The New England Journal of Statistics in Data Science Volume 2, 72–85 (2024)
DOI: https://doi.org/10.51387/22-NEJSDS7

The Total i3+3 (Ti3+3) Design for Assessing Multiple Types and Grades of Toxicity in Phase I Trials

Meizi LIU, Yuan JI*, and Ji LIN

Abstract

Phase I trials investigate the toxicity profile of a new treatment and identify the maximum tolerated dose for further evaluation. Most phase I trials use a binary dose-limiting toxicity endpoint to summarize the toxicity profile of a dose. In reality, reported toxicity information is much more abundant, including various types and grades of adverse events. Building upon the i3+3 design (Liu *et al.*, 2020), we propose the Ti3+3 design, in which the letter "T" represents "total" toxicity. The proposed design takes into account multiple toxicity types and grades by computing the toxicity burden at each dose. The Ti3+3 design aims to achieve desirable operating characteristics using a simple statistics framework that utilizes"toxicity burden interval" (TBI). Simulation results show that Ti3+3 demonstrates comparable performance with existing more complex designs.

KEYWORDS AND PHRASES: Interval design, Multiple toxicity grades, Rule-based design, Toxicity burden.

1. INTRODUCTION

Phase I clinical trials are first-in-human studies evaluating the toxicity profile of a new treatment. Given a series of candidate dose levels, the goal of a phase I trial in oncology is to determine the maximum tolerated dose (MTD), defined as the highest dose having a dose-limiting toxicity (DLT) probability close to or not higher than a target toxicity probability, say $p_T = 30\%$. In most dose-finding studies, a DLT is typically defined as the occurrence of a grade 3 or higher toxicity event according to the Common Terminology Criterion for Adverse Events (CTCAE) by the National Cancer Institute [1]. Toxicity events of lower grades are called moderate toxicities and are often not modeled in oncology dose-finding designs. Instead, most dose-finding designs consider DLT as the only endpoint, including the 3+3 design [2], the continual reassessment method [3], the interval-based designs such as mTPI and mTPI-2 [4, 5], the BOIN and keyboard designs [6, 7], and the i3+3 design [8].

Recently, the FDA Oncology Center of Excellence launched an initiative, named "Project Optimus" FDA (2022) [9], to improve the dose optimization and dose selection paradigm in oncology drug development as the conventional goal of identifying the MTD is no longer applicable for modern molecular targeted agents and immunotherapies. Specifically, these modern agents do not necessarily exhibit monotone dose-response relationships, rendering MTD a potentially sub-optimal dose for patient care. Moreover, emerging evidence shows that these new treatments often induce moderate adverse events rather than DLTs [10, 11]. As noted in Shah *et al.* (2021), it is important to evaluate the negative health impact of different adverse toxicity events, including both DLT and lower-grade toxicity events [12]. For example, patients who experienced a large number of moderate toxicity events may suffer from a comparable toxicity burden as patients with DLTs but minimum moderate toxicity.

To address this challenge, several statistical approaches have been proposed to incorporate a more comprehensive measure of a patient's toxicity burden as the endpoint. Bekele and Thall (2004) first introduced the concept of total toxicity burden, which is the sum of severity weights of all toxicities experienced by a patient [13]. Yuan *et al.* (2007) developed a quasi-likelihood CRM approach based on equivalent toxicity scores by converting the toxicity grades into a single outcome [14]. Lee *et al.* (2009) proposed an alternative measure, the toxicity burden score (TBS), which is estimated by fitting linear mixed-effect models using historical data [15]. Later, Lee *et al.* (2011) developed the continual reassessment method with multiple constraints (CRM-MC), which allows for the specification of various toxicity thresholds with a continuous or ordinal toxicity measure such as the TBS [16]. Van Meter *et al.* (2012) extended the CRM to incorporate toxicity severity using proportional odds models [17]. Ezzalfani *et al.* (2013) introduced the total toxicity profile, defined as the Euclidean norm of the weights of toxicities experienced by a patient, to summarize the overall severity of multiple types and grades of toxicities [18]. More recently, Mu *et al.* (2019) proposed a generalized Bayesian optimal interval design (gBOIN) that extended the BOIN design to account for toxicity grades, binary or continuous toxicity endpoints under a unified framework [19]. All of the

*Corresponding author.

methods aforementioned rely on a model-based or model-assisted inference for dose recommendation.

The concept of toxicity burden has also been implemented beyond phase I dose-finding trials. Hobbs *et al.* (2015) [20] proposed a Bayesian group sequential design using the total toxicity burden and progression-free survival as co-primary endpoints. A randomized Phase II trials (NCT01512589) has been completed using the design [21, 22].

Motivated by these early developments and the use of interval designs such as i3+3 [8], we propose a model-free design called Ti3+3 that also uses toxicity burden to summarize patients' toxicity profile but adopts a simple dose-finding algorithm based on toxicity burden interval (TBI). Utilizing TBI, the Ti3+3 design greatly simplifies the application in practice.

The remainder of the paper is organized as follows. In Section 2, we define the toxicity burden and describe details the proposed Ti3+3 design, including the dose-finding algorithm and MTD selection criteria. In Section 3, we perform simulations to compare the performance of the proposed method with an existing design and present the simulation results as well as a sensitivity analysis in Section 4. We end the paper with a discussion in Section 5.

2. METHODS

2.1 Toxicity Burden

In order to comprehensively evaluate the effects of toxicities to patients, Bekele and Thall (2004) [13] first proposed the toxicity burden as a weighted sum of individual toxicity types and grades, where the severity weights reflect the relative health impact of each grade and type of toxicity. The proposed Ti3+3 design utilizes type-specific and overall toxicity burdens similar to Bekele and Thall (2004), which is described next.

Let $d \in \{1, \ldots, T\}$ denote a set of ascending doses explored in a phase I trial. Assume multiple types of treatment-related toxicity $j \in \{1, \ldots, J\}$ are observed for patients, and each type of toxicity is classified into toxicity grades $k \in \{0, \ldots, K\}$ using a standard reference, such as the CTCAE [1]. For example, neurotoxicity and GI toxicity are two different types of adverse events, each consisting of five grades with grade 0 denoting no toxicity, 1-2 moderate toxicity, 3-4 severe toxicity, and 5 death. Toxicities with grade 5, corresponding to treatment-related death, require trial suspension and direct intervention from the safety committee and therefore are not considered in the proposed design.

Specifically, let $Y_{ij} \in \{0, \ldots, K\}$ denote the observed toxicity grade of type j for patient i; and let $\{X_i = d\}$ denote the event that patient i is treated at dose d. Denote $p_{djk} = Pr(Y_{ij} = k | X_i = d)$ the toxicity probability of grade k for type j at dose d, and apparently $\sum_{k=0}^{K} p_{djk} = 1$, $0 \le p_{djk} \le 1$. The proposed Ti3+3 design relies on a weight matrix $\boldsymbol{W}$ that is elicited through consultation with clinicians. Let $\boldsymbol{W} = \{w_{jk}\}$ be a standardized matrix of weights,

$$\boldsymbol{W} = \begin{pmatrix} w_{10} & \ldots & w_{1K} \\ \vdots & \ddots & \vdots \\ w_{J0} & \ldots & w_{JK} \end{pmatrix},$$

where $\sum_{j,k} w_{jk} = 1$. Here $w_{jk} \ge 0$ quantifies the relative health impact assigned by physicians for the toxicity event of type j and grade k. Denote the row sum $c_j = \sum_{k=0}^{K} w_{jk}$. The magnitudes of $\{c_j\}$'s reflect the relative severity of different types of toxicity events; within each type j, the magnitudes of $\{w_{jk}\}$'s reflect the relative average severity of different grades of toxicity events. These interpretations are exploited in the definitions of the toxicity burdens. In Appendix A, we describe an algorithm to guide the elicitation of the $\boldsymbol{W}$ matrix by statisticians and clinicians. To reflect the belief that higher grade of a type of toxicity is more impactful to a patient's health, we assume the monotonicity $0 = w_{j0} < \cdots < w_{jK}$ for any type j, which implies an increasing toxicity burden for high toxicity grades.

Next, we define a type-specific toxicity burden for toxicity type j at dose d as

$$TB_d^j = \sum_{k=0}^{K} \frac{w_{jk}}{w_{jK}} p_{djk}, \tag{2.1}$$

where $TB_d^j \in (0, 1)$ is analogous to the toxicity probability in DLT-based dose-finding trials. To see this, TB_d^j in (2.1) is a weighted sum of the type-specific toxicity probabilities of different grades, with the weight equal to 1 for the highest grade K, and $\frac{w_{jk}}{w_{jK}}$ for grade $k < K$. Since $w_{jk} < w_{jK}$ for $k < K$, $\frac{w_{jk}}{w_{jK}} < 1$. Therefore, (2.1) implies that TB_d^j is a re-scaled probability of toxicity of the highest grade K for type j at dose d, converting all the lower grade toxicities to the highest grade by a weight factor of $\frac{w_{jk}}{w_{jK}}$. Note that TB_d^j is a parameter. To estimate it, we consider the following statistics, $\widehat{TB}_{id}^j$, the observed type-specific toxicity burden for patient i treated at dose d, given by

$$\widehat{TB}_{id}^j = \sum_{k=0}^{K} \frac{w_{jk}}{w_{jK}} \boldsymbol{I}(Y_{ij} = k) \boldsymbol{I}(X_i = d), \tag{2.2}$$

where $\boldsymbol{I}(\cdot)$ is an indicator function and

$$\boldsymbol{I}(Y_{ij} = k) = \begin{cases} 1, & \text{if patient } i \text{ experiences toxicity type } j \\ & \text{with grade } k, \\ 0, & \text{otherwise.} \end{cases}$$

Here $\widehat{TB}_{id}^j$ is based on observed data $\{Y_{ij}, X_i\}$. Since a patient may experience multiple adverse events associated with multiple types and grades of toxicity, we only use the most severe grade of each type in defining the toxicity burdens. In other words, grade "k" in (2.2) is the most severe toxicity grade among all the toxicity events of type j for patient i. In addition, for patient i,

$$\widehat{TB}_{id} = \sum_{j=1}^{J} c_j \sum_{k=0}^{K} \frac{w_{jk}}{w_{jK}} \boldsymbol{I}(Y_{ij} = k)\boldsymbol{I}(X_i = d)$$

is the observed toxicity burden for patient i. Assuming a total of n_d patients have been treated at the dose d, then TB_d^j can be estimated by

$$\widehat{TB}_d^j = \frac{1}{n_d} \sum_{i=1}^{n_d} \widehat{TB}_{id}^j. \tag{2.3}$$

It is trivial to show that $\widehat{TB}_d^j$ is unbiased, i.e., $E(\widehat{TB}_d^j) = TB_d^j$, assuming w_{jk} is given and fixed.

Finally, given the type-specific toxicity burdens TB_d^j, the overall toxicity burden at dose d can be defined as

$$TB_d = \sum_{j=1}^{J} c_j TB_d^j = \sum_{j=1}^{J} c_j \sum_{k=0}^{K} \frac{w_{jk}}{w_{jK}} p_{djk}, \tag{2.4}$$

where $c_j = \sum_{k=0}^{K} w_{jk}$ are constants that reflect the relative severity between toxicity types. Apparently, $\sum_j c_j = \sum_{j,k} w_{jk} = 1$. Similarly, the observed overall toxicity burden at dose d is calculated as follows

$$\widehat{TB}_d = \sum_{j=1}^{J} c_j \widehat{TB}_d^j. \tag{2.5}$$

An interval dose-finding design like i3+3 needs to specify the target toxicity probability (of DLTs), p_T, and an equivalence interval (EI) to facilitate the dose-finding decisions. Similarly, the proposed Ti3+3 design needs a target toxicity burden (TTB) and an associated EI for TTB. To start, we define a type-specific TTB^j for toxicity type j. The TTB^j can be viewed as the toxicity target for the MTD if only toxicity events of type j are considered as outcome; similarly, the EI^j is the equivalence interval for the MTD when only type j toxicity events are modeled. For example, assume that there are two types of toxicities $J = 2$, and the targets for type-specific toxicity burden could be set to $TTB^1 = 0.3$ and $TTB^2 = 0.25$. In addition, denote $EI^j = (TTB^j - \epsilon_1^j, TTB^j + \epsilon_2^j)$ the equivalence interval of TTB^j for toxicity type j, which is an interval centered around the TTB^j. The upper bound and lower bound of the EI^j reflect the highest and lowest value of toxicity burden that the clinicians would consider to be acceptable for MTD if only type j events are considered. Similar to how the definition of the DLT is tailored case-by-case for conventional dose-finding trials, these values should be elicited with the clinical team based on the specific context of a trial. Given the specified TTB^j and Equation 2.4, an overall target toxicity burden, denoted as TTB, is defined as a weighted sum of TTB^j by

$$TTB = \sum_{j=1}^{J} c_j TTB^j. \tag{2.6}$$

And similarly, the equivalence interval for the overall toxicity burden, EI, is derived as a weighted average of EI^j with weights c_j. That is, $EI = \sum_{j=1}^{J} c_j EI^j$.

In summary, the main effort in applying the proposed Ti3+3 is in the initial setup, which requires the specification of the weight matrix, $\boldsymbol{W}$, the target toxicity burden, TTB^j, and the equivalence interval, EI^j, for each toxicity type j. Once they are determined, dose finding proceeds based on a simple algorithm next.

2.2 Dose-Finding Algorithm

Assume patients are assigned sequentially in cohorts, starting with the lowest dose. The next cohort of patients will not be enrolled until toxicity outcomes have been observed for the present cohort. Suppose dose d is the current dose used to treat patients and n_d patients have been treated at the dose, Ti3+3 extends the dose-finding algorithm in the i3+3 design to accommodate the new toxicity burden endpoints. The dose-finding algorithm of Ti3+3 are first applied to toxicity type j to generate a type-specific decision, denoted as $\mathcal{A}^j \in \{$"E", "S", "D"$\}$, where "E", "S", and "D" denote "Escalation", "Stay", and "De-escalation", respectively, and to the overall toxicity burden to generate $\mathcal{A}^0 \in \{$"E", "S", "D"$\}$. The next cohort of patients is assigned to the minimum dose level indicated by decisions $\mathcal{A}^j$'s and $\mathcal{A}^0$.

Below we introduce the dose-finding algorithm.

Algorithm 1 Ti3+3 dose-finding algorithm.

for $j \in \{0, 1, \ldots, J\}$ **do**
  if $\widehat{TB}_d^j$ is below the EI^j **then**
    dose d is considered safe and escalate ("E") to dose $(d+1)$ (i.e., $\mathcal{A}^j =$"E")
  else if $\widehat{TB}_d^j$ is inside the EI^j **then**
    dose d is considered to be close to the MTD and stay ("S") at dose d (i.e., $\mathcal{A}^j =$"S")
  else if $\widehat{TB}^j$ is above the EI^j **then**
    if $\widehat{TB}_{d,-1}^j$ (defined next) is below the EI^j **then**
      the decision is to stay ("S") at dose d (i.e., $\mathcal{A}^j =$"S")
    else if $\widehat{TB}_{d,-1}^j$ is inside or above the EI^j **then**
      dose d is considered toxic, and the decision is to de-escalate ("D") to dose $(d-1)$ (i.e., $\mathcal{A}^j =$"D")
    end if
  end if
end for
The next cohort is assigned based on the decisions $\{\mathcal{A}^j : j = 0, 1, \ldots, J\}$. Since decisions "$E$", "$S$", "$D$" correspond to doses $(d+1)$, d, and $(d-1)$, respectively, we denote $\mathcal{B}^j = 1, 0,$ or -1 if $\mathcal{A}^j =$ "E", "S", or "D", respectively. Assign the next cohort of patients to $d + \min_j\{\mathcal{B}^0, \ldots, \mathcal{B}^J\}$.

Algorithm 1 requires computing a new quantity $\widehat{TB}_{d,-1}^j$, which is defined as a hypothetical observed toxicity burden assuming the patient in the cohort who experienced the lowest toxicity burden would experience no toxicity at all. In other words, if we remove all the toxicity events from the

Table 1. The dose-finding algorithm of the Ti3+3 design.

Condition	Decision $\mathcal{B}^j$ $(\mathcal{A}^j)$
$\widehat{TB}_d^j < EI^j$	1 (E^*)
$\widehat{TB}_d^j \in EI^j$	0 (S)
$\widehat{TB}_d^j > EI^j$ & $\widehat{TB}_{d,-1}^j < EI^j$	0 (S)
$\widehat{TB}_d^j > EI^j$ & $\widehat{TB}_{d,-1}^j \in EI^j$	-1 (D^*)
$\widehat{TB}_d^j > EI^j$ & $\widehat{TB}_{d,-1}^j > EI^j$	-1 (D^*)
Final Decision:	$d + \min_j\{\mathcal{B}^j\}$

$*$: when d is the highest dose ($d = T$) or the lowest dose ($d = 1$), the decisions D and E should be replaced by S accordingly.

patient with the lowest burden from the data at dose d, $\widehat{TB}_{d,-1}^j$ is the new observed toxicity burden for type j. This idea is similar to that of the i3+3 design, the difference being that i3+3 only considers DLT while Ti3+3 considers different grades and types of toxicity.

Table 1 summarizes the decisions of Ti3+3. The type-specific decision $\mathcal{B}^j(\mathcal{A}^j)$ for each toxicity type j (or overall burden if $j = 0$) is listed, and the final decision is to assign the next cohort to dose $(d + \min_j\{\mathcal{B}^j\})$. Similar to i3+3, in Ti3+3, there is a special rule. When $\widehat{TB}_d^j > EI^j$ and $\widehat{TB}_{d,-1}^j < EI^j$, it indicates that removing the toxicity events from a single patient in the observed data renders the observed toxicity burden from being above the equivalence interval to below the interval. In other words, changing one-patient worth of information of the data would result in a reversal of the decision from de-escalation (since $\widehat{TB}_d^j > EI^j$) to escalation since $(\widehat{TB}_{d,-1}^j < EI^j)$. This implies that the information in the observed data is sparse and small change of the data results in reversal of decisions. Therefore, Ti3+3, in this case, does not de-escalate due to lack of confidence in the data, and instead, continues to treat patients at the current dose, i.e., "S" stay. The other rules in Table 1 are straightforward, following the idea of de-escalation if observed toxicity burden is above the EI, stay if inside the EI, or escalation if below the EI.

In Table 2, we provide three examples to illustrate the proposed algorithm. Suppose there are two types of toxicities, and a pre-determined standardized weight matrix is given below

$$\boldsymbol{W} = \begin{pmatrix} 0 & 0.03 & 0.11 & 0.17 & 0.42 \\ 0 & 0.03 & 0.03 & 0.07 & 0.14 \end{pmatrix}.$$

Moreover, assume $TTB^1 = TTB^2 = TTB = 0.30$ and the $EI^1 = EI^2 = EI = (0.25, 0.33)$. Based on the observed toxicity types and grades from patients, $\widehat{TB}_{id}$ and $\widehat{TB}_d$ can be calculated based on (2.5). In case 1, $n_d = 3$ patients are treated at the current dose d, and $\widehat{TB}_d = 0.34$, which is greater than the upper bound 0.33. However, since $\widehat{TB}_{d,-1} = 0.24$ falls below the EI, according to the Ti3+3 algorithm, the decision is, $\mathcal{A}^0$ = "S", stay at the current dose. Therefore, even though $\widehat{TB}_d = 0.34$ is above the EI, the decision is to enroll more patients at the same dose d since $\widehat{TB}_{d,-1}$ is below EI. We see that there are only three patients at dose d and the data is sparse. In the second case, $n_d = 5$ and $\widehat{TB}_d = 0.34$, which is greater than 0.33. And $\widehat{TB}_{d,-1} = 0.28$ falls inside of the EI, therefore, $\mathcal{A}^0$ = "D", de-escalate to the next lower dose $(d-1)$. In the third case, the two quantities $\widehat{TB}_d$, $\widehat{TB}_{d,-1}$ are the same and below the EI, and the decision is then $\mathcal{A}^0$ = "E", escalate to the dose $(d+1)$. The examples demonstrate the simplicity of the proposed decision rules using the overall toxicity burden. Algorithm 1 applies these rules to each toxicity type j as well.

In addition, the Ti3+3 design consists of a few safety rules for practical and ethical concerns. Again, these safety rules are applied iteratively to each type-specific and overall toxicity burdens.

- Safety rule 1 (early termination): At any moment during the trial, if $n_1 \geq 3$, and $Pr(TB_1^j > TTB^j|\text{data}) > 0.95$ or $Pr(TB_1 > TTB|\text{data}) > 0.95$, terminate the trial due to excessive toxicity. This rule stops the trial whenever the lowest dose $(d = 1)$ is deemed overly toxic.
- Safety rule 2 (dose exclusion): At any moment during the trial, suppose the current dose is d. If $n_d \geq 3$, and $Pr(TB_d^j > TTB^j|\text{data}) > 0.95$ or $Pr(TB_d > TTB|\text{data}) > 0.95$, remove dose d and higher doses from the trial. In other words, if a sufficient number $(n_d \geq 3)$ of patients has been treated at a dose d, and their outcomes suggest that dose d is deemed overly toxic, dose d and higher doses are excluded from the trial. Any future escalation to dose d will be changed to "S", stay.

The calculation of the posterior distributions $Pr(TB_d^j > TTB^j|\text{data})$ and $Pr(TB_d > TTB|\text{data})$ are discussed below.

2.3 A working model and MTD selection

Once all patients finish their followup at the end of a trial, the Ti3+3 design first selects the type-specific MTD, denoted as d_j^*, and the MTD based on the overall toxicity burden, denoted as d_0^*.

Next, we propose a working statistical model to calculate the posterior probabilities $Pr(TB_d^j > TTB^j|\text{data})$ and $Pr(TB_d > TTB|\text{data})$ and select the MTD based on the observed data. Recall p_{djk} denote the probability of toxicity grade k for type j at dose d. For a given dose d, assume different types of toxicities are independent. Let $\boldsymbol{p_{dj}} = \{p_{dj0} \ldots, p_{djK}\}$ represent the vector of the probabilities associated with different toxicity grades for type j at dose d, and $\boldsymbol{y_{dj}} = \{y_{dj0} \ldots, y_{djK}\}$ the vector of patient counts across different grades, i.e., $y_{djk} = \sum_i \boldsymbol{I}(Y_{ij} = k)\boldsymbol{I}(X_i = d)$. Then $\boldsymbol{y_{dj}}$ is assumed to follow a the multinomial sampling

Table 2. Three hypothetical cases to illustrate the dose-finding decisions of the Ti3+3 design for a trial with two toxicity types and five grades. Notation: 1) Tox data: (k_1, k_2) if the patient experiences grade k_1 of the first type of toxicity and grade k_2 of the second type of toxicity; 2) $\widehat{TB}_{id}$: observed toxicity burden for patient i who is treated on dose d; 3) $\widehat{TB}_d$:observed toxicity burden for the current dose d; 4) $\widehat{TB}_{d,-1}$: $\widehat{TB}_d$ calculated assuming the patient with the lowest $\widehat{TB}_{id}$ experienced no toxicity. Suppose $TTB = 0.30$, and the EI is $(0.25, 0.33)$. The patient with the lowest $\widehat{TB}_{id}$ in each case is ***bolded****.*

Case #	Patient #	Tox data	$\widehat{TB}_{id}$	$\widehat{TB}_d$	$\widehat{TB}_{d,-1}$	$\mathcal{A}^0$
	1	**(2, 3)**	**0.33**			
1	2	(3, 2)	0.35	0.34	0.23	S
	3	(3, 2)	0.35	$0.34 > EI$	$0.23 < EI$	
	1	(2, 2)	0.25			
	2	**(2, 1)**	**0.25**			
2	3	(2, 3)	0.33	0.34	0.28	D
	4	(3, 0)	0.30	$0.34 > EI$	$0.28 \in EI$	
	5	(3, 4)	0.56			
	1	**(0, 0)**	**0.00**			
3	2	(1, 1)	0.11	0.07	0.07	E
	3	(1, 2)	0.11	$0.07 < EI$	$0.07 < EI$	

distribution given by

$$\boldsymbol{y_{dj}}|\boldsymbol{p_{dj}} \sim \text{Multinomial}(\boldsymbol{p_{dj}}, n_d), \tag{2.7}$$

where for any j, $\sum_{k=0}^{K} p_{djk} = 1$, $p_{djk} \in (0,1)$, and $\sum_{k=0}^{K} y_{djk} = n_d$. Assume a conjugate Dirichlet prior distribution of $\boldsymbol{p_{dj}}$, i.e.,

$$\boldsymbol{p_{dj}} \sim \text{Dirichlet}(\boldsymbol{\alpha}), \tag{2.8}$$

where $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_K)$ are positive values. We set $\alpha_0 = \cdots = \alpha_K = 0.1$ for different toxicity types across doses. Following Morita *et al.* (2011) [23], since $\sum_{k=0}^{4} \alpha_k = 0.5$, a small value, the Dirichlet prior in (2.8) is deemed vague and has little impact on the posterior distribution given by

$$\boldsymbol{p_{dj}}|\boldsymbol{y_{dj}} \sim \text{Dirichlet}(y_{dj0} + \alpha_0, \ldots, y_{djK} + \alpha_K). \tag{2.9}$$

Therefore, the posterior of TB_d^j and TB_d can be computed numerically by sampling $f_j(\boldsymbol{p_{dj}}|\boldsymbol{y_{dj}})$ in 2.9. To calculate the posterior probabilities used in safety rules, we have

$$Pr(TB_d^j > TTB^j|\text{data}) \approx \frac{1}{S}\sum_{s=1}^{S} \boldsymbol{I}(\sum_{k=0}^{K} \frac{w_{jk}}{w_{jK}} p_{djk}^{(s)} > TTB^j), \tag{2.10}$$

and

$$Pr(TB_d > TTB|\text{data}) \approx \frac{1}{S}\sum_{s=1}^{S} \boldsymbol{I}(\sum_{j=1}^{J} c_j \sum_{k=0}^{K} \frac{w_{jk}}{w_{jK}} p_{djk}^{(s)} > TTB), \tag{2.11}$$

where $p_{djk}^{(s)}$ is the s-th random draw from $f_j(\boldsymbol{p_{dj}}|\boldsymbol{y_{dj}})$.

To impose monotonicity assumption of dose-toxicity relationship, we apply isotonic regression to the posterior means of TB_d^j and TB_d via the pool adjacent violators algorithm [24]. Let $\widetilde{TB}_d^j$ and $\widetilde{TB}_d$ be the isotonic transformed posterior means for all dose levels. Among all the tried doses ($n_d > 0$) for which satisfy the safety rules, the estimated MTDs based on the type-specific and overall toxicity burden are defined as

$$d_j^* = \arg\min_d |\widetilde{TB}_d^j - TTB|, j = 0, 1, \ldots, J. \tag{2.12}$$

If more than one dose of d_j^* exists, only one dose is selected based on following rules:

1. If $\widetilde{TB}_{d_j^*} > TTB^j$, choose the lowest dose among the tied doses as the final d_j^*;
2. If $\widetilde{TB}_{d_j^*} \leq TTB^j$, choose the highest dose among the tied doses as the final d_j^*.

Lastly, the final MTD is the smallest dose among the estimated type-specific and overall MTDs, i.e.,

$$d^* = min\{d_j^*, j = 0, .., J\}. \tag{2.13}$$

3. SIMULATIONS

3.1 Comparison with CRM-MC

We simulate clinical trials to assess the operating characteristics of the Ti3+3 design. We first compare to the CRM-MC method by Lee *et al.* (2011) [16]. The Ti3+3 design requires close collaboration between statisticians and clinicians to define the numerical weights $\boldsymbol{W}$ and the EI's at the study design stage. For simulation purpose, we adopt the general setting of the bortezomib trial described in Lee *et al.* (2011) [16]. Under the bortezomib trial, two main types ($J = 2$) of toxicities are identified as related to the treatment. The first type ($j = 1$) is neuropathy and the second ($j = 2$) is low platelet count. The standardized matrix of weights $\boldsymbol{W}$, provided in Lee *et al.* (2011), is given by:

$$\boldsymbol{W} = \begin{pmatrix} 0 & 0.03 & 0.11 & 0.17 & 0.42 \\ 0 & 0.03 & 0.03 & 0.07 & 0.14 \end{pmatrix}.$$

The first and second rows correspond to the weights for grade 0 to grade 4 of each type, neuropathy and low platelet count, respectively. The DLT is defined as the occurrence of a grade 3 or 4 neuropathy or a grade 4 low platelet count. Given this matrix and the definition of TB, a patient has a $\widehat{TB}_{id}$ value 0.30 for a grade 3 neuropathy or 0.27 for a grade 4 low platelet count. And a patient experiencing a grade 3 neuropathy has similar $\widehat{TB}_{id} = 0.30$ as a patient experiencing a grade 2 neuropathy plus a grade 3 low platelet count $\widehat{TB}_{id} = 0.31$. These severity weights reflect the relative clinical importance between different toxicity types and grades, and less severe toxicities are given lower weights. In addition, $c_1 = \sum_{k=0}^{4} w_{1k} = 0.73$ and $c_2 = \sum_{k=0}^{4} w_{2k} = 0.27$, which implies that overall neuropathy is a much more severe type of toxicity than low platelet count.

To implement the Ti3+3 design, the TTB^1, TTB^2, and TTB are set at 0.30, which suggests (by comparing to W) that the occurrence of a grade 3 or higher neuropathy, or a grade 4 low platelet count are considered clinically unacceptable. Moreover, the equivalence intervals used are $EI^1 = EI^2 = EI = (0.25, 0.35)$. For the CRM-MC method, the toxicity burden of patient i is calculated as $T_i = 0.03\boldsymbol{I}(Y_{i1} = 1) + 0.11\boldsymbol{I}(Y_{i1} = 2) + 0.17\boldsymbol{I}(Y_{i1} = 3) + 0.42\boldsymbol{I}(Y_{i1} = 4) + 0.03\boldsymbol{I}(Y_{i2} = 1, 2) + 0.07\boldsymbol{I}(Y_{i2} = 3) + 0.14\boldsymbol{I}(Y_{i2} = 4)$, where Y_{i1} and Y_{i2} are the grades of toxicity type 1 and 2, respectively, for the patient. CRM-MC applies primary and secondary constraints $Pr(T_i \geq 0.25|d) \leq 0.10$ and $Pr(T_i \geq 0.17|d) \leq 0.25$, respectively, to decide the dose for next patients. The vector of scaled doses is obtained via backward substitution with dose 3 as the prior guess of MTD, 0.08 as the indifference interval parameter, 0.69 as the prior median of the slope parameter, and a probit model with an intercept equal to 3. The prior distributions of the probit model slope parameter β and thresholds $(\gamma_l - \gamma_{l-1})$ follow independent exponential distributions with mean 1, as suggested by the authors. Refer to Lee *et al.* (2011) for more details.

Following the original bortezomib trial, Lee *et al.* (2011) implemented a simulation with a sample size 18, cohort size 1, and a starting dose at level 3. We consider a more common simulation setup and apply the following parameters to both designs. The sample size is fixed at 30 with a total of 10 cohorts. A total of five dose levels are investigated, and both designs, Ti3+3 and CRM-MC will start at the lowest dose level. We construct a total of eight scenarios, including the first four scenarios from Lee *et al.* (2011) and additional four scenarios. According to Ti3+3, the true MTD is defined as the highest dose with the TB_d^j falls below or inside the equivalence interval of TTB^j for all type-specific and overall toxicity burdens, that is, the dose with $TB_d^j \leq 0.35$ for $j = 1, 2$ and $TB_d \leq 0.35$. Different from Ti3+3, the CRM-MC design selects MTD as the highest dose that satisfies the primary and secondary toxicity constraints. Therefore, CRM-MC and Ti3+3 sometimes would consider different doses in the same scenario as the true MTD, for example, in scenario 1, Ti3+3 considers dose level 3 as the true MTD while based on CRM-MC, dose level 2 is the true MTD. A full description of all dosing scenarios is provided in Appendix B. For each scenario, we simulate 1,000 trials.

The simulation results are presented in Table 3, which shows the percentage of recommending a particular dose (Selection %) and the percentage of patients assigned to each dose level (Allocation %). A desirable design should demonstrate a good balance between the ability to correctly identify the MTD and patient safety. Compared with the CRM-MC design, Ti3+3 has higher percentage of correct selection (PCS) among scenarios where both design consider the same dose as the true MTD (scenarios $2, 3, 4, 6$, and 8). In terms of patient allocation, Ti3+3 seems to perform better at assigning patients to the target dose (PCA) in scenarios 2-6, and is less likely to assign patients to a dose above the MTD (POA) in majority of the eight scenarios. Overall, the operating characteristics of our proposed Ti3+3 design are comparable to the CRM-MC design, which relies on a model-based inference for dose assignment decisions.

3.2 Comparison with gBOIN

Additionally, we evaluate the performance of Ti3+3 by comparing it with the gBOIN design by Mu *et al.* (2019) [19]. The gBOIN design generalizes the BOIN design and provides a unified framework to incorporate non-binary toxicity outcomes. We follow the simulation settings described in Section 3.2 in Mu *et al.* (2019). Only one type of toxicity is considered, and the standardized weight matrix is defined as

$$\boldsymbol{W} = \begin{pmatrix} 0 & 0 & 0.16 & 0.33 & 0.50 \end{pmatrix}.$$

In words, grades 0 and 1 are of no concern, grade 2 toxicity is considered equivalent to half grade 3 toxicity, and grade 4 toxicity is equivalent to one and half grade 3 toxicity. Based on Mu *et al.* (2019), the TTB is 0.31, and the recommended default escalation and de-escalation boundaries are $\lambda_e^* = 0.249$ and $\lambda_d^* = 0.377$. The same set of target burden and EI is adopted in Ti3+3. A total sample of 30 patients and a cohort size of 3 is used in the simulation, with a starting dose at dose level 1. The details of the ten dose-toxicity scenarios is provided in Appendix B.

Table 4 shows the results based on 4,000 simulated trials. In general, the two designs demonstrate comparable operating characteristics, and Ti3+3 shows superior performance in terms of trial safety. Even though gBOIN yields higher PCS in scenarios 1 through 7, the differences in PCS are no greater than 0.05 except in scenario 7. And Ti3+3 generates higher PCS in scenarios 8, 9 and 10. It is worthnoting that Ti3+3 is less likely to recommend a overly toxic dose as the MTD across all scenarios. The patient allocation of the two designs are very close as shown in Table 4. The Ti3+3 yields higher or similar PCA in 7 out of 10 scenarios, and shows consistently lower POA across all scenarios. The POS and POA are two important safety metrics considered in early phase trials, and Ti3+3 shows relatively strong performance in the safety metrics.

*Table 3. Performance of the Ti3+3 design compared with CRM-MC. The $TTB^1 = TTB^2 = TTB$ is 0.30, and the $EI^1 = EI^2 = EI$ is $(0.25, 0.35)$. The MTD in each scenario is in **bold**.*

	Selection %					Allocation %				
	1	2	3	4	5	1	2	3	4	5
Scenario 1										
TB^1	*0.10*	*0.21*	***0.28***	*0.34*	*0.42*					
TB^2	*0.11*	*0.23*	***0.34***	*0.46*	*0.59*					
TB	*0.11*	*0.22*	***0.30***	*0.37*	*0.46*					
CRM-MC	0.13	**0.71**	0.15	0.01	0.00	0.25	**0.53**	0.18	0.03	0.00
Ti3+3	0.05	0.58	**0.35**	0.02	0.00	0.23	0.49	**0.24**	0.04	0.00
Scenario 2										
TB^1	*0.10*	*0.11*	***0.21***	*0.29*	*0.39*					
TB^2	*0.08*	*0.11*	***0.23***	*0.46*	*0.57*					
TB	*0.10*	*0.11*	***0.22***	*0.34*	*0.44*					
CRM-MC	0.00	0.15	**0.74**	0.10	0.00	0.14	0.24	**0.45**	0.15	0.02
Ti3+3	0.00	0.02	**0.80**	0.18	0.00	0.11	0.20	**0.48**	0.19	0.01
Scenario 3										
TB^1	*0.11*	*0.10*	0.12	***0.21***	*0.29*					
TB^2	*0.08*	*0.11*	0.16	***0.23***	*0.46*					
TB	*0.10*	*0.11*	0.13	***0.22***	*0.34*					
CRM-MC	0.00	0.04	0.29	**0.58**	0.09	0.14	0.16	0.24	**0.34**	0.12
Ti3+3	0.00	0.00	0.07	**0.77**	0.16	0.12	0.14	0.23	**0.38**	0.14
Scenario 4										
TB^1	*0.10*	*0.11*	*0.12*	*0.14*	***0.21***					
TB^2	*0.05*	*0.08*	*0.16*	*0.21*	***0.23***					
TB	*0.09*	*0.10*	*0.13*	*0.16*	***0.22***					
CRM-MC	0.00	0.03	0.15	0.32	**0.50**	0.13	0.16	0.20	0.23	**0.29**
Ti3+3	0.00	0.00	0.05	0.21	**0.74**	0.11	0.13	0.19	0.24	**0.33**
Scenario 5										
TB^1	*0.22*	***0.32***	*0.36*	*0.48*	*0.52*					
TB^2	*0.17*	***0.25***	*0.45*	*0.52*	*0.60*					
TB	*0.21*	***0.30***	*0.39*	*0.49*	*0.54*					
CRM-MC	**0.73**	0.26	0.01	0.00	0.00	**0.64**	0.28	0.06	0.01	0.00
Ti3+3	0.38	**0.57**	0.05	0.00	0.00	0.52	**0.41**	0.07	0.01	0.00
Scenario 6										
TB^1	*0.13*	***0.23***	*0.30*	*0.34*	*0.45*					
TB^2	*0.11*	***0.21***	*0.43*	*0.44*	*0.79*					
TB	*0.13*	***0.22***	*0.33*	*0.37*	*0.54*					
CRM-MC	0.24	**0.66**	0.10	0.00	0.00	0.33	**0.47**	0.15	0.04	0.01
Ti3+3	0.02	**0.73**	0.25	0.01	0.00	0.25	**0.52**	0.21	0.01	0.00
Scenario 7										
TB^1	*0.10*	*0.12*	***0.18***	*0.31*	*0.48*					
TB^2	*0.18*	*0.28*	***0.29***	*0.58*	*0.73*					
TB	*0.12*	*0.16*	***0.21***	*0.38*	*0.55*					
CRM-MC	0.04	0.32	**0.60**	0.04	0.00	0.14	0.31	**0.46**	0.08	0.01
Ti3+3	0.07	0.45	**0.46**	0.02	0.00	0.28	0.39	**0.27**	0.06	0.00
Scenario 8										
TB^1	*0.04*	*0.06*	*0.14*	***0.25***	*0.29*					
TB^2	*0.10*	*0.12*	*0.16*	***0.17***	*0.44*					
TB	*0.06*	*0.08*	*0.15*	***0.23***	*0.33*					
CRM-MC	0.01	0.06	0.26	**0.56**	0.11	0.08	0.11	0.24	**0.40**	0.18
Ti3+3	0.00	0.00	0.06	**0.71**	0.23	0.12	0.15	0.21	**0.36**	0.16

4. SENSITIVITY ANALYSIS

Sensitivity analysis is conducted to further evaluate the Ti3+3 design using the first eight scenarios considered in Section 3.2. First, we investigate the effect of EI length on the performance of the proposed method. Specifically, with TTB^j and TTB fixed at 0.30, we select four different EI's (represented as Cases A, B, C, and D), and for each EI value and each of the eight scenarios, we simulate 1,000 trials each with a cohort size 3 and total sam-

Table 4. Performance of the Ti3+3 design compared with gBOIN. The TTB is 0.31, and the EI is $(0.249, 0.377)$. *The MTD in each scenario is in* ***bold***.

	Selection %						Allocation %					
	1	2	3	4	5	6	1	2	3	4	5	6
Scenario 1												
TB	*0.08*	*0.13*	*0.22*	***0.32***	*0.50*	*0.70*						
gBOIN	0.00	0.02	0.29	**0.56**	0.12	0.01	0.12	0.17	0.28	**0.30**	0.12	0.01
Ti3+3	0.00	0.07	0.34	**0.50**	0.08	0.00	0.12	0.20	0.31	**0.27**	0.09	0.01
Scenario 2												
TB	*0.05*	*0.09*	*0.19*	***0.28***	*0.47*	*0.66*						
gBOIN	0.00	0.00	0.16	**0.66**	0.18	0.00	0.12	0.15	0.25	**0.33**	0.14	0.01
Ti3+3	0.00	0.04	0.24	**0.62**	0.11	0.00	0.12	0.16	0.27	**0.33**	0.11	0.01
Scenario 3												
TB	*0.11*	***0.26***	*0.33*	*0.44*	*0.55*	*0.75*						
gBOIN	0.02	**0.47**	0.36	0.13	0.00	0.00	0.21	**0.41**	0.26	0.11	0.02	0.00
Ti3+3	0.12	**0.43**	0.33	0.10	0.01	0.00	0.21	**0.41**	0.26	0.10	0.02	0.00
Scenario 4												
TB	*0.07*	*0.22*	***0.30***	*0.40*	*0.52*	*0.71*						
gBOIN	0.01	0.28	**0.47**	0.23	0.01	0.00	0.16	0.30	**0.36**	0.15	0.03	0.00
Ti3+3	0.05	0.32	**0.46**	0.16	0.01	0.00	0.18	0.34	**0.31**	0.13	0.03	0.00
Scenario 5												
TB	*0.00*	*0.04*	*0.06*	*0.07*	*0.11*	***0.22***						
gBOIN	0.00	0.00	0.00	0.00	0.04	**0.96**	0.10	0.10	0.11	0.11	0.15	**0.42**
Ti3+3	0.00	0.00	0.00	0.01	0.08	**0.91**	0.10	0.10	0.11	0.12	0.16	**0.41**
Scenario 6												
TB	***0.30***	*0.42*	*0.53*	*0.67*	*0.77*	*0.86*						
gBOIN	**0.69**	0.29	0.01	0.00	0.00	0.00	**0.66**	0.27	0.06	0.01	0.00	0.00
Ti3+3	**0.62**	0.19	0.01	0.00	0.00	0.00	**0.69**	0.26	0.05	0.00	0.00	0.00
Scenario 7												
TB	*0.13*	***0.30***	*0.38*	*0.49*	*0.60*	*0.78*						
gBOIN	0.11	**0.55**	0.28	0.06	0.00	0.00	0.30	**0.43**	0.21	0.06	0.01	0.00
Ti3+3	0.15	**0.54**	0.25	0.05	0.01	0.00	0.31	**0.42**	0.20	0.05	0.01	0.00
Scenario 8												
TB	*0.05*	*0.16*	*0.21*	***0.29***	*0.37*	*0.50*						
gBOIN	0.00	0.06	0.22	**0.41**	0.28	0.04	0.09	0.19	0.24	**0.28**	0.15	0.05
Ti3+3	0.00	0.07	0.25	**0.41**	0.25	0.02	0.13	0.19	0.25	**0.24**	0.15	0.04
Scenario 9												
TB	*0.11*	***0.30***	*0.38*	*0.49*	*0.6*	*0.78*						
gBOIN	0.00	**0.91**	0.00	0.00	0.00	0.00	0.11	**0.76**	0.14	0.00	0.00	0.00
Ti3+3	0.00	**0.94**	0.06	0.00	0.00	0.00	0.11	**0.79**	0.11	0.00	0.00	0.00
Scenario 10												
TB	*0.05*	*0.16*	*0.21*	***0.29***	*0.37*	*0.50*						
gBOIN	0.00	0.00	0.03	**0.71**	0.23	0.03	0.10	0.12	0.17	**0.44**	0.13	0.04
Ti3+3	0.00	0.00	0.06	**0.79**	0.14	0.01	0.10	0.12	0.19	**0.47**	0.09	0.03

ple size 30. Simulation results with varying EI lengths are shown in Table 5. Note that in Case B ($EI = (0.20, 0.40)$) and D ($EI = (0.25, 0.35)$), EIs are symmetric around the TTB, whereas in Case A ($EI = (0.20, 0.35)$) and C ($EI = (0.25, 0.40)$), the EIs are asymmetric around TTB. For all scenarios, patient safety (the percentage of patients treated at or below the true MTD) and PCS tend to improve with wider EIs. This is because a wider EI allows a wider range of doses to be considered as the MTD. However, the EI cannot be too wide to become clinically meaningless. Moreover, across all scenarios, we observe that Case A allocates fewer patients to doses over MTD than Case B and C since the upper bound of the EI in A is smaller than that of the EI in B and C. The overall performances observed across the four cases are comparable, and the proposed Ti3+3 design seems robust against various choices of EI lengths.

We fix the total sample size of 30 and conduct another sensitivity analysis using Ti3+3 with cohort sizes 1, 2, or 3. Results with different cohort sizes are shown in Table 6. For all scenarios, the results for cohort sizes 1 and 2 are less

Table 5. Sensitivity analysis results of the Ti3+3 with different EI lengths. The TTB is 0.30. The different EIs: $A = (0.20, 0.35)$, $B = (0.20, 0.40)$, $C = (0.25, 0.40)$, $D = (0.25, 0.35)$. The MTD in each scenario is in ***bold****.*

	Selection %						Allocation %					
	1	2	3	4	5	6	1	2	3	4	5	6
Scenario 1												
TB	*0.08*	*0.13*	*0.22*	*0.32*	*0.50*	*0.70*						
A	0.01	0.10	**0.46**	**0.38**	0.04	0.00	0.14	0.25	**0.36**	**0.21**	0.05	0.00
B	0.02	0.09	**0.44**	**0.40**	0.05	0.00	0.14	0.23	**0.35**	**0.23**	0.05	0.00
C	0.01	0.09	0.35	**0.49**	0.07	0.00	0.12	0.20	0.30	**0.28**	0.09	0.01
D	0.02	0.08	0.37	**0.45**	0.08	0.00	0.13	0.20	0.31	**0.26**	0.09	0.01
Scenario 2												
TB	*0.05*	*0.09*	*0.19*	*0.28*	*0.47*	*0.66*						
A	0.00	0.05	0.33	**0.53**	0.10	0.00	0.12	0.19	0.33	**0.27**	0.07	0.00
B	0.00	0.03	0.35	**0.54**	0.07	0.00	0.13	0.18	0.33	**0.30**	0.06	0.00
C	0.00	0.04	0.23	**0.63**	0.09	0.00	0.11	0.15	0.27	**0.34**	0.12	0.01
D	0.01	0.05	0.25	**0.61**	0.09	0.00	0.11	0.17	0.27	**0.32**	0.11	0.01
Scenario 3												
TB	*0.11*	*0.26*	*0.33*	*0.44*	*0.55*	*0.75*						
A	0.14	**0.55**	**0.24**	0.05	0.00	0.00	0.32	**0.45**	**0.18**	0.04	0.01	0.00
B	0.14	**0.53**	**0.26**	0.06	0.01	0.00	0.28	**0.47**	**0.20**	0.05	0.00	0.00
C	0.12	**0.41**	**0.35**	0.10	0.01	0.00	0.23	**0.38**	**0.26**	0.11	0.02	0.00
D	0.14	**0.44**	**0.32**	0.08	0.01	0.00	0.27	**0.38**	**0.24**	0.09	0.02	0.00
Scenario 4												
TB	*0.07*	*0.22*	*0.30*	*0.40*	*0.52*	*0.71*						
A	0.07	**0.46**	**0.38**	0.09	0.00	0.00	0.23	**0.44**	**0.26**	0.07	0.01	0.00
B	0.07	**0.39**	**0.43**	0.12	0.01	0.00	0.21	**0.41**	**0.29**	0.09	0.01	0.00
C	0.06	0.31	**0.46**	0.16	0.01	0.00	0.18	0.34	**0.32**	0.14	0.03	0.00
D	0.07	0.36	**0.42**	0.13	0.01	0.00	0.21	0.36	**0.29**	0.12	0.02	0.00
Scenario 5												
TB	*0.00*	*0.04*	*0.06*	*0.07*	*0.11*	*0.22*						
A	0.00	0.00	0.00	0.01	0.12	**0.86**	0.10	0.11	0.12	0.13	0.18	**0.35**
B	0.00	0.00	0.00	0.02	0.11	**0.87**	0.10	0.11	0.12	0.14	0.18	**0.35**
C	0.00	0.00	0.00	0.01	0.10	**0.89**	0.10	0.10	0.11	0.11	0.16	**0.42**
D	0.00	0.00	0.00	0.01	0.10	**0.89**	0.10	0.10	0.11	0.11	0.17	**0.41**
Scenario 6												
TB	*0.30*	*0.42*	*0.53*	*0.67*	*0.77*	*0.86*						
A	**0.67**	0.12	0.01	0.00	0.00	0.00	**0.80**	0.18	0.02	0.00	0.00	0.00
B	**0.66**	0.12	0.01	0.00	0.00	0.00	**0.79**	0.18	0.02	0.00	0.00	0.00
C	**0.64**	0.17	0.01	0.00	0.00	0.00	**0.67**	0.27	0.05	0.00	0.00	0.00
D	**0.63**	0.17	0.01	0.00	0.00	0.00	**0.69**	0.25	0.05	0.01	0.00	0.00
Scenario 7												
TB	*0.13*	*0.30*	*0.38*	*0.49*	*0.60*	*0.78*						
A	0.20	**0.57**	0.20	0.02	0.00	0.00	0.36	**0.43**	0.17	0.04	0.00	0.00
B	0.18	**0.57**	**0.22**	0.03	0.00	0.00	0.33	**0.45**	**0.17**	0.04	0.00	0.00
C	0.17	**0.52**	**0.24**	0.06	0.00	0.00	0.28	**0.41**	**0.22**	0.07	0.01	0.00
D	0.18	**0.53**	0.23	0.04	0.00	0.00	0.33	**0.41**	0.20	0.05	0.01	0.00
Scenario 8												
TB	*0.05*	*0.16*	*0.21*	*0.29*	*0.37*	*0.50*						
A	0.02	0.18	**0.43**	**0.29**	0.08	0.01	0.17	0.34	**0.31**	**0.15**	0.04	0.00
B	0.03	0.18	**0.42**	**0.29**	**0.09**	0.00	0.17	0.33	**0.32**	**0.15**	**0.04**	0.00
C	0.01	0.05	0.28	**0.42**	**0.23**	0.01	0.13	0.18	0.26	**0.25**	**0.15**	0.04
D	0.01	0.07	0.26	**0.43**	0.22	0.02	0.13	0.19	0.24	**0.25**	0.14	0.04

desirable when comparing to cohort size 3. Specifically, the percentage of patients treated above the true MTD tends to increase with smaller cohort size, and PCS tends to decrease with smaller cohort size. As more information is needed for the estimation of the toxicity profile at each dose TB_d as opposed to the probability of binary DLT, we recommend implementing this proposed method with cohort size 3 or above.

Table 6. Sensitivity analysis results of the Ti3+3 with different cohort sizes. The TTB is 0.30, and the EI is (0.25,0.35). The MTD in each scenario is in ***bold****.*

	Selection %						Allocation %					
	1	2	3	4	5	6	1	2	3	4	5	6
Scenario 1												
TB	*0.08*	*0.13*	*0.22*	***0.32***	*0.50*	*0.70*						
Cohort size 1	0.04	0.14	0.35	**0.43**	0.03	0.00	0.09	0.17	0.30	**0.31**	0.10	0.02
Cohort size 2	0.04	0.15	0.36	**0.40**	0.06	0.00	0.12	0.22	0.30	**0.27**	0.08	0.01
Cohort size 3	0.02	0.08	0.37	**0.45**	0.08	0.00	0.13	0.20	0.31	**0.26**	0.09	0.01
Scenario 2												
TB	*0.05*	*0.09*	*0.19*	***0.28***	*0.47*	*0.66*						
Cohort size 1	0.03	0.10	0.26	**0.56**	0.05	0.00	0.07	0.14	0.25	0.40	**0.12**	0.02
Cohort size 2	0.02	0.09	0.29	**0.53**	0.08	0.00	0.11	0.18	0.28	0.33	**0.10**	0.01
Cohort size 3	0.01	0.05	0.25	**0.61**	0.09	0.00	0.11	0.17	0.27	0.32	**0.11**	0.01
Scenario 3												
TB	*0.11*	***0.26***	***0.33***	*0.44*	*0.55*	*0.75*						
Cohort size 1	0.22	**0.38**	**0.29**	0.10	0.00	0.00	0.26	**0.32**	**0.26**	0.12	0.03	0.01
Cohort size 2	0.20	**0.43**	**0.30**	0.06	0.01	0.00	0.30	**0.37**	**0.23**	0.08	0.02	0.00
Cohort size 3	0.14	**0.44**	**0.32**	0.08	0.01	0.00	0.27	**0.38**	**0.24**	0.09	0.02	0.00
Scenario 4												
TB	*0.07*	*0.22*	***0.30***	*0.40*	*0.52*	*0.71*						
Cohort size 1	0.12	0.33	**0.42**	0.12	0.00	0.00	0.16	0.31	**0.32**	0.16	0.04	0.01
Cohort size 2	0.15	0.34	**0.39**	0.12	0.01	0.00	0.24	0.34	**0.29**	0.11	0.02	0.00
Cohort size 3	0.07	0.36	**0.42**	0.13	0.01	0.00	0.21	0.36	**0.29**	0.12	0.02	0.00
Scenario 5												
TB	*0.00*	*0.04*	*0.06*	*0.07*	*0.11*	***0.22***						
Cohort size 1	0.00	0.01	0.01	0.03	0.12	**0.82**	0.04	0.05	0.05	0.07	0.16	**0.64**
Cohort size 2	0.01	0.01	0.02	0.03	0.14	**0.80**	0.07	0.08	0.09	0.10	0.17	**0.48**
Cohort size 3	0.00	0.00	0.00	0.01	0.10	**0.89**	0.10	0.10	0.11	0.11	0.17	**0.41**
Scenario 6												
TB	***0.30***	*0.42*	*0.53*	*0.67*	*0.77*	*0.86*						
Cohort size 1	**0.61**	0.15	0.01	0.00	0.00	0.00	**0.65**	0.26	0.07	0.02	0.00	0.00
Cohort size 2	**0.66**	0.15	0.01	0.00	0.00	0.00	**0.71**	0.23	0.05	0.01	0.00	0.00
Cohort size 3	**0.63**	0.17	0.01	0.00	0.00	0.00	**0.69**	0.25	0.05	0.01	0.00	0.00
Scenario 7												
TB	*0.13*	***0.30***	*0.38*	*0.49*	*0.60*	*0.78*						
Cohort size 1	0.26	**0.44**	0.24	0.05	0.00	0.00	0.30	**0.34**	0.22	0.10	0.03	0.01
Cohort size 2	0.27	**0.46**	0.22	0.05	0.00	0.00	0.36	**0.36**	0.20	0.07	0.01	0.00
Cohort size 3	0.18	**0.53**	0.23	0.04	0.00	0.00	0.33	**0.41**	0.20	0.05	0.01	0.00
Scenario 8												
TB	*0.05*	*0.16*	*0.21*	***0.29***	*0.37*	*0.50*						
Cohort size 1	0.07	0.12	0.27	**0.33**	0.21	0.01	0.11	0.16	0.25	**0.26**	0.18	0.05
Cohort size 2	0.05	0.13	0.25	**0.37**	0.19	0.01	0.13	0.22	0.26	**0.25**	0.13	0.02
Cohort size 3	0.01	0.07	0.26	**0.43**	0.22	0.02	0.13	0.19	0.24	**0.25**	0.14	0.04

5. DISCUSSION

We propose a practical rule-based Ti3+3 design that extends the i3+3 design by incorporating toxicity outcomes with multiple toxicity types and grades to improve the efficacy and safety of phase I trials. The Ti3+3 adopts a similar dose-finding algorithm as i3+3, which is simple and straightforward. In addition, we show that it exhibits desirable operating characteristics by extensive simulations. Compared with the existing methods such as CRM-MC and gBOIN, the Ti3+3 demonstrates similar PCS with better safety performance. We provide an RShiny tool freely available at https://i3design.shinyapps.io/ti3plus3/ that generates dose-escalation decisions based on the Ti3+3 design and conducts simulation given toxicity scenarios provided by the users.

A major advantage of using Ti3+3 instead of the model-based and model-assisted designs for practical trials might be its simplicity, especially the simplicity of the dose-finding rules. In particular, the up-and-down rules can be directly assessed and easily understood and executed by clinicians. It is important for clinicians to understand the dose-finding rules since they are the final decision makers for dose selec-

tion for each patient, to whom they might need to explain these decisions. Even though the decision rules of the model-assisted design gBOIN are also straightforward, these rules are based on complex statistical inference that is usually not easy to explain to clinicians. The Ti3+3 design, on the other hand, is based on a set of simple rules, which could be easily conveyed to physicians.

The proposed Ti3+3 design incorporates type-specific toxicity burdens to quantify the impact of a particular type of toxicity events on patients' health and thereby reduce the risk of selecting a toxic dose which may be considered "safe" when evaluated solely based on the overall toxicity burden. Furthermore, the type-specific target toxicity burdens as well as their equivalence intervals used in the proposed method are relatively easy to specify and interpret, i.e., they can be considered as a re-scaled toxicity probability in the conventional DLT-based dose-finding studies. Specifying the target and EI for the overall toxicity burden is also straightforward given the pre-determined weight matrix.

The lower bound of the EI can be considered the smallest toxicity probability that clinicians would not want to escalate the dose, and the upper bound the highest toxicity probability that clinicians would not want to de-escalate. Here, even though the EI is for toxicity burden instead of toxicity probability, because the way we constructed the burden (Equation (2.1)), the toxicity burden is essentially re-scaled as toxicity probability for the highest grade K. Therefore, the rescaling facilitates the elicitation and interpretation of the EIs as we have now explained.

The use of toxicity burden or multiple toxicity grades in dose-finding trials has been limited in practice, mostly due to the need for extensive collaboration between statisticians and clinicians required for the elicitation of weights and target in the design stage. The numerical value of each severity weight reflects the relative effect on patients' survival/quality of life that is associated with experiencing the toxicity at the given grade. The severity weights elicited via interactions between statisticians and clinicians are intrinsically subjective. Alternatively, the elicitation of weights can be facilitated by utilizing existing trial data (for the same drug or drug class) and/or medical databases, such as the FDA Adverse Event Reporting System (FAERS) [25]. Moreover, as noted in Mu *et al.* (2019), to ensure good operating characteristics of the design, the elicitation process should be an iterative process. Simulation studies conducted by statisticians and inputs from multiple physicians are required to ensure that the dose assignment decisions reflect appropriate clinical significance and the design is calibrated appropriately.

APPENDIX A. AN ALGORITHM FOR WEIGHTS ELICITATION

We propose the following algorithm for elicitation of the numerical weights.

1. First, ask the investigators to assign scores on the scale of 1 to 10 to each type of toxicity based on the following three categories 1) the potential impact on patients' survival, 2) the potential impact on patients' quality of life, 3) the potential impact on patients' opportunity to receive new treatments following progression. Then the average score of each toxicity type is calculated. For example, two types of toxicities, neuropathy and low platelet count, are identified as related to the treatment. The average score for neuropathy is calculated to be 8, and the average score for low platelet count is 4.
2. Second, ask the investigators to select a reference toxicity type j. For example, the investigators may choose neuropathy as the reference toxicity. Next, assign severity weights to each grade of the reference toxicity
   (a) Select a reference grade k. Suppose in our example, grade $k = 3$ of neuropathy is considered as the reference grade and is assigned with a weight 1.0.
   (b) Ask the investigators to specify numerical weights for other grades by comparing the clinical importance with the reference grade. For instance, ask a question like "How many grade 1 neuropathy adverse events do you expect to have a similar toxicity effect as a grade 3 neuropathy?" If the answer is 5, then the weight for grade 1 is $\frac{1}{5} = 0.2$. Repeat this step for each grade of toxicity type j. For example, $(0, 0.2, 0.6, 1.0, 2.5)$ may be the resulting numerical weights for grade 0 to 4 neuropathy.
3. Last, determine if the relative toxicity effects across different grades of a specific toxicity type are the same as that of the reference type.
   (a) If the answer is yes, then the numerical weights for the rest of the toxicities can be calculated proportionally to the weights of the reference type based on the ratio between the average scores calculated in step 1. In our example, the weights for low platelet would be $(0, 0.2, 0.6, 1.0, 2.5)/(8/4) = (0, 0.1, 0.3, 0.5, 1.25)$.
   (b) If the answer is no, then go through step 2(a) and 2(b) again and assign weight to each grade of the particular toxicity.
4. List the weights obtained earlier as a matrix and standardized by dividing by the sum of the matrix. For example, the standardized weight matrix in our example is given by,

$$\boldsymbol{W} = \begin{pmatrix} 0 & 0.03 & 0.09 & 0.16 & 0.39 \\ 0 & 0.02 & 0.05 & 0.07 & 0.19 \end{pmatrix}.$$

Repeating the elicitation process with another group of physicians would be useful. If so, the final weight matrix may be the average of the weight matrices from the physicians.

APPENDIX B. SIMULATION SCENARIOS

Table A1: True probabilities of the 2 types of toxicities for 8 scenarios used in Section 3.1.

Toxicity type	Toxicity grade	Dose level					Dose level					Dose level					Dose level				
		1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
		Scenario 1					Scenario 2					Scenario 3					Scenario 4				
Neuropathy	Grade 0	0.47	0.37	0.36	0.36	0.34	0.47	0.47	0.37	0.35	0.30	0.47	0.47	0.39	0.37	0.35	0.47	0.47	0.39	0.38	0.37
	Grade 1	0.23	0.20	0.20	0.20	0.20	0.23	0.23	0.20	0.20	0.20	0.23	0.23	0.30	0.20	0.20	0.23	0.23	0.30	0.25	0.20
	Grade 2	0.27	0.25	0.12	0.11	0.02	0.27	0.27	0.25	0.12	0.11	0.27	0.27	0.27	0.25	0.12	0.27	0.27	0.27	0.32	0.25
	Grade 3	0.02	0.09	0.14	0.07	0.07	0.03	0.02	0.09	0.14	0.07	0.03	0.02	0.03	0.09	0.14	0.03	0.03	0.02	0.02	0.09
	Grade 4	0.01	0.10	0.18	0.27	0.37	0.01	0.01	0.10	0.19	0.32	0.01	0.01	0.02	0.10	0.19	0.00	0.01	0.02	0.03	0.10
Low platelets	Grade 0	0.56	0.31	0.30	0.25	0.20	0.66	0.56	0.31	0.25	0.15	0.66	0.56	0.41	0.31	0.25	0.81	0.66	0.41	0.32	0.31
	Grade 1 or 2	0.40	0.50	0.35	0.30	0.20	0.30	0.40	0.50	0.30	0.25	0.30	0.40	0.50	0.50	0.30	0.16	0.30	0.50	0.50	0.50
	Grade 3	0.02	0.12	0.15	0.10	0.10	0.03	0.02	0.12	0.10	0.15	0.03	0.02	0.05	0.12	0.10	0.02	0.03	0.05	0.13	0.12
	Grade 4	0.02	0.07	0.20	0.35	0.50	0.01	0.02	0.07	0.35	0.45	0.01	0.02	0.04	0.07	0.35	0.01	0.01	0.04	0.05	0.07
		Scenario 5					Scenario 6					Scenario 7					Scenario 8				
Neuropathy	Grade 0	0.06	0.04	0.03	0.02	0.02	0.34	0.20	0.13	0.12	0.11	0.46	0.35	0.26	0.12	0.10	0.88	0.82	0.32	0.15	0.13
	Grade 1	0.44	0.36	0.31	0.17	0.13	0.35	0.31	0.31	0.37	0.17	0.34	0.40	0.31	0.25	0.07	0.04	0.07	0.38	0.32	0.44
	Grade 2	0.38	0.36	0.35	0.35	0.22	0.23	0.32	0.31	0.18	0.19	0.15	0.19	0.32	0.24	0.12	0.03	0.04	0.20	0.37	0.21
	Grade 3	0.05	0.07	0.10	0.13	0.29	0.05	0.08	0.09	0.12	0.26	0.02	0.03	0.05	0.27	0.44	0.03	0.04	0.06	0.07	0.02
	Grade 4	0.07	0.17	0.21	0.33	0.34	0.03	0.09	0.16	0.22	0.28	0.03	0.03	0.06	0.12	0.27	0.02	0.03	0.04	0.10	0.20
Low platelets	Grade 0	0.49	0.37	0.28	0.23	0.20	0.76	0.38	0.18	0.18	0.13	0.39	0.29	0.27	0.21	0.12	0.57	0.53	0.46	0.40	0.13
	Grade 1 or 2	0.39	0.41	0.28	0.12	0.10	0.10	0.47	0.37	0.39	0.05	0.43	0.35	0.41	0.15	0.09	0.40	0.41	0.44	0.46	0.37
	Grade 3	0.06	0.09	0.11	0.30	0.22	0.09	0.07	0.17	0.13	0.07	0.16	0.28	0.21	0.17	0.14	0.02	0.04	0.08	0.11	0.26
	Grade 4	0.06	0.13	0.34	0.35	0.48	0.05	0.08	0.28	0.30	0.75	0.02	0.08	0.11	0.47	0.65	0.01	0.02	0.03	0.03	0.24

Table A2: True probability of each toxicity grade for 10 scenarios used in Section 3.2.

	Dose level						Dose level					
	1	2	3	4	5	6	1	2	3	4	5	6
	Scenario 1						Scenario 2					
Grade 0, 1	0.83	0.75	0.62	0.51	0.34	0.19	0.92	0.85	0.70	0.55	0.24	0.00
Grade 2	0.12	0.15	0.18	0.19	0.16	0.11	0.03	0.05	0.10	0.15	0.26	0.36
Grade 3	0.04	0.07	0.11	0.14	0.15	0.11	0.03	0.07	0.14	0.21	0.35	0.49
Grade 4	0.01	0.03	0.09	0.16	0.35	0.59	0.02	0.03	0.06	0.09	0.15	0.21
	Scenario 3						Scenario 4					
Grade 0, 1	0.78	0.56	0.50	0.40	0.30	0.16	0.88	0.64	0.52	0.35	0.17	0.00
Grade 2	0.14	0.19	0.18	0.17	0.15	0.09	0.04	0.12	0.16	0.22	0.28	0.39
Grade 3	0.06	0.12	0.14	0.15	0.14	0.10	0.06	0.17	0.22	0.30	0.38	0.52
Grade 4	0.02	0.12	0.18	0.28	0.41	0.65	0.02	0.07	0.10	0.13	0.17	0.23
	Scenario 5						Scenario 6					
Grade 0, 1	1.00	0.91	0.88	0.86	0.80	0.65	0.50	0.38	0.29	0.19	0.13	0.08
Grade 2	0.00	0.06	0.07	0.08	0.10	0.13	0.25	0.24	0.21	0.16	0.11	0.07
Grade 3	0.00	0.03	0.04	0.05	0.08	0.14	0.11	0.12	0.12	0.10	0.08	0.05
Grade 4	0.00	0.00	0.01	0.01	0.02	0.08	0.14	0.26	0.38	0.55	0.68	0.80
	Scenario 7						Scenario 8					
Grade 0, 1	0.78	0.58	0.50	0.40	0.30	0.16	0.92	0.76	0.68	0.57	0.45	0.25
Grade 2	0.14	0.18	0.18	0.17	0.15	0.09	0.00	0.00	0.00	0.00	0.00	0.00
Grade 3	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.24	0.32	0.43	0.55	0.75
Grade 4	0.08	0.24	0.32	0.43	0.55	0.75	0.00	0.00	0.00	0.00	0.00	0.00
	Scenario 9						Scenario 10					
Grade 0, 1	0.66	0.10	0.00	0.00	0.00	0.00	0.84	0.52	0.36	0.14	0.45	0.25
Grade 2	0.34	0.90	0.86	0.54	0.20	0.33	0.16	0.48	0.64	0.86	0.00	0.00
Grade 3	0.00	0.00	0.14	0.46	0.80	0.00	0.00	0.00	0.00	0.00	0.55	0.75
Grade 4	0.00	0.00	0.00	0.00	0.00	0.67	0.00	0.00	0.00	0.00	0.00	0.00

REFERENCES

[1] National Cancer Institute. (2018). Common terminology criteria for adverse events (ctcae). Version 3.0.

[2] Storer, B. E. (1989). Design and analysis of phase i clinical trials. *Biometrics* 925–937. https://doi.org/10.2307/2531693. MR1029610

[3] O'Quigley, J., Pepe, M. and Fisher, L. (1990). Continual reassessment method: a practical design for phase 1 clinical trials in cancer. *Biometrics* 33–48. https://doi.org/10.2307/2531628. MR1059105

[4] Ji, Y., Liu, P., Li, Y. and Bekele, B. N. (2010). A modified toxicity probability interval method for dose-finding trials. *Clinical Trials* **7**(6) 653–663.

[5] Ji, Y. and Wang, S. -J. (2013). Modified toxicity probability interval design: a safer and more reliable method than the 3+ 3 design for practical phase i trials. *journal of Clinical Oncology* **31**(14) 1785.

[6] Liu, S. and Yuan, Y. (2015). Bayesian optimal interval designs for phase i clinical trials. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **64**(3) 507–523. https://doi.org/10.1111/rssc.12089. MR3325461

[7] Yan, F., Mandrekar, S. J. and Yuan, Y. (2017). Keyboard: a novel bayesian toxicity probability interval design for phase i clinical trials. *Clinical Cancer Research* **23**(15) 3994–4003.

[8] Liu, M., Wang, S.-J. and Ji, Y. (2020). The i3+3 design for phase i clinical trials. *Journal of Biopharmaceutical Statistics* **30**(2) 294–304.

[9] U.S. Food and Drug Administration (2022). Project optimus. https://www.fda.gov/about-fda/oncology-center-excellence/project-optimus

[10] Penel, N., Adenis, A., Clisant, S. and Bonneterre, J. (2011). Nature and subjectivity of dose-limiting toxicities in contemporary phase 1 trials: comparison of cytotoxic versus non-cytotoxic drugs. *Investigational new drugs* **29**(6) 1414–1419.

[11] Le Tourneau, C., Diéras, V., Tresca, P., Cacheux, W. and Paoletti, X. (2010). Current challenges for the early clinical development of anticancer drugs in the era of molecularly targeted agents. *Targeted oncology* **5**(1) 65–72.

[12] Shah, M., Rahman, A., Theoret, M. R. and Pazdur, R. (2021). The drug-dosing conundrum in oncology-when less is more. *The New England journal of medicine* **385**(16) 1445–1447.

[13] Bekele, B. N. and Thall, P. F. (2004). Dose-finding based on multiple toxicities in a soft tissue sarcoma trial. *Journal of the American Statistical Association* **99**(465) 26–35. https://doi.org/10.1198/016214504000000043. MR2061885

[14] Yuan, Z., Chappell, R. and Bailey, H. (2007). The continual reassessment method for multiple toxicity grades: a bayesian quasi-likelihood approach. *Biometrics* **63**(1) 173–179. https://doi.org/10.1111/j.1541-0420.2006.00666.x. MR2345586

[15] Lee, S. M., Hershman, D., Martin, P., Leonard, J. and Cheung, K. (2009). Validation of toxicity burden score for use in phase i clinical trials. *Journal of Clinical Oncology* **27**(15_suppl) 2514.

[16] Lee, S. M., Cheng, B. and Cheung, Y. K. (2011). Continual reassessment method with multiple toxicity constraints. *Biostatistics* **12**(2) 386–398.

[17] Van Meter, E. M., Garrett-Mayer, E. and Bandyopadhyay, D. (2012). Dose-finding clinical trial design for ordinal toxicity grades using the continuation ratio model: an extension of the continual reassessment method. *Clinical trials* **9**(3) 303–313.

[18] Ezzalfani, M., Zohar, S., Qin, R., Mandrekar, S. J. and Le Deley, M. -C. (2013). Dose-finding designs using a novel quasi-continuous endpoint for multiple toxicities. *Statistics in medicine* **32**(16) 2728–2746. https://doi.org/10.1002/sim.5737. MR3069903

[19] Mu, R., Yuan, Y., Xu, J., Mandrekar, S. J. and Yin, J. (2019). gboin: a unified model-assisted phase i trial design accounting for toxicity grades, and binary or continuous end points. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **68**(2) 289–308. https://doi.org/10.1111/rssc.12263. MR3902995

[20] Hobbs, B. P., Thall, P. F. and Lin, S. H. (2016). Bayesian group sequential clinical trial design using total toxicity burden and progression-free survival. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **65**(2) 273–297. https://doi.org/10.1111/rssc.12117. MR3456689

[21] Milton, D. R., Lin, S. H. and Hobbs, B. P. (2020). Comparing radiation modalities with trimodality therapy using total toxicity burden. *International Journal of Radiation Oncology, Biology, Physics* **107**(5) 1001–1005.

[22] Lin, S. H., Hobbs, B. P., Verma, V., Tidwell, R. S., Smith, G. L., Lei, X., Corsini, E. M., Mok, I., Wei, X., Yao, L. et al. (2020). Randomized phase iib trial of proton beam therapy versus intensity-modulated radiation therapy for locally advanced esophageal cancer. *Journal of Clinical Oncology* **38**(14) 1569.

[23] Morita, S., Thall, P. F. and Müller, P. (2008). Determining the effective sample size of a parametric prior. *Biometrics* **64**(2) 595–602. https://doi.org/10.1111/j.1541-0420.2007.00888.x. MR2432433

[24] Robertson, T. (1988). Order restricted statistical inference. Technical report.

[25] U.S. Food and Drug Administration (2022). Adverse event reporting system.

[26] Guo, W., Wang, S. -J., Yang, S., Lynn, H. and Ji, Y. (2017). A bayesian interval dose-finding design addressing ockham's razor: mtpi-2. *Contemporary clinical trials* **58**. 23–33.

[27] Ivanova, A., Flournoy, N. and Chung, Y. (2007). Cumulative cohort design for dose-finding. *journal of Statistical Planning and Inference* **137**(7) 2316–2327. https://doi.org/10.1016/j.jspi.2006.07.009. MR2325437

[28] Mair, P., Hornik, K. and de Leeuw, J. (2009). Isotone optimization in r: pool-adjacent-violators algorithm (pava) and active set methods. *Journal of statistical software* **32**(5) 1–24.

[29] Lee, S. M. and Cheung, Y. K. (2009). Model calibration in the continual reassessment method. *Clinical Trials* **3**(6) 227–238.

Meizi Liu. Takeda Pharmaceuticals.
E-mail address: meizi.liu@takeda.com

Yuan Ji. Department of Public Health Sciences, University of Chicago.
E-mail address: koaeraser@gmail.com

Ji Lin. Sanofi U.S.
E-mail address: ji.lin@sanofi.com