Published in final edited form as:

Contemp Clin Trials. 2010 May ; 31(3): 259–270. doi:10.1016/j.cct.2010.03.006.

A Simulation Based Comparison of the Traditional Method, Rolling-6 Design and a Frequentist Version of the Continual Reassessment Method with Special Attention to Trial Duration in Pediatric Phase I Oncology Trials

Arzu Onar-Thomas^a and Zang Xiong^{b,1}

Arzu Onar-Thomas: arzu.onar@stjude.org; Zang Xiong: Ann.Xiong@AmericanMedicalSystems.com

^a St Jude Children's Research Hospital, 332 Danny Thomas Place, MS 768, Memphis, TN 38105, Telephone: 901-595-5499, FAX: 901-595-8843

^b St Jude Children's Research Hospital, 332 Danny Thomas Place, MS 768, Memphis, TN 38105

Abstract

The Traditional Method (TM), also known as the 3+3 up-and-down design, and the Continual Reassessment Method (CRM) are commonly used in Phase I Oncology Trials to identify the maximum tolerated dose (MTD). The Rolling-6 is a relative newcomer which was developed to shorten trial duration by minimizing the period of time during which the trial is closed to accrual for toxicity assessment. In this manuscript we have compared the performance of these three approaches via simulations not only with respect to the usual parameters such as overall toxicity, sample size and percentage of patients treated at doses above the MTD but also in terms of trial duration and the dose chosen as the MTD. Our results indicate that the toxicity rates are comparable across the three designs, but the TM and the Rolling-6 tend to treat a higher percentage of patients at doses below the MTD. With respect to trial duration, Rolling-6 leads to shorter trials compared to the TM but not compared to the CRM. Additionally, the doses identified as the MTD by the TM and the Rolling-6 differ in a large percentage of trials. Our results also indicate that the body surface area-based dosing used in pediatric trials can make a difference in dose escalation/de-escalation patterns in the CRM compared to the cases where such variations are not taken into account in the calculations, even leading to different MTDs in some cases.

Keywords

Body Surface Area Based Dosing; Pediatric Trials; Dose Finding; Maximum Tolerated Dose

Correspondence to: Arzu Onar-Thomas, arzu.onar@stjude.org.

¹A majority of this work was done when the 2nd author was employed at St Jude Children's Research Hospital. The present address of the second author is: American Medical Systems Inc. 10700 Bren Rd W. Minnetonka, MN 55343
Ann.Xiong@AmericanMedicalSystems.com

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Phase I trials represent an important step in oncology drug development process and are often used to determine the maximum tolerated dose (MTD) to be studied in later phase trials for further toxicity assessments and evidence of efficacy. In classical Phase I trials, a short list of escalating doses are investigated to determine the highest dose with tolerable rate of toxicity i.e. the maximum tolerated dose (MTD). While with some molecularly targeted or immunogenic agents, the toxicity may be too low and thus the definition of the target dose may be based on different criteria such as adequate biologic activity, this paper focuses on agents for which toxicity would be the main determinant of the dose to be carried forward. Further our focus in this manuscript is on pediatric oncology phase I trials which differ in important ways from their adult counterparts.

Currently there are two main approaches to dose finding algorithms in pediatric oncology: empirical methods and model based methods. Perhaps the most widely used empirical method is the so-called traditional method (TM), also known as the 3+3 up-and-down design. Recently a modified version of the TM has been proposed by Skolnik et al.[1] called the Rolling-6 design, which is currently being used in all Phase I trials conducted by the Children's Oncology Group's (COG) Phase I Consortium. The preferred method to dose finding by the Pediatric Brain Tumor Consortium (PBTC) on the other hand is the Continual Reassessment Method (CRM) introduced by O'Quigley et al. [2] which is a model based. The operating characteristics of the CRM and TM have been widely studied [3-6]. Our intent here is to add the Rolling-6 into the mixture and compare its performance to the other two in a pediatric context. A need for such a comparison was initially articulated by Hartford et al.[7] in an editorial that accompanied the Skolnik et al.'s[1] JCO paper which introduced the Rolling 6 design. In the simulations presented here, the versions of the CRM and TM currently employed by the PBTC were studied. Similar to other published simulation-based studies these three designs were compared with respect to sample size, dose chosen as the MTD, overall toxicity and assignment profile of patients to various dose levels. Unlike most other published studies however, we also compared the designs with respect to trial duration and overall agreement in the estimated MTD. The intent behind these simulations was to learn more about the operating characteristics of these designs in a pediatric oncology context and to identify the circumstances in which the use of one approach may be advantageous over the others.

In Section 2 below we briefly summarize the properties of the three dose finding designs we wish to compare and in Section 3 we discuss some special challenges that need to be accommodated in pediatric Phase I trials. In section 4 we provide details on our simulations whose results are summarized in Section 5. In the last section we discuss the overall results as well as offer some practical guidance on the use of these designs.

2. Brief Description of the TM, the Rolling-6 and the CRM

In Phase I trials where the target dose is determined based on toxicity often a pre-determined list of dose-related toxicities is used to assess the outcome of patients. Patients who experience one or more of these pre-determined dose limiting toxicities (DLT) are counted as failures. Perhaps one of the most commonly used dose-finding designs is the TM also known as the 3+3 up-and-down design. The approach is empirical and utilizes fixed cohort sizes, most common of which is three. A widespread version of the algorithm is initiated by assigning three patients to the lowest (or one of the lowest) proposed dose level and escalation to the next dose occurs if none of these three patients experience a dose limiting toxicity. If one of these patients experiences a DLT then the cohort is expanded by treating 3 more patients at the same dose level. If no other DLTs are observed at that dose level, then escalation to the next dose level occurs. If $\geq 2/3$ or $\geq 2/6$ patients experience DLTs at a dose level then that dose is declared

as 'too toxic' and the next lower dose level is investigated using the same rules. The MTD is declared when 6 patients are safely treated at a dose level (i.e. no more than 1 DLT is observed in 6 patients) and the next dose level is too toxic; otherwise no MTD determination is made.

The Rolling-6 design was recently introduced by Skolnik et al.[1] and it carries the motivation to shorten the duration of pediatric Phase I trials by minimizing the time the trial would be closed to accrual for toxicity monitoring. The approach is quite similar to the TM in that it is empirical, requires a pre-set list of doses to study and uses the same criteria to declare a dose too toxic (i.e. more than 1 DLT at any dose level with 3-6 patients) and to pronounce the MTD as estimated. The difference is that the Rolling-6 design allows enrolling anywhere from 3 to 6 patients at a dose level without requiring that the DLT status of the patients already assigned to the same dose level are known; hence reducing the number of patients who would have to be turned away due to unavailability of open slots. The justification for this more aggressive enrollment strategy was provided by Lee et al.[8] who based on a literature study noted that most pediatric Phase I trials in oncology have not produced excessive toxicities possibly because they were often preceded by their adult counterparts and the knowledge gained in the latter could be utilized towards ensuring safety in pediatric trials. The simulations in Skolnik et al.[1] indicate that the Rolling-6 design decreases the duration of Phase I trials compared to the TM.

The CRM was initially introduced by O'Quigley et al.[2] and it is a model based design. The original design has gone through many revisions leading to a variety of versions [see 4-6, 9-11 for examples] both in Bayesian and frequentist settings. The basic principle common to a majority of these versions is that a model is specified that relates dose level to probability of toxicity and the data which arrives sequentially is used to estimate the model parameters. The MTD is defined as the dose level with a pre-set target toxicity probability, usually in the 20-40% range. Since at the start of the trial very little data is available, prior information is utilized to stabilize the calculations, even in frequentist versions of the algorithm. Here we will limit our detailed description of the algorithm to one version that has been adopted by the PBTC [6], since this is the one which we have employed in our simulations. Similar to [4], we use a frequentist, likelihood-based version of the CRM based on a two-parameter logistic function to model the relationship between toxicity and dose, with pre-specified list of dose levels usually 30% apart. In a consortium setting pre-set dose levels are both more acceptable to clinicians and are easier to manage operationally as specific dosing tables can be incorporated into the protocol.

At the beginning of the trial since no data is available, certain assumptions are needed to initiate model fitting. Instead of making assumptions, similar to the approach employed in [4] we introduce some prior information. Specifically we identify two dose levels, one presumed to have a very low toxicity level (e.g. 1%) and the other a very high toxicity level (e.g. 99%), and we assume that 5 patients have been treated at each of these two levels and that the expected number of toxicities have occurred. The toxicity probability and the cohort size associated with the prior information were based on extensive simulations (results not shown). The values ultimately chosen led to good operating characteristics with respect to estimating the model parameters and moderating the influence of these prior parameters on the final outcome. By incorporating such prior information into the model, we effectively tie the extremes of the logistic curve and let the observed data determine the shape of the curve between these two points.

Similar to [5], we initiate the trial at the first proposed dose level, usually 80% of the adult MTD, and escalate one dose level at a time with no limitation on dose de-escalation. In our list of dose levels we also include a dose level 0, often designated as 30% below the starting dose level, to account for the possibility that the starting dose level may be too toxic and to avoid

having to amend the trial in the event that this does happen. At each dose level we open 3 slots and require that the DLT information is available from at least two patients before an escalation/de-escalation decision can be made. We consider the MTD as estimated when at least 6 patients have been treated at the proposed MTD, and treating two more hypothetical patients at that dose level would not lead to escalation. This stopping rule incorporates a mixture of approaches proposed by Korn et al. [3], O'Quigley and Reiner [10] and O'Quigley [9]. We have studied the operating characteristics of this version of the CRM in extensive simulations which have been reported in [6].

3. BSA Adjusted Dosing in Pediatric Trials

As indicated above, in most pediatric oncology Phase I trials patients are dosed based on their BSA. This is done in order to accommodate the wide variation in sizes among patients whose ages may vary from 0 to 21 years. Hence a patient with a BSA of 1.12m^2 who may be assigned to $150\text{mg}/\text{m}^2/\text{day}$ would need $168\text{mg}/\text{day}$ of drug. In cases where the drug under study is oral, available pill sizes would have to be accommodated in these calculations. For example if the available pill sizes are in multiples of 20mg with the smallest pill size being 20mg for the agent in the example above then the closest deliverable dose for a patient with a BSA of 1.12m^2 is $160\text{mg}/\text{dose}/\text{day}$ and therefore the patient will receive $142\text{mg}/\text{m}^2/\text{dose}/\text{day}$ (i.e. 160mg divided by 1.12m^2).

Oral drugs are desirable both in adult and pediatric oncology settings as unlike IV drugs, they do not require hospitalizations; hence they potentially provide better quality of life for patients and their families. However it is uncommon to have pediatric formulations of oral agents available at the time of a pediatric phase I trial and this can lead to wide variations from the targeted levels in doses received due to the BSA adjustment (see 12-14 for examples). Since a child is often defined as anyone younger than 21 years old for the purposes of pediatric trials [12-15], wide ranges of BSAs may be observed. The median BSA in the PBTC database based on patients treated during the past 10 years is 1.15m^2 with 25th and 75th percentiles of 0.82m^2 and 1.51m^2 , respectively; and a range of 0.36m^2 to 2.78m^2 . In contrast a 'typical' adult BSA is 1.7m^2 . Figure 1 displays the proposed dose levels as well as the closest deliverable BSA-adjusted doses for a recently published PBTC Phase I trial (PBTC-006) of STI571 [15] where total daily doses ranging from $100\text{mg}/\text{m}^2$ to $465\text{mg}/\text{m}^2$ given once daily were studied. The dosing approach required that the total daily dose would be administered as close as possible to the target level. The available pill sizes for this agent at the time of the Phase I trial were 50mg and 100mg . As it is evident from Figure 1, wide variations from the targeted doses were observed. Areas of overlap (marked by the arcs) are of particular concern as these are the regions where the actual deliverable doses would be the same for the dose levels involved. For example patients with BSAs in the range $0.50\text{--}0.62\text{m}^2$ would receive the same actual mg/m^2 dose whether they would be assigned to $200\text{mg}/\text{m}^2$ or to $150\text{mg}/\text{m}^2$. Hence a de-escalation from $200\text{mg}/\text{m}^2$ to $150\text{mg}/\text{m}^2$ would not make any difference for these patients in terms of the amount of drug they would receive and this may raise safety concerns. See [19] for additional comments and possible strategies for accommodating these variations during the actual conduct of a trial.

Here we focus on another consequence of such dosing variations from the targeted dose levels i.e. their effect on the MTD estimate. Clearly empirical methods such as the TM or the Rolling-6 cannot accommodate the actual dose levels studied as they only use the rank of the targeted dose level as part of the algorithm. The CRM on the other hand can incorporate the values of the dose levels themselves both targeted as well as the actually delivered values. Hence in the simulation studies that we will discuss below, two sets of results for the CRM design will be provided, one based on the targeted level, ignoring the BSA-based variation and the other based on the actual delivered dose as adjusted for BSA. Our aim is to document the magnitude of the

effect such BSA-based dosing may have on the accuracy of the MTD estimate, as well as on various other properties of the CRM such as the overall toxicity, sample size etc. For the rest of this manuscript when we use the word design we will mean the three dose finding approaches i.e. TM, Rolling-6 and the CRM; whereas when we use the term algorithm we will refer to the four versions studied here i.e. TM, Rolling-6, CRM and CRM incorporating BSA-based dosing which we will call CRM-BSA.

4. Details on the Simulation Set-Up

As indicated above the aim of our simulations is to compare the TM, the Rolling-6 and the CRM based on various criteria of interest in a Phase I pediatric oncology trial setting. Conditional on the assumption that the other commonly used criteria such as toxicity and sample size are comparable, our two main foci in this work are estimating the expected duration of the trial under these three designs as well as capturing the effect of BSA-based dosing on the MTD estimate for the CRM. Figure 2 provides a visual display of a subset of the dose-toxicity relationships we have investigated. They were chosen to span a variety of scenarios including cases where the starting dose may be too toxic as well as cases where the highest dose level's toxicity is below the targeted toxicity probability. While all six distributions in Figure 2 were generated from a 2-parameter logistic distribution, in an attempt to investigate the sensitivity of the CRM to model misspecification, we have also studied distributions defined by power curves and arc-tangent functions. However given the flexibility of the 2-parameter logistic model, the results were essentially unchanged and thus these cases will not be discussed further here.

In order to simulate the trials we assumed that patients arrive at a uniform rate of 18/year based on a Poisson process. This rate is consistent with what we have observed in actual trials, though we have also studied slower (10 patients/year) and faster (36 patients/year) accrual rates. Using a Poisson process, for each trial we generated a large group of patients (several fold more than what would ordinarily be needed to complete the trial) and used the same set of patients for all four algorithms i.e. TM, Rolling-6, CRM and CRM-BSA. For each of these simulated patients, we generated an indicator regarding whether the patient would be evaluable (based on a Bernoulli distribution with $p=0.9$) and for the inevaluable patients, we used a normal distribution with mean 14 days from start of treatment and standard deviation 2 days to model the time it would take to process the data and to declare a patient inevaluable. Once a patient was declared inevaluable, a slot was made available to replace him/her. We also generated DLT outcomes for each patient at all possible dose levels based on a Bernoulli distribution with the toxicity probability as specified for each dose level, since depending on the algorithm the patient could be assigned to different dose levels. We used a normal distribution with mean 7 days from end of dose finding period and standard deviation 1.5 days to model the time to data sign-off. Barring the cases where the dose finding period would end early due to a DLT, we used 28 days as the dose finding period, which is typical for many of our trials [12, 16-18]. Within each simulated trial, for all four dose finding algorithms, we used the same set of data generated as described above, where we enrolled a patient on the trial if a slot was available at the time of his/her arrival; otherwise we assumed that the patient would be lost. While in most settings a stand-by queue is available for patients to wait, we have ignored this possibility in our simulations as in our experience it has been difficult to reliably predict whether patients would wait for slots to open or seek other alternatives. For each of the simulated cases we used a target toxicity probability of 20%, dose levels 140mg/m², 180 mg/m², 230 mg/m², 300 mg/m², 400 mg/m², 520 mg/m² with various toxicity probabilities as summarized in Tables I-VI. Dose level 180 mg/m² was designated as the starting dose. A minimum pill size of 50mg was assumed in simulating trials for CRM-BSA. The dose levels used in the simulations as well as the 50mg minimum pill size are very similar to the set-up

studied in the PBTC003 trial [12]. For each distribution 1000 trials were simulated using SAS ver. 9.1.

5. Results

The simulation results are summarized in Tables I-VI where the probability of choosing each of the proposed dose levels as the MTD, the percentage of patients treated at each dose level, median sample size, median percent toxicity, percentage of trials with ≥ 3 DLTs at a dose level and percentage of trials with at least one change in the dose escalation/de-escalation decisions between the CRM and CRM-BSA are presented. As the results in the tables indicate, the TM and Rolling-6 tend to choose lower doses as the MTD compared to the CRM or CRM-BSA for which the target toxicity probability was set at 20%. Despite this however the median toxicity probabilities for the 4 algorithms, including the two versions of the CRM, considered here tend to be very comparable. Note that the median percent toxicity is the median across all simulated trials of the percent toxicity observed per trial where percent toxicity is defined as number of patients with DLTs divided by the total sample size in a given trial.

As it is evident from Figure 2, the toxicity profiles of the six distributions considered here are variable hence the median percent toxicity expected for each distribution is different. For example since the starting dose level of Distribution 4 has 40% toxicity probability and even the lowest proposed dose level has 33% toxicity probability, we expect small sample sizes with a large number of DLTs and hence a large median percent toxicity and indeed the simulation results confirm this expectation. On the other hand the toxicity profiles of Distribution 2 and 3 are relatively flat and thus we expect to escalate to the higher dose levels leading to larger sample sizes without too much toxicity. Once again the simulation results are in agreement with this expectation. The other three distributions, namely Distributions 1, 5 and 6 start with low probability of toxicity which increases with increasing doses. The rates of increase are different among these distributions which translate to different median percent toxicities. For example distribution 6 has the fastest increase in toxicity probability early in the life of the trial leading to smaller sample sizes with larger toxicity probabilities compared to Distribution 5 whose toxicity probabilities stay relatively flat through the first few dose levels leading to larger sample sizes and lower toxicity probabilities, whereas Distribution 1 falls somewhere in between Distributions 5 and 6.

Similarly the estimated experimentation percentage values, i.e. the percentage of patients treated at each dose level averaged across the 1000 simulated trials, do not reveal a striking difference across the 4 algorithms. We also tracked the percentage of trials with ≥ 3 DLTs at a dose due to initial objections from our clinical colleagues that this may be more likely for the CRM. As noted elsewhere[6], the value of this parameter obtained for the CRM or the CRM-BSA is comparable to the value observed for the TM; however as the tables I-VI clearly show, it may be somewhat higher for the Rolling-6 design. This is not surprising given the larger cohort sizes employed by the Rolling-6. With respect to sample size, again not surprisingly, the largest values are observed for the Rolling-6 design, whereas CRM and CRM-BSA tend to have lowest sample sizes overall. These observations are consistent with results published in the literature by us as well as by others [4-6].

One of our main motivations for conducting this simulation exercise was to investigate the properties of trial duration under the four algorithms. Consistent with the results reported in [1], our simulation results summarized in Table VII also showed that while the average difference in duration varies based on the dose toxicity relationship, the Rolling-6 design leads to shorter trials compared to the TM. Interestingly however the average trial duration associated with the CRM is comparable and often shorter than the average trial duration observed under the Rolling-6 design. In fact in some cases the gains in the trial duration by using the CRM

instead of the Rolling-6 approach can be substantial, as evidenced by the results in Table VII. The advantage of using the Rolling-6 design in reducing the study duration is more substantial when accrual rate is fast. In our simulations where the accrual rate was assumed to be 36 patients/year, which is quite high for most PBTC studies, the Rolling-6 design led to shorter trials on average in 3 out of the 6 cases studied compared to the CRM. In these three cases the median differences in duration ranged from 10-39 days. On the other hand, the simulated results based on 10 patients/year revealed shorter trials on average for the CRM where the differences in median duration across the six scenarios ranged from 23 days to 183 days. This can be largely attributed to the fact that our version of the CRM uses 2- or 3-patient cohorts for the initial visit to a dose level, after which depending on the accrual rate, escalation/de-escalation decisions can be made based on a single patient's toxicity outcome. In contrast, the Rolling-6 requires at least 3 patients per cohort before an escalation/de-escalation decision can be made.

We have also investigated the agreement across the 4 algorithms to dose finding in terms of the final MTD estimate and these results are summarized in Table VIII. Naturally the level of agreement on the MTD estimate among the 4 algorithms varies across the six scenarios considered here but overall the agreement is quite low. In some ways this is not surprising as the definition of the MTD is not consistent between the TM and the CRM for example, nevertheless all of these designs share a common objective: identifying a dose to be studied in later phase trials. Considering that the MTD estimates across the 4 algorithms in Table VIII were obtained based on simulated data which was kept as similar as possible, quantifying the extent of these differences is informative, especially between the Rolling-6 and the TM, since these two empirical designs are quite similar and the former was motivated by the desire to shorten the trial duration associated with the latter. While this was not mentioned by Skolnik et al. [1] it is notable that the agreement on the estimated MTD between the TM and the Rolling-6 is nearly as poor as the agreement between the TM and the CRM. The average disagreement rate between the TM and the Rolling-6 ranged between 19%-52% of the simulated trials with no apparent trend in the MTD estimates of one over the other. It may come as a surprise to our clinical colleagues to observe that by allowing larger cohorts in an effort to reduce the duration of the trial, the MTD estimate will likely be altered as well. The disagreement rate between the TM and the CRM and the Rolling-6 vs. the CRM on the other hand were 16%-68% and 23-74%, respectively.

Table IX provides the pair-wise Spearman's Rank Correlations of the MTD estimates across the 4 algorithms and the 6 distributions as estimated based on the 1,000 simulated trials. As it is evident from the table, the agreement in the MTD estimates between the CRM and CRM-BSA is quite high but the correlations within the rest of the pairs are substantially lower. Table X provides the breakdown of the MTD estimates between the TM and the Rolling 6, the TM and the CRM as well as between the CRM and the Rolling 6 for Distributions 3 and 5. Distribution 3 was included in this table as it is one of the most extreme distributions in terms of disagreement. For Distribution 3, the main source of this disagreement between the two empirical designs, i.e. TM and the Rolling-6, and the CRM is a design characteristic of the former. More specifically, in 1000 simulated trials dose level 520 was never chosen as the MTD by either the TM or the Rolling-6. This is expected as the last level cannot be the MTD for either of these two designs since both require that before a dose level can be chosen as the MTD, the next dose level must be studied and have to be declared too toxic. For this distribution since the overall toxicity level was quite low, in a large percentage of trials both the TM and the Rolling-6 led to the decision that the MTD was not determined whereas the CRM chose doses 400 or 520 as the MTD in more than half of the cases. The disagreement between the TM and the Rolling-6 on the other hand is due to the variation in the doses that were chosen as the MTD. Distribution 5 is another case where the correlation values for the MTD estimates across the CRM, TM and Rolling-6 are low. As Table X shows however all three designs

choose one of three doses as the MTD in almost all simulated trials and the disagreement in this case is simply due to differing choices among these three designs.

Though to a lesser extent, our simulations indicate that such a discrepancy in the MTD estimates can also occur when BSA-based dosing is taken into account in the MTD estimation procedure. The last column in Table VIII provides the percentage of simulated trials where the MTD estimate was different between CRM and CRM-BSA. These values range between 6.4% - 13.8% across the six distributions presented here. Similarly Table IX shows the rank correlations of the MTD estimates obtained for the CRM and CRM-BSA. In all six cases the correlations are quite high but not perfect. While the disagreement in the MTD estimates between the CRM and CRM-BSA is substantially lower than the disagreement observed for any of the other 5 pairings, it is still notable as it is caused by the BSA-based dosing variations alone. Additionally, the last row in Tables I-VI provides an estimate of the percentage of trials where the dose escalation/de-escalation decisions were not identical between the CRM and the CRM-BSA. These percentages vary from 10.7% to 28.3%, which is non-ignorable. On the other hand the simulation results presented in Tables I-VI do not indicate any notable advantages in using the CRM-BSA over CRM in terms of improved accuracy, lower toxicity, shorter trial duration or lower sample size. In all these parameters, the two model-based approaches seem highly similar. Having said that, the mean absolute deviation values of the estimated MTDs from the true MTDs are overall lower for the CRM-BSA compared to the CRM (results not shown). In scrutinizing the simulation results based on 10 patients/year and 36 patients/year and comparing them to the results obtained for 18 patients/year reveals that the disagreement between the CRM and CRM-BSA with respect to the MTD estimate as well as with respect to escalation/de-escalation decisions decrease with increasing accrual rate; whereas the reverse is true with respect to the disagreement in the MTD estimate for the TM vs. the Rolling-6 as well as the CRM vs. the Rolling-6. This can be partially explained by the increasing difference in average sample sizes between the Rolling-6 and the TM and the Rolling-6 and the CRM with increasing accrual rate.

The results presented above are specific to the distributions that were studied in the simulations but we believe the overall observations would hold in other settings as well. The distributions studied here cover a variety of common and plausible scenarios and the design parameters such as the number of doses studied as well as 20% target toxicity probability for the CRM are typical in our experience. Naturally changing some or all of these parameters would lead to different results. For example if the target toxicity probability is set at 10% or 30% in the designs that we have studied, we expect that the MTD estimates between the CRM and the two empirical designs would differ even more frequently. As it is also evident from the results in Tables I-VI, this is because 20% seems to be the approximate toxicity probability associated with the MTD for the TM and the Rolling-6 design. With a lower toxicity target for the MTD, say 10%, we expect that the sample size would be smaller for the CRM compared to the TM and the Rolling-6 design and the overall toxicity will also be lower; whereas the opposite is true if the toxicity target for the MTD is raised to 30%. If one considers cases where a fewer dose levels are proposed to be investigated, it may not be possible to employ the CRM as fitting a logistic regression to data based on a few dose levels may not be reasonable. The TM and the Rolling-6 Design do not suffer from this shortcoming. Within the two empirical designs we expect that the trends discussed above regarding the operating characteristics studied would hold even with fewer dose levels.

6. Discussion

In this manuscript three common dose finding designs in pediatric oncology trials, namely TM, Rolling-6 and the CRM, were studied in an effort to learn more about their operating characteristics and to add to the existing literature. Except for the original paper by Skolnik et

al. [1] which introduced the algorithm, to our knowledge this is the first large scale comparison study of the Rolling-6 design to the TM and the CRM. Further the simulations presented here are different from many others in the literature in the sense that an attempt was made to closely simulate the actual trial environment by incorporating patient arrival times, inevaluability, time to data sign-off etc. and the same set of simulated patients were used as much as possible for each of the dose finding algorithms per trial within the constraints of the differing slot availability and escalation/de-escalation rules. In addition to the usual parameters of interest, i.e. estimated MTD, toxicity probabilities, experimentation percentage, average sample size etc., comparisons based on study duration were also made. The results presented here confirm that Rolling-6 designs tend to declare the MTD as estimated faster than the TM, but not faster than the CRM. In other words while both the Rolling-6 and the CRM seem to be associated with shorter trial durations compared to the TM, the winner between the Rolling-6 and the CRM is not as clear-cut. For the accrual rates typical for pediatric brain tumor trials, CRM is often faster than the Rolling 6 though the simulation results also indicate that this advantage may shift to the Rolling-6 design if much faster accrual rates than the ones studied here are attainable.

Note that all three of these designs are ultimately used to determine a dose to be carried forward to later phase trials hence it is of interest to know by how much the MTD estimates may differ among them. Since the TM and the Rolling-6 do not have a toxicity target, it is not surprising that the MTD estimate based on either of these two designs tends to be different from the one produced by the CRM but perhaps it is surprising that the MTD estimate based on the TM and the Rolling-6 differ almost as often. We believe it is important to realize that Rolling-6 does not merely offer a faster way to reach the same MTD as the TM despite the fact that these two designs appear quite similar with respect to their escalation/de-escalation rules. It may be comforting however that a closer scrutiny of the MTD estimates obtained by the TM and the Rolling-6 in the simulated trials did not reveal any patterns regarding higher or lower MTD estimates for one of the designs. Variations in the escalation/de-escalation decisions as well as in the MTD estimates between CRM and CRM-BSA were also investigated, with the intent to determine the impact of BSA based dosing on these parameters. In a notable percentage of cases, taking BSA-based dosing into account led to differences in escalation/de-escalation decisions throughout the life of the trial and could in fact result in different MTD estimates.

Perhaps the most interesting issue is determining which design to use when. The CRM is favored among many statisticians for its model based approach and well defined MTD whereas the TM is well-liked by many clinicians for its simplicity. The Rolling-6 design which retains most of the simplicity of the TM while reducing the duration of the clinical trial compared to the TM appears to be an attractive option. The simulation results presented here indicate that the toxicity probabilities are comparable across all three designs. The number of patients treated at doses above the MTD also seems to be well controlled in all three designs but the TM and the Rolling-6 seem to have the tendency to treat more patients at lower dose levels compared to the CRM with a target toxicity probability of 20%. Thus from a safety perspective there does not appear to be a disadvantage in using the Rolling-6 design over the TM and as stated by Skolnik et al. [1] the Rolling-6 design is indeed superior to the TM with respect to trial duration. One caveat here is that while the percentage of toxicity observed for the Rolling-6 design is not worse than the percentage of toxicity observed for the TM, since the Rolling-6 design tends to have larger sample sizes, the number of toxicities observed can be larger especially when the accrual is fast. So if the agent turns out to be toxic, having 3/3 toxicities at the 1st dose level may be more palatable than having 6/6 toxicities.

The implications of the simulation results are not as clear cut when one ponders comparisons between the Rolling-6 and the CRM, which was not investigated in [1]. CRM did not seem to have been considered in Skolnik et al.'s motivation in proposing the Rolling-6 design but we

believe it is important to compare its operating characteristics to the Rolling-6 as the CRM is also a common Phase I design and it is often the design of choice for statisticians, including the ones serving the PBTC. Though sensitive to the unobserved dose-toxicity relationship, the simulations presented here reveal that the number of patients needed to determine the MTD is typically lowest for the CRM and highest for the Rolling-6 design, as expected, with the TM falling in between the two. As mentioned above, for slow to moderate accrual, the trial duration appears to be shorter for the CRM compared to the Rolling-6 design though for trials with fast accrual, Rolling-6 may lead to shorter duration than the CRM. Hence one of the new insights that was revealed from these simulations is that neither from a safety perspective nor from an accrual and trial duration stand point is there an advantage in using the Rolling-6 design over the CRM. Further the fact that the CRM utilizes all available data in estimating an MTD which is associated with a pre-determined toxicity probability and can incorporate variations in deliverable doses due to BSA-based dosing, make it an attractive option.

There are some scenarios however where empirical designs, such as the Rolling-6 may be advantageous over model based designs such as the CRM. One notable case which is common is when only 2-3 dose levels are proposed as a brief safety trial prior to proceeding to a Phase II study. In such cases a model based design may not be suitable since very little data would be available to fit the model. Another scenario favoring the Rolling-6 design is trials based on agents for which no or very little toxicity is expected, such as some molecularly targeted or immunogenic agents. Additionally Phase I trials with extended dose-finding periods where accrual rate is high may lend themselves well to designs such as the Rolling-6, as a way of shortening the trial duration and allowing more patients to participate in Phase I trials. One such example is trials where radiotherapy and chemotherapy are being administered concurrently in a Phase I setting and the dose finding period spans the entire radiotherapy treatment (usually 6-8 weeks) plus 2-4 weeks of additional observation period for delayed toxicities. The PBTC has successfully utilized the CRM for such trials also but the extended suspension of accrual while awaiting dose finding data is often a source of frustration for clinicians as well as patients and families. Thus such trials may create a favorable setting for the use of the Rolling-6 design, especially if only a few doses are proposed for the study and the expected toxicity incidence is low.

In conclusion, believe the Rolling-6 design provides an attractive alternative over the TM since it's safety profile is comparable to the TM yet it shortens the duration of the trial. One must recognize however that the MTD estimate obtained by the Rolling 6 design will not necessarily be the same as the MTD that would have been estimated using the TM for the same trial. In cases where 5, 6 or more dose levels are proposed to be studied and some toxicities are expected, model based designs, such as the CRM, have distinct advantages in being able to use the data from all dose levels in estimating the MTD, in accommodating patient-specific dosing and in providing an MTD estimate that is associated with a toxicity probability. The simulation-based comparison between the Rolling-6 and the CRM presented here indicates that the latter leads to shorter trial durations for slow to medium accrual rates whereas the former may have an advantage if the accrual rate is fast. The Rolling-6 may also be preferable over the CRM if very few or no toxicity is expected with the agent under study and if the dose finding period is long.

Acknowledgments

This work was supported in part by NIH grant U01 CA81457 for the Pediatric Brain Tumor Consortium (PBTC) and the American Lebanese Syrian Associated Charities. The authors acknowledge helpful discussions with the staff of the Operations and Biostatistics Center (OBC) for the PBTC in particular, Dr. James M. Boyett, Principal Investigator of the PBTC and the Executive Director of the OBC, as well as the support of the PBTC investigators.

References

1. Skolnik JM, Barrett JS, Jayaraman B, Patel D, Adamson PC. Shortening the timeline of pediatric phase I trials: The rolling six design. *J Clin Oncol* 2008;26:190–195. [PubMed: 18182661]
2. O'Quigley J, Pepe M, Fisher L. Continual Reassessment Method: A practical design for Phase I clinical studies in cancer. *Biometrics* 1990;46:33–48. [PubMed: 2350571]
3. Korn EL, Midthune D, Chen TT, Rubinstein LV, Christian MC, Simon RM. A Comparison of two Phase I trial designs. *Stat Med* 1994;14:1799–1806. [PubMed: 7997713]
4. Piantadosi S, Fisher JD, Grossman S. Practical implementation of a modified continual reassessment method for dose finding trials. *Cancer Chemother Pharmacol* 1998;41:429–436. [PubMed: 9554585]
5. Goodman SN, Zahurak ML, Piantadosi S. Some practical improvements in the continual reassessment method for Phase I studies. *Stat Med* 1995;14:1149–1161. [PubMed: 7667557]
6. Onar A, Kocak M, Boyett JM. Modified Continual Reassessment Method versus the Traditional Empirically-Based Design for Phase I Trials in Pediatric Oncology: Experiences of the Pediatric Brain Tumor Consortium. *J Biopharm Stat* 2009;19:437–455. [PubMed: 19384687]
7. Hartford C, Volchenbom SL, Cohn SL. 3+3≠ (Rolling) 6. *J Clin Oncol* 2008;26:170–171. [PubMed: 18182656]
8. Lee DP, Skolnik JM, Adamson PC. Pediatric Phase I Trials in Oncology: An Analysis of Study Conduct Efficiency. *J Clin Oncol* 2005;23:8431–8441. [PubMed: 16293874]
9. O'Quigley J. Continual reassessment designs with early termination. *Biostatistics* 2002;3:87–99. [PubMed: 12933626]
10. O'Quigley J, Rainer E. A stopping rule for the continual reassessment method. *Biometrika* 1998;85:741–748.
11. Heyd J, Carlin B. Adaptive design improvements in the continual reassessment method for phase I studies. *Stat Med* 1999;18:1307–1321. [PubMed: 10399198]
12. Kieran M, Packer R, Onar A, Blaney SM, Phillips P, Pollack IF, Geyer JR, Gururangan S, Banerjee A, Goldman S, Turner CD, Belasco JB, Broniscer A, Zhu Y, Frank E, Kirschmeier P, Statkevich P, Yver A, Boyett JM, Kun LE. Phase I and Pharmacokinetic Study of the Oral Farnesyltransferase Inhibitor Sarasar™ (Lonafarnib–SCH66336) Given Twice Daily To Pediatric Patients with Advanced Central Nervous System Tumors: A Pediatric Brain Tumor Consortium (PBTC) Study. *J Clin Oncol* 2007;25:3137–3143. [PubMed: 17634493]
13. Villablanca JG, Krailo MD, Ames MM, Reid JM, Reaman GH, Reynolds CP. Phase I Trial of Oral Fenretinide in Children With High-Risk Solid Tumors: A Report From the Children's Oncology Group (CCG 09709). *J Clin Oncol* 2006;24:3423–3430. [PubMed: 16849757]
14. Widemann BC, Salzer WL, Arceci RJ, Blaney SM, Fox E, end D, Gillespie A, Whitcomb P, Palumbo JS, Pitney A, Jayaprakash N, Zannikos P, Balis FM. Phase I Trial and Pharmacokinetic Study of the Farnesyltransferase Inhibitor Tipifarnib in Children With Refractory Solid Tumors or Neurofibromatosis Type I and Plexiform Neurofibromas. *J Clin Oncol* 2006;24:507–516. [PubMed: 16421428]
15. Pollack IF, Jakacki RI, Blaney SM, Hancock ML, Kieran MW, Phillips P, Kun LE, Friedman H, Packer RJ, Banerjee A, Geyer JR, Goldman S, Poussaint TY, Krasin MJ, Wang Y, Hayes M, Murgu A, Weiner S, Boyett JM. Phase I Trial of Imatinib in Children with Newly Diagnosed Brainstem and Recurrent Malignant Gliomas: A Pediatric Brain Tumor Consortium Report. *Neuro Oncol* 2007;9:145–60. [PubMed: 17293590]
16. Broniscer A, Gururangan S, MacDonald TJ, Goldman S, Packer RJ, Stewart CF, Wallace D, Danks MK, Friedman HS, Poussaint TY, Kun LE, Boyett JM, Gajjar A. Phase I Trial of Single-Dose Temozolomide and Continuous Administration of O6-Benzylguanine in Children with Brain Tumors: a Pediatric Brain Tumor Consortium Report. *Clin Cancer Res* 2007 Nov 15;13(22):6712–6718. [PubMed: 18006772]
17. MacDonald TJ, Stewart CF, Kocak M, Goldman S, Ellenbogen RG, Phillips P, Lafond D, Poussaint TY, Kieran MW, Boyett JM, Kun LE. Phase I Clinical Trial of Cilengitide (EMD 121974) in Children with Refractory Brain Tumors: A Pediatric Brain Tumor Consortium Study (PBTC-012). *J Clin Oncol* 2008 Feb 20;26(6):919–24. [PubMed: 18281665]

18. Gururangan S, Turner CD, Stewart CF, O'Shaughnessy M, Kocak M, Poussaint TY, Phillips PC, Goldman S, Packer RJ, Pollack IF, Blaney SM, Karsten V, Gerson SL, Boyett JM, Friedman HS, Kun LE. Phase I trial of VNP40101M (CLORETAZINE®) in children with recurrent brain tumors - A Pediatric Brain Tumor Consortium (PBTC) study. Clin Cancer Res 2008 Feb 15;14(4):1124–30. [PubMed: 18281546]
19. Onar A, Ramamurthy U, Wallace D, Boyett JM. An Operational Perspective of Challenging Statistical Dogma while Establishing a Modern, Secure Distributed Data Management and Imaging Transport System – The Pediatric Brain Tumor Consortium Phase I Experience. Clin Transl Sci 2009;2:143–149.

Figure 1. Variations from Target Doses: BSA Adjusted Actual Daily Dose (mg/m²) vs. Body Surface Area(m²) for PBTC-006 (Phase I Trial of STI571 in Children with Newly Diagnosed Brainstem Gliomas).

Figure 2.
The six sets of distributions used in the simulations

Table 1

Results Comparing Operating Characteristics of the Traditional Method, the Rolling-6 Design and the Continual Reassessment Method Based on 1,000 Simulated Trials for Distribution 1

DISTRIBUTION 1		Traditional Method		Rolling 6		CRM		CRM-BSA	
Dosage (Toxicity Probability)		% Chosen	Exp%#	% Chosen	Exp%	% Chosen	Exp%	% Chosen	Exp %
140 (9.0%)		14.6%	--	17.2%	--	17.8%	--	18.4%	--
180 (12.0%) -Starting dose		22.2%	41.1%	22.5%	46.7%	11.6%	41.5%	9.9%	41.1%
230 (18.0%)		34.6%	30.4%	33.0%	29.5%	32.3%	29.9%	32.8%	30.4%
300 (29.0%)		26.0%	20.6%	23.5%	17.3%	32.6%	21.8%	33.1%	21.8%
400 (49.0%)		2.6%	7.3%	3.7%	5.9%	5.7%	6.5%	5.8%	6.4%
520 (74.0%)		0.0%	0.6%	0.0%	0.6%	0.0%	0.3%	0.0%	0.3%
MTD Undetermined		0.0%	--	0.1%	--	0.0%	--	0.0%	--
Median Sample Size (Empirical 95% CI)		14 (3 - 22)		15 (3 - 25)		12 (2 - 20)		12 (2 - 20)	
Median %Toxicity (Empirical 95% CI)		22 (13 - 67)		22 (11 - 67)		22.22 (13 - 67)		22.22 (13 - 67)	
% of Trials with >= 3 DL'Ts at a Dose		12.2%		18.6%		13.5%		14.1%	
% of Trials CRM and CRMBSA Differed in Escalation/De-escalation Decisions		--		--		26.5%		--	

Abbreviations: CRM: Continual Reassessment Method; BSA: Body Surface Area; CRM-BSA: Continual Reassessment Method where instead of assigned doses, patient specific BSA adjusted doses were used in the calculations; MTD: Maximum Tolerated Dose; DLT: Dose Limiting Toxicity; %Chosen: Percentage among the 1,000 simulated trials the given dose was chosen as the MTD; Exp%: Experimentation Percentage i.e. percentage of patients treated at each dose level averaged across all 1,000 simulated trials; Median Sample Size: Median sample size needed to reach the MTD across all 1000 simulated trials; Median % Toxicity: median across all simulated trials of the toxicity percentage observed per trial where percent toxicity is defined as number of patients with DLTs divided by the total sample size.

Table II
Results Comparing Operating Characteristics of the Traditional Method, the Rolling-6 Design and the Continual Reassessment Method Based on 1,000 Simulated Trials for Distribution 2

DISTRIBUTION 2	Traditional Method		Rolling 6		CRM		CRM BSA	
	% Chosen	Exp ^a %	% Chosen	Exp ^a %	% Chosen	Exp ^a %	% Chosen	Exp ^a %
Dosage (Toxicity Probability)								
140 (9.0%)	9.2%	--	12.5%	--	14.2%	--	14.7%	--
180 (10.0%) - Starting dose	11.7%	31.2%	11.0%	37.1%	5.7%	34.2%	4.6%	34.4%
230 (13.0%)	16.3%	23.5%	19.9%	24.8%	15.8%	23.5%	15.9%	23.2%
300 (17.0%)	27.7%	20.6%	25.2%	19.3%	29.7%	23.0%	29.0%	22.5%
400 (25.0%)	24.5%	16.1%	22.1%	12.7%	27.2%	14.6%	28.3%	15.1%
520 (37.0%)	0.0%	8.7%	0.0%	6.1%	6.7%	4.7%	6.9%	4.8%
MTD Undetermined	10.6%	--	9.3%	--	0.7%	--	0.6%	--
Median Sample Size (Empirical 95% CI)	18 (3 - 26)		20 (4 - 30)		15 (2 - 23)		15 (2 - 24)	
Median % Toxicity (Empirical 95% CI)	18 (6 - 67)		17 (7 - 50)		18 (6 - 50)		18 (6 - 50)	
% of Trials with ≥ 3 DLTs at a Dose	7.2%		11.7%		8.1%		7.8%	
% of Trials CRM and CRMBSA Differed in Escalation/De-escalation Decisions	--		--		28.3%		--	

Abbreviations: CRM: Continual Reassessment Method; BSA: Body Surface Area; CRM-BSA: Continual Reassessment Method where instead of assigned doses, patient specific BSA adjusted doses were used in the calculations; MTD: Maximum Tolerated Dose; DLT: Dose Limiting Toxicity; %Chosen: Percentage among the 1,000 simulated trials the given dose was chosen as the MTD; Exp%: Experimentation Percentage i.e. percentage of patients treated at each dose level averaged across all 1,000 simulated trials; Median Sample Size: Median sample size needed to reach the MTD across all 1000 simulated trials; Median % Toxicity: median across all simulated trials of the toxicity percentage observed per trial where percent toxicity is defined as number of patients with DLTs divided by the total sample size.

Table III

Results Comparing Operating Characteristics of the Traditional Method, the Rolling-6 Design and the Continual Reassessment Method Based on 1,000 Simulated Trials for Distribution 3

DISTRIBUTION 3	Traditional Method		Rolling 6		CRM		CRM BSA	
	% Chosen	Exp ^a %	% Chosen	Exp ^a %	% Chosen	Exp ^a %	% Chosen	Exp ^a %
Dosage (Toxicity Probability)								
140 (5.0%)	2.9%	--	3.1%	--	5.6%	--	5.8%	--
180 (6.0%) - Starting dose	4.1%	21.5%	4.4%	24.9%	2.9%	23.8%	2.0%	23.8%
230 (7.0%)	6.8%	19.5%	8.4%	21.9%	5.7%	19.7%	5.8%	19.7%
300 (9.0%)	12.3%	19.1%	13.0%	20.1%	14.4%	21.1%	15.0%	21.3%
400 (12.0%)	23.4%	19.3%	22.5%	17.8%	33.5%	20.8%	34.1%	21.0%
520 (19.0%)	0.0%	20.6%	0.0%	15.3%	25.5%	14.5%	25.6%	14.3%
MTD Undetermined	50.5%	--	48.6%	--	12.4%	--	11.7%	--
Median Sample Size (Empirical 95% CI)	18 (6 - 24)		25 (6 - 30)		17 (2 - 25)		17 (2 - 26)	
Median % Toxicity (Empirical 95% CI)	10 (0 - 33)		10 (0 - 33)		11 (0 - 50)		11 (0 - 50)	
% of Trials with ≥ 3 DLTs at a Dose	2.6%		3.7%		2.3%		2.6%	
% of Trials CRM and CRMBSA Differed in Escalation/De-escalation Decisions	--		--		24.2%		--	

Abbreviations: CRM: Continual Reassessment Method; BSA: Body Surface Area; CRM-BSA: Continual Reassessment Method where instead of assigned doses, patient specific BSA adjusted doses were used in the calculations; MTD: Maximum Tolerated Dose; DLT: Dose Limiting Toxicity; %Chosen: Percentage among the 1,000 simulated trials the given dose was chosen as the MTD; Exp%: Experimentation Percentage i.e. percentage of patients treated at each dose level averaged across all 1,000 simulated trials; Median Sample Size: Median sample size needed to reach the MTD across all 1000 simulated trials; Median % Toxicity: median across all simulated trials of the toxicity percentage observed per trial where percent toxicity is defined as number of patients with DLTs divided by the total sample size.

Table IV

Results Comparing Operating Characteristics of the Traditional Method, the Rolling-6 Design and the Continual Reassessment Method Based on 1,000 Simulated Trials for Distribution 4

DISTRIBUTION 4	Traditional Method		Rolling 6		CRM		CRM BSA	
	% Chosen	Exp % [#]	% Chosen	Exp %	% Chosen	Exp %	% Chosen	Exp %
Dosage (Toxicity Probability)								
140 (33.0%)	75.5%	--	77.3%	--	80.3%	--	81.4%	--
180 (40.0%) - Starting dose	19.5%	86.2%	19.3%	89.3%	12.7%	83.9%	11.3%	84.1%
230 (48.0%)	4.9%	12.0%	3.4%	9.7%	6.4%	14.0%	7.0%	13.8%
300 (61.0%)	0.1%	1.7%	0.0%	1.0%	0.6%	2.0%	0.3%	2.1%
400 (76.0%)	0.0%	0.1%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%
520 (88.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
MTD Undetermined	0.0%	--	0.0%	--	0.0%	--	0.0%	--
Median Sample Size (Empirical 95% CI)	5 (2 - 15)		5 (2 - 14)		4 (2 - 14)		4 (2 - 14)	
Median % Toxicity (Empirical 95% CI)	50 (22 - 100)		48 (22 - 100)		50 (22 - 100)		50 (22 - 100)	
% of Trials with ≥ 3 DLTs at a Dose	15.6%		26.3%		16.5%		16.9%	
% of Trials CRM and CRMBSA Differed in Escalation/De-escalation Decisions	--		--		10.7%		--	

Abbreviations: CRM: Continual Reassessment Method; BSA: Body Surface Area; CRM-BSA: Continual Reassessment Method where instead of assigned doses, patient specific BSA adjusted doses were used in the calculations; MTD: Maximum Tolerated Dose; DLT: Dose Limiting Toxicity; %Chosen: Percentage among the 1,000 simulated trials the given dose was chosen as the MTD; Exp%: Experimentation Percentage i.e. percentage of patients treated at each dose level averaged across all 1,000 simulated trials; Median Sample Size: Median sample size needed to reach the MTD across all 1000 simulated trials; Median % Toxicity: median across all simulated trials of the toxicity percentage observed per trial where percent toxicity is defined as number of patients with DLTs divided by the total sample size.

Table V

Results Comparing Operating Characteristics of the Traditional Method, the Rolling-6 Design and the Continual Reassessment Method Based on 1,000 Simulated Trials for Distribution 5

DISTRIBUTION 5	Traditional Method		Rolling 6		CRM		CRM BSA	
	% Chosen	Exp % ^a	% Chosen	Exp % ^a	% Chosen	Exp % ^a	% Chosen	Exp % ^a
Dosage (Toxicity Probability)								
140 (0.0%)	0.1%	--	0.1%	--	0.7%	--	0.8%	--
180 (1.0%) - Starting dose	0.2%	17.9%	0.8%	22.9%	0.1%	18.0%	0.0%	18.1%
230 (2.0%)	6.0%	19.3%	5.6%	23.5%	2.3%	18.6%	2.3%	18.6%
300 (7.0%)	55.4%	28.4%	55.3%	25.8%	36.5%	28.9%	35.1%	28.7%
400 (30.0%)	38.0%	26.9%	38.1%	22.1%	59.8%	29.4%	61.2%	29.7%
520 (78.0%)	0.0%	7.5%	0.0%	5.6%	0.6%	5.0%	0.6%	5.0%
MTD Undetermined	0.3%	--	0.1%	--	0.0%	--	0.0%	--
Median Sample Size (Empirical 95% CI)	18 (13 - 23)		22 (15 - 28)		16 (12 - 21)		16 (12 - 21)	
Median % Toxicity (Empirical 95% CI)	17 (11 - 25)		14 (8 - 23)		14 (7 - 28)		15 (7 - 29)	
% of Trials with ≥ 3 DLTs at a Dose	22.9%		30.9%		17.9%		18.2%	
% of Trials CRM and CRMBSA Differed in Escalation/De-escalation Decisions	--		--		24.2%		--	

Abbreviations: CRM: Continual Reassessment Method; BSA: Body Surface Area; CRM-BSA: Continual Reassessment Method where instead of assigned doses, patient specific BSA adjusted doses were used in the calculations; MTD: Maximum Tolerated Dose; DLT: Dose Limiting Toxicity; %Chosen: Percentage among the 1,000 simulated trials the given dose was chosen as the MTD; Exp%: Experimentation Percentage i.e. percentage of patients treated at each dose level averaged across all 1,000 simulated trials; Median Sample Size: Median sample size needed to reach the MTD across all 1000 simulated trials; Median % Toxicity: median across all simulated trials of the toxicity percentage observed per trial where percent toxicity is defined as number of patients with DLTs divided by the total sample size.

Table VI

Comparing Operating Characteristics of the Traditional Method, the Rolling-6 Design and the Continual Reassessment Method Results Based on 1,000 Simulated Trials for Distribution 6

DISTRIBUTION 6		Traditional Method		Rolling 6		CRM		CRM BSA	
Dosage (Toxicity Probability)		% Chosen	Exp%#	% Chosen	Exp%	% Chosen	Exp%	% Chosen	Exp%
140 (12.0%)		34.0%	--	35.2%	--	41.0%	--	41.9%	--
180 (20.0%) - Starting dose		45.0%	64.9%	44.8%	67.4%	27.7%	63.6%	24.9%	63.3%
230 (35.0%)		19.9%	28.1%	19.1%	26.9%	28.8%	28.7%	30.3%	28.8%
300 (60.0%)		1.1%	6.7%	0.9%	5.5%	2.5%	7.3%	2.9%	7.5%
400 (87.0%)		0.0%	0.3%	0.0%	0.2%	0.0%	0.4%	0.0%	0.4%
520 (98.0%)		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
MTD Undetermined		0.0%	--	0.0%	--	0.0%	--	0.0%	--
Median Sample Size (Empirical 95% CI)		9 (2 - 17)		10 (3 - 18)		9 (2 - 17)		9 (2 - 17)	
Median %Toxicity (Empirical 95% CI)		33 (17 - 100)		33 (17 - 100)		33 (17 - 100)		33 (17 - 100)	
% of Trials with >= 3 DLTs at a Dose		13.2%		20.0%		14.6%		16.7%	
% of Trials CRM and CRMBSA Differed in Escalation/De-escalation Decisions		--		--		21.9%		--	

Abbreviations: CRM: Continual Reassessment Method; BSA: Body Surface Area; CRM-BSA: Continual Reassessment Method where instead of assigned doses, patient specific BSA adjusted doses were used in the calculations; MTD: Maximum Tolerated Dose; DLT: Dose Limiting Toxicity; %Chosen: Percentage among the 1,000 simulated trials the given dose was chosen as the MTD; Exp%: Experimentation Percentage i.e. percentage of patients treated at each dose level averaged across all 1,000 simulated trials; Median Sample Size: Median sample size needed to reach the MTD across all 1000 simulated trials; Median % Toxicity: median across all simulated trials of the toxicity percentage observed per trial where percent toxicity is defined as number of patients with DLTs divided by the total sample size.

Table VII

Comparison of Trial Durations Across the Four Algorithms and the Six Distributions based on three Different Accrual Rates
The values in the table provide the median trial duration (95% Empirical CI on the trial duration) in days observed across the 1000 simulations.

Accrual Rate		Traditional Method	Rolling-6	CRM	CRM-BSA
10 patients/year	Dist 1	692 (114-1329)	631 (107-1165)	572 (60-1074)	574 (60-1072)
	Dist 2	886 (144-1494)	828 (123-1395)	668 (64-1194)	674 (65-1181)
	Dist 3	982 (302-1512)	963 (301-1498)	776 (89-1312)	779 (89-1282)
	Dist 4	224 (52-819)	207 (52-676)	172 (43-719)	172 (43-719)
	Dist 5	896 (564-1387)	864 (526-1326)	766 (450-1192)	764 (441-1183)
	Dist 6	450 (76-1042)	404 (77-908)	381 (55-906)	377 (54-906)
18 patients/year	Dist 1	461 (85-847)	401 (79-768)	401 (57-695)	406 (57-706)
	Dist 2	593 (109-993)	541 (84-913)	482 (58-806)	485 (58-807)
	Dist 3	658 (213-994)	650 (178-950)	540 (72-877)	539 (72-888)
	Dist 4	168 (44-543)	129 (47-442)	133 (40-503)	133 (40-507)
	Dist 5	605 (395-874)	571 (365-842)	523 (336-788)	523 (340-766)
	Dist 6	308 (59-633)	252 (58-553)	273 (47-610)	269 (47-608)
36 patients/year	Dist 1	327 (66-568)	258 (64-486)	294 (49-509)	296 (49-505)
	Dist 2	413 (94-665)	348 (71-599)	345 (55-574)	346 (53-581)
	Dist 3	466 (146-701)	447 (114-613)	399 (100-625)	402 (93-625)
	Dist 4	118 (39-378)	89 (43-254)	107 (39-343)	106 (39-347)
	Dist 5	422 (282-610)	393 (253-556)	382 (248-540)	382 (251-543)
	Dist 6	228 (47-431)	173 (55-360)	212 (44-403)	213 (44-404)

Abbreviations: CRM: Continual Reassessment Method; BSA: Body Surface Area; CRM-BSA: Continual Reassessment Method where instead of assigned doses, patient specific BSA adjusted doses were used in the calculations;

Table VIII
Concordance in the Maximum Tolerated Dose Estimates Across the Four Algorithms and the Six Distributions

	All 4 designs agree	TM, Rolling-6 and CRM agree, CRM-BSA disagrees	TM and Rolling-6 disagree	TM and CRM disagree	TM and CRM-BSA disagree	Rolling-6 and CRM disagree	Rolling-6 and CRM-BSA disagree	CRM and CRM-BSA disagree
Dist 1	32.0%	3.90%	41.2%	47.2%	49.0%	54.1%	54.5%	13.2%
Dist 2	21.7%	2.90%	52.1%	57.8%	59.7%	63.1%	64.5%	13.8%
Dist 3	17.3%	1.10%	40.9%	68.9%	69.4%	74.7%	74.9%	9.90%
Dist 4	70.2%	2.10%	19.3%	16.0%	17.0%	23.1%	23.7%	6.40%
Dist 5	34.3%	3.90%	35.7%	44.1%	45.3%	47.7%	49.7%	9.00%
Dist 6	45.0%	4.90%	29.7%	35.7%	38.7%	41.7%	45.9%	11.7%

Abbreviations: TM: traditional Method; CRM: Continual Reassessment Method; BSA: Body Surface Area; CRM-BSA: Continual Reassessment Method where instead of assigned doses, patient specific BSA adjusted doses were used in the calculations;

Spearman's Rank Correlations of MTD Estimates Based on the 1000 Simulated Trials Across the Four Algorithms and the Six Distributions

Table IX

	CRM vs. CRM-BSA	CRM vs. Rolling 6	CRM vs. TM	CRM-BSA vs. Rolling 6	CRM-BSA vs. TM	Rolling 6 vs. TM
Dist 1	0.9063	0.4561	0.5530	0.4770	0.5654	0.5346
Dist 2	0.9291	0.3688	0.4929	0.3704	0.4933	0.4183
Dist 3	0.9552	0.2229	0.3363	0.222	0.3449	0.4161
Dist 4	0.8358	0.4351	0.6852	0.3982	0.6562	0.5443
Dist 5	0.8376	0.2116	0.3062	0.1807	0.2793	0.3086
Dist 6	0.9185	0.4998	0.6019	0.4672	0.5882	0.6080

Abbreviations: TM: traditional Method; CRM: Continual Reassessment Method; BSA: Body Surface Area; CRM-BSA: Continual Reassessment Method where instead of assigned doses, patient specific BSA adjusted doses were used in the calculations;

Comparison of Doses Declared as the MTD Based on the Traditional Method, the Continual Reassessment Method and the Rolling -6 Design across the 1000 Simulated Trials for Distributions 3 and 5 - Entries in Cells Represent Counts

TM (rows) vs. Rolling 6 (column s)									TM(rows) vs. CRM (columns)									CRM (rows) vs. Rolling 6 (columns)									
		140	180	230	300	400	MTD>	All			140	180	230	300	400	520	MTD>	All			140	180	230	300	400	MTD>	All
Dist 3	140	15	0	0	3	5	6	29	140	24	0	0	2	2	1	0	29	140	12	1	3	5	13	22	56		
	180	0	24	3	1	5	8	41	180	0	19	10	1	8	3	0	41	180	0	12	2	2	3	10	29		
	230	0	3	32	6	9	18	68	230	0	0	23	22	8	10	5	68	230	0	8	14	7	9	19	57		
	300	3	2	5	54	19	40	123	300	5	1	4	32	47	22	12	123	300	4	1	24	28	25	62	144		
	400	4	6	20	16	120	68	234	400	6	1	7	23	116	71	10	234	400	9	14	21	42	105	144	335		
	MTD>	9	9	24	50	67	346	505	MTD>	21	8	13	64	154	148	97	505	MTD>	4	8	14	27	55	147	255		
All		31	44	84	130	225	486	1000	All	56	29	57	144	335	255	124	1000	MTD>	2	0	6	19	15	82	124		
																		All	31	44	84	130	225	486	1000		
Dist 5	140	1	0	0	0	0	0	1	140	1	0	0	0	0	0	1	140	1	0	1	4	1	0	0	7		
	180	0	2	0	0	0	0	2	180	0	0	2	0	0	0	2	180	0	0	0	1	0	0	1			
	230	0	0	25	19	15	1	60	230	0	0	7	33	20	0	60	230	0	3	4	10	6	0	23			
	300	0	3	12	394	145	0	554	300	4	1	9	256	282	2	554	300	0	3	22	244	96	0	365			
	400	0	3	18	138	221	0	380	400	2	0	5	74	295	4	380	400	0	2	29	292	274	1	598			
	MTD>	0	0	1	2	0	0	3	MTD>	0	0	0	2	1	0	3	MTD>	0	0	0	2	4	0	6			
All		1	8	56	553	381	1	1000	All	7	1	23	365	598	6	1000	All	1	8	56	553	381	1	1000			

Abbreviations: TM: traditional Method; CRM: Continual Reassessment Method where instead of assigned doses, patient specific BSA adjusted doses were used in the calculations; CRM-BSA: Continual Reassessment Method where instead of assigned doses, patient specific BSA