VOLUME 31 · NUMBER 14 · MAY 10 2013

JOURNAL OF CLINICAL ONCOLOGY

STATISTICS IN ONCOLOGY

Modified Toxicity Probability Interval Design: A Safer and More Reliable Method Than the 3 + 3 Design for Practical Phase I Trials

Yuan Ji and Sue-Jane Wang

Yuan Ji, Center for Clinical and Research Informatics, NorthShore University HealthSystem, Evanston, IL; and Sue-Jane Wang, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD.

Published online ahead of print at www.jco.org on April 8, 2013.

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Corresponding author: Yuan Ji, PhD, Center for Clinical and Research Informatics, NorthShore University HealthSystem, 1001 University Place, Evanston, IL 60201; e-mail: yji@northshore.org.

© 2013 by American Society of Clinical Oncology

0732-183X/13/3114w-1785w/$20.00

DOI: 10.1200/JCO.2012.45.7903

ABSTRACT

The 3 + 3 design is the most common choice among clinicians for phase I dose-escalation oncology trials. In recent reviews, more than 95% of phase I trials have been based on the 3 + 3 design. Given that it is intuitive and its implementation does not require a computer program, clinicians can conduct 3 + 3 dose escalations in practice with virtually no logistic cost, and trial protocols based on the 3 + 3 design pass institutional review board and biostatistics reviews quickly. However, the performance of the 3 + 3 design has rarely been compared with model-based designs in simulation studies with matched sample sizes. In the vast majority of statistical literature, the 3 + 3 design has been shown to be inferior in identifying true maximum-tolerated doses (MTDs), although the sample size required by the 3 + 3 design is often orders-of-magnitude smaller than model-based designs. In this article, through comparative simulation studies with matched sample sizes, we demonstrate that the 3 + 3 design has higher risks of exposing patients to toxic doses above the MTD than the modified toxicity probability interval (mTPI) design, a newly developed adaptive method. In addition, compared with the mTPI design, the 3 + 3 design does not yield higher probabilities in identifying the correct MTD, even when the sample size is matched. Given that the mTPI design is equally transparent, costless to implement with free software, and more flexible in practical situations, we highly encourage its adoption in early dose-escalation studies whenever the 3 + 3 design is also considered. We provide free software to allow direct comparisons of the 3 + 3 design with other model-based designs in simulation studies with matched sample sizes.

J Clin Oncol 31:1785-1791. © 2013 by American Society of Clinical Oncology

INTRODUCTION

A phase I oncology trial aims to determine the maximum-tolerated dose (MTD) of a drug—the highest dose with a toxicity rate close to a prespecified target level, p_T. The 3 + 3 design[1,2] is the leading method for phase I dose-escalation trials in oncology; over 95% of published phase I trials in the past two decades have been based on this design.[3-5] The popularity of the 3 + 3 design is striking, because numerous model-based dose-escalation methods have been developed by biostatisticians during the same time period, and almost all of the new methods seem to exhibit better performance than the 3 + 3 design.[6-9]

The main reason for the popularity of the 3 + 3 design results from its simplicity (no computer program is required), transparency, and costless implementation in practice. In contrast, most model-based designs often require considerable logistic burden and complexity to implement. Even if the practical burden could be overcome, protocols based on model-based designs are often subject to more thorough reviews by institutional review boards or among biostatisticians, because these new designs require computer-generated operating characteristics. Conversely, if a protocol is based on the 3 + 3 design, this requirement disappears, because the design has been widely used. As a result, despite the acceleration in the development of adaptive model-based designs, the lower standard in the review process and cost-free implementation in practice makes the 3 + 3 design persistently popular among physicians. Setting aside logistic issues, we seek to assess the advantages of model-based designs over the 3 + 3 design. In reviewing the statistical literature on phase I adaptive designs compared with the 3 + 3 design, we found that most studies did not match sample sizes across the designs. For example, Ji et al[10] showed that 3 + 3 design exhibited a smaller average sample size in computer simulations than model-based designs, and consequently, the 3 + 3 design also identified the true MTD a smaller percentage of the time. Because sample size was

Downloaded from jco.ascopubs.org on May 9, 2013. For personal use only. No other uses without permission.
Copyright © 2013 American Society of Clinical Oncology. All rights reserved.

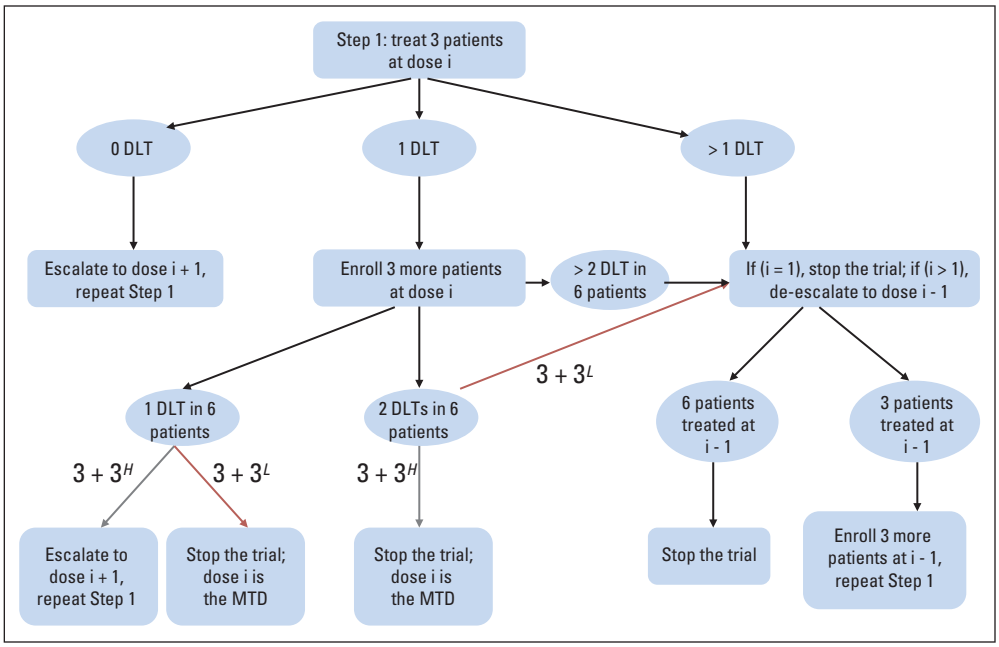

Fig 1. Schema of the enhanced 3 + 3 design. The two versions of 3 + 3^L and 3 + 3^H represent cases where the maximum-tolerated dose (MTD) is defined as the highest dose at which no more than one and two dose-limiting toxicities (DLTs) are observed among six patients, respectively.

not matched in the comparison, it is difficult to determine the reason for the reduced percentages under the 3 + 3 design. More importantly, because phase I trials focus on patient safety, comparisons without matching sample sizes cannot provide accurate assessments of the safety characteristics of designs. In fact, designs resulting in larger sample sizes are usually safer, because patients enrolled onto trials during the latter stages will be better protected as a result of more-precise statistical inference.

In this article, we report a comprehensive simulation study to evaluate the operating characteristics of the 3 + 3 design and a newly developed adaptive design known as the modified toxicity probability interval (mTPI) method.[5,10] In doing so, we match the sample size of both designs. Compared with the 3 + 3 design, the mTPI design is equally as simple, transparent, and costless to implement. In other words, the logistic burden of both designs is negligible, allowing us to focus on the simulation performance. Although the mTPI design is a novel approach, it has already received attention from both research and industry entities.[11,12] For example, almost all phase I oncology trials conducted at Merck (Whitehouse Station, NJ) in the past 2 years have been based on the mTPI design or its variations (K. Anderson, personal communication, June 2012). Recently, phase I trials based on the mTPI design have been published in the clinical community.[13,14] Considering the short time period since the publication of the mTPI design, this popularity is encouraging.

The 3 + 3 design consists of a set of deterministic rules that dictate dose-escalation decisions based on observed patient outcomes (Fig 1). For example, if zero, one, or > one toxicities are observed in three treated patients, the 3 + 3 design will recommend escalating the dose level, continuing at the same dose level, or de-escalating the dose level, respectively (as described by Ibrahim et al[15] and Strumberg et al[16]). The mTPI design uses a Bayesian statistical framework and a beta/binomial hierarchic model to compute the posterior probabilities of three intervals that reflect the relative distance between the toxicity rate of each dose level to the target probability, p_T. That is, the mTPI design replaces the 3 + 3 rules that are based on events such as zero of three, one of three, two of three, and three of three with a model-based inference on the toxicity probability intervals. For example, for each of the three p_T values, the mTPI design can be applied to three corresponding dosing intervals listed in Table 1 (ie, underdosing, proper dosing, and overdosing). Ji et al[10] describe the proper dosing interval as the equivalence interval (EI).

Dose-escalation decisions are easily made based on the three dosing intervals. If the toxicity rate of the currently used dose level is within the underdosing interval, the mTPI design will recommend escalating the dose level; in the case of proper dosing, the design will recommend continuing at the current dose; for overdosing, the mTPI design will recommend de-escalating the dose level. These rules are conceptually similar to those used in the 3 + 3 design; however, the decisions of the mTPI design are based on posterior probabilities calculated under a coherent probability model. As a model-based design, the mTPI design automatically and appropriately tailors dose-escalation decisions for different trials with different toxicity parameters. More importantly, all the dose-escalation decisions for a given trial can be precalculated under the mTPI design and presented in a two-way table (Fig 2). Once the trial starts, clinicians can easily monitor the trial and select the appropriate doses following the precalculated table. A hypothetical example is presented in Table 2. The simplicity and transparency of the mTPI design make it a strong

Table 1. Examples of Toxicity Probability Intervals

p_T	Underdosing	Proper Dosing	Overdosing
0.10	0 to 0.04	0.05 to 0.15	0.16 to 1.00
0.20	0 to 0.14	0.15 to 0.25	0.26 to 1.00
0.30	0 to 0.24	0.25 to 0.35	0.36 to 1.00

 © 2013 by American Society of Clinical Oncology

Downloaded from jco.ascopubs.org on May 9, 2013. For personal use only. No other uses without permission.
Copyright © 2013 American Society of Clinical Oncology. All rights reserved.

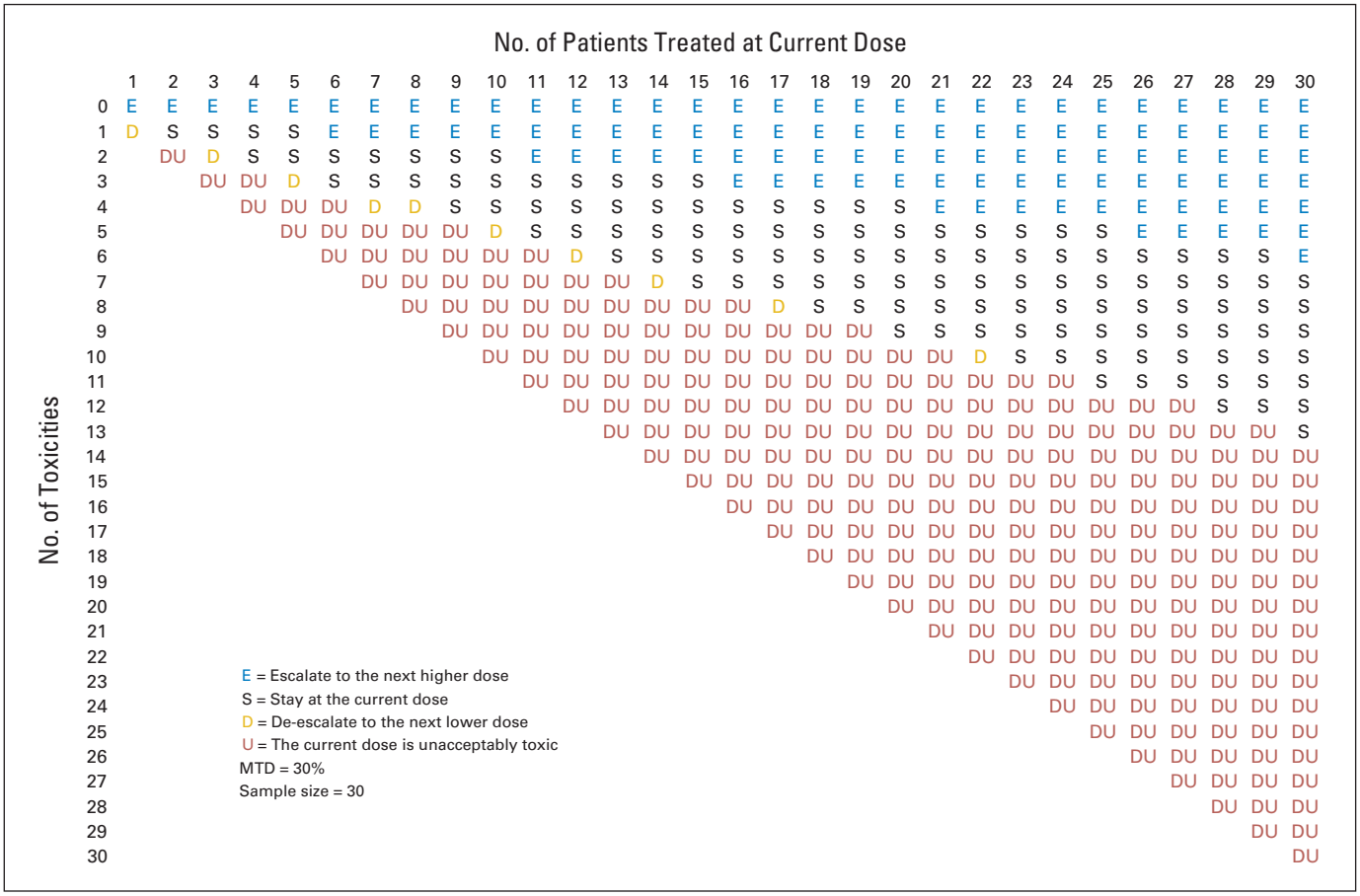

Fig 2. Dose-finding spreadsheet of the modified toxicity probability interval (mTPI) method. The spreadsheet is generated based on a beta/binomial model and precalculated before a trial starts. The letters in different colors are computed based on the decision rules under the mTPI method and represent different dose-finding actions. In addition to actions de-escalate the dose (D), stay at the same dose (S), and escalate the dose (E), the table includes action unacceptable toxicity (U), which is defined as the execution of the dose-exclusion rule in mTPI. MTD, maximum-tolerated dose.

candidate for the model-based counterpart of the 3 + 3 design in practice. Important details of the 3 + 3 and mTPI designs are presented here and in the Appendix (online only). We will show surprising and important findings and make a recommendation for the use of the mTPI design in future phase I trials. We provide free software at http://www.northshore.org/research/investigators/Yuan-Ji-PhD/ that allows users to examine the 3 + 3 and mTPI designs side by side based on simulated trials.

Table 2. Hypothetical Trial Based on the mTPI Design

	Dose Level										
	1		2		3*		4		5		
Cohort	No. of DLTs	No. of Patients	No. of DLTs	No. of Patients	No. of DLTs	No. of Patients	No. of DLTs	No. of Patients	No. of DLTs	No. of Patients	Decision
1	**0**	**3**	—	—	—	—	—	—	—	—	E
2	0	3	**0**	**3**	—	—	—	—	—	—	E
3	0	3	0	3	**1**	**3**	—	—	—	—	S
4	0	3	0	3	**1 + 0**†	**3 + 3**†	—	—	—	—	E
5	0	3	0	3	1	6	**2**	**3**	—	—	D
6	0	3	0	3	**1 + 1**	**6 + 3**	2	3	—	—	S
Total	0	3	0	3	2	9	2	3	—	—	

NOTE. Assumes $p^T = 0.3$. Bold font indicates current dose level being used.
Abbreviations: D, de-escalation; DLT, dose-limiting toxicity; E, escalation; MTD, maximum-tolerated dose; mTPI, modified toxicity probability interval; S, stay at same dose.
*Final estimated MTD.
†"1 + 0" DLTs among "3 + 3" patients means that there were two cohorts for that dose, and results are one DLT among three patients for the first cohort and zero DLTs among three patients for the second cohort. Therefore, the overall result is one DLT among six patients, used to reach the decision E.

© 2013 by American Society of Clinical Oncology

Downloaded from jco.ascopubs.org on May 9, 2013. For personal use only. No other uses without permission.
Copyright © 2013 American Society of Clinical Oncology. All rights reserved.

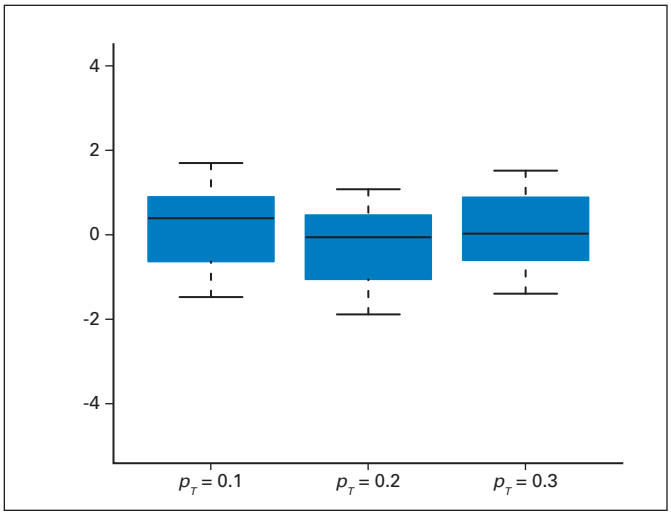

Fig 3. Difference in the average sample size per trial between the 3 + 3 and modified toxicity probability interval designs. Each boxplot summarizes the differences for 14 scenarios for a given target toxicity value p_T.

3 + 3 AND mTPI DESIGNS

Algorithmic and Model-Based Designs

Algorithmic designs such as the 3 + 3 design are based on predetermined and fixed rules. They rely on heuristic reasoning and argument rather than rigorous mathematic modeling. Typically, an algorithmic design has limited generality, because the heuristic argument is often based on strong assumptions. For example, the 3 + 3 design assumes that the target toxicity probability p_T takes values close to either one of six or one of three. In contrast, model-based designs such as the mTPI design use quantitative models to describe and account for the uncertainties in the observed data. The models are usually general, allowing the designs to be applied in a variety of cases. More importantly, medical decisions are based on principled mathematic inference, reducing the level of subjectivity in the decision-making process.

3 + 3 Design

We provide a comprehensive and improved algorithm for implementing the 3 + 3 design in practice (Fig 1; full algorithm provided in the Appendix, online only). With this enhanced version of the 3 + 3 design, many important steps (eg, those described in Lin and Shih[17]) are often ignored in real-life trials. The algorithm indicates that it is necessary to consider at least two variations of the 3 + 3 design, depending on the specifics of the trial. The first variation—the 3 + 3^H design—targets the MTD with ≤ two toxicities among six patients ($p_T \leq$ two of six); the second variation—the 3 + 3^L design—targets the MTD with ≤ one toxicity among six patients ($p_T \leq$ one of six). Both MTD definitions are commonly used.

mTPI Design

The full mTPI design is provided in the Appendix (online only) and also described in detail by Ji et al.[10] A trial based on the mTPI design can be monitored using a spreadsheet in the form shown in Figure 2. The spreadsheet consists of all the dose-escalation rules for a

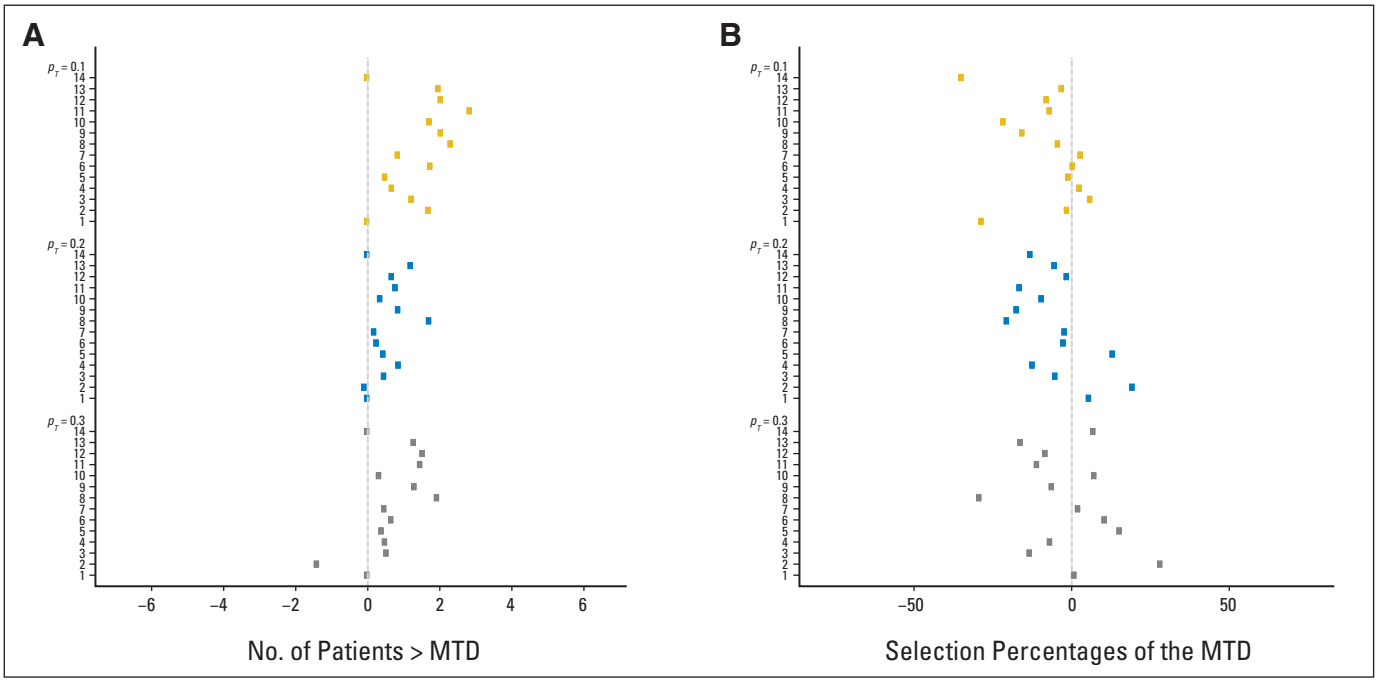

Fig 4. Comparison between the 3 + 3 and modified toxicity probability interval (mTPI) designs based on matched sample sizes. (A) Points to the right of 0 correspond to the 3 + 3 design being less safe than the mTPI design; (B) points to the left of 0 correspond to the 3 + 3 design being less reliable than the mTPI design. (A) Differences in the numbers of patients treated at doses above the maximum-tolerated dose (MTD; $n_{>\ MTD}$; ie, values of [$n_{>\ MTD}$ 3 + 3 − $n_{>\ MTD}$ mTPI] for all 42 scenarios); (B) differences in the selection percentages of the true MTD ($\%Sel_{MTD}$; ie, values of [$\%Sel_{MTD}$ 3 + 3 − $\%Sel_{MTD}$ mTPI] for all 42 scenarios). Colors indicate results corresponding to the three different p_T values.

 © 2013 by American Society of Clinical Oncology

Downloaded from jco.ascopubs.org on May 9, 2013. For personal use only. No other uses without permission.
Copyright © 2013 American Society of Clinical Oncology. All rights reserved.

trial and can be generated using an Excel macro (https://biostatistics.mdanderson.org/SoftwareDownload/SingleSoftware.aspx?Software_Id=72). In Figure 2, the columns are the numbers of patients treated so far at the current dose level, and the rows are the corresponding numbers of patients experiencing toxicity. The entries of the table are dose-finding decisions—E, S, and D—representing escalating the dose, staying at the same dose, and de-escalating the dose. In addition, decision U means that the current dose level is unacceptable because of high toxicity and should be excluded from the trial. For example, when one of three patients experiences toxicity, the decision can be located at row 1 and column 3, which is S—to stay at the current dose level. This means that the next cohort of patients will be treated at the same dose level currently being used. If zero of three patients experiences toxicity, the decision is at row 0 and column 3, which is E—to escalate. This means that the next cohort of patients will be treated at the next-higher dose level. If three of three patients experiences toxicity, the decision is DU—to de-escalate to the next-lower dose level and exclude the current dose from the trial, because the high toxicity level is unacceptable. Note that these decisions are also in agreement with the 3 + 3 method. However, when different target p_T values are used, the decisions in the mTPI table will automatically adjust to the values, which is an advantage of the model-based design over algorithmic approaches.

COMPARISON OF THE 3 + 3 AND mTPI DESIGNS

Simulation Setup

We performed computer simulations of phase I trials based on the 3 + 3 and mTPI designs and compared their operating characteristics. A total of 2,000 trials was simulated for each dose-toxicity scenario using each method.

Values of p_T. In practice, the target p_T values are rarely larger than 30%, because this implies unnecessary exposure of patients to doses with high toxicity. Here we study three choices of p_T: 0.1, 0.2, and 0.3 (ie, the target toxicity rates of the MTD in our simulated trials are 10%, 20%, and 30%). In practice, the specification of p_T greatly depends on the disease and the definition of toxicity. For example, for severe diseases and high-risk procedures such as stem-cell transplantation for leukemia, a higher p_T value (more tolerance to toxicity) is acceptable, given the high toxicity rate associated with the procedure. In other words, high-grade toxicity events can be tolerated to avoid death. In contrast, in treatment for diseases such as sarcomas or prostate cancer, a lower p_T value is specified to impose stricter control of toxicity.

Clinical scenarios. We considered six doses in the simulated trials, with 14 scenarios for each of the three target p_T values, resulting in a total of 42 scenarios (Appendix Table A1 and Fig A1, online only). Details are provided in the Appendix (online only).

Matching sample size. An important feature of our comparison is that we attempted to match the average sample size of the 3 + 3 and mTPI designs for each of the clinical scenarios used in the simulation study. To achieve this, for each scenario, we first applied the 3 + 3 design to 2,000 computer-simulated trials and obtained the mean of the 2,000 sample sizes. We then applied the mTPI design, for which we needed to specify the maximum sample size. The mTPI design stops a trial when the total number of patients enrolled is equal to or larger than the maximum sample size. We calibrated the maximum sample sizes of the mTPI design so that the average sample sizes over simulated trials under both designs were similar across all the scenarios. Figure 3 shows the differences of the average sample sizes (> 2,000 simulated trials) between the 3+ 3 and mTPI designs. The two designs exhibited comparable sample sizes overall. Our calibration of the mTPI design involved only varying the maximum sample size while keeping all the other design features unchanged.

Variations of the 3 + 3 design. To account for different target p_T values, we used one of the two 3 + 3 variations (ie, 3 + 3^L and 3 + 3^H designs). These two variations are illustrated in Figure 1 and are also fully described in the Appendix (online only). We used the 3 + 3^L design for trials with $p_T = 0.1$ or $p_T = 0.2$ and the 3 + 3^H variation for trials with $p_T = 0.3$.

Performance Evaluation

Summarizing results from 42 scenarios over three different p_T values can be subjective, depending on the criteria used in the comparison. Because the average sample sizes between the two methods were roughly matched, we focused our comparison on two summary statistics:

- $n_{>\,MTD}$ = mean number of patients treated above the true MTD
- $\%Sel_{MTD}$ = percentage of simulated trials selecting the true MTD

The value $n_{>\,MTD}$ directly evaluates the safety of the design, because under a matched sample size, a smaller $n_{>\,MTD}$ value implies less toxicity. The value of $\%Sel_{MTD}$ measures the reliability of the design (Appendix Fig A2, online only). To calculate $\%Sel_{MTD}$, we need

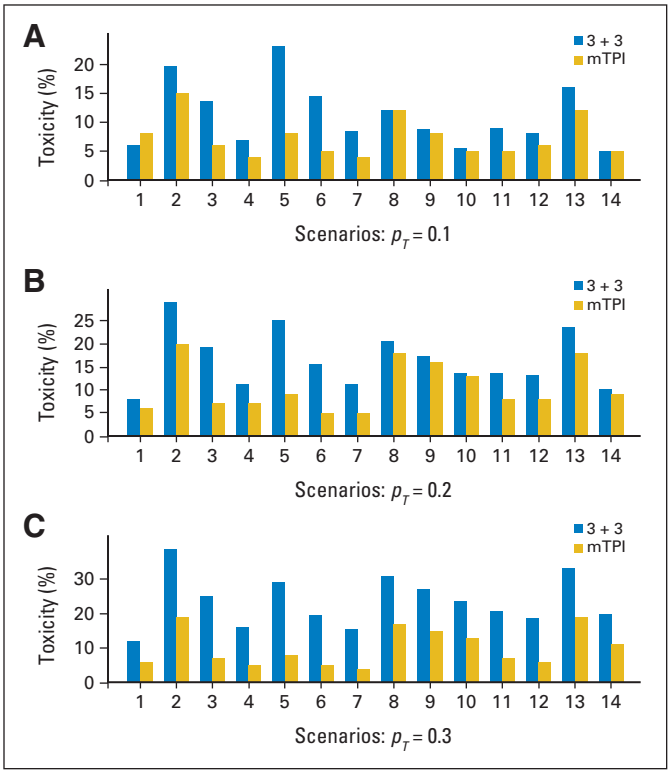

Fig 5. Overall toxicity percentages for the 3 + 3 and modified toxicity probability interval (mTPI) designs across all simulated trials. Scenarios in which (A) $p_T = 0.1$, (B) $p_T = 0.2$, and (C) $p_T = 0.3$. The mTPI design represents a safer and more reliable method than the 3 + 3 design for practical phase I trials.

© 2013 by American Society of Clinical Oncology

Downloaded from jco.ascopubs.org on May 9, 2013. For personal use only. No other uses without permission.
Copyright © 2013 American Society of Clinical Oncology. All rights reserved.

to decide which dose will be considered as the MTD for each scenario. Note that in many scenarios (eg, scenarios five to 11), none of the doses have toxicity probabilities exactly equal to p_T. However, some dose-toxicity probabilities are close enough to be declared the MTDs. This is consistent with the real-world belief that the exact value p_T is almost never achieved for a dose. To this end, we assumed that any doses with true toxicity probabilities in the EI $(p_T - \varepsilon_1, p_T + \varepsilon_2)$ would be considered as the MTDs. In our simulation, we set $\varepsilon_1 = \varepsilon_2 = 0.05$ so that the EI was (0.05, 0.15) when $p_T = 0.1$, (0.15, 0.25) when $p_T = 0.2$, and (0.25, 0.35) when $p_T = 0.3$. If no dose was in the EI, the highest dose with true toxicity probability $< p_T$ was considered the MTD. If the MTD could not be identified (eg, if all the doses have toxicity probabilities $> p_T$), the correct decision was not to select any dose. In that case, the percentage of "none" selection (ie, no dose selected as the MTD) was used to compare the designs.

Main Results

Figure 4 summarizes the comparisons of the 3 + 3 and mTPI designs in terms of safety $n_{>MTD}$ and reliability $\%Sel_{MTD}$. In Figure 4A, points to the right of 0 correspond to the 3 + 3 design being less safe than the mTPI design; in Figure 4B, points to the left of 0 correspond to the 3 + 3 design being less reliable than the mTPI design. The 3 + 3 design had lower $n_{>MTD}$ values for two scenarios, higher $n_{>MTD}$ for 34 scenarios, and the same $n_{>MTD}$ for six scenarios. Therefore, in 40 of 42 times, the mTPI design treated fewer or the same number of patients at doses higher than the MTD compared with the 3 + 3 design. Additionally, Figures 5A to 5C examines the overall toxicity percentage, defined as:

A/B × 100%, where A = total number of toxicities over all simulated trials, and B = total number of patients treated over all simulated trials

The 3 + 3 design exhibited a lower overall toxicity percentage than the mTPI design in only one of 42 scenarios.

Figure 4B compares the reliability measure $\%Sel_{MTD}$ between the two designs. In 10 of 42 scenarios, the 3 + 3 design had a higher selection percentage of the true MTD than the mTPI design. Among these scenarios, the 3 + 3 design selected the MTD up to approximately 25% more often than the mTPI design (scenario two for $p_T = 0.3$). In the remaining 32 scenarios, the mTPI design selected the MTD more often than the 3 + 3 design, up to > 40% (scenario 14 for $p_T = 0.1$). A closer examination revealed that the 3 + 3 design had higher $\%Sel_{MTD}$ values only in scenarios where none of the doses had toxicity probabilities close to p_T or where the MTD was at the lower or higher end of the dosing set. We performed additional simulations and confirmed this finding. We found that the 3 + 3 design was a better method when none of the investigational doses was close to the true MTD. This advantage seems to be of limited utility in practice, because doses are usually chosen based on scientific and historical data, and it is anticipated that some of them are close to the MTD, not the opposite.

Summarizing the two plots in Figure 4 and considering that first, the overall sample sizes between the two designs are roughly matched for all the scenarios, and second, the 42 scenarios were constructed to cover a wide range of practical dose-response shapes, we conclude that the 3 + 3 design is more likely to treat patients at toxic doses above the MTD and less likely to identify the true MTD than the mTPI design. In other words, the 3 + 3 design is less safe and less reliable than the mTPI design.

DISCUSSION

The mTPI design has all the attractive properties that the 3 + 3 design enjoys with regard to practical considerations and implementation. Furthermore, compared with the 3 + 3 design, the mTPI design is safer, because it treats fewer patients at doses above the MTD and generally yields higher probabilities in identifying the true MTD.

In practice, a single value n must be provided as the maximum sample size for the mTPI design in any dose-escalation study. In implementing the mTPI design, we recommend a sample size of $n = k \times (d + 1)$ to ensure that the design will reach the highest dose if needed and an additional cohort is available for use. Here k is the cohort size, and d is the number of doses. For example, when $k = 3$ and $d = 5$, $n = 18$.

It is commonly accepted that phase I trials should be small. However, the consequences of this methodology are poorly addressed in the literature. Small phase I trials often provide suboptimal recommended doses for phase II, resulting in either low efficacy or high toxicity if the recommended doses are too low or too high, respectively. More discussion and investigation of appropriate sample sizes for phase I trials are needed. For example, a streamlined and seamless phase I/II design may result in higher power in the identification of safe and effective doses[18] because of increased sample sizes from the seamless features.

We note that comparisons between the continual reassessment method (CRM) and the 3+ 3 design have been investigated by various authors[19-21] and thus are not included in this article. One downside of the CRM is the lack of an easy means of implementation in practice. We have included the CRM design in our software to allow interested users to examine all three designs together (ie, the 3 + 3, CRM, and mTPI designs).

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The author(s) indicated no potential conflicts of interest.

AUTHOR CONTRIBUTIONS

Manuscript writing: All authors
Final approval of manuscript: All authors

REFERENCES

1. Storer BE: An evaluation of phase I clinical trials designs in the continuous dose-response setting. Stat Med 20:2399-2408, 2001
2. Storer BE: Design and analysis of phase I clinical trials. Biometrics 45:925-937, 1989
3. Rogatko A, Schoeneck D, Jonas W, et al: Translation of innovative designs into phase I trials. J Clin Oncol 25:4982-4986, 2007
4. Le Tourneau C, Lee JJ, Siu LL: Dose escalation methods in phase I cancer clinical trials. J Natl Cancer Inst 101:708-720, 2009
5. Ji Y, Li Y, Nebiyou Bekele B: Dose-finding in phase I clinical trials based on toxicity probability intervals. Clin Trials 4:235-244, 2007
6. O'Quigley J, Pepe M, Fisher L: Continual reassessment method: A practical design for phase I clinical trials in cancer. Biometrics 46:33-48, 1990
7. Berry SM, Carlin BP, Lee JJ, et al: Bayesian Adaptive Methods for Clinical Trials. Boca Raton, FL, CRC Press, 2011
8. Cheung YK: Dose Finding by the Continual Reassessment Method. Boca Raton, FL, CRC Press, 2011
9. Neuenschwander B, Branson M, Gsponer T: Critical aspects of the Bayesian approach to phase I cancer trials. Stat Med 27:2420-2439, 2008
10. Ji Y, Liu P, Li Y, et al: A modified toxicity probability interval method for dose-finding trials. Clin Trials 7:653-663, 2010

 © 2013 by American Society of Clinical Oncology

Downloaded from jco.ascopubs.org on May 9, 2013. For personal use only. No other uses without permission.
Copyright © 2013 American Society of Clinical Oncology. All rights reserved.

11. Blanchard MS, Longmate JA: Toxicity equivalence range design (TEQR): A practical phase I design. Contemp Clin Trials 32:114-121, 2011

12. Hather GJ, Mackey H: Some notable properties of the standard oncology phase I design. J Biopharm Stat 19:543-555, 2009

13. Fanale M, Fayad L, Pro B, et al: Phase I study of bortezomib plus ICE (BICE) for the treatment of relapsed/refractory Hodgkin lymphoma. Br J Haematol 154:284-286, 2011

14. Yap TA, Yan L, Patnaik A, et al: First-in-man clinical trial of the oral pan-AKT inhibitor MK-2206 in patients with advanced solid tumors. J Clin Oncol 29:4688-4695, 2011

15. Ibrahim NK, Desai N, Legha S, et al: Phase I and pharmacokinetic study of ABI-007, a cremophor-free, protein-stablized, nanoparticle formulation of paclitaxel. Clin Cancer Res 8:1038-1044, 2002

16. Strumberg D, Richly H, Hilger RA, et al: Phase I clinical and pharmacokinetic study of the novel Raf kinase and vascular endothelial growth factor receptor inhibitor BAY 43-9006 in patients with advanced refractor solid tumors. J Clin Oncol 23:965-972, 2005

17. Lin Y, Shih WJ: Statistical properties of the traditional algorithm-based designs for phase I cancer clinical trials. Biostatistics 2:203-215, 2001

18. Xie F, Ji Y, Tremmel L: A Bayesian adaptive design for multi-dose, randomized, placebo-controlled phase I/II trials. Contemp Clin Trials 33: 739-748, 2012

19. O'Quigley J: Another look at two phase I clinical trial designs. Stat Med 18:2683-2690, 1999

20. Korn EL, Midthune D, Chen TT, et al: A comparison of two phase I designs. Stat Med 13: 1799-1806, 1994

21. Goodman SN, Zahurak ML, Piantadosi S: Some practical improvements in the continual reassessment method for phase I studies. Stat Med 14:1149-1161, 1995

Shape Your Future and Ours: Apply for ASCO's Leadership Development Program

If you completed your final subspecialty training between 2002 and 2007 and are interested in becoming a future leader in both ASCO and the practice of oncology, ASCO's Leadership Development Program is for you. If selected, you will:

- Network with and receive mentorship from ASCO leadership
- Gain exposure to ASCO committees and government research agencies
- Enhance leadership skills through interactive sessions
- Receive first-hand advocacy experience on Capitol Hill

To apply or learn about the eligibility requirements for the Leadership Development Program, please visit **asco.org/leadership**.

ASCO
American Society of Clinical Oncology

© 2013 by American Society of Clinical Oncology

Downloaded from jco.ascopubs.org on May 9, 2013. For personal use only. No other uses without permission.
Copyright © 2013 American Society of Clinical Oncology. All rights reserved.

Acknowledgment

We thank Clift Norris and John Venier at the University of Texas MD Anderson Cancer Center for providing the program for the 3 + 3 design and participating in the software development. We thank the late John Johnson of the US Food and Drug Administration for an insightful discussion on the 3 + 3 method and its historical background.

Appendix

Clinical Scenarios Used in the Simulation

In each scenario, true toxicity probabilities are specified for six doses. They are presented in Appendix Table A1 and Fig A1. These scenarios are set up to capture a wide range of dose-response shapes in practice, as discussed previously (Ji Y et al: J Biopharm Stat 22:1206-1219, 2012). Specifically, scenarios one and two represent two cases where all the doses are lower and higher than the maximum-tolerated dose (MTD), respectively; scenarios three and four cover a wide range of toxicities, and one dose is the true MTD with toxicity rate equal to p_T; scenarios five to seven also cover a wide range of toxicities, but the true MTD is bracketed by two adjacent doses; in scenarios eight to 10, toxicity does not vary much and centers around the target p_T; scenarios 11 and 12 are similar to scenarios eight to 10, except doses have a wider range of toxicity; lastly, scenarios 13 and 14 represent two rare cases in which the MTD is the lowest and highest dose, respectively.

Software

We have released free software that allows investigators to conduct the same type of simulations we have described for their own practical trials. The software allows investigators to examine the performance of the 3 + 3 design, modified toxicity probability interval (mTPI) design, and continual reassessment method based on well-matched sample sizes for a variety of scenarios. The software is available at http://www.northshore.org/research/investigators/Yuan-Ji-PhD/.

3 + 3 Design

We summarized various versions of the 3 + 3 design in practice and present a comprehensive and enhanced algorithm. The new 3 + 3 design consists of two variations and is illustrated in Fig 1. The two variations of the 3 + 3 design differ in steps 5b and 5c, when one and two toxicities are observed among six patients, respectively. The algorithm is as follows:

1. Start trial by treating patients at the initial dose.
2. Denote the dose level being used to treat patients as the current dose level. Accrue and treat three patients at the current dose level.
   2a. If the maximum number of patients (eg, because of limited resources) has been accrued, stop the trial. The MTD is inconclusive.
3. Check the number of patients at the current dose level.
   3a. If there are three patients, go to four.
   3b. If there are four patients, go to five.
4. Check the number of toxicities (among three patients) at the current dose level.
   4a. If there are zero toxicities, escalate and go to seven.
   4b. If there is one toxicity, stay at the current dose and go to two.
   4c. If there are two or three toxicities, declare that the MTD has been exceeded and go to six.
5. Check the number of toxicities (among six patients) at the current dose level.
   5a. If there are zero toxicities, stop the trial and declare that the MTD is the current dose.
   5b. If there is one toxicity, and the MTD has been exceeded, stop the trial and declare that the MTD is the current dose; otherwise, escalate and go to seven ($3 + 3^H$) or stop the trial and declare that the MTD is the current dose ($3 + 3^L$).
   5c. If there are two toxicities, stop the trial; the MTD is the current dose ($3 + 3^H$) or the MTD has been exceeded ($3 + 3^L$); go to six.
   5d. If there are three or four toxicities, declare that the MTD has been exceeded and go to six (it is impossible to have five or six toxicities among six patients at the same dose level under this 3 + 3 design).
6. The MTD has been exceeded.
   6a. If the current dose is the lowest dose, stop the trial; declare that the MTD is lower than the lowest dose level.
   6b. If the next-lower dose level has six patients, stop the trial and declare that the MTD is the next lower dose level; otherwise, the next-lower dose level has three patients; set the current dose level to be the next-lower dose level and go to two.
7. Escalate if possible.
   7a. If the current dose level is the highest dose level; stop the trial and declare that the MTD is higher than the highest dose level.
   7b. Otherwise, escalate to the next-higher dose level; go to two.

mTPI Design

To briefly describe the mTPI design, we first introduce some notations for ease of exposition. Let $p = (p_1, \ldots, p_d)$ denote the toxicity probabilities for doses $i = 1, \ldots, d$, where d is the total number of candidate doses in the trial. The observed data include n_i, the number of patients treated at dose i, and x_i, the number of patients experiencing toxicity. The likelihood function for data $\{(x_i, n_i), i = 1, \ldots, d\}$ is a

© 2013 by American Society of Clinical Oncology

Downloaded from jco.ascopubs.org on May 9, 2013. For personal use only. No other uses without permission.
Copyright © 2013 American Society of Clinical Oncology. All rights reserved.

product of binomial densities. The estimates of these toxicity probabilities p are sequentially updated and are used to decide if some of the doses studied would be close to the true MTD.

The mTPI design[10] is an extension of the toxicity probability interval method[5] and employs a simple beta-binomial hierarchic model. Decision rules are based on calculating the unit probability mass (UPM) of three intervals corresponding to underdosing, proper dosing, and overdosing in terms of toxicity. Specifically, the underdosing interval is defined as $(0, p_T - \varepsilon_1)$, the overdosing interval as $(p_T + \varepsilon_2, 1)$, and the proper dosing interval as $(p_T - \varepsilon_1, p_T + \varepsilon_2)$, where ε_1 and ε_2 are small fractions, such as 0.05, to account for the uncertainty around the true target toxicity. A sensitivity analysis reported by Ji et al[10] showed that the mTPI design is robust to the specification of ε values. In addition, ε_1 and ε_2 could take different values to reflect physician preference and the nature of the disease. For advanced diseases with few treatment options, higher toxicity rates might be considered acceptable, implying a specification of $\varepsilon_2 > \varepsilon_1$. For less-advanced diseases, the two ε values could be identical or ε_1 could be $> \varepsilon_2$. The three dosing intervals are associated with three different dose-escalation decisions. The underdosing interval corresponds to a dose escalation (E), overdosing corresponds to a dose de-escalation (D), and proper dosing corresponds to staying at the current dose (S). Given an interval and a probability distribution, the UPM of that interval is defined as the probability of the interval divided by the length of the interval. The mTPI design calculates the UPMs for the three dosing intervals, and the one with the largest UPM implies the corresponding dose-finding decision. That decision provides the dose level to be used for future patients. For example, if the underdosing interval has the largest UPM, decision E, to escalate, will be executed, and the next cohort of patients will be treated at the next-higher dose level. Ji et al show that the decision based on the UPM is optimal in that it minimizes a subsequent expected loss. Under the mTPI design, a trial is terminated when either the lowest dose is above the MTD or a prespecified maximum sample size is reached. The article by Ji et al provides more detail.

Scenarios

Appendix Table A1 lists the true toxicity probabilities for all the doses under the 42 scenarios, with p_T taking three values: 0.1, 0.2, and 0.3. They are also plotted in Figure A1.

Sample Size and Its Effects

We conducted a small simulation study to evaluate the effect of sample size on the identification of the right MTD. We ran two standard scenarios, each with 5,000 simulated trials. The MTD had a target toxicity probability of $p_T = 0.3$. The scenarios assumed the true probabilities of six doses as follows: scenario one = (0.05, 0.1, 0.3, 0.5, 0.55, 0.6) and scenario two = (0.05, 0.1, 0.15, 0.25, 0.3, 0.4). Appendix Figure A2 plots the percentage of time the true MTD was selected against the sample size allowed in the trial. In both scenarios, we identified an increasing trend; when sample size increased, the percentage of time the true MTD was selected increased as well. We observed this same trend in other scenarios (results not shown). This demonstrates the importance of considering the tradeoff between increases in sample size and acceptable percentages in selecting the right MTD.

© 2013 by American Society of Clinical Oncology

Downloaded from jco.ascopubs.org on May 9, 2013. For personal use only. No other uses without permission.
Copyright © 2013 American Society of Clinical Oncology. All rights reserved.

Table A1. Fourteen Scenarios for Three Target p_T *Values Representing Different Practical Dose Responses*

	Dose					
$p_T = 0.1$ (scenario)	1	2	3	4	5	6
1	0.04	0.05	0.06	0.07	0.08	0.09
2	0.15	0.2	0.25	0.3	0.35	0.4
3	0.01	0.1	0.2	0.25	0.3	0.35
4	0.01	0.02	0.03	0.04	0.1	0.25
5	0.05	0.4	0.5	0.6	0.65	0.7
6	0.01	0.03	0.05	0.4	0.5	0.6
7	0.01	0.02	0.03	0.04	0.05	0.4
8	0.09	0.11	0.13	0.15	0.17	0.19
9	0.05	0.07	0.09	0.11	0.13	0.15
10	0.01	0.03	0.05	0.07	0.09	0.11
11	0.02	0.04	0.08	0.12	0.17	0.25
12	0.02	0.04	0.07	0.1	0.15	0.2
13	0.1	0.15	0.2	0.25	0.3	0.35
14	0.01	0.03	0.05	0.06	0.08	0.1
	Dose					
$p_T = 0.2$ (scenario)	1	2	3	4	5	6
1	0.02	0.05	0.08	0.11	0.14	0.17
2	0.25	0.35	0.4	0.5	0.6	0.7
3	0.01	0.2	0.4	0.6	0.8	0.95
4	0.04	0.06	0.08	0.1	0.2	0.5
5	0.05	0.5	0.8	0.9	0.95	0.99
6	0.01	0.05	0.1	0.5	0.7	0.9
7	0.01	0.03	0.07	0.1	0.15	0.7
8	0.19	0.21	0.23	0.25	0.27	0.29
9	0.15	0.17	0.19	0.21	0.23	0.25
10	0.11	0.13	0.15	0.17	0.19	0.21
11	0.05	0.11	0.17	0.23	0.29	0.35
12	0.05	0.1	0.15	0.2	0.3	0.4
13	0.2	0.25	0.3	0.35	0.4	0.45
14	0.05	0.08	0.11	0.14	0.17	0.2
	Dose					
$p_T = 0.3$ (scenario)	1	2	3	4	5	6
1	0.02	0.05	0.1	0.15	0.2	0.25
2	0.35	0.45	0.5	0.6	0.7	0.8
3	0.01	0.3	0.55	0.65	0.8	0.95
4	0.04	0.06	0.08	0.1	0.3	0.6
5	0.05	0.6	0.8	0.9	0.95	0.99
6	0.01	0.05	0.1	0.6	0.7	0.9
7	0.01	0.03	0.07	0.1	0.15	0.75
8	0.29	0.31	0.33	0.35	0.37	0.39
9	0.25	0.27	0.29	0.31	0.33	0.35
10	0.21	0.23	0.25	0.27	0.29	0.31
11	0.05	0.2	0.27	0.33	0.39	0.45
12	0.05	0.1	0.2	0.3	0.4	0.4
13	0.3	0.35	0.4	0.45	0.5	0.55
14	0.15	0.18	0.21	0.24	0.27	0.3

NOTE. Data represent true toxicity probabilities.

© 2013 by American Society of Clinical Oncology

Downloaded from jco.ascopubs.org on May 9, 2013. For personal use only. No other uses without permission.
Copyright © 2013 American Society of Clinical Oncology. All rights reserved.

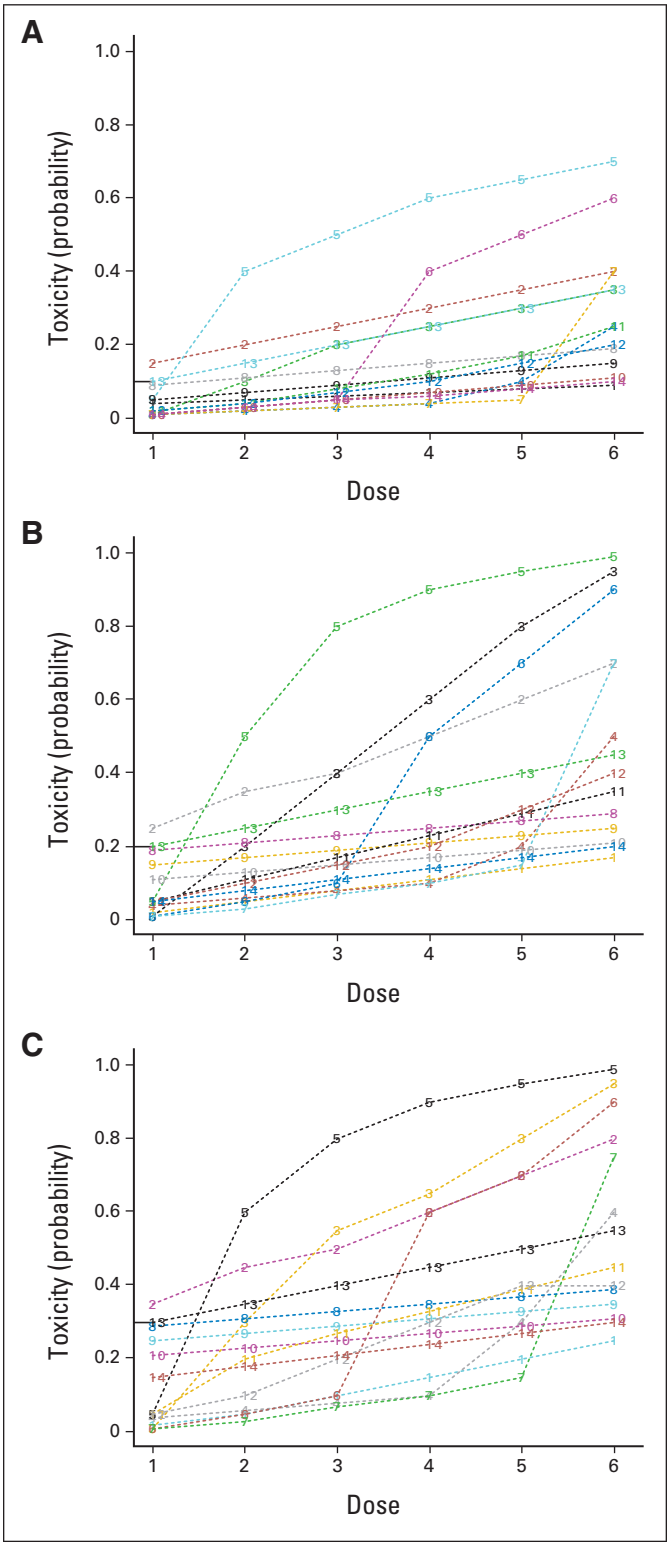

Fig A1. Dose-response patterns for the 42 clinical scenarios in which (A) $p_T = 0.1$, (B) $p_T = 0.2$, and (C) $p_T = 0.3$; 14 scenarios were constructed for each.

© 2013 by American Society of Clinical Oncology
Downloaded from jco.ascopubs.org on May 9, 2013. For personal use only. No other uses without permission.
Copyright © 2013 American Society of Clinical Oncology. All rights reserved.

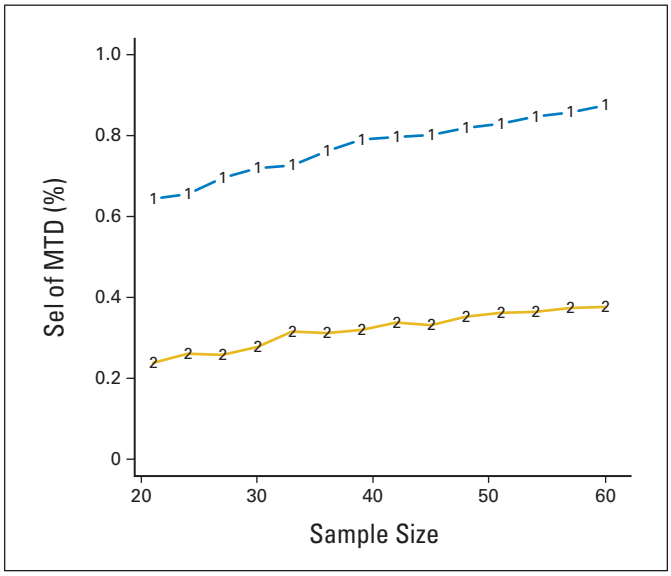

Fig A2. Relationship between sample size and the percentage of time the true maximum-tolerated dose (MTD) is selected in two standard scenarios. Sel, selection.

© 2013 by American Society of Clinical Oncology

Downloaded from jco.ascopubs.org on May 9, 2013. For personal use only. No other uses without permission.
Copyright © 2013 American Society of Clinical Oncology. All rights reserved.