RE-PARAMETERIZATION OF THE DOSE-TOXICITY MODEL OF THE TIME-TO-EVENT CONTINUAL REASSESSMENT METHOD TO ACCOMMODATE PATIENT HETEROGENEITY

by

Xinxin Sun

MEng, Shanghai Jiao Tong University, China, 2014

Submitted to the Graduate Faculty of

the Department of Biostatistics

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Xinxin Sun

It was defended on

July 13, 2018

and approved by

Maria Mori Brooks, PhD, Professor, Department of Epidemiology
Graduate School of Public Health, University of Pittsburgh

Ada Youk, PhD, Associate Professor, Department of Biostatistics
Graduate School of Public Health, University of Pittsburgh

Thesis Advisor:
Daniel Normolle, PhD, Associate Professor, Department of Biostatistics
Graduate School of Public Health, University of Pittsburgh

Daniel Normolle, PhD

RE-PARAMETERIZATION OF THE DOSE-TOXICITY MODEL OF THE TIME-TO-EVENT CONTINUAL REASSESSMENT METHOD TO ACCOMMODATE PATIENT HETEROGENEITY

Xinxin Sun, MS

University of Pittsburgh, 2018

ABSTRACT

The objective of this thesis is to develop a new method, re-parameterizing the logistic dose-toxicity model of the time-to-event continual reassessment method (TITE-CRM) for phase 1 dose-escalation cancer trials, to account for patient heterogeneity.

In a phase 1 dose-escalation clinical trial, the goal is to identify a dose (the maximum tolerated dose or MTD) that is associated with a pre-specified probability of unacceptable toxicity. TITE-CRM is a Bayesian, model-based dose-escalation trial design specificially intended for situations where participants must be observed for toxicity a long period time compared to the rate at which patients present for treatment.

The risk of toxicity may vary due to particpants' inherent characteristics. For example, patients with lung cancer and concomitant renal disease are not able to tolerate the same dose level of the cancer treatment as the patients without concomitant renal disease. If patient heterogeneity exists, the one-group study design will ignore the patient heterogeneity and give an average MTD, which could result in excess toxicity in the higher risk patients and suboptimal dosing in the lower risk patients.

We have developed a new method to address this issue by re-parameterizing the logisitic model of the TITE-CRM and adding dose-escalation rules to reflect risk group ordering information and control aggressive dose escalation. We assessed the operating characteristics of this design in simulations assuming three risk groups, although the model is trivially extensible to a greater number. We compared this method to parallel trials that use independent one-parameter logistic models. We investigated scenarios with equal numbers of patients in the risk groups and more patients in the higher toxicity risk group, three different true dose-toxicity models and three sample sizes, for a total of 5,400 trials.

According to the results, the TITE-CRM with re-parameterized logistic model used in 3-group trials performs similarly to the TITE-CRM with one-parameter logistic model used in the parallel trials in terms of operating characteristics. However, the TITE-CRM with the re-parameterized logistic model worked better in terms of in-trial dose allocation of patients and recommending the correct final dose in the scenario where there were more patients with higher toxicity risk in a trial. However, estimations of the group risk difference were not precise and the coverages of credible intervals of the group difference parameters were excessively conservative. The method shows promise, and could be implemented now, but further improvements are required. Cancer is one of the major public health concerns in the modern society and the new method can be used in the phase 1 dose-escalation cancer trial and have positive impact on the public health.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

The purpose of a phase 1 dose-escalation clinical trial is to find the maximum tolerated dose (MTD), defined as the highest dose of a drug or treatment associated with unacceptable toxicity in a prespecified proportion of patients [1]. Currently, there are algorithm-based designs, such as 3+3, and model-based designs, such as the continual assessment method (CRM) [2] for these trials.

Algorithmic designs such as 3+3 uses a set of prespecified rules to decide the dose level for the next cohort of participants, depending on the toxicities observed in the cohort just completed. It treats 3 participants at each dose level and raises the dose for the next 3 participants if no toxicity is observed; if there is a toxicity, another cohort of 3 participants is treated at the same dose level. If there is no toxicity among the 3 additional participants, the next cohort of 3 at a dose level up, if there is toxicity, the next cohort of 3 is treated at one dose level down. Simulations demonstrate this design will, on average, stop at the dose level at which the probability of toxicity is 1/6.

CRM uses the dosing and toxicity data of all the participants who have been treated to decide the dose level of the next participant by means of a mathematical function relating dose to toxicity. Compared to the 3+3 method, CRM and other model-based designs are more flexible, in that they can accomodate larger sample sizes (improving the precision of estimation) and can be calibrated to converge on target probabilities smaller or larger than 1/6, as required by the perceived risk-benefit tradeoff (e.g., patients with higher-risk disease may tolerate treatments with

higher risk of toxicity), and they tend to treat more patients at doses closer to the MTD [3], which improves the efficiency of estimation and potentially provides more effective treatments to the participants of the trial. According to a literature review for model-based designs conducted by Love [4], model-based designs were used in 1.6% of phase 1 trials published between 1991 and 2006, increasing to 6.4% in trials published between 2012 and 2014. Although model-based designs are not yet as prevalent as the algorithm-based methods, they are becoming more widely used.

CRM is a Bayesian design that has five components: prior; dose-toxicity model; target probability of toxicity; pre-specified probability of toxicity at all dose levels; patient doses and toxicity already observed. The prior and model reflect the assumed dose-toxicity relationship and the certainty it is correct based on the information available before the first patient is treated. The first patient is treated at a preselected dose (or treatment level), and, after observing if the patient experiences dose-limiting toxicity (DLT) or not, the dose-toxicity curve is fit. The next patient will be treated at the dose at which the toxicity probability predicted by the dose-toxicity curve is close to the target toxicity probability, subject to rules that moderate the speed of escalation. The step is then repeated until the last patient is treated. The number of patients is pre-specified.

Extended CRM was proposed after O'Quigley's first introduction of CRM in 1990. O'Quigley et al. [5] proposed a two-group likelihood-based CRM in 1999 to determine the MTD for two groups with different toxicity risks. In many phase 1 oncology studies, the risk for toxicity among patients is different due to variation in disease or other predispositions. For example, it has been observed that younger patients better tolerate higher doses of treatment for acute leukemia [5]. If patient heterogeneity exists in the study sample, the one-parameter CRM design will ignore the patient heterogeneity and give an average estimated MTD. O'Quigley et al.'s proposed two-

group CRM design applied two-parameter models and maximum likelihood estimation to locate the MTD for each group without conducting two separate trials. Economically, a two-group design has advantages regarding human and financial resources over conducting two separate trials. Later, O'Quigley and Paoletti [6] studied CRM for two groups with ordered risk. In this study, O'Quigley and Paoletti included the ordering information into the dose allocation strategy in the first stage and the prior for the parameter representing the group difference, to locate the MTD for each group more efficiently, especially with small sample sizes.

The time-to-event continual reassessment method (TITE-CRM) proposed by Cheung and Chappell [7] in 2000 is an extension of CRM appropriate for treatments where the toxicities are delayed in onset. In most designs, a participant cannot be recruited until the previous participant (or the paticipants in the previous cohort) has been fully observed. In trials for cancer treatments such as radiation, it may take months to observe an event for a participant, resulting in impractically long trials. On the other hand, if the observation period defined in the study protocol is not long enough to observe all likely events, some toxicity events will not be considered with respect to the dose escalation decision. TITE-CRM manages late-onset toxicities by allowing subjects to be enrolled as they present, irrespective of the observational status of patients already on trials, thereby increasing trial efficiency.

Salter et al.[8] further studied the TITE-CRM design with two-parameter models between two risk groups using likelihood estimation. The likelihood estimation approach required at least one toxic and one non-toxic observation in each group. Hence, a two-stage trial, an initial one-group dose escalation stage followed by a two-group TITE-CRM stage, was employed to meet the requirements in the study. The study results showed that the TITE-CRM can accommodate two risk groups with different MTDs within a trial.

The objective of our study is to re-parameterize the dose-toxicity model used in the TITE-CRM among groups (two or more than two) with ordered toxicity risk, using Bayesian estimation. In our study, we will use the same rules throughout the study, without separating the trial into different stages.

CRM and TITE-CRM will be reviewed in more details in Section 2. Our study method will be described in Section 3. Simulations will be conducted and the results will be presented in Section 4. The results are to be further discussed in Section 5.

2.0 LITERATURE REVIEW

CRM was proposed to allocate more patients to dose levels with estimated probability of toxicity close to a predefined target probability, to increase the therapeutic benefit in cancer trials at an acceptable risk level [2]. While the 3+3 uses the information gathered at the current dose level to decide which dose level the next patient will receive, CRM designs use all the information available to assess the relationship between dose and toxicity and recommend a dose level for the next patient. The methods of CRM and its extensions are reviewed in this section.

2.1 CRM

In all phase 1 dose-escalation trials, dose-limiting toxicity is a dichotomous event defined in the trial protocol by a class of adverse events (generally characterized in terms of severity of event and relatedness to treatment) occurring in a specified time frame starting at the beginning of treatment. A patient who has experienced toxicity is denoted by y_j=1 and a patient who has not by y_j=0. Suppose there are k dose levels denoted d_i (i = 1, …, k) on clinical scale and n patients recruited in a trial. Suppose π^* denotes the target probability of toxicity, which is assumed to be given (generally specified based on assumptions about the costs of likely toxicities compared to the perceived benefit of treatment). A function $\pi(x, \alpha)$ is assumed to describe the relationship between dose and the probability of toxicity, where π is monotonic in x and α, where $\alpha>0$, and x is a numerical relabled dose representing clinical dose or treatment level d used in the dose-toxicity function, where $x_i=\pi(\pi_{0i},E(\alpha))^{-1},\{\pi_{01},...,\pi_{0k}\}$ is the *skeleton*, the expected probabilities of toxicity

before any data are collected and E(α) is the expected value of α under the prior. The values $x_1,...x_k$ are not updated as data are collected, instead, the dose-toxicity function is updated. It is assumed there exist α and x such that $\pi(x, \alpha) = \pi^*$. The true α in the dose-toxicity function is contantly estimated by the posterior mean $\hat{\alpha}$ when a toxicity outcome is observed. Suppose $f(\alpha)$ is the prior distribution of α and contains all the information about the dose-toxicity function at the intitiation of the trial. Let x_j denote the relabeled dose for jth participant. After we have observed outcomes for J patients, the posterior distribution of α is assumed to be Equation (1). Then, we plug posterior mean $\hat{\alpha}$, Equation (2), into the dose-toxicity equation to get the plug-in estimates of π_{iJ+1}, Equation (3), of the probability of toxicity for each dose level i for the J+1th patient. The recommended dose level for patient J+1 will be the level that minimizes $(\pi_{iJ+1} - \pi^*)^2$.

$$f(\alpha, y_1, \ldots, y_J) \propto f(\alpha) \prod_{j=1}^{J} \pi(x_j, \alpha)^{y_j} \{1 - \pi(x_j, \alpha)\}^{1-y_j} \qquad (1)$$

$$\hat{\alpha} = \int_{\Omega_\alpha} \alpha f(\alpha, y_1, \ldots, y_J) d\alpha \qquad (2)$$

$$\pi_{i,J+1} = \pi(x_{i,J+1}, \hat{\alpha}) \qquad (i = 1, .., k) \qquad (3)$$

These steps are repeated until the last patient has been fully observed.

According to O'Quigley [2], the prior distribution of the parameter α should reflect the knowledge of dose-toxicity before the trial begins. In his study, $f(\alpha) = \exp(-\alpha)$ was chosen as the prior distribution for its positive parameter space and simplicity; truncated normal and log-normal were also considered by O'Quigley. A one-parameter hyperbolic model (4) and a two-parameter logistic model were considered as dose-toxicity functions and the operating characteristics were compared among these models. In his simulations, he concluded that the one-parameter hyperbolic model outperformed the two-parameter logistic model in terms of operating characteristics.

$$\pi(x_i, \alpha) = \{\frac{(tanh\, x_i + 1)}{2}\}^{\alpha} \qquad (4)$$

The sample size for O'Quigley's simulation study was 25, as the author considered the typical sample sizes of 3+3 trials (generally, 12-18) would be too small. The starting dose was the level at which the probability of toxicity was equal to the target probability. The doses given to patients were dominated by the CRM. However, the proposed method in the O'Quigley's study cannot be fully practiced in reality due to the patient safety issue of overdose. In typical clinical practice, the starting dose is the lowest dose level and no dose should be skipped between two adjacent patients for safety considerations.

O'Quigley's following study involved point estimation and interval estimation for CRM [9], and the mean squared error and normal approximated confidence intervals for both maximum likelihood estimators and Bayesian estimators were used to assess the precision of the estimators. It was concluded that the simulations for both estimators had similar results and the Bayesian estimates and intervals slightly outperformed the maximum likelihood estimates and intervals.

After the introduction of CRM, Korn et al.[10] conducted a study to compare the CRM and the 3+3 and concluded that the CRM trial took too long and tended to treat patients at higher dose levels than the true MTD. He suggested recruiting more patients than one at a time at one dose level and choosing a dose at which the posterior probability is less than and closest to the target probability in the CRM trials.

2.2 CRM FOR PATIENT HETEROGENEITY

2.2.1 CRM for two groups

As an extension of the original CRM, a two-group continual reassessment method using maximum likelihood estimation for patients with toxicity risk heterogeneity was introduced by O'Quigley et al. [5] in 1999. The two-group CRM was referred as two-sample CRM. The purpose of the two-sample CRM was to find the appropriate dose for each group using one working model.

Suppose there are k dose levels d_i ($i = 1, \ldots, k$) and assume two groups share the dose levels. Suppose there are n patients that have received doses ($j = 1, \ldots, n$), where n_1 patients are in group 1 and n-n_1 patients are in group 2. Let x_j denote the relabeled dose for jth patient and I denote an indicator variable for group, $I = 1$ for group 1 and $I = 2$ for group 2. Let y_j denote the toxicity outcome for the jth patient. The dose-toxicity model satisfies Equation (5),

$$\begin{cases} P(Y = 1|x, I = 1) = \pi_1(x, \alpha) \\ P(Y = 1|x, I = 2) = \pi_2(x, \alpha, \beta) \end{cases} \tag{5}$$

where π_1 is the dose-toxicity equation for group 1 and π_2 is the dose-toxicity equation for group 2, α represents the common dose-toxicity information shared between groups and β represents the group difference. There exists (α, β) such that $\pi_1(x_i, \alpha) = \pi^*$ and $\pi_2(x_l, \alpha, \beta) = \pi^*$, where π^* is the target toxicity probability. The true α and β were estimated by the maximum likelihood estimators. The likelihood for the n observations can be written as Equation (6).

$$\prod_{j=1}^{n_1} \pi_1(x_j, \alpha)^{y_j} \{1 - \pi_1(x_j, \alpha)\}^{1-y_j} \prod_{j=n_1+1}^{n} \pi_2(x_j, \alpha, \beta)^{y_j} \{1 - \pi_2(x_j, \alpha, \beta)\}^{1-y_j} \tag{6}$$

The recommended dose for the n+1th patient will be the dose level that minimizes $|\pi_1(x_{n+1}, \hat{\alpha}) - \pi^*|$ if the n+1th patient goes to group 1, and the dose level that minimizes $|\pi_2(x_{n+1}, \hat{\alpha}, \hat{\beta}) - \pi^*|$ if the n+1th patient goes to group 2, where $\hat{\alpha}$ and $\hat{\beta}$ are the maximum likelihood estimates.

In the O'Quigley proposal [5], patients were recruited randomly with equal probability to two groups. Because their method used maximum likelihood estimation that required heterogeneity among patient responses to ensure the existence of maximum likelihood estimate [11], a two-stage design was employed, with a first stage similar to the 3+3 dose-escalation design and a second stage as CRM. Both groups started at dose level 1 and moved one level up if two consecutive non-toxicity outcomes were observed. Dose escalation proceeded separately in the two groups until toxicity was observed in both groups and then the trial proceeded as described above.

Two variants of the hyperbolic model (4) used in the original CRM paper shown below were used to estimate the probability of toxicity at each dose level.

$$\begin{cases} \pi(x,\alpha) = \{\frac{tanh\, x+1}{2}\}^{\alpha} & \alpha > 0 \\ \pi(x,\alpha,\beta) = \{\frac{tanh\, x+1}{2}\}^{(\alpha+\beta)} & \alpha > 0, -\alpha < \beta < \infty \end{cases} \tag{7}$$

$$\begin{cases} \pi(x,\alpha) = \{\frac{tanh x+1}{2}\}^{\alpha} & \alpha > 0 \\ \pi(x,\alpha,\beta) = \{\frac{tanh(x-\beta)+1}{2}\}^{\alpha} & \alpha > 0, -\infty < \beta < \infty \end{cases} \tag{8}$$

The sample size was 32, and each group had at least 10 participants. The recommended dose level was the one after the last patient was treated. According to the study results, the two-group CRM performed better than two separate CRM trials or one-merged CRM trial when two groups of patients had differences while sharing some common features.

2.2.2 CRM for ordered groups

The two-group CRM was extended further by O'Quigley and Paoletti [6] using patient heterogeneity with ordered risk. In the two-sample CRM, the ordering information of the groups is unknown. For the ordered group study, the ordering information was included into the design to increase trial efficiency. For the two ordered groups, a two-stage design was applied using a model that was similar to Equation (5). Patients from the lower risk group were treated with dose levels greater than or equal to the higher risk group in the first-stage dose escalation, until dose limiting toxicity was observed. The second stage featured either a maximum likelihood-based or Bayesian CRM. The ordering information was included in the prior when Bayesian CRM was applied at the second stage. For example, a gamma or normal distribution was used as the prior of β to reflect that the risk difference was positive in the study. Given the toxicity risk of group 1 was larger than that of group 2, the prior of β should satisfy

$$\pi_1(x_k, \alpha) \simeq \pi_2(x_{k+l}, \alpha, \mu_\beta) \simeq \pi* \qquad (9)$$

where k and $k+l$ are dose levels, $l \geq 0$, representing the difference between two dose levels, π^* is target probability of toxicity, α and μ_β are unknown and need to be estimated. If μ_β is 0, it is assumed that there is no risk difference between the groups and the result depends largely on the data. σ_β represents the standard deviation of β and reflects how informative the prior is. The larger σ_β, the vaguer the prior. A set of σ_β was tested in the simulation and the impact of prior was assessed.

Results showed that under correct modelling assumptions, the ordered group CRM outperformed the two-group likelihood-based CRM. When the assumptions were incorrect and the prior was not informative, the ordered group CRM performed better due to the information

included in the first-stage dose escalation. When the assumptions were incorrect and the prior was informative, the assumptions impacted the running of trial, but the final recommendation dose was informative and not incorrect.

2.3 TITE-CRM

In the CRM design, a patient is recruited and treated based on the toxicity results of all previously treated patients (although small cohorts could be introduced, they are not part of the modeling). If the current patient has not experienced DLT or reached the end of observation period, the next patient's treatment must be delayed, creating recruitment (eligible patients might not be enrollment) and trial management (the trial must be repeatedly opened and closed to accrual) issues. TITE-CRM was introduced for clinical trials with late-onset toxicities by Cheung and Chappell [7] to deal with this problem.

Cheung and Chappell extended the CRM by including a weighted dose-response model. In their study [7], a weight w, which is a function of the time since treatment, was added to the parametric model used in the CRM. Let $g(x, w, \alpha) = w \cdot \pi(x, \alpha)$, where $0 \leq w \leq 1$, g is monotonically increasing in w, and $\pi(x, \alpha)$ is the parametric dose-toxicity model. w is the fraction of the observation time for patient over the total observation period pre-specified in protocol. Hence, the likelihood for weighted CRM for J observations is

$$\prod_{j=1}^{J} g(x_j, w_j, \alpha)^{y_j} \{1 - g(x_j, w_j, \alpha)\}^{1-y_j} \qquad (10)$$

.

Additionally, Cheung and Chappell used an initial dose escalation stage for both maximum likelihood estimation and Bayesian estimation, letting three patients start at the lowest dose level

and escalating to the next dose if no toxicity was observed until the first toxicity occurred, then switching to TITE-CRM. In their simulated trials, the total observation period was set to 6 months, and in the initial stage, 3 patients were recruited at 6-month intervals and in the TITE-CRM stage, a patient was recruited every 0.5 months. The power model $\pi(x, \alpha) = x^{\alpha}$ was used in the simulation. An exponential distribution was used for the prior. The patients' failure times were generated from a uniform distribution, log-logistic distribution and Weibull distribution in different simulations. The results showed that the correct dose recommendation percentages were similar under failure time generated using different distributions. The results of the TITE-CRM were comparable to that of the CRM, whereas the trial duration was shortened under the TITE-CRM scheme. The authors also suggested that the two-stage designs should be used for safety concern.

Later, Normolle and Lawrence [12] simulated trials to further study the operating characteristics of the one-stage TITE-CRM compared to the 3+3. In their study, a logistic model and a normal prior with mean 0 and standard deviation as 0.1 or 0.3 were used for the Bayesian TITE-CRM. There was no initial cohorts-of-3 escalation. The usual rules to control aggressive dose escalation in CRM stated in Section 2.1 were applied. Additionally, the rule that dose escalation for the next patient was not allowed until the previous patient had experienced DLT or had finished the whole observation period without DLT was also added. The sample size was 36. The results indicated that TITE-CRM was susceptible to underestimation of toxicity when there were abrupt jumps in the true dose-toxicity function. When the assumed dose-toxicity function is close to the true state of nature, the toxicity risk for the TITE-CRM was not greater than that for the 3+3.

2.4 TITE-CRM FOR PATIENT HETEROGENEITY

Patient heterogeneity regarding toxicity risk exists in phase 1 oncology trials. Previously, O'Quigley et al. studied the two-group CRM which was based on the maximum likelihood method. In 2015, Salter et al. [8] extended the TITE-CRM to two groups with toxicity risk.

A two-parameter dose-toxicity model with additive relationship between two groups was used,

$$\begin{cases} P(Y=1|x,I=1) = g(x,w,\alpha) \\ P(Y=1|x,I=2) = g(x,w,\alpha+\tau) \end{cases} \tag{11}$$

where *I* was the indicator variable for group 1 and group 2, *g* was weighted dose-toxicity function, α represented the shared information between two groups and τ represented the difference between the two groups. Parameter estimation was based on the maximum likelihood approach and the two-stage method proposed by O'Quigley [5] was applied. Therefore, the likelihood after *n* observations is given as:

$$L_n = \prod_{j=1}^{n_1} g(x_j,w_j,\alpha)^{y_j}\{1-g(x_j,w_j,\alpha)\}^{1-y_j} \prod_{j=n_1+1}^{n} g(x_j,w_j,\alpha+\tau)^{y_j}\{1-g(x_j,w_j,\alpha+\tau)\}^{1-y_j} \tag{12}$$

where n_1 the number of patients observed in group 1, n-n_1 the number of patients observed in group 2. The maximum likelihood estimates $\hat{\alpha}$ and $\hat{\tau}$ are the solutions to the Equations (13).

$$\begin{cases} \frac{\partial log(L_n)}{\partial \alpha} = 0 \\ \frac{\partial log(L_n)}{\partial \tau} = 0 \end{cases} \tag{13}$$

The recommended dose for n+1th patient is the dose level that minimizes $|\pi(x_{n+1}, \hat{\alpha}) - \pi^*|$ if n+1th patient is in group 1, and the dose level that minimizes $|\pi(x_{n+1}, \hat{\alpha} + \hat{\tau}) - \pi^*|$ if n+1th patient is in group 2, where π^* is the target probability of toxicity. The steps are repeated until all subjects are observed.

Empiric functions $f(x, \alpha) = x^{\exp(\alpha)}$ and $f(x, \alpha+\tau) = x^{\exp(\alpha+\tau)}$ were used as the dose-toxicity functions in the simulation. Sample size was 32, with a balanced scenario of 16 patients in each group and an unbalanced scenario of 20 and 12 patients in two groups. The results of the simulation comparing the one-group and two-group scenarios showed that the TITE-CRM can accommodate two groups with different MTDs using maximum likelihood method and the results were comparable, even though the proportion that a true MTD was recommended in the two-group scenario was lower than the one-group scenario due to smaller sample sizes.

2.5 WORKING MODEL AND PRIOR FOR BAYESIAN METHOD

The studies referenced above used various dose-toxicity models, including the one-parameter hyperbolic model (4) which is equivalent to power model $\pi(x, \alpha) = x^{\alpha}$, empiric model (14) and one-parameter and two-parameter logistic models (15).

$$\pi(x_i) = \theta_{i1}^{\alpha} \qquad (i = 1, \ldots, k) \tag{14}$$

$$\pi(x_i) = \frac{exp(\beta + \alpha \cdot x_i)}{1 + exp(\beta + \alpha \cdot x_i)} \qquad (i = 1, \ldots, k) \tag{15}$$

For the empiric model (14), θ_{i1} was the initial estimate for the toxicity probability at each dose level i, which are determined by physicians based on experience before trial. The logistic model (15) is often used due to its flexibility and familiarity to physicians [13]. When β is a constant, Equation (15) is an intercept-fixed one-parameter logistic model. When α is a constant, Equation (15) is a slope-fixed one-parameter logistic model. When both parameters are not fixed, the equation becomes a two-parameter logistic model, which is more flexible than the one-parameter logistic model.

According to the simulated study, compared to the one-parameter hyperbolic model, the empiric model and the logistic model by Chevret [13], the intercept-fixed one-parameter logistic model worked better in terms of the probability of identifying the true MTD, especially when the intercept was equal to 3 for the logistic model. The study by Chevret [13] showed that when intercept was equal to 3, the bias and MSE of estimated probability of DLT were the smallest compared to the model with intercept smaller or larger than 3. The study also compared the impact of different choices of priors such as Gamma(1,2), Exponential(1), Lognormal(1,2), Uniform(0,3) to the CRM and the results showed that as long as the prior was not informative, they had similar effects on the CRM.

3.0 METHOD

The objective of this thesis is to re-parameterize the dose-toxicity model used in the TITE-CRM to accommodate patients with ordered toxicity risk. In order to introduce ordering information into the model, we divided the sample subjects into 2 or more groups according to their toxicity risk instead of 2 groups as referenced in the literature [5, 6, 8]. We conducted simulated studies using the re-parameterized model and parallel one-parameter models to compare the differences of the operating characteristics of one trial with several risk groups versus those of several separate trials. The model will be explained and designed using three groups, but the model is easily extended to more than three.

3.1 BAYESIAN ESTIMATION

Suppose there are k dose levels d_i ($i = 1, \ldots, k$) shared among 3 groups of patients and n patients are recruited in total. Let y_j denote the toxicity outcome for the jth patient. y_j takes the value of 0 for not observing a toxicity event and 1 for observing a toxicity event. Let I denote the group indicator, I =1, 2 or 3. Suppose π^* is the target probability of toxicity. The dose-toxicity function π for the three-group CRM is specified such that $0 \leq \pi(x, I) \leq 1$, and π is increasing in x. The dose-toxicity function for an observation is expressed as $P(d, I, y) = \pi(x, \alpha, \delta_1, \delta_2)$, where α represents the shared information among 3 groups, δ_1 represents the difference between group 1 and group 2, δ_2 represents the difference between group 2 and group 3. The ordering of risk of 3 groups is reflected in Equation (16):

$$\begin{cases} P(d_1, I = l) \le P(d_2, I = l) \le \cdots \le P(d_k, I = l) & (l = 1, 2, 3) \\ P(d_m, I = l) \le P(d_m, I = 2) \le P(d_m, I = 3) & (m = 1, \ldots, k) \end{cases} \tag{16}$$

where *l* denotes the *l*th group, *m* denotes the *m*th dose level. It is assumed that priors of α, δ_1, δ_2 are independent of each other.

For the assumed dose-toxicity function $\pi(x_i, \alpha, \delta_1, \delta_2)$, there exist α, δ_1, δ_2 such that $\pi(x_k, \alpha) = \pi(x_l, \alpha, \delta_1) = \pi(x_m, \alpha, \delta_1, \delta_2) = \pi^*$ $(k \ge l \ge m)$. The true α, δ_1, δ_2 in the dose-toxicity function are estimated continuously throughout the trial when a toxicity outome is observed or not. Because TITE-CRM was used in our study, we added a weight *w* to the dose-toxicity function and let $g(x, w, \alpha) = w \cdot \pi(x, \alpha)$ denote the weighted dose-toxicity function for group 1, $g(x, w, \alpha, \delta_1)$ for group 2 and $g(x, w, \alpha, \delta_1, \delta_2)$ for group 3.

Suppose $f_1(\alpha)$ is the prior distribution of α, $f_2(\delta_1)$ the prior of δ_1 and $f_3(\delta_2)$ the prior of δ_2, and these priors contain the pre-trial information about the dose-toxicity function. The prior joint distribution is then $f_1(\alpha) \cdot f_2(\delta_1) \cdot f_3(\delta_2)$. After we have observed outcomes for *J* out of n patients, where J_1, J_2 and J_3 are the within-group sample sizes, the joint posterior distribution of α, δ_1, δ_2 and posterior mean estimates are as below.

$$\begin{aligned} f(\alpha, \delta_1, \delta_2, y_1, \ldots, y_J) \propto\ & f(\alpha) f(\delta_1) f(\delta_2) \textstyle\prod_{j=1}^{J_1} g(x_j, \alpha)^{y_j} \{1 - g(x_j, \alpha)\}^{1-y_j} \\ & \cdot \textstyle\prod_{j=J_1+1}^{J_1+J_2} g(x_j, \alpha, \delta_1)^{y_j} \{1 - g(x_j, \alpha, \delta_1)\}^{1-y_j} \\ & \cdot \textstyle\prod_{j=J_1+J_2+1}^{J} g(x_j, \alpha, \delta_1, \delta_2)^{y_j} \{1 - g(x_j, \alpha, \delta_1, \delta_2)\}^{1-y_j} \end{aligned} \tag{17}$$

$$\begin{aligned} \hat{\alpha} &= \textstyle\int_{\Omega_\alpha} \int_{\Omega_{\delta_2}} \int_{\Omega_{\delta_1}} \alpha\, f(\alpha, \delta_1, \delta_2, y_1, \ldots, y_J) d\delta_1 d\delta_2 d\alpha \\ \hat{\delta_1} &= \textstyle\int_{\Omega_{\delta_1}} \int_{\Omega_{\delta_2}} \int_{\Omega_\alpha} \delta_1\, f(\alpha, \delta_1, \delta_2, y_1, \ldots, y_J) d\alpha d\delta_2 d\delta_1 \\ \hat{\delta_2} &= \textstyle\int_{\Omega_{\delta_2}} \int_{\Omega_{\delta_1}} \int_{\Omega_\alpha} \delta_2\, f(\alpha, \delta_1, \delta_2, y_1, \ldots, y_J) d\alpha d\delta_1 d\delta_2 \end{aligned} \tag{18}$$

Then we substitue $\hat{\alpha}, \widehat{\delta_1}, \widehat{\delta_2}$ into the dose-toxicity function π to get the estimate π_{n+1} at all dose levels in all risk groups for *n*+1th patient. The recommended dose level for the *n*+1th patient will be the level that minimizes $|\pi_{n+1} - \pi^*|$ for the group the patient is in.

3.2 SIMULATION STUDY

3.2.1 Model re-parameterization

For the simulation study, we used the intercept-fixed logistic model with intercept equal to 3 based on previous studies [13]. Other functions can be used using a similar parametric framework, as long as the oredering of the risk group-specific dose-toxicity functions is guaranteed. The function of the logistic model for three groups is given as.

$$\begin{cases} \pi(x,\alpha,I) = \frac{exp(3+\alpha\cdot x)}{1+exp(3+\alpha\cdot x)} & (I=1) \\ \pi(x,\alpha,\delta_1,I) = \frac{exp\{3+(\alpha-\delta_1)\cdot x\}}{1+exp\{3+(\alpha-\delta_1)\cdot x\}} & (I=2) \\ \pi(x,\alpha,\delta_1,\delta_2,I) = \frac{exp\{3+(\alpha-\delta_1-\delta_2)\cdot x\}}{1+exp\{3+(\alpha-\delta_1-\delta_2)\cdot x\}} & (I=3) \end{cases} \tag{19}$$

We assumed there were eight dose levels shared among groups because there were three risk groups in the trial and the starting dose levels were different for each group. The probabilities of toxicity at each dose in each group that would be given by the physicians at the outset of the trial, are in Table 1.

Table 1. Assumed probabilities of toxicity

Dose	Group		
	1	2	3
1	0.01	0.014	0.020
2	0.025	0.034	0.045
3	0.05	0.065	0.082
4	0.1	0.125	0.151
5	0.15	0.181	0.213
6	0.2	0.235	0.271
7	0.25	0.289	0.325
8	0.3	0.34	0.377

The relabeled doses are given as $x_1, \ldots, x_8$. Those for group 1 are determined by inverting the first equation in (19) when α=1, that is, x_k=log(P_k/(1-P_k))-3 for k=1,...8 and P_k is an entry in the second column (Group 1) of Table 1. The relabeled doses are displayed in Table 2.

Table 2. Relabeled doses from Table 1

Dose	Relabeled Dose
1	-7.595120
2	-6.663562
3	-5.944439
4	-5.197225
5	-4.734601
6	-4.386294
7	-4.098612
8	-3.847298

The same relabeled doses from Group 1 are used in all other risk groups. The difference in P(DLT) between Group 1 and 2 is incorporated into the model by determining the value of δ_1 that minimizes the squared differences between the Group 2 skeleton and a dose-toxicity function with parameter α-δ_1:

$$\sum_{i=1}^{8} \left\{ P(d, I = 2) - \frac{e^{3+(\alpha-\delta_1)\times x_i}}{1 + e^{3+(\alpha-\delta_1)\times x_i}} \right\}^2 \qquad (20)$$

The same method is used to get the initial value for δ_2 by minimizing the function below with the relabeled doses, α (which equals 1) and δ_1, (estimated from the Group 2 probabilities and the Group 1 relabeled doses) and P(DLT) from Group 3 defined in Table 1:

$$\sum_{i=1}^{8} \left\{ P(d, I = 3) - \frac{e^{3+(\alpha-\delta_1-\delta_2)\times x_i}}{1 + e^{3+(\alpha-\delta_1-\delta_2)\times x_i}} \right\}^2 \tag{21}$$

We assume independent normal prior of the parameters[6, 12]; δ_1 and δ_2 are constrained to be greater than or equal to 0:

$$\begin{cases} \alpha \sim N(\mu_\alpha, \sigma_\alpha^2) \\ \delta_1 \sim N(\mu_{\delta_1}, \sigma_{\delta_1}^2) \qquad (\delta_1 \geq 0) \\ \delta_2 \sim N(\mu_{\delta_2}, \sigma_{\delta_2}^2) \qquad (\delta_2 \geq 0) \end{cases} \tag{22}$$

The hyperparameters in Equation (22) are:

$$\mu_\alpha = 1, \sigma_\alpha = 0.3, \mu_{\delta_1} = 0, \sigma_{\delta_1} = 0.1, \mu_{\delta_2} = 0, \sigma_{\delta_2} = 0.1$$

The prior for α is relatively uninformative. The standard deviations for δ_1 and δ_2 are set to 0.1 to ensure that the probability for toxicity would not become unrealistically low.

3.2.2 Trial setup

We conducted 100 simulated trials for three situations: 1) the true probability of toxicity at a dose level P(DLT) is the same as the assumptions (Table 1); 2) the true P(DLT) is higher than the assumption; 3) the true P(DLT) is higher than the assumption and there is a bigger difference of probability between the dose level 4 and 5. The true P(DLT) for group 1 (that is, the probabilities of DLT used to simulate toxicities) were manually specified and that for group 2 and 3 were calculated based on a pre-specified true α, δ_1 and δ_2 as indicated in Table 3 using Equation (19). The P(DLT) of dose 4 and 5 in simulation 3 were manually set to have a larger difference. We

used the assumed probabilities and relabeled dose to estimate the P(DLT) in 3 simulations and compared those estimates to the true probabilities.

Table 3. True parameters

Parameters	Sim 1	Sim 2	Sim 3
α	1	0.94	0.9
δ_1	0.048	0.035	0.078
δ_2	0.042	0.07	0.047

The true probability of toxicity are stated in Table 4 and can be visualized in Figure 1. In Figure 1, the brown reference line indicates the target P(DLT). There is a big jump of P(DLT) from dose 4 to 5 across the target P(DLT) in simulation 3.

Table 4. True probabilities of toxicity for 3 groups in 3 simulations

Dose	Simulation 1			Simulation 2			Simulation 3		
	1	2	3	1	2	3	1	2	3
1	0.01	0.014	0.020	0.016	0.02	0.034	0.021	0.038	0.053
2	0.025	0.034	0.045	0.037	0.046	0.071	0.047	0.077	0.103
3	0.05	0.065	0.082	0.07	0.085	0.123	0.087	0.132	0.167
4	0.1	0.125	0.151	0.132	0.154	0.208	0.127	0.182	0.224
5	0.15	0.181	0.213	0.19	0.217	0.278	0.237	0.308	0.357
6	0.2	0.235	0.271	0.245	0.275	0.34	0.279	0.353	0.401
7	0.25	0.289	0.325	0.299	0.33	0.396	0.334	0.404	0.452
8	0.3	0.34	0.377	0.351	0.381	0.447	0.386	0.453	0.499

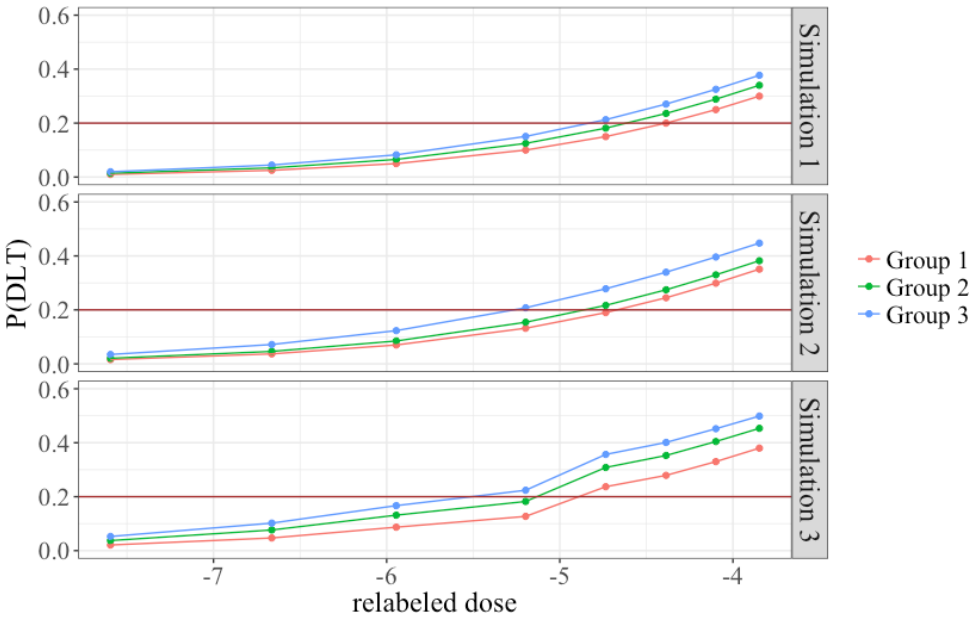

Figure 1. True dose-toxicity relationship generated for 3 simulations

The sample sizes were set to be 40, 80 and 120 and each group were constrained to have at least 10 patients in all groups at the end of the study. For multi-group studies, sample sizes targeting between 15 and 30 participants per risk group should be considered; it is unrealistic to expect useful parameter estimates with smaller sample sizes.

We assumed that patients with different characteristics came to trials at random. The toxicity risk factors are pre-specified according to the patient characteristics and the patients can be grouped according to their characteristics. As an example, in radiotherapy for lung cancer, larger volumes of normal lung tissue exposed to radiation (determined by the radiation treatment plan) is associated with greater risk of toxicity. Patients recruited for lung cancer treatment could be grouped into several risk groups determined by how much normal lung tissue will be under radiation therapy. For our simulation, we considered these as two scenarios: 1. The patients with different risk come to each risk group with the same probability (0.33), which means there are same number of patients in each risk group; 2. The patients with different risk come to each risk

group with different probabilities (0.2, 0.35, 0.45), which means more patients are in the higher risk groups.

Uniform distribution, Unif (0, b), where b is the pre-specified observation days, was used to generate the time of arrival. Bernoulli(p) distribution was used to generate toxicity event where p is the true probabilities of toxicity of each risk group.

The same rules used for TITE-CRM [12] and CRM trials were employed in the simulation for the highest risk group. Additionally, the following rules were applied to the simulated trials to reflect ordering information.

Suppose there are k dose levels d_i (i = 1, …, k) shared among groups I (I = 1, 2, 3) and n patients x_j (j = 1, …, n) recruited in total as mentioned in section 3.1.

1. The starting dose level is set to 1 for the highest risk group, 2 for the medium risk group and 3 for the lowest risk group. If no patient has been treated and observed in group I or any group larger than I, the first patient in that group will be assigned to that starting level.

2. When participant j in group I has completed observation at level d_i, participant j+1 in group I may be assigned to level d_i+1.

3. When participant j in group I_j>1 has completed observation at level d_i, participant j+1 in any group less than or equal to I_j may be assigned to level d_i+1.

In parallel trials for different risk groups, we applied the usual rules used in TITE-CRM trial to constrain aggressive dose escalation [12]. The sample sizes for a separate trial were set as 13, 26 and 40 for scenario 1. The sample size for separate trials were 8, 16 and 24 for group 1, 14, 28 and 42 for group 2, and 18, 36, 54 for group 3 for scenario 2.

Point and interval estimates of α, δ_1 and δ_2 were calculated by using Gibbs sampling to draw samples from the posterior distribution (17) using JAGS (Just Another Gibbs Sampler, v.

4.3.0), called through the R package rjags (v. 4-6) under R v. 3.4.2. The mean of the posterior sample was used as the point estimate of parameters. The interval estimates of the parameters were the 90% credible intervals from the posterior samples. The point and 90% interval estimates of P(DLT) were drawn from the P(DLT) samples calculated using parameter samples. Finally, the estimates for the last observation in each trial were used for analysis. A sample of R code and jags code for the simulation is provided in Appendix B.

4.0 RESULTS

We conducted the simulation study using two scenarios for both a group trial and a parallel trial. For scenario 1, there were same number of patients in each risk group. For scenario 2, there were more patients in the higher risk groups. For each scenario, we performed 3 simulations and for each simulation, we performed 100 trials for each of three different sample sizes. The results were analyzed and compared in terms of the estimated parameters and P(DLT), the MTD recommendation, the in-trial dose distribution and the distribution of DLTs. The sections that contain the results for each situation are summarized in Table 5.

Table 5. Sections of results corresponding to each situation

	Results	Section
Scenario 1	Estimation:	4.1.1
Group trials	α	4.1.1.1
	δ_1	4.1.1.2
	δ_2	4.1.1.3
	P(DLT)	4.1.1.4
	MTD recommendation	4.1.1.5
	In-trial dose distribution	4.1.1.6
	Distribution of DLT	4.1.1.7
Parallel trials	Estimation:	4.1.2
Group 1	α	4.1.2.1
Group 2	α	4.1.2.2
Group 3	α	4.1.2.3

Table 5 Cont'd		
	P(DLT)	4.1.2.4
	MTD recommendation	4.1.2.5
	In-trial dose distribution	4.1.2.6
	Distribution of DLT	4.1.2.7
Scenario 2	Estimation:	4.2.1
Group trials	α	4.2.1.1
	δ_1	4.2.1.2
	δ_2	4.2.1.3
	P(DLT)	4.2.1.4
	MTD recommendation	4.2.1.5
	In-trial dose distribution	4.2.1.6
	Distribution of DLT	4.2.1.7
Parallel trials	Estimation:	4.2.2
Group 1	α	4.2.2.1
Group 2	α	4.2.2.2
Group 3	α	4.2.2.3
	P(DLT)	4.2.2.4
	MTD recommendation	4.2.2.5
	In-trial dose distribution	4.2.2.6
	Distribution of DLT	4.2.2.7

4.1 SCENARIO 1: EQUAL NUMBER OF PATIENTS IN EACH RISK GROUP

4.1.1 Group trials

4.1.1.1 Estimation of α

The distribution of estimated α is displayed in Figure 2. The black square represents the true α in each simulation. As sample size increases, the range of the α estimates gets narrower, and the median and mean sample α get closer to the true α. The difference between mean estimated α and true α is the smallest in simulation 3, where there is an abrupt jump in P(DLT) between dose 4 and 5.

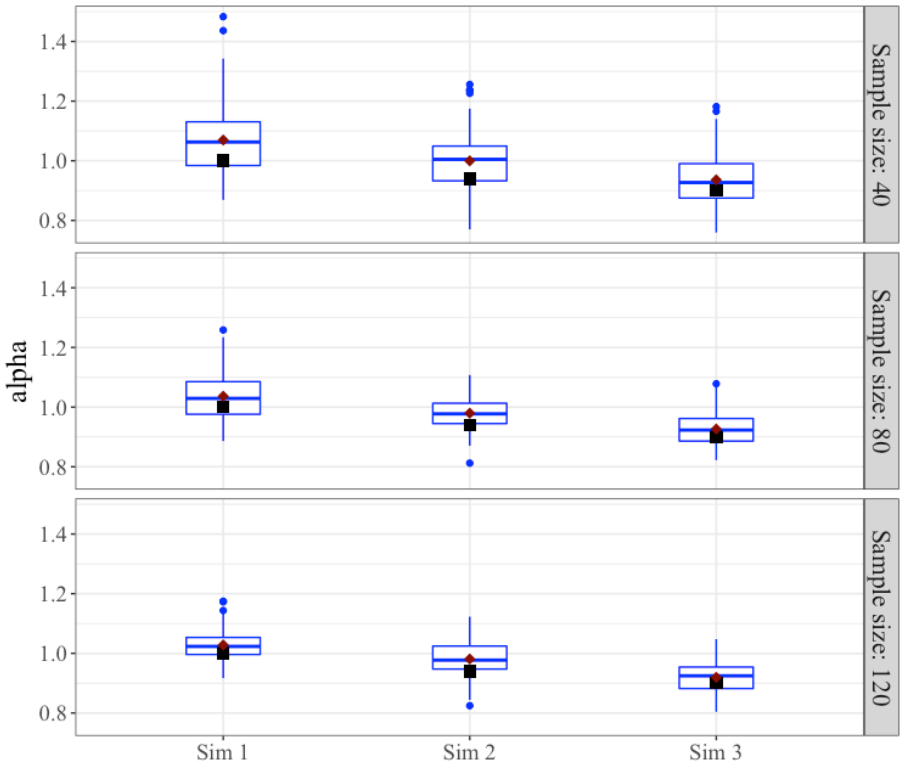

Figure 2. Distribution of estimated α in group trials
Black squares represent the true α and red hearts represent sample means. The box indicates 25th, 50th and 75th percentiles and whisker limits show the minimum and maximum. The dots are outliers.

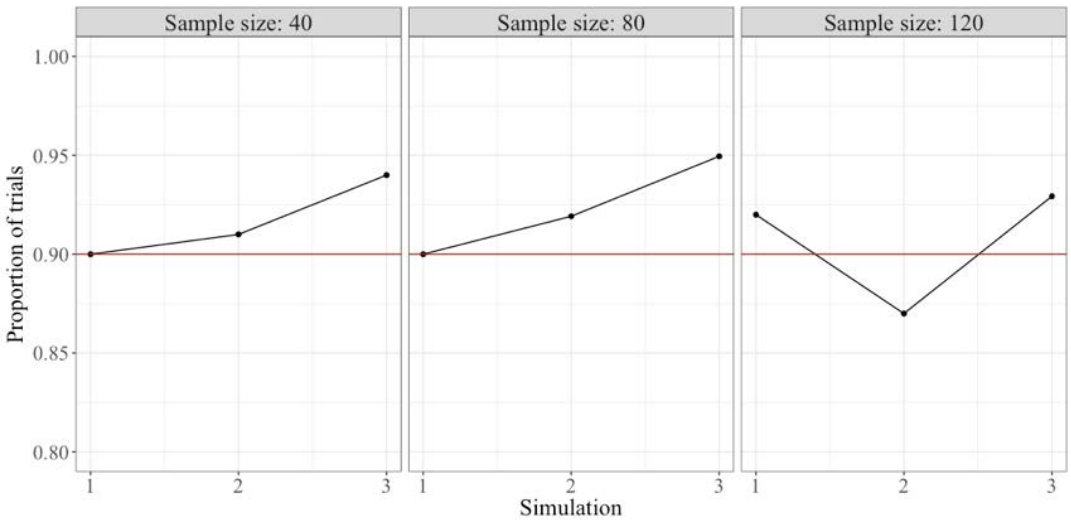

Figure 3. Coverage of 90% credible interval of true α in group trials

The coverage of the 90% credible interval of true α is displayed in Figure 3. The coverage of α for group trials is above or around 90% in all simulations of all sizes, indicating a good quality of interval estimation.

4.1.1.2 Estimation of δ_1

The distribution of estimated δ_1 is displayed in Figure 4. The black square represents the true δ_1 in each simulation. As sample size increases, the range of δ_1 estimates remains wide. The median and mean sample δ_1 get closer to the true δ_1 in simulation 3 only.

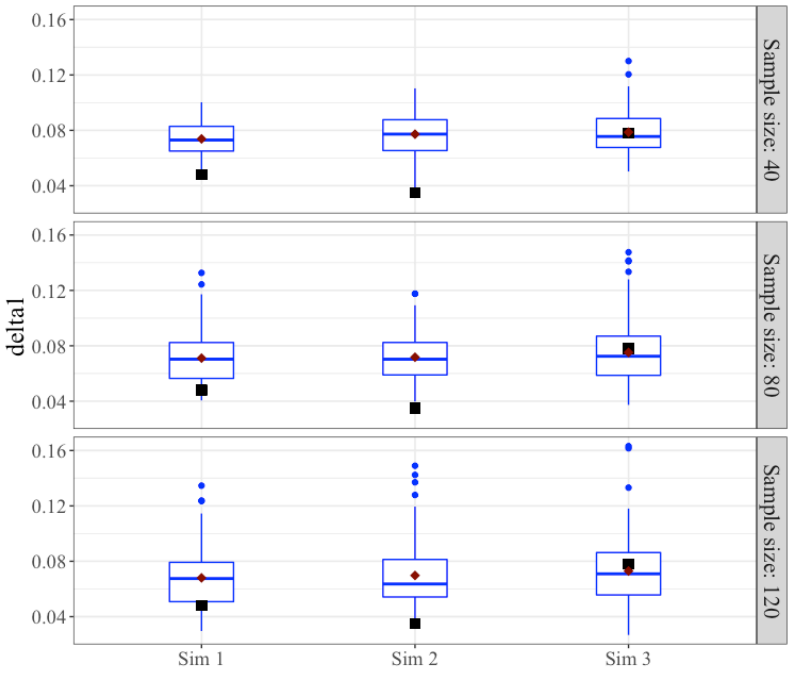

Figure 4. Distribution of estimated δ_1 in group trials

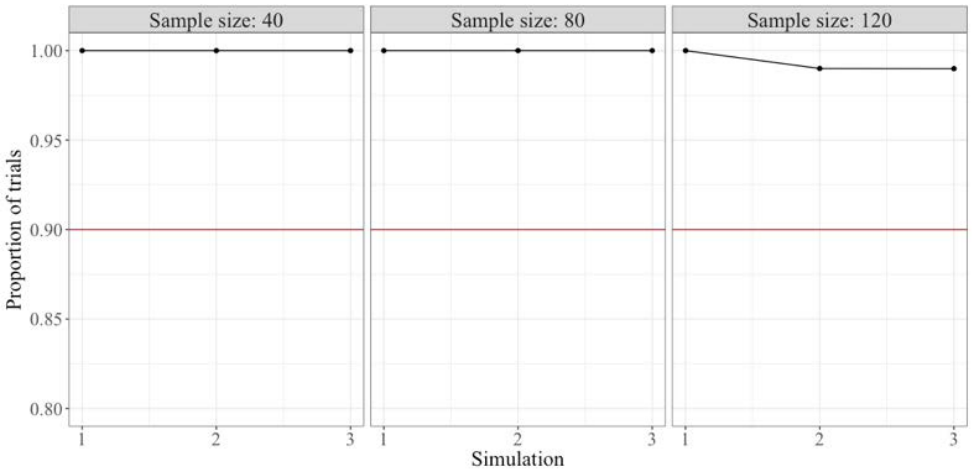

Figure 5. Coverage of 90% credible interval of true δ_1 in group trials

The coverage of 90% credible interval of δ_1 is close to 100% due to wide credible intervals, indicating an imprecise estimation of δ_1.

4.1.1.3 Estimation of δ_2

The distribution of estimated δ_2 is displayed in Figure 6. The black square represents the true δ_2 in each simulation. As sample size increases, the range of δ_2 estimates still remains wide. The median and mean sample δ_2 get closer to the true δ_2 in simulation 2 only. The true values are less than the sample mean in all situations.

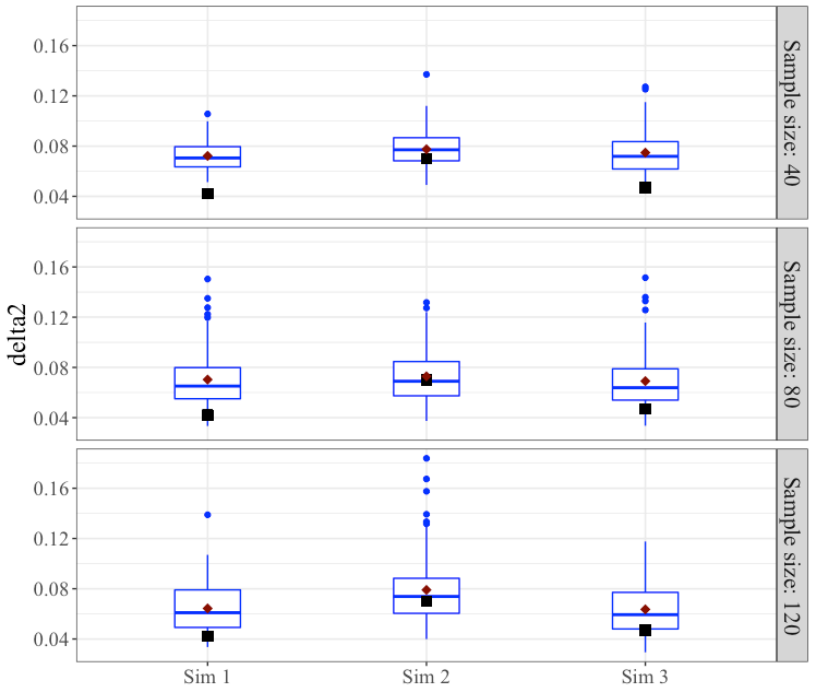

Figure 6. Distribution of estimated δ_2 in group trials

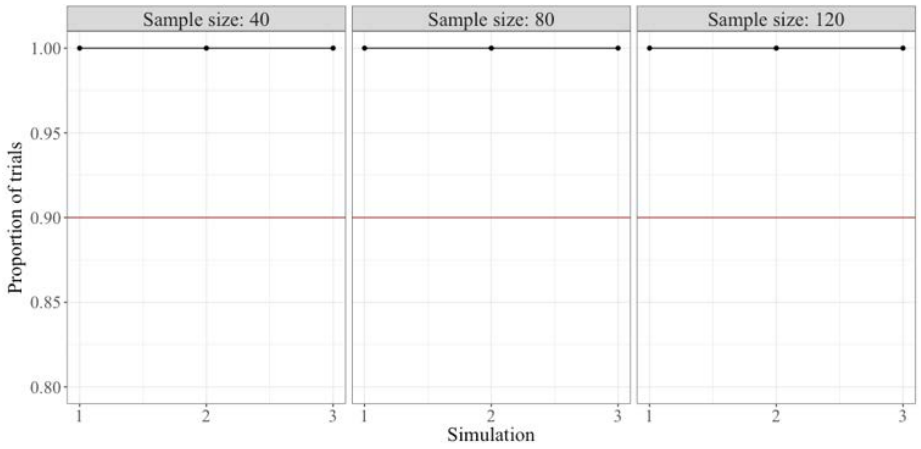

Figure 7. Coverage of 90% credible interval of true δ_2 in group trials

The coverage of 90% credible interval of δ_2 is 100% due to the wide credible intervals, indicating an imprecise estimation of δ_2.

4.1.1.4 Estimation of P(DLT)

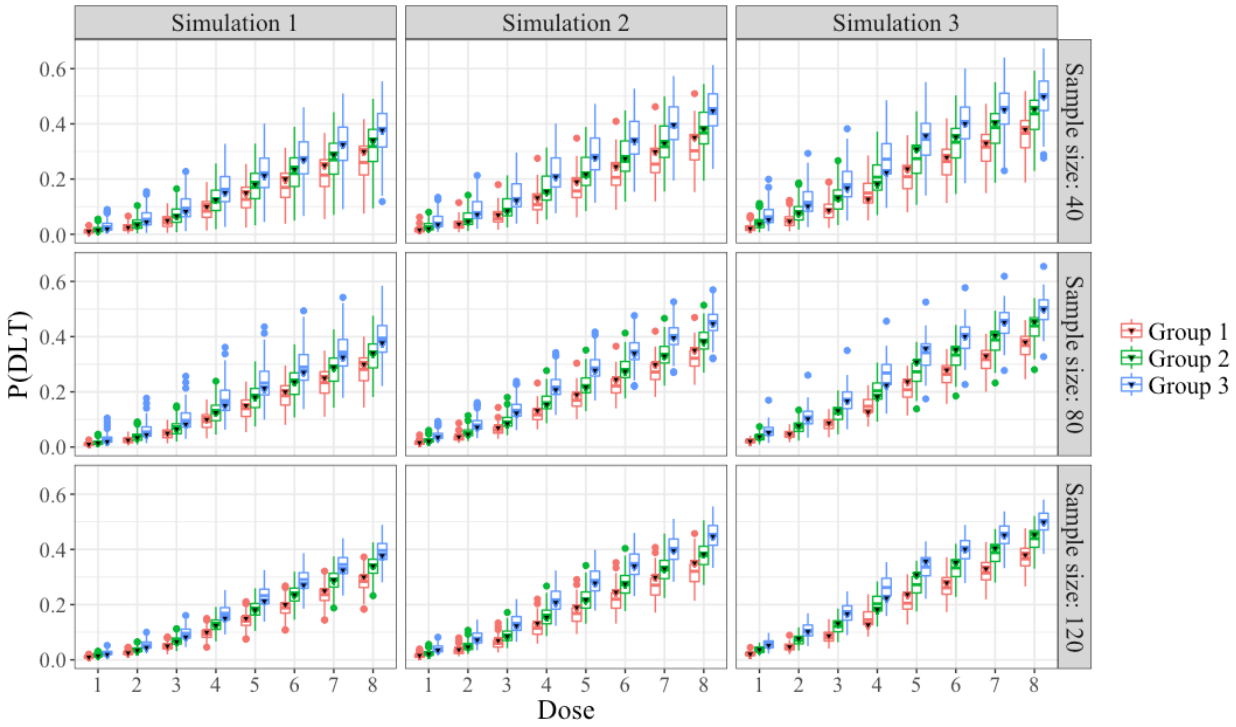

Figure 8. Distribution of estimated P(DLT) for group trials.
The black triangle refers to the true P(DLT)

The distribution of P(DLT) is displayed in Figure 8. The black triangle represents the true P(DLT) in each group at each dose level. The range of estimates gets narrower as sample size increases in each simulation. The mean P(DLT)s are close to the true P(DLT)s except for dose 4 and 5 in simulation 3.

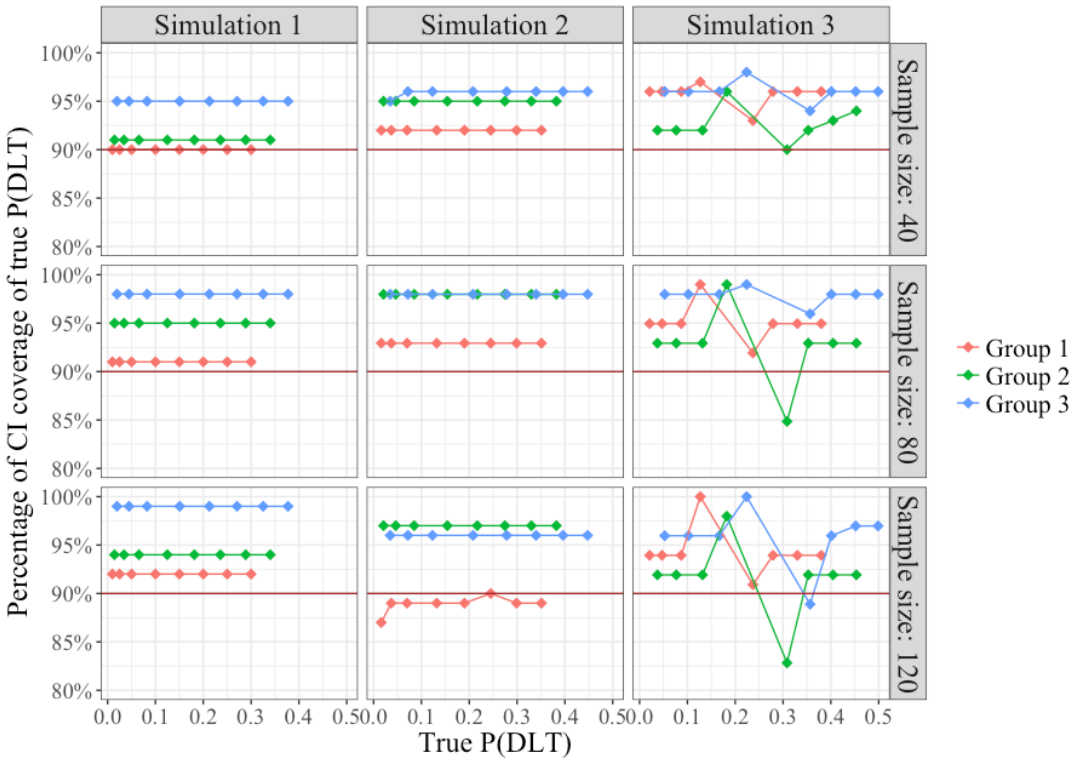

Figure 9. Coverage of 90% credible interval of true P(DLT) for group trials (equal number of patients in each risk group)

The coverage of the 90% credible intervals of true P(DLT) at each dose in each group are displayed in Figure 9. The coverages of true P(DLT) at most doses in all simulations for all sample sizes are around or above 90%, except the estimated P(DLT) at dose 5 in simulation 3. As the sample size increases in simulation 3, the coverage of true P(DLT) at dose 5 in simulation 3 seems to be getting worse, while the coverage of true P(DLT) at dose 4 is becoming better. One possible reason is the big difference of P(DLT) between dose 4 and dose 5. The method cannot produce a

good estimate of P(DLT) at dose 5 in simulation 3 probably because, as the sample size increases, more patients are assigned to dose 4 at which the P(DLT) is around 0.2, much less toxic than dose 5.

4.1.1.5 MTD recommendation

The proportion of trials in which the true MTD is recommended is summarized in Figure 10. The largest dose with P(DLT) ≤ target P(DLT) in the graph below refers to the maximum dose level at which the P(DLT) ≤ 0.2 in a group in a trial. The proportion of trials in a single simulation that the recommended dose level is the MTD is displayed. In simulation 1, most trials recommend the true MTD in group 2 and 3 while most trials in group 1 tend to recommend a higher or lower dose level than the true MTD with different sample sizes. In simulation 2, most trials in group 2 recommend the true MTD and most trials in group 3 recommend a lower dose level than the true MTD with all sample sizes, while most trials in group 1 recommend the true MTD for sample size 40 and 80. In simulation 3, the sample size 80 has the best result in terms of the proportion of trials recommending the true MTD. Most trials with sample size 120 in simulation 3 tend to recommend a lower dose level than the true MTD for group 1 and 3, because dose level 5 in simulation 3 for sample size 120 see most DLTs as Figure 12 shows. Ideally, the highest proportion of trials should recommend the true MTD, but it is acceptable from a safety perspective to recommend a slightly lower dose level than the true MTD.

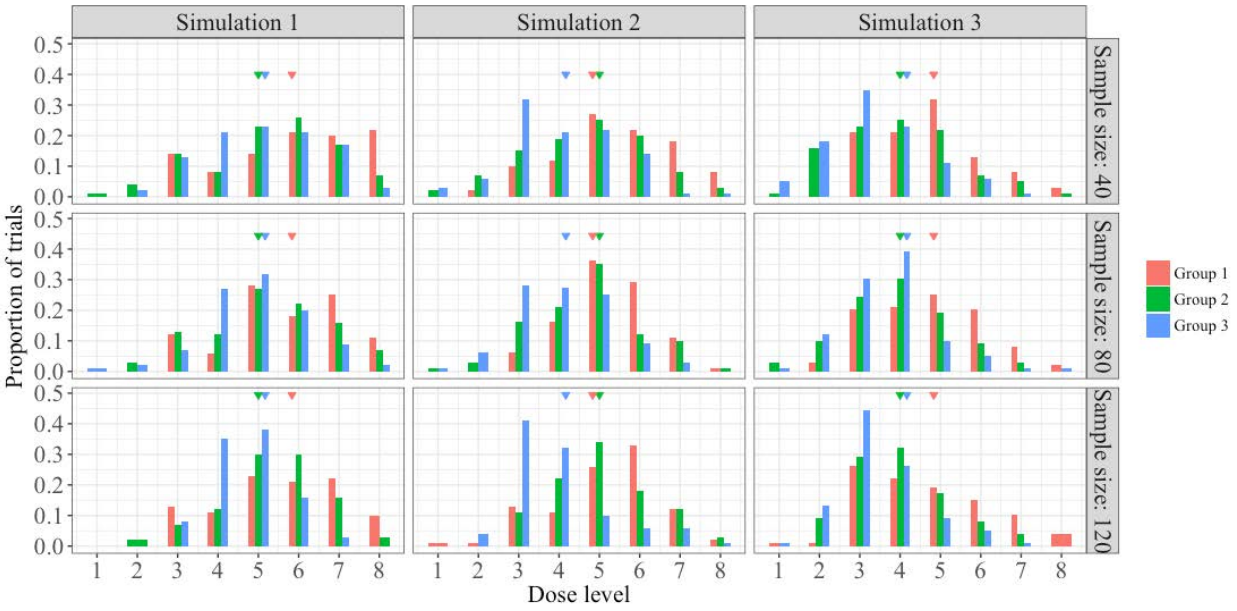

Figure 10. Proportion of trials recommending the true MTD for group trials (equal number of patients in each risk group)
Triangles represent the true MTD at which P(DLT) is around 0.2. Largest dose with P(DLT) ≤ target P(DLT)

4.1.1.6 In-trial dose distribution

Dose allocation of patients in a trial is one of the operating characteristics. As we stated previously, a major advantage of CRM is that more patients in the trial are allocated to dose levels close to the MTD. The in-trial dose allocation of patients in group trials is shown in Figure 11. The proportion in the graph represents the proportion of patients in a group of a trial allocated to a specific dose level. The true MTD in our study is defined as the dose level at which the true P(DLT) is closest to target P(DLT) and is represented by a triangle in the graphs.

As sample size increases, more patients are assigned to the true MTD in 3 groups. As shown in the graph, many patients in group 3 simulation 3 are assigned to a lower dose level than the true MTD. The reason is the wide difference between dose 4 and 5, and our method chose the safer dose level for the patients. According to our pre-specified true probabilities of toxicity, if the true MTD is defined as the dose level at which the P(DLT) is closest to but less than the target P(DLT), then dose level 3 will be the true MTD for the group 3 in simulation 3.

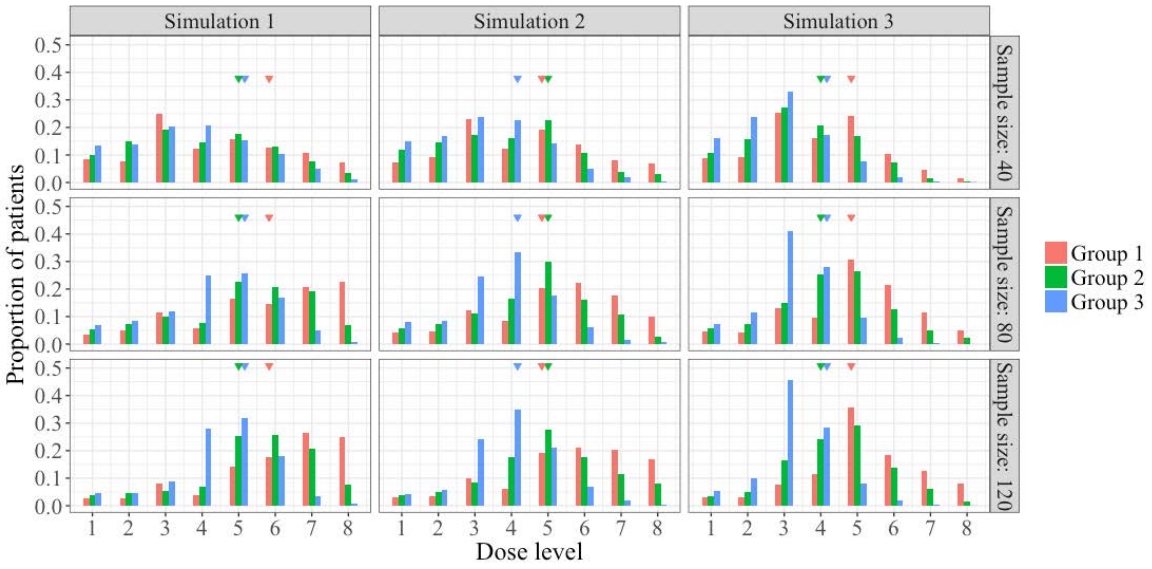

Figure 11. In-trial dose allocation of patients in group trials (equal number of patients in each risk group)
Triangles represent the true MTD at which P(DLT) is around 0.2.

4.1.1.7 In-trial DLT distribution

The amounts of DLTs per trial are summarized in Figure 12. The proportions of DLTs per trial are summarized in Figure 13. There are more DLTs in trials as sample size increases as Figure 12 shows. In Figure 13, the mean DLT proportion is around or below 0.2.

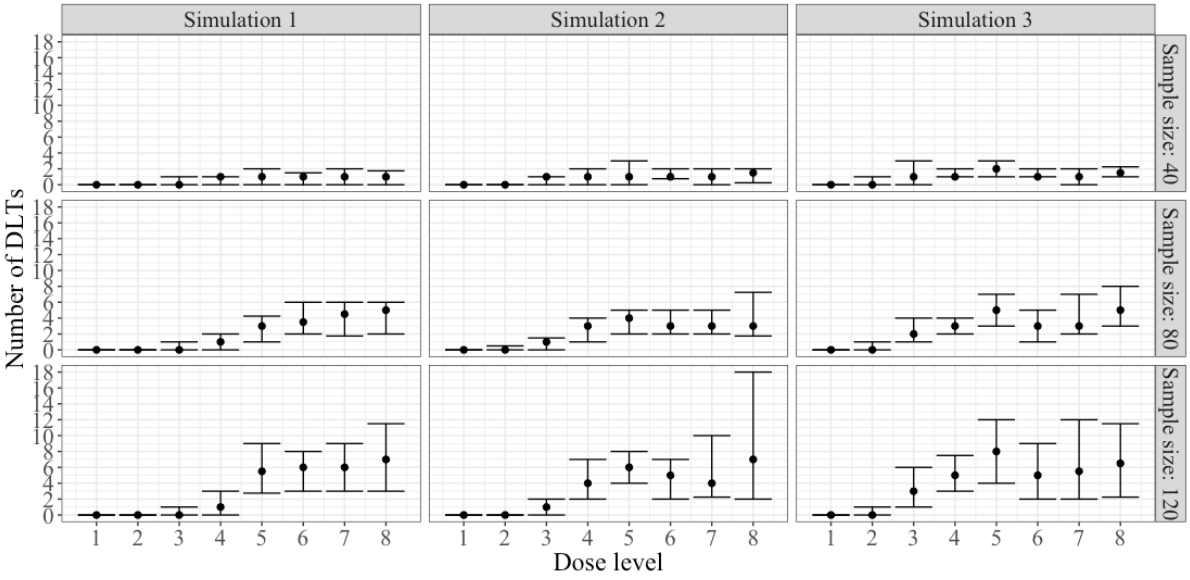

Figure 12. DLTs at all dose levels in group trials (equal number of patients in each risk group)
The error bars represent 25th and 75th percentiles. The dots represent the median numbers of DLTs.

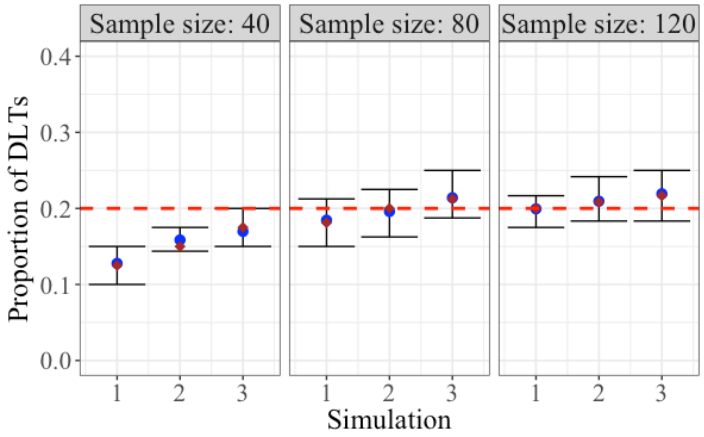

Figure 13. DLT proportion per trial in group trials (equal number of patients in each risk group)
The error bars represent 25th and 75th quantiles. Blue dots represent the mean proportions and red hearts represent the median proportions.

4.1.2 Parallel trials

4.1.2.1 Estimation of α of group 1

The distribution of estimated α of group 1 is displayed in Figure 14. The black square represents the true α in each simulation. As sample size increases, the range of α estimates gets narrower, and the median and mean sample α get closer to, but still higher than the true α.

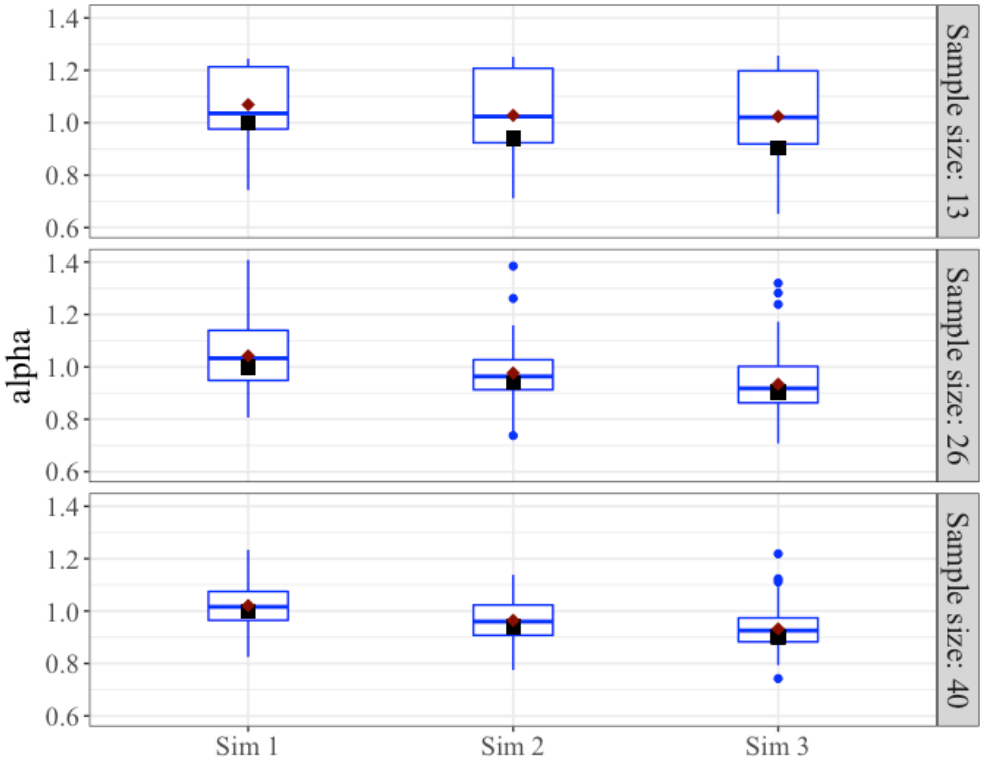

Figure 14. Distribution of estimated α of group 1 in parallel trials

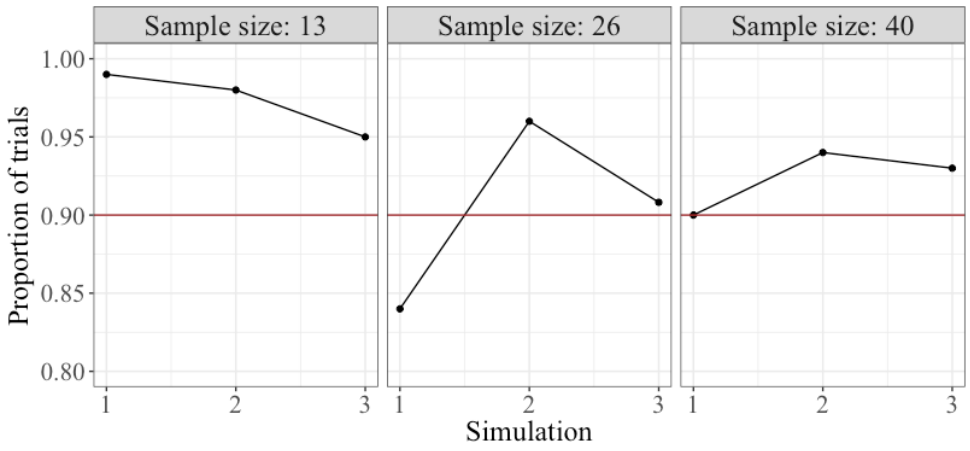

Figure 15. Coverage of 90% credible interval of true α of group 1 in parallel trials

The coverage of 90% credible interval of true α is displayed in Figure 15. The coverage of α for is getting closer to 90% in all simulations as sample size increases, indicating a good quality of interval estimation.

4.1.2.2 Estimation of α of group 2

Figure 16 shows the distribution of estimated α in group 2. As sample size increases, the range of α estimates gets narrower, and the median and mean sample α get closer to the true α. Simulation 3 has the largest range of sample α. In Figure 17, the coverage of 90% credible interval of true α is getting closer to 90% as sample goes from 13 to 26. The coverage is farther from 0.9 when sample size is 40 than 26.

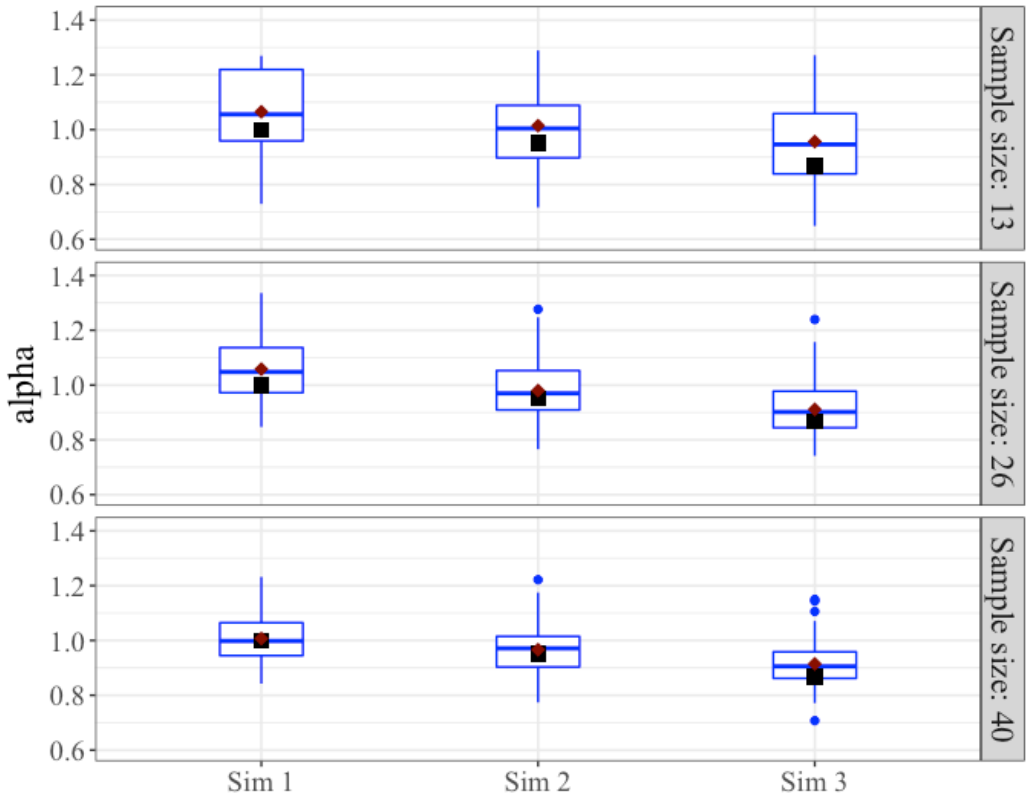

Figure 16. Distribution of estimated α of group 2 in parallel trials

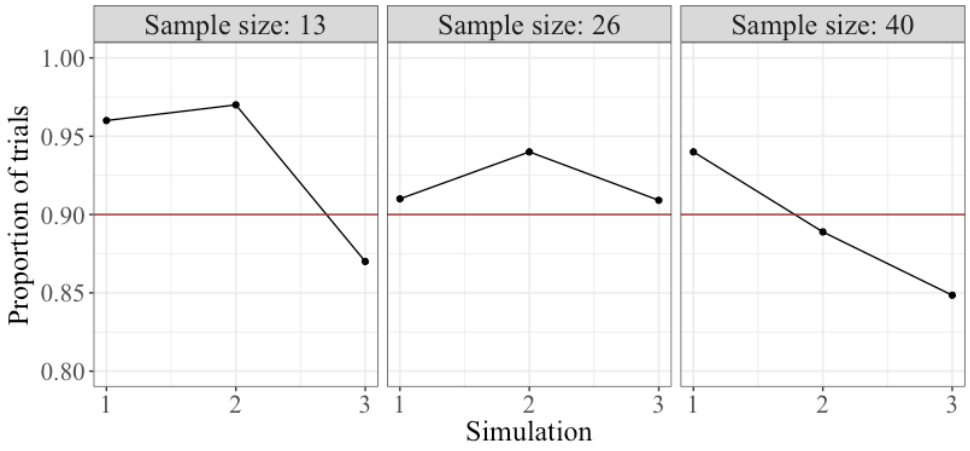

Figure 17. Coverage of 90% credible interval of true α of group 2 in parallel trials

4.1.2.3 Estimation of α of group 3

Figure 18 shows the distribution of estimated α in group 3. As sample size increases, the range of α estimates gets narrower, and the median and mean sample α get closer to the true α.

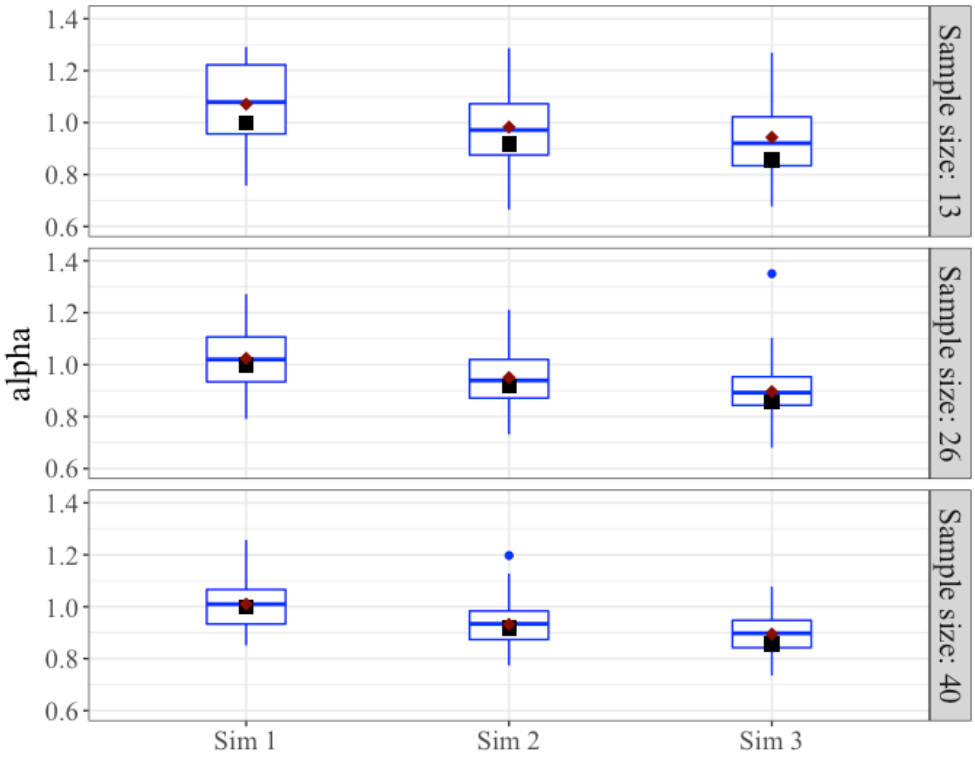

Figure 18. Distribution of estimated α of group 3 in parallel trials

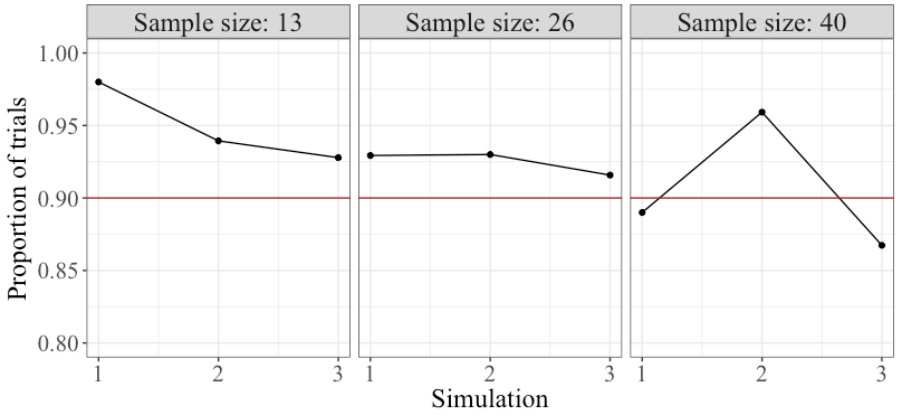

Figure 19. Coverage of 90% credible interval of true α of group 3 in parallel trials

According to Figure 19, sample size of 26 has the best coverage of 90% credible intervals among all three sample sizes.

4.1.2.4 Estimation of P(DLT)

The distribution of estimated P(DLT) of 3 groups in parallel trials is displayed in Figure 20. The black triangles are the true P(DLT) in each group at each dose level. The range of estimated P(DLT) at each dose level in each simulation gets narrower and the median P(DLT) is getting closer to the true P(DLT) as sample size increases. The true P(DLT) is close to the third quantile of estimated P(DLT) at dose 5 in simulation 3, indicating the estimated P(DLT) might not be very precise in this situation.

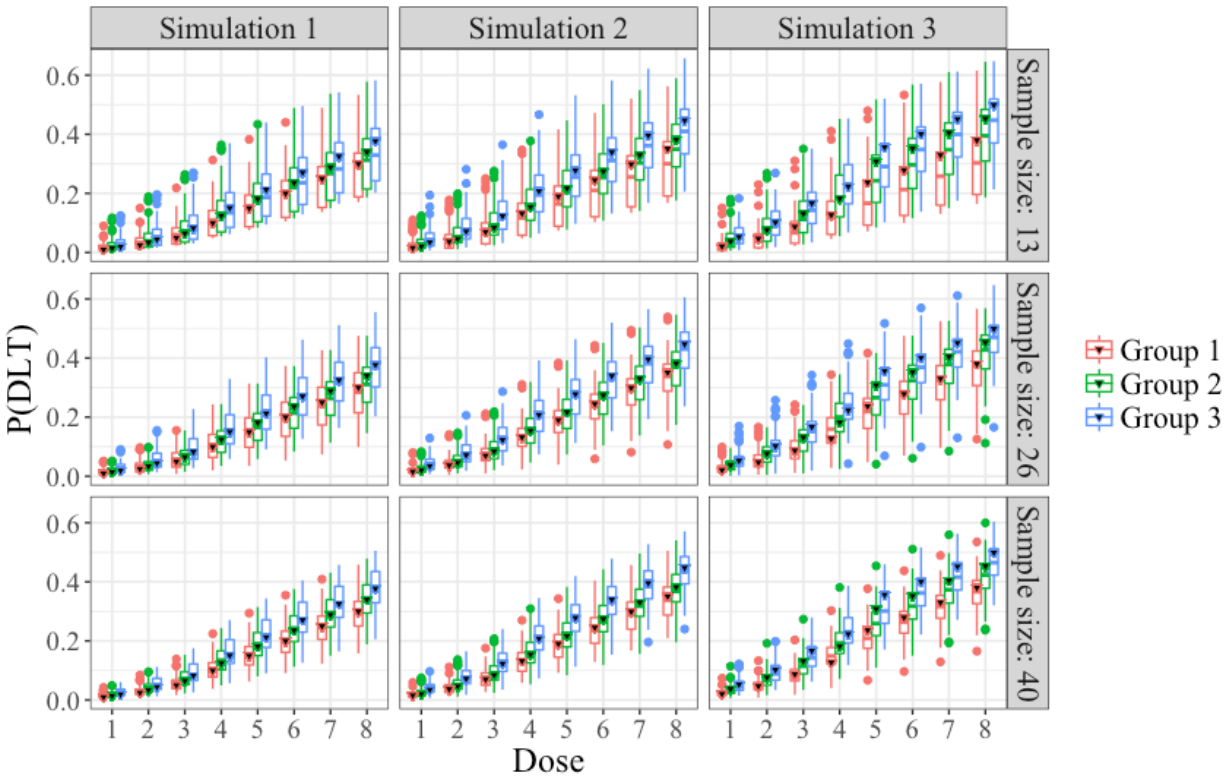

Figure 20. Distribution of estimated P(DLT) in parallel trials
The black triangles refer to the true P(DLT)s.

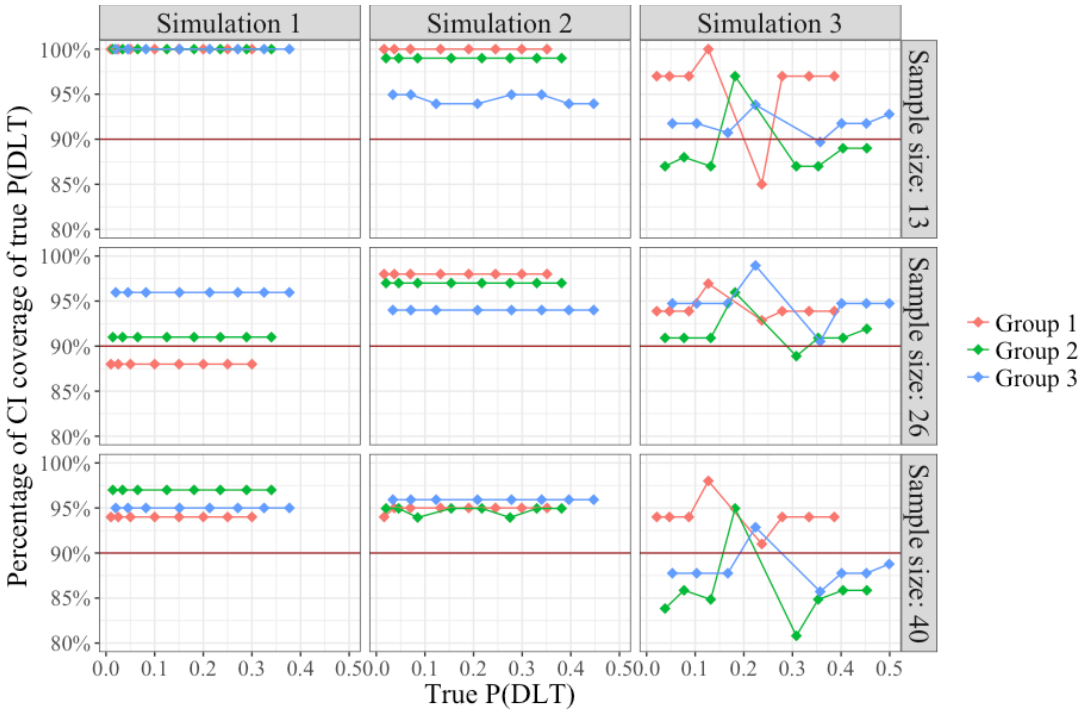

Figure 21. Coverage of 90% credible interval of true P(DLT) in parallel trials (equal number of patients in each risk group)

The coverages of true P(DLT) at most doses in all simulations for sample sizes 26 and 40 are around 85% to 95%. The credible intervals of P(DLT) for sample size 13 are too wide, indicating low precision. As the sample size increases the coverages are getting better, except that of dose 5 in simulation 3.

4.1.2.5 MTD recommendation

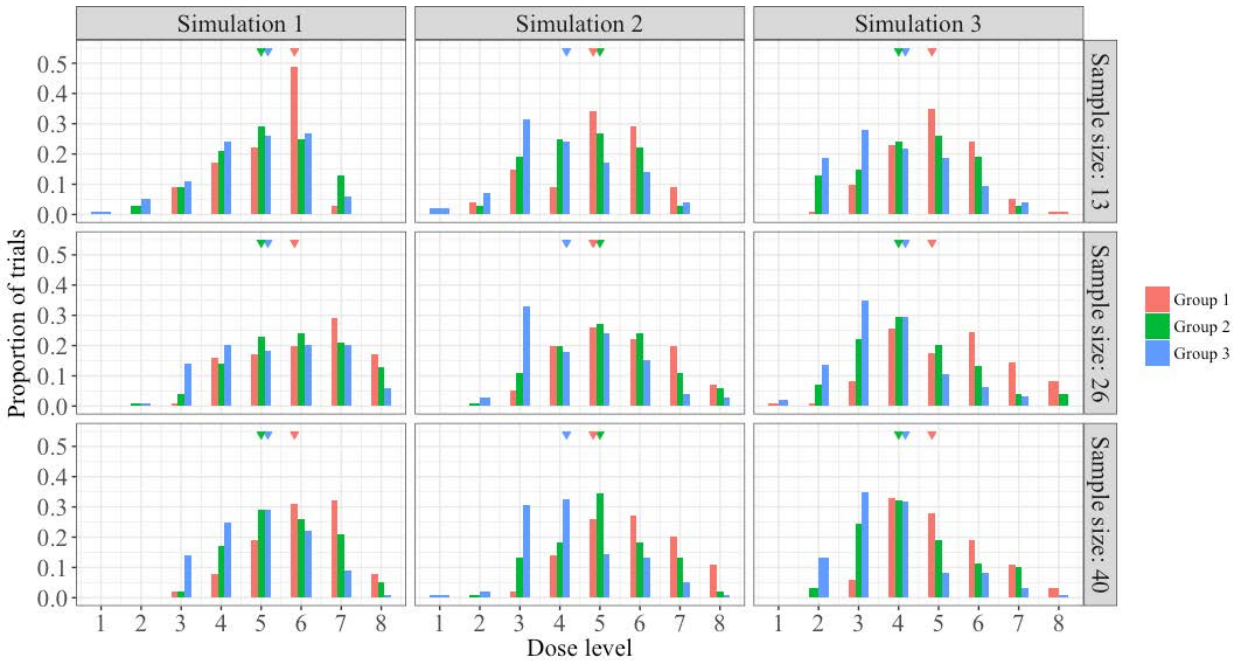

Figure 22. Proportion of trials recommending the true MTD for parallel trials (equal number of patients in each risk group)

As Figure 22 shows, most trials recommend the true MTD in simulation 1 with smallest sample size. As sample size increases, most trials for group 1 in all simulations tend to recommend a higher dose level, and most trials in group 3 tend to recommend a lower dose level in simulation 3.

4.1.2.6 In-trial dose distribution

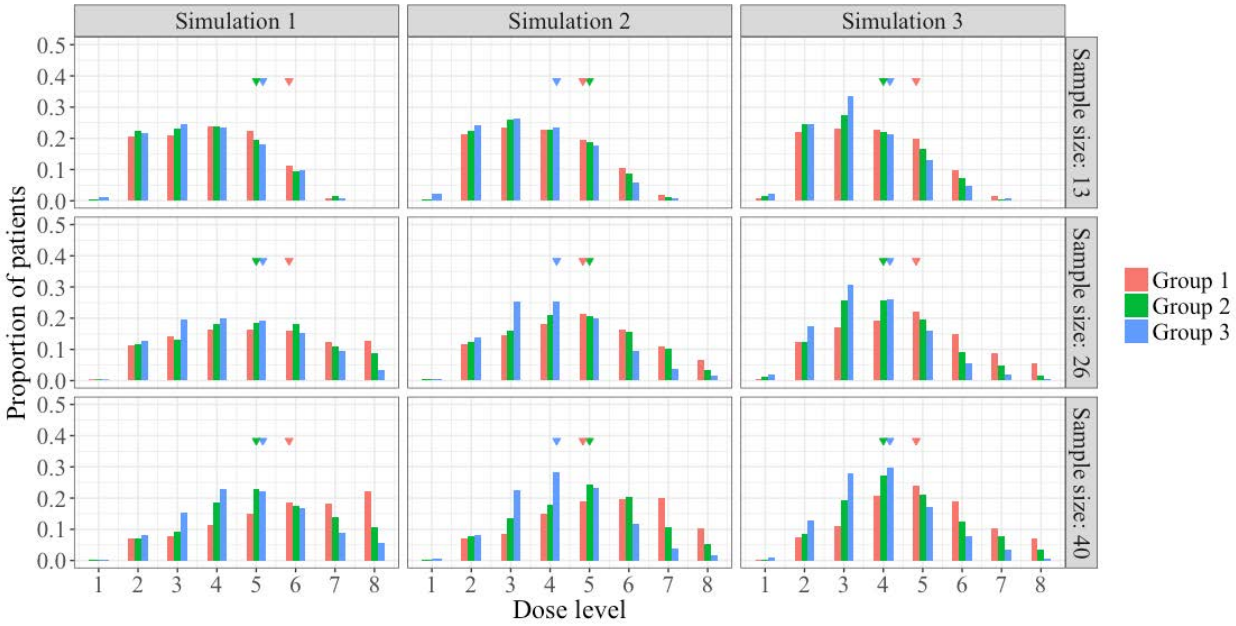

Figure 23. In-trial dose allocation of patients in parallel trials (equal number of patients in each risk group)

When sample size is small, most patients are assigned to lower dose levels than true MTD. As sample size increases, more patients are given the true MTD doses. For the largest sample size, a lot of patients in group 1 are assigned to dose level 8.

4.1.2.7 DLT distribution

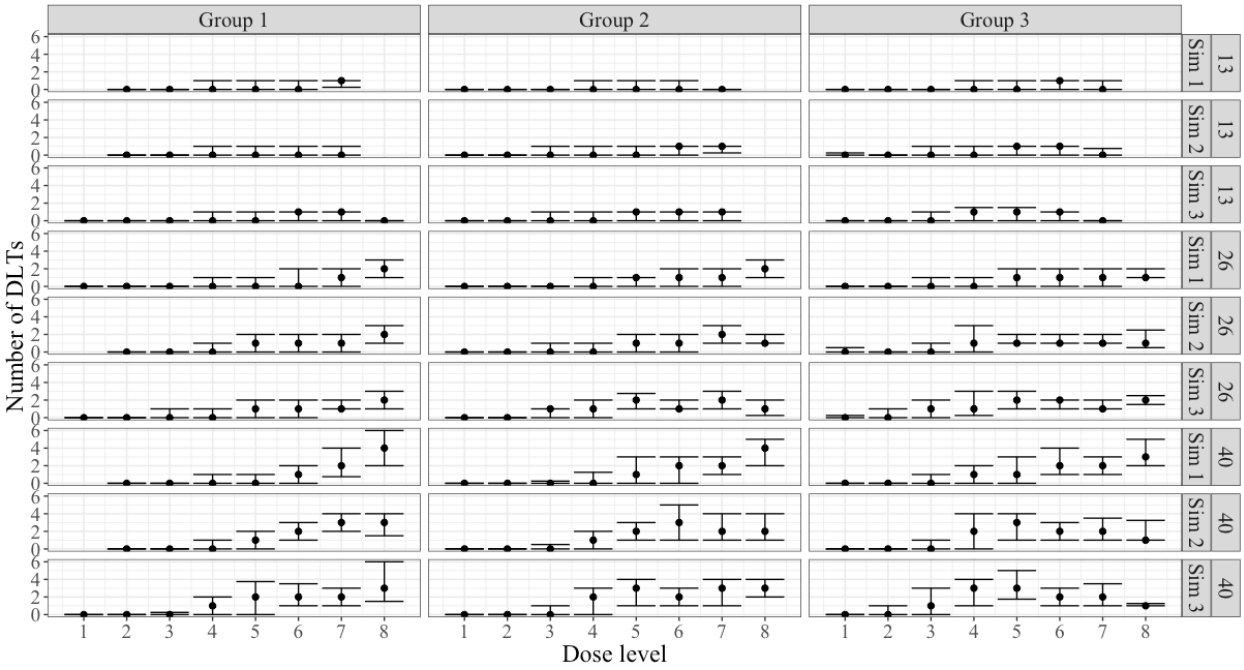

Figure 24. DLTs at all dose levels in parallel trials (equal number of patients in each risk group)

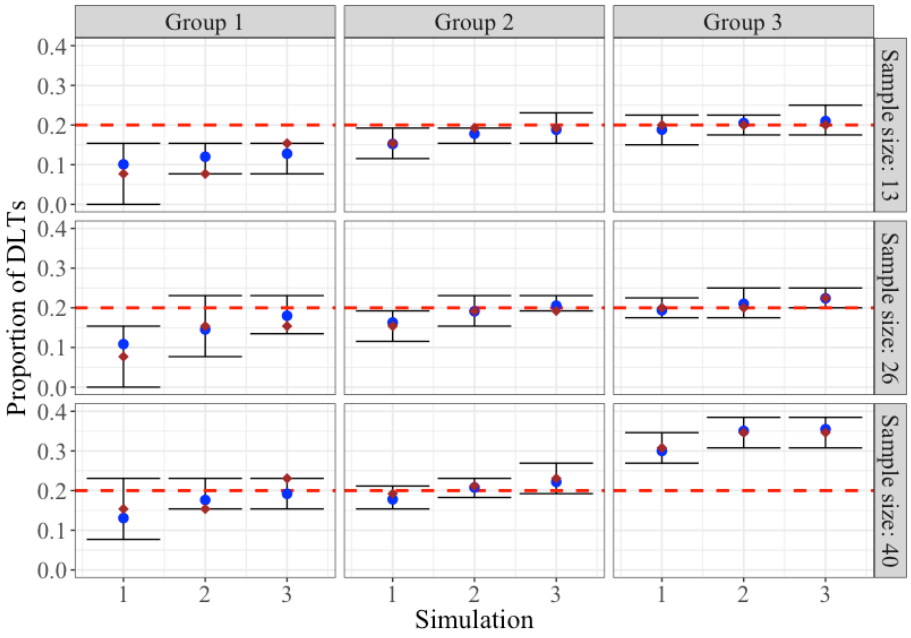

Figure 25. DLT proportion in a trial for parallel trials (equal number of patients in each risk group)

The DLT proportion in most simulations are around and below 0.2. The DLT proportion of group 3 in simulation 3 with sample size 40 is higher than the average DLT proportion in 3-group trial, probably because more patients are assigned to dose level 4 and above.

4.2 SCENARIO 2: MORE PATIENTS ARE IN HIGHER RISK GROUPS

Simulation studies of scenario 2 were performed to explore the effect of more patients with higher toxicity risk. Additionally, as we use rules to include ordering information, the higher risk groups should provide more information than the lower risk group. Hence, it is expected to observe more rapid dose-escalation in the lower risk groups.

The sample size for the parallel trials are 8, 16 and 24 for group 1, 14, 28 and 42 for group 2, and 18, 36, 54 for group 3. In the following graphic results, sample 1 refers to the smallest sample size for each group, sample 2 the medium sample size and sample 3 the largest sample size.

4.2.1 Group trials

4.2.1.1 Estimation of α

The distribution of the estimated α is displayed in Figure 26. The black square represents the true α in each simulation. As sample size increases, the range of α estimates gets narrower, and the median and mean sample α get closer to, but still higher than the true α. It is noted that the difference between mean estimated α and true α is the smallest in simulation 3, where there is an

abrupt jump in P(DLT) between dose 4 and 5. The ranges of α estimates are similar in all simulation for one sample size.

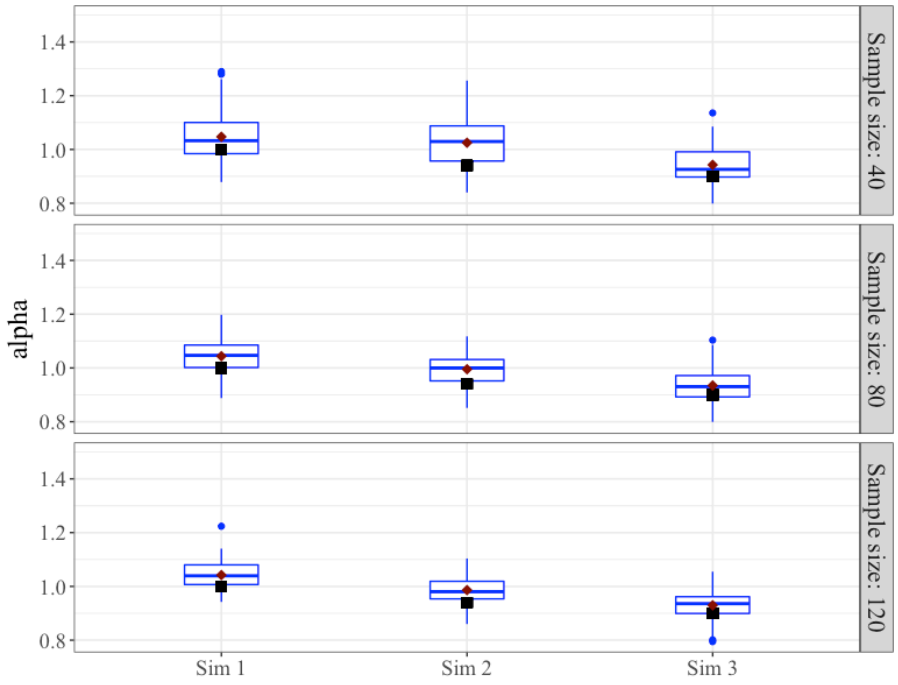

Figure 26. Distribution of estimated α in group trials

The coverage of 90% credible interval of true α is displayed in Figure 27. The coverage of α is around 90% for all simulations.

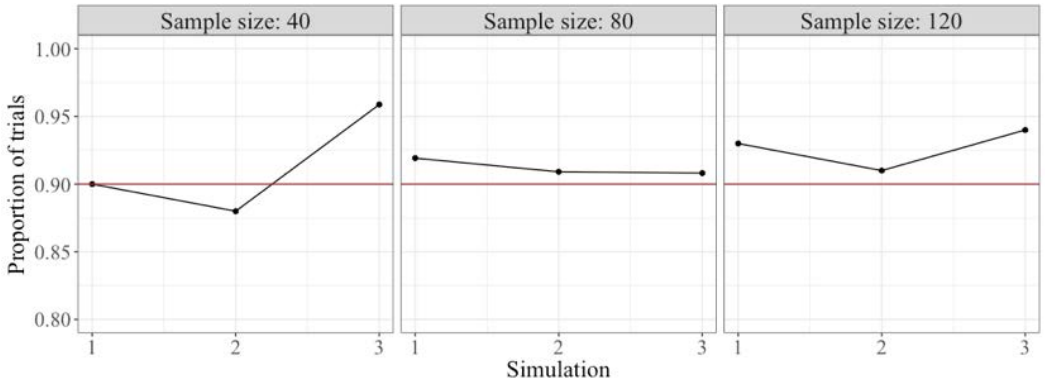

Figure 27. Coverage of 90% credible interval of true α in group trials

4.2.1.2 Estimation of δ_1

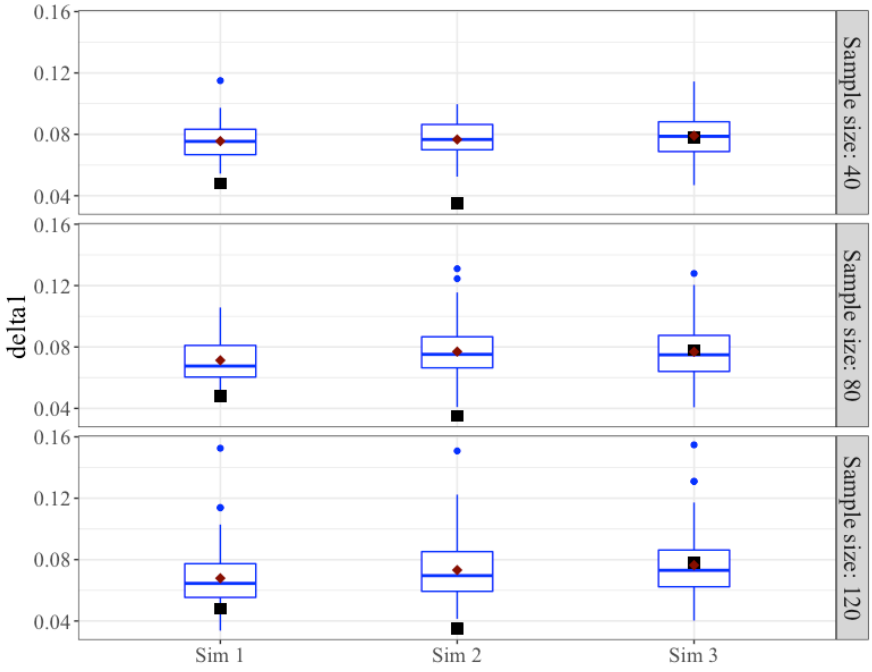

Figure 28. Distribution of estimated δ_1 in group trials

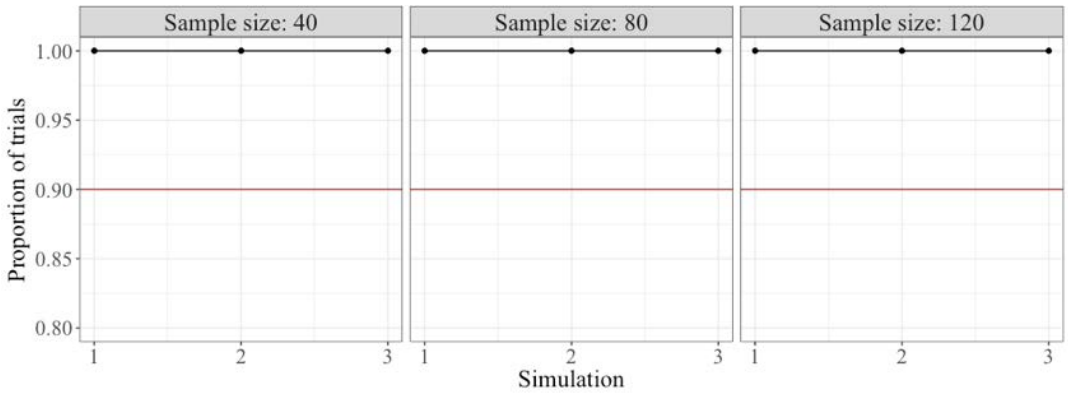

Figure 29. Coverage of 90% credible interval of true δ_1 in group trials

The mean sample δ_1 is close to true δ_1 in simulation 3 only. In the first and second simulation, the true value is much lower than the estimated δ_1. As the sample size increases, the ranges of sample

δ_1 do not get narrower.The coverage of 90% credible interval of true δ_1 is 100% due to wide credible intervals.

4.2.1.3 Estimation of δ_2

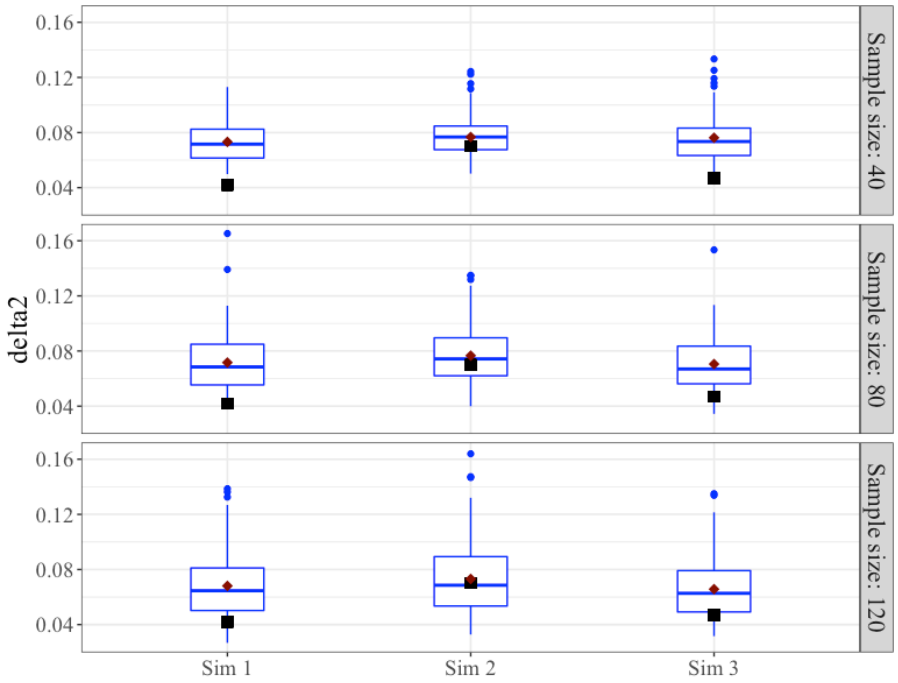

Figure 30. Distribution of estimated δ_2 in group trials

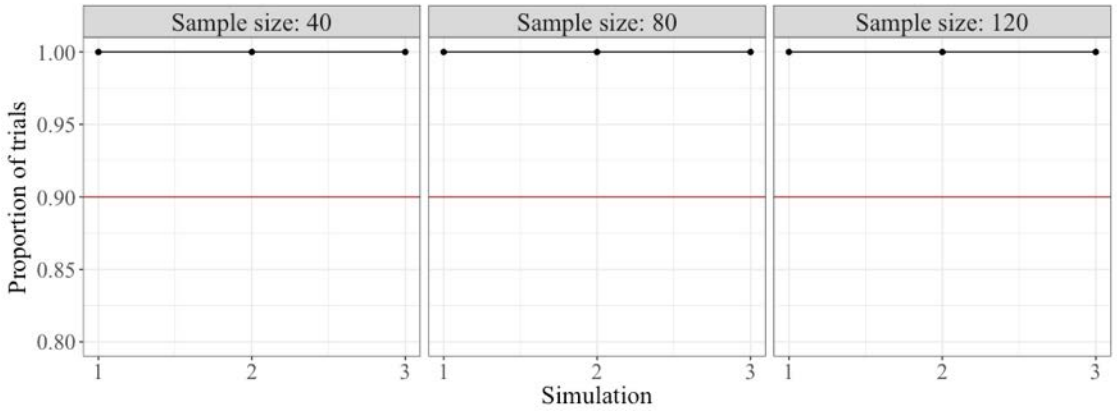

Figure 31. Coverage of 90% credible interval of true δ_2 in group trials

The mean sample δ_2 is close to true δ_2 in simulation 2 only. In the first and third simulation, the true value is much lower than the mean δ_2. As the sample size increases, the ranges of sample δ_2 do not get narrower. The coverage of 90% credible interval of true δ_2 is 100% due to wide credible intervals.

4.2.1.4 Estimation of P(DLT)

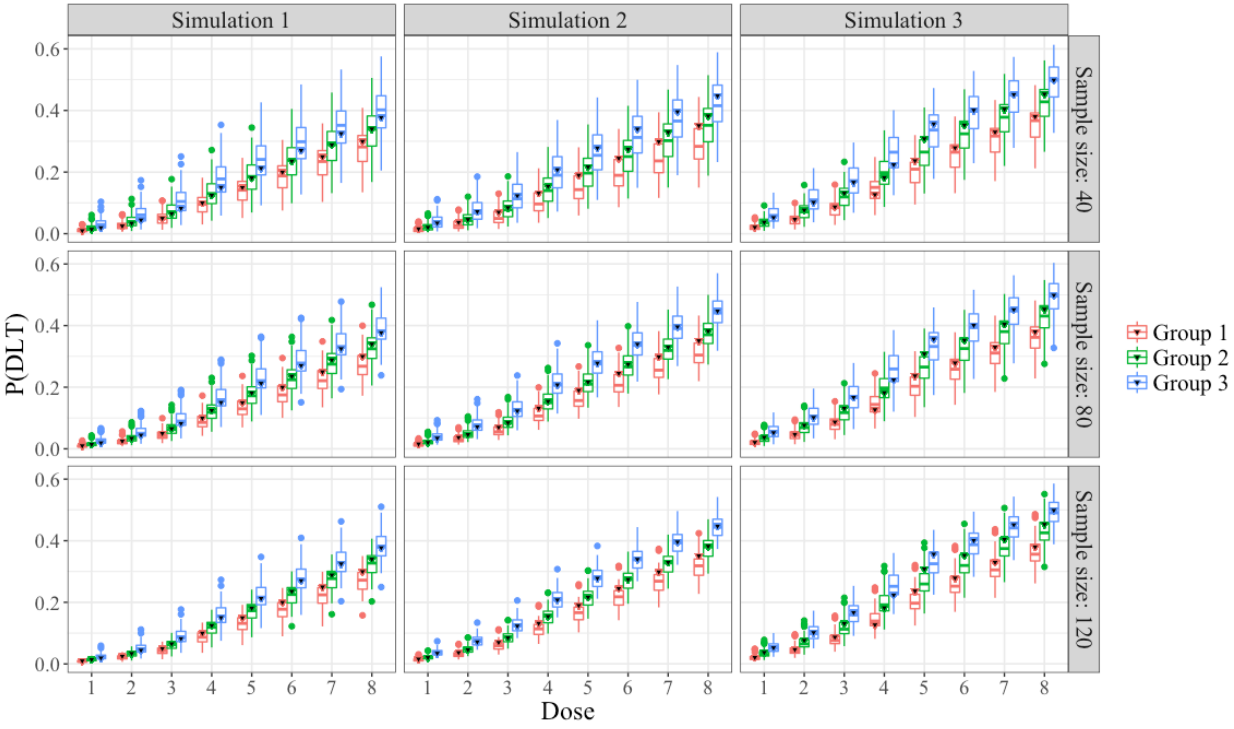

Figure 32. Distribution of estimated P(DLT) for group trials
The black triangle refers to the true P(DLT).

The distribution of estimated P(DLT) is displayed in Figure 32. The black triangles are the true P(DLT). The range of estimated P(DLT) at each dose level in each simulation gets narrower and the median P(DLT) is getting close to the true P(DLT) as sample size increases. The median sample P(DLT) is much lower than the true for group 1 in simulation 2, compared to the rest

situations. The true P(DLT) is close to the third quantile of estimated P(DLT) at dose 5 in simulation 3, indicating the estimated P(DLT) might not be very precise in this case.

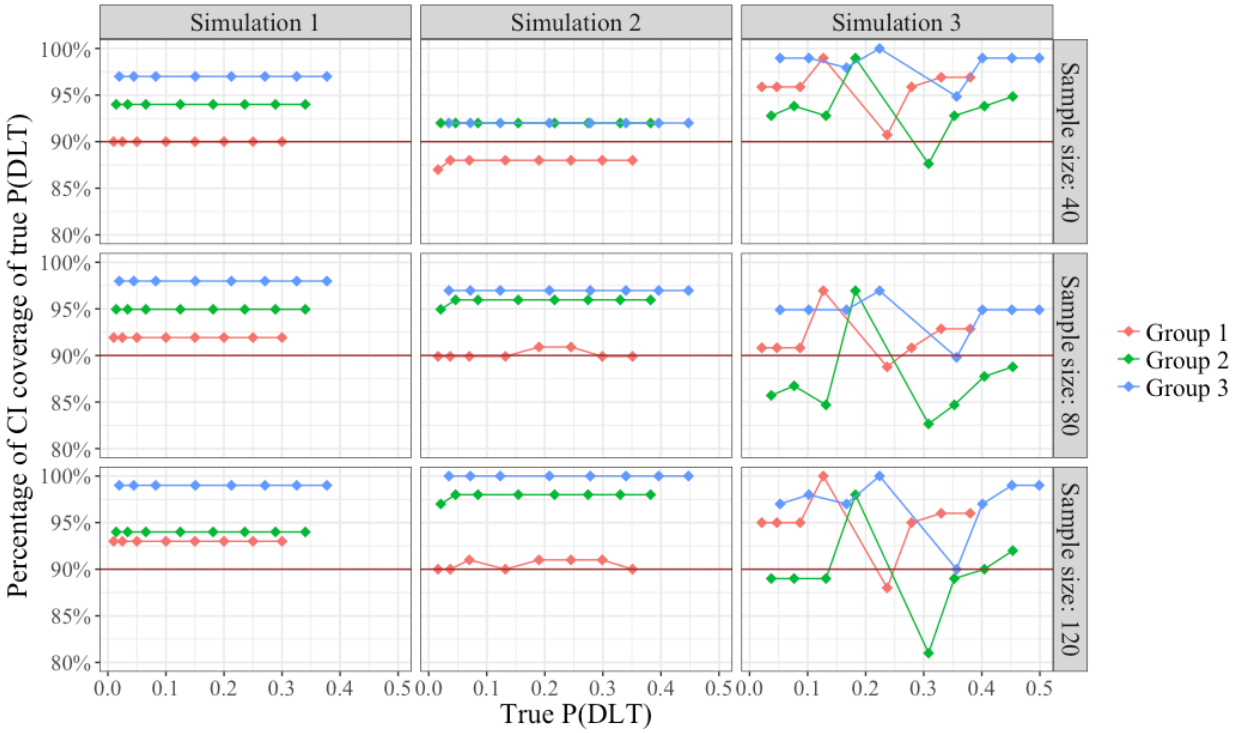

Figure 33. Coverage of 90% credible interval of true P(DLT) in group trials

The coverages of true P(DLT) in each simulation are not getting better as sample size increases. Most coverages are around 85% to 97%.

4.2.1.5 MTD recommendation

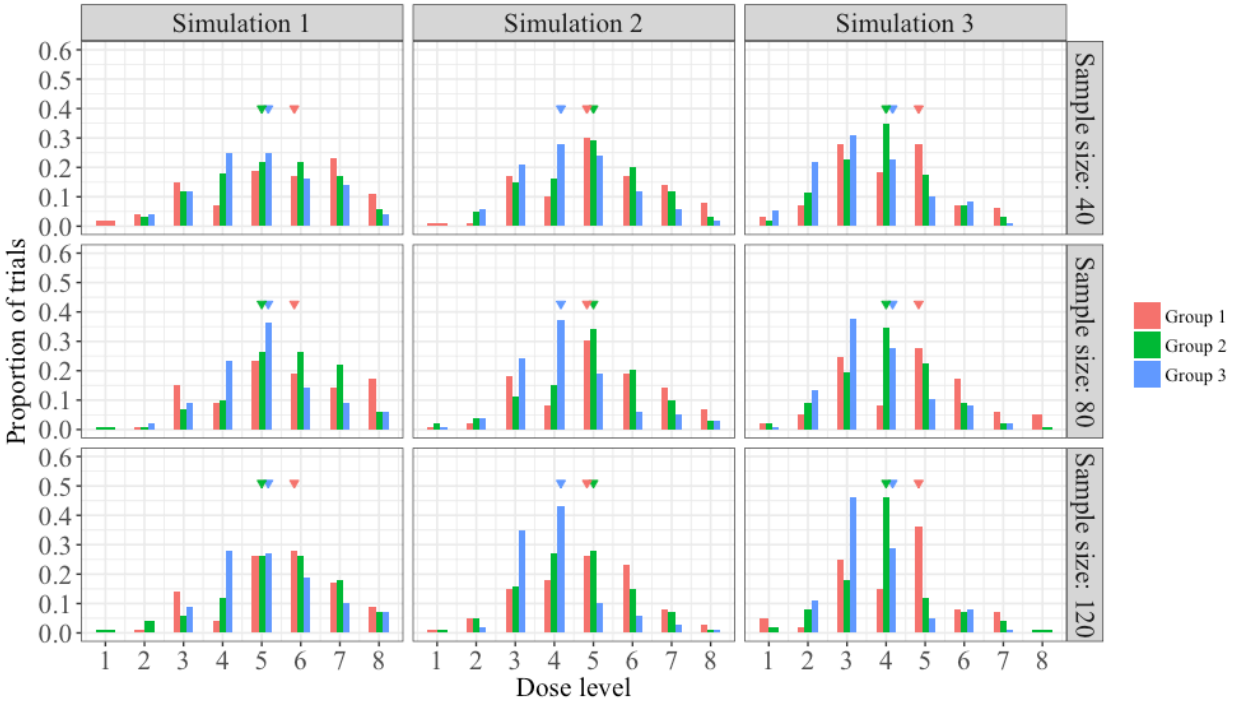

Figure 34. Proportion of trials recommending the true MTD in group trials (more patients in higher risk groups)

As Figure 34 shows, most trials recommend the true MTD except group 3 in simulation 3 and group 1 in simulation 1 with the smallest sample size. In simulation 3, where there is an abrupt rise of P(DLT) between dose 4 and 5, dose 4 is the largest safe dose with P(DLT) < 0.2. It is acceptable to have a lower level recommended in this case from a safety perspective.

4.2.1.6 In-trial dose distribution

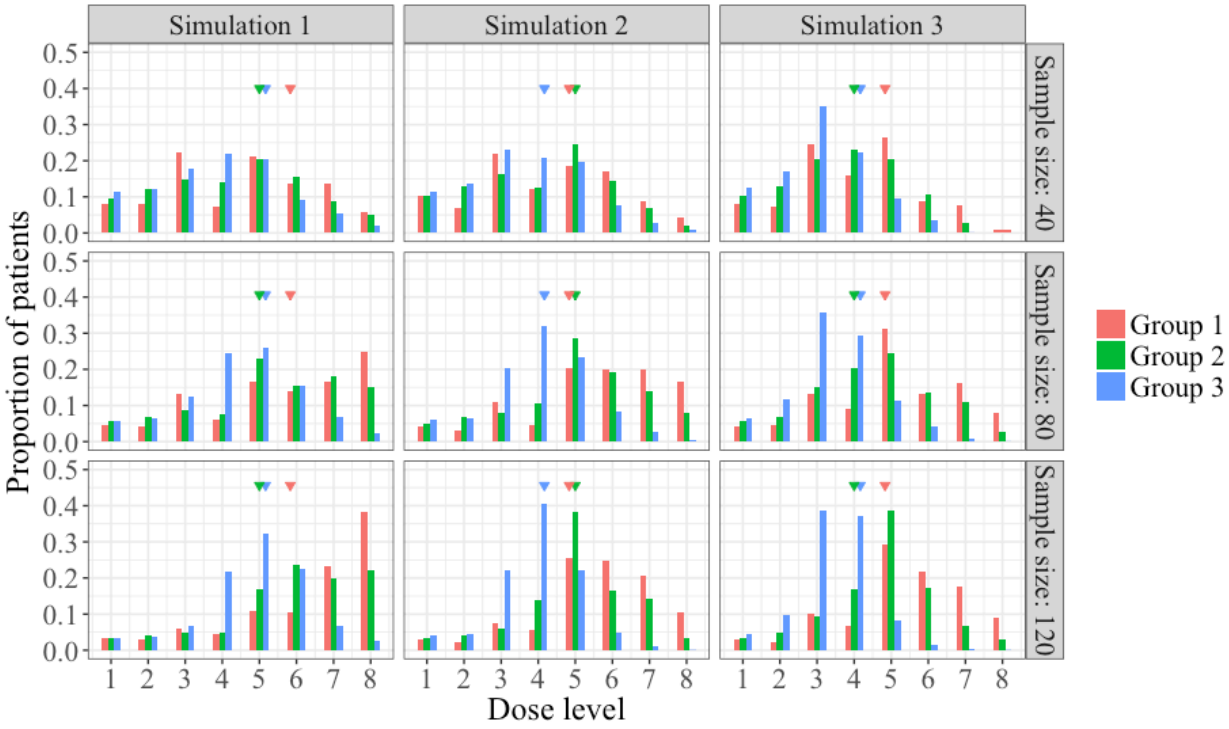

Figure 35. In-trial dose allocation of patients in 3-group trials (more patients in higher risk groups)

The in-trial dose allocation of patients is shown in Figure 35. Groups 1 and 2 see a more rapid dose-escalation, since more full observations in the higher risk group provide dose-toxicity information to lower risk groups. For simulation 1, there are too many patients are assigned to higher dose level such as level 8 in group 1.

4.2.1.7 DLT distribution

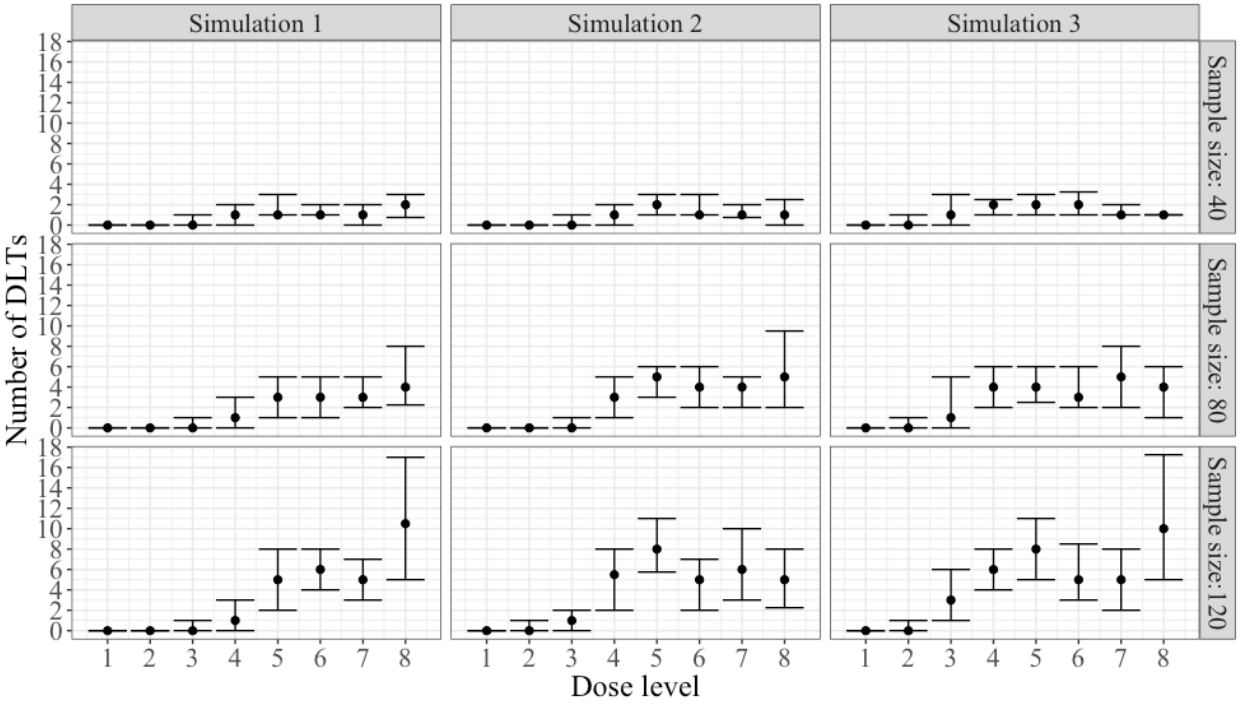

Figure 36. DLTs at all dose levels in group trials (more patients in higher risk groups)

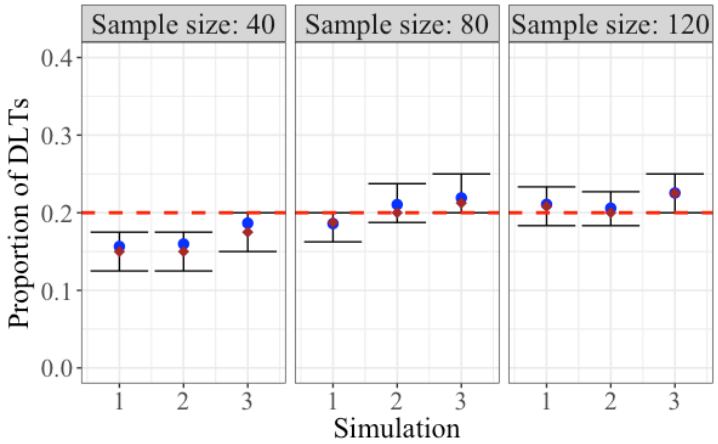

Figure 37. DLT proportion in a trial for group trials (more patients in higher risk groups)

The number of DLTs in a trial are summarized in Figure 36. The proportion of DLTs in a trial are summarized in Figure 37. The mean proportion of DLTs is around or below 0.2.

4.2.2 Parallel trials

4.2.2.1 Estimation of α of group 1

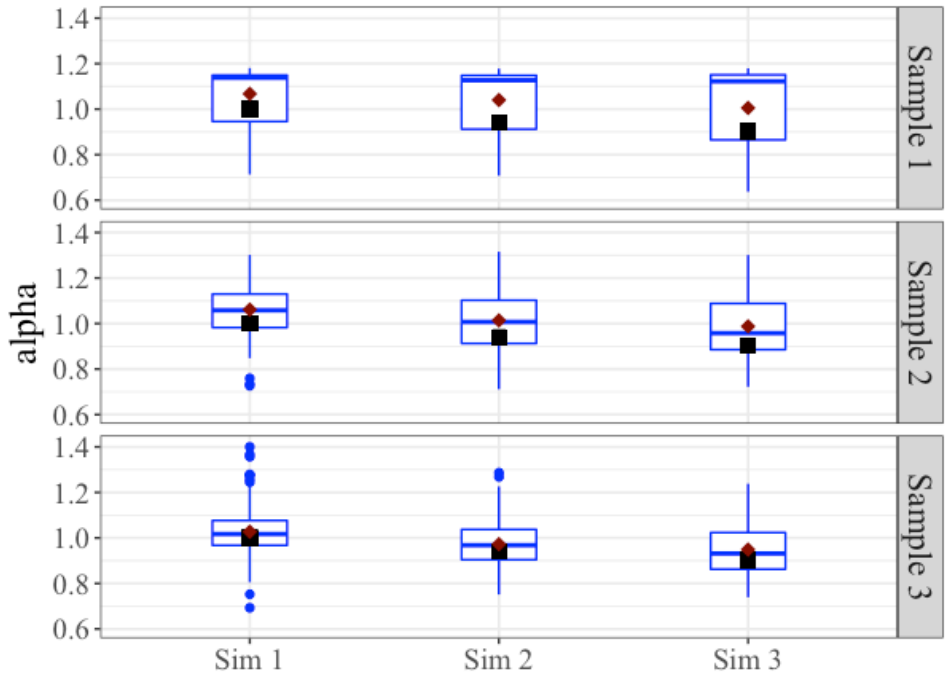

Figure 38. Distribution of estimated α of group 1 in parallel trials

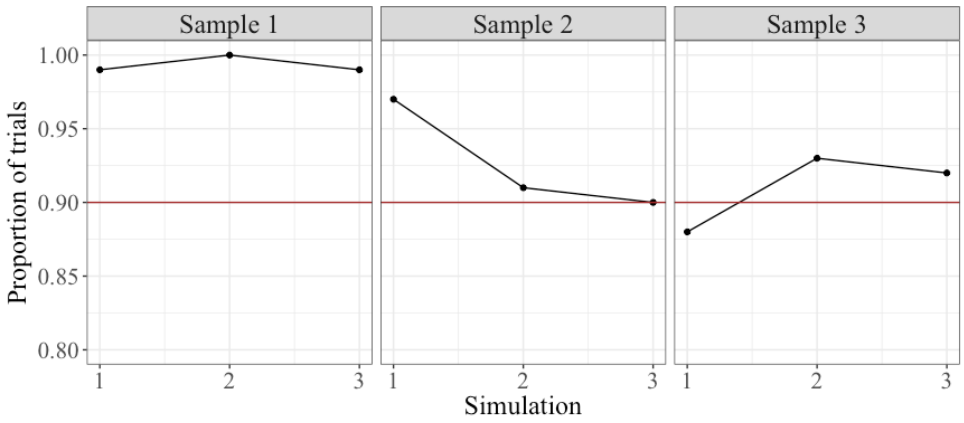

Figure 39. Coverage of 90% credible interval of true α of group 1 in parallel trials

The distribution of estimated α of group 1 is displayed in Figure 38. The black square represents the true α in each simulation. As sample size increases, the interquartile range of α estimates gets narrower, and the median and mean sample α get closer to the true α. The coverage of 90% credible interval of true α is displayed in Figure 39. The coverage of α is around 90% for larger sample sizes. The coverage of α of the first sample size is large due to wide credible intervals and small sample size.

4.2.2.2 Estimation of α of group 2

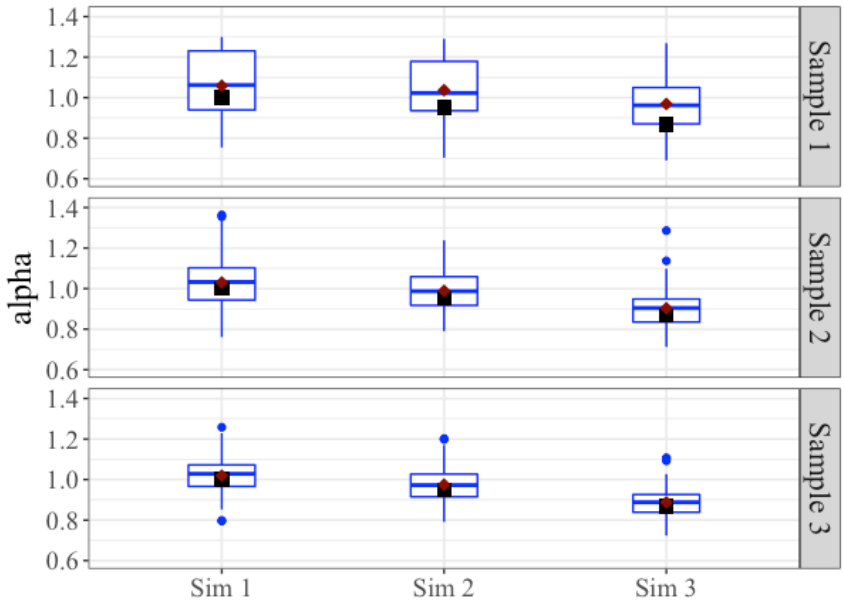

Figure 40. Distribution of estimated α of group 2 in parallel trials

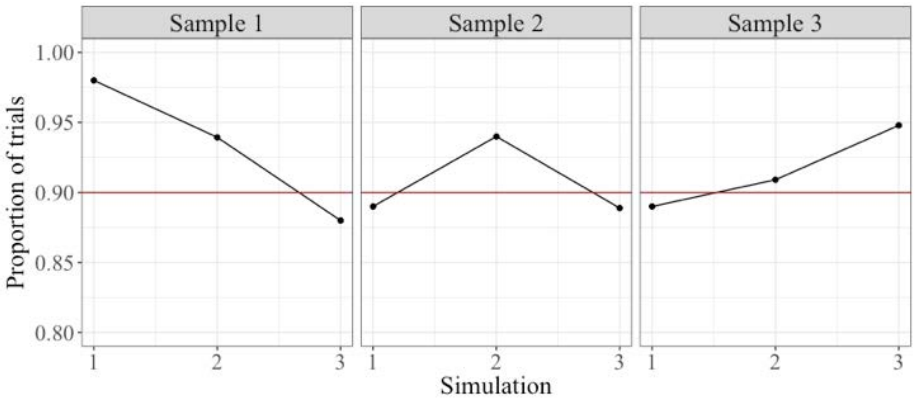

Figure 41. Coverage of 90% credible interval of true α of group 2 in parallel trials

The distribution of estimated α of group 2 is displayed in Figure 40. As sample size increases, the range of α estimates gets narrower, and the median and mean sample α get closer to the true α. The coverage of α is around 90%. The coverage of α is better in group 2 due to larger sample size.

4.2.2.3 Estimation of α of group 3

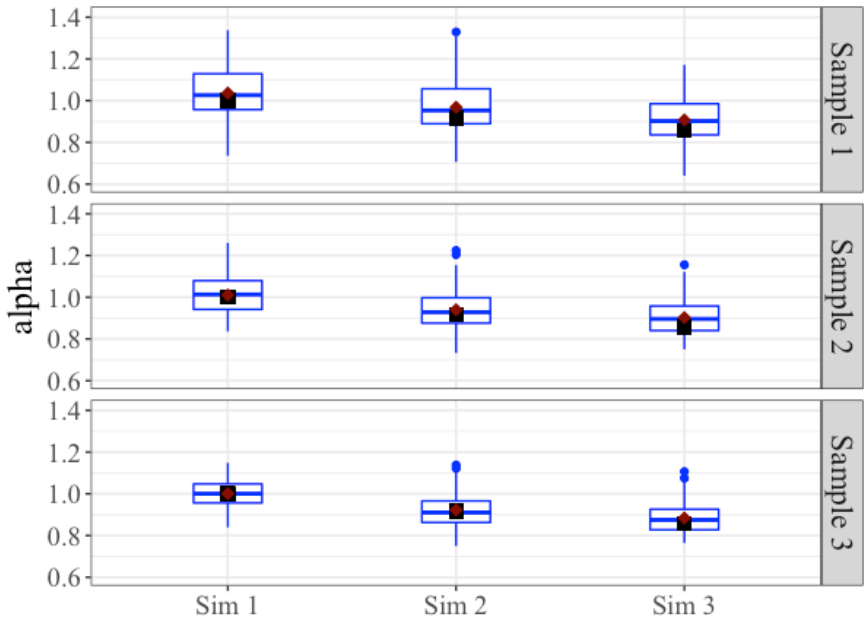

Figure 42. Distribution of estimated α of group 3 in parallel trials

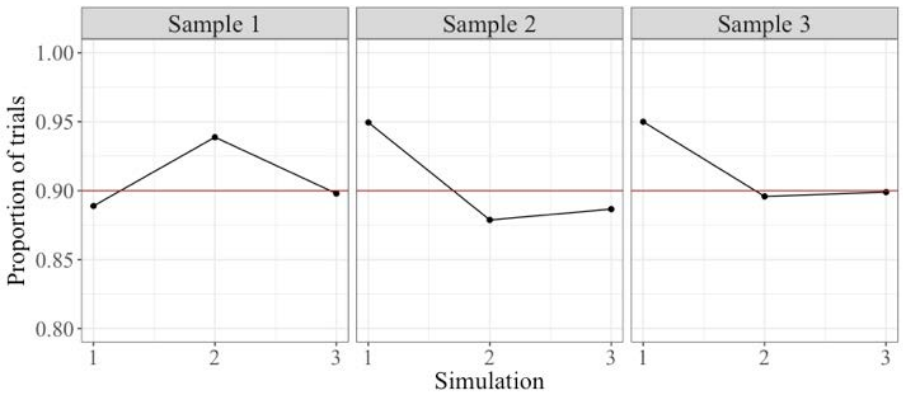

Figure 43. Coverage of 90% credible interval of true α of group 3 in parallel trials

The distribution of estimated α of group 3 is displayed in Figure 42. As sample size increases, the range of α estimates gets narrower. The sample mean α and the true α overlap for the largest sample. The coverage of α is around 90%. The coverage of α is better in group 2 and 3 due to larger sample size.

4.2.2.4 Estimation of P(DLT)

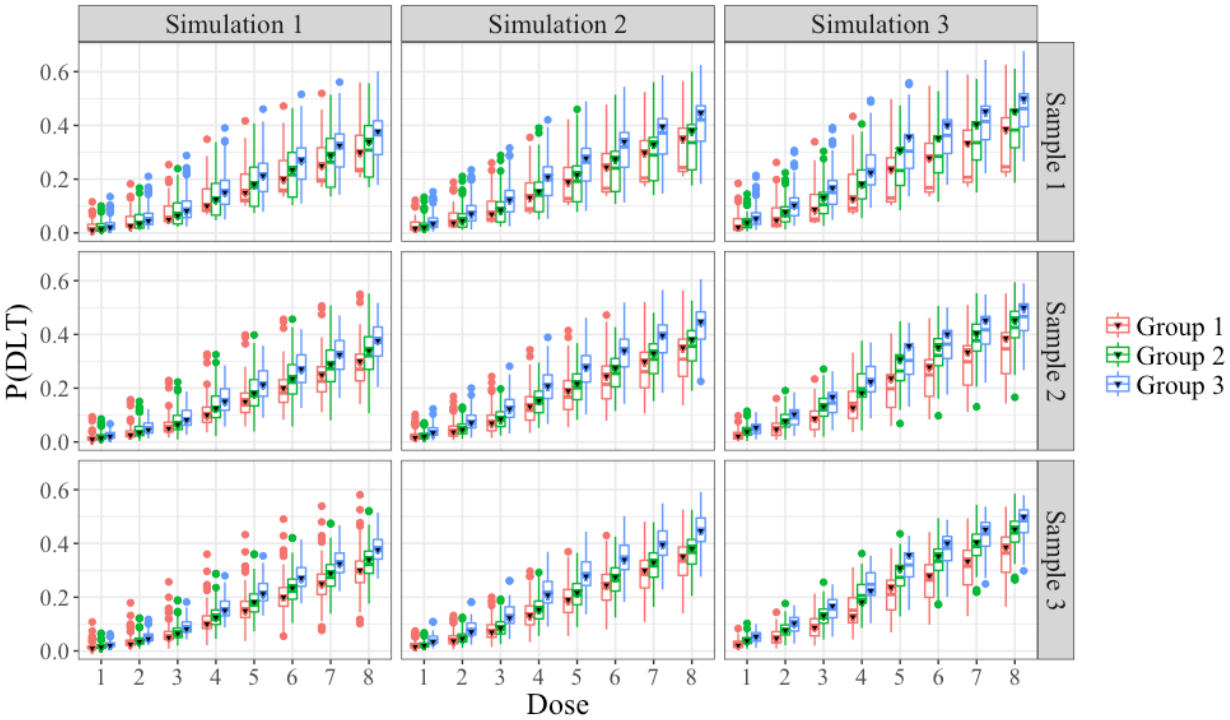

Figure 44. Distribution of estimated P(DLT) in parallel trials

The distribution of estimated P(DLT) is displayed in Figure 44. The black triangles are the true P(DLT). The range of estimated P(DLT) at each dose level in each simulation gets narrower and the median P(DLT) is getting closer to the true P(DLT) as sample size increases. The true P(DLT) is close to the third quantile of estimated P(DLT) at dose 5 in simulation 3, indicating the estimated P(DLT) might not be very precise in this case.

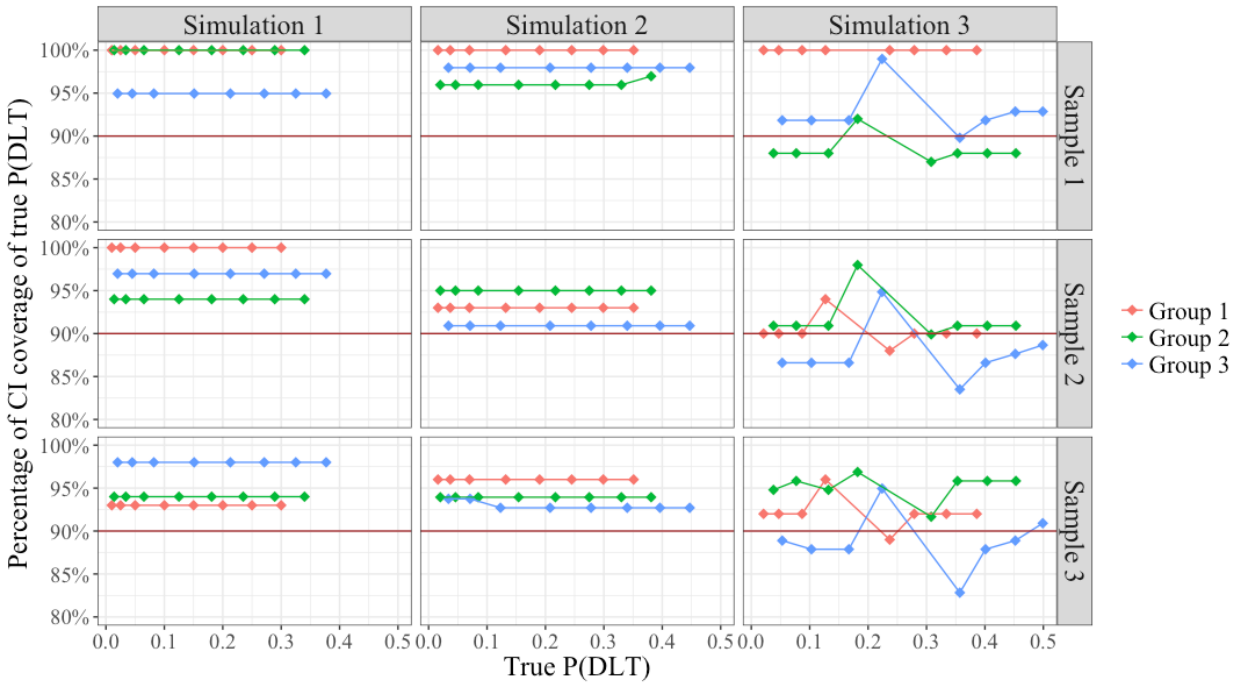

Figure 45. Coverage of 90% credible interval of true P(DLT) in parallel trials

The coverages of true P(DLT) at are getting better as sample size increases. Most coverages are around 85% to 95% for medium and large sample sizes. The coverages of sample 1 so high because smaller sample size compared to the rest samples.

4.2.2.5 MTD recommendation

The proportion of trials that the true MTD is recommended is summarized in Figure 46. In simulation 1 with the least sample size, most trials in group 1 recommend a lower dose level than

the true MTD while most trials in group 2 recommend a higher dose level than the true MTD. As sample size increases, most trials recommend the true MTD in group 1 and 3 while most trials in group 2 recommend a higher dose level than the true MTD. In simulation 2, most trials in group 1 and 2 with the least and largest sample size recommend the true MTD while most trials in group 3 recommend a level next to the true MTD as the MTD. In simulation 3, most trials in group 2 and 3 recommend the true MTD while most trials in group 1tend to recommend higher dose level than the true MTD.

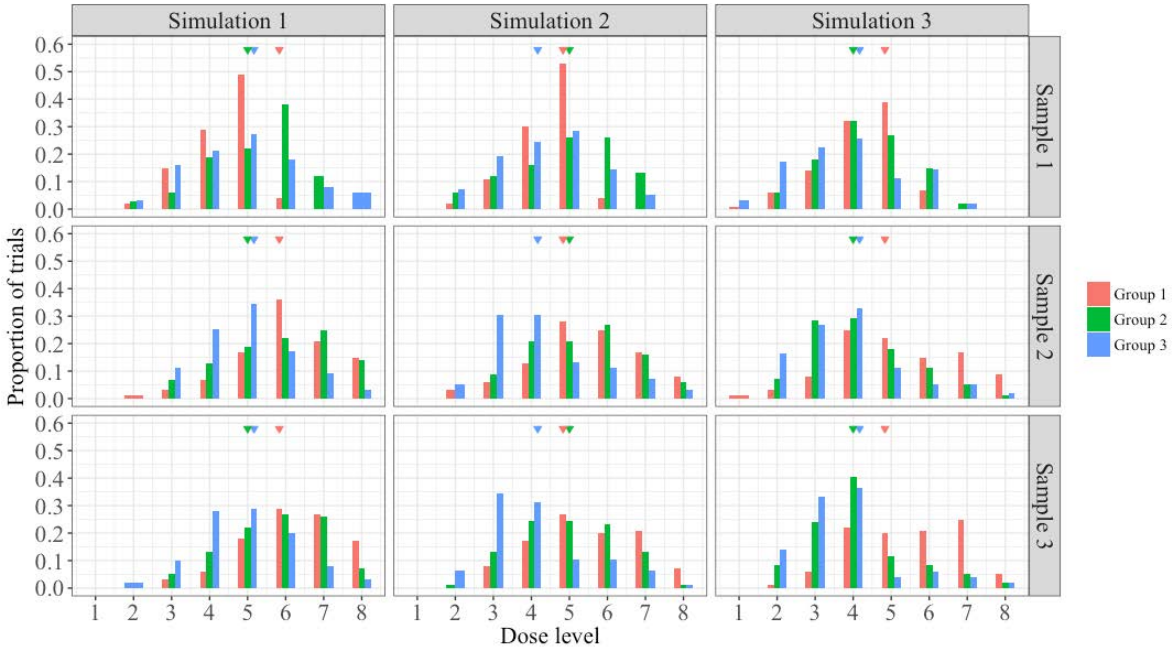

Figure 46. Proportion of trials recommending the true MTD in parallel trials (more patients in higher risk groups)

4.2.2.6 In-trial dose distribution

The in-trial dose allocation of patients is shown is Figure 47. For smallest sample size, there is a great portion of patients assigned to suboptimal dose levels. The distribution of patients of sample

size 2 and 3 are similar, largest portion of patients are assigned to the true MTD, except for group 3 in simulation 3 where most trials recommend dose 4 as the true MTD due to safety.

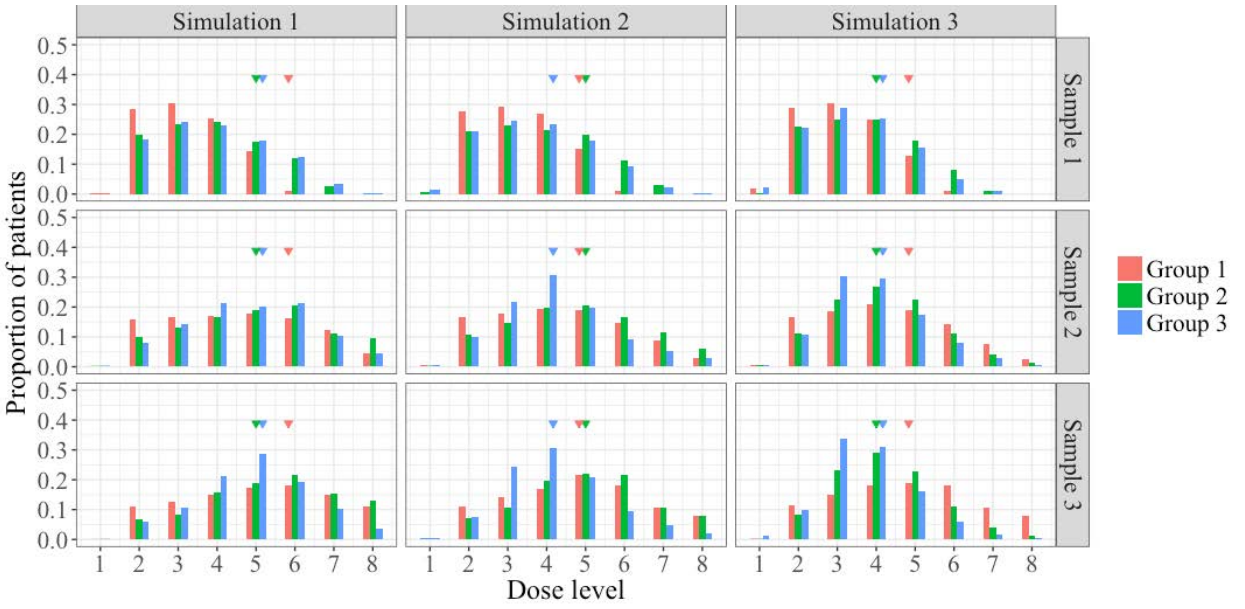

Figure 47. In-trial dose allocation of patients in parallel trials (more patients in higher risk groups)

4.2.2.7 DLT distribution

The numbers of DLTs are summarized in Figure 48. The proportion of DLTs in a trial are summarized in Figure 49. The mean DLT proportion in a trial is around or below 0.2. Group 1 has the lowest DLT proportion per trial possibly because of the smaller sample sizes.

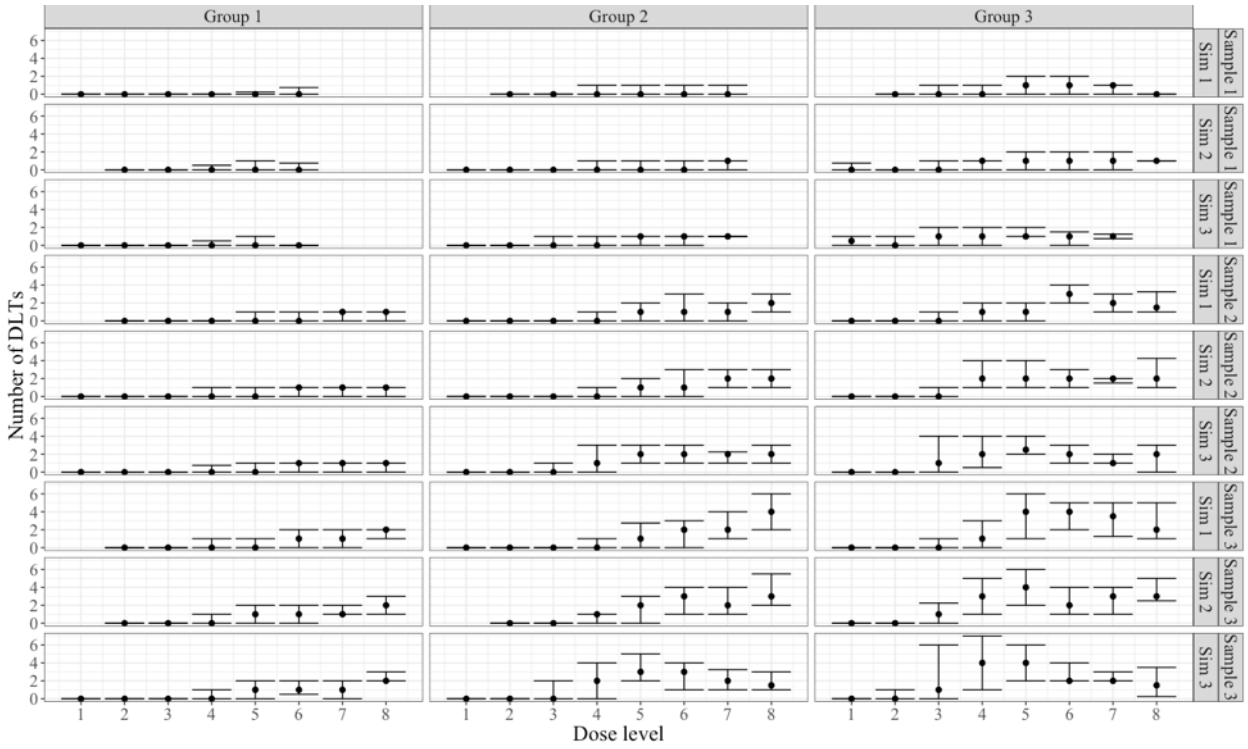

Figure 48. DLTs at all dose levels in parallel trials (more patients in higher risk groups)

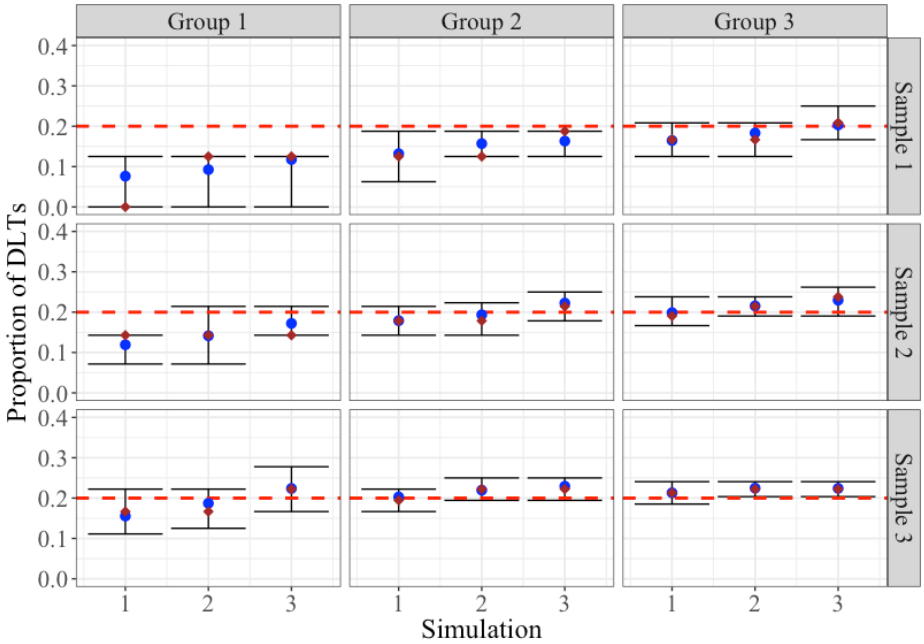

Figure 49. DLT proportion in a trial for parallel trials (more patients in higher risk groups)

5.0 DISCUSSION

In this thesis, we re-parameterized the one-parameter logistic model used in the TITE-CRM dose-escalation trials to account for patient heterogeneity and studied the operating characteristics of the re-parameterized model. Based on the results of the simulations, parameter α is well estimated in all simulations in both group trials using three-parameter logistic model and parallel trials using one-parameter logistic model. However, the estimations of δ_1 and δ_2 of the 3-parameter model are not as good as that of α. The point estimates are not close to the true values and the credible intervals are too wide. Possible reasons for the imprecise estimation could be either the prior for δ_1 and δ_2 are too informative or the estimation is more focused on α - δ_1 than δ_1. In the future, following methods could be used for imporvements:

1. Try different dose-toxicity models such as power model with various prior settings.
2. Try different ways of reparameterizing the logistic model, such as adding the difference variable to the intercept of logistic model.
3. Use the posterior median instead of mean as the parameter estimator for δ_1 and δ_2 because the sample data for both parameters are skewed.

Both group trials and parallel trials achieved a good coverage of true P(DLT) in simulation 1 and 2. However, estimating the true P(DLT) did not perform well when there was a big jump of true P(DLT) between the adjacent two dose levels in simulation 3.

The results when recommending the correct MTD were somewhat similar in both group trials and parallel trials when there were same number of patients in each group. However, when there were more high risk patients in a trial, the result of correct final recommendation in group

trials was better than the parallel trials in terms of the proportion of trials that recommended the true MTD. Generally, the parallel trials tend to recommend higher dose levels than the group trials. Additionally, when there were more high risk patients in a trial, the final recommendation resulted in group trials seemed to be better than when there were same number of patients in each group in terms of the proportion of trials that recommended the true MTD.

Regarding the in-trial dose distribution of patients, more patients were assigned to a more effective dose levels in group trials than the parallel trials in both scenarios, especially when there were more patients with higher risk of toxicity in a trial. More patients were allocated to a dose level of or close to the true MTD when the sample size was 26 or above in a parallel trial.

The proportion of patients experiencing DLT in group trials and parallel trials were both around 0.2 in all simulations, except that proportion of patients experiencing DLT in the parallel trial for simulation 3 group 3 with the largest sample size. From the simulation results, we can see that TITE-CRM tended to allocate patients to higher dose levels as sample size increased. Hence, it is not surprising to see a higher proportion of DLT in the larger-sample-sized trials. However, in reality a phase 1 dose-escalation trial does not recruit that many subjects and DLT should not be an issue in the TITE-CRM trials with dose-escalation constraining rules mentioned in section 3.

The new method can be applied to the phase 1 dose-escalation trial for patients with different toxicity risk to estimate the MTD for each risk group. The method uses TITE-CRM which can shorten the trial and increase the trial efficiency. Compared to the previous TITE-CRM method with two groups [8], this method does not separate a trial into two stages. Rules to control rapid dose-escalation and to reflect ordering information are implemented in the design and are conducted throughout the trial.

However, the thesis work only used a fixed intercept logistic dose-toxicity model. Other commonly models previously metioned could be used for this purpose. Moreover, when using the logistic model, there are not many appropriate options for the prior. Future work could incorporate different prior settings into this new method.

In summary, the TITE-CRM with re-parameterized logistic model performs acceptably in accommodating patient heterogeneity in a trial in terms of operating characteristics. There are improvements that can be made for more precise estimation of true dose-toxicity relationship accommodating patient toxicity risk difference. Different dose-toxicity models with various prior settings or different ways of re-parameterization or using the posterior median as estimate can be studied to see if more precise estimation could be achieved for the risk difference between groups.

PUBLIC HEALTH SIGNIFICANCE

Cancer is one of the major public health concerns in the modern society. The clinical trial is the gold standard to test the safety and efficacy of cancer treatments. The phase 1 clinical trial is the initial stage at which the safety of the treatment is tested and the recommended dose is located. Our study reparameterized the logistic dose-toxicity model to identify the shared information and differences among different risk groups, and located the maximum tolerated dose (MTD) for each risk group. Using the new method in the phase 1 dose-escalation cancer trials could give patients in different risk groups a recommended dose with maximum effectiveness at a corresponding acceptable toxicity level, which can reduce the number of patients exposed to a suboptimal dose or a too toxic dose. In conclusion, the method will have a positive impact on public health.

APPENDIX A: ACRONYMS

3+3: A rule-based dose-escalation method. Treat 3 patients at each dose level, escalate and de-esclate dose level based on the toxicity outcome of the previous cohort of 3 patients.

CRM: Continual Reassement Method

DLT: Dose Limiting Toxicity

MTD: Maximum Tolerated Dose

P(DLT): Probability of a DLT at a dose level

TITE-CRM: Time-to-Event Continual Reassessment Method

APPENDIX B: PROGRAMMING CODE

B.1 CODE FOR SIMULATION

B.1.1 R code (R 3.4.2)

```
rm(list=ls())
library(rjags)
library(plyr)
n.core <- 10
library(doMC); registerDoMC(n.core)
setwd("/ihome/dnormolle/xis54/script")
#-------------------------------------------------------------------------#
# I. Simulation
#----------get random numbers
seeds <- sample(1:100, n.core, replace=F)
#---alphahat function calls the JAGS program to estimate alpha and delta
based on the doses, dlts and days of #observation
w <- foreach(i.core =
             1:n.core,.verbose=T,.errorhandling='stop',.export=ls(.GlobalE
             nv),.packages=c('rjags','plyr')) %dopar% {

 set.seed(seeds[i.core])

 alphahat <-
 function(levs,dlt,ndays,ts,priorstd1,priorstd2,obsdays,tite,group) {
     N <- length(levs)
     dosage <- ts[levs[]]
     w <- rep(1,N)
     if (tite=="Y" | tite=="y") {
     w[dlt==0 & ndays<obsdays] <- ndays[dlt==0 & ndays<obsdays]/obsdays
     }
 data <-
 list(N=N,precision1=1/priorstd1/priorstd1,precision2=1/priorstd2/priorst
      d2,dose=dosage,dlt=dlt,w=w,cohort=group)
 inits <- list(alpha1=1,delta1=delta1[1],delta2=delta2[1])
 n.mcmciter <- 5000
 sink(file="/dev/null")
 model <- jags.model("titecrm-3-
                     parameter.jags",data=data,inits=inits,n.adapt=n.mcmc
                     iter,n.chains=4,quiet=T)
 update(model,n.iter=n.mcmciter,quiet=T)
 samples <-
 jags.samples(model,c("alpha1","delta1","delta2"),n.iter=n.mcmciter,quiet
              =T,n.chains=4)
 sink()
```

```
#---Return ahat, d1hat, d2hat
ahat <- mean(samples$alpha1)
ahat.05 <- quantile(samples$alpha1, 0.05,na.rm = T)
ahat.95 <- quantile(samples$alpha1, 0.95,na.rm = T)

delta1hat <- mean(samples$delta1)
d1hat.05 <- quantile(samples$delta1, 0.05,na.rm = T)
d1hat.95 <- quantile(samples$delta1, 0.95,na.rm = T)

delta2hat <- mean(samples$delta2)
d2hat.05 <- quantile(samples$delta2, 0.05,na.rm = T)
d2hat.95 <- quantile(samples$delta2, 0.95,na.rm = T)

alphas <- cbind(samples$alpha1[1:(n.mcmciter*4)],
                samples$alpha1[1:(n.mcmciter*4)]-
                samples$delta1[1:(n.mcmciter*4)],
                samples$alpha1[1:(n.mcmciter*4)]-
                samples$delta1[1:(n.mcmciter*4)]-
                samples$delta2[1:(n.mcmciter*4)])
    p <- array(NA,dim=c(n.mcmciter*4,nlevel,ncohort))
    for(i in 1:nrow(alphas)){
      for(j in 1:ncohort){
        for(k in 1:nlevel){
            p[i,k,j] <-
            (exp(3+alphas[i,j]*ts1[k]))/(1+exp(3+alphas[i,j]*ts1[k]))
        }
      }
    }
psummary <- data.frame(matrix(0, ncol = 6, nrow =
                       nlevel*ncohort),stringsAsFactors = F)
names(psummary) <- c("mean","p.05","median","p.95","cohort","dose")
psummary$dose <- rep(1:nlevel,ncohort)
psummary$cohort <- rep(1:ncohort,each=nlevel)
for(i in 1:ncohort){
    for(j in 1:nlevel){
                      psummary$mean[psummary$cohort==i&psummary$dose==j]
                      <- mean(p[,j,i],na.rm = T)
                      psummary$p.05[psummary$cohort==i&psummary$dose==j]
                      <- quantile(p[,j,i], 0.05,na.rm = T)
                      psummary$median[psummary$cohort==i&psummary$dose==
                      j] <- quantile(p[,j,i], 0.5,na.rm = T)
                      psummary$p.95[psummary$cohort==i&psummary$dose==j]
                      <- quantile(p[,j,i], 0.95, na.rm = T)
      }
    }

  list(ahat,delta1hat,delta2hat,ahat.05,ahat.95,d1hat.05,d1hat.95,d2hat.
       05,d2hat.95,psummary)
}

#-----------------------------------------------------------------------#
#---assign.dose assigns the dose to the newly presented (this.id)
patient, based on the accumulated data, by calling alphahat
```

```
assign.dose <-
           function(this.id,id,dose,date.on,dlt.date,level1,obsdays,obsp
           eriod,obstype,pi.target,adm.marg,priorstd1,priorstd2,pi.0,tit
           e,cohort) {

#-------The first patient is treated at level1
  if (sharing==0){
    if (this.id==1) {
      if (cohort[this.id]==(ncohort-2)){
        trt <- 3}
      if (cohort[this.id]==(ncohort-1)){
        trt <- 2}
      if (cohort[this.id]==ncohort){
        trt <- 1}
      alpha1.est <- 1
      f.delta1 <- function(delta1) (sum((pi.0[,ncohort-1]-
                  exp(3+(alpha1.est-delta1)*ts1)/(1+exp(3+(alpha1.est-
                  delta1)*ts1)))^2))
      delta1.est <- optimize(f.delta1, c(0, 1), tol = 0.000001,
                    maximum=FALSE)$minimum
      f.delta2 <- function(delta2) (sum((pi.0[,ncohort]-
                  exp(3+(alpha1.est-delta1.est-
                  delta2)*ts1)/(1+exp(3+(alpha1.est-delta1.est-
                  delta2)*ts1)))^2))
      delta2.est <- optimize(f.delta2, c(0, 1), tol = 0.000001,
              maximum=FALSE)$minimum
      ahat.05 <- 0
      ahat.95 <- 0
      d1hat.05 <- 0
      d1hat.95 <- 0
      d2hat.05 <- 0
      d2hat.95 <- 0
      psummary <- NA
    }

    #-------Otherwise,
    if (this.id>1) {
          all <- data.frame(id,dose,date.on,dlt.date,cohort)
          npats <- this.id-1
          all <- all[1:npats,]
          today <- date.on[this.id]
          all$dlt <-  as.numeric(all$dlt.date<=today)
          nlevel <- dim.pi.0[1]

    #-------Add number of days patients have been on study to all
          all$ndays <- obsdays
          all$ndays[all$dlt==0] <- today - all$date.on[all$dlt==0]
          all$ndays[all$ndays>obsdays] <- obsdays

    #-------Estimate alpha & beta given current data
          ests <- alphahat(all$dose,all$dlt,all$ndays,ts,
                           priorstd1,priorstd2,obsdays,tite,all$cohort)
          alpha1.est <- ests[[1]]
```

```
          delta1.est <- ests[[2]]
          delta2.est <- ests[[3]]
          ahat.05 <- ests[[4]]
          ahat.95 <- ests[[5]]
          d1hat.05 <- ests[[6]]
          d1hat.95 <- ests[[7]]
          d2hat.05 <- ests[[8]]
          d2hat.95 <- ests[[9]]
          psummary <- ests[[10]]
          alpha.now <- c(alpha1.est,alpha1.est-delta1.est,alpha1.est-
                         delta1.est-delta2.est)
#-------Determine number of patients, number of days and number of
#complete
#-------patients at each level and put them in data frame expose
          expose <- array(0,dim=c(nlevel,ncohort,2))
          for (i.all in 1:nrow(all)) {
               expose[all[i.all,]$dose,all[i.all,]$cohort,2] <-
               expose[all[i.all,]$dose,all[i.all,]$cohort,2] +
               all[i.all,]$ndays
               if (all[i.all,]$ndays>=obsdays | all[i.all,]$dlt==1)
                  expose[all[i.all,]$dose,all[i.all,]$cohort,1] <-
                  expose[all[i.all,]$dose,all[i.all,]$cohort,1] + 1
               }

    #-------calculate pi.now
               pi.now <- matrix(rep(NA,ncohort*nlevel),ncol=ncohort)
               for (col in 1:ncohort) {
                    pi.now[,col] <-
                    (exp(3+alpha.now[col]*ts1))/(1+exp(3+alpha.now[col]*
                    ts1))
               }

    #-------assign first dose to each cohort
        if(is.element(cohort[this.id],head(cohort,n=this.id-1))==FALSE)
              {
                 if (cohort[this.id]==(ncohort-2)){
                   trt <- 3}
                 if (cohort[this.id]==(ncohort-1)){
                   trt <- 2}
                 if (cohort[this.id]==ncohort){
                   trt <- 1}
               }

    #-------assign second and following doses to each cohort
    #-------Calculate P(DLT) given current estimate of alpha & beta

if(cohort[this.id] %in% head(cohort,n=this.id-1)) {

    #-------Select dose based on alpha & beta estimate and the RULES
    #     First, just based on alpha and beta
            trt <- 0
            for (d in 1:nlevel) {
              if (pi.now[d,cohort[this.id]]<=pi.target+adm.marg) {trt <-
                  d}
```

```
            }

            all <- all[order(all$date.on),]
            l.d.a <- rep(NA,ncohort)
            l.d.a[cohort[this.id]] <-
            tail(all[which(all$cohort==cohort[this.id]),],n=1)$dose

#-------RULE 1:  Jump only one dose between patients
            if (trt > l.d.a[cohort[this.id]]+1){
                trt <- l.d.a[cohort[this.id]]+1}
#-------RULE 2:  If a lower dose needs more exposure,
            lodose <-
            head(all[which(all$cohort==cohort[this.id]),],n=1)$dose
            if (any(all$dose<level1)) {
                  lodose <- min(all$dose,na.rm=T)
                  level1 <- lodose
            }
            if (trt>lodose) {
                 sflag <- 0

    #-----------Count complete patients
                if (obstype==1) {
                      for (d in lodose:(trt-1)) {
                        if (sflag==0 &
                            expose[d,cohort[this.id],1]<obsperiod) {
                            trt <- d
                            sflag <- 1
                           }
                      }
                } #---End obstype==1

    #-----------Count observation periods
                if (obstype==2) {
                      for (d in lodose:(trt-1)) {
                           if (sflag==0 &
                               expose[d,cohort[this.id],2]<obsdays*obsper
                               iod) {
                                 trt <- d
                                 sflag <- 1
                           }
                      }
                } #---End obstype==2
           } #--End trt>lodose

#-------If trt=0, set it to the lowest dose until k patients #observed.
            if (trt == 0 & this.id <= k) {
              trt <- 1
            }
       } #--End of assign 2nd and following dose to each cohort
    } #---End this is at least the second patient
  } #--End of sharing == 0

###-----------------------------------------------------------------###
if (sharing==1){
```

```
if (this.id==1) {

  if (cohort[this.id]==(ncohort-2)){
    trt <- 3}
  if (cohort[this.id]==(ncohort-1)){
    trt <- 2}
  if (cohort[this.id]==ncohort){
    trt <- 1}

   alpha1.est <- 1
   f.delta1 <- function(delta1) (sum((pi.0[,ncohort-1]-
               exp(3+(alpha1.est-delta1)*ts1)/(1+exp(3+(alpha1.est-
               delta1)*ts1)))^2))
   delta1.est <- optimize(f.delta1, c(0, 1), tol = 0.000001,
                  maximum=FALSE)$minimum
   f.delta2 <- function(delta2) (sum((pi.0[,ncohort]-
               exp(3+(alpha1.est-delta1.est-
               delta2)*ts1)/(1+exp(3+(alpha1.est-delta1.est-
               delta2)*ts1)))^2))
   delta2.est <- optimize(f.delta2, c(0, 1), tol = 0.000001,
                  maximum=FALSE)$minimum

  ahat.05 <- 0
  ahat.95 <- 0
  d1hat.05 <- 0
  d1hat.95 <- 0
  d2hat.05 <- 0
  d2hat.95 <- 0
  psummary <- NA
} #-- end of this.id ==1

#-------Otherwise, we have to think

if (this.id>1) {

  all <- data.frame(id,dose,date.on,dlt.date,cohort)
  npats <- this.id-1
  all <- all[1:npats,]
  today <- date.on[this.id]
  all$dlt <-  as.numeric(all$dlt.date<=today)
  nlevel <- dim.pi.0[1]

  #-------Add number of days patients have been on study to all

  all$ndays <- obsdays
  all$ndays[all$dlt==0] <- today - all$date.on[all$dlt==0]
  all$ndays[all$ndays>obsdays] <- obsdays

  #-------Estimate alpha & beta given current data

  ests <- alphahat(all$dose,all$dlt,all$ndays,ts,
                   priorstd1,priorstd2,obsdays,tite,all$cohort)
```

```
  alpha1.est <- ests[[1]]
  delta1.est <- ests[[2]]
  delta2.est <- ests[[3]]
  ahat.05 <- ests[[4]]
  ahat.95 <- ests[[5]]
  d1hat.05 <- ests[[6]]
  d1hat.95 <- ests[[7]]
  d2hat.05 <- ests[[8]]
  d2hat.95 <- ests[[9]]
  psummary <- ests[[10]]
  alpha.now <- c(alpha1.est,alpha1.est-delta1.est,alpha1.est-
                 delta1.est-delta2.est)

#-------Determine number of patients, number of days and number of
#complete  patients at each level and #put them in data frame expose

  expose <- array(0,dim=c(nlevel,ncohort,2))
  for (i.all in 1:nrow(all)) {
        expose[all[i.all,]$dose,all[i.all,]$cohort,2] <-
        expose[all[i.all,]$dose,all[i.all,]$cohort,2]
        +all[i.all,]$ndays
        if (all[i.all,]$ndays>=obsdays | all[i.all,]$dlt==1)
            expose[all[i.all,]$dose,all[i.all,]$cohort,1] <-
            expose[all[i.all,]$dose,all[i.all,]$cohort,1] + 1
  }

  #-------calculate pi.now
  pi.now <- matrix(rep(NA,ncohort*nlevel),ncol=ncohort)
  for (col in 1:ncohort) {
       pi.now[,col] <-
       (exp(3+alpha.now[col]*ts1))/(1+exp(3+alpha.now[col]*ts1))
  }

  #-------assign first dose to each cohort
   if(is.element(cohort[this.id],head(cohort,n=this.id-1))==FALSE) {
     if (cohort[this.id]==(ncohort-2)){
       trt <- 3}
     if (cohort[this.id]==(ncohort-1)){
       trt <- 2}
     if (cohort[this.id]==ncohort){
       trt <- 1}

     if (obstype==1) {
       for(i in 1:(this.id-1)){
           if (expose[dose[i],cohort[i],1]>obsperiod &
           cohort[this.id]<cohort[i]) {trt <- dose[i]}
       }
     }
     if (obstype==2) {
       for(i in 1:(this.id-1)){
           if (expose[dose[i],cohort[i],2]>obsdays*obsperiod &
           cohort[this.id]<cohort[i]) {trt <- dose[i]}
       }
     }
```

```
        } #end of assigning first dose to each cohort when this.id >1

#-------assign second and following doses to each cohort
  #-------Calculate P(DLT) given current estimate of alpha & delta

      if(cohort[this.id] %in% head(cohort,n=this.id-1)) {

#-------Select dose based on alpha & beta estimate and the RULES
        #First, just based on alpha and beta

        trt <- 0

        for (d in 1:nlevel) {
          if (pi.now[d,cohort[this.id]]<=pi.target+adm.marg) {trt <- d}
        }

        all <- all[order(all$date.on),]
        l.d.a <- rep(NA,ncohort)
        l.d.a[cohort[this.id]] <-
        tail(all[which(all$cohort==cohort[this.id]),],n=1)$dose

#-------RULE 1:  Jump only one dose between patients within a #cohort
#(rule 1 will be overridden by rule 2 and 3 if there is a conflict)

          if (trt > l.d.a[cohort[this.id]]+1){
              trt <- l.d.a[cohort[this.id]]+1
          }

#-------RULE 2:  If a lower dose needs more exposure, go there

        lodose <-
        head(all[which(all$cohort==cohort[this.id]),],n=1)$dose
        if
        (any(all$dose<head(all[which(all$cohort==cohort[this.id]),],n=1)
        $dose)) {
                  lodose <- min(all$dose,na.rm=T)
                  level1 <- lodose
          }

        if (trt>lodose) {
          sflag <- 0

         #-----------Count complete patients

          if (obstype==1) {
            for (d in lodose:(trt-1)) {
              if (sflag==0 & expose[d,cohort[this.id],1]<obsperiod) {
                trt <- d
                sflag <- 1
              }
            }
          } #---End obstype==1

         #-----------Count observation periods
```

```
          if (obstype==2) {
            for (d in lodose:(trt-1)) {
                 if (sflag==0 &
                     expose[d,cohort[this.id],2]<obsdays*obsperiod) {
                     trt <- d
                     sflag <- 1
              }
            }
          } #---End obstype==2
        } #--End trt>lodose


#-------RULE 3: The dose level of lower risk dose are not lower than
#the most recent higher risk dose level.

        if (obstype==1) {
          for(i in 1:(this.id-1)){
              if (expose[dose[i],cohort[i],1]>obsperiod &
              cohort[this.id]<cohort[i] & dose[i]>trt) {trt <-
              dose[i]+1}
          }
        }
        if (obstype==2) {
          for(i in 1:(this.id-1)){
              if (expose[dose[i],cohort[i],2]>obsdays*obsperiod &
              cohort[this.id]<cohort[i] & dose[i]>trt) {trt <-
              dose[i]+1}
          }
        }

        if (trt > nlevel) {trt <- nlevel}

#-------If trt=0, set it to the lowest dose until k patients observed.
        if (trt == 0 & this.id <= k) {
          trt <- 1
        }

      } #--End of assign 2nd and following dose to each cohort
    } #---End this is at least the second patient
  } #--End of sharing == 1


list(trt,alpha1.est,delta1.est,delta2.est,ahat.05,ahat.95,d1hat.05,d1hat
     .95,d2hat.05,d2hat.95,psummary)
} #---End assign.dose

#-----------------------------------------------------------------------#
#---Trial Descriptors
#   tite is the trial TITE-CRM (Y/N)
#   level1 is the first dose level
#   obsdays is the number of days the participants are observed for DLT
#   obsperiod is the multiple of units must be observed at a level prior
#to escalation to the next level
```

```
#   obstype 1:  units=completed patients  2:  units=observation periods
#   pi.target is the target rate
#   target dose is the highest dose with P(DLT)<pi.target+adm.marg.  Set
#adm.marg to 0
#   to never exceed target rate (very conservative)
#   priorstd is the standard deviation on the prior.  Ususally 0.3 for
#alpha. Set 0.1 for delta
#   n.max is the number of patients/trial
#   recruit.days is the number of days to recruit n.max patients
#   pi.0 is the assumed skeleton
#   sharing is the whether or not the rules of ordering information are
#applied
#   k is the least number of patients to be treated in a trial, control
#for failed trial
#   a1, d1 and d2 are true parameters pre-specified

tite <- "Y"
level1 <- 2
obsdays <- 42
obsperiod <- 1
obstype <- 2
pi.target <- 0.2
adm.marg <- 0.05
priorstd1 <- 0.3
priorstd2 <- 0.1
sharing <- 1
a1 <- c(1,0.94,0.9)
d1 <- c(0.048,0.035,0.078)
d2 <- c(0.042,0.07,0.047)

#---define n.max and id and at least k patients receiving treatment in
#each trial.

n.max <- 80
id <- 1:n.max
k <- 10
recruit.days <- 600

#---Rescale dose and P(DLT)

pi.0.1 <- c(0.01,0.025,0.05,0.1,0.15,0.2,0.25,0.3)
ts1 <- log(pi.0.1/(1-pi.0.1))-3

pi.0.2 <- exp(3+(a1[1]-d1[1])*ts1)/(1+exp(3+(a1[1]-d1[1])*ts1))
pi.0.3 <- exp(3+(a1[1]-d1[1]-d2[1])*ts1)/(1+exp(3+(a1[1]-d1[1]-
          d2[1])*ts1))
pi.0 <- cbind(pi.0.1,pi.0.2,pi.0.3)
dim.pi.0 <- dim(pi.0)
nlevel <- dim.pi.0[1]

#-------------------------------------------------------------------------#

#---Simulation Controls
#
```

```
#   n.sim is the number of simulation skeletons
#   Number the skeletons pi.1, etc. and concatenate them into the matrix
#pi.true.all
#   n.trial is the number of trials per skeleton

n.sim <- 3
ncohort <- 3
pi.1 <- pi.0

pi.2.1 <- c(0.016,0.037,0.07,0.132,0.19,0.245,0.299,0.351)
ts2 <- (log(pi.2.1/(1-pi.2.1))-3)/a1[2]
pi.2.2 <- exp(3+(a1[2]-d1[2])*ts2)/(1+exp(3+(a1[2]-d1[2])*ts2))
pi.2.3 <- exp(3+(a1[2]-d1[2]-d2[2])*ts2)/(1+exp(3+(a1[2]-d1[2]-
          d2[2])*ts2))
pi.2 <- cbind(pi.2.1,pi.2.2,pi.2.3)

pi.3.1 <- c(0.021,0.047,0.087,0.127,0.237,0.279,0.33,0.38)
ts3 <- (log(pi.3.1/(1-pi.3.1))-3)/a1[3]
pi.3.2 <- exp(3+(a1[3]-d1[3])*ts3)/(1+exp(3+(a1[3]-d1[3])*ts3))
pi.3.3 <- exp(3+(a1[3]-d1[3]-d2[3])*ts3)/(1+exp(3+(a1[3]-d1[3]-
          d2[3])*ts3))
pi.3 <- cbind(pi.3.1,pi.3.2,pi.3.3)

pi.true.all <- array(c(pi.1,pi.2,pi.3),c(nlevel,ncohort,n.sim) )

ts.all <- cbind(ts1,ts2,ts3)

n.trial <- 10
#-----------------------------------------------------------------------#
#---Simulation loop. If assign.dose returns 0, stop.flag is set to 1 and
no more doses are assigned for that trial

for (i.sim in 1:n.sim) {
    pi.true <- pi.true.all[,,i.sim]
    sim <- rep(i.sim,n.max)
    starget <- rep(pi.target,n.max)
    core <- rep(i.core,n.max)
    ts <- ts.all[,1]

    for (i.trial in 1:n.trial) {

          dose <- rep(0,n.max)
          date.on <-ceiling(sort(runif(n.max,0,recruit.days)))
          dlt <- rep(NA,n.max)
          alpha1 <- rep(NA,n.max)
          delta1 <- rep(NA,n.max)
          delta2 <- rep(NA,n.max)
          alpha1.05 <- rep(NA,n.max)
          alpha1.95 <- rep(NA,n.max)
          delta1.05 <- rep(NA,n.max)
          delta1.95 <- rep(NA,n.max)
          delta2.05 <- rep(NA,n.max)
          delta2.95 <- rep(NA,n.max)
          dlt.date <- rep(9999,n.max)
```

```
cohort<- sample(1:3,n.max,replace = T)
trial <- rep(i.trial,n.max)
psum <-
data.frame(matrix(NA,nrow=ncohort*nlevel*n.max,ncol=10))
colnames(psum) <-
c("mean","p.05","median","p.95","cohort","dose","id","trial","
sim","core")
psum$core <- i.core
psum$sim <- i.sim
psum$trial <- i.trial
stop.flag <- 0

for (this.id in 1:n.max) {
      if (stop.flag==0) {
        hat <-
        assign.dose(this.id,id,dose,date.on,dlt.date,level1,ob
        sdays,obsperiod,obstype,pi.target,adm.marg,priorstd1,p
        riorstd2,pi.0,tite,cohort)
        dose[this.id] <- hat[[1]]
        alpha1[this.id] <- hat[[2]]
        delta1[this.id] <- hat[[3]]
        delta2[this.id] <- hat[[4]]
        alpha1.05[this.id] <- hat[[5]]
        alpha1.95[this.id] <- hat[[6]]
        delta1.05[this.id] <- hat[[7]]
        delta1.95[this.id] <- hat[[8]]
        delta2.05[this.id] <- hat[[9]]
        delta2.95[this.id] <- hat[[10]]
        psum[((this.id-
        1)*ncohort*nlevel+1):(this.id*ncohort*nlevel),1:6] <-
        hat[[11]]
        psum$id[((this.id-
        1)*ncohort*nlevel+1):(this.id*ncohort*nlevel)] <-
        this.id
      }

      if (dose[this.id]>0 & stop.flag==0) {
          dlt[this.id] <-
          rbinom(1,1,pi.true[dose[this.id],cohort[this.id]])
          if (dlt[this.id]==1) dlt.date[this.id] <-
          date.on[this.id]+ceiling(runif(1,0,obsdays))
      } else {
           stop.flag <- 1
      }
} # end of this.id loop

 this.trial <-
 data.frame(core,sim,trial,cohort,id,dose,date.on,dlt,dlt.date,
            alpha1,delta1,delta2,alpha1.05,alpha1.95,delta1.05
            ,delta1.95,delta2.05,delta2.95,starget)

if (i.sim==1 & i.trial==1) {
      all.trial <- this.trial
```

```
            } else {
                 all.trial <- rbind(all.trial,this.trial)
            }


            if (i.sim==1 & i.trial==1) {
              all.psum <- psum
            } else {
              all.psum <- rbind(all.psum,psum)
            }
            print(paste("sim=",i.sim,"  trial=",i.trial,sep=""),quote=F)
        }# end of i.trial loop
    }# end of i.sim loop

    all.trial$dlt.date[all.trial$dlt.date==9999] <- NA

    #-----------------------------------------------------------------------#
    #---Determine proportion of trials that fail.

    trial.fail <-
    ddply(all.trial,.(core,sim,trial),summarise,fail=any(dose==0))

    all.trial <- merge(all.trial,trial.fail,by=c("core","sim","trial"))

    ok.trial <- all.trial[!all.trial$fail,]
    list(all.psum,trial.fail,all.trial,ok.trial)
} # end of i.core loop

# merge datasets from different cores
all.pi <- w[[1]][[1]]
for (i.core in 2:n.core) {
  all.pi <- rbind(all.pi,w[[i.core]][[1]])
}

all.fail <- w[[1]][[2]]
for (i.core in 2:n.core) {
  all.fail <- rbind(all.fail,w[[i.core]][[2]])
}

all.trial <- w[[1]][[3]]
for (i.core in 2:n.core) {
  all.trial <- rbind(all.trial,w[[i.core]][[3]])
}

ok.trial <- w[[1]][[4]]
for (i.core in 2:n.core) {
  ok.trial <- rbind(ok.trial,w[[i.core]][[4]])
}
#-----------------------------------------------------------------------#
# export datasets needed
write.csv(all.trial, "3cohort_all_trial80.csv",row.names = FALSE)
write.csv(ok.trial, "3cohort_oktrial80.csv",row.names = FALSE)
write.csv(all.fail, "3cohort_failtrial80.csv",row.names = FALSE)
write.csv(all.pi, "3cohort_all_pi80.csv",row.names = FALSE)
```

B.1.2 JAGS code (JAGS 4.3.0)

```
model {
      for ( i in 1:N ){
            p[i] <-
(w[i]*exp(3+alpha[cohort[i]]*dose[i]))/(1+exp(3+alpha[cohort[i]]*dose[i]))
            dlt[i] ~ dbern(p[i])
            }
alpha <- c(alpha1,alpha1-delta1,alpha1-delta1-delta2)
      alpha1 ~ dnorm(1,precision1)
      delta1 ~ dnorm(0,precision2) T(0, )
      delta2 ~ dnorm(0,precision2) T(0, )
}
```

B.2 CODE FOR ANALYSIS

B.2.1 R code (R 3.4.2)

```
library(plyr)
library(dplyr)
library(ggplot2)

#---Clear leftover variables out

rm(list=ls())

all.trial <- read.csv("3cohort_all_trial.csv")
trial.fail <- read.csv("3cohort_failtrial.csv")
ok.trial <- read.csv("3cohort_oktrial.csv")
all.psum <- read.csv("3cohort_all_pi.csv")

#I. Data management in general
# manage ok.trial

oktrial.number1 <-
ddply(ok.trial,.(sim,core),summarise,trial.number=length(unique(trial)))
oktrial.number <-
ddply(oktrial.number1,.(sim),summarise,trial.number=sum(trial.number))
ok.trial <-
ok.trial[order(ok.trial$sim,ok.trial$core,ok.trial$trial,ok.trial$id),]

for(i in 1:n.sim){
    ok.trial$tr[ok.trial$sim==i] <-
    rep(1:oktrial.number$trial.number[i],each=n.max)
}
```

```
# manage all.psum
all.psum <-
all.psum[order(all.psum$sim,all.psum$core,all.psum$trial,all.psum$id),]
for(i in 1:n.sim){
  all.psum$tr[all.psum$sim==i] <- rep(1:n.trial,each=n.max*nlevel*ncohort)
}

#---calculate mean interarrival time for one trial
interArrival <- rep(NA,n.max-1)
for(i in 2:n.max){
  interArrival[i-1] <- all.trial$date.on[i]-all.trial$date.on[i-1]
}
interArrival.bar <- mean(interArrival)

# II.Analysis of operation characteristics based on 100 trials per
#simulation for group trials. Three simulations in total.

# 1. Compare true pi and estimated pi for diagnostics.

# ---data management for pi.true
for (i in 1:nrow(ok.trial)) {
     ok.trial$pi.true[i] <-
     pi.true.all[ok.trial$dose[i],ok.trial$cohort[i],ok.trial$sim[i]]
}

#create a dataset for pi.true
piTrue <- data.frame(matrix(nrow=nlevel*ncohort*n.sim, ncol=4))
colnames(piTrue) <- c("sim","cohort","dose","pi")
piTrue$sim <- rep(1:n.sim,each=nlevel*ncohort)
piTrue$cohort <- rep(1:ncohort,each=nlevel)
piTrue$dose <- rep(1:nlevel)
for(i.sim in 1:n.sim){
  for(i.cohort in 1:ncohort){
    for(i.dose in 1:nlevel){
        piTrue$pi[piTrue$sim==i.sim&piTrue$cohort==i.cohort&piTrue$dose==i
        .dose] <- pi.true.all[i.dose,i.cohort,i.sim]
    }
  }
}

## -- True RP2D
### 1. rp2d is the dose of which pi <= pi.target
for(i.sim in 1:n.sim){
  for(i.cohort in 1:ncohort){
      piTrue$rp2d.1[piTrue$sim==i.sim&piTrue$cohort==i.cohort] <-
      piTrue$dose[which.max(piTrue$pi[piTrue$sim==i.sim&piTrue$cohort==i.
      cohort][piTrue$pi[piTrue$sim==i.sim&piTrue$cohort==i.cohort] <=
      pi.target])]
  }
}

### 2. rp2d is the dose of which pi is closest pi.target
piTrue$diff <- abs(piTrue$pi-pi.target)
```

```
for(i.sim in 1:n.sim){
  for(i.cohort in 1:ncohort){
      piTrue$rp2d.2[piTrue$sim==i.sim&piTrue$cohort==i.cohort] <-
      piTrue$dose[which.min(piTrue$diff[piTrue$sim==i.sim&piTrue$cohort==i
      .cohort])]
  }
}

write.csv(piTrue, "3cohort_piTrue.csv",row.names = FALSE)

piTrue <- read.csv("3cohort_piTrue.csv")

## for pi estimates of last observation
lastObs.p <- filter(all.psum,id==n.max)

for(i in 1:nrow(lastObs.p)){
    lastObs.p$pi.true[i] <-
    piTrue$pi[piTrue$sim==lastObs.p$sim[i]&piTrue$cohort==lastObs.p$cohor
    t[i]&piTrue$dose==lastObs.p$dose[i]]
}
## check coverage of true pi.
for(i in 1:nrow(lastObs.p)){
     if(lastObs.p$pi.true[i] <= lastObs.p$p.95[i] & lastObs.p$mean[i] >=
        lastObs.p$p.05[i]){
           lastObs.p$contain[i] <- 1 }
    else{
           lastObs.p$contain[i] <- 0
  }
}

coverage.pi <-
ddply(lastObs.p,.(sim,cohort,dose),summarise,contain.pi=length(mean[contai
n==1]),trial=length(tr))
coverage.pi$coverage <- coverage.pi$contain.pi/coverage.pi$trial
write.csv(coverage.pi, "3cohort_rule_coverage_pi_120.csv",row.names =
FALSE)

coverage.40 <- read.csv("3cohort_rule_coverage_pi_40.csv")
coverage.80 <- read.csv("3cohort_rule_coverage_pi_80.csv")
coverage.120 <- read.csv("3cohort_rule_coverage_pi_120.csv")

###plot for coverage of true pi.
### data management

coverage.40$sample <- 1 # 1 means sample size is 40
coverage.80$sample <- 2 # 2 means sample size is 80
coverage.120$sample <- 3 # 3 means sample size is 120

coverage.40.80.120 <- rbind(coverage.40,coverage.80,coverage.120)

for(i in 1:nrow(coverage.40.80.120)){
    coverage.40.80.120$pi.true[i] <-
    piTrue$pi[piTrue$sim==coverage.40.80.120$sim[i] &
```

```
    piTrue$cohort==coverage.40.80.120$cohort[i] &
    piTrue$dose==coverage.40.80.120$dose[i]]
}

write.csv(coverage.40.80.120,
"3cohort_coverage_pi_40_80_120.csv",row.names = FALSE)

coverage.40.80.120 <- read.csv("3cohort_coverage_pi_40_80_120.csv")

### plot
coverage.40.80.120$sample <- factor(coverage.40.80.120$sample,
                                    levels = c(1:3),
                                    labels = c("Sample 1", "Sample
                                    2","Sample 3"))
coverage.40.80.120$sim <- factor(coverage.40.80.120$sim,
                                 levels = c(1:n.sim),
                                 labels = c("Simulation 1", "Simulation
                                 2","Simulation 3"))
coverage.40.80.120$cohort <- factor(coverage.40.80.120$cohort,
                                    levels = c(1:ncohort),
                                    labels = c("Group 1", "Group
                                    2","Group 3"))

 ggplot(data=coverage.40.80.120,
        aes(x=pi.true,y=coverage,colour=cohort,group=cohort))+
        facet_grid(sample~sim)+theme_bw()+
        geom_point(size=3,pch=18)+geom_line()+
        theme(legend.title=element_blank(),legend.text=element_text(size=1
              6,family="Times New Roman"),
        plot.title = element_text(hjust = 0.5,size=20,family="Times New
                                  Roman"),
        plot.subtitle = element_text(size=18,family = "Times New Roman"),
        axis.title = element_text(size=18,family="Times New Roman"),
        axis.text = element_text(size=14,family="Times New Roman"),
        strip.text=element_text(size=16,family="Times New Roman"))+
        labs(list(x="True P(DLT)",y="Percentage of CI coverage of true
             P(DLT)",
        subtitle="Patient's probability to be assigned to each group =
        0.2, 0.35, 0.45 \n 90% Credible interval coverage"))+
        scale_y_continuous(limits = c(0.8,1),labels = scales::percent)+
        scale_x_continuous(breaks=pretty(coverage.40.80.120$pi.true, n =
                           6))+
        geom_hline(yintercept=0.9,colour="brown")


# 2. estimate parameters.

## created dataset with last observations.
lastObs <- filter(ok.trial,id==n.max)

for(j in 1:n.sim){
  lastObs$a1[lastObs$sim==j] <- a1[j]
  lastObs$d1[lastObs$sim==j] <- d1[j]
  lastObs$d2[lastObs$sim==j] <- d2[j]
```

```
}

## to see if alpha1 credible interval contains true alpha1,delta1 and
#delta2
for(i in 1:nrow(lastObs)){
    if(lastObs$alpha1.05[i] <= lastObs$a1[i] &
    lastObs$a1[i]<=lastObs$alpha1.95[i]){
        lastObs$contain.a1[i] <- 1}
    else{
          lastObs$contain.a1[i] <- 0}


if(lastObs$delta1.05[i]<=lastObs$d1[i]&lastObs$d1[i]<=lastObs$delta1.95[i]
){
        lastObs$contain.d1[i] <- 1}
  else{
        lastObs$contain.d1[i] <- 0}

if(lastObs$delta2.05[i]<=lastObs$d2[i]&lastObs$d2[i]<=lastObs$delta2.95[i]
){
        lastObs$contain.d2[i] <- 1}
  else{
        lastObs$contain.d2[i] <- 0}
}

coverage <-
ddply(lastObs,.(sim),summarise,alpha1=length(alpha1[contain.a1==1]),
      delta1=length(delta1[contain.d1==1]),delta2=length(delta2[contain.d
      2==1]))
coverage$a1p <- coverage$alpha1/oktrial.number$trial.number
coverage$d1p <- coverage$delta1/oktrial.number$trial.number
coverage$d2p <- coverage$delta2/oktrial.number$trial.number

write.csv(coverage, "3cohort_coverage_parameters_120.csv",
          row.names = FALSE)


###plot for 3 sample sizes
cov.para.40 <- read.csv("3cohort_coverage_parameters_40.csv")
cov.para.80 <- read.csv("3cohort_coverage_parameters_80.csv")
cov.para.120 <- read.csv("3cohort_coverage_parameters_120.csv")

cov.para.40$sample <- 1 # sample size: 40
cov.para.80$sample <- 2 # sample size: 40
cov.para.120$sample <- 3 # sample size: 120

cov.para <- rbind(cov.para.40,cov.para.80,cov.para.120)

write.csv(cov.para,"coverage_parameter.csv",row.names = FALSE)

cov.para$sample <- factor(cov.para$sample,
                          levels = c(1:3),
                          labels = c("Sample 1", "Sample 2", "Sample 3"))
```

```
ggplot(cov.para, aes(sim,a1p)) + theme_bw()+
geom_point(stat = "identity") + facet_grid(.~sample)+
      labs(list(title = "Coverage of alpha1", x = "Simulation", y =
           "Proportion of trials",
      subtitle="Patient's probability to be assigned to each group = 0.2,
                0.35 0.45"))+
      ylim(0.8,1)+
      scale_x_continuous(breaks=seq(0,n.sim,1)) +
      theme(legend.title=element_blank(),legend.text=element_text(size=16,
            family="Times New Roman"),
      plot.title = element_text(hjust = 0.5,size=20,family="Times New
                                Roman"),
      plot.subtitle = element_text(size=18,family = "Times New Roman"),
      axis.title = element_text(size=18,family="Times New Roman"),
      axis.text = element_text(size=16,family="Times New Roman"),
      strip.text=element_text(size=18,family="Times New Roman"))+
      geom_line()+
      geom_hline(yintercept=0.9, colour="brown")

ggplot(cov.para, aes(sim,d2p)) + theme_bw()+
      geom_point(stat = "identity") + facet_grid(.~sample)+
      labs(list(title = "Coverage of delta2", x = "Simulation", y =
           "Proportion of trials",
      subtitle="Patient's probability to be assigned to each group = 0.2,
                0.35, 0.45"))+
      ylim(0.8,1)+
      scale_x_continuous(breaks=seq(0,n.sim,1)) +
      theme(legend.title=element_blank(),legend.text=element_text(size=16,
            family="Times New Roman"),
      plot.title = element_text(hjust = 0.5,size=20,family="Times New
                                Roman"),
      plot.subtitle = element_text(size=18,family = "Times New Roman"),
      axis.title = element_text(size=18,family="Times New Roman"),
      axis.text = element_text(size=16,family="Times New Roman"),
      strip.text=element_text(size=18,family="Times New Roman"))+
      geom_line()+
      geom_hline(yintercept=0.9, colour="brown")

#2.in-trial patient allocation
### calcualate patient proportion for at each dose level for each cohort
patient.dlt.proportion <-
ddply(ok.trial,.(sim,cohort,dose),summarise,patients=length(id),
      dlt.number=length(dlt[dlt==1]))
patient.dlt.proportion2 <-
ddply(patient.dlt.proportion,.(sim,cohort),summarise,
      patients=sum(patients),dlt=sum(dlt.number))
proportion <-
merge(patient.dlt.proportion,patient.dlt.proportion2,by=c("sim","cohort"))
proportion$patient.prop <- proportion$patients.x/proportion$patients.y
proportion$dlt.prop <- proportion$dlt.number/proportion$dlt

for(i in 1:nrow(proportion)){
  proportion$rp2d.1[i] <- piTrue$rp2d.1[i]
  proportion$rp2d.2[i] <- piTrue$rp2d.2[i]
```

```
}

for (i in 1:nrow(proportion)){
  if(proportion$dose[i] == proportion$rp2d.2[i]){
    proportion$match1[i] <- max(proportion$patient.prop)+0.05
  } else{
    proportion$match1[i] <- NA
  }
}

write.csv(proportion, "3cohort_patient_proportion.csv",row.names = FALSE)

#### patient proportion plot
proportion$sim <- factor(proportion$sim,
                   levels = c(1:n.sim),
                   labels = c("Sim 1", "Sim 2", "Sim 3"))
proportion$cohort <- factor(proportion$cohort,
                      levels = c(1:ncohort),
                      labels = c("Group 1", "Group 2", "Group 3"))


#### patient proportion plot for 3 sample sizes

proportion.40 <- read.csv("3cohort_patient_proportion40.csv")
proportion.80 <- read.csv("3cohort_patient_proportion80.csv")
proportion.120 <- read.csv("3cohort_patient_proportion120.csv")

proportion.40$sample <- 1 # sample size: 40
proportion.80$sample <- 2 # sample size: 80
proportion.120$sample <- 3 # sample size: 120

proportion.40.80.120 <- rbind(proportion.40,proportion.80,proportion.120)

write.csv(proportion.40.80.120,"proportion_4080120.csv",row.names = FALSE)
proportion.40.80.120 <- read.csv("proportion_4080120.csv")

proportion.40.80.120$sim <- factor(proportion.40.80.120$sim,
                                   levels = c(1:n.sim),
                                   labels = c("Simulation 1", "Simulation
                                   2", "Simulation 3"))
proportion.40.80.120$cohort <- factor(proportion.40.80.120$cohort,
                                      levels = c(1:ncohort),
                                      labels = c("Group 1", "Group 2",
                                      "Group 3"))

proportion.40.80.120$sample <- factor(proportion.40.80.120$sample,
                                      levels = c(1:3),
                                      labels = c("Sample 1", "Sample 2",
                                      "Sample 3"))

ggplot(proportion.40.80.120, aes(dose,patient.prop, fill=cohort)) +
       theme_bw()+
       geom_bar(position="dodge",stat = "identity",width = 0.5) +
                facet_grid(sample~sim)+
```

```
        labs(list(title = "In-trial dose allocation", x = "Dose level", y =
                  "Proportion of patients",
        subtitle="Patient's probability to be assigned to each group = 0.2,
                  0.35, 0.45"))+
        scale_x_continuous(breaks=seq(0,nlevel,1)) +
        theme(legend.title=element_blank(),legend.text=element_text(size=16
              ,family="Times New Roman"),
        plot.title = element_text(hjust = 0.5,size=20,family="Times New
                                  Roman"),
        plot.subtitle = element_text(size=18,family = "Times New Roman"),
        axis.title = element_text(size=18,family="Times New Roman"),
        axis.text = element_text(size=16,family="Times New Roman"),
        strip.text=element_text(size=15.5,family="Times New Roman"))+
        geom_point(aes(dose,match1,colour=cohort),shape=25,size=1.5,positio
                   n=position_dodge(.5),stat = "identity")+
        ylim(0,0.5)

# 3--- calculate and plot number of DLTs/proportion for all dose levels in
3 simulations
## 3.1 --calculate number dlt for all dose levels and cohorts
dlt.trial <-
ddply(ok.trial,.(sim,tr,dose),summarise,dlt=length(dlt[dlt==1]))
dlt.trial.sum <- ddply(dlt.trial,.(sim,dose),summarise,dlt.5=median(dlt),
                       dlt.25=quantile(dlt,
                       0.25),dlt.75=quantile(dlt,0.75))

overall.dltprop <- ddply(dlt.trial,.(sim,tr),summarise,dlt=sum(dlt))
overall.dltprop$dltprop <- overall.dltprop$dlt/n.max
overall.dltprop.sum <-
ddply(overall.dltprop,.(sim),summarise,mean=mean(dltprop),
      median=median(dltprop),p25=quantile(dltprop,
      0.25),p75=quantile(dltprop,0.75))

## export datasets
write.csv(dlt.trial, "3cohort_dlt_trial.csv",row.names = FALSE)
write.csv(dlt.trial.sum, "3cohort_dlt_trial_sum.csv",row.names = FALSE)
write.csv(overall.dltprop.sum, "3cohort_dltprop_sum.csv",row.names =
          FALSE)


### 3.2--DLT plot
dlt.trial.sum$sim <- factor(dlt.trial.sum$sim,
                            levels = c(1:n.sim),
                            labels = c("Simulation 1", "Simulation 2",
                                       "Simulation 3"))

#### DLT plot for 3 sample sizes
dlt.trial.sum.40 <- read.csv("3cohort_40_dlt_trial_sum.csv")
dlt.trial.sum.80 <- read.csv("3cohort_80_dlt_trial_sum.csv")
dlt.trial.sum.120 <- read.csv("3cohort_120_dlt_trial_sum.csv")

dlt.trial.sum.40$sample <- 1 # sample size: 40
dlt.trial.sum.80$sample <- 2 # sample size: 80
dlt.trial.sum.120$sample <- 3 # sample size: 120
```

```
dlt.trial.sum.40.80.120 <-
rbind(dlt.trial.sum.40,dlt.trial.sum.80,dlt.trial.sum.120)

write.csv(dlt.trial.sum.40.80.120,"dltsum_4080120.csv",row.names = FALSE)

dlt.trial.sum.40.80.120 <- read.csv("dltsum_4080120.csv")

dlt.trial.sum.40.80.120$sim <- factor(dlt.trial.sum.40.80.120$sim,
                                      levels = c(1:n.sim),
                                      labels = c("Simulation 1",
                                                 "Simulation 2",
                                                 "Simulation 3"))
dlt.trial.sum.40.80.120$sample <- factor(dlt.trial.sum.40.80.120$sample,
                                         levels = c(1:3),
                                         labels = c("Sample 1", "Sample
                                                    2", "Sample 3"))

ggplot(dlt.trial.sum.40.80.120, aes(dose,dlt.5)) +
       theme_bw()+
       geom_errorbar(aes(x=dose, y=dlt.5, ymin=dlt.25,
                         ymax=dlt.75),position="dodge",stat = "identity")+
                         facet_grid(sample~sim)+
       labs(list(title = "DLTs at all dose levels among 100 trials", x =
            "Dose level", y = "Number of DLTs",
       subtitle="Patient's probability to be assigned to each group = 0.2,
                 0.35, 0.45"))+
       scale_x_continuous(breaks=seq(0,nlevel,1)) +
       scale_y_continuous(breaks=pretty(dlt.trial.sum.40.80.120$dlt.75,
                                        n = 8))+
       theme(legend.title=element_blank(),legend.text=element_text(family=
             "Times New Roman",size=16),
       plot.title = element_text(hjust = 0.5,family="Times New
                                 Roman",size=20),
       plot.subtitle = element_text(size=18,family = "Times New Roman"),
       axis.title = element_text(size=18,family="Times New Roman"),
       axis.text = element_text(size=16,family="Times New Roman"),
       strip.text=element_text(size=15.5,family="Times New Roman"))+
       geom_point(aes(dose,dlt.5),shape=20,size=3,stat = "identity")

#### overall DLT prop plot 3 sample sizes
overall.dltprop.sum.40 <- read.csv("3cohort_40_dltprop_sum.csv")
overall.dltprop.sum.80 <- read.csv("3cohort_80_dltprop_sum.csv")
overall.dltprop.sum.120 <- read.csv("3cohort_120_dltprop_sum.csv")

overall.dltprop.sum.40$sample <- 1 # sample size: 40
overall.dltprop.sum.80$sample <- 2 # sample size: 40
overall.dltprop.sum.120$sample <- 3 # sample size: 120
overall.dltprop.sum.40.80.120 <-
rbind(overall.dltprop.sum.40,overall.dltprop.sum.80,overall.dltprop.sum.12
      0)

write.csv(overall.dltprop.sum.40.80.120,"dltprop_sum.csv",
          row.names = FALSE)
```

```
overall.dltprop.sum.40.80.120 <- read.csv("dltprop_sum.csv")

overall.dltprop.sum.40.80.120$sample <-
factor(overall.dltprop.sum.40.80.120$sample,
       levels = c(1:3),
       labels = c("Sample 1", "Sample 2", "Sample 3"))

ggplot(overall.dltprop.sum.40.80.120, aes(sim,mean)) +
      theme_bw()+
      geom_errorbar(aes(x=sim, y=mean, ymin=p25,
      ymax=p75),position="dodge",stat = "identity")+
      facet_grid(.~sample)+
      labs(list(title = "Overall DLT proportion among 100 trials in each
           simulaiton", x = "Simulation", y = "Proportion of DLTs",
      subtitle="Patient's probability to be assigned to each group = 0.2,
                0.35, 0.45"))+
      scale_x_continuous(breaks=seq(0,n.sim,1)) +
      ylim(0,0.4)+
      theme(legend.title=element_blank(),legend.text=element_text(family=
            "Times New Roman",size=16),
      plot.title = element_text(hjust = 0.5,family="Times New
                                Roman",size=20),
      plot.subtitle = element_text(size=18,family = "Times New Roman"),
      axis.title = element_text(size=18,family="Times New Roman"),
      axis.text = element_text(size=16,family="Times New Roman"),
      strip.text=element_text(size=18,family="Times New Roman"))+
      geom_point(aes(x=sim,y=mean),colour="blue",shape=16,size=3)+
      geom_point(aes(x=sim,y=median),colour="brown",shape=18,size=3)

# 4---% Trial each dose is chosen as target
#---Largest dose with P(DLT)<=pi.target
ps <- ddply(ok.trial,.(sim,tr,dose,cohort),summarise,ps=mean(dlt))
rp2d.largest <- ddply(ps[ps$ps<=pi.target,],.(sim,tr,cohort),summarise,
                      rp2d.largest=max(dose))
colnames(rp2d.largest)[4] <- 'dose'
trial.target.largest <- ddply(rp2d.largest,.(sim,cohort,dose),summarise,
                              trial.number=length(unique(tr)))
for(i in 1:n.sim){
  for(j in 1:ncohort){
      trial.target.largest$target[trial.target.largest$sim==i &
      trial.target.largest$cohort==j] <-
      piTrue$rp2d.2[piTrue$sim==i&piTrue$cohort==j][1]
  }
}

for(i in 1:n.sim){
  trial.target.largest$totaltrial[trial.target.largest$sim==i] <-
  oktrial.number$trial.number[oktrial.number$sim==i]
}

trial.target.largest$trial.prop <-
trial.target.largest$trial.number/trial.target.largest$totaltrial

for(i in 1:nrow(trial.target.largest)){
```

```
  if(trial.target.largest$dose[i]==trial.target.largest$target[i]){
     trial.target.largest$match[i] <-
     max(trial.target.largest$trial.prop)+0.05}
  else {trial.target.largest$match[i] <- NA}
}

write.csv(trial.target.largest,
          "3cohort_recommend_largest120.csv",row.names = FALSE)

### plot Largest dose with P(DLT)<=pi.target for 3 sample sizes
rec.largest.40 <- read.csv("3cohort_recommend_largest40.csv")
rec.largest.80 <- read.csv("3cohort_recommend_largest80.csv")
rec.largest.120 <- read.csv("3cohort_recommend_largest120.csv")

rec.largest.40$sample <- 1 # sample size: 40
rec.largest.80$sample <- 2 # sample size: 40
rec.largest.120$sample <- 3 # sample size: 120

rec.largest.40.80.120 <-
rbind(rec.largest.40,rec.largest.80,rec.largest.120)

write.csv(rec.largest.40.80.120,"recommend_largest_sum.csv",
          row.names = FALSE)

#rec.largest.40.80.120 <- read.csv("recommend_largest_sum.csv")

rec.largest.40.80.120$sim <- factor(rec.largest.40.80.120$sim,
                                    levels = c(1:n.sim),
                                    labels = c("Simulation 1",
                                               "Simulation 2",
                                               "Simulation 3"))
rec.largest.40.80.120$sample <- factor(rec.largest.40.80.120$sample,
                                       levels = c(1:3),
                                       labels = c("Sample 1",
                                                  "Sample 2",
                                                  "Sample 3"))
rec.largest.40.80.120$cohort <- factor(rec.largest.40.80.120$cohort,
                                       levels = c(1:ncohort),
                                       labels = c("Group 1",
                                                  "Group 2",
                                                  "Group 3"))

ggplot(rec.largest.40.80.120, aes(dose,trial.prop, fill=cohort)) +
   theme_bw()+
   geom_bar(position="dodge",stat = "identity",width = 0.5) +
   facet_grid(sample~sim)+
   labs(list(title = "Final recommended dose distribution among 3 groups",
        subtitle="Patient's probability to be assigned to each group =
        0.2, 0.35, 0.45 \nLargest dose with P(DLT) <= target P(DLT)",x =
        "Dose level", y = "Proportion of trials")) +
   scale_x_continuous(breaks=seq(0,nlevel,1)) +
   scale_y_continuous(limits = c(0,0.6),breaks = seq(0, 0.6, by = 0.1))+
   theme(legend.title=element_blank(), plot.title = element_text(hjust =
         0.5))+
```

```
geom_point(aes(dose,match,colour=cohort),shape=25,size=1.5,
           position=position_dodge(.5),stat ="identity")+
theme(legend.title=element_blank(),legend.text=element_text(family="Tim
      es New Roman",size=11),
plot.title = element_text(hjust = 0.5,family="Times New
                          Roman",size=20),
plot.subtitle = element_text(size=16,family = "Times New Roman"),
axis.title = element_text(size=16,family="Times New Roman"),
axis.text = element_text(size=16,family="Times New Roman"),
strip.text=element_text(size=16,family="Times New Roman"))
```

BIBLIOGRAPHY

1. *NCI Dictionary of Cancer Terms*, https://www.cancer.gov/publications/dictionaries/cancer-term.
2. O'Quigley, J., Pepe, M., and Fisher, L. *Continual reassessment method: a practical design for phase 1 clinical trials in cancer.* Biometrics, 1990. **46**: p. 33-48.
3. Le Tourneau, C., Lee, J., and Siu, L. *Dose escalation methods in phase I cancer clinical trials.* J Natl Cancer Inst, 2009. **101**(10): p. 708–720.
4. Love, S.B., Brown,S., and Weir, C.J. *Enbracing model-based designs for dose-finding trials.* British Journal of Cancer, 2017. **117**: p. 332-339.
5. O'Quigley, J., Shen, L.Z., and Gamst, A. *Two-sample continual reassessment method.* Journal of Biopharmaceutical Statistics, 1999. **9**(1): p. 17-44.
6. O'Quigley, J. and Paoletti, X. *Continual reassessment method for ordered group.* Biometrics, 2003. **59**: p. 430-440.
7. Cheung, Y.K. and Chappell, R. *Sequential designs for phase I clinical trials with late-onset toxicities.* Biometrics, 2000. **56**: p. 1177-1182.
8. Salter, A., et al., *Two-group time-to-event continual reassessment method using likelihood estimation.* Contemporary Clinical Trials, 2015. **45**: p. 340-345.
9. O'Quigley, J., *Estimating the probabiity of toxicity at the recommended dose following a phase I clinical trial in cancer.* Biometrics, 1992. **48**: p. 853-862.
10. Korn, E.L., et al., *A comparison of two phase I trial designs.* Statistics in medicine, 1994. **13**: p. 1799-1806.
11. O'Quigley, J. and Shen, L.Z. *Continual reassessment method: a likelihood approach.* Biometrics, 1996. **52**: p. 673-684.
12. Normolle, D. and Lawrence, T. *Designing dose-escalation trials with late-onset toxicities using the time-to-event continual reassessment method.* Journal of Clinical Oncology, 2006. **24**(27): p. 4426-4433.
13. Chevret, S. *The continual reassessment method in cancer phase I clinical trials: a simulation study.* Statistics in medicine, 1993. **12**: p. 1093-1108.
14. Paoletti, X. and Kramar, A. *A comparison of model choices for the continual reassessment method in phase I cancer trials.* Statistics in medicine, 2009(28): p. 3012-3028.