

## Designing Dose-Escalation Trials With Late-Onset Toxicities Using the Time-to-Event Continual Reassessment Method

Daniel Normolle and Theodore Lawrence

From the Department of Radiation Oncology, University of Michigan, Ann Arbor, MI.

Submitted September 26, 2005; accepted July 13, 2006.

Supported by National Cancer Institute/ National Institutes of Health Grants No. P01 CA59827 and RO1 CA78554.

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

Address reprint requests to Daniel Normolle, PhD, 8D22, 300 N Ingalls, Ann Arbor, MI 48109-0473; e-mail: monk@umich.edu.

© 2006 by American Society of Clinical Oncology

0732-183X/06/2427-4426/\$20.00

DOI: 10.1200/JCO.2005.04.3844

### A B S T R A C T

#### Purpose

The standard design for phase I trials of combined chemotherapy and radiation, which enters either three or six patients per dose level, has little statistical basis and is subject to opening and closing because of delayed toxicities that disrupt patient accrual. We compared the operating characteristics of this standard design and the time-to-event continual reassessment method (TITE-CRM) for dose-escalation trials of combination chemotherapy and radiation.

#### Methods

The operating characteristics were determined by Monte Carlo simulation of 60,000 phase I trials.

#### Results

Compared with the standard trial design, in studies with delayed toxicity (ie, where four or more patients are expected to enter onto the study during a single previously enrolled patient's observation for toxicity), TITE-CRM trials are significantly shorter when toxicity observation times are long, treat more patients at or above the maximum-tolerated dose, identify the maximum-tolerated dose (MTD) more accurately, and provide phase II information, but do not expose patients to significant additional risk. Estimation precision and overdose control of TITE-CRM increase as the design assumptions more closely resemble the true state of nature, but are reduced if, for instance, the toxicity of treatment has been grossly underestimated.

#### Conclusion

Compared with the standard design, if there is any prior knowledge concerning the toxicity profile of a treatment, TITE-CRM can leverage it to produce more accurate estimates of the MTD and does not expose patients to significant excess risk, but requires timely communication between clinical investigators, data managers, and study statisticians.

*J Clin Oncol* 24:4426-4433. © 2006 by American Society of Clinical Oncology

### INTRODUCTION

A number of alternatives have been proposed to the standard three-or-six-patients-per-dose paradigm (3 + 3 design) for phase I dose-escalation trials that improve on its efficiency and statistical validity, including the continual reassessment method (CRM),<sup>1,2</sup> and its variations,<sup>3,4</sup> time-to-event CRM (TITE-CRM),<sup>5</sup> accelerated titration,<sup>6</sup> up and down,<sup>7</sup> overdose control,<sup>8</sup> and the isotonic design.<sup>9</sup> The standard 3 + 3 design is poorly suited for phase I trials of radiochemotherapy, for which toxicities such as radiation pneumonitis or radiation-induced liver disease may occur up to several months after treatment. In this case, the 3 + 3 design, or any design that requires all patients to have completed observation for toxicity, is subject to openings and closings as patients present after a dose level potentially has filled, but before sufficient time

has elapsed to be certain that treatment has not produced dose-limiting toxicity (DLT).<sup>10,11</sup>

We were interested in a dose-escalation design that did not require the trial to be closed to accrual while patients were observed, and provided a statistically valid assessment of the maximum-tolerated dose (MTD), but did not expose patients to unnecessary risk. TITE-CRM, a variant of the original CRM paradigm, is open to accrual continually, thus fulfilling the first of our requirements. Indeed, we have completed a clinical trial of combination gemcitabine, cisplatin, and radiation to treat patients with pancreatic cancer using the TITE-CRM paradigm, thereby demonstrating the feasibility of the TITE-CRM design.<sup>12</sup> Like CRM, TITE-CRM seeks to determine the target dose, defined as the dose most closely identified with the target rate, which is the largest acceptable probability of toxicity, determined a priori by the investigators based on the

relative costs and benefits of the treatment (typically between 5% and 25%). As the trial progresses and patients do or do not experience toxicities at different doses, the estimates of probability of toxicity are recalculated using a Bayesian expectation, and subsequent patients are assigned to doses under the principle to always treat at the target dose (which can be modified, if necessary, to address the concerns of review boards).

We hypothesized that, in addition to permitting more consistent accrual than the 3 + 3 design, TITE-CRM would provide a better estimate of the MTD, but without a significant increase in risk. In this report, we present Monte Carlo simulations of clinical trials that demonstrate the superior properties of the TITE-CRM algorithm in terms of efficient and ethical use of scarce patient resources and improved characterization of the dose-toxicity relationship. We show that TITE-CRM selects the correct dose more often than the 3 + 3 design. Furthermore, for treatments with late-onset toxicity, TITE-CRM is significantly more efficient than either the 3 + 3 or the standard CRM design, and it achieves this efficiency without a significant increase in observed toxicity.

Dose	Risk Model			
	1	2	3	4
1	.01	.01	.01	.01
2	.04	.05	.05	.04
3	.07	.10	.10	.07
4	.10	.20	.14	.10
5	.15	.32	.20	.50
6	.20	.50	.26	.70
7	.25	.70	.34	.90

NOTE. All TITE-CRM trials initially assumed risk model 1 as the true state of nature.  
Abbreviation: TITE-CRM, time-to-event continual reassessment method.

**METHODS**

We describe the TITE-CRM paradigm as used in the simulations and the overall structure of the Monte Carlo experiment.

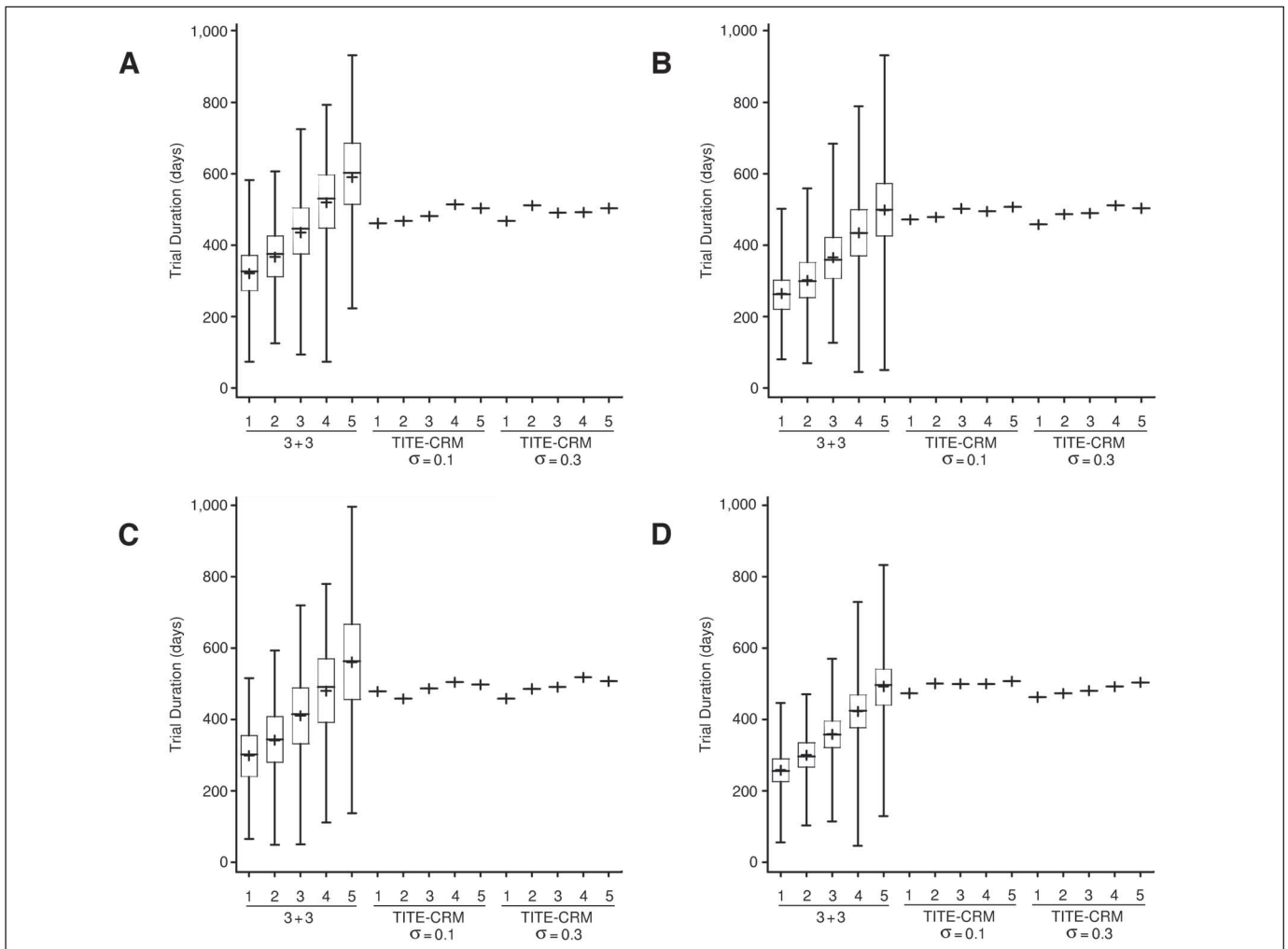


Fig 1. Duration of simulated trials. The box plots are skeletal; the lower and upper boundaries of the box represent the first and third quartiles, respectively; the middle line indicates the median, (+) denotes mean, and the vertical lines are drawn to the minimum and maximum values. The numbers on the horizontal axis indicate the ratio of the observation period to the mean patient interarrival time. 3 + 3, standard dose escalation design; TITE-CRM, time-to-event continual reassessment method.

**Phase I Algorithms**

The 3 + 3 design was implemented according to the rules summarized by Storer.<sup>7</sup> Where possible, shortcuts that would be used in a real trial were employed (eg, once two patients had experienced toxicity at a given dose, no additional patients were enrolled at that dose, but the next available patient was immediately assigned to the next lower dose, unless six patients had already been tested at that dose).

The TITE-CRM was implemented as described by Cheung and Chappell<sup>5</sup> and Muller et al.<sup>12</sup> The acceptable frequency of toxicity (target rate) was set at 0.20. A logistic dose-toxicity model was used<sup>1-4</sup>:

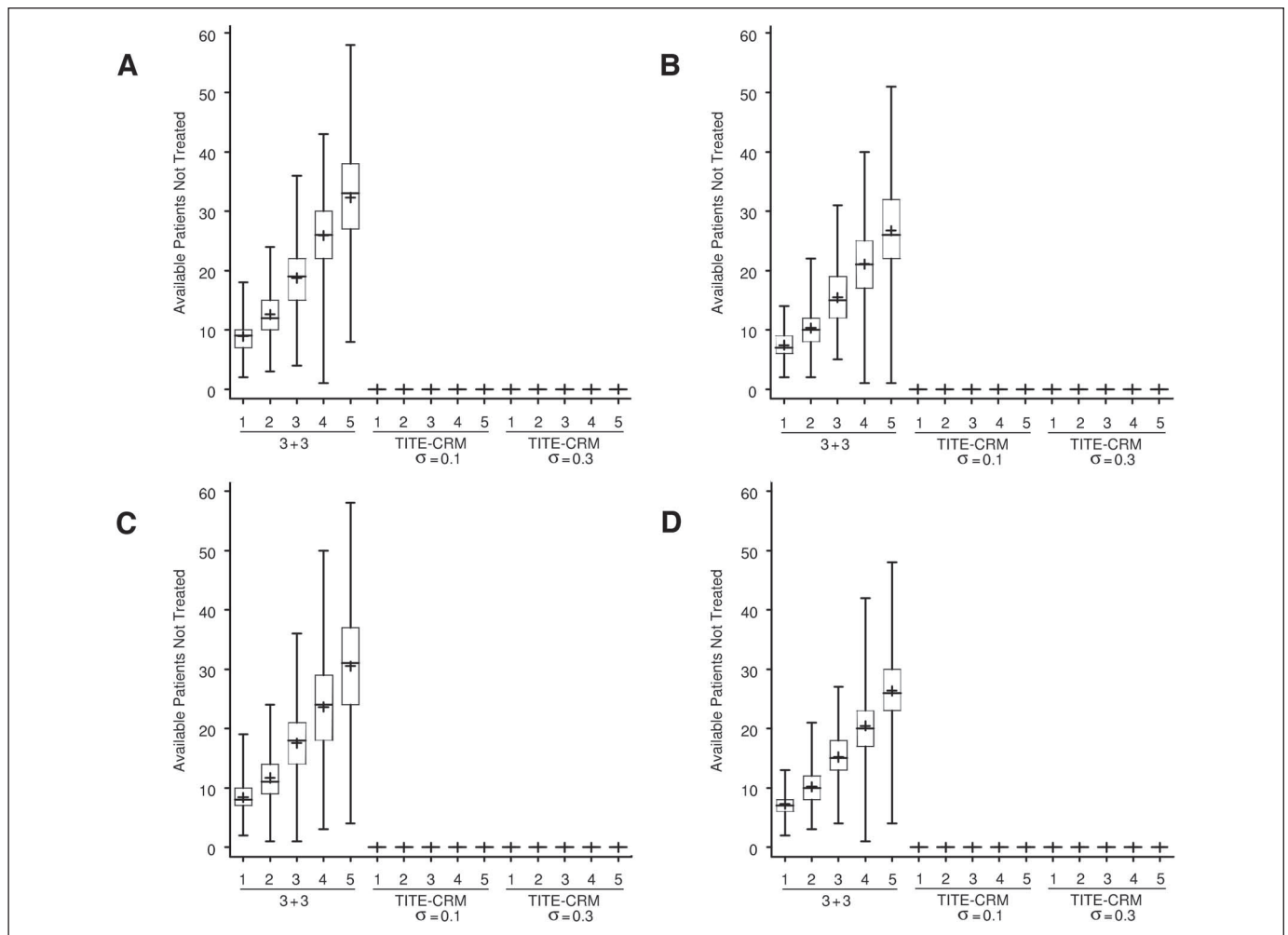
$$\varphi(d; \alpha) = e^{3+ad} / (1 + e^{3+ad})$$

where  $d$  is the relabeled dose [eg, if  $P = .10$ , then  $d$  equals  $-5.197$ , given that  $\varphi(-5.197; 1) = 0.10$ ], and  $\alpha$  is the estimand, the dose-toxicity parameter. For each simulated trial, following the standard CRM practice described by O'Quigley et al<sup>1</sup> and Cheung and Chappell,<sup>5</sup> the estimate of  $\alpha$  was initially set to 1, and the doses were relabeled so that the model fit the initial estimates of toxicity.

In this experiment, the investigators' initial estimates are always equal to column 1 in Table 1, so that the investigators' estimates are correct for risk model 1 and wrong in different degrees for the other risk models. In Bayesian

estimation (used by the CRM algorithms), in addition to an initial estimate, each parameter is associated by a prior distribution, which expresses the researchers' belief in the quality of the initial estimate. As the trial progresses, this prior distribution is combined mathematically with the observed data to yield the posterior distribution of the parameter, from which, in this case, estimates of  $\alpha$  are derived. We used a normal prior distribution for  $\alpha$  with mean equal to 1 and standard deviation of either  $\sigma = 0.1$  or  $\sigma = 0.3$ . The prior distribution determines the responsiveness of CRM to the accumulated data:  $\sigma = 0.1$  will result in a stiff estimator, where the toxicity probability estimates will remain close to the initial estimates unless significant data otherwise accrues, whereas  $\sigma = 0.3$  will tend to be more immediately responsive to data, for instance, reducing the target dose faster when a toxicity is observed. As in all TITE-CRM trials, whenever a patient presented for enrollment, the estimate of  $\alpha$  was determined from a weighted likelihood,<sup>5</sup> using weights uniform over the observation period for that trial; all previously enrolled patients who had experienced toxicity before the new patient's enrollment or had completed the observation period without toxicity were assigned weights of 1; otherwise, they were assigned weights equal to the proportion of the observation period they had completed.

Thirty-six patients were enrolled onto each TITE-CRM trial, which is a typical sample size. In both 3 + 3 and TITE-CRM trials, the first patient was



**Fig 2.** Number of patients untreated, equal to the number of patients who arrived before the final enrolled patient, but could not be treated because cohorts were under observation. Time-to-event continual reassessment method (TITE-CRM) trials can enroll patients at any time, so their values are uniformly equal to zero. The box plots are skeletal; the lower and upper boundaries of the box represent the first and third quartiles, respectively; the middle line indicates the median, (+) denotes mean, and the vertical lines are drawn to the minimum and maximum values. The numbers on the horizontal axis indicate the ratio of the observation period to the mean patient interarrival time. 3+3, standard dose escalation design.

assigned to dose 1. In the TITE-CRM trials, dose escalation was restricted to one dose between adjacent patients, but dose de-escalation was unrestricted. Escalation was not allowed until at least one patient had completed the observation period at the previous dose without experiencing toxicity. These rules, which are not inherent in the TITE-CRM paradigm, are conservative, but common in dose-escalation trials, and their implications are explored in the Discussion.

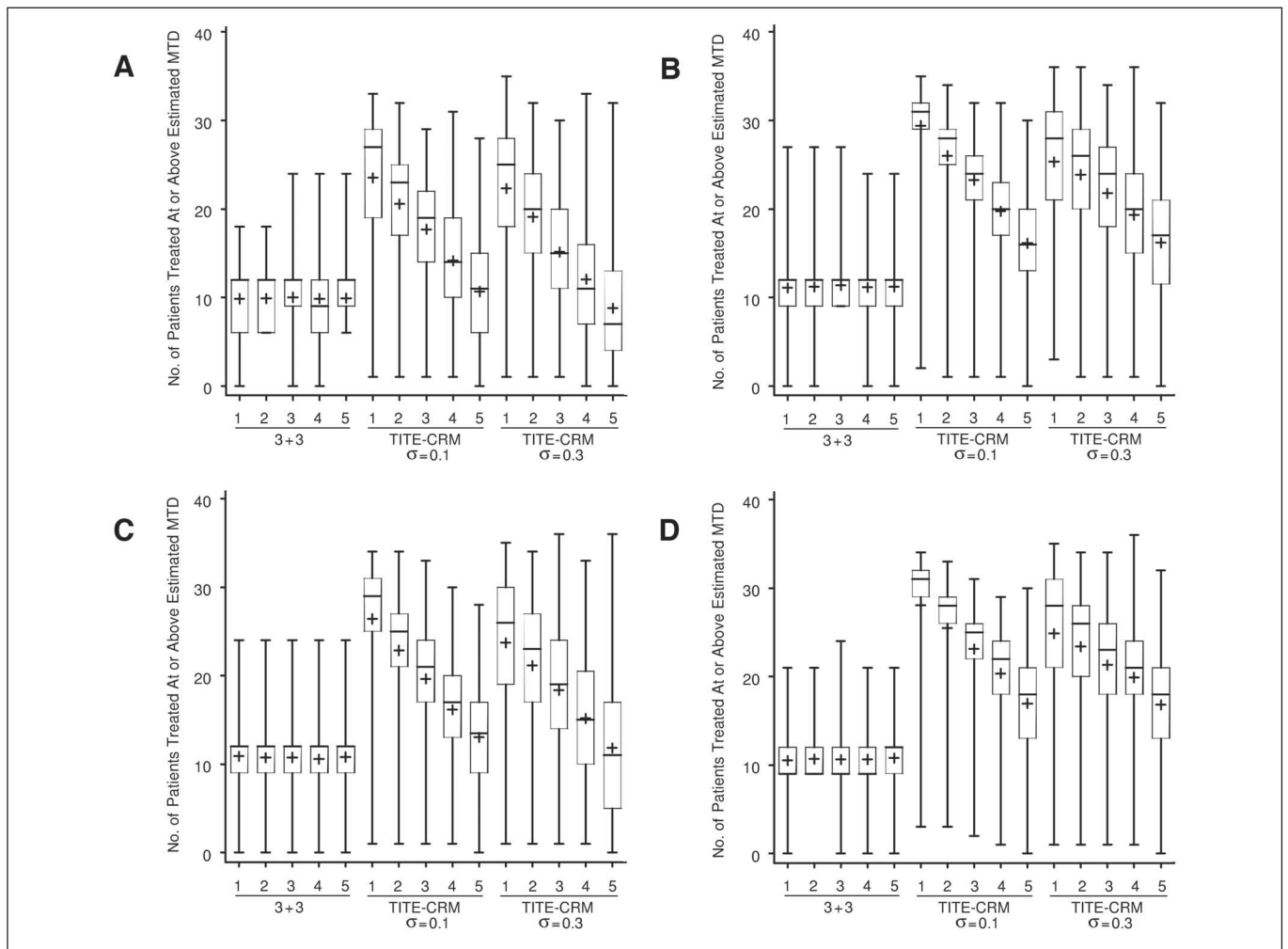
**Monte Carlo Experiment**

We assumed the objective of each simulated phase I trial was to test seven dose levels to determine the MTD, or, in CRM terminology, the target dose. The simulation experiment is arranged in a  $4 \times 5$  factorial design, with five observation periods (10 to 50 days in increments of 10 days) and four population dose-toxicity patterns (Table 1). The five observation periods represent trial designs for increasingly delayed toxicities. The units of the observation periods, and the exact numbers of days, were chosen arbitrarily for ease of presentation; the ratio of the observation period to the mean patient interarrival time (O/I ratio) determines the relative efficiencies of the TITE-CRM and the 3 + 3 designs, and the probability that previously enrolled patients will not have completed observation when a new patient presents for enrollment. In our chemoradiotherapy trials, O/I ratios of 2 to 3 are common, and have run as

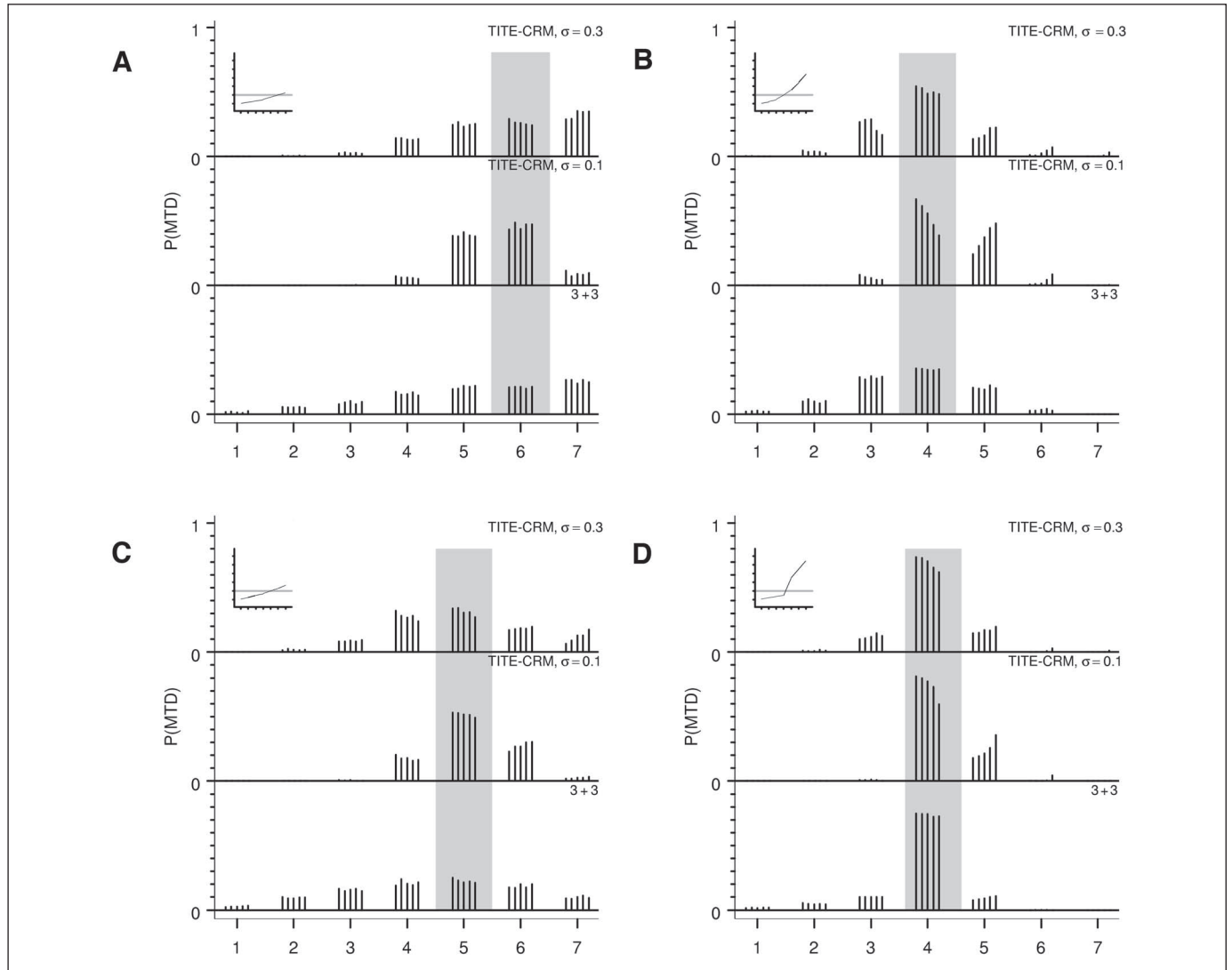
high as 12. The four risk models describe the true relationship between dose and toxicity. Risk model 1 has a gradual increase in toxicity with dose; risk model 2 has an abrupt jump after dose 3 and high toxicity at doses 5 through 7; risk model 3 displays a gradual increase in toxicity, but higher than is initially assumed; and risk model 4 has an extreme jump in toxicity over the MTD after dose 4. Risk model 1 is representative of a situation in which there is some experience with the treatment or its components, and the doses selected are close to the true MTD.

Each of the 20 design cells has 1,000 replicated sets of patient data. Both 3 + 3 and TITE-CRM trials draw from the same pool of 90 potential patients with interarrival times uniformly distributed between 1 and 19 days, for a mean interarrival time of 10 days. A patient was assigned to a trial if it was open for accrual when the patient arrived; given that the TITE-CRM trial was always open to accrual, it always drew the first 36 patients in the set, whereas the 3 + 3 trials accepted patients only as the algorithm permitted.

Each patient's response was either toxicity or no toxicity, where the dose-specific probability of toxicity is presented in Table 1. If a patient's response was toxicity, the time to observed toxicity was drawn uniformly between 1 day and the length of the observation period of the design cell. If a patient's response was no toxicity, that response became available at the end of the observation period.



**Fig 3.** Number of patients treated at or above the selected maximum-tolerated dose (MTD). The box plots are skeletal; the lower and upper boundaries of the box represent the first and third quartiles, respectively; the middle line indicates the median, (+) denotes mean, and the vertical lines are drawn to the minimum and maximum values. The numbers on the horizontal axis indicate the ratio of the observation period to the mean patient interarrival time. 3 + 3, standard dose escalation design; TITE-CRM, time-to-event continual reassessment method.



**Fig 4.** Proportions of trials that select the dose indicated on the horizontal axis to be the maximum-tolerated dose (MTD; or target) dose. The true target dose for each risk model is indicated by the shading. The five bars over each dose indicate the increasing (from 1 to 5) ratio of the observation period to the mean patient interarrival time (O/I ratio). The inset graph indicates the true dose-toxicity function for each risk model, where the horizontal line indicates the target rate of 0.20. TITE-CRM, time-to-event continual reassessment method.

All simulations were performed using SAS Version 9.1 (SAS Institute, Cary, NC), using its random number generators, RANBIN and RANUNI.

## RESULTS

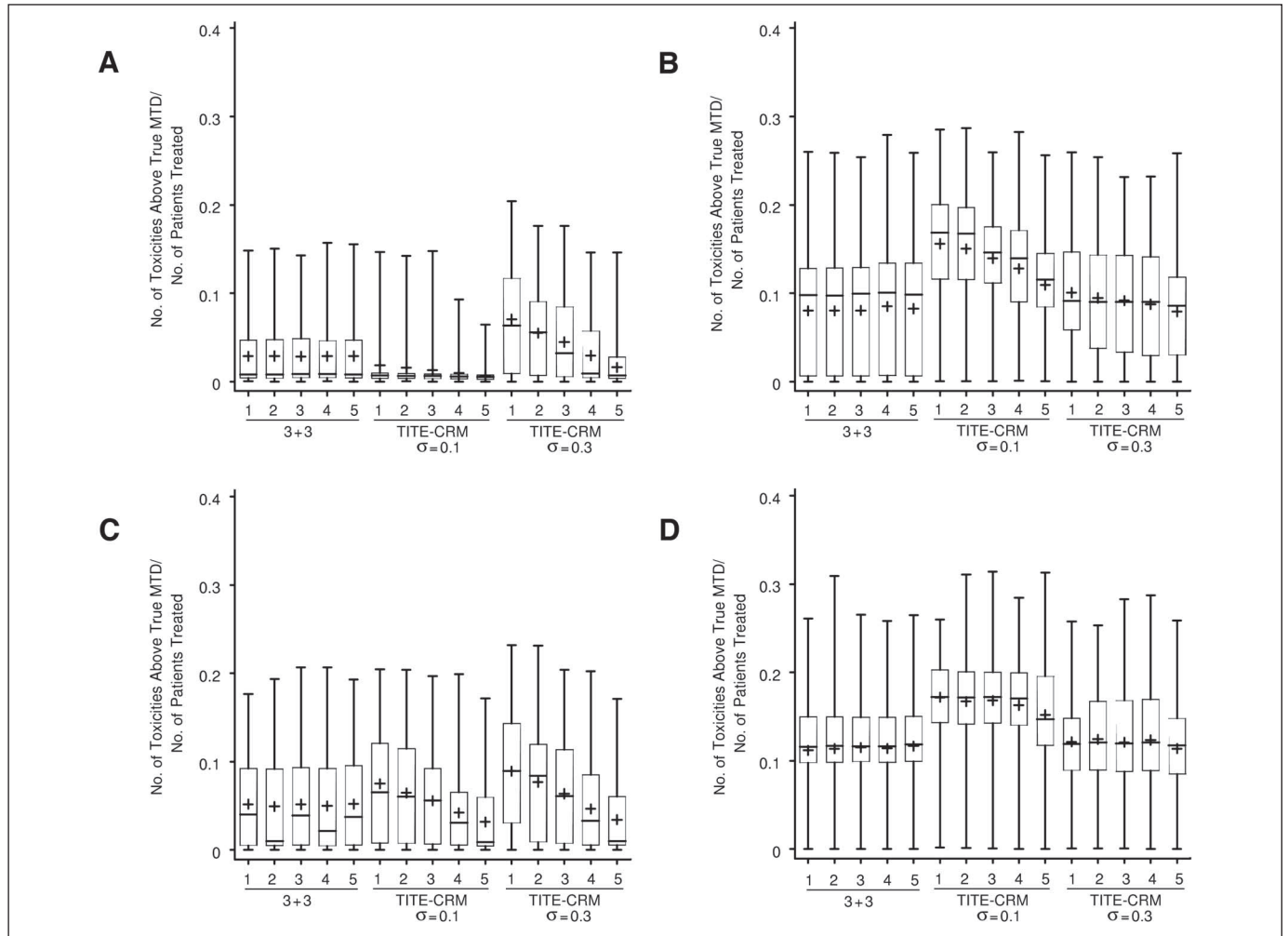
### Efficiency

All TITE-CRM trials accrued 36 patients; 3 + 3 trials accrued a median of 24 patients under risk models 1 and 3, and 21 patients under risk models 2 and 4. Figure 1 demonstrates that, as the O/I ratio increases above 3, durations of TITE-CRM trials are reduced compared with 3 + 3 trials. Figure 2 displays the number of patients who arrive while the trials are closed to accrual and are therefore unavailable for analysis. The number of patients who cannot be enrolled onto 3 + 3 trials increases linearly with the O/I ratio to more than the total number of patients enrolled onto a trial. TITE-CRM trials are always open for enrollment. Figure 3 demonstrates the numbers of patients

treated at or above the MTD, who could be used in an efficacy analysis. For low O/I ratios, the numbers of patients available are equivalent to that of an early, single-arm, phase II efficacy trial.

### Identification of MTD

There are significant differences between the TITE-CRM and 3 + 3 designs in the estimation of the recommended dose (Fig 4). In risk model 1, the true target dose, dose 6, is identified as the target dose in half of the trials by the TITE-CRM with  $\sigma = 0.1$ , and one third of the trials when  $\sigma = 0.3$ ; in both TITE-CRM cases, the next lowest dose is chosen in 40% of the trials; the advantage of the tighter prior distribution on  $\alpha$  when the model is correct is demonstrated clearly. When the toxicity increases gradually with dose (risk models 1 and 3), the 3 + 3 design is almost uninformative. When using  $\sigma = 0.3$ , TITE-CRM behaves similarly to the 3 + 3 design in risk model 3, and dose 7 [true  $P(\text{DLT}) = .34$ ] is chosen in 15% of the trials, as opposed to 10% chosen by the 3 + 3 design. The 3 + 3 design is more robust when



**Fig 5.** The number of toxicities occurring in patients treated above the true target dose (indicated in Fig 4), normalized by the total number of patients enrolled. The box plots are skeletal; the lower and upper boundaries of the box represent the first and third quartiles, respectively; the middle line indicates the median, (+) denotes mean, and the vertical lines are drawn to the minimum and maximum values. The numbers on the horizontal axis indicate the ratio of the observation period to the mean patient interarrival time. MTD, maximum-tolerated dose; 3 + 3, standard dose escalation design; TITE-CRM, time-to-event continual reassessment method.

sudden increases in the true probability of toxicity are present. In risk model 2 (using  $\sigma = 0.1$ ) the prior estimates are overweighted, and cause TITE-CRM to choose dose 5 [ $P(\text{DLT}) = .32$ ] in between 25% and 45% of the trials, as opposed to 25%, irrespective of the O/I ratio, by the 3 + 3 design and TITE-CRM using  $\sigma = 0.3$ . All of the methods are actually more accurate when the jump in toxicity is extreme (risk model 4), given that toxicities are highly likely to occur in three or more patients at dose 5 [ $P(\text{DLT}) = .5$ ], causing all of the algorithms to avoid testing that and higher doses. The TITE-CRM method is most susceptible to underestimation of toxicity when there are jumps in the dose-toxicity function and the O/I ratio is high.

### Overdose Control

The proportions of patients who are treated above the true MTD and experience toxicity are shown in Figure 5. In risk models 1 and 3, where the assumed model is correct, or nearly so, using TITE-CRM the tight prior distribution ( $\sigma = 0.1$ ) results in toxicity no greater than that of the 3 + 3 design (significantly less in risk model 1). Decreasing excess toxicity associated with increasing O/I ratio when using the looser prior distribution ( $\sigma = 0.3$ ) is caused by the escalation con-

straint; starting at dose 1, when the observation period is long, TITE-CRM takes a long time to get to the target dose, which reduces toxicity but adversely affects estimation, as seen in Figure 4. The trade-off is reversed in risk models 2 and 4, where the stiffer prior distribution causes overallocation to higher doses, and a median toxicity index 50% higher for TITE-CRM than for 3 + 3. TITE-CRM produced uniformly less excess toxicity, relative to the estimated MTD and normalized to the total number of patients treated, than 3 + 3 (not shown) in all risk models, irrespective of the I/O ratio and prior distribution employed.

## DISCUSSION

CRM and TITE-CRM trials generally recruit between 24 and 36 participants.<sup>13</sup> We did not vary the sample size in our Monte Carlo experiment in the interests of simplicity of presentation, but, when designing a TITE-CRM trial, the effect of the sample size on the operating characteristics should be determined by simulation. Larger

sample sizes are associated with smaller credible intervals around the estimate of the dose-toxicity parameter  $\alpha$  and its functions, resulting in better estimates of the true toxicity probabilities, especially close to the target dose. We consider TITE-CRM trials as phase I/II designs with an improved dose-selection phase; given that TITE-CRM concentrates most of the accrual around the target dose, early estimates of efficacy are possible. The phase II segment of a phase I/II design is sometimes used to augment the phase I segment trial with a 3 + 3 design, but the poor dose-selection characteristics of the phase I component make it likely that the phase II cohort will be treated at the wrong dose; efforts to further adjust the dose in the phase II cohort results in a design that, essentially, is CRM.

We found that the TITE-CRM design produces statistically sound estimates of the MTD that are closer to the truth than the estimates from the 3 + 3 design when the doses were well chosen, as in risk models 1 and 3. When the doses were too widely spaced (risk models 2 and 4), causing large increases in toxicity between adjacent doses, the advantage of the TITE-CRM trials was reduced, leading to possibility of an excessively high target dose, especially if the prior distribution is too tight. This possibility can be reduced by increasing the number of dose levels, increasing the sample size, and relaxing the prior distribution on  $\alpha$ . Conversely, when the increase in risk is gradual, the 3 + 3 design has an increased likelihood of choosing a dose that is too low, leading to reduced efficacy in subsequent phase II trials. The operating characteristics of any dose-escalation trial design should be evaluated carefully by simulation if the possibility of abrupt increases in toxicity is significant.

CRM and TITE-CRM trials treat more patients at doses closer to the finally selected dose (Fig 3). Therefore, more patients are likely to be treated at therapeutically effective doses, and more information is obtained about the toxicity at doses that subsequent patients are likely to receive. The lower number of patients treated at or above the MTD in trials with higher O/I ratios are a result of starting the trial at dose 1, and the delayed dose-escalation constraint, both of which violate the CRM paradigm of always treating at the target dose. Although the 3 +

3 trials appear to achieve lower toxicity under risk models 2 and 4, they do so at the cost of reduced information about where the true MTD is, thereby delaying the observed risk to subsequent efficacy trials. Statistical characteristics of TITE-CRM trials could be improved further by starting trials at higher doses and relaxing dose-escalation constraints that are not inherent to the paradigm.

These findings are consistent with our experiences with 3 + 3 designs for chemoradiotherapy studies. For instance, we carried out a phase I radiation dose escalation trial with fixed full-dose gemcitabine (1,000 mg/m<sup>2</sup>). We ultimately required 37 patients to evaluate seven dose levels, which took us more than 2.5 years to complete.<sup>14</sup> Similarly, a trial by Akerley et al<sup>15</sup> required 26 patients and six dose levels during 2 years to establish the MTD for combining paclitaxel with radiation for the treatment of patients with locally advanced lung cancer. In both of these trials, the great majority of the patients were treated using doses of either radiation<sup>14</sup> or chemotherapy<sup>15</sup> well below that which were ultimately determined to be the proper phase II dose. These practical examples demonstrate some of the difficulties and inadequacies of the 3 + 3 models for chemoradiotherapy trials, and encourage exploration of more efficient algorithms such as TITE-CRM, which are able to generate both toxicity and response data and treat the majority of patients at or near the target dose.<sup>12</sup>

Many clinical researchers have been hesitant to use TITE-CRM or other newer phase I designs because of discomfort with Bayesian methods, fear that the design is less safe than the standard approach, and the added burden of convincing oversight committees of the validity of a new approach. We have demonstrated in practice that clinical investigators can be acclimatized to Bayesian methods, that TITE-CRM design does not put patients at significant excess risk, and that oversight committees will accept Bayesian designs if they are explained in detail. We believe that the benefit to patients of defining more accurately and efficiently the appropriate doses for chemoradiotherapy, and of administering these doses to a higher fraction of patients within the trial, more than outweigh these concerns.

## REFERENCES

- 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
- O'Quigley J: Another look at two phase I clinical trial designs. *Stat Med* 18:2683-2690, 1999
- Piantadosi S, Fisher J, Grossman S: Practical implementation of a modified continual reassessment method for dose-finding trials. *Cancer Chemother Pharmacol* 41:429-436, 1998
- Goodman S, Zahurak M, Piantadosi S: Some practical improvements in the continual reassessment method for phase I studies. *Stat Med* 14:1149-1161, 1995
- Cheung K, Chappell R: Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics* 56:1177-1182, 2000
- Simon R, Friedlin B, Rubinstein L, et al: Accelerated titration designs for phase I clinical trials in oncology. *J Natl Cancer Inst* 89:1138-1147, 1997
- Storer B: Design and analysis of phase I clinical trials. *Biometrics* 45:925-937, 1989
- Babb J, Rogatko A, Zacks S: Cancer phase I clinical trials: Efficient dose escalation with overdose control. *Stat Med* 17:1103-1130, 1998
- Leung D, Wang Y: Isotonic designs for phase I trials. *Control Clin Trials* 22:126-138, 2001
- Thall P, Lee S: Practical model-based dose-finding in phase I clinical trials: Methods based on toxicity. *Int J Gynecol Cancer* 13:251-261, 2003
- Korn E, Midthune D, Chen T, et al: A comparison of two phase I trial designs. *Stat Med* 13:1799-1806, 1994
- Muller J, McGinn C, Normolle D, et al: A phase I trial using the time-to-event continual reassessment strategy to escalate cisplatin with gemcitabine and radiation therapy for pancreatic cancer. *J Clin Oncol* 22:238-243, 2004
- Eisenhauer E, O'Dwyer P, Christian M, et al: Phase I clinical trial design in cancer drug development. *J Clin Oncol* 18:684-692, 2000
- McGinn C, Zalupski M, Shureiqi I, et al: Phase I trial of radiation dose escalation with concurrent weekly full-dose gemcitabine in patients with advanced pancreatic cancer. *J Clin Oncol* 19:4202-4208, 2001
- Akerley W, Glantz M, Choy H, et al: Phase I trial of weekly paclitaxel in advanced lung cancer. *J Clin Oncol* 16:153-158, 1998

**Authors' Disclosures of Potential Conflicts of Interest**

The authors indicated no potential conflicts of interest.

**Author Contributions**

**Conception and design:** Daniel Normolle, Theodore Lawrence  
**Collection and assembly of data:** Daniel Normolle  
**Data analysis and interpretation:** Daniel Normolle  
**Manuscript writing:** Daniel Normolle, Theodore Lawrence  
**Final approval of manuscript:** Daniel Normolle, Theodore Lawrence