However, when using trial and error approach or ad-hoc approaches such as the Bayesian one-stage calibration coupled with a cohort of three design, it is important to check for coherence. For example, it should be noted that in the oncology trial that we chose as motivating example if the prior guess of the MTD were to be dose level 1, the optimal δ would be 0.05, for which the cohort of three initial design would be incoherent.

Our method provides a systematic approach to obtain a coherent initial design for the two-stage CRM, which yields good operating characteristics across a wide range of calibration scenarios. While the selection of design parameters in general is very challenging and time consuming by the trial and error approach given the high dimension of design parameters, the method we propose is fast and reproducible, which greatly simplifies the selection of the initial design and the design parameters for the two-stage CRM.

Acknowledgments

SM Lee was supported by the American Cancer Society (grant number MRSG-13-146-01- CPHPS). A Ivanova was supported in part by NIH grant P01 CA142538.

7. Appendix A

Table 4

Optimal indifference interval width δ and initial design D_0 using toxicity calibration scenarios based on an odds ratio of 2 in the proposed algorithm, target probability of DLT p = 0.10, K dose levels and sample size N.

K	λ	N=25		N=30		N=35		N=40	
		δ	D_0	δ	D_0	δ	D_0	δ	D_0
4	1	**0.03**	**(4,5,6,10)**	**0.02**	**(5,5,6,14)**	**0.02**	**(5,5,6,19)**	**0.02**	**(5,5,6,24)**
				0.03	(6,7,7,10)				
	2	**0.04**	**(1,2,2,20)**	**0.03**	**(2,4,4,20)**	**0.03**	**(4,5,6,20)**	**0.02**	**(5,5,6,24)**
		0.02	(1,2,2,20)	0.04	(3,3,4,20)	0.02	(4,5,6,20)	0.03	(6,7,7,20)
		0.03	(1,2,2,20)			0.04			
		0.05	(1,2,2,20)						
		0.06	(1,2,2,20)						

K	*λ*	N=25		N=30		N=35		N=40	
		δ	D_0	*δ*	D_0	*δ*	D_0	*δ*	D_0
5	1	**0.03**	**(3,3,4,5,10)**	**0.03**	**(4,5,5,6,10)**	**0.02**	**(5,5,5,5,15)**	**0.02**	**(5,5,5,5,20)**
		0.04	(3,3,4,5,10)	0.02	(5,5,5,5,10)				
		0.05	(3,4,4,4,10)	0.04	(4,4,6,6,10)				
		0.06	(3,4,4,4,10)						
	2	**0.03**	**(1,1,1,2,20)**	**0.04**	**(2,2,3,3,20)**	**0.04**	**(3,3,4,5,20)**	**0.03**	**(4,5,5,6,20)**
				0.03	(2,2,2,3,20)	0.02	(3,4,4,4,20)	0.02	(5,5,5,5,20)
				0.05	(2,2,3,3,20)	0.03	(3,3,4,5,20)	0.04	(4,4,6,6,20)
6	1	**0.04**	**(2,2,3,4,4,10)**	**0.02**	**(3,3,4,5,5,10)**	**0.02**	**(4,4,4,5,5,13)**	**0.02**	**(4,4,4,5,5,18)**
	2	**0.03**	**(1,1,1,1,1,20)**	**0.04**	**(1,1,2,3,3,20)**	**0.02**	**(2,2,3,4,4,20)**	**0.02**	**(3,3,4,5,5,20)**
7	1	**0.04**	**(2,2,2,2,3,4,10)**	**0.04**	**(3,3,3,3,4,4,10)**	**0.02**	**(3,4,4,4,5,5,10)**	**0.02**	**(4,4,4,4,5,5,14)**
		0.02	(2,2,2,2,3,4,10)	0.02	(3,3,3,3,4,4,10)				
		0.03	(2,2,2,3,3,3,10)	0.03	(3,3,3,3,4,4,10)				
		0.05	(2,2,2,3,3,3,10)	0.05	(3,3,3,3,4,4,10)				
		0.06	(2,2,2,3,3,3,10)	0.06	(3,3,3,3,4,4,10)				
	2	*		**0.04**	**(1,1,1,1,3,3,20)**	**0.03**	**(2,2,2,3,3,3,20)**	**0.02**	**(3,3,3,3,4,4,20)**
				0.02	(1,1,1,1,3,3,20)	0.02	(2,2,2,2,3,4,20)	0.03	(3,3,3,3,4,4,20)
a				0.03	(1,1,2,2,2,2,20)	0.04	(2,2,2,2,3,4,20)		
				0.05	(1,1,2,2,2,2,20)	0.05	(2,2,2,3,3,3,20)		

[a]The additional designs yield average percentage of correct selection within 1 percentage point of the optimal indifference interval width δ and initial design D_0. λ = 1 and 2 imply that the numbers of patients reserved on the highest dose level K are 10 and 20, respectively.

[*]No valid initial designs available.

Table 5

Optimal indifference interval width δ and initial design D_0 using toxicity calibration scenarios based on an odds ratio of 2 in the proposed algorithm, target probability of DLT p = 0.20, K dose levels and sample size N.

K	*λ*	N=25		N=30		N=35		N=40	
		δ	D_0	*δ*	D_0	*δ*	D_0	*δ*	D_0
4	1	**0.04**	**(2,3,3,17)**	**0.04**	**(2,3,3,22)**	**0.04**	**(2,3,3,27)**	**0.04**	**(2,3,3,32)**
		0.03	(2,2,3,18)	0.03	(2,2,3,23)				
	2	**0.04**	**(2,3,3,17)**	**0.04**	**(2,3,3,22)**	**0.04**	**(2,3,3,27)**	**0.04**	**(2,3,3,32)**
		0.03	(2,2,3,18)	0.03	(2,2,3,23)				
5	1	**0.04**	**(2,2,2,3,16)**	**0.04**	**(2,2,2,3,21)**	**0.04**	**(2,2,2,3,26)**	**0.04**	**(2,2,2,3,31)**
		0.03	(2,2,2,2,17)	0.03	(2,2,2,2,22)	0.03	(2,2,2,2,27)	0.03	(2,2,2,2,32)
	2	**0.04**	**(2,2,2,3,16)**	**0.04**	**(2,2,2,3,21)**	**0.04**	**(2,2,2,3,26)**	**0.04**	**(2,2,2,3,31)**
		0.03	(2,2,2,2,17)	0.03	(2,2,2,2,22)	0.03	(2,2,2,2,27)	0.03	(2,2,2,2,32)
6	1	**0.03**	**(1,2,2,2,2,16)**	**0.03**	**(1,2,2,2,2,21)**	**0.03**	**(1,2,2,2,2,26)**	**0.03**	**(1,2,2,2,2,31)**
		0.04	(2,2,2,2,3,14)	0.04	(2,2,2,2,3,19)	0.04	(2,2,2,2,3,24)	0.04	(2,2,2,2,3,29)
	2	**0.03**	**(1,2,2,2,2,16)**	**0.03**	**(1,2,2,2,2,21)**	**0.03**	**(1,2,2,2,2,26)**	**0.03**	**(1,2,2,2,2,31)**
		0.04	(2,2,2,2,3,14)	0.04	(2,2,2,2,3,19)	0.04	(2,2,2,2,3,24)	0.04	(2,2,2,2,3,29)

	K	λ	N=25		N=30		N=35		N=40	
			δ	D_0	δ	D_0	δ	D_0	δ	D_0
[a]	7	1	**0.03**	**(1,1,1,2,2,2,16)**	**0.03**	**(1,1,1,2,2,2,21)**	**0.03**	**(1,1,1,2,2,2,26)**	**0.03**	**(1,1,1,2,2,2,31)**
		2	**0.03**	**(1,1,1,2,2,2,16)**	**0.03**	**(1,1,1,2,2,2,21)**	**0.03**	**(1,1,1,2,2,2,26)**	**0.03**	**(1,1,1,2,2,2,31)**

[a]The additional designs yield average percentage of correct selection within 1 percentage point of the optimal indifference interval width δ and initial design D_0. λ = 1 and 2 imply that the numbers of patients reserved on the highest dose level K are 5 and 10, respectively.

Table 6

Optimal indifference interval width δ and initial design D_0 using toxicity calibration scenarios based on an odds ratio of 2 in the proposed algorithm, target probability of DLT p = 0.25, K dose levels and sample size N.

	K	λ	N=25		N=30		N=35		N=40	
			δ	D_0	δ	D_0	δ	D_0	δ	D_0
	4	1	**0.06**	**(2,2,3,18)**	**0.04**	**(2,2,2,24)**	**0.05**	**(2,2,2,29)**	**0.04**	**(2,2,2,34)**
			0.03	(1,2,2,20)	0.05	(2,2,2,24)	0.04	(2,2,2,29)	0.05	(2,2,2,34)
			0.04	(2,2,2,19)	0.06	(2,2,3,23)	0.06	(2,2,3,28)	0.06	(2,2,3,33)
			0.05	(2,2,2,19)						
		2	**0.06**	**(2,2,3,18)**	**0.04**	**(2,2,2,24)**	**0.05**	**(2,2,2,29)**	**0.04**	**(2,2,2,34)**
			0.03	(1,2,2,20)	0.05	(2,2,2,24)	0.04	(2,2,2,29)	0.05	(2,2,2,34)
			0.04	(2,2,2,19)			0.06	(2,2,3,28)	0.06	(2,2,3,33)
	5	1	**0.04**	**(1,1,2,2,19)**	**0.04**	**(1,1,2,2,24)**	**0.04**	**(1,1,2,2,29)**	**0.04**	**(1,1,2,2,34)**
							0.05	(2,2,2,2,27)		
		2	**0.04**	**(1,1,2,2,19)**	**0.04**	**(1,1,2,2,24)**	**0.04**	**(1,1,2,2,29)**	**0.04**	**(1,1,2,2,34)**
							0.05	(2,2,2,2,27)		
	6	1	**0.04**	**(1,1,1,2,2,18)**	**0.04**	**(1,1,1,2,2,23)**	**0.04**	**(1,1,1,2,2,28)**	**0.04**	**(1,1,1,2,2,33)**
							0.03	(1,1,1,1,2,29)	0.03	(1,1,1,1,2,34)
		2	**0.04**	**(1,1,1,2,2,18)**	**0.04**	**(1,1,1,2,2,23)**	**0.04**	**(1,1,1,2,2,28)**	**0.04**	**(1,1,1,2,2,33)**
							0.03	(1,1,1,1,2,29)	0.03	(1,1,1,1,2,34)
	7	1	**0.04**	**(1,1,1,1,2,2,17)**	**0.04**	**(1,1,1,1,2,2,22)**	**0.04**	**(1,1,1,1,2,2,27)**	**0.04**	**(1,1,1,1,2,2,32)**
							0.03	(1,1,1,1,1,1,29)	0.03	(1,1,1,1,1,1,34)
[a]		2	**0.04**	**(1,1,1,1,2,2,17)**	**0.04**	**(1,1,1,1,2,2,22)**	**0.04**	**(1,1,1,1,2,2,27)**	**0.04**	**(1,1,1,1,2,2,32)**
							0.03	(1,1,1,1,1,1,29)	0.03	(1,1,1,1,1,1,34)

[a]The additional designs yield average percentage of correct selection within 1 percentage point of the optimal indifference interval width δ and initial design D_0. λ = 1 and 2 imply that the numbers of patients reserved on the highest dose level K are 4 and 8, respectively.

Table 7

Optimal indifference interval width δ and initial design D_0 using toxicity calibration scenarios based on an odds ratio of 2 in the proposed algorithm, target probability of DLT p = 0.33, K dose levels and sample size N.

K	*λ*	N=25		N=30		N=35		N=40	
		δ	D_0	*δ*	D_0	*δ*	D_0	*δ*	D_0
4	1	**0.06**	**(1,1,2,21)**	**0.07**	**(1,2,2,25)**	**0.06**	**(1,1,2,31)**	**0.05**	**(1,1,2,36)**
		0.05	(1,1,2,21)	0.06	(1,1,2,26)	0.05	(1,1,2,31)	0.06	(1,1,2,36)
		0.07	(1,2,2,20)			0.07	(1,2,2,25)	0.07	(1,2,2,35)
	2	**0.06**	**(1,1,2,21)**	**0.07**	**(1,2,2,25)**	**0.06**	**(1,1,2,31)**	**0.05**	**(1,1,2,36)**
		0.05	(1,1,2,21)	0.06	(1,1,2,26)	0.05	(1,1,2,31)	0.06	(1,1,2,36)
		0.07	(1,2,2,20)			0.07	(1,2,2,30)	0.07	(1,2,2,35)
5	1	**0.07**	**(1,1,2,2,19)**	**0.06**	**(1,1,1,2,25)**	**0.05**	**(1,1,1,1,31)**	**0.05**	**(1,1,1,1,36)**
		0.05	(1,1,1,1,21)	0.04	(1,1,1,1,26)	0.04	(1,1,1,1,31)	0.04	(1,1,1,1,36)
		0.06	(1,1,1,2,20)	0.05	(1,1,1,1,26)	0.06	(1,1,1,2,30)	0.06	(1,1,1,2,35)
				0.07	(1,1,2,2,24)	0.07	(1,1,2,2,29)	0.07	(1,1,2,2,34)
	2	**0.07**	**(1,1,2,2,19)**	**0.06**	**(1,1,1,2,25)**	**0.05**	**(1,1,1,1,31)**	**0.05**	**(1,1,1,1,36)**
		0.05	(1,1,1,1,21)	0.04	(1,1,1,1,26)	0.04	(1,1,1,1,31)	0.04	(1,1,1,1,36)
		0.06	(1,1,1,2,20)	0.05	(1,1,1,1,26)	0.06	(1,1,1,2,30)	0.06	(1,1,1,2,35)
				0.07	(1,1,2,2,24)	0.07	(1,1,2,2,29)	0.07	(1,1,2,2,34)
6	1	**0.05**	**(1,1,1,1,1,20)**	**0.06**	**(1,1,1,1,2,24)**	**0.06**	**(1,1,1,1,2,29)**	**0.05**	**(1,1,1,1,1,35)**
		0.04	(1,1,1,1,1,20)	0.04	(1,1,1,1,1,25)	0.04	(1,1,1,1,1,30)	0.04	(1,1,1,1,1,35)
		0.06	(1,1,1,1,2,19)	0.05	(1,1,1,1,1,25)	0.05	(1,1,1,1,1,30)	0.06	(1,1,1,1,2,34)
		0.07	(1,1,1,2,2,18)	0.07	(1,1,1,2,2,23)	0.07	(1,1,1,2,2,28)	0.07	(1,1,1,2,2,33)
	2	**0.05**	**(1,1,1,1,1,20)**	**0.06**	**(1,1,1,1,2,24)**	**0.06**	**(1,1,1,1,2,29)**	**0.05**	**(1,1,1,1,1,35)**
		0.04	(1,1,1,1,1,20)	0.04	(1,1,1,1,1,25)	0.04	(1,1,1,1,1,30)	.04	(1,1,1,1,1,35)
		0.06	(1,1,1,1,2,19)	0.05	(1,1,1,1,1,25)	0.05	(1,1,1,1,1,31)	0.06	(1,1,1,1,2,34)
		0.07	(1,1,1,2,2,18)	0.07	(1,1,1,2,2,23)	0.07	(1,1,1,2,2,28)	0.07	(1,1,1,2,2,33)
7	1	**0.05**	**(1,1,1,1,1,1,19)**	**0.05**	**(1,1,1,1,1,1,24)**	**0.05**	**(1,1,1,1,1,1,29)**	**0.05**	**(1,1,1,1,1,1,34)**
		0.06	(1,1,1,1,1,2,18)	0.06	(1,1,1,1,1,2,23)	0.06	(1,1,1,1,1,2,28)	0.06	(1,1,1,1,1,2,33)
		0.07	(1,1,1,1,1,2,18)	0.07	(1,1,1,1,1,2,23)				
	2	**0.05**	**(1,1,1,1,1,1,19)**	**0.05**	**(1,1,1,1,1,1,24)**	**0.05**	**(1,1,1,1,1,1,29)**	**0.05**	**(1,1,1,1,1,1,34)**
a		0.06	(1,1,1,1,1,2,18)	0.06	(1,1,1,1,1,2,23)	0.06	(1,1,1,1,1,2,28)	0.06	(1,1,1,1,1,2,33)
		0.07	(1,1,1,1,1,2,18)	0.07	(1,1,1,1,1,2,23)				

*a*The additional designs yield average percentage of correct selection within 1 percentage point of the optimal indifference interval width *δ* and initial design D_0. *λ* = 1 and 2 imply that the numbers of patients reserved on the highest dose level *K* are 3 and 6, respectively.

8. Appendix B: Sample R Code

```
library(dfcrm)

    n=24 # sample size is 24

    K=5 # number of doses is 5
```

```
target=0.25 # target toxicity probability is 0.25

delta=0.07 # half width of the indifference interval is 0.07

nu=3 # prior guess of the MTD is dose level 3

# generate the skeleton, i.e., initial estimate of dose toxicity using Lee and Cheung algorithm

skeleton=getprior(delta, target, nu, K, model="empiric")

# generate the most conservative and coherent initial design, prune if needed
initial=getinit(skeleton, target, n, nK=10, method="mle") # reserve 10 pts on the highest dose
```

References

Braun TM. The current design of oncology phase I clinical trials: progressing from algorithms to statistical models. Chinese Clinical Oncology. 2014; 3(1)

Cheung YK. Coherence principles in dose-finding studies. Biometrika. 2005; 92(4):863–873.

Cheung, YK. Dose-finding by the continual reassessment method. New York: Chapman & Hall; 2011.

Cheung, YK. dfcrm: Dose-finding by the continual reassessment method. R package version 0.2-2. 2013. http://www.columbia.edu/~yc632

Cheung YK, Chappell R. A simple technique to evaluate model sensitivity in the Continual Reassessment Method. Biometrics. 2002; 58:671–674. [PubMed: 12230003]

Ivanova A, Wang K. Bivariate isotonic design for dose-finding with ordered groups. Statistics in Medicine. 2006; 25:2018–2026. [PubMed: 16220476]

Ivy SP, Siu LL, Garrett-Mayer E, Rubinstein L. Approaches to phase 1 clinical trial design focused on safety, efficiency, and selected patient populations: a report from the clinical trial design task force of the National Cancer Institute investigational drug steering committee. Clin Cancer Res. 2010; 16(6):1726–36. [PubMed: 20215542]

Jia X, Lee SM, Cheung YK. Characterization of likelihood continual reassessment method. Biometrika. 2014; 101(3):599–612.

Lee SM, Cheung YK. Model calibration in the continual reassessment method. Clinical Trials. 2009; 6:227–238. [PubMed: 19528132]

Lee SM, Cheung YK. Calibration of prior variance in the Bayesian continual reassessment method. Statistics in Medicine. 2011; 30(17):2081–2089. [PubMed: 21413054]

National Cancer Institute. Common Terminology Criteria for Adverse Events v4.0. NIH publication 09-7473. 2009

O'Quigley J, Shen L. Continual reassessment method: A likelihood approach. Biometrics. 1996; 52:673–684. [PubMed: 8672707]

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]

O'Quigley J, Paoletti X. Continual reassessment method for ordered groups. Biometrics. 2003; 59:430–440. [PubMed: 12926728]

O'Quigley J, Conaway M. Continual reassessment and related dose-finding designs. Statistical Science. 2010; 24(2):202–216.

R Development Core Team. R Foundation for Statistical Computing. Vienna, Austria: 2008. R: A language and environment for statistical computing. URL http://www.R-project.org

Shen L, O'Quigley J. Consistency of the continual reassessment method under model mispecification. Biometrika. 1996; 83:395–405.

Storer BE. Design and analysis of phase I clinical trials. Biometrics. 1989; 45:925–937. [PubMed: 2790129]

Table 1

Dose levels for the oncology trial and calibrated initial design.

Dose level	Gemcitabine	Cabazitaxel	Cisplatin	Patient number
1	2g	2.5mg	0	1,2
2	2g	5mg	0	3,4
3	2g	5mg	66mg	5,6
4	2g	5mg	80mg	7,8
5	2g	5mg	100mg	9 to 24

Table 2

Optimal indifference interval width δ and initial design D_0 using toxicity calibration scenarios based on an odds ratio of 2 in the proposed algorithm, given target probability of DLT p, number of dose levels K and sample size N.

p	*K*	*λ*	N=25		N=30		N=35		N=40	
			δ	D_0	*δ*	D_0	*δ*	D_0	*δ*	D_0
0.10	4	1	0.03	(4,5,6,10)	0.02	(5,5,6,14)	0.02	(5,5,6,19)	0.02	(5,5,6,24)
		2	0.04	(1,2,2,20)	0.03	(2,4,4,20)	0.03	(4,5,6,20)	0.02	(5,5,6,24)
	5	1	0.03	(3,3,4,5,10)	0.03	(4,5,5,6,10)	0.02	(5,5,5,5,15)	0.02	(5,5,5,5,20)
		2	0.03	(1,1,1,2,20)	0.04	(2,2,3,3,20)	0.04	(3,3,4,5,20)	0.03	(4,5,5,6,20)
	6	1	0.04	(2,2,3,4,4,10)	0.02	(3,3,4,5,5,10)	0.02	(4,4,4,5,5,13)	0.02	(4,4,4,5,5,18)
		2	0.03	(1,1,1,1,1,20)	0.04	(1,1,2,3,3,20)	0.02	(2,2,3,4,4,20)	0.02	(3,3,4,5,5,20)
	7	1	0.04	(2,2,2,2,3,4,10)	0.04	(3,3,3,3,4,4,10)	0.02	(3,4,4,4,5,5,10)	0.02	(4,4,4,4,5,5,14)
		2	*		0.04	(1,1,1,1,3,3,20)	0.03	(2,2,2,3,3,3,20)	0.02	(3,3,3,3,4,4,20)
0.20	4	1	0.04	(2,3,3,17)	0.04	(2,3,3,22)	0.04	(2,3,3,27)	0.04	(2,3,3,32)
		2	0.04	(2,3,3,17)	0.04	(2,3,3,22)	0.04	(2,3,3,27)	0.04	(2,3,3,32)
	5	1	0.04	(2,2,2,3,16)	0.04	(2,2,2,3,21)	0.04	(2,2,2,3,26)	0.04	(2,2,2,3,31)
		2	0.04	(2,2,2,3,16)	0.04	(2,2,2,3,21)	0.04	(2,2,2,3,26)	0.04	(2,2,2,3,31)
	6	1	0.03	(1,2,2,2,2,16)	0.03	(1,2,2,2,2,21)	0.03	(1,2,2,2,2,26)	0.03	(1,2,2,2,2,31)
		2	0.03	(1,2,2,2,2,16)	0.03	(1,2,2,2,2,21)	0.03	(1,2,2,2,2,26)	0.03	(1,2,2,2,2,31)
	7	1	0.03	(1,1,1,2,2,2,16)	0.03	(1,1,1,2,2,2,21)	0.03	(1,1,1,2,2,2,26)	0.03	(1,1,1,2,2,2,31)
0.25	4	1	0.06	(2,2,3,18)	0.04	(2,2,2,24)	0.05	(2,2,2,29)	0.04	(2,2,2,34)
		2	0.06	(2,2,3,18)	0.04	(2,2,2,24)	0.05	(2,2,2,29)	0.04	(2,2,2,34)
	5	1	0.04	(1,1,2,2,19)	0.04	(1,1,2,2,24)	0.04	(1,1,2,2,29)	0.04	(1,1,2,2,34)
		2	0.04	(1,1,2,2,19)	0.04	(1,1,2,2,24)	0.04	(1,1,2,2,29)	0.04	(1,1,2,2,34)
	6	1	0.04	(1,1,1,2,2,18)	0.04	(1,1,1,2,2,23)	0.04	(1,1,1,2,2,28)	0.04	(1,1,1,2,2,33)
		2	0.04	(1,1,1,2,2,18)	0.04	(1,1,1,2,2,23)	0.04	(1,1,1,2,2,28)	0.04	(1,1,1,2,2,33)
	7	1	0.04	(1,1,1,1,2,2,17)	0.04	(1,1,1,1,2,2,22)	0.04	(1,1,1,1,2,2,27)	0.04	(1,1,1,1,2,2,32)
		2	0.04	(1,1,1,1,2,2,17)	0.04	(1,1,1,1,2,2,22)	0.04	(1,1,1,1,2,2,27)	0.04	(1,1,1,1,2,2,32)
0.33	4	1	0.06	(1,1,2,21)	0.07	(1,2,2,25)	0.06	(1,1,2,31)	0.05	(1,1,2,36)
		2	0.06	(1,1,2,21)	0.07	(1,2,2,25)	0.06	(1,1,2,31)	0.05	(1,1,2,36)

p	K	λ	N=25		N=30		N=35		N=40	
			δ	D_0	δ	D_0	δ	D_0	δ	D_0
	5	1	0.07	(1,1,2,2,19)	0.06	(1,1,1,2,25)	0.05	(1,1,1,1,31)	0.05	(1,1,1,1,36)
		2	0.07	(1,1,2,2,19)	0.06	(1,1,1,2,25)	0.05	(1,1,1,1,31)	0.05	(1,1,1,1,36)
	6	1	0.05	(1,1,1,1,1,20)	0.06	(1,1,1,1,2,24)	0.06	(1,1,1,1,2,29)	0.05	(1,1,1,1,1,35)
		2	0.05	(1,1,1,1,1,20)	0.06	(1,1,1,1,2,24)	0.06	(1,1,1,1,2,29)	0.05	(1,1,1,1,1,35)
	7	1	0.05	(1,1,1,1,1,1,19)	0.05	(1,1,1,1,1,1,24)	0.05	(1,1,1,1,1,1,29)	0.05	(1,1,1,1,1,1,34)
		2	0.05	(1,1,1,1,1,1,19)	0.05	(1,1,1,1,1,1,24)	0.05	(1,1,1,1,1,1,29)	0.05	(1,1,1,1,1,1,34)

* indicates that no valid initial designs are available.

Table 3

Operating characteristics of the various approaches for selecting initial design for the oncology trial using two-stage Bayesian CRM. The bolded columns represent the true maximum tolerated doses.

Method	Initial design	Proportion of recommendation						$\bar{N}$	DLT Proportion
Dose level		MTD below level 1	1	2	3	4	5		
pr(DLT)			**0.25**	0.35	0.50	0.65	0.80		
3+3 Design		0.44	**0.36**	0.18	0.02	0.00	0.00	9	0.40
General	(1,1,2,2,18)	0.06	**0.64**	0.26	0.04	0.00	0.00	24	0.36
Tailored	(2,2,2,2,16)	0.06	**0.60**	0.29	0.05	0.00	0.00	24	0.35
Fixed design 1	(1,1,1,1,20)	0.06	**0.54**	0.35	0.05	0.00	0.00	24	0.37
Fixed design 2	(3,3,3,3,12)	0.06	**0.54**	0.35	0.04	0.00	0.00	24	0.35
pr(DLT)			0.15	**0.25**	0.40	0.55	0.70		
3+3 Design		0.21	0.34	**0.33**	0.11	0.01	0.00	11	0.30
General	(1,1,2,2,18)	0.02	0.27	**0.48**	0.21	0.02	0.00	24	0.29
Tailored	(2,2,2,2,16)	0.02	0.22	**0.52**	0.22	0.02	0.00	24	0.28
Fixed design 1	(1,1,1,1,20)	0.02	0.18	**0.57**	0.22	0.01	0.00	24	0.30
Fixed design 3	(3,3,3,3,12)	0.02	0.19	**0.56**	0.21	0.01	0.00	24	0.27
pr(DLT)			0.10	0.15	**0.25**	0.40	0.55		
3+3 Design		0.10	0.19	0.30	**0.30**	0.10	0.02	14	0.23
General	(1,1,2,2,18)	0.01	0.04	0.27	**0.46**	0.21	0.02	24	0.25
Tailored	(2,2,2,2,16)	0.01	0.03	0.27	**0.47**	0.20	0.02	24	0.24
Fixed design 1	(1,1,1,1,20)	0.01	0.02	0.25	**0.54**	0.17	0.01	24	0.25
Fixed design 2	(3,3,3,3,12)	0.01	0.02	0.28	**0.52**	0.16	0.01	24	0.22
pr(DLT)			0.03	0.07	0.15	**0.25**	0.40		
3+3 Design		0.01	0.06	0.17	0.36	**0.29**	0.10	18	0.16
General	(1,1,2,2,18)	0.00	0.00	0.03	0.27	**0.47**	0.23	24	0.22
Tailored	(2,2,2,2,16)	0.00	0.00	0.03	0.28	**0.48**	0.21	24	0.20
Fixed design 1	(1,1,1,1,20)	0.00	0.00	0.03	0.34	**0.51**	0.13	24	0.21
Fixed design 2	(3,3,3,3,12)	0.00	0.00	0.03	0.37	**0.47**	0.14	24	0.17

Method	Initial design	Proportion of recommendation						N	DLT Proportion
pr(DLT)		0.01	0.03	0.07	0.15	**0.25**			
3+3 Design		0.00	0.01	0.05	0.17	0.32	**0.46**	19	0.11
General	(1,1,2,2,18)	0.00	0.00	0.00	0.04	0.27	**0.69**	24	0.17
Tailored	(2,2,2,2,16)	0.00	0.00	0.00	0.04	0.31	**0.65**	24	0.15
Fixed design 1	(1,1,1,1,20)	0.00	0.00	0.00	0.06	0.39	**0.56**	24	0.17
Fixed design 2	(3,3,3,3,12)	0.00	0.00	0.00	0.07	0.42	**0.51**	24	0.12