



Published in final edited form as:

Biometrics. 2010 June ; 66(2): 541–548. doi:10.1111/j.1541-0420.2009.01297.x.

Risk-Group-Specific Dose Finding Based on an Average Toxicity Score

B. Nebiyou Bekele^{1,*}, Yisheng Li¹, and Yuan Ji²

¹Department of Biostatistics, Division of Quantitative Sciences, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, Texas 77030, USA

²Department of Bioinformatics and Computational Biology, Division of Quantitative Sciences, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, Texas 77030, USA

Summary

We propose a Bayesian dose-finding design that accounts for two important factors, the severity of toxicity and heterogeneity in patients' susceptibility to toxicity. We consider toxicity outcomes with various levels of severity and define appropriate scores for these severity levels. We then use a multinomial likelihood function and a Dirichlet prior to model the probabilities of these toxicity scores at each dose, and characterize the overall toxicity using an average toxicity score parameter. To address the issue of heterogeneity in patients' susceptibility to toxicity, we categorize patients into different risk groups based on their susceptibility. A Bayesian isotonic transformation is applied to induce an order-restricted posterior inference on the average toxicity scores. We demonstrate the performance of the proposed dose-finding design using simulations based on a clinical trial in multiple myeloma.

Keywords

Adaptive design; Bayesian inference; Isotonic regression; Average toxicity score

1 Introduction

The purpose of phase I oncology trial designs is to determine the highest dose with acceptable toxicity, the maximum tolerated dose (MTD), by sequentially assigning doses to patient cohorts. In these designs, the dose assigned to the next cohort is chosen adaptively based on the toxicity response data from previously treated patients in the trial.

Numerous statistical designs in which toxicity is assumed to be binary have been developed for phase I oncology trials (See Thall and Lee (2003) for a review). However, it is often useful to distinguish between levels of toxicity severity, which requires modeling toxicity as an ordinal, categorical, or continuous outcome (Bekele and Thall, 2004; Yuan, Chappell,

*bbekele@mdanderson.org.

⁷ Supplementary Materials

Supplementary Materials can be accessed at <http://www.tibs.org/biometrics>.

and Bailey, 2007; Ivanova and Kim, 2008). In addition, a common assumption of most dose-finding methods is that patients are homogeneous in their susceptibility to toxicity. In reality, patients may actually be heterogeneous due to their underlying disease status, clinical characteristics, or demographics (Extermann et al., 2004; Rogatko et al., 2004). For example, patients with good performance status (a measure of general well-being) may be less likely to experience toxicity at a given dose than those with poor performance status. Similarly, younger patients may tend to tolerate higher doses than older patients. Throughout this paper, we use the term “risk group” to denote a group of patients considered homogeneous in their susceptibility to toxicity. To our knowledge, a few papers have attempted to incorporate risk groups into statistical models for dose finding (Wijesinha and Piantadosi, 1995; Babb and Rogatko, 2001; Legedza and Ibrahim, 2001; O’Quigley and Paoletti, 2003; Yuan and Chappell, 2004; Ivanova and Wang, 2006), but all of these treat toxicity as a binary outcome.

In this paper, we consider the design of a clinical trial in which 1) toxicity is measured by severity levels, and 2) patients from various risk groups will be enrolled. Correspondingly, our method consists of two major components. First, after eliciting a score for each toxicity severity level, we adopt a Bayesian multinomial-Dirichlet model to estimate the probabilities of observing these toxicity scores. We then define the *average toxicity score* as a measure of the overall toxicity at each dose. Second, assuming a partial ordering of the average toxicity scores across dose/risk-group combinations, we borrow strength by modeling these scores using a Bayesian isotonic regression transformation (Robertson et al., 1988; Dunson and Neelon, 2003; Li et al., 2008). We propose decision rules for dose finding based on the posterior distributions of the order-constrained average toxicity scores.

There are various clinical scenarios in which this method could be used. These include situations in which investigators have identified clinical characteristics (gender, performance status, age category); biomarker characteristics (for example, investigators may know that patients with specific genetic characteristics may be more susceptible to toxicity due to the drug’s mechanism of action); and factors such as prior or current treatment (for example heavily pre-treated patients may be more susceptible to toxicity relative to patients exposed to less prior therapy).

The remainder of the paper is organized as follows. Section 2 describes a motivating trial. Section 3 presents the probability model, and Section 4 outlines the dose-finding algorithm. Section 5 compares the proposed method to another recently developed approach via computer simulations. Section 6 ends the paper with a discussion.

2 Motivating trial

The method developed in this paper is motivated by a phase I trial of a new treatment from the class of immunomodulatory drugs (that is, the compound is in the same therapeutic class as thalidomide) in combination with dexamethasone (a steroid) for multiple myeloma (MM) patients with renal insufficiency. This new compound has both immunomodulatory and anti-angiogenic properties, which could confer antitumor and antimetastatic effects. The main objective of the study is to determine if this compound can be safely given to patients

with MM. MM is a type of hematologic malignancy in which the immune system cells known as plasma cells become cancerous. Although MM may be amenable to treatment with this compound, a common negative consequence associated with this disease is renal insufficiency— a condition in which the kidneys fail to function properly. Renal functioning is measured via the glomerular filtration rate (GFR); patients with GFRs of 60 mL/min or less are categorized as having renal insufficiency. In the context of phase I clinical trials, the degree of renal insufficiency is believed to have an impact on how well a patient tolerates treatment. The principal investigators (PIs) of this phase I trial wanted to protect patients with renal insufficiency against the possible risk of increased toxicity. To this end, the PIs decided to stratify patients into three risk groups: patients with $40 < \text{GFR} \leq 60$ mL/min were deemed low risk, patients with $20 < \text{GFR} \leq 40$ mL/min were considered moderate risk, and patients with $\text{GFR} \leq 20$ mL/min were categorized as high risk. The PIs planned to evaluate up to four doses in the low risk group, but only evaluate the three lowest doses in the moderate risk group and the two lowest doses in the high risk group. Specifically, maximum sample sizes of 21, 18, and 12 were set for the low, moderate, and high risk groups, respectively, hence resulting in a total maximum sample size of 51 patients. Guidelines for how to choose the sample size is given in section 5 below.

In addition to defining the risk groups above, the PIs believed that the various toxicities expected in the trial could be classified into several toxicity severity categories and they believed that incorporating these toxicity severities into the dose-finding algorithm would be helpful and important. The toxicity severity categories are: Category 1 (no toxicity); Category 2 (grade 3 or 4 fatigue, and respiratory infection); Category 3 (anemia, thrombocytopenia, and neutropenia); Category 4 (cardiotoxicity); Category 5 (blood clots and myelosuppression with fever). Since a patient may experience two or more toxicity events from different categories, they decided to record the highest severity observed for that patient.

3 Methodology

3.1 Probability model

Let (Y_1, \dots, Y_n) denote the vector of toxicity outcomes on n patients currently enrolled in the trial. Note that n increases as new patients are enrolled in the study. The toxicity outcome for the i th patient, Y_i , can take on one of $K + 1$ ordered values. Without loss of generality, we rank these ordered values as $0, 1, \dots, K$, where by convention, 0 represents no toxicity. In the MM trial example, $K = 4$. Assume that there are M risk groups, and let r_i be the risk-group membership of the i th patient for $r_i = 1, \dots, M$. Assume that J_h doses are under investigation in the h th risk group, $h = 1, \dots, M$. Define d_i to be the dose received by the i th patient for $d_i = 1, \dots, J_{r_i}$. Lastly, let $p_{kjh} = \Pr(Y_i = k \mid d_i = j, r_i = h)$ be the probability of toxicity outcome k at the j th dose for the h th risk group. For a fixed dose j and a risk group

h , we assume that $\sum_{k=0}^K p_{kjh} = 1$.

Let $n_{kjh} = \sum_{i=1}^n I(Y_i = k, d_i = j, r_i = h)$ be the number of patients in risk group h assigned to dose j who experienced toxicity of severity category k . The likelihood for $\{n_{kjh}\}$ with fixed j and h is a multinomial probability

$$L_{jh} = \prod_{k=0}^K p_{kjh}^{n_{kjh}},$$

and the full likelihood is given by $L = \prod_{j=1}^{J_h} \prod_{h=1}^M L_{jh}$. Defining the vector $\mathbf{P}_{jh} = (p_{0jh}, p_{1jh}, \dots, p_{Kjh})$, we assume *a priori* that \mathbf{P}_{jh} follows a Dirichlet distribution given by

$$\mathbf{P}_{jh} \sim \text{Dirichlet}(\alpha_{0jh}, \alpha_{1jh}, \dots, \alpha_{Kjh}).$$

Due to conjugacy, the posterior distribution of \mathbf{P}_{jh} is also a Dirichlet,

$$\mathbf{P}_{jh} \sim \text{Dirichlet}(\alpha_{0jh} + n_{0jh}, \alpha_{1jh} + n_{1jh}, \dots, \alpha_{Kjh} + n_{Kjh}). \quad (1)$$

We will discuss the choice of the parameters $\alpha_{0jh}, \dots, \alpha_{Kjh}$ in Section 5.

3.2 Average toxicity score

To incorporate the impact of each toxicity severity category, we take an approach similar to that of Bekele and Thall (2004). Specifically, we first elicit a score for each toxicity severity category from the PIs. Let s_k be the elicited score for toxicity severity category k , and denote $\mathbf{s} = (s_1, \dots, s_K)$, where $0 = s_0 < s_1 < s_2 < \dots < s_K$. Here, the ordering of s_k reflects the ordering of the toxicity severity level. To measure the overall toxicity at a dose, we define the average toxicity score (ATS) at each dose/risk-group combination as

$$\psi(\mathbf{s}, j, h) = \sum_{k=1}^K s_k \cdot p_{kjh}. \quad (2)$$

For the MM trial, toxicity is defined as an ordinal variable with five levels (no toxicity, mild toxicity, moderate toxicity, severe toxicity, and extremely severe toxicity). The toxicity score s_k represents the PIs' expert opinion and may take on any positive values subject to the monotonicity constraint $s_0 < s_1 < s_2 < \dots < s_K$. Because only the relative magnitudes of the toxicity scores matter in our dose-finding algorithm, we elicited the s_k values from the PIs s_k and scaled them to be between 0 and 1. For the MM trial, we elicited the following scores for the five toxicity severity levels identified in Section 2: Category 1 (no toxicity) with $s_0 = 0$; Category 2 with $s_1 = 0.25$; Category 3 with $s_2 = 0.5$; Category 4 with $s_3 = 0.75$; Category 5 with $s_4 = 1.0$. Note that different investigators may prefer different scores. For example, one could use the National Cancer Institute common toxicity criteria grading system – a scale used to classify toxicities observed during the conduct of clinical trials.

3.3 Isotonic regression transformation

The ATS $\psi(\mathbf{s}, j, h)$ is a parameter summarizing the overall toxicity at dose j and risk group h . An underlying assumption in dose-finding trials is that a higher dose and a risk group with

higher susceptibility should lead to a higher toxicity rate. Therefore, a reasonable statistical model should constrain $\psi(\mathbf{s}, j, h)$ such that it increases in j and h . We apply a Bayesian isotonic transformation to posterior samples of $\psi(\mathbf{s}, j, h)$ so that the transformed samples can be considered from a posterior distribution that obeys the desired order constraints (Li et al., 2008). In implementing the Bayesian isotonic transformation, we first draw posterior samples of \mathbf{P}_{jh} from (1), which results in posterior samples of $\psi(\mathbf{s}, j, h)$ after the application of (2). We then apply the minimum lower sets algorithm (MLSA) (Robertson et al., 1988) to the posterior samples of $\psi(\mathbf{s}, j, h)$ and define $\psi^*(\mathbf{s}, j, h)$ to be the transformed ATS's for the j th dose and h th risk group. The transformed ATS parameters follow the partial ordering constraints such that

$$\psi^*(\mathbf{s}, j, h) \leq \psi^*(\mathbf{s}, j+1, h) \quad \text{and} \quad \psi^*(\mathbf{s}, j, h) \leq \psi^*(\mathbf{s}, j, h+1),$$

i.e., $\psi^*(\mathbf{s}, j, h)$ are monotonically increasing in dose given risk group and in risk group given dose. A rationale for using the MLSA in a similar matrix-order setting, can be found in Li et al. (2008). Details on the MLSA are in the *Biometrics* journal Supplementary Materials website.

4 Proposed dose-finding scheme

4.1 Dose-finding algorithm

As a first step the researchers will need to determine the target average toxicity score. Determination of this quantity (in a slightly different context) is given in Bekele and Thall (2004). Specific details on how to calculate this quantity in the context of the method proposed in this article can be found in the *Biometrics* journal Supplementary Materials website. Once we have determined the target ATS ψ_T , we develop decision rules for dose assignment and MTD estimation based on the posterior probabilities $\xi_{jh} = \Pr\{\psi^*(\mathbf{s}, j, h) > \psi_T \mid \text{data}\}$. To start, we partition the unit interval into three subintervals determined by the probability cut-offs $0 < \underline{\xi} < \bar{\xi} < 1$.

Definition—For a fixed risk group h , the toxicity at dose j is *negligible* if $\xi_{jh} < \underline{\xi}$, *acceptable* if $\underline{\xi} \leq \xi_{jh} \leq \bar{\xi}$, and *excessive* if $\xi_{jh} > \bar{\xi}$.

Reasonable values for these cut-offs are $.05 \leq \underline{\xi} \leq .30$ and $.70 \leq \bar{\xi} \leq .95$ (Bekele et al., 2008), with the particular values chosen to reflect how aggressively or conservatively the PIs wish the algorithm to behave. In addition, the choices of $\underline{\xi}$ and $\bar{\xi}$ should be guided by preliminary computer simulations and adjusted to obtain a design with desirable operating characteristics for the proposed trial. Given ψ_T , $\underline{\xi}$, and $\bar{\xi}$, we describe below a dose-finding algorithm for trials with multiple toxicity categories and risk groups.

1. Within any risk group, if the current dose exhibits negligible toxicity, then escalate to the next higher dose in that risk group.
2. Within any risk group, if the current dose exhibits acceptable toxicity, then stay at the current dose for that risk group.

3. Within a risk group, if the j th dose exhibits excessive toxicity and $j > 1$, then de-escalate to the highest dose less than j that is not excessively toxic.
4. At any point in the trial when a cohort of patients is fully assessed for toxicity within a given risk group, reassess all risk groups for toxicity and for dose-assignment decisions to de-escalate, stay, or escalate.
5. At the end of the trial, select from within each risk group a dose with either negligible or acceptable toxicity having a posterior mean of $\psi^*(\mathbf{s}, j, h)$ closest to the target ψ_T .

The algorithm above guides dose finding when using our proposed method. In addition to these steps, we include additional trial-conduct rules due to practical considerations. These rules are: 1) treat the first cohort of patients in each risk group at the lowest dose for that risk group (note, while the clinical trial for which the method was originally designed used this rule, the method is not constrained to require it), and 2) at any point in the trial, if the starting dose for the lowest risk group is deemed unsafe, then stop the trial. To assess the safety of the starting dose one could use ξ or alternatively, since the negative impact of stopping enrollment to a risk group without choosing a dose is high one may choose to define a risk group stopping probability ξ^* such that if $\xi_{1h} > \xi^*$ stop the trial. These rules are usually incorporated into a practical trial to protect the patients safety, even though they are not directly related to the statistical model used for dose finding.

5 Simulation study

For our simulation study, we constructed six scenarios with different true ATS across dose/risk-group combinations (see Figure 1). As stated in Section 2, the maximum sample size chosen for this simulation was 51, with maximum risk-group sample sizes of 21, 18, and 12 for the low, moderate, and high groups, respectively. In general the number of subjects chosen for a multiple risk groups trial is a function of the number of doses and risk-groups being studied. As a rule of thumb, an investigator could initially consider 6 patients per dose/risk-group combination (In this trial that number would be 54 subjects). The sample size obtained from this rule is equivalent to the maximum possible sample size under 3 independent trials using the standard 3+3 algorithm. From this starting point the investigator could subsequently identify the sample size to achieve desirable operating characteristics via simulations, in conjunction with the selection of tuning parameters ξ , ξ , and ξ^* . For each dose/risk-group combination, we set the parameters $(\alpha_{0jh}, \dots, \alpha_{Kjh}) = (0.604, 0.178, 0.089, 0.071, 0.059)$ for the Dirichlet prior of \mathbf{P}_{jh} . Under this prior, it can be shown that *a priori* $\psi(\mathbf{s}, j, h)$ is centered around 0.20 and has an effective sample size approximately equivalent to one patient.

In our simulation, we used the target ATS $\psi_T = 0.25$, based on the elicitation process described in Section 4.1. In addition, we used cutoff probabilities $\xi = 0.25$, $\xi = 0.90$ and risk group stopping probability $\xi^* = 0.95$. We examined six scenarios for the MM trial, and for each scenario we simulated 1,000 trials.

For comparison purposes, we also implemented the modified continual reassessment method (M-CRM) by Yuan, Chappell, and Bailey (2007). They model the toxicity scores $0 = w_0 <$

$w_1 < \dots < w_K = 1$ as fractional responses using a Bernoulli quasi-likelihood. Let s_i^* be the observed toxicity score for the i th patient. If the i th patient experiences the k th category of toxicity, then $s_i^* = s_k$. Within each risk group, the model parameters for the M-CRM are $\theta_{J,h} = (\hat{s}_{1,h}, \dots, \hat{s}_{J,h,h})$ and α_h , which are estimated by the quasi-likelihood given by

$$\tilde{L}_{n_h}(\theta_{J,h}, \alpha_h) = \prod_{i=1}^{n_h} \left(\hat{s}_{d_i,h}^{\alpha_h} \right)^{s_i^*} (1 - \hat{s}_{d_i,h}^{\alpha_h})^{1-s_i^*}. \quad (3)$$

The quantity $\theta_{J,h}$ can be considered a “toxicity score skeleton” in which $\hat{s}_{j,h}$ is fixed with $0 < \hat{s}_{1,h} < \dots < \hat{s}_{J,h,h} < 1$. In addition, $\alpha_h = e^{\beta_h}$, where β_h has a normal prior distribution with mean 0 and variance 100. In the simulations, we used $\theta_{4,1} = (0.05, 0.10, 0.3, 0.40)$ for the lowest risk group, $\theta_{3,2} = (0.15, 0.3, 0.40)$ for the middle risk group, and $\theta_{2,3} = (0.25, 0.35)$ for the highest risk group. These skeletons were chosen to reflect the PIs’ belief that toxicity would increase with dose and risk group. In addition, we attempted to calibrate the prior means and variances of β_h so that the reported M-CRM design had reasonable operating characteristics.

Within each risk group, we used the M-CRM to choose each successive dose, and determined the MTD by minimizing $|s_{j,h}^{\hat{\alpha}_h} - \psi_T|$. To ensure a fair comparison, we modified this method so that it would stop early for excessive toxicity, since our method does the same. Specifically, we inserted an additional rule when implementing the M-CRM that stopped the accrual of new patients to a risk group if $Pr(s_{1,h}^\alpha > \pi^* | \theta, D_n) > 0.90$.

The operating characteristics for our proposed method and the M-CRM are summarized in Tables 1 and 2, which are organized into scenario sections. For each scenario section, italicized rows represent the true ATS for that dose/risk-group combination multiplied by 100. Rows in bold face show the selection percentages and average number of patients assigned to a given dose/risk-group combination. The columns “Scen” and “Risk Grp” indicate the scenario and risk group being evaluated, while the column “None” represents the percentages of the 1,000 simulated trials in which no doses were selected from any of the risk groups due to excessive toxicity.

Recall that there are four doses for risk group 1 (low risk), three for risk group 2 (moderate risk), and two for risk group 3 (high risk). Our proposed method generally treats more patients at doses closer to the target ATS $\psi_T = 0.25$, and tends to be more conservative for higher risk groups (a positive feature in dose finding) because it uses information regarding toxicity risk across risk groups. For example, under Scenario 1, our method is more likely than the M-CRM to choose a dose closest to the target ATS (except for risk group 3), and on average places more patients at these doses. Scenario 2 is characterized by a steep increase in the dose-toxicity curve. Our method selects the highest non-toxic doses (dose 3 for risk group 1, dose 2 for risk group 2, and dose 1 for risk group 3) at least **35%** more often than the M-CRM. In Scenario 3, both methods choose doses closest to the target ATS over 50% of the time, the M-CRM outperforms our method with respect to correct selection percentage for risk groups 1 and 3 while our method outperforms the M-CRM for risk group 2. In scenarios 4–6 our method compares favorably to the M-CRM.

In the *Biometrics* journal Supplementary Materials website we provide additional simulation results which show how the method behaves under a wide variety of situations. These include additional simulation results evaluating the prior; spacing of the elicited toxicity scores; a set of cases in which the elicited toxicity scores are restricted to take on values of either 0 or 1 (equivalent to the binary toxicity case); additional simulation scenarios; scenarios comparing the proposed method and the M-CRM to a case in which the middle and lower risk groups are allowed to escalate to higher doses; cases in which the posterior mean toxicity score is isotonized rather than incorporating the isotonic regression as part of the Monte Carlo simulation; and lastly, cases in which concurrent independent phase I trials in each risk group are monitored using our proposed Dirichlet model in conjunction with the ATS framework and then using the MLSA only at the end of the trial to combine the results and make final dose selections. In general we found that our method works well. Of particular note is that the modified versions of the method using either the posterior mean isotonized ATS or the independent clinical trials ATS also work reasonably well.

6 Discussion

We have described a statistical framework for dose finding that incorporates ordinal toxicity severity levels. Our method also accommodates the degree of patients' susceptibility to toxicity via a Bayesian isotonic regression transformation for an average toxicity score across dose/risk-group combinations. Our method requires close interaction between the PIs and statisticians to establish the risk groups, toxicity severity categories, toxicity severity scores, and target ATS.

There are various alternative to modeling ordinal toxicity severity levels, including the cumulative link ordinal logistic regression, the adjacent categories logistic regression, and the continuation ratio models. However, each of these approaches requires that specific assumptions be met, and it is unclear how they would perform in the context of risk-group-specific dose finding.

In this article we compared our method to three independent applications of the M-CRM to each risk group. We note that it may be possible to extend the original M-CRM method to better handle clinical trials in which risk-groups have been identified. There are two ways to extend the M-CRM model to the multiple risk group setting. The first involves adding a risk-group effect to the M-CRM model in a manner analogous to O'Quigley and Paoletti (2003). The second approach would be to calculate the expected posterior toxicity score using the M-CRM within each risk group and then use the minimum lower sets algorithm on these expected posterior toxicity scores to obtain partially ordered estimates for dose finding. Although the model would be different, the dose-escalation rules would be similar to the rules used in our proposed method.

Our current design assumes fixed sample sizes with risk groups. If an investigator fills up one risk group and another patient of that risk comes along at a later timepoint, the investigators may want to enter him/her into the trial anyway. We note that our method can be modified to relax the restriction on fixed within risk-group sample sizes. Having noted this and although the concern over possibly wasting patients is a valid one, it is not clear that

an investigator will necessarily want to enter an excess patient into an arm which has already met its target sample size. Specifically, the investigators must balance the desire to not waste a possible patient against the competing desire to ensure that the toxicity profile of the compound is adequately characterized across all risk groups. Given that most trials are limited in the number of patients they can accrue, putting an excess number of patients on a filled risk group automatically ensures that fewer patients in the other risk groups can be enrolled into the study (under the usual assumption that the total sample is fixed due to monetary cost or study duration constraints). While in our trial the PI and sponsor both agreed that characterizing the toxicity profile for each group was most important, this is not an open-and-shut issue. This issue should be reviewed by the PI and statistician; potentially even running simulations comparing the two alternative enrollment strategies with respect to various operating characteristics.

As suggested by one reviewer, some may consider toxicity grades a more important factor in assessing toxicity than risk group. In this case, one could assume that the dose response curves are different for each risk group then it may possible to experiment with more dose levels over a finer grid rather than a subset of a few doses. While this is a valid concern, practical and statistical limitations may recommend use of risk-group based designs. First, safely escalating over a fine grid poses challenges because it may not be possible to implement a “do-not-skip a dose” without having an excessively large trial. In the absence of a do-not-skip rule, one could decide to implement a “do-not-jump-more-than-X-dose-units” rule. The researcher would have to carefully assess the operating characteristics of such a design. Second, it may not be possible to physically manufacture doses over a fine grid (due to financial constraints). This statement is especially true for compounds given in pill form. Specifically, the process of preparing and combining the various components of a new treatment can complex, expensive, and feasible only for a few doses because each of these doses should meet Regulatory Good Manufacturing Practice standards. Lastly, there is still something to be gained (either at the end of the trial or during the trial) from borrowing strength across doses. Specifically, were we to use only the information in a particular risk group to make dose-escalation decisions, we could possibly end-up in an situation in which the MTD for a higher risk groups is higher than the MTD for a lower risk group.

Our supplemental simulations showed that the method based on a non-isotonized dose-escalation scheme could be calibrated to provide similar operating characteristics to the method proposed in this article. While simple, care should be given in its application as it could lead to undesirable situations during trial conduct in which patients with higher risk profiles are assigned to higher doses than patients with lower risk profiles. The trade-off between simplicity and possible undesirable outcome should be explored carefully before the non-isotonized version of the method is implemented.

Extentions of this research include developing models to characterize continuous toxicity outcomes and ordinal-competing-risks outcomes (ordinal-competing-risk allow for enrollment before the current patient has been completely assessed for toxicity). We believe that the method presented in this paper could possibly be used in the case of continuous toxicity outcomes via an extension of Ivanova and Kim (2008). In the case of ordinal-

competing-risk outcomes, the method proposed in this paper could be extended using models developed by Cheung and Chappell (2000), Cheung and Thall (2002).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

- Babb J, Rogatko A. Patient specific dosing in a cancer phase I clinical trials. *Statistics in Medicine*. 2001; 20:2079–2090. [PubMed: 11439422]
- Bekele BN, Thall PF. Dose-finding based on multiple toxicities in a soft tissue sarcoma trial. *Journal of the American Statistical Association*. 2004; 99:26–35.
- Bekele BN, Ji Y, Shen Y, Thall PF. Monitoring late-onset toxicity in phase I trials using predicted risks. *Biostatistics*, bf. 2008; 9:442–457.
- Cheung YK, Chappell R. Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics*. 2000; 56:1177–1182. [PubMed: 11129476]
- Cheung YK, Thall PF. Monitoring the rates of composite events with censored data in phase II clinical trials. *Biometrics*. 2002; 58:89–97. [PubMed: 11890331]
- Dunson DB, Neelon B. Bayesian Inference on Order-Constrained Parameters in Generalized Linear Models. *Biometrics*. 2003; 59:286–295. [PubMed: 12926713]
- Extermann M, Bonetti M, Sledge GW, O'Dwyer PJ, Bonomi P, Benson AB. MAX2—a convenient index to estimate the average per patient risk for chemotherapy toxicity; validation in ECOG trials. *European Journal of Cancer*. 2004; 40:1193–8. [PubMed: 15110883]
- Ivanova A, Wang K. Bivariate isotonic design for dose-finding with ordered groups. *Statistics in Medicine*. 2006; 25:2018–2026. [PubMed: 16220476]
- Ivanova A, Kim SH. Dose finding for continuous and ordinal outcomes with monotone objective function: a unified approach. *Biometrics*. 2008 to appear.
- Legedza A, Ibrahim J. Heterogeneity in phase I clinical trials: prior elicitation and computation using the continual reassessment method. *Statistics in Medicine*. 2001; 20:867–882. [PubMed: 11252009]
- Li Y, Bekele BN, Ji Y, Cook JD. Dose-schedule finding in phase I/II clinical trials using a Bayesian isotonic transformation. *Statistics in Medicine*. 2008; 27:4895–4913. [PubMed: 18563789]
- O'Quigley J, Paoletti X. Continual reassessment method for ordered groups. *Biometrics*. 2003; 59:430–440. [PubMed: 12926728]
- Robertson, T.; Wright, FT.; Dykstra, R. *Order restricted statistical inference*. New York: Wiley; 1988. p. 12-26.
- Rogatko A, Babb JS, Wang H, Slifker MJ, Hudes GR. Patient characteristics compete with dose as predictors of acute treatment toxicity in early phase clinical trials. *Clinical Cancer Research*. 2004; 10:4645–4651. [PubMed: 15269136]
- Thall PF, Lee SJ. Practical model-based dose-finding in phase I clinical trials: Methods based on toxicity. *International Journal of Gynecologic Cancer*. 2003; 13:251–261.
- Wijesinha MC, Piantadosi S. Dose-response models with covariates. *Biometrics*. 1995; 51:977–987. [PubMed: 7548713]
- Yuan Z, Chappell R. Isotonic designs for phase I cancer clinical trials with multiple risk groups. *Clinical Trials*. 2004; 1:499–508. [PubMed: 16279290]
- Yuan Z, Chappell R, Bailey H. The Continual Reassessment Method for Multiple Toxicity Grades: A Bayesian Quasi-Likelihood Approach. *Biometrics*. 2007; 63:173–179. [PubMed: 17447942]

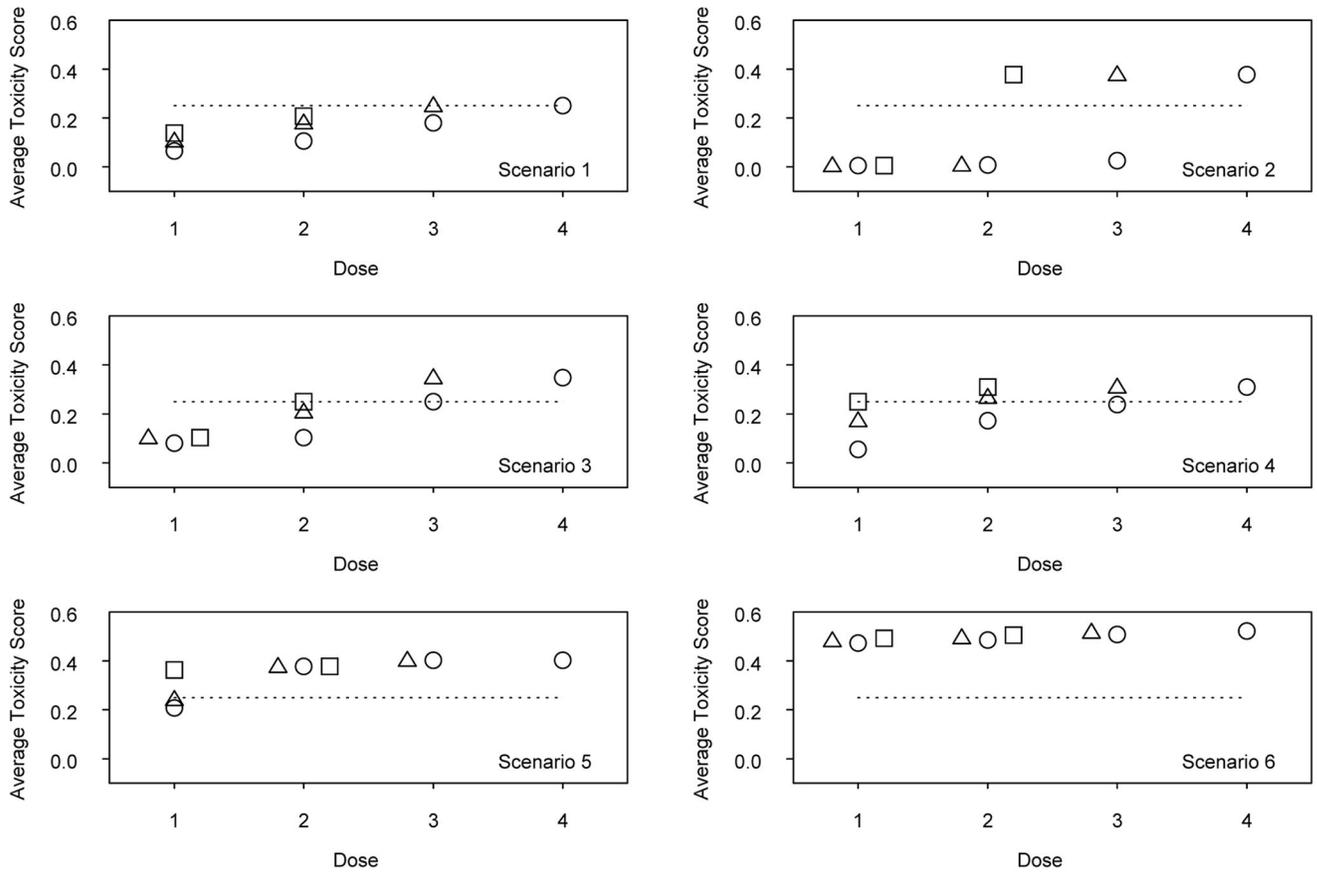


Figure 1.

Scenarios in the simulation study for the multiple myeloma trial. The horizontal and vertical axes are dose and average toxicity score, respectively. The $ATS \equiv 0.25$ is represented by the horizontal dotted line. True ATS's for the low risk group ($40 < GFR \leq 60$ mL/min), moderate risk group ($20 < GFR \leq 40$ mL/min), and high risk group ($GFR \leq 20$ mL/min) are represented by a circle, triangle, and square, respectively. Table 1: Simulation results using the proposed design for dose finding with risk groups. For each scenario we provide the true ATS by risk group (italicized numbers in table) and report the correct selection probabilities and the average number of patients per dose for each risk group.

Simulation results using the proposed design for dose finding with risk groups. For each scenario we provide the true ATS by risk group (italicized numbers in table) and report the correct selection probabilities and the average number of patients per dose for each risk group.

Table 1

Seen	Risk Grp	Dose				None
		1	2	3	4	
1	1	0.07	0.11	0.18	0.25	0
		0.0 (3.0)	3.2 (3.6)	26.7 (5.5)	70.1 (8.9)	
	2	0.11	0.18	0.25	—	0
		4.0 (3.7)	31.5 (6.0)	64.5 (8.4)	—	—
	3	0.14	0.21	—	—	2.2
		26.2 (5.5)	71.6 (6.3)	—	—	—
2	1	0.01	0.01	0.03	0.38	0
		0.0 (3.0)	0.0 (3.0)	71.4 (7.4)	28.6 (7.6)	
	2	0.01	0.01	0.38	—	0
		0.0 (3.0)	70.0 (7.3)	30.0 (7.7)	—	—
	3	0.01	0.38	—	—	0
		59.8 (5.5)	40.2 (6.5)	—	—	—
3	1	0.08	0.1	0.25	0.35	0
		0.2 (3.1)	14.3 (4.5)	56.0 (8.1)	29.5 (5.3)	
	2	0.1	0.21	0.35	—	0
		6.0 (3.8)	67.0 (8.5)	27.0 (5.7)	—	—
	3	0.1	0.25	—	—	0.4
		31.4 (5.4)	68.2 (6.6)	—	—	—
4	1	0.06	0.17	0.24	0.31	0
		3.3 (3.5)	18.3 (5.6)	43.7 (7.0)	34.7 (4.9)	
	2	0.17	0.27	0.31	—	1.4
		30.1 (6.8)	46.5 (7.6)	22.0 (3.4)	—	—
	3	0.25	0.31	—	—	14.9
		62.6 (8.2)	22.5 (2.9)	—	—	—

Selection percentage for each dose/risk-group combination (average # of patients)

Seen	Risk Grp	Dose				
		1	2	3	4	None
5	1	0.21	0.38	0.4	0.4	1.1
		73.0 (11.5)	21.8 (7.9)	3.4 (1.3)	0.7 (0.2)	
	2	0.24	0.38	0.4	—	12.9
		81.4 (13.6)	5.4 (2.9)	0.3 (0.2)	— (—)	
6	3	0.36	0.38	—	—	64.8
		34.5 (7.8)	0.7 (0.5)	— (—)	— (—)	
	1	0.47	0.49	0.51	0.52	92.2
		7.3 (8.9)	0.3 (0.8)	0.2 (0.1)	0.0 (0.0)	
6	2	0.48	0.5	0.52	—	99.4
		0.5 (4.8)	0.1 (0.2)	0.0 (0.0)	— (—)	
	3	0.49	0.51	—	—	100
		0.0 (3.1)	0.0 (0.1)	— (—)	— (—)	

Simulation results using the M-CRM for dose finding with risk groups. For each scenario we provide the true ATS by risk group (italicized numbers in table) and report the correct selection probabilities and the average number of patients per dose for each risk group.

Table 2

Seen	Risk Grp	Dose			
		1	2	3	4
1	1	0.07	0.11	0.18	0.25
	2	0.1 (3.5)	11.2 (5.5)	40.9 (7.1)	47.5 (4.8)
	3	5.4 (4.6)	33.9 (6.4)	58.0 (6.6)	— (—)
2	1	0.01	0.01	0.03	0.38
	2	0.0 (3.0)	0.0 (3.0)	28.6 (4.7)	71.4 (10.3)
	3	0.0 (3.0)	22.8 (4.5)	77.2 (10.5)	— (—)
3	1	0.08	0.1	0.25	0.35
	2	0.1 (3.7)	16.5 (6.5)	68.0 (8.5)	15.4 (2.4)
	3	5.1 (4.7)	53.6 (7.8)	38.7 (5.0)	— (—)
4	1	0.06	0.17	0.24	0.31
	2	0.5 (3.7)	31.7 (8.2)	48.7 (6.8)	19.1 (2.3)
	3	32.8 (8.1)	39.7 (6.3)	22.7 (2.9)	— (—)
None	1	0.14	0.21	—	—
	2	10.6 (4.5)	82.3 (6.8)	— (—)	7.1
	3	37.5 (6.2)	40.2 (3.9)	— (—)	22.3

Selection percentage for each dose/risk-group combination (Average # of Patients)

Seen	Risk Grp	Dose			
		1	2	3	4 None
5	1	0.21 63.7 (13)	0.38 27.1 (6.8)	0.4 1.8 (0.4)	0.4 0.0 (0.0)
	2	0.24 77.5 (13)	0.38 15.0 (4.0)	0.4 1.3 (0.3)	— — (—)
	3	0.36 41.8 (7.1)	0.38 10.4 (1.5)	— — (—)	— 47.8
6	1	0.47 6.3 (7.7)	0.49 0.0 (0.4)	0.51 0.0 (0.0)	0.52 0.0 (0.0)
	2	0.48 7.9 (6.7)	0.5 0.2 (0.3)	0.52 0.0 (0.0)	— — (—)
	3	0.49 14.1 (5.5)	0.51 0.2 (0.3)	— — (—)	— 85.7