

# Shift Models for Dose-Finding in Ordered Groups with Late-Onset Toxicity

RAMI HAWILA\* AND NOLAN A. WAGES

---

## Abstract

Historically, the primary objective of Phase I clinical trials has been to pick an optimal dose in terms of patient safety, referred to as the maximum tolerated dose (MTD). Most of these trials recommend a “one-size-fits-all” dose for the patient population being studied, while also solely focusing on short-term adverse events. Often patient heterogeneity exists so that group-specific dose selection is of interest. To address the issue of patient heterogeneity, several dose-finding methods have been proposed, including the shift model framework based on the Continual Reassessment Method (CRM). Additionally, for many cancer therapies, relevant toxicities may be defined by participants experiencing adverse events at any point over a long evaluation window, resulting in pending outcomes when new participants need to be assigned a dose. By leveraging the CRM, the time-to-event continual reassessment method (TITE-CRM) provides a feasible approach for addressing this issue. Motivated by a Phase I trial involving radiotherapy that included two patient groups conducted at the University of Virginia, we have developed a hybrid design that combines elements from the TITE-CRM and the shift model framework. This approach helps address patient heterogeneity and late-onset toxicity simultaneously. We illustrate how to perform a dose-finding trial using the proposed design, and compare its operating characteristics to other suggested methods in the field by conducting a simulation study. The shift model TITE-CRM is shown to be a practical design with good operating characteristics in regard to selecting the correct MTD in each group. An **R** package is also being developed, allowing investigators to provide group-specific MTD recommendations by applying the proposed design, in addition to providing operating characteristics for custom simulation settings.

KEYWORDS AND PHRASES: Dose finding designs, Late onset toxicities, Patient heterogeneity, Cancer.

---

## 1. INTRODUCTION

This article considers the problem of designing dose-finding trials that simultaneously account for patient heterogeneity and late-onset toxicities. The design described herein is applied to a Phase I clinical trial conducted at the University of Virginia. Participants were classified into two groups (good prognosis and poor prognosis) according to their expected tolerance to a radiation treatment [15]. Instead of conducting two separate, parallel trials for the groups, the design used in Muller et al. [15] borrowed accumulating information across groups to identify group-specific maximum tolerated doses (MTDs) within a single trial. In addition, the dose-limiting toxicity (DLT) observation window is up to 90 days, and on average 1-2 patients are accrued monthly. Therefore, there is also a potential for late-onset toxicities, which the original trial design did not incorporate, motivating the new design proposed in this work. Although such phase I trials are increasing in modern oncology development, there is little practical statistical methodology for designing these trials. More details of the trial are provided below.

It is not uncommon for participants to be separated into prognostic groups where they differ in terms of their expected reaction to the treatment [12]. For instance, prior studies have shown that children diagnosed with Acute Leukemia have a higher tolerance to the treatment relative to adults [16]. In this case, we say that the groups are ordered in that one group has a higher expected probability of DLT than the other group when receiving the same dose. Using a standard Phase I design to conduct two independent trials is to be avoided due to the possibility of observing a reversal [8]. A reversal occurs when the MTD selection violates the known group order. For instance, the MTD for a group known to have a poor prognosis should not be greater than that of the good prognosis group. Additionally, if the investigators decide on conducting a standard design, disregarding heterogeneity, the dose level that is recommended would be too toxic for part of the population and sub-optimal for the other. O’Quigley et al. [19] proposed a straightforward extension of the Continual Reassessment Method (CRM) to two groups in which two parameters are being estimated rather than one. The idea is that the additional parameter models the relationship between the two groups. Furthermore, every additional distinct group requires an extra parameter to be estimated ac-

---

\*Corresponding author.

ording to the multi-sample CRM. Due to relatively small sample sizes and limited resources in Phase I trials, it is more efficient to exploit under-parameterized models rather than estimating additional continuous parameters. An intuitive and simple design extending the CRM to “shift” type models [20] accommodates for patient heterogeneity and is practical in trials [15]. The method argues that since we are dealing with a discrete set of dose levels  $\{d_1, d_2, \dots, d_k\}$ , the clinical difference between the groups can be accounted for by shifting the final recommended dose level of one group by one, two, or more levels away from the other group.

The CRM, and its patient heterogeneity extensions, assume complete information on the DLT status of each participant when assigning doses. Participants can experience a DLT at any point in the evaluation window, but they have to complete the full evaluation window without DLT to be classified as a non-DLT outcome. Due to the short observation window in Phase I clinical trials, many Phase I methods face a challenge in situations where toxicities are defined over a long evaluation window relative to patient accrual rate. For instance, suppose that DLT for a particular treatment can be observed at any point over 15 weeks, and the accrual rate is one patient per week. Since we have to wait for 15 weeks to score a patient’s outcome as a non-DLT, as many as 15 patients may be accrued before the first patient’s DLT outcome is observed. So what doses will these patients be administered? Waiting for 15 weeks until the evaluation of the previous patient is recorded would result in a long trial duration. Alternatively, putting them on the same dose as the first participant is an efficient use of resources in a small Phase I trial. Multiple publications have addressed late-onset toxicities, such as [23, 1, 13, 30, 29, 27]. The two most popular designs in practice are the time-to-event (TITE)-CRM [6] and the TITE-BOIN [31].

There are limited options for designing Phase I trials that simultaneously account for patient heterogeneity and late-onset toxicities. Salter et al. [22] proposed a modification to the TITE-CRM [6] using the two-parameter model of O’Quigley et al. [19] and maximum likelihood estimation. Chapple and Thall [2] presented a Bayesian design called Sub-TITE which makes sequentially adaptive subgroup-specific decisions while possibly combining subgroups that have similar estimated dose-toxicity curves. A logistic regression model for the probability of toxicity is utilized to make the decisions based on computed posterior quantities. This scheme allows for different subgroups to be combined for dose finding if the accumulated data indicates homogeneous groups. The Two-Stage Sub-TITE (2S-Sub-TITE) was subsequently presented by McGovern et al. [14], which delays borrowing strength and dynamic clustering across subgroups when the trial starts to improve trial accuracy. In the 1st stage, separate models are estimated for each subgroup. The 2nd stage is initiated at some pre-specified point of patient accrual, and the Sub-TITE design is subsequently followed.

The objective of this paper is to develop a practical hybrid design that combines elements from the TITE-CRM and the shift model framework (Shift TITE-CRM) to address patient heterogeneity and late-onset toxicities simultaneously. The rationale for extending the shift models framework to the TITE-CRM emanates from the computational simplicity exhibited by each method relative to the alternative methods described above. The rest of this article is laid out as follows. The following sub-section provides a motivating Phase 1/2 trial in which participants were separated into two different groups and late-onset toxicities were possible. In the **Methods** section, we demonstrate how the Shift TITE-CRM is executed by building on methodology from the TITE-CRM and the shift model CRM. We discuss the simulation design and investigate the operating characteristics of the proposed method in the **Simulation Studies** section. Finally, we will conclude with remarks and future research in the **Conclusion and Discussion** section.

## Motivating Trial

Muller et al. [15] conducted a single-arm Phase I/II trial at the University of Virginia where patients received a radiation treatment via a novel real-time radiation oncology workflow for the delivery of high-dose single-fraction stereotactic body radiation therapy (SBRT). The study team believed that the therapy treatment at a dose higher than 8 Gy would lead to rapid, durable, and significant pain relief, etc. Among the 46 eligible patients for treatment, 31 were considered poor prognosis and classified as Group 1. 15 patients were considered good prognosis and classified as Group 2. DLTs were defined at grade  $\geq 4$  toxicity occurring within 90 days of treatment. Patients in both groups were administered one of  $k = 4$  available dose levels, and dose allocation was conducted using a 2-stage shift model CRM for heterogeneous groups [26], but the potential for late-onset toxicity was not accounted for in the design. Acceptable safety was defined by any estimated DLT probability less than or equal to a maximum toxicity tolerance of  $\Theta = 0.20$ .

In Group  $g$ , the trial is investigating four ordered dose levels:  $\{d_{g1}, d_{g2}, d_{g3}, d_{g4}\}$ , and the primary objective is to identify the MTD in each of the  $g$  groups. Since patients in Group 1 have a poorer prognosis, the expected probability of toxicity should be higher than that of Group 2 at each dose level. Thus, the MTD for Group 1 should be lower than the MTD for Group 2. For instance, if the MTD was defined to be dose  $d_{23}$  for Group 2, then the MTD for Group 1 could be one or two levels below  $d_{23}$ , given by the following set  $\{d_{11}, d_{12}\}$ . For a fixed dose  $d_{2j}$  administered in Group 2, the possible doses administered in Group 1 are  $\{d_{11}, d_{12}, \dots, d_{1,j-1}\}$ . So the total number of possible shifts in the MTD between the two groups would be  $j - 1$ , where the “true” shift could be any of these values. The true shift is sequentially estimated from a class of CRM shift models using the accumulated data from each new cohort of participants.

## 2. METHODS

Implementation of the proposed design can be done using two approaches; a likelihood-based approach, or a Bayesian approach. We will be focusing on the latter since a likelihood approach would require an additional stage of rule-based allocation until at least one DLT and one non-DLT are observed (outcome heterogeneity) so that one can start modeling. Now, suppose there are  $K$  available dose levels and  $G$  distinct groups. Let  $d_k$  denote the dose at level  $k$  for some Group  $g$ , where  $k \in \{1, 2, \dots, K\}$  and  $g \in \{1, 2, \dots, G\}$ . We utilize the idea of the shift between group dose levels to incorporate the clinical relationship between the groups [9]. The shifts in the MTD then range from 0 to  $K - 1$  dose levels according to the clinical difference of one group relative to the other.

In general, assume that we have  $M$  possible shifts, indexed by  $m = 1, \dots, M$ , between the two groups. For every possible shift model  $m$ , the probability of DLT  $R_{mg}(d_k)$  at dose  $k$  in Group  $g$  is modeled by

$$R_{mg}(d_k) = \Pr(Y_{gk} = 1 | d_k, g, m) = \Psi_{mg}(d_k, a_m) = P_{mgk}^{\exp(a_m)}$$

The “ $P_{mgk}$ ” values construct the skeleton, which is a set of pre-specified constants representing an initial guess for the probability of toxicity at each dose level. As mentioned before, the group with a poorer prognosis will have relatively larger skeleton values. The relative location of the MTDs is accounted for by selecting various skeletons with group-specific MTD positions reflecting all possible scenarios. Then, the skeletons are constructed to be both shifted and monotonic. The term “shift” refers to the relative location of MTDs between groups, while within each group, the skeleton is monotonic, assuming that DLT probabilities increase with dose. In the motivating trial with  $K = 4$  dose levels in each of  $G = 2$  groups, there were three shift models considered given in Table 1.

Since a Bayesian form of the CRM is used, a prior probability distribution  $g(a_m)$  of the parameter  $a_m$ , and a prior probability  $p(m)$  for each shift model should be pre-specified. For the power model, it is recommended to use a zero mean normal prior distribution for the parameter  $a$  for each shift model  $m$ , such as  $g_m(a_m) \sim N(0, \sigma_{a_m}^2)$  [17] and the concept of the least informative variance has shown good performance across a broad array of dose-toxicity scenarios [11]. Let  $n_g$  be the number of patients in Group  $g$  at the end of the study, so that the total number of patients accrued is  $n = \sum_{g=1}^G n_g$ . The likelihood under shift model  $m$  is then given by,

$$L_m(\Omega_j | a_m) = \prod_{g=1}^G \prod_{j=1}^{n_g} (w_{jg} \Psi_{mg}(x_{jg}, a_m))^{y_{jg}} (1 - w_{jg} \Psi_{mg}(x_{jg}, a_m))^{(1-y_{jg})}$$

where  $\Omega_j$  denotes the data accumulated up to the  $j$ th patient,  $w_{jg}$  denotes the weight for the  $j$ th patient in group  $g$ ,  $\Psi_{mg}(x_{jg}, a_m)$  represents the probability of detecting a DLT in Group  $g$  for shift model  $m$ , corresponding to the dose

$x_{jg}$  administered to patient  $j$ , and  $y_{jg}$  denotes the DLT status of patient  $j$  in Group  $g$ . The weight  $w$  is a function of the observation of each patient and has a linear association with the dose-toxicity model. Utilizing a linear function for the weights has been shown to perform well in most scenarios [6]. Similar to the TITE-CRM [6], the weights are constructed using a linear function where  $w_{jg} = \frac{u_{jg}}{T}$ ,  $u_{jg}$  is the time to observe a DLT for patient  $j$  in Group  $g$ , and  $T$  is the length of time that defines the DLT evaluation window. The posterior density of  $a$  given the data accumulated after  $j$  participants,

$$f_m(a_m | \Omega_j) = \frac{L_m(\Omega_j | a_m) \times g(a_m)}{\int_{-\infty}^{\infty} L_m(\Omega_j | a_m) \times g(a_m) da_m}$$

and is used to generate an estimate  $\hat{a}_m$  for  $a_m$  for each shift model. Further, the posterior probabilities of the models given the data can be established and given by:

$$\pi(m | \Omega_j) = \frac{p(m) \int_{-\infty}^{\infty} L_m(\Omega_j | a_m) \times g(a_m) da_m}{\sum_{m=1}^M p(m) \int_{-\infty}^{\infty} L_m(\Omega_j | a_m) \times g(a_m) da_m}$$

The design will choose the shift model with the largest posterior probability,  $p(m)$ , among the  $M$  models. It is expected that the more the data support a particular shift model  $m$ , the greater its posterior probability will be. So for Group  $g$ , a particular shift model is chosen, say  $h$ , that has the highest posterior probability among the models. Then, we take the working model  $\Psi_{hg}(d_k, a)$  associated with this shift model, and estimate  $a$  by applying the Bayesian form of the CRM

$$\hat{R}_{hg}(d_k) = P_{hgk}^{\exp(\hat{a}_h)} \quad \text{where,}$$

$$\hat{a}_h = \frac{\int_{-\infty}^{\infty} a_h \times L_h(\Omega_j | a_h) \times g_h(a_h) da_h}{\int_{-\infty}^{\infty} L_h(\Omega_j | a_h) \times g_h(a_h) da_h}$$

We choose to estimate  $R_{hg}(d_k)$  based on the “plug-in” estimate [18] for the MTD, which has been studied more thoroughly and systematically than other estimators in the literature [5]. When escalating, the design is restricted to not skipping any dose levels that haven’t been administered. Dose allocation for the next cohort of patients, or MTD selection if at the end of the trial, is found by selecting the dose in each group which minimizes the difference in the probability of toxicity and target DLT rate,  $|\hat{R}_{hg}(d_k) - \Theta|$ .

## 3. SIMULATION STUDIES

### 3.1 Single Simulated Trial

This subsection exemplifies the proposed design in a single simulated trial with a similar setting to the motivat-

ing trial by Muller et al. [15]. Participants are classified into  $G = 2$  distinct groups with  $K = 4$  available dose levels,  $d_{g1}, \dots, d_{g4}$ . Participants with poor prognosis are classified as Group 1, and participants with good prognosis are classified as Group 2. The magnitude of the clinical difference is then translated by the number of shifts in dose levels between the groups. The true toxicity curves for Groups 1 and 2 are given by  $\{0.03, 0.11, \mathbf{0.21}, 0.33\}$  and  $\{0.01, 0.03, 0.11, \mathbf{0.21}\}$  respectively; where we assumed a one level shift in the MTD location between the two groups. The target DLT probability that defines the MTD is specified at  $\Theta = 0.20$ , indicating that  $d_3$  is the correct dose for Group 1, and  $d_4$  is the correct dose for Group 2. The study is set to enroll a total of 46 participants, with an anticipated distribution of 50% in Group 1 and 50% in Group 2. The DLT endpoint was defined with respect to a 3-month follow-up period with one participant being enrolled every 0.5 months. We established the skeleton values using a systematic approach [10], and adjusted the position of these values to correspond to each of the three possible shift models given in Table 1. For each shift model, we assigned a prior probability  $p(m) = \frac{1}{M}$  and implemented a normal prior  $N(0, \sigma_a^2 = 1.34)$  on the parameter  $a$ . Failure times were generated under a conditionally uniform model. After each included participant, the model-based approach described in Section 3 is employed to assign the dose for the next participant. The MTD selection is generated at the end of the trial for each group, and a summary of the simulated trial is provided in Table 2. The initial patient, classified as being at poor risk, received dose level  $d_1$  and exhibited a non-DLT response. Consequently, the subsequent recommended dose for the second patient, also categorized as poor risk, was  $d_2$ , resulting in a non-DLT response as well. When the third participant, categorized as being at good risk, enrolled in the study, the design recommended escalating the dose to  $d_3$ , taking into account the improved risk profile of participant 3 compared to participants 1 and 2. At the end of the trial we observed four consecutive non-DLTs in Group 1, and three consecutive non-DLTs in Group 2. The final estimated DLT probabilities for Group 1 were given by  $\hat{R}(d_1) = 0.067$ ,  $\hat{R}(d_2) = 0.125$ ,  $\hat{R}(d_3) = 0.194$ , and  $\hat{R}(d_4) = 0.284$ . The final estimated DLT probabilities for Group 2 were given by  $\hat{R}(d_1) = 0.028$ ,  $\hat{R}(d_2) = 0.067$ ,  $\hat{R}(d_3) = 0.125$ , and  $\hat{R}(d_4) = 0.194$ . Therefore, the recommended dose levels are  $d_3$  and  $d_4$  for Groups 1 and 2 respectively.

### 3.2 Simulation Setting

In this section, we evaluate the operating characteristics of our method using a simulation study that consists of three scenarios used the 2-sample TITE-CRM paper [22]. We compare the performance of the two methods using these scenarios. Salter et al. [22] specified a target probability of DLT at  $\Theta = 0.20$  for a treatment that consists of  $k = 6$  dose levels,  $d_1, \dots, d_6$ . A total of 32 participants were accrued to the study, with an anticipated distribution of 50% in Group 1

Table 1. Skeleton values by dose level for each possible shift model in the Motivating Trial.

Model	Prognosis group	Doses in Gy			
		8	10	12.5	15
$m = 1$	1 - Poor	0.07	0.13	<b>0.20</b>	0.29
	2 - Good	0.03	0.07	0.13	<b>0.20</b>
$m = 2$	1 - Poor	0.13	<b>0.20</b>	0.29	0.38
	2 - Good	0.03	0.07	0.13	<b>0.20</b>
$m = 3$	1 - Poor	<b>0.20</b>	0.29	0.38	0.47
	2 - Good	0.03	0.07	0.13	<b>0.20</b>

Table 2. A single simulated trial of 46 patients where “ $j$ ” denotes patient ID, “ $g$ ” denotes group designation, “ $u_j$ ” denotes time to toxicity, “ $x_j$ ” denotes dose administered, and “ $y_j$ ” denotes DLT status.

$j$	$g$	$u_j$	$x_j$	$y_j$	$j$	$g$	$u_j$	$x_j$	$y_j$
1	1	-	$d_1$	0	24	1	-	$d_2$	0
2	1	-	$d_2$	0	25	2	-	$d_3$	0
3	2	-	$d_3$	0	26	2	-	$d_3$	0
4	2	-	$d_4$	0	27	1	-	$d_2$	0
5	1	1.33	$d_4$	1	28	2	2.28	$d_3$	1
6	1	1.22	$d_4$	1	29	2	-	$d_3$	0
7	2	1.82	$d_4$	1	30	2	-	$d_3$	0
8	1	-	$d_2$	0	31	2	-	$d_4$	0
9	1	-	$d_1$	0	32	1	-	$d_3$	0
10	2	2.01	$d_4$	1	33	2	-	$d_3$	0
11	1	-	$d_1$	0	34	2	-	$d_3$	0
12	1	-	$d_1$	0	35	2	-	$d_3$	0
13	1	-	$d_1$	0	36	2	-	$d_3$	0
14	2	-	$d_3$	0	37	2	-	$d_4$	0
15	1	-	$d_1$	0	38	1	-	$d_3$	0
16	1	-	$d_1$	0	39	1	2.97	$d_3$	1
17	2	-	$d_3$	0	40	2	-	$d_4$	0
18	1	-	$d_2$	0	41	2	-	$d_4$	0
19	2	-	$d_3$	0	42	1	-	$d_3$	0
20	1	1.55	$d_2$	1	43	2	-	$d_4$	0
21	1	-	$d_2$	0	44	1	-	$d_3$	0
22	2	-	$d_3$	0	45	1	-	$d_2$	0
23	1	-	$d_2$	0	46	1	-	$d_2$	0
$\hat{a} = 0.017$		Recommended dose for group 1: $d_3$							
		Recommended dose for group 2: $d_4$							

and 50% in Group 2. The DLT endpoint was defined over a 6-months follow-up period with a new participant being accrued every 0.5 months. The true DLT probabilities for each of these scenarios are provided in the lines labeled  $R(d_k)$  in Table 4. The proposed skeleton values in Salter et al. [22] were used as prior guess of DLT probabilities.

Next, we examined the performance of the proposed design in trials in which three patient groups were being studied. Across seven different toxicity scenarios, we simulated trials accruing a total of 36 participants, with an expected allocation of one-third in each of the three groups. The

Table 3. Skeleton value by dose level for three groups.

Model	Group 1				Group 2				Group 3			
	$d_1$	$d_2$	$d_3$	$d_4$	$d_1$	$d_2$	$d_3$	$d_4$	$d_1$	$d_2$	$d_3$	$d_4$
$m = 1$	0.05	0.15	<b>0.25</b>	0.35	0.05	0.15	<b>0.25</b>	0.35	0.05	0.15	<b>0.25</b>	0.35
$m = 2$	0.15	<b>0.25</b>	0.35	0.45	0.05	0.15	<b>0.25</b>	0.35	0.05	0.15	<b>0.25</b>	0.35
$m = 3$	<b>0.25</b>	0.35	0.45	0.55	0.05	0.15	<b>0.25</b>	0.35	0.05	0.15	<b>0.25</b>	0.35
$m = 4$	0.15	<b>0.25</b>	0.35	0.45	0.15	<b>0.25</b>	0.35	0.45	0.05	0.15	<b>0.25</b>	0.35
$m = 5$	<b>0.25</b>	0.35	0.45	0.55	0.15	<b>0.25</b>	0.35	0.45	0.05	0.15	<b>0.25</b>	0.35
$m = 6$	<b>0.25</b>	0.35	0.45	0.55	<b>0.25</b>	0.35	0.45	0.55	0.05	0.15	<b>0.25</b>	0.35

true toxicity probabilities for each of these scenarios are provided in the lines labeled  $R(d_k)$  in Table 5. The scenarios were constructed in a way to accommodate all possible clinical differences between the groups knowing that  $MTD_1 \leq MTD_2 \leq MTD_3$ . The proportion of MTD selection for all scenarios is a result of 1000 simulated trials generated using the programming language **R**. In each scenario, the starting dose was  $d_1$  for all groups. Participants were followed for 6 months, and a new patient is accrued on a fixed scheme of every 0.5 months. For each shift model, we assigned a prior probability  $p(m) = \frac{1}{M}$  and implemented a normal prior  $N(0, \sigma_a^2 = 1.34)$  on the parameter  $a$ . We investigated two different failure time models of the Shift TITE-CRM. S-TITE U generated the failure times under a conditionally uniform model, and S-TITE E generated failure times under a truncated exponential model, with an upper limit restricted to less than the follow-up period of 6 months. We established the skeleton using a systematic approach [10], and adjusted the position of these values to correspond to each of the seven possible orders consistent with the shifts given in Table 3. We also provide simulation results evaluating patient allocation, varying sample sizes, target DLT rates, and accrual rates. These additional results are given in the supplementary material.

### 3.3 Simulation Results

In the two group setting, the proportion of MTD selection over the three scenarios are provided in Table 4 where the results of the Shift TITE-CRM and the 2-sample TITE CRM are compared. The two methods had comparable ranges of proportion of correct selection (PCS) of the group-specific MTDs over the 3 scenarios. The 2-sample TITE CRM generated a PCS range of  $\{0.37, 0.58\}$  and the Shift TITE-CRM generated a PCS range of  $\{0.426, 0.579\}$ . In the 1st scenario,  $d_2$  and  $d_3$  corresponded to true probabilities of DLT closest to  $\Theta = 0.20$  for Groups 1 and 2 respectively. The PCS in both groups were slightly higher with the 2-sample TITE CRM approach compared to the Shift-TITE CRM approach. In the 2nd scenario, the simulated trial assumed a greater clinical difference between the two groups, where  $d_2$  and  $d_4$  corresponded to true probabilities of DLT closest to the target DLT rate for Groups 1 and 2 respectively. The Shift TITE-CRM resulted in a higher  $PCS_S = 0.493$  relative to the 2-sample TITE CRM  $PCS_{2Sample} = 0.44$  in

Group 1. In Group 2, the Shift TITE-CRM outperforms the 2-sample TITE CRM with approximately 10% difference in PCS (0.466 vs. 0.37). This specific scenario indicates that the proposed design is outperforming the 2-sample TITE CRM when there is a two level shift in MTD between the two groups. In the 3rd scenario, the two groups were assumed the same in terms of their probabilities of DLT over the entire range of dose levels. Therefore, the dose that corresponded to probability of DLT closest to  $\Theta = 0.20$  is  $d_2$  in both groups. The two methods resulted in similar operating characteristics with  $PCS_{2Sample} = 0.45$  for both groups,  $PCS_S = 0.492$  and  $PCS_S = 0.426$  for Groups 1 and 2 respectively.

In the three group setting, the proportion of MTD selection across the seven scenarios are provided in Table 5. The range of PCS of the group-specific MTDs over the seven scenarios is  $\{0.380, 0.709\}$ . In scenario 1, it is assumed that there is no difference between the groups. In Group 1, the S-TITE U generated a PCS = 0.472 similar to that of Group 2 with PCS = 0.496 and Group 3 with PCS = 0.437. As for the S-TITE E, all of the groups showed a slight reduction in PCS. In scenario 2, it is assumed that there is a one-shift difference in Groups 2 and 3 relative to Group 1. The PCS decreased on average across all three groups compared to scenario 1, while still maintaining the highest percentage of MTD selection at the correct doses across all groups. In scenario 3, we assumed a greater shift for Groups 2 and 3 relative to Group 1. S-TITE U generated a PCS = 0.472 for Group 1, PCS = 0.437 for Group 2, and PCS = 0.474 for Group 3. In scenario 4, we assumed that only Group 3 is clinically different to Group 1. The simulations generated a PCS range of  $\{0.415, 0.522\}$ . The 5th and 7th scenarios are considered the most heterogeneous in regards to their group-specific MTDs; all groups are clinically different from one another. For these scenarios, the proposed method would indicate that the “true” shift model is  $m = 5$  provided in Table 3. Across these two scenarios, the PCS for Group 1 is 0.522 and 0.696 under S-TITE U, and the PCS is 0.522 and 0.607 under S-TITE E. Groups 2 and 3 generated a smaller PCS that ranges from 0.415 to 0.513 across the two methods. In the 6th scenario, Groups 1 and 2 are assumed to have  $d_1$  as the correct dose, shifted two levels away from that of Group 3  $d_3$ . The PCS of the MTD was high for Group 1 under both uniform and exponential assumptions. S-TITE U

Table 4. The proportion of MTD Selection for Shift Models TITE and 2-sample TITE-CRM at each dose where “ $R(d_k)$ ” denotes the true probability of DLT, “S-TITE” denotes the Shift TITE-CRM design, and “2-Sample TITE” denotes the 2-Sample TITE CRM design.

Scenario		Group 1						Group 2					
		$d_1$	$d_2$	$d_3$	$d_4$	$d_5$	$d_6$	$d_1$	$d_2$	$d_3$	$d_4$	$d_5$	$d_6$
1	$R(d_k)$	0.08	<b>0.20</b>	0.35	0.50	0.70	0.80	0.01	0.05	<b>0.18</b>	0.40	0.55	0.70
	S-TITE	0.303	<b>0.458</b>	0.223	0.016	0.000	0.000	0.008	0.193	<b>0.579</b>	0.205	0.015	0.000
	2-Sample TITE	0.22	<b>0.49</b>	0.25	0.05	0.000	0.000	0.001	0.15	<b>0.58</b>	0.20	0.04	0.000
2	$R(d_k)$	0.02	<b>0.19</b>	0.31	0.45	0.51	0.63	0.03	0.05	0.11	<b>0.21</b>	0.39	0.50
	S-TITE	0.105	<b>0.493</b>	0.319	0.078	0.005	0.000	0.000	0.057	0.354	<b>0.466</b>	0.117	0.006
	2-Sample TITE	0.12	<b>0.44</b>	0.35	0.09	0.02	0.000	0.000	0.05	0.35	<b>0.37</b>	0.18	0.05
3	$R(d_k)$	0.07	<b>0.23</b>	0.31	0.35	0.45	0.57	0.07	<b>0.23</b>	0.31	0.35	0.45	0.57
	S-TITE	0.276	<b>0.492</b>	0.191	0.037	0.004	0.000	0.122	<b>0.426</b>	0.345	0.083	0.023	0.001
	2-Sample TITE	0.25	<b>0.45</b>	0.23	0.07	0.03	0.00	0.25	<b>0.45</b>	0.25	0.04	0.01	0.000

Table 5. The proportion of MTD selection for each dose across the 7 scenarios where “ $R(d_k)$ ” denotes the true probability of toxicity, “S-TITE U” denotes Shift TITE-CRM design with failure times generated uniformly, and “S-TITE E” denotes Shift TITE-CRM design with failure times generated exponentially.

Scenario		Group 1				Group 2				Group 3			
		$d_1$	$d_2$	$d_3$	$d_4$	$d_1$	$d_2$	$d_3$	$d_4$	$d_1$	$d_2$	$d_3$	$d_4$
1	$R(d_k)$	0.05	0.15	<b>0.25</b>	0.35	0.05	0.15	<b>0.25</b>	0.35	0.05	0.15	<b>0.25</b>	0.35
	S-TITE U	0.064	0.324	<b>0.472</b>	0.140	0.008	0.243	<b>0.496</b>	0.253	0.002	0.142	<b>0.437</b>	0.419
	S-TITE E	0.031	0.338	<b>0.458</b>	0.173	0.005	0.231	<b>0.477</b>	0.287	0.001	0.156	<b>0.418</b>	0.425
2	$R(d_k)$	0.15	<b>0.25</b>	0.35	0.45	0.05	0.15	<b>0.25</b>	0.35	0.05	0.15	<b>0.25</b>	0.35
	S-TITE U	0.262	<b>0.416</b>	0.269	0.053	0.035	0.321	<b>0.443</b>	0.201	0.009	0.214	<b>0.403</b>	0.374
	S-TITE E	0.245	<b>0.432</b>	0.252	0.071	0.018	0.316	<b>0.444</b>	0.222	0.007	0.206	<b>0.426</b>	0.361
3	$R(d_k)$	<b>0.22</b>	0.33	0.42	0.52	0.05	0.17	<b>0.27</b>	0.37	0.03	0.12	<b>0.24</b>	0.40
	S-TITE U	<b>0.472</b>	0.364	0.142	0.022	0.067	0.361	<b>0.437</b>	0.135	0.009	0.250	<b>0.474</b>	0.267
	S-TITE E	<b>0.442</b>	0.395	0.142	0.021	0.035	0.341	<b>0.435</b>	0.189	0.000	0.199	<b>0.472</b>	0.329
4	$R(d_k)$	0.13	<b>0.27</b>	0.45	0.55	0.16	<b>0.26</b>	0.36	0.46	0.05	0.15	<b>0.25</b>	0.35
	S-TITE U	0.388	<b>0.480</b>	0.121	0.011	0.222	<b>0.460</b>	0.258	0.060	0.034	0.313	<b>0.403</b>	0.250
	S-TITE E	0.338	<b>0.506</b>	0.145	0.011	0.168	<b>0.488</b>	0.272	0.072	0.028	0.303	<b>0.380</b>	0.289
5	$R(d_k)$	<b>0.22</b>	0.32	0.40	0.52	0.15	<b>0.24</b>	0.33	0.43	0.05	0.15	<b>0.25</b>	0.40
	S-TITE U	<b>0.522</b>	0.358	0.110	0.010	0.219	<b>0.415</b>	0.309	0.057	0.029	0.348	<b>0.427</b>	0.196
	S-TITE E	<b>0.522</b>	0.348	0.117	0.013	0.224	<b>0.416</b>	0.290	0.070	0.019	0.315	<b>0.455</b>	0.211
6	$R(d_k)$	<b>0.25</b>	0.35	0.45	0.55	<b>0.25</b>	0.35	0.45	0.55	0.05	0.15	<b>0.25</b>	0.35
	S-TITE U	<b>0.709</b>	0.244	0.045	0.002	<b>0.494</b>	0.375	0.118	0.013	0.060	0.415	<b>0.386</b>	0.139
	S-TITE E	<b>0.698</b>	0.251	0.049	0.002	<b>0.509</b>	0.357	0.120	0.014	0.053	0.409	<b>0.391</b>	0.147
7	$R(d_k)$	<b>0.24</b>	0.40	0.54	0.66	0.11	<b>0.26</b>	0.41	0.55	0.03	0.10	<b>0.25</b>	0.40
	S-TITE U	<b>0.696</b>	0.266	0.037	0.001	0.293	<b>0.461</b>	0.227	0.019	0.023	0.359	<b>0.493</b>	0.125
	S-TITE E	<b>0.607</b>	0.336	0.055	0.002	0.187	<b>0.501</b>	0.280	0.032	0.010	0.280	<b>0.513</b>	0.197

generated a PCS = 0.709, and S-TITE E generated a PCS = 0.698. In Groups 2 and 3, the PCS range was {0.386,0.509} across the two methods. Overall, the results indicate that Shift TITE-CRM is performing well in terms of correctly recommending group-specific MTDs. The Shift TITE-CRM also demonstrates good operating characteristics in terms of patient allocation to doses at and around group-specific MTDs, as detailed in the supplementary material.

#### 4. CONCLUSIONS

The hybrid design (Shift TITE-CRM) presented in this article is a practical tool in Phase I clinical trials that ad-

resses the issue of patient heterogeneity and late-onset toxicities simultaneously in a single trial. The proposed design is presented in a Bayesian framework since using likelihood estimation presents additional challenges in more than one group setting. Prior to estimation, at least one toxic and one non-toxic outcome have to be observed to fit the model. So the design for the first stage has to be specified and there is not a clear optimal design to use, especially in presence of patient heterogeneity. We examined the performance of the Shift TITE-CRM in terms of correctly selecting group-specific MTDs across seven different scenarios and compared its performance to the 2-sample TITE CRM. Each of the

three 2-sample TITE CRM applications were competitive with the Shift TITE-CRM in terms of the PCS except when there was a shift greater than one in MTD level. In this case, our proposed Shift TITE-CRM demonstrated superior performance. We also examined the performance with different sample sizes, accrual rates, and target toxicity rates provided in the supplementary material. We expect to have at most three groups in Phase I trials and the proposed design generates good operating characteristics in a reasonable range of sample sizes for such trials. In our simulation study, we assumed a sample size of 36 patients separated into three groups. In a real-life phase I trial, it is recommended to have a larger sample size when patients are separated into more than two groups which will yield better precision in selecting the correct MTD as shown in the supplemental material.

For future work, we are exploring the use of this method for partially ordered groups, where group order is known between some but not all groups. We will also build an **R** package which will serve as a tool to conduct the Shift TITE-CRM design in Phase I clinical trials. Salter et al. [21] developed a SAS program to accommodate two groups using the TITE-CRM and likelihood estimation. In the **R** package, statisticians and clinicians will be able to simulate operating characteristics at the design stage and input their new data after accruing patients to recommend a dose to the next cohort of patients at the trial conduct stage. The skeleton values should be constructed according to the specific trial which should be straightforward when knowing the complete ordering of the groups. Note that an increase in the number of groups and/or dose levels leads to an increase in the required sample size to ensure good operating characteristics. An increase in the dimension also leads to the need to specify more skeleton values for the shift models, which can also add complexity. In partially ordered groups, the construction of the skeleton values could become more complex in some scenarios. In trials accounting for patient heterogeneity, borrowing information across groups early in the trial can lead to rapid dose escalation in less sensitive groups. Building on the approach of [28], future implementations are exploring the integration of asymmetric loss functions within the CRM shift model framework to impose penalties on overdosing. This approach aims to mitigate the risk of overdosing in each group, particularly during the early stages of the study when sample sizes are small.

## SUPPLEMENTARY MATERIAL

The supplementary material includes additional simulation results, detailing patient allocation, varying sample sizes, target DLT rates, and accrual rates.

## ACKNOWLEDGEMENTS

The authors would like to acknowledge the helpful comments of the reviewer that improved the original submission.

## FUNDING

The research of Wages is supported by the National Institute of Health grant R01CA247932. The research of Hawila is supported by funding from Massey Comprehensive Cancer Center.

Accepted 21 May 2025

## REFERENCES

- [1] BEKELE, B.N., JI, Y., SHEN, Y. and THALL, P. F. Monitoring late-onset toxicities in phase I trials using predicted risks. *Biostatistics* **9**(3) 442–457 (2008).
- [2] CHAPPLE, A. G. and THALL, P. F. Subgroup-specific dose finding in phase I clinical trials based on time to toxicity allowing adaptive subgroup combination. *Pharmaceutical statistics* **17**(6) 734–749 (2018).
- [3] CHEUNG, K. and DUONG, M. J. Package ‘dfcrm’ (2013).
- [4] CHEUNG, Y. K. Coherence principles in dose-finding studies. *Biometrika* **92**(4) 863–873 (2005). <https://doi.org/10.1093/biomet/92.4.863>. MR2234191
- [5] CHEUNG, Y. K. *Dose finding by the continual reassessment method*. CRC Press (2011).
- [6] CHEUNG, Y. K. and CHAPPELL, R. Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics* **56**(4) 1177–1182 (2000). <https://doi.org/10.1111/j.0006-341X.2000.01177.x>. MR1815616
- [7] HORTON, B. J. and WAGES, N. A. In *Continual Reassessment Method*. *Wiley StatsRef: Statistics Reference Online* 1–7 (2014).
- [8] HORTON, B. J., O’QUIGLEY, J. and CONAWAY, M. R. Consequences of performing parallel dose finding trials in heterogeneous groups of patients. *JNCI Cancer Spectrum* **3**(2):pkz013 (2019).
- [9] HORTON, B. J., WAGES, N. A. and CONAWAY, M. R. Shift models for dose-finding in partially ordered groups. *Clinical Trials* **16**(1) 32–40 (2019).
- [10] LEE, S. M. and CHEUNG, Y. K. Model calibration in the continual reassessment method. *Clinical Trials* **6**(3) 227–238 (2009).
- [11] LEE, S. M. and CHEUNG, Y. K. Calibration of prior variance in the Bayesian continual reassessment method. *Statistics in medicine* **30**(17) 2081–2089 (2011). <https://doi.org/10.1002/sim.4139>. MR2829158
- [12] LEGEDZA, A. T. R. and IBRAHIM, J. G. Heterogeneity in phase I clinical trials: prior elicitation and computation using the continual reassessment method. *Statistics in Medicine* **20**(6) 867–882 (2001). <https://doi.org/10.5351/KJAS.2002.15.2.323>. MR1998919
- [13] LIU, S., YIN, G. and YUAN, Y. Bayesian data augmentation dose finding with continual reassessment method and delayed toxicity. *The annals of applied statistics* **7**(4) 1837 (2013). <https://doi.org/10.1214/13-AOAS661>. MR3161716
- [14] MCGOVERN, A., CHAPPLE, A. G. and MA, C. Two-stage subgroup-specific time-to-event (2s-sub-tite): An adaptive two-stage time-to-toxicity design for subgroup-specific dose finding in phase I oncology trials. *Pharmaceutical Statistics* **21**(6) 1138–1148 (2022).
- [15] MULLER, D. A., WAGES, N. A., WILSON, D. D., DUTTA, S. W., ALONSO, C. E., HANDSFIELD, L. L., CHEN, Q., SMITH, A. B., ROMANO, K. D., JANOWSKI, E. M. et al. STAT RAD: prospective dose escalation clinical trial of single fraction scan-plan-QA-treat stereotactic body radiation therapy for painful osseous metastases. *Practical Radiation Oncology* **10**(6) e444–e451 (2020).
- [16] NEAGA, A., JIMBU, L., MESAROS, O., BOTA, M., LAZAR, D., CAINAP, S., BLAG, C. and ZDRENGHEA, M. Why do children with acute lymphoblastic leukemia fare better than adults? *Cancers* **13**(15) 3886 (2021).
- [17] O’QUIGLEY, J. and SHEN, L. Z. Continual reassessment method: a likelihood approach. *Biometrics*. 673–684 (1996).

- [18] O'QUIGLEY, J., PEPE, M. and FISHER, L. Continual reassessment method: a practical design for phase I clinical trials in cancer. *Biometrics* **33**–48 (1990). <https://doi.org/10.2307/2531628>. MR1059105
- [19] O'QUIGLEY, J., SHEN, L. Z. and GAMST, A. Two-sample continual reassessment method. *Journal of biopharmaceutical statistics* **9**(1) 17–44 (1999).
- [20] O'QUIGLEY, J. Theoretical study of the continual reassessment method. *Journal of Statistical Planning and Inference* **136**(6) 1765–1780 (2006). <https://doi.org/10.1016/j.jspi.2005.08.003>. MR2255595
- [21] SALTER, A., MORGAN, C. and ABAN, I. B. Implementation of a two-group likelihood time-to-event continual reassessment method using sas. *Computer methods and programs in biomedicine* **121**(3) 189–196 (2015).
- [22] SALTER, A., O'QUIGLEY, J., CUTTER, G. R. and ABAN, I. B. Two-group time-to-event continual reassessment method using likelihood estimation. *Contemporary clinical trials* **45**. 340–345 (2015).
- [23] THALL, P. F., LEE, J. J., TSENG, C.-H. and ESTEY, E. H. Accrual strategies for phase I trials with delayed patient outcome. *Statistics in Medicine* **18**(10) 1155–1169 (1999).
- [24] WAGES, N. A. and PETRONI, G. R. A web tool for designing and conducting phase I trials using the continual reassessment method. *BMC cancer* **18**(1) 1–8 (2018). MR3593209
- [25] WAGES, N. A., CONAWAY, M. R. and O'QUIGLEY, J. Performance of two-stage continual reassessment method relative to an optimal benchmark. *Clinical Trials* **10**(6) 862–875 (2013).
- [26] WAGES, N. A., READ, P. W. and PETRONI, G. R. A phase I/II adaptive design for heterogeneous groups with application to a stereotactic body radiation therapy trial. *Pharmaceutical statistics* **14**(4) 302–310 (2015). <https://doi.org/10.1080/10543406.2010.514461>. MR2758355
- [27] WAGES, N. A., BRAUN, T. M. and CONAWAY, M. R. Isotonic design for phase I cancer clinical trials with late-onset toxicities. *Journal of Biopharmaceutical Statistics* **33**(3) 357–370 (2023).
- [28] WHEELER, G. M., SWEETING, M. J. and MANDER, A. P. Toxicity-dependent feasibility bounds for the escalation with overdose control approach in phase I cancer trials. *Statistics in medicine* **36**(16) 2499–2513 (2017). <https://doi.org/10.1002/sim.7280>. MR3660146
- [29] XUE, X., FOSTER, M. C. and IVANOVA, A. Rapid enrollment design for finding the optimal dose in immunotherapy trials with ordered groups. *Journal of biopharmaceutical statistics* **29**(4) 625–634 (2019).
- [30] YIN, G., ZHENG, S. and XU, J. Fractional dose-finding methods with late-onset toxicity in phase I clinical trials. *Journal of Biopharmaceutical Statistics* **23**(4) 856–870 (2013). <https://doi.org/10.1080/10543406.2013.789892>. MR3196080
- [31] YUAN, Y., LIN, R., LI, D., NIE, L. and WARREN, K. E. Time-to-event Bayesian optimal interval design to accelerate phase I trials. *Clinical Cancer Research* **24**(20) 4921–4930 (2018).

Rami Hawila. Virginia Commonwealth University, USA.

E-mail address: [hawilar@vcu.edu](mailto:hawilar@vcu.edu)

Nolan A. Wages. Virginia Commonwealth University, USA.

E-mail address: [wagesn@vcu.edu](mailto:wagesn@vcu.edu)