STATISTICS IN MEDICINE
Statist. Med. 2006; **25**:2365–2383
Published online 11 October 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.2325

An adaptive dose-finding design incorporating both toxicity and efficacy

Wei Zhang[1], Daniel J. Sargent[2,*,†] and Sumithra Mandrekar[2]

[1]*Department of Biostatistics, University of Iowa, Iowa City, IA 52242, U.S.A.*
[2]*Division of Biostatistics, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, U.S.A.*

SUMMARY

Novel therapies are challenging the standards of drug development. Agents with specific biologic targets and limited toxicity require novel designs to determine doses to be taken forward into larger studies. In this paper, we describe an approach that incorporates both toxicity and efficacy data into the estimation of the biologically optimal dose of an agent in a phase I trial. The approach is based on the flexible continuation-ratio model, and uses straightforward optimal dose selection criteria. Dose selection is based on all patients treated up until that time point, using a continual reassessment method approach. Dose–outcome curves considered include monotonically increasing, monotonically decreasing, and unimodal curves. Our simulation studies demonstrate that the proposed design, which we call TriCRM, has favourable operating characteristics. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: phase I trial; Bayesian method; continual reassessment method; trinomial; continuation-ratio model; proportional odds model

1. INTRODUCTION

Phase I clinical trials are generally the first test of new agents in humans. The primary purpose of these trials is to determine the dose of an agent (or a combination of agents) for further study. Historically, in trials of cancer chemotherapy agents, the determination of the dosage for further study has been based solely on toxicity considerations, under the assumption that the dose–toxicity and dose–efficacy relationships are monotone non-decreasing. Based on this assumption, the maximum tolerable dose is assumed to be the dose level with the most promising prospects for efficacy.

*Correspondence to: Daniel J. Sargent, Division of Biostatistics, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, U.S.A.
†E-mail: sargent.daniel@mayo.edu

Contract/grant sponsor: Mayo Clinic Cancer Center; contract/grant number: CA-15083
Contract/grant sponsor: North Central Cancer Treatment Group; contract/grant number: CA-25224

Received 8 September 2004
Accepted 6 June 2005

Recently, the paradigm of cancer drug development has shifted towards less toxic approaches, such as 'targeted' therapy, where an agent is developed to target a specific characteristic of the tumour, such as a receptor, and immunotherapy, where a treatment is designed to stimulate the body's own immune system to attack the tumour. Clinical trials of many such agents are underway, and several such agents are already in clinical use [1]. Pharmacologically, these agents do not follow the standard toxicity/efficacy model, and thus the determination of an optimal dosage from a phase I trial for further testing in phase II or III trials is challenging. Particularly, in immunotherapy trials, often the dose–efficacy relationship is unimodal: at low dose levels there is no efficacy, at some 'optimal' dose level(s) there is a beneficial effect, but at dose levels that exceed the optimal level(s), the therapy causes 'over-stimulation', with either no efficacy, or even life-threatening toxicity.

In this paper we consider dose finding studies of agents where the monotone non-decreasing dose–outcome relationship may not hold. In such cases, the dose-finding algorithm needs to incorporate a measure of efficacy along and no toxicity. We therefore consider the outcome in each patient to be trinomial: no efficacy and no toxicity, efficacy without toxicity, and toxicity. We note that 'efficacy' may take the form of a usual tumour response (shrinkage), or may be based on a presumed surrogate marker, such as the level of a marker from a blood test.

A Bayesian adaptive design for conducting single-arm clinical trials for trinomial outcomes was proposed by Thall and Russell (TR) [2], in which the goals were (1) to find a dose of a new agent that satisfies specific safety and efficacy requirements, (2) to stop the trial early if it is likely that no dose is both safe and efficacious, and otherwise (3) to treat as many subjects as possible at the optimal doses, so participants receive maximal benefit. TR defined the 'no response' outcome, y_0, as no efficacy and no severe toxicity, the 'success' outcome, y_1, as efficacy and no severe toxicity, and the 'toxicity' outcome, y_2, as severe toxicity. Thus, the outcome $y=(y_0, y_1, y_2)$ is trinomial when c subjects are treated at a dose x, with $y_0 + y_1 + y_2 = c$ and $\Psi(x;\theta)=\{\psi_0(x;\theta), \psi_1(x;\theta), \psi_2(x;\theta)\}$ as the corresponding cell probabilities. TR used a proportional odds (PO) model [3] for the relationship between dose and toxicity/efficacy, given by

$$\begin{aligned} \text{logit}\{\psi_1(x;\theta) + \psi_2(x;\theta)\} &= \alpha + \mu + \beta x \\ \text{logit}\{\psi_2(x;\theta)\} &= \alpha + \beta x \end{aligned} \tag{1}$$

where $\theta=(\mu, \alpha, \beta)$, $\text{logit}(x)=\log[x/(1-x)]$, $\mu>0$ to ensure that $\psi_j(x;\theta)$ is a probability distribution on $j=0,1,2$, and $\beta>0$ to ensure the monotonicity assumption for the toxicity outcome. This model assumes that the effect of dose is the same across the cumulative logits, which might sometimes fail. Fan and Chaloner [4] instead proposed the continuation-ratio (CR) model [5] for a trinomial outcome in their non-sequential optimal designs. The CR model does not have the PO assumption. Fan and Chaloner further demonstrated that the CR model fits much better than the PO model for a data set given by Agresti [6].

In this paper, we describe an approach to incorporate both toxicity and efficacy data into the process of estimating the biologically optimal dose (BOD) of an agent. The approach uses a modification of the design proposed by TR [2], but employs the CR model, different optimal dose selection criteria, and different stopping rules. The remainder of this paper is organized

as follows. Section 2 gives a specific motivating example and introduces the CR model and the dose-finding algorithm. In Section 3, we present simulation results comparing the proposed design using the CR model to the design using the PO model in various dose–outcome (i.e. dose–success and dose–toxicity) scenarios for different decision criteria. Section 4 concludes with a discussion of the limitations and strengths.

2. METHODS

2.1. The dose–outcome model

Consider a clinical trial to find the BOD of a new agent which is theorized to boost haemoglobin levels (HbL) for patients whose HbL are below the normal range (14–18 g/dL for men; 12–16 g/dL for women). Following the administration of the therapeutic agent, outcomes of patients could be classified as: 'no response', meaning patients are under-stimulated and their HbL are still below the normal range; 'success', meaning patients are properly stimulated and their HbL are raised to the normal range; and 'toxicity', meaning patients are over-stimulated and their HbL are raised beyond the normal range. Obviously, these 3 outcomes are exclusive and exhaustive, and have probabilities that sum to one. The agent is not expected to cause any non-haemoglobin-related effects, but if such an adverse event were encountered, the patient would be considered to have a 'toxicity' outcome, regardless of his/her HbL measurement. Using the notation defined in Section 1, $\psi_0(x;\theta)+\psi_1(x;\theta)+\psi_2(x;\theta)=1$ for any dose x and some parameter set θ. We can reasonably assume (1) the probability of no response $\psi_0(x;\theta)$ decreases monotonically with dose, and (2) the probability of toxicity $\psi_2(x;\theta)$ increases monotonically with dose. However, we must allow the probability of the desired success outcome, $\psi_1(x;\theta)$, to be non-monotone in dose. This can be accomplished by the CR model given by

$$\begin{aligned} \log\left\{\frac{\psi_1(x;\theta)}{\psi_0(x;\theta)}\right\} &= \alpha_1+\beta_1 x \\ \operatorname{logit}\{\psi_2(x;\theta)\} &= \alpha_2+\beta_2 x \end{aligned} \tag{2}$$

where $\theta=(\mu,\beta_1,\alpha_2,\beta_2)$, $\alpha_2=\alpha_1-\mu$, and $\mu,\beta_1,\beta_2>0$. Solving the above two equations, we have

$$\begin{aligned} \psi_2(x;\theta) &= \frac{\exp(\alpha_2+\beta_2 x)}{1+\exp(\alpha_2+\beta_2 x)} \\ \psi_1(x;\theta) &= \frac{\exp(\alpha_1+\beta_1 x)}{[1+\exp(\alpha_1+\beta_1 x)][1+\exp(\alpha_2+\beta_2 x)]} \\ \psi_0(x;\theta) &= \frac{1}{[1+\exp(\alpha_1+\beta_1 x)][1+\exp(\alpha_2+\beta_2 x)]} \end{aligned} \tag{3}$$

It is clear that the toxicity probability $\psi_2(x;\theta)$ is a monotone increasing function of dose x; the no-response probability $\psi_0(x;\theta)$ is a monotone non-increasing function of dose; and the success probability $\psi_1(x;\theta)$ is unimodal, but can be monotone increasing or monotone decreasing for a dose range to be studied depending on the value of the parameter vector θ. Thus, equation (3) satisfies exactly our assumptions.

O'Quigley *et al.* [7] introduced the Bayesian approach into sequential phase I design using the continual reassessment method (CRM). Since this introduction, the concept of generating decision rules using Bayesian criteria and then evaluating the operating characteristics of the design obtained via simulation is becoming widely accepted [8–15]. Bayesian methods for dose-finding trials are attractive due to the sequential nature of the data collection. In general, the required elements for the posterior distribution are the CR model representing the dose–outcome relationship, and a prior distribution for the parameter vector θ. As the data (x, y) become available from each successive subject (or cohort of subjects), the posterior distribution of θ is updated, and then the posterior mean of θ is used to estimate the probabilities of success and toxicity at each dose level. The next patient (or cohort) is then allocated to the dose level at which the estimated probabilities of both success and toxicity are desirable. When pre-specified stopping rules have been met, the BOD is selected according to pre-defined decision criteria, based on the data from all the patients treated.

Let n be the number of cohorts treated at the current time, $\mathbf{x}$ be an $n \times 1$ dose vector with element x_i, and $\mathbf{y}$ be an $n \times 3$ outcome matrix with y_i as the ith row, for $i = 1, \ldots, n$. Given the dose–outcome model and current data $(\mathbf{x}, \mathbf{y})$, the likelihood function is

$$L(\theta|\mathbf{x},\mathbf{y}) \propto \prod_{i=1}^{n} \psi_0(x_i;\theta)^{y_{0i}} \psi_1(x_i;\theta)^{y_{1i}} \psi_2(x_i;\theta)^{y_{2i}} \tag{4}$$

where x_i is the dose administered to the ith cohort, $y_i = (y_{0i}, y_{1i}, y_{2i})$ is the trinomial outcome of the ith cohort with size $c_i = (y_{0i} + y_{1i} + y_{2i})$ at dose x_i, and $\psi_j(\cdot)$, $j = 0, 1, 2$, are the probabilities given by (3). This likelihood function is updated as data from each successive cohort becomes available. If only a binary toxicity outcome measure (toxic *versus* non-toxic) is available, we can simplify (4) as

$$L^*(\theta_2|\mathbf{x},\mathbf{y}) \propto \prod_{i=1}^{n} \psi_2(x_i;\theta_2)^{y_{2i}} [1 - \psi_2(x_i;\theta_2)]^{c_i - y_{2i}} \tag{5}$$

where $\theta_2 = (\alpha_2, \beta_2)$. This is the likelihood function used in the usual CRM based on a two-parameter logistic model without consideration of efficacy data.

Finally, the prior joint probability density function (PDF), $f(\theta)$, for the parameter vector θ needs to be specified. If strong prior information is available regarding the dose–outcome relationship, this information could be incorporated through this prior. Alternatively, the prior joint PDF of θ could be a multivariate normal distribution with some mean vector and appropriate variance matrix, such that the prior belief is vague. However, for simplicity and to lessen the computational burden, we use a suitable non-informative prior, the uniform distribution, for each parameter. Specifically,

$$\mu \sim \text{Unif}[0, u_1], \quad \beta_1 \sim \text{Unif}[0, u_2], \quad \alpha_2 \sim \text{Unif}[l, u_3], \quad \beta_2 \sim \text{Unif}[0, u_4] \tag{6}$$

where (l, u_1, u_2, u_3, u_4) are constants. This way, the posterior joint distribution of θ, $f(\theta|\mathbf{x}, \mathbf{y})$, is proportional to the likelihood function given in (4) over the bounded region.

2.2. Bayesian parameter estimation

We employ a Bayesian procedure to update the estimate of the parameter vector θ. In particular, the posterior mean of θ is estimated by $\hat{\theta} = (\hat{\mu}, \hat{\beta}_1, \hat{\alpha}_2, \hat{\beta}_2)$, where

$$\hat{\mu} = \frac{\int_{\mu}\int_{\beta_2}\int_{\beta_1}\int_{\alpha_2} \mu f(\theta) L(\theta|\mathbf{x}, \mathbf{y})\, \mathrm{d}\alpha_2\, \mathrm{d}\beta_1\, \mathrm{d}\beta_2\, \mathrm{d}\mu}{\int_{\theta} f(\theta) L(\theta|\mathbf{x}, \mathbf{y})\, \mathrm{d}\theta} \tag{7}$$

Estimates $\hat{\beta}_1$, $\hat{\alpha}_2$, and $\hat{\beta}_2$ could be obtained similarly. However, the numerators and denominators in (7) involve four-dimension integration, which is computationally intensive and potentially unstable, particularly when sample size n is small in early stage of the trial.

Let $f_1(\theta_1)$ and $f_2(\theta_2)$ be prior PDFs for $\theta_1 = (\mu, \beta_1)$ and $\theta_2 = (\alpha_2, \beta_2)$, respectively. An alternative estimation approach first estimates θ_2 using the likelihood function $L^*(\cdot)$ in (5) for toxicity data, then obtains the estimate of θ_1 based on the pseudo-likelihood $L(\theta_1, \hat{\theta}_2|\mathbf{x}, \mathbf{y})$. Specifically, we can estimate θ_2 by $\hat{\theta}_2 = (\hat{\alpha}_2, \hat{\beta}_2)$, where

$$\begin{aligned}
\hat{\alpha}_2 &= \frac{\int_{\alpha_2}\int_{\beta_2} \alpha_2 f_2(\theta_2) L^*(\theta_2|\mathbf{x}, \mathbf{y})\, \mathrm{d}\beta_2\, \mathrm{d}\alpha_2}{\int_{\theta_2} f_2(\theta_2) L^*(\theta_2|\mathbf{x}, \mathbf{y})\, \mathrm{d}\theta_2} \\
\hat{\beta}_2 &= \frac{\int_{\beta_2}\int_{\alpha_2} \beta_2 f_2(\theta_2) L^*(\theta_2|\mathbf{x}, \mathbf{y})\, \mathrm{d}\alpha_2\, \mathrm{d}\beta_2}{\int_{\theta_2} f_2(\theta_2) L^*(\theta_2|\mathbf{x}, \mathbf{y})\, \mathrm{d}\theta_2}
\end{aligned} \tag{8}$$

Given $\hat{\theta}_2$, θ_1 is estimated by $\hat{\theta}_1 = (\hat{\mu}, \hat{\beta}_1)$, where

$$\begin{aligned}
\hat{\mu} &= \frac{\int_{\mu}\int_{\beta_1} \mu f_1(\theta_1) L(\theta_1, \hat{\theta}_2|\mathbf{x}, \mathbf{y})\, \mathrm{d}\beta_1\, \mathrm{d}\mu}{\int_{\theta_1} f_1(\theta_1) L(\theta_1, \hat{\theta}_2|\mathbf{x}, \mathbf{y})\, \mathrm{d}\theta_1} \\
\hat{\beta}_1 &= \frac{\int_{\beta_1}\int_{\mu} \beta_1 f_1(\theta_1) L(\theta_1, \hat{\theta}_2|\mathbf{x}, \mathbf{y})\, \mathrm{d}\mu\, \mathrm{d}\beta_1}{\int_{\theta_1} f_1(\theta_1) L(\theta_1, \hat{\theta}_2|\mathbf{x}, \mathbf{y})\, \mathrm{d}\theta_1}
\end{aligned} \tag{9}$$

We performed the integration in (8) and (9) numerically using adaptive Gauss-quadrature [16] via the *R* package *adapt* contributed by Genz *et al.* [17]. When using the uniform distribution specified in (6) for the θ_1 and θ_2, the prior joint distributions $f_1(\theta_1)$ and $f_2(\theta_2)$ in the numerators and denominators in (8) and (9) are constants and cancel out.

2.3. Proposed dose-finding algorithm

Let π_0 be the maximum tolerable toxicity probability allowed, which is specified *a priori* by the clinical investigators. We define two decision functions $\delta_1(x;\theta)$ and $\delta_2(x;\theta)$ to be

$$\begin{aligned}\delta_1(x;\theta)&=I_{[\psi_2(x;\theta)<\pi_0]}\\ \delta_2(x;\theta)&=\psi_1(x;\theta)-\lambda\cdot\psi_2(x;\theta)\end{aligned} \tag{10}$$

where I is an indicator function, and $0\leqslant\lambda\leqslant 1$ is the weight for the toxicity probability $\psi_2(x;\theta)$. These decision functions allow us to jointly specify a toxicity and an efficacy decision criteria as follows: the toxicity criterion requires $\delta_1(x;\theta)=1$, that is, the toxicity probability at the dose x is smaller than the pre-specified π_0, and the efficacy criterion is $\delta_2(x^*;\theta)=\max_{x\in C(x)}\{\delta_2(x;\theta)\}$, that is, in the range of a given dose set, the dose x^* maximizes either the difference between the success probability and toxicity probability if $\lambda=1$, or the success probability if $\lambda=0$. We first use the toxicity criterion as a screening rule to provide a subset of dose levels $C(x)$ satisfying $\delta_1(x;\theta)=1$. Among $C(x)$ we then find the dose x^* that satisfies the efficacy criterion $\delta_2(x^*;\theta)=\max_{x\in C(x)}\{\delta_2(x;\theta)\}$. The BOD is the recommended dose level when the trial ends. If restricting toxicity to be less than π_0 is sufficient for a trial, we can specify $\lambda=0$, which defines the BOD as the dose x^* that maximizes the success probability subject to $\delta_1(x^*;\theta)=1$. The value of λ can be varied to include toxicity in the efficacy criterion.

Specifically, we propose that the trial be conducted based on the following rules:

1. Assign a cohort of size c to a single dose level at a time, starting from the lowest dose level.
2. Dose levels cannot be skipped during escalation, however there is no restriction on de-escalations.
3. Given the accumulated data at every interim point in the trial, update $\hat{\theta}$ using equations (8) and (9), and evaluate $\delta_1(x;\hat{\theta})$ for each dose level.
   (a) If $\delta_1(x;\hat{\theta})=0$ for all x (i.e. no dose level satisfies the toxicity criterion) and the current dose level is the lowest level, the trial is terminated and no dose is recommended as the BOD.
   (b) If $\delta_1(x;\hat{\theta})=0$ for all x and the current dose level is not the lowest dose level, then assign the next cohort to the lowest dose level.
   (c) Otherwise, find the subset $C(x)$.
4. Assign the next cohort to the dose x^* from $C(x)$ that satisfies $\delta_2(x^*;\hat{\theta})=\max_{x\in C(x)}\{\delta_2(x;\hat{\theta})\}$.
5. The trial is terminated after treating at least n_1 patients, provided at least m subjects are treated at the recommended dose level, or a maximum of n_2 subjects are treated, whichever comes first.

Rules 1, 2 and 5 are similar to those used in the modified CRM design based on binary toxicity data for dose-finding in a conventional phase I trial [9, 11, 12]. Rule 3(a) enhances the safety of a trial, and takes precedence over rule 5. Our design is more general than the modified CRM in that we summarize patient outcome by a trinomial ordinal variable that accounts for both efficacy and toxicity outcomes, and we require that the trial be terminated early if it is likely that all of the dose levels being considered are unacceptable in terms of

toxicity. This design is a modified version of TR's design [1], in that we use the CR model instead of PO model, and adopt different decision criteria and stopping rules. For convenience, we will refer to our design as the TriCRM design in what follows.

3. SIMULATIONS

3.1. Simulation settings

We assess the operating characteristics of the TriCRM design under nine different scenarios that are relevant to the HbL trial described in Section 2.1. These are scenarios that we feel are meaningful and practical. We consider six dose levels ranging from 0.60 to 1.35 mcg/kg, with the fixed doses $x_i = 0.60, 0.75, 0.90, 1.05, 1.20$ and 1.35. For the purpose of comparison, we also implemented the TriCRM design using the PO model instead of the CR model. The true toxicity and success curves for the nine scenarios investigated are displayed in Figure 1.

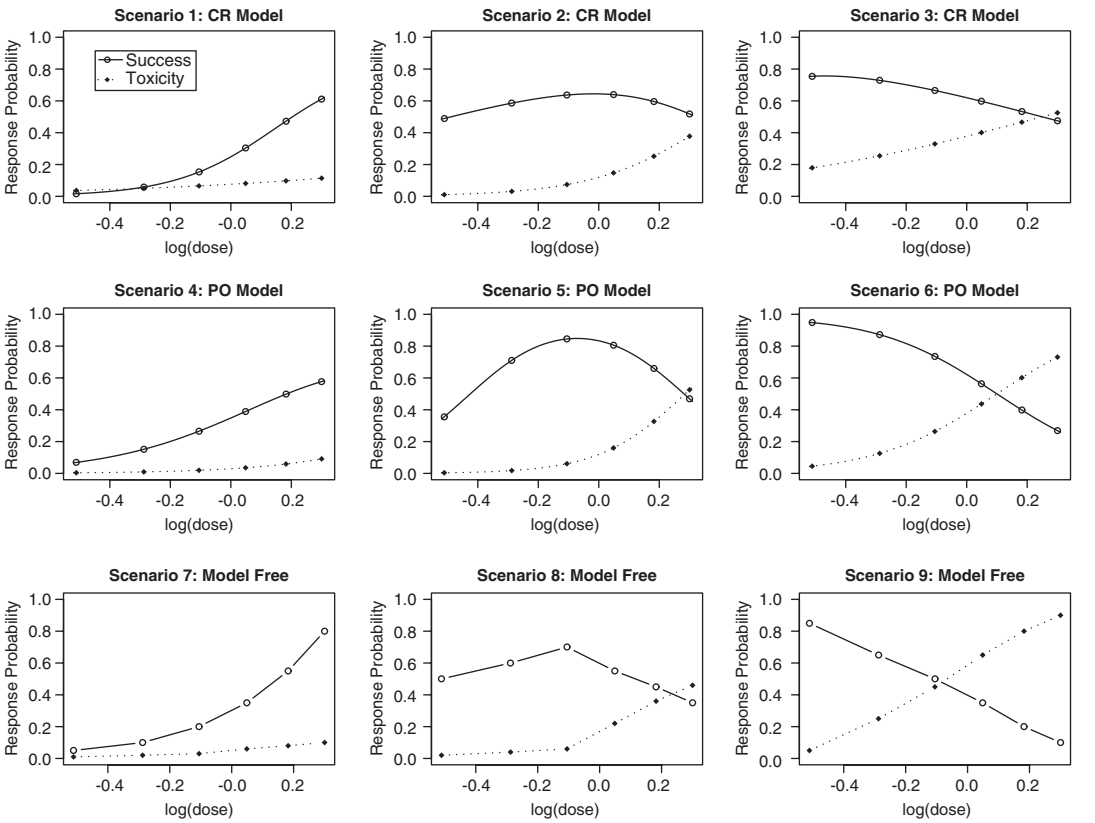

Figure 1. Dose–toxicity and dose–success curves for the nine scenarios.

Dose–outcome curves in scenarios 1–3 were generated based on the CR model; curves in scenarios 4–6 were generated from the PO model; and curves in scenarios 7–9 were not based on any model. These nine scenarios cover three monotone increasing (scenarios 1, 4 and 7), three unimodal (scenarios 2, 5 and 8), and three monotone decreasing (scenarios 3, 6, and 9) dose–success curves. The dose–toxicity curves are monotone increasing in all nine scenarios.

We employed fairly broad uniform priors for parameters (μ, β_1, α_2, and β_2) to encompass these scenarios and also some more extreme situations that might possibly occur. In particular, we chose $u_1 = u_2 = u_4 = 10$, $u_3 = 5$, and $l = -10$ in (6) for the simulations. The priors for parameters (μ, α, β) of the PO model are the same as the ones for parameters (μ, α_2, β_2) of the CR model, respectively. The highest tolerable toxicity probability π_0 was set at 0.33. The simulations were performed for the efficacy function $\delta_2(x;\theta) = \psi_1(x;\theta) - \lambda \cdot \psi_2(x;\theta)$ with $\lambda = 1$ and 0. We used the cohort size $c = 3$, the minimum sample size $n_1 = 18$, the maximum sample size $n_2 = 30$, and the required number of subjects treated at the BOD $m = 9$. The outcomes of each subject in the simulation were generated via a random trinomial generation function. Each simulation consisted of 5000 simulated trials. For each simulation, we computed the percentage of experimentation at each dose, the percentage of recommendation for each dose, the percentage of no recommendation for any dose, the average and interquartile range (IQR) sample sizes, the overall percentage of patients with a toxic outcome observed, and the overall percentage of patients with a successful outcome observed.

3.2. Simulation results

3.2.1. Maximizing the difference between $\psi_1(x;\theta)$ and $\psi_2(x;\theta)$ subject to $\psi_2(x;\theta) < 0.33$ ($\lambda = 1$). Results for the nine scenarios using the efficacy function $\delta_2(x;\theta) = \psi_1(x;\theta) - \psi_2(x;\theta)$ are shown in Table I.

The dose–success curves are monotone increasing in scenarios 1, 4, and 7 with dose level 6 as the BOD; the dose–success curves are unimodal in scenarios 2, 5, and 8 with dose level 3 as the BOD (for scenario 2, both dose levels 2 and 3 are the BODs); the dose–success curves are monotone decreasing in scenarios 3, 6 and 9 with dose level 1 as the BOD.

In scenarios 1, 2 and 3, the dose–outcome curves are based on the CR model. For scenario 1, dose level 6 was recommended in 68 per cent of the simulated trials by the CR model, and 47 per cent by the PO model. In scenario 2, the CR model recommended either dose level 2 or 3 in 73 per cent of simulations, compared to 61 per cent under the PO model. Both models correctly recommended dose level 1 as the BOD 75 per cent of the times for scenario 3. The dose–outcome curves are generated from the PO model in scenarios 4–6. Both models identified the BOD in about 80 per cent of the simulated trials for scenario 4; both models correctly recommended the BOD in about 70 per cent of the simulated trials for scenario 5; the CR model correctly identified the BOD 5 per cent more times than the PO model in scenarios 6. In the model free scenarios 7 and 9, the CR model performed better than the PO model by 9 and 4 per cent, respectively, in terms of identifying the BOD. Both models correctly identified the BOD about 57 per cent of the times in scenario 8.

3.2.2. Maximizing $\psi_1(x;\theta)$ subject to $\psi_2(x;\theta) < 0.33$ ($\lambda = 0$). Table II displays the results for the nine scenarios using the efficacy function $\delta_2(x;\theta) = \psi_1(x;\theta)$. In terms of identifying the BOD, the performance of the CR model was similar to the PO model for scenarios 1–7.

Table I. TriCRM performance for $\lambda = 1$.

Model	Scenario	Dose level 1	2	3	4	5	6	No recom. (%)	Sample size Mean	IQR	Observed (%) Success	Toxicity
	1	(0.02,0.04)*	(0.06,0.05)	(0.15,0.07)	(0.30,0.08)	(0.47,0.10)	(0.61,0.11)					
		$\delta_2 = -0.02$	0.01	0.08	0.22	0.37	**0.50**					
PO		30.7(28.9)†	1.9(13.8)	2.9(12.3)	7.6(12.6)	9.4(12.8)	**46.8**(19.5)	0.7	23.1	{18,27}	25	7
CR		5.1(15.1)	1.7(14.4)	7.8(15.7)	7.9(15.0)	9.5(14.8)	**67.5**(24.9)	0.5	25.0	{24,27}	30	8
	2	(0.49,0.01)	(0.59,0.03)	(0.64,0.07)	(0.64,0.15)	(0.60,0.25)	(0.52,0.38)					
		$\delta_2 = 0.48$	**0.56**	**0.57**	0.49	0.35	0.14					
PO		0.8(15.3)	**14.3**(21.5)	**46.8**(28.8)	32.3(21.7)	4.7(9.0)	1.1(3.7)	0.0	22.5	{18,24}	60	10
CR		1.5(15.8)	**24.4**(23.8)	**48.9**(29.5)	22.0(19.1)	2.4(8.1)	0.8(3.6)	0.1	22.5	{18,24}	60	9
	3	(0.75,0.18)	(0.73,0.25)	(0.67,0.33)	(0.60,0.40)	(0.53,0.47)	(0.47,0.53)					
		$\delta_2 = \mathbf{0.57}$	0.48	0.34	0.20	0.06	−0.06					
PO		**75.3**(80.6)	1.6(13.4)	0.0(4.4)	0.0(1.3)	0.0(0.3)	0.0(0.0)	23.1	15.7	{18,18}	74	20
CR		**75.0**(80.9)	2.0(13.0)	0.0(4.4)	0.0(1.3)	0.0(0.3)	0.0(0.0)	23.0	15.7	{18,18}	74	20
	4	(0.07,0.00)	(0.15,0.01)	(0.27,0.02)	(0.39,0.04)	(0.50,0.06)	(0.58,0.09)					
		$\delta_2 = 0.07$	0.14	0.25	0.35	0.44	**0.49**					
PO		1.7(13.1)	0.4(12.7)	1.1(12.9)	5.2(14.6)	11.7(16.8)	**79.8**(30.0)	0.0	25.1	{24,27}	38	5
CR		0.0(11.9)	0.2(12.4)	3.2(13.7)	7.2(15.3)	11.4(16.8)	**78.0**(29.9)	0.0	25.2	{24,27}	38	5
	5	(0.36,0.00)	(0.71,0.02)	(0.84,0.06)	(0.81,0.16)	(0.66,0.33)	(0.47,0.53)					
		$\delta_2 = 0.36$	0.69	**0.78**	0.65	0.33	−0.06					
PO		0.2(14.6)	19.3(24.3)	**70.9**(36.3)	9.4(16.2)	0.2(6.5)	0.0(2.1)	0.0	21.3	{18,24}	72	9
CR		0.1(14.7)	20.5(24.0)	**68.7**(36.0)	10.6(16.7)	0.1(6.5)	0.0(2.0)	0.0	21.3	{18,24}	71	9
	6	(0.95,0.05)	(0.87,0.13)	(0.74,0.26)	(0.56,0.44)	(0.40,0.60)	(0.27,0.73)					
		$\delta_2 = \mathbf{0.90}$	0.74	0.48	0.12	−0.20	−0.46					

Table I. *Continued.*

Model	Scenario	Dose level 1	2	3	4	5	6	No recom. (%)	Sample size Mean	IQR	Observed (%) Success	Toxicity
PO		**93.0**(68.7)	5.7(17.6)	0.0(9.1)	0.0(3.8)	0.0(0.7)	0.0(0.0)	1.2	18.3	{18,18}	90	10
CR		**98.1**(71.5)	0.5(14.7)	0.0(9.4)	0.0(3.8)	0.0(0.6)	0.0(0.0)	1.4	18.0	{18,18}	90	10
	7	(0.05,0.01)	(0.10,0.02)	(0.20,0.03)	(0.35,0.06)	(0.55,0.08)	(0.80,0.10)					
		$\delta_2 = 0.04$	0.08	0.17	0.29	0.47	**0.70**					
PO		7.3(16.4)	1.1(13.2)	2.5(13.3)	8.6(14.9)	16.3(16.8)	**64.3**(25.4)	0.0	24.9	{24,27}	39	6
CR		0.4(12.4)	0.6(12.9)	5.2(14.5)	6.9(15.4)	13.3(16.8)	**73.6**(28.0)	0.0	25.2	{24,27}	42	6
	8	(0.50,0.02)	(0.60,0.04)	(0.70,0.06)	(0.55,0.22)	(0.45,0.36)	(0.35,0.46)					
		$\delta_2 = 0.48$	0.56	**0.64**	0.33	0.09	−0.11					
PO		2.5(17.6)	17.8(23.6)	**57.7**(31.3)	19.6(19.4)	1.8(6.4)	0.4(1.7)	0.2	21.6	{18,24}	59	10
CR		3.2(18.1)	29.6(26.7)	**56.2**(31.3)	10.1(16.8)	0.6(5.6)	0.2(1.5)	0.1	21.7	{18,24}	59	10
	9	(0.85,0.05)	(0.65,0.25)	(0.50,0.45)	(0.35,0.65)	(0.20,0.80)	(0.10,0.90)					
		$\delta_2 = \mathbf{0.80}$	0.40	0.05	−0.30	−0.60	−0.80					
PO		**89.3**(70.5)	9.0(22.2)	0.0(6.3)	0.0(1.0)	0.0(0.0)	0.0(0.0)	1.7	18.3	{18,18}	78	12
CR		**92.8**(73.4)	5.5(19.1)	0.0(6.3)	0.0(1.0)	0.0(0.1)	0.0(0.0)	1.7	18.1	{18,18}	78	12

Note: Scenarios 1–3 are based on the CR model; scenarios 4–6 are based on the PO model; scenarios 7–9 are model free.

*The success probability at this dose level, the toxicity probability at this dose level.

†The percentage of times this dose level is identified as the BOD (the percentage of subjects treated at this dose level during the trial).

Table II. TriCRM performance for $\lambda = 0$.

Model	Scenario	Dose level 1	2	3	4	5	6	No recom. (%)	Sample size Mean	IQR	Observed (%) Success	Toxicity
	1	(0.02,0.04)*	(0.06,0.05)	(0.15,0.07)	(0.30,0.08)	(0.47,0.10)	(0.61,0.11)					
		$\delta_2 = 0.02$	0.06	0.15	0.30	0.47	**0.61**					
PO		0.3(13.8)†	0.9(14.4)	1.8(12.9)	1.9(13.2)	5.5(14.3)	**88.9**(31.4)	0.8	25.0	{24,27}	33	8
CR		0.4(13.7)	0.9(14.5)	1.7(12.9)	1.7(13.1)	5.5(14.4)	**89.0**(31.5)	0.7	25.0	{24,27}	33	8
	2	(0.49,0.01)	(0.59,0.03)	(0.64,0.07)	(0.64,0.15)	(0.60,0.25)	(0.52,0.38)					
		$\delta_2 = 0.49$	0.59	**0.64**	**0.64**	0.60	0.52					
PO		0.1(13.5)	3.1(15.9)	**25.6**(23.1)	**50.7**(26.9)	17.4(15.0)	3.2(5.4)	0.0	23.5	{21,27}	60	12
CR		0.1(13.6)	5.3(16.8)	**33.5**(24.7)	**45.7**(25.7)	12.6(13.6)	2.8(5.5)	0.0	23.6	{21,27}	60	12
	3	(0.75,0.18)	(0.73,0.25)	(0.67,0.33)	(0.60,0.40)	(0.53,0.47)	(0.47,0.53)					
		$\delta_2 = \mathbf{0.75}$	0.73	0.67	0.60	0.53	47					
PO		**71.0**(76.0)	6.0(17.9)	0.2(4.5)	0.0(1.3)	0.0(0.3)	0.0(0.1)	22.8	15.8	{18,18}	74	20
CR		**69.7**(75.6)	7.3(18.1)	0.2(4.6)	0.0(1.3)	0.0(0.3)	0.0(0.1)	22.8	15.8	{18,18}	74	20
	4	(0.07,0.00)	(0.15,0.01)	(0.27,0.02)	(0.39,0.04)	(0.50,0.06)	(0.58,0.09)					
		$\delta_2 = 0.07$	0.15	0.27	0.39	0.50	**0.58**					
PO		0.0(12.3)	0.0(12.7)	0.1(12.4)	0.3(12.7)	3.4(14.4)	**96.2**(35.5)	0.0	24.5	{24,24}	39	5
CR		0.0(12.3)	0.0(12.7)	0.1(12.4)	0.6(12.8)	3.3(14.4)	**96.0**(35.5)	0.0	24.5	{24,24}	39	5
	5	(0.36,0.00)	(0.71,0.02)	(0.84,0.06)	(0.81,0.16)	(0.66,0.33)	(0.47,0.53)					
		$\delta_2 = 0.36$	0.71	**0.84**	0.81	0.66	0.47					
PO		0.0(13.9)	5.2(17.8)	**63.2**(33.9)	30.9(24.0)	0.7(8.0)	0.0(2.3)	0.0	21.8	{18,24}	72	10
CR		0.0(13.9)	5.3(18.1)	**60.7**(33.4)	33.3(24.7)	0.8(7.6)	0.0(2.2)	0.0	21.8	{18,24}	72	10
	6	(0.95,0.05)	(0.87,0.13)	(0.74,0.26)	(0.56,0.44)	(0.40,0.60)	(0.27,0.73)					
		$\delta_2 = \mathbf{0.95}$	0.87	0.74	0.56	0.40	0.27					

Table II. *Continued.*

Model	Scenario	Dose level 1	2	3	4	5	6	No recom. (%)	Sample size Mean	IQR	Observed (%) Success	Toxicity
PO		**92.8**(68.3)	5.8(17.8)	0.1(9.4)	0.0(3.8)	0.0(0.7)	0.0(0.0)	1.2	18.3	{18,18}	90	10
CR		**93.7**(69.2)	4.9(17.1)	0.0(9.1)	0.0(3.7)	0.0(0.7)	0.0(0.0)	1.4	18.2	{18,18}	90	10
	7	(0.05,0.01)	(0.10,0.02)	(0.20,0.03)	(0.35,0.06)	(0.55,0.08)	(0.80,0.10)					
		$\delta_2 = 0.05$	0.10	0.20	0.35	0.55	**0.80**					
PO		0.0(12.5)	0.1(13.0)	0.3(12.4)	1.1(13.0)	6.1(15.2)	**92.3**(33.9)	0.1	24.7	{24,24}	44	6
CR		0.0(12.6)	0.1(13.0)	0.3(12.5)	1.1(13.0)	5.0(14.8)	**93.4**(34.2)	0.1	24.7	{24,24}	44	6
	8	(0.50,0.02)	(0.60,0.04)	(0.70,0.06)	(0.55,0.22)	(0.45,0.36)	(0.35,0.46)					
		$\delta_2 = 0.50$	0.60	**0.70**	0.55	0.45	0.35					
PO		0.2(15.0)	5.1(18.0)	**41.2**(27.6)	43.5(26.4)	8.3(10.4)	1.4(2.5)	0.2	22.5	{18,24}	58	13
CR		0.2(14.8)	7.4(18.6)	**51.5**(29.3)	34.8(25.0)	4.8(9.7)	1.1(2.5)	0.2	22.8	{21,24}	58	13
	9	(0.85,0.05)	(0.65,0.25)	(0.50,0.45)	(0.35,0.65)	(0.20,0.80)	(0.10,0.90)					
		$\delta_2 = \mathbf{0.85}$	0.65	0.50	0.35	0.20	0.10					
PO		**70.4**(59.4)	27.5(32.1)	0.5(7.4)	0.0(1.0)	0.0(0.0)	0.0(0.0)	1.5	18.5	{18,18}	75	15
CR		**74.1**(61.8)	24.2(29.8)	0.4(7.3)	0.0(1.0)	0.0(0.0)	0.0(0.0)	1.3	18.5	{18,18}	76	14

Note: Scenarios 1–3 are based on the CR model; scenarios 4–6 are based on the PO model; scenarios 7–9 are model free.

*The success probability at this dose level, the toxicity probability at this dose level.

†The percentage of times this dose level is identified as the BOD (the percentage of subjects treated at this dose level during the trial).

For the model free scenarios 8 and 9, the CR model outperformed the PO model by 10 and 4 per cent, respectively, in terms of identifying the BOD.

For both $\lambda = 1$ and 0:

1. The percentage of recommending no dose is ignorable when the BOD exists except in scenario 3 where about 23 per cent of the simulated trials failed to recommend the BOD due to relatively high toxicity at low dose levels.
2. The average sample size ranges from 16 to 25 depending on the BOD.
3. The CR model and the PO model differ little in other aspects compared: the allocation pattern, the sample sizes (mean, IQR), and the observed success and toxicity percentages.

3.3. *Design variation*

The TriCRM design described in Section 2.3 does not permit skipping dose levels during dose escalation. As a result, in scenarios 1, 4 and 7, where the BOD is the last dose level, the percentage of subjects allocated to the BOD by the TriCRM is not satisfactory. For scenarios 1, 4 and 7, the experimentation percentage for the BOD is only about 25–30 per cent for $\lambda = 1$, which is relatively low comparing to other scenarios (Table I). We therefore repeated the simulation in scenario 1 for $\lambda = 1$, but without the restriction that dose escalation proceeds at only one level at a time. The recommendation percentage of the BOD increased from 68 per cent with restricted escalation to 79 per cent without restriction, and the experimentation percentage improved from 25 per cent with restriction to 59 per cent without restriction (Table III), i.e. 34 per cent more subjects would be treated at the BOD if the restriction on escalation were removed. In addition, the required average sample size decreased from 25 to 19. Figure 2 displays the trace plot of dose allocation with and without restriction for 10 simulated trials. When the restriction is in place, in general the dose escalates steadily one level at one time, and stays steady at dose 6. When the restriction is released, subjects are treated at the highest dose level (level 6) immediately after the first cohort. This illustrates the common dilemma of the CRM: increased efficiency at the cost of assigning subsequent subjects to high dose levels immediately (or after the first cohort) when restrictions are not in place. A modification of the design can be adopted in practice taking both efficiency and safety into account, e.g. allowing the design to skip a single dose level, but not two or more, during escalation.

3.4. *Comparison with the randomization design* (*RD*)

An alternative to a model-based design is the RD, which is adopted in some phase I trials due to its simplicity. The RD is carried out as follows: (1) assign c subjects to each dose level; (2) define the empirical estimators for trinomial outcome probabilities as $\hat{\psi}_j(x_i) = y_{ji}/c$, for $j = 0, 1, 2$, and $i = 1, \ldots, 6$; (3) evaluate $\delta_1(x; \hat{\theta})$ for each dose; (4) if $\delta_1(x; \hat{\theta}) = 0$ for all x (no dose level satisfies the toxicity criteria), then no dose is recommended as the BOD; and (5) otherwise, find the subset $C(x)$, and the BOD is the dose x^* among $C(x)$ satisfying $\delta_2(x^*; \hat{\theta}) = \max_{x \in C(x)}\{\delta_2(x; \hat{\theta})\}$. Due to the frequency of its use and its simplicity, we compared the operating characteristics of the RD to those of the TriCRM design. The maximum average sample size is about 25 for the TriCRM design based on the simulation results in Tables I and II. We therefore used 30 subjects (five at each dose level) for the RD design. The TriCRM design is superior to the RD in all the nine scenarios for almost every

Table III. Restriction on dose escalation *versus* no restriction in the TriCRM design ($\lambda = 1$) for scenario 1.

	Dose level							Sample size		Observed (%)	
Restriction	1	2	3	4	5	6	No recom. (%)	Mean	IQR	Success	Toxicity
	(0.02,0.04)* $\delta_2 = -0.02$	(0.06,0.05) 0.01	(0.15,0.07) 0.08	(0.30,0.08) 0.22	(0.47,0.10) 0.37	(0.61,0.11) **0.50**					
Yes	5.1(15.1)†	1.7(14.4)	7.8(15.7)	7.9(15.0)	9.5(14.8)	**67.5**(24.9)	0.5	25.0	{24,27}	30	8
No	2.1(18.4)	0.5(1.9)	2.6(2.6)	4.7(5.9)	10.4(12.0)	**79.3**(59.2)	0.4	19.3	{18,21}	44	9

*The success probability at this dose level, the toxicity probability at this dose level.

†The percentage of times this dose level is identified as the BOD (the percentage of subjects treated at this dose level during the trial).

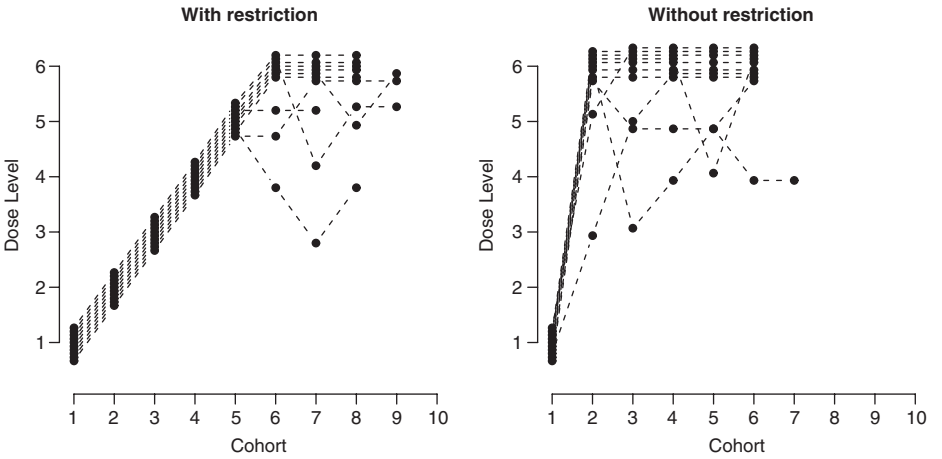

Figure 2. Trace plot of dose allocation with and without restriction on dose escalation for the 10 simulated trials.

aspect compared: (1) the TriCRM design has a higher correct recommendation percentage for the BOD and a higher percentage of subjects treated at the BOD with less subjects; (2) overall, the TriCRM design has higher observed success percentage, with lower observed toxicity percentage. Only the results for the model free scenarios are listed in Table IV due to the space limitations.

4. DISCUSSION

The TriCRM design we propose is conceptually similar to the TR design [1]. However, we employed a more flexible CR model, different decision criteria and simplified stopping rules. Nine dose–outcome scenarios are considered in our simulations with three scenarios each based on the CR model, the PO model and no model. Each set of three scenarios covers one monotone increasing, one unimodal, and one monotone decreasing dose–success curves.

The major objective of a phase I trial of a biologic or targeted agent is to identify the BOD. In general, our simulations show that the CR model has an advantage compared to the PO model in recommending the correct BOD. The CR model outperforms the PO model in scenarios based on the CR model for $\lambda = 1$ and model free scenarios for both $\lambda = 0$ and 1. For the PO model-based scenarios, the CR model is not inferior to the PO model. The performance of both the CR and the PO models are similar in terms of allocating subjects, the sample sizes (mean, IQR), and the observed success and toxicity percentages.

Recently, Thall and Cook (TC) [18] extended the design of TR to use the CR model. However, the TC method is focused on phase I/II trials with proposed sample sizes of 72 patients. The TC approach also proposes a utility-based decision criteria, with considerable effort expended to the elicitation of priors. Our approach offers a simple alternative to the TC approach. We believe that the decision criteria for our design are very intuitive, in that the BOD is defined as the dose maximizing the difference between success probability and the

Table IV. Comparison of the TriCRM design ($\lambda = 1$) with the RD for the model free scenarios 7–9.

Design	Scenario	Dose level 1	2	3	4	5	6	No recom. (%)	Mean sample size	Observed (%) Success	Toxicity
	7	(0.05,0.01)*	(0.10,0.02)	(0.20,0.03)	(0.35,0.06)	(0.55,0.08)	(0.80,0.10)				
		$\delta_2 = 0.04$	0.08	0.17	0.29	0.47	**0.70**				
RD		0.8(16.7)†	1.7(16.7)	5.2(16.7)	12.9(16.7)	27.7(16.7)	**51.7**(16.7)	0.0	30	34	5
TriCRM		0.4(12.4)	0.6(12.9)	5.2(14.5)	6.9(15.4)	13.3(16.8)	**73.6**(28.0)	0.0	25	42	6
	8	(0.50,0.02)	(0.60,0.04)	(0.70,0.06)	(0.55,0.22)	(0.45,0.36)	(0.35,0.46)				
		$\delta_2 = 0.48$	0.56	**0.64**	0.33	0.09	−0.11				
RD		20.9(16.7)	28.4(16.7)	**36.6**(16.7)	9.3(16.7)	3.6(16.7)	1.2(16.7)	0.0	30	52	19
TriCRM		3.2(18.1)	29.6(26.7)	**56.2**(31.3)	10.1(16.8)	0.6(5.6)	0.2(1.5)	0.1	22	59	10
	9	(0.85,0.05)	(0.65,0.25)	(0.50,0.45)	(0.35,0.65)	(0.20,0.80)	(0.10,0.90)				
		$\delta_2 = \mathbf{0.80}$	0.40	0.05	−0.30	−0.60	−0.80				
RD		**81.8**(16.7)	13.3(16.7)	3.6(16.7)	0.6(16.7)	0.1(16.7)	0.0(16.7)	0.6	30	44	52
TriCRM		**92.8**(73.4)	5.5(19.1)	0.0(6.3)	0.0(1.0)	0.0(0.1)	0.0(0.0)	1.7	18	78	12

*The success probability at this dose level, the toxicity probability at this dose level.

†The percentage of times this dose level is identified as the BOD (the percentage of subjects treated at this dose level during the trial).

product of λ and toxicity probability, given the pre-specified toxicity requirement is satisfied. In addition, our focus is on smaller, initial dose-finding studies limited to a maximum of 30 patients, with the intention that such studies will be followed by a 'standard' phase II trial. For scenarios 5 and 8 in the present paper, the recommendation percentages of the TriCRM for the BOD are 69 and 56 per cent (Table I) with only about 21 subjects for an agent with six dose levels. TC demonstrated in two scenarios where the BOD is inside the dose domain (scenarios 2 and 3 of their Table I) that the recommendation percentages for their BOD are 62 and 50 per cent with 72 subjects for an agent with five dose levels.

O'Quigley *et al.* [19] proposed a two-stage dose-finding design (OHF) with two binary outcomes, toxicity and efficacy, assuming separate CRM models for the toxicity probability $R(x;\theta)$ and the probability of efficacy given no toxicity $Q(x;\theta)$. They define overall success as efficacy without toxicity, so that $P(x;\theta)=Q(x;\theta)[1-R(x;\theta)]$ is the probability of success at dose x. An acceptable level of toxicity is determined in the first stage, starting with a low toxicity target that later may be increased, and a sequential probability ratio test is used in the second stage to find the dose maximizing $P(x;\theta)$. This is similar to the TriCRM design maximizing $\delta_2(x;\theta)$ with $\lambda=0$. While it would be desirable to compare the TriCRM with the OHF design, discrepancies between the two designs in other aspects make this impractical; the OHF design is a two-stage design which uses variable target toxicity π_0, variable dose levels, and has a target success rate while the TriCRM design evaluates toxicity and efficacy simultaneously, has a fixed toxicity threshold and fixed number of dose levels, and searches for a dose with the highest possible success rate within the maximum tolerable toxicity. Moreover, the assumption that $\psi_0(x;\theta)$ is a monotone non-increasing function of dose is inherent to the CR model for the TriCRM, but cannot be guaranteed by the OHF design.

We use a Bayesian approach to evaluate the parameters in the CR model. One alternative to Bayesian estimation would be maximum-likelihood estimation (MLE). Although the MLE has wide use in many contexts, it is unsuitable in this case due to the small sample size on which we must base our estimation, particularly early in the dose finding. O'Quigley and Shen [20] proposed a maximum-likelihood alternative for the univariate CRM paradigm, but an extension of their method from the one-parameter model to the four-parameter CR model is not straightforward.

The choice of the prior distribution is always a concern for a Bayesian approach. We use a multivariate uniform distribution for the required prior parameters for the sake of simplicity and to facilitate rapid numerical computation of posteriors. The lower and upper bounds for the uniform distribution we chose work well for the logarithm of the dose ranging from 0.60 to 1.35. We can transform any dose units into this range and apply the same uniform distributions used in our simulations. Widening the prior range for the parameters has little impact on our simulation results due to the already vague prior information, but does increase simulation time.

The choice of the weight λ reflects the extent to which the toxicity is taken into account in the efficacy decision function $\delta_2(x;\theta)=\psi_1(x;\theta)-\lambda\cdot\psi_2(x;\theta)$. No general recommendation can be made for the choice of λ; this decision must be made at the initiation of trial in conjunction with the clinical investigators. The stopping rules we adopt allow for the early termination of the trial in the case of excessive toxicity. We investigated a scenario where the toxicity probability is greater than 0.4 for all dose levels. In this case, the CR model recommended no dose as BOD in over 90 per cent of simulations, with an average sample size of only eight subjects.

The strength of the CR model lies in its flexibility to fit a variety of dose–outcome curves. The simulations for scenarios 2, 5 and 8 show that the CR model works well for the unimodal dose–success curves. The CR model likely outperforms the PO model as it uses four parameters compared to the three-parameter PO model. TC [18] suggest but discard as intractable an extension of the PO model into a four-parameter model by allowing β to be different in equation (1). We do not pursue this here.

The average sample size for the TriCRM design is about 25 for scenarios 1, 4 and 7 where most subjects are needed due to the fact that the BOD is the highest dose level. The average sample size can be as few as 16–18 for scenarios 3, 6 and 9 where the BOD is the first dose level.

In practice, the TriCRM design can be carried out more flexibly. The size of cohort c can be 2, 3, or 4 subjects. Furthermore, the number of subjects allocated to a dose level at one time can be varied. In particular, the concept of a cohort could be discarded, and the dose for each patient determined by the data available at the time the dosing decision is made (cohorts of 1). The dose can also be 'continuous' within a given range instead of selected as one of several pre-specified levels. The required number of subjects treated at the BOD m, the minimum/maximum sample size n_1/n_2 of the trial, and the toxicity requirement π_0 can all be tailored according to the particular clinical situation.

ACKNOWLEDGEMENTS

The authors would like to thank Dr Kathryn Chaloner for helpful suggestions and the reviewers for suggesting several improvements to the manuscript. This work was partially supported by the Mayo Clinic Cancer Center (CA-15083) and the North Central Cancer Treatment Group (CA-25224).

REFERENCES

1. Cunningham D, Humblet Y, Siena S *et al*. Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. *New England Journal of the Medicine* 2004; **351**:337–345.
2. Thall PF, Russell KE. 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.
3. McCullagh P. Regression models for ordinal data (with discussion). *Journal of the Royal Statistical Society, Series B* 1980; **42**:109–142.
4. Fan SK, Chaloner K. Optimal designs and limiting optimal designs for a trinomial response. *Journal of Statistical Planning and Inference* 2004; **126**:347–360.
5. McCullagh P, Nelder JA. *Generalized Linear Models* (2nd edn). Chapman & Hall: New York, 1989; 160–164.
6. Agresti A. *Categorical Data Analysis*. Wiley: New York, 2002; 290.
7. O'Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase 1 clinical trails in cancer. *Biometrics* 1990; **46**:33–48.
8. Chevret S. The continual reassessment method in cancer phase I clinical trials: a simulation study. *Statistics in Medicine* 1993; **12**:1093–1108.
9. Korn EL, Midthune D, Chen TT *et al*. A comparison of two phase I trails designs. *Statistics in Medicine* 1994; **13**:1799–1806.
10. Gooley TA, Martin PJ, Fisher LD *et al*. Simulation as a design tool for phase I/II clinical trials: an example from bone marrow transplantation. *Controlled Clinical Trials* 1994; **15**:450–462.
11. Goodman SN, Zahurak ML, Piantadosi S. Some practical improvements in the continual reassessment method for phase I studies. *Statistics in Medicine* 1995; **14**:1149–1161.
12. Møller 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**:911–922.
13. Heyd JM, Carlin BP. Adaptive design improvements in the continual reassessment method for phase I studies. *Statistics in Medicine* 1999; **18**:1307–1321.

14. Cheung YK, Chappell R. Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics* 2000; **56**:1177–1182.
15. 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.
16. Berntsen J, Espelid TO, Genz A. An adaptive multidimensional integration routine for a vector of integrals. *ACM Transactions on Mathematical Software* 1991; **17**:452–456.
17. Genz A, Meyer M, Lumley T *et al. The adapt Package* (ver. 1.0-3), 2002. Available at: http://www.r-project.org
18. Thall PF, Cook JD. Dose-finding based on efficacy–toxicity trade-offs. *Biometrics* 2004; **60**:684–693.
19. O'Quigley J, Hughes MD, Fenton T. Dose-finding designs for HIV studies. *Biometrics* 2001; **57**:1018–1029.
20. O'Quigley J, Shen LZ. Continual reassessment method: a likelihood approach. *Biometrics* 1996; **52**:673–684.

