STATISTICS IN MEDICINE
Statist. Med. **18**, 1155–1169 (1999)

ACCRUAL STRATEGIES FOR PHASE I TRIALS WITH DELAYED PATIENT OUTCOME

PETER F. THALL[1]*, J. JACK LEE[1], CHI-HONG TSENG[1] AND ELIHU H. ESTEY[2]

[1]*Department of Biomathematics, Box 237, The University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, Texas 77030, U.S.A.*
[2]*Department of Leukemia, Box 061, The University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030, U.S.A.*

SUMMARY

Phase I dose-finding trials typically are conducted using adaptive rules that select dose levels for successive patient cohorts based on the outcomes of patients treated previously in the trial. When patient outcome cannot be observed immediately after treatment, the problem arises of how to deal with new patients while waiting to observe the current patient cohort's outcomes. We consider two alternative approaches to this problem in the context of a phase I trial conducted using the continual reassessment method. With the first approach, a patient requiring treatment before the next cohort opens is treated off protocol with standard therapy, and otherwise waits until the next cohort opens. The second approach treats each patient immediately upon arrival at the dose recommended based on currently available data. We compare these two approaches by simulation under varying dose–toxicity curves, accrual rates, cohort sizes and early stopping rules. We evaluate patient waiting time, trial duration, number of patients treated off protocol and the probabilities of toxicity and of selecting the correct dose. We also study three strategies for assigning patients to trials when two or more phase I trials may be ongoing simultaneously. Based on our results, we provide practical guidelines for deciding among these approaches and strategies in a given clinical setting.

1. INTRODUCTION

The scientific goal of a phase I clinical trial of an experimental chemotherapeutic regimen in oncology is to determine a maximum tolerated dose (MTD). Statistical designs for such trials typically are based on an adverse patient outcome, called 'toxicity', that occurs relatively soon after initiation of treatment. For ethical reasons, such trials are conducted adaptively, since toxicity may involve permanent organ damage or life-threatening events. Patients are treated in cohorts at successive dose levels, with each cohort given a dose determined by the outcomes of patients who have been treated previously in the trial. When toxicity cannot be observed immediately after treatment, the problem of deciding how to deal with new patients arises. This is because strict adherence to a given sequential dose-finding rule requires observation of the outcomes of all patients in the current cohort in order to determine the recommended dose for patients in the next cohort. This seems to imply that new patients must be made to wait until all outcomes of patients in the current cohort have been observed. In settings where it takes several weeks or months to observe whether a patient has experienced toxicity, a sequential dose selection rule may create an ethical dilemma. This is due to the conflict between the desire to treat

* Correspondence to: Peter F. Thall, Department of Biomathematics, Box 237, The University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, Texas 77030, U.S.A. E-mail: rex@mdanderson.org

CCC 0277–6715/99/101155–15$17.50
Received December 1997
Accepted August 1998

patients at the dose recommended by the most recent data and the desire to treat patients in a timely fashion. The manner in which this problem is addressed in a given trial has scientific, logistical and ethical consequences.

The problem described above is complicated further if several phase I trials, each appropriate for the same patient group, are ongoing simultaneously. In a given clinic or co-operative group, phase I trials of different agents are usually initiated in sequence over time based on a variety of criteria, including scientific motivation, drug availability, patient accrual and monetary cost. The different treatments are rarely the arms of a randomized phase I trial due to the fact that new drugs typically become available at different times. Moreover, the scientific goal in phase I is to determine an acceptable dose rather than to compare toxicity rates of different agents. Thus, the problem of deciding among two or more available treatments for each new patient may arise.

In this paper, we consider these problems in the context of one or more phase I trials conducted using the continual reassessment method, or CRM[1]. Aside from some discussion by O'Quigley *et al.*,[1] the problem of dealing with delayed patient outcome when implementing the CRM has not been dealt with in the literature. We first define and compare two alternative approaches for dealing with the problem of delayed observation of patient outcome in the context of a single phase I trial conducted using the CRM. Under the first approach, if a new patient requires treatment before the next cohort in the trial opens, then the patient is treated off protocol using standard therapy. Otherwise, the patient waits and is treated on protocol when the next cohort opens. The second approach does not delay accrual, with each patient treated immediately upon arrival at the dose that is recommended based on the most recently available data. We compare these two approaches under several different dose–toxicity curves, patient accrual rates, cohort sizes and early stopping rules. We evaluate the effect of these parameters on the scientific criteria consisting of probabilities of toxicity and of selecting the correct dose and on the logistical and ethical criteria consisting of patient waiting time, trial duration and number of patients treated off protocol.

We carry out all evaluations by computer simulation. It is often very useful to simulate a clinical trial over a range of clinical scenarios before the trial is conducted. Aside from simple settings where a design's operating characteristics are well known or can be computed analytically, computer simulation is the only tool we know of for evaluating how a particular design may behave during actual trial conduct. Based on simulation results, design parameters may be adjusted to obtain a design with good operating characteristics. The use of simulation in designing early-phase clinical trials has been proposed recently by a number of authors.[2–9] While properties of various versions of the CRM have been studied extensively by simulation,[3,4,10–13] none of these studies have addressed the issues of delayed patient outcome and multiple trials considered here. Our goals are to propose methods for dealing with these commonly occurring problems when implementing the CRM, to explore their effects on the operating characteristics noted above, and to provide practical guidelines. We provide two freely available computer programs for trial conduct and for conducting simulations as a basis for choosing between the proposed strategies in specific settings.

The simulation study reported here was originally motivated by a smaller simulation conducted in the process of designing a particular phase I trial using the CRM. When questions pertaining to the effects of delaying accrual arose, we extended the simulation to address the problem more generally. The study of strategies for dealing with multiple trials was conducted subsequently in response to a referee's request for a more complete picture of the clinical environment and a more general basis for making practical recommendations.

The remainder of the paper is organized as follows. In Section 2 we provide brief descriptions of phase I trials, the CRM, and two methods for dealing with delayed patient outcome when using the CRM. We describe a particular phase I trial that serves as a basis for the simulation in Section 3. In Sections 4 and 5 we summarize the simulation results. We provide practical guidelines for trial conduct and information for obtaining computer programs in Section 6.

2. BACKGROUND AND MOTIVATION

2.1. Phase I Trials

Patients in phase I oncology trials typically have very poor prognosis,[14,15] usually having failed to achieve substantive responses to multiple prior treatments. The scientific goal in phase I is determination of an appropriate dose for subsequent, better prognosis patients, rather than evaluation of efficacy *per se*. This approach is ethical because, although the patient's likelihood of achieving an efficacy outcome in phase I is very low, these patients still are willing to undergo possibly toxic treatment in the hope of achieving a remission. While a treatment with no toxicity whatsoever is of course most desirable, determination of an MTD is motivated by the implicit assumption that higher dose levels are associated with higher rates of a desired efficacy outcome, such as $\geqslant 50$ per cent shrinkage of a solid tumour or complete remission of leukaemia. Thus, the idea underlying phase I dose-finding is to determine the highest dose level which does not have unacceptably high toxicity.

Virtually all designs for dose-finding in phase I characterize patient outcome by a binary[16,17] or possibly ordinal[18] toxicity variable. In practice, some amount of time is required to administer the treatment and the particular patient events comprising toxicity must be observed over a specified time period after initiation of treatment. We shall refer to the period of time between the start of treatment and scoring of the patient's outcome in a given trial as the 'evaluation window', and we denote its length by w. Because the dose level of each successive cohort is chosen based on the outcomes of previous patients, unless w is very small a practical problem arises when new patients arrive in the clinic to be treated but some outcomes of the most recent cohort have not yet been scored. The question of what to do with these new patients involves logistical, scientific and ethical issues.

2.2. The Continual Reassessment Method

The continual reassessment method, or CRM,[1,3,4,10–13,18] is a Bayesian model-based algorithm that aims to select the dose level having toxicity probability closest to a specified fixed target probability ψ^*. The basic idea underlying the CRM is quite simple.[1] Given k dose levels, a one-parameter model $\psi_j(\alpha)$ for the probability of toxicity at level j is assumed, and a non-informative or weakly-informative prior is typically used for α. Denote the posterior probability of toxicity at the jth dose level given the current data in the trial by $\psi_j(\alpha|\text{data})$. As the data from each successive cohort of patients are observed and the posteriors of α and hence of $\{\psi_j(\alpha|\text{data}), j = 1, \ldots, k\}$ are updated, the CRM chooses the dose level having $E[\psi_j(\alpha|\text{data})]$ closest to the target probability ψ^*. We will refer to this as the 'best' or 'current recommended' dose level, denoted by j^*. Naturally, j^* changes during the course of the trial. The best level j^* at the end of the trial is declared the MTD. Various stopping rules are used in practice, including stopping when k patients have already been treated at j^*, with typical values of k ranging from 6 to 10, requiring a minimum sample size in addition to the 'stop at k' rule, treating a fixed maximum

number of patients, running the trial for a maximum duration, or using some combination of these.

2.3. Methods for Dealing With Delayed Outcome

Strict adherence to any sequential dose-finding design requires that, until the responses of all patients in the current cohort have been observed, and, based on this, the next recommended dose level has been determined, no additional patients may be treated in the trial. This is the case regardless of the cohort size, since even with $c = 1$ under the CRM new patients may arrive during the current patient's evaluation window. We first consider the problem of dealing with patients who arrive in the clinic and are ready to be treated before all of the outcomes of patients in the current cohort have been evaluated. We address this in the context of a single trial conducted using the CRM, and we propose the following two approaches. Subsequently, we will deal with the problem of multiple trials.

The first method, design 1, requires that the maximum time each patient can wait for treatment, W, be determined at the time the patient arrives in the clinic. In practice, some patients require immediate treatment ($W = 0$) while, medically, it may be reasonable or even desirable for others to wait ($W > 0$). This suggests that the distribution of W should be a mixture with point mass $p = \Pr[W = 0]$ at 0 and the remaining probability $1 - p = \mathrm{pr}[W > 0]$ following some reasonable continuous distribution. Denoting the time from the patient's arrival until the next cohort in the trial opens by T, if $W < T$ then the patient is treated off protocol using standard therapy. If $W \geqslant T$ then the patient waits for the period T until the next dose level is determined and is then treated on protocol at that dose. An important proviso in determining T is that, in some cases, the recommended dose will not be affected by the outcomes of patients in the current cohort whose responses are as yet unknown. For example, say $c = 3$, the outcome of the first two patients in the current cohort have been recorded, and the next recommended dose level will be j^* regardless of whether the remaining patient in the current cohort experiences toxicity. In this case, there is no reason to wait for the third patient's outcome to determine the recommended dose for the next cohort, since it is already known. This may be implemented by computing j^* for each possible configuration of outcomes of patients who are currently in their evaluation windows. If the values of j^* so determined are all identical to the value based on currently available data, then $T = 0$ and treatment of the next cohort at level j^* may begin as soon as new patients arrive. We will refer to this as the 'look-ahead' option. To avoid a possible conflict between the safety modification that dose levels may not be skipped when escalating and the look-ahead option, we impose the former restriction at a higher level of priority. Thus, the look-ahead option is used only if no untried levels are skipped when escalating.

Design 2 is motivated by the consideration that, under the Bayesian paradigm, all information needed for making decisions adaptively at any point during a trial is contained in the most recent posterior probability distribution. This suggests that, if a new patient arrives in the clinic and is ready to be treated before all of the outcomes of patients in the current cohort have been evaluated, then the appropriate action is simply to treat the new patient at the current j^*, as determined by the most recently available data. Thus, design 2 does not delay accrual and by definition has a cohort size of 1. This approach is discussed in Section 3 of O'Quigley *et al.*,[1] who also suggest an extended method obtained by defining a range of acceptable dose levels near j^* for treatment of currently accrued patients. We do not explore this extension here, however.

Table I. Dose levels for the GEMCY trial

Dose level (j)	Gemcytabine	Cyclophosphamide	p_j
1	16	300	0·15
2	16	400	0·20
3	16	500	0·40
4	20	500	0·50
5	25	500	0·60
6	25	625	0·70
7	25	750	0·80

3. THE GEMCY TRIAL

Our simulation study is based on a trial of gemcytabine + cyclophosphamide (GEMCY) to treat patients with refractory AML conducted at M.D. Anderson Cancer Center. The aim of this trial was to determine the dose among the seven levels given in Table I having toxicity probability closest to $\psi^* = 0{\cdot}40$.

While the selection of $\psi^* = 0{\cdot}40$ may seem unusually high, it was chosen for two reasons. First, AML is fatal within a few months if treated ineffectively. Consequently, AML patients are almost always willing to receive a dose relatively likely to produce toxicity in the hope that it will increase the likelihood of response. Second, the principal toxicity of many AML therapies is myelosuppression. Based on our experience, the duration of myelosupression is influenced by the number of prior therapies received, and, possibly, by the anti-leukaemia response to treatment. Since patients entered into phase I trials have invariably received one or more other regimens and are likely to have persistent leukaemia after therapy, they are more likely to experience prolonged myelosupression than patients entered into phase II, who are generally less heavily pre-treated and are more likely to respond. Thus, the dose that produces a 0·40 probability of toxicity in phase I is likely to produce a lower toxicity probability in phase II.

The choice of dose level 3 as the initial dose also reflects the desire to avoid ineffective doses. Indeed, one of the advantages of the CRM is that it is less likely to treat patients at ineffective dose levels, compared to conventional dose-finding methods used in phase I. In general, our clinical experience is that several dose levels are usually investigated before an MTD is discovered. This is particularly true in AML, where the starting dose is frequently identical to the MTD determined on the basis of myelosupression in patients with solid tumours. However, the degree of myelosuppression considered intolerable in solid tumour patients is almost always considered tolerable, and therapeutically necessary, in AML patients. In the GEMCY trial, the successive combinations for the dose levels were chosen so that level $j + 1$ represents a 20 to 33 per cent total increase over level j. This range of escalation is typical of most phase I trials.

We applied the CRM with exponential dose–toxicity model $\psi_j(\alpha) = p_j^{\exp(\alpha)}$, where $p_1 < p_2 < \ldots < p_k$ are fixed probabilities of toxicity at the respective dose levels corresponding to $\alpha = 0$, and α is assumed to follow a normal prior with mean 0 and $\sigma = 1{\cdot}34$ as recommended by O'Quigley and Shen.[4] The numerical values of $p_1, \ldots, p_7$ are given in the last column of Table I. In practice, the p_j's are chosen to obtain a design with good operating characteristics, and they do not necessarily reflect prior belief regarding the probabilities of toxicity at the successive dose levels. We began with the first cohort at dose level 3 and imposed the safety constraint that no

untried dose levels could be skipped when escalating. A fixed sample size of 30 patients was used, with no provision for early stopping. Our reason for not employing an early stopping rule will be explained in Section 4.2. Using these design parameters, we simulated the trial with cohort sizes of $c = 1$, 2 or 3.

To account for the 5-week patient evaluation window required to administer and evaluate this treatment regimen, we further assumed that patients would arrive in the clinic over time according to a Poisson process. Based on historical experience with this patient group at M. D. Anderson, we assumed an accrual rate of $a = 1{\cdot}15$ patients per week, equivalently 60 per year. The results of this initial simulation raised questions pertaining to the effects of delaying accrual while waiting for the current cohort's outcomes, and how different values of a might affect the design's properties. This motivated the more general simulation which we now describe.

4. SIMULATION STUDY OF ONE TRIAL

In this section, we summarize a simulation study of an array of generalizations of the GEMCY trial. We evaluated various properties of designs 1 and 2 for the CRM under each of the four dose–toxicity curves S_1, S_2, S_3, S_4 illustrated in Figure 1. The correct dose level under each of the four scenarios is that having the targeted 40 per cent toxicity, levels 1, 3, 5 and 7, respectively. We considered accrual rates $a = 0{\cdot}5$, $1{\cdot}15$, 2 or 3 patients per week and used an evaluation window with $w = 5$ weeks throughout, since varying w is the same as varying a proportionally, and hence provides no additional insight. For design 1, we considered cohort sizes of $c = 1$, 2 or 3, and, based on medical experience, we assumed that the maximum patient waiting time was 0 for 20 per cent of patients and, among the remaining 80 per cent of patients who could wait, that it followed a Weibull distribution determined by the equalities $\Pr[W > 2 \text{ weeks } | W > 0] = 1/2$ and $\Pr[W > 6 \text{ weeks } | W > 0] = 1/80$. All simulations were carried out in S-plus 3·3 on a DEC AlphaServer 2100 5/250 running Digital UNIX V4·0B. Each case was simulated 1000 times, and each tabulated value is the mean over the simulations.

4.1. Effects of the look-ahead option under design 1

We first examined the effects of using the look-ahead option to define $T = 0$ when the outcomes of patients currently in their evaluation windows will not alter j^*, as described in Section 2.4. For convenience, we denote the version of design 1 without the look-ahead option as design 1^0. Figure 2 illustrates the effects on the number of patients turned away, and hence treated off protocol, on the trial duration, and on individual patient waiting time. This last quantity is the actual time that the patient waits, and should not be confused with the maximum waiting time W that is elicited when the patient arrives. The box plots in Figure 2 and elsewhere have a box running from the 25th to 75th percentiles, midline at the median, and whiskers extending to the 5th and 95th percentiles. The reductions gained from using the look-ahead option are, as intuitively expected, most pronounced for a cohort size of 1, with much smaller relative drops for $c = 2$ or 3. Although these results are for $a = 1{\cdot}15$, the results are very similar for $a = 2$ or 3.

4.2. Effects of Early Stopping

We next examined the effects of using an early stopping rule on the probability of correct selection (PCS) and sample size for each of the two CRM designs. Table II summarizes the PCS, specifically the probability of selecting the dose having toxicity 0·40, under each scenario. The

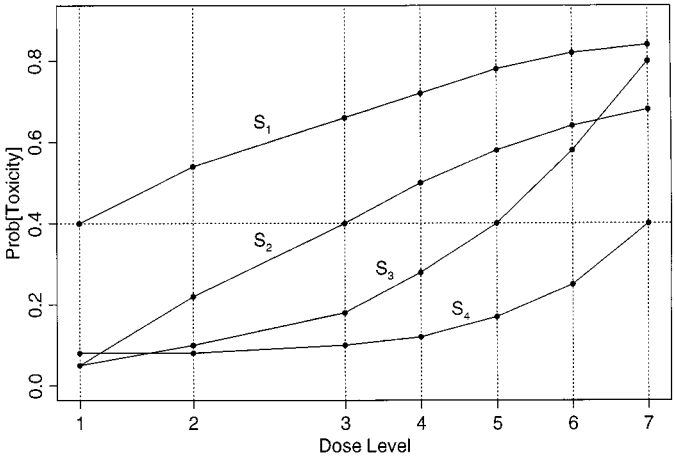

Figure 1. The four dose–toxicity scenarios

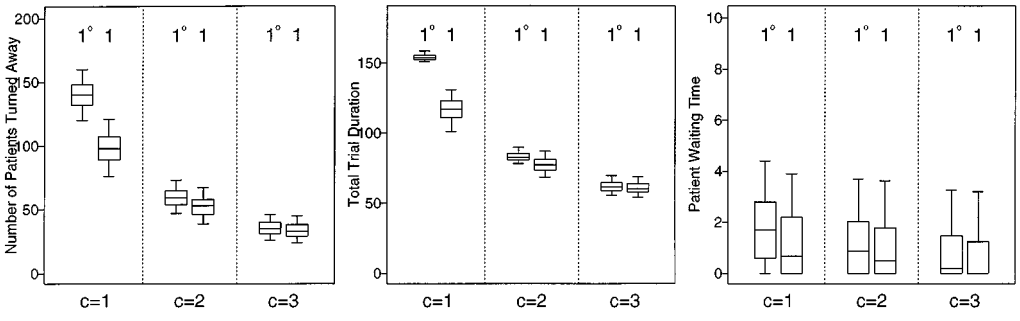

Figure 2. Effects of the 'look-ahead' rule on number of patients turned away, trial duration and patient waiting time, under dose–toxicity curve S_3 with accrual rate 1·15 patients per week. Design 1 without the look-ahead option is denoted by 1^0

Table II. Effects of early stopping on per cent correct selection

Scenario	Stopping rule	CRM design 1			CRM design 2			
		$c = 1$	$c = 2$	$c = 3$	$a = 0{\cdot}5$	$a = 1{\cdot}15$	$a = 2$	$a = 3$
1	30 patients	70	70	68	66	64	59	57
	$\geqslant 6$ at MTD	59	52	49	56	53	51	50
	35 weeks*	50	53	58	62	65	59	57
2	30 patients	58	63	59	59	57	59	61
	$\geqslant 6$ at MTD	47	48	54	53	75	85	88
	35 weeks	34	47	49	42	55	58	60
3	30 patients	55	58	54	56	54	55	51
	$\geqslant 6$ at MTD	37	37	38	40	39	41	40
	35 weeks	30	42	45	44	57	56	53
4	30 patients	69	62	57	53	53	40	33
	$\geqslant 6$ at MTD	53	53	44	50	48	41	37
	35 weeks	44	48	45	47	53	42	35

*$a = 1{\cdot}15$ for design 1 with maximum duration 35 weeks

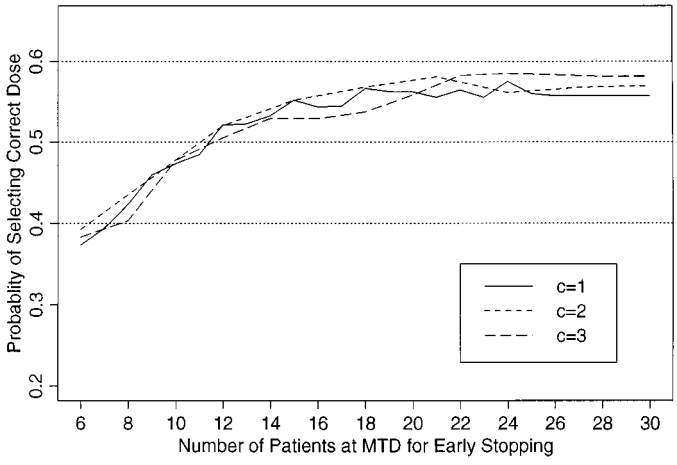

Figure 3. Effects of early stopping on the probability of selecting the correct dose for design 1 and dose–toxicity curve S_3 with cohort sizes $c = 1$, 2 or 3

tabulated values are percentages. Using a maximum sample size of 30 patients throughout, we evaluated each of the two CRM designs applied: (i) with the full sample size of 30 patients, that is, no early stopping rule; (ii) with the rule that the trial should be terminated if 6 or more patients have already been treated at the next recommended dose level; (iii) with the rule that the trial should run no longer than 35 weeks. Accrual rate has no effect on the PCS under design 1 with either fixed sample size or a 'stop at k' rule, while design 2 has a cohort size of 1 by definition. Thus, values for design 1 with $a = 1{\cdot}15$ and $c = 1$, 2 or 3 and for design 2 with $a = 0{\cdot}5$, $1{\cdot}15$, 2 or 3 are tabulated.

Table II shows that the 'stop at 6' rule causes a drop in PCS under all four scenarios for design 1. For design 2, the 'stop at 6' rule causes a drop in PCS under either S_1 or S_3, has little effect under S_4 where the target is achieved at the highest dose level, and causes an increase in PCS in the case of S_2 and $a \geqslant 1{\cdot}15$. This is due to the fact that the starting dose level happens to be the desired target. Essentially, if one is lucky enough to start at the correct dose level, then stopping quickly leads to a high PCS. If the starting dose is not the target, however, then the drop in PCS due to early stopping may be substantial. This phenomenon was also seen by Korn *et al.*[11] when studying the CRM under a different array of dose–toxicity curves. To assess the sensitivity of the 'stop at k' rule's effect on PCS to the value of k, we evaluated the PCS of the CRM design 1 with this rule for values of k ranging from 6 to 30, under scenario 3. The results are summarized in Figure 3. This figure indicates that, regardless of cohort size, even for values of k as large as 16, early stopping may cause a substantial degradation of the PCS. The plateau in PCS at 57 per cent for values of $k \geqslant 20$ is due to the maximum sample size limitation of 30, since PCS $\rightarrow 1$ if the sample size is allowed to increase without limit. For example, PCS = 83 per cent if the sample size is 100 patients.

The use of early stopping rules in phase I may be rationalized on the basis of savings in sample size. Table III, which corresponds to Table II, gives the sample sizes obtained under each CRM design with each of the two early stopping rules. While these sample sizes are much smaller than the maximum of 30 patients, it should be noted that, for example, under design 1 and scenario 3,

Table III. Effects of early stopping on sample size

Scenario	Stopping rule	CRM design 1			CRM design 2			
		$c = 1$	$c = 2$	$c = 3$	$a = 0{\cdot}5$	$a = 1{\cdot}15$	$a = 2$	$a = 3$
1	$\geqslant 6$ at MTD	10·3	10·6	11·1	12·2	14·2	17·1	20·5
	35 weeks	11·1	17·6	21·3	17·4	29·8	30·0	30·0
2	$\geqslant 6$ at MTD	12·0	11·4	10·7	12·9	11·7	11·1	11·7
	35 weeks	7·3	13·7	18·4	17·4	29·8	30·0	30·0
3	$\geqslant 6$ at MTD	13·0	13·8	13·2	15·2	16·7	19·1	20·5
	35 weeks	7·1	13·5	18·1	17·4	29·8	30·0	30·0
4	$\geqslant 6$ at MTD	13·1	15·2	17·0	15·3	18·4	21·7	24·9
	35 weeks	7·7	14·7	19·0	17·4	29·8	30·0	30·0

the expected sample size of roughly 13 to 14 patients results in a substantially reduced PCS (Figure 3). It is doubtful that the saving in sample size warrants the drop in PCS. Based on these results, we implemented each of the two CRM designs with the full sample size of 30 patients in all subsequent simulations.

4.3. Selection Probabilities

Figure 4 provides a graphical illustration of the selection probabilities of CRM designs 1 and 2 for each dose level under each of the four scenarios. The case $a = 2$ is omitted for CRM design 2 to simplify the figure. Designs 1 and 2 have roughly the same PCS under S_2, for which the starting dose is the target. Design 1 with $c = 2$ or 3 has a slightly higher PCS under S_3. Perhaps the most notable pattern is that the PCS under design 2 decreases with accrual rate under either S_1 or S_4, while the probability of selecting an incorrect dose level adjacent to the desired target increases with a. Under S_1, where the target is at the lowest level, design 2 has a higher probability of selecting dose level 2, which has 54 per cent toxicity. These results suggest that, if the target is not close to the starting level and patient accrual is relatively high, the pure Bayesian approach of design 2 may result in a relatively large decrease in PCS, compared to design 1.

There is a smaller but potentially important effect of c on the performance of design 1. It appears that, under S_4 where the target is at the highest level, the probability of correctly selecting this level decreases with increasing c. Thus, $c = 1$ appears to be best at finding a target farther away from the starting point.

4.4. Toxicity

Figure 5 summarizes the toxicity probability distributions under each of the four dose–toxicity curves for design 1 with $c = 1$, 2 or 3 and for design 2 with the four accrual rates $a = 0{\cdot}50$, 1·15, 2 and 3. Figure 5 indicates that design 2 risks much higher toxicity rates if the starting level has a high rate of toxicity and ψ^* occurs at the lowest dose level, but achieves lower toxicity rates if ψ^* occurs at the highest dose level. The toxicity rates of designs 1 and 2 are about the same if the target is at the starting level. For design 1, the toxicity probabilities decrease with cohort size when the target is high among the dose levels. Thus, under S_4 a cohort size of $c = 3$ has the lowest toxicity under design 1.

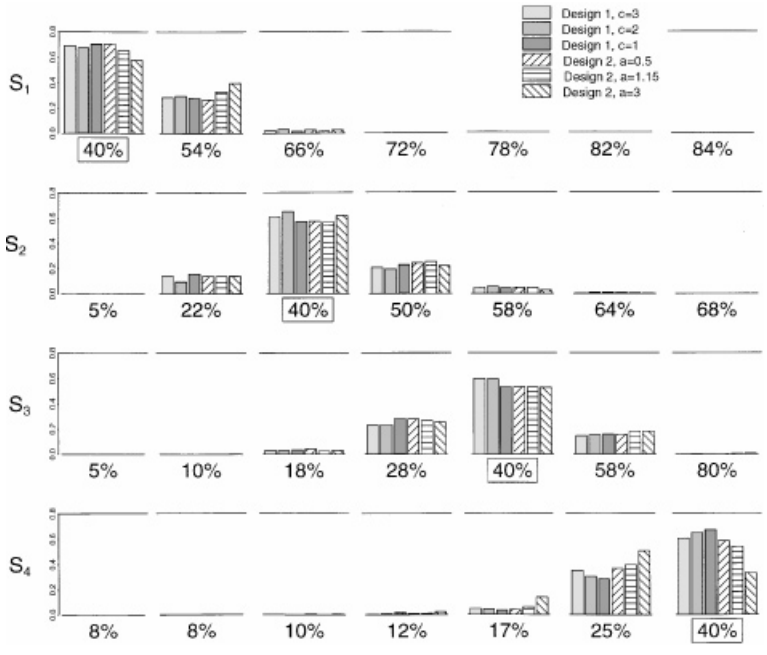

Figure 4. Dose selection probabilities for design 1 with cohort sizes $c = 1$, 2 or 3 and design 2 with accrual rates $a = 0{\cdot}5$, 1·15 or 3, under each of the four dose–toxicity scenarios S_1, S_2, S_3, S_4

4.5. Trial Duration

Figure 6 summarizes the trial duration distributions of designs 1 and 2 for varying c and a. As expected, treating patients immediately leads to a shorter trial. For design 1, the average and the variability in trial duration both decrease with increasing accrual rate and cohort size. For design 1, regardless of accrual rate, the largest drop in trial duration is achieved as c is increased from 1 to 2, with a smaller relative drop going from $c = 2$ to $c = 3$.

4.6. Number Turned Away and Patient Waiting Time

Figure 7 summarizes the distributions of the number of patients turned away and individual patient waiting time under design 1 for varying a and c. The number of patients turned away increases with accrual rate and decreases with cohort size. Moreover, for the range of a and c studied the differences are quite large. There is also an increase in the variability of the number turned away with increasing accrual rate. The effect of cohort size here is analogous to that on trial duration in that the greatest drop in the number of patients turned away is achieved by increasing c from 1 to 2, with a much smaller decrease achieved by going from $c = 2$ to $c = 3$. Patient waiting time increases with a and decreases with c, while the variability of patient waiting time decreases with c.

5. MULTIPLE TRIALS

As noted earlier, phase I trials are initiated sequentially over time within a given clinic or co-operative group as a consequence of several factors. Typically, it is not feasible, or necessarily

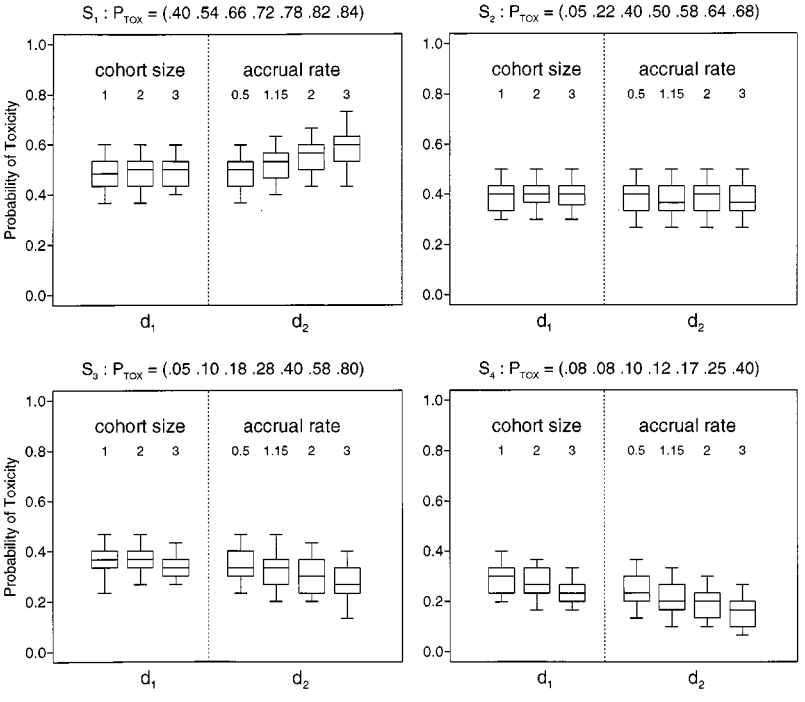

Figure 5. Toxicity probabilities for design 1 with cohort sizes $c = 1$, 2 or 3 and design 2 with accrual rates $a = 0{\cdot}5$, $1{\cdot}15$ or 3, under each of the four dose–toxicity scenarios

desirable, to study multiple phase I agents as arms of a randomized trial. Thus, if two or more phase I trials are ongoing at the same time, the question of how patients should be assigned to trials arises.

In this section, we study three possible ways to deal with this problem. The first method aims at achieving balance by enrolling new patients in the trial that has the fewest patients. The second method takes the opposite approach of enrolling patients in the trial that is nearest to completion. With this method, patients are accrued to trials other than the one closest to completion only if it is temporarily closed while waiting for a given cohort's toxicity outcomes to be observed. The third method randomizes patients fairly among the open trials. To evaluate these three approaches, we applied each to a sequence of 13 hypothetical trials that began at the actual starting dates of the 13 phase I trials in AML conducted at M.D. Anderson from 1993 to the first quarter of 1998. An accrual rate of 1.15 patients per week was assumed, with patients arriving according to a Poisson process over time. To focus on the effects of the three methods, all of the hypothetical trials were of the same form as the GEMCY trial, conducted using design 1 with a sample size of 30 patients.

The three approaches produce strikingly different results. Figure 8 plots the trial durations as functions of calendar time under each of the three methods. Figure 8(*b*) shows that enrolling new patients in the trial with the fewest patients produces clusters of trials that run nearly simultaneously. This effect is so extreme that the trial begun in the last quarter of 1994 takes 8 years to complete, while the last trial begun in the first quarter of 1998 runs nearly simultaneously with the

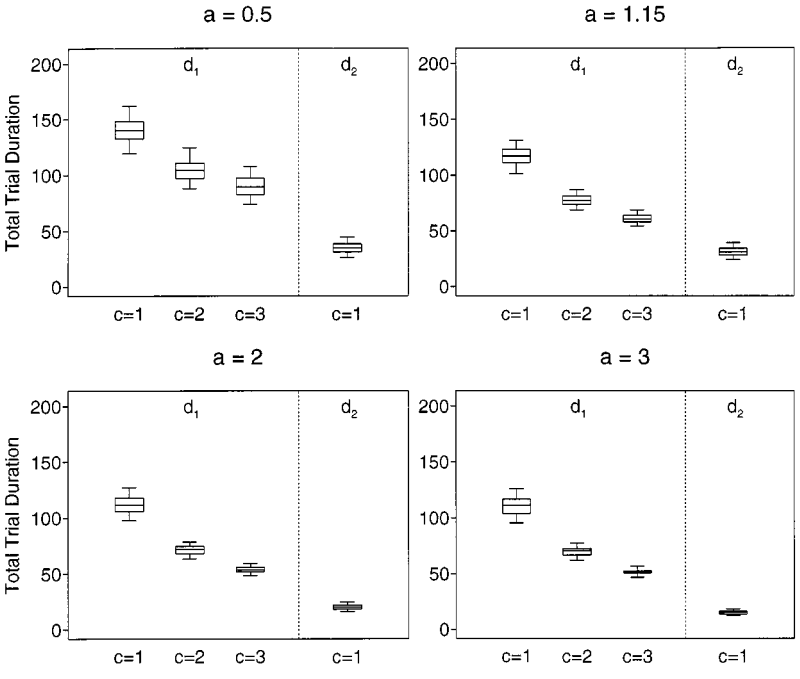

Figure 6. Trial duration for designs 1 and 2 with accrual rates $a = 0{\cdot}5$, 1·15 or 3, under dose–toxicity curve S_3

preceding 10 trials. Figure 8(*c*) shows that enrolling new patients in the trial nearest completion produces trials that have little or no overlap in time, while Figure 8(*a*) shows that randomizing patients greatly increases the variability in the durations of many of the trials.

Naturally, fewer patients are treated off any protocol and patient waiting time decreases if two or more trials are ongoing simultaneously. Thus, the disadvantages of design 1 compared to design 2 are reduced whenever multiple trials overlap in time.

6. PRACTICAL GUIDELINES

Computer programs are available via anonymous ftp from odin.mdacc.tmc.edu as the compressed files crm-0·5.tar.gz in the subdirectory/pub/source and acdel.sim.shar.gz in the subdirectory /pub/S. A Windows executable version of the first file is available as crm-0.5.exe in the subdirectory /pub/msdos. The 'crm' file contains a menu-driven program for conducting trials using the CRM that also allows the user to simulate trials. The 'acdel' file contains S-plus programs to simulate CRM designs 1 and 2 while varying a and w. It provides empirical estimates of the probabilities of selection and toxicity by dose, trial duration, and, for CRM design 1, patient waiting time and the number of patients treated off protocol.

While our simulation results pertain to the array of design parameters and clinical scenarios considered here, their implications are likely to obtain more generally. In any case, we strongly suggest that in practice one should simulate the particular trial being planned. Given this proviso, our recommendations for applying the CRM are as follows:

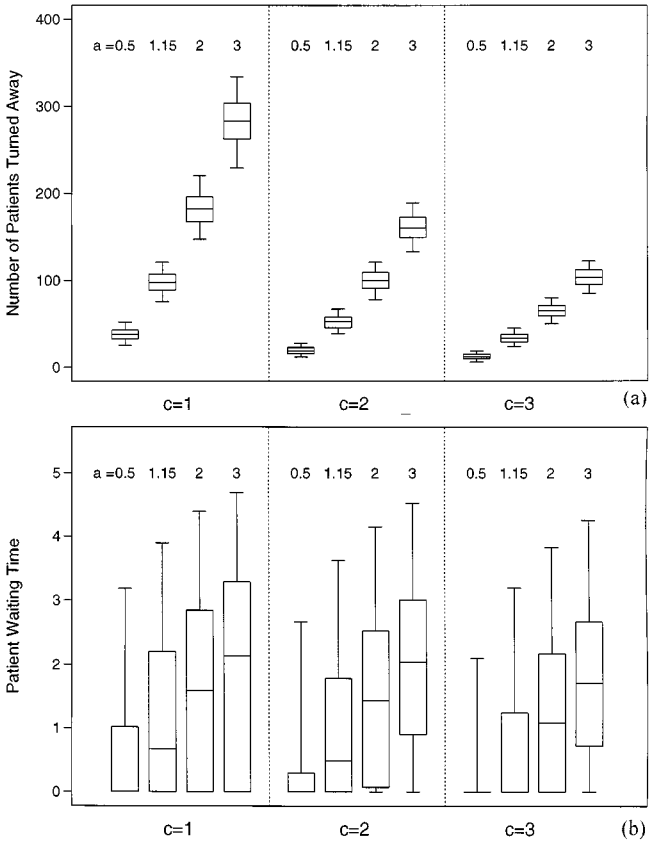

Figure 7. (*a*) Number of patients treated off protocol, and (b) patient waiting time, for design 1 as functions of cohort size and accrual rate under dose–toxicity curve S_3

1. Select the starting dose carefully, and do not skip dose levels when escalating.
2. Never use an early stopping rule. Instead, study the effects of different sample sizes by simulation and choose a fixed sample size on that basis.
3. Always use the 'look ahead' option with CRM design 1.
4. Decide between CRM designs 1 and 2 based on the following criteria, after simulating the trial using both designs with the values of a and w for your trial:
   (a) PCS: design 1 generally has higher PCS. Design 2 suffers a drop in PCS if ψ^* is not achieved near the starting dose, particularly if accrual is high.
   (b) Toxicity: design 2 risks high toxicity if ψ^* is achieved at a low dose level and accrual is high. Design 1 achieves lower toxicity for $c = 3$ compared to $c = 1$ or 2 if ψ^* is at a high dose level.
   (c) Logistics: design 2 has the lowest trial duration, no patient waiting time and treats all patients on protocol. Design 1 achieves its shortest trial duration and treats fewest patients off protocol for $c = 3$ compared to $c = 1$ or $c = 2$, and treats more patients off protocol if the accrual rate is high.

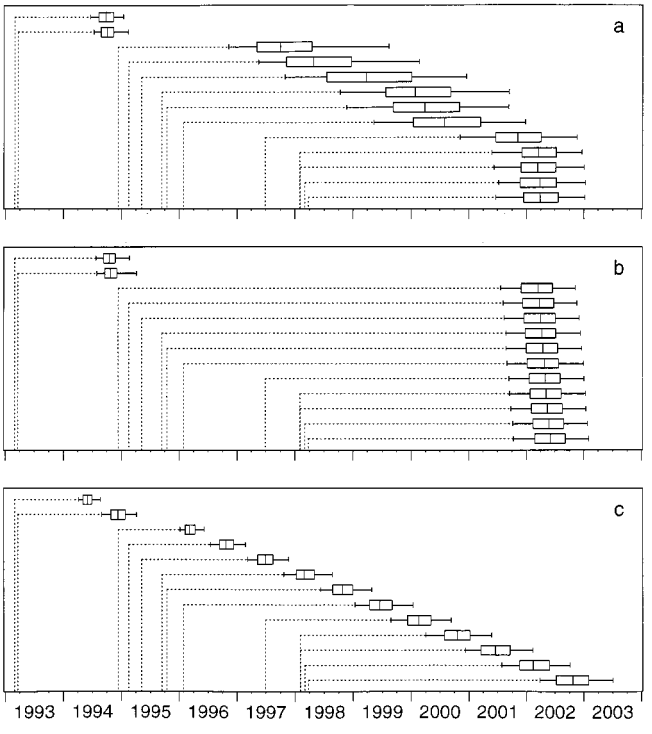

Figure 8. Multiple trials: trial durations under design 1 with cohort size $c = 3$, accrual rate 1·15 patient per week, under dose–toxicity curve S_3, if (*a*) new patients are randomized among open trials, (*b*) new patients are enrolled in the trial with fewest patients, or (*c*) new patients are enrolled in the trial nearest completion. The starting date for each trial is indicated by a dotted line connecting the box plot of its duration distribution to the horizontal axis

5. If two or more trials are open simultaneously, treat new patients on the trial that is closest to completion. Under design 1, the number of patients treated off any protocol and patient waiting both decrease with the number of simultaneous trials.

ACKNOWLEDGEMENTS

The authors thank two referees for their detailed and constructive comments on two earlier drafts of this paper.

REFERENCES

1. O'Quigley, J., Pepe, M. and Fisher, L. 'Continual reassessment method: a practical design for phase I clinical trials in cancer', *Biometrics*, **46**, 33–48 (1990).
2. Gooley, T. A., Martin, P. J., Fisher, L. D. and Pettinger, M. 'Simulation as a design tool for phase I/II clinical trials: An example from bone marrow transplantation', *Controlled Clinical Trials*, **15**, 450–462 (1994).
3. O'Quigley, J. and Chevret, S. 'Methods for dose-finding in cancer clinical trials: A review and results of a Monte-Carlo study', *Statistics in Medicine*, **10**, 1647–1664 (1991).
4. O'Quigley, J. and Shen, L. Z. 'Continual reassessment method: a likelihood approach', *Biometrics*, **52**, 673–684 (1996).

5. Thall, P. F., Simon, R. and Estey, E. H. 'Bayesian sequential monitoring designs for single-arm clinical trials with multiple outcomes', *Statistics in Medicine*, **14**, 357–379 (1995).
6. Thall, P. F., Simon, R. and Estey, E. H. 'New statistical strategy for monitoring safety and efficacy in single-arm clinical trials', *Journal of Clinical Oncology*, **14**, 296–303 (1996).
7. Thall, P. F. and Russel, K. E. 'A strategy for dose-finding and safety monitoring based on efficacy and adverse outcomes in phase I/II clinical trials', *Biometrics*, **54**, 251–264 (1998).
8. Thall, P. F. and Sung, H-G. 'Some extensions and applications of a Bayesian strategy for monitoring multiple outcomes in clinical trials', *Statistics in Medicine*, **17**, 1563–1580 (1998).
9. Lazaridis, E. and Gonin, R. 'Continuously monitored stopping boundary methodologies: The issues of sensitivity, association and trial suspension', *Statistics in Medicine*, **16**, 1925–1941 (1997).
10. Faries, D. 'Practical modifications of the continual reassessment method for phase I cancer clinical trials', *Journal of Biopharmaceutical Statistics*, **4**, 147–164 (1994).
11. Korn, E. L., Midthune, D., Chen, T. T., Rubinstein, L. V., Christian, M. C. and Simon, R. M. 'A comparison of two phase I trial designs', *Statistics in Medicine*, **13**, 1799–1806 (1994).
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*, **14**, 911–922 (1995).
13. Goodman, S. N., Zahurak, M. L. and Piantadosi, S. 'Some practical movements in the continual reassessment method for phase I studies', *Statistics in Medicine*, **14**, 1149–1161 (1995).
14. Estey, E. H., Hoth, D., Simon, R., Marsoni, S., Leyland-Jones, B. and Wittes, R. 'Therapeutic response in phase I trials of antineoplastic agents', *Cancer Treatment Reports*, **70**, 1105–1115 (1986).
15. Smith, T. L., Lee, J. J., Kantarjian, H. M., Legha, S. S. and Raber, M. N. 'Design and results of phase I cancer clinical trials: three–year experience at M. D. Anderson Cancer Center', *Journal of Clinical Oncology*, **14**, 287–295 (1996).
16. Geller, N. L. 'Design of phase I and II clinical trials in cancer: a statistician's view', *Cancer Investigation*, **2**, 483–491 (1984).
17. Storer, B. 'Design and analysis of phase I clinical trials', *Biometrics*, **45**, 925–939 (1989).
18. deMoor, C. A., Higdon, D. M., Hilsenbeck, S. G., Clark, G. M. and von Hoff, D. D. 'Incorporating toxicity grade information in the continual reassessment method for phase I clinical trials', unpublished manuscript (1996).

