FOR MASTERS THESES: A SIMULATED EXPLORATION OF ALTERNATIVES TO STANDARD EXPANSION COHORTS IN PHASE 1 TRIALS

By

Justin Leon Jacobs

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Biostatistics

December 17, 2022

Nashville, Tennessee

Approved:

Tatsuki Koyama, Ph.D.

Fei Ye, Ph.D.

ACKNOWLEDGMENTS

I am extremely grateful to my thesis advisor and committee chair, Tatsuki Koyama, Ph.D., for his direction, constant feedback, and assistance throughout this process. I am also thankful for the assistance I received from the second member of my thesis committee, Fei Ye, Ph.D., for helping review this project and for her flexibility in working around my work schedule. And I could not have completed this whole graduate school process without the encouragement and advice from Robert Greevy, Ph.D., Director of Graduate Studies for the Biostatistics department throughout my time at Vanderbilt, who believed in me even when I did not believe in myself.

Thanks should also go to my cohort members and classmates. From late-night study sessions to venting over shared meals, they provided support like only someone going through the same process could. I would also like to acknowledge the teams I worked with during my internship with and employment at Sarah Cannon Research Institute, Genospace, and HCA Healthcare, who were always encouraging and accommodating around prioritizing completing my degree, and who provided valuable insight into real-world applications of the material I learned from the program and gave me opportunities to put what I had learned into practice.

Lastly, I would like to thank my wife, Klhee Jacobs, NP. I never would have even attempted graduate school without her support and encouragement throughout the entire experience.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction and Background

Over the years, clinical trials have changed and evolved to better understand the effects of different treatments on the human body. Each advance has allowed investigators to address the scientific question of interest, from the introduction of placebo controls to the use of personalized medicine to determine which treatment would be best based on a patient's genetic makeup. One such design consideration is using multiple phases in a clinical trial to determine how best to assess the treatment in question at different stages of understanding its effects on patients. In particular, phase 1 trials are used for first-in-human studies to assess toxicity and determine the maximum tolerated dose (MTD) for subsequent phases to use as a guide when determining efficacy. Generally, a phase 1 trial will attempt to determine the MTD by exposing as few patients as possible to escalating dose levels and comparing the number of patients at each dose level with dose-limiting toxicities (DLT) and a prespecified limit. Using the MTD, researchers will ascertain a recommended phase 2 dose (RP2D), and in phase 2 a larger group of patients will be treated with the RP2D established in phase 1 to assess efficacy.

Phase 1 trials are sometimes organized into two main components, a dose-escalation component, and a dose-expansion component. In the dose-escalation component, patients are treated in cohorts at multiple escalating dose levels, and multiple methods of determining the MTD have been proposed and studied, with varying properties, strengths, and weaknesses. In a dose-expansion component, additional cohorts of patients are treated at the MTD established during escalation and depending on the number of DLTs an RP2D is recommended either at or near the MTD.

1.1 3+3 Method

The most conventional method for dose escalation, the 3+3 method, involves enrolling a cohort of three patients at the lowest available dose level and then determining how many DLTs occurred at that dose level. Based on the number of DLTs, the study would escalate to the next dose level, enroll three more patients at the current dose level, or de-escalate and either enroll three more patients at the previous dose level or set the previous dose level as the MTD.

If at a given dose level, zero of the three patients enrolled had DLTs, the next three would be enrolled at the next dose level. If one patient of the three experiences a DLT, then a second cohort of three patients will be enrolled at the same dose level, and if two patients experience a DLT between both cohorts, then that dose and all higher ones would be discontinued, and if there is still a lower dose with only three patients treated then a second set of three will be enrolled at that dose. If after the second cohort of three patients at a given dose, the total number of DLTs is still one, then that dose is selected as the MTD. If the total is two DLTS, then the dose is either treated as the MTD, or discontinued, depending on rules established at the start of the study. And if the total after six patients is three or more DLTs, the dose is discontinued. The study ends once all doses have no more patients to enroll, either from having six patients enrolled, having been discontinued due to too many DLTs at that dose or a lower one, or once a dose has been chosen as the MTD.

This method is especially conservative, and previous simulation studies have shown that it does not consistently choose the correct MTD and tends to underestimate the true MTD. (He et al. 2006 and Iasonos et al. 2008) We will explore this further in establishing our methodology.

1.2 Continual Reassessment Method

Due to the conservative nature of the 3+3 method and its tendency to treat multiple patients at dose levels below the MTD when the true MTD is in one of the higher proposed dose levels, O'Quigley asserted that it is not particularly effective in situations like phase 1 oncology trials, where concerns about providing inferior treatment and the benefits of providing appropriate treatment quickly can outweigh the risks of escalating dose level due to the nature of the disease. (O'Quigley et al. 1990) For this reason, O'Quigley proposed a continual reassessment method (CRM), a Bayesian method that updates its prior estimations for the dose toxicity curve with information collected from each cohort of patients enrolled in dose escalation before recommending a dose based on the new posterior mean estimate of the MTD for the next cohort. In this way, if the model predicts that the MTD will be one of the higher dose levels based on information from the earlier dose levels, it can skip dose levels unlikely to be the true MTD and arrive at a recommended MTD close to the truth more quickly. Modifications on CRM include rules that prevent skipping dose levels or stopping the trial if the model predicts that the lowest dose level has an unacceptable toxicity rate, which may be more appropriate for trials where the risk of overdosing is greater than the risk of undertreating.

CRM works by assuming a parametric model for the actual dose toxicity curve. The original formulation of CRM utilized a one-parameter model, where a function of dose level and some arbitrary parameter a is set up to provide the expected probability of response for patients enrolled on a given dose level. Let x_j represent the dose level, and Y_i represent a binary random variable $(0, 1)$ where 1 denotes DLT for patient i entered into the trial. If the target toxicity rate is θ, then there would be some $x^\star$ that represents the dose level with $\mathrm{E}(\mathrm{Y_i}) = \theta$. $\mathrm{E}(Y_i)$ can be represented by some function $\psi(x_j, a)$, therefore, there must exist a $\psi(x^*, a_0)$ where a_0 represents the true natural state where dose $x^\star$ gives a response probability of θ. Let $\Omega_i = \{y_1, \dots, y_{i-1}\}$ and$f(a, \Omega_i)$ be a nonnegative function that summarizes all available information about parameter a_0. Giving all possible values of a as $\mathrm{A} = (0, \infty)$, we find that

$$\int_0^\infty f(a, \Omega_i) da = 1 \quad (i = 1, \dots, n).$$

Using Bayes theorem, it is possible to obtain $f(a, \Omega_{i+1})$ from $f(a, \Omega_i)$, allowing estimation of $\theta_{(i,j)}$ at dose level j given information about the first $i - 1$ patients. If patient i is enrolled on dose level $x(i)$, this allows for updating our prior for a with information taken from the i^{th} patient enrolled.

Let

$$\theta_{(i,j)} = \int_0^\infty \psi(x_j, a)\mathrm{f}(a, \Omega_i)\mathrm{da} \quad (j = 1, \dots, k)$$

$$\theta'_{(i,j)} = \psi[x_j, \mu(i)] \quad \mu(i) = \int_0^\infty a\, f(a, \Omega_i) da$$

$$\phi(x(i), y_i, a) = \psi^{y_i}(x(i), a)[1 - \psi(x(i), a)]^{(1-y_i)}$$

$$\therefore f(a, \Omega_{i+1}) \propto \phi(x(i), y_i, a) f(a, \Omega_i)$$

Prediction of the correct dose level for the $i + 1^{th}$ patient enrolled by updating a_0 based on the responses from patients 1 through i can be repeated for each patient enrolled to recommend the dose level of the next patient.

This can then be used with some function to minimize the distance between $\theta_{(i,j)}$ and θ, $\theta'_{(i,j)}$and θ, or $x_j$and $\psi^{-1}(\mu(i), \theta)$, whichever criterion is selected. Once all patients are enrolled in this way, the recommended dose is the dose that gives the smallest distance between those variables in the criterion selected. Let $g(a) = f(a, \Omega_1)$ represent the prior distribution of a_0 before experimentation, or our prior knowledge of the relationship between dose level and toxicity. This function can take multiple forms, such as gamma, log-normal, or exponential, and is even sometimes modified to have a two-parameter form.

A commonly used functional model for CRM is the power model, with α as the unknown parameter. This uses an exponential function to describe the prior relationship between dose level and toxicity. Let $\mathrm{p_j}$ denote a true DLT probability at dose level j, and let ϕ represent the target DLT probability.

$$\mathrm{p_j} = \mathrm{a_j^{\exp(\alpha)}}, \text{ for } \mathrm{j} = 1, \dots, \mathrm{j}$$

The prior guesses for DLT probability at a given dose, $0 < \mathrm{a_1} < \cdots < \mathrm{a_j} < 1$, represent the initial estimates for the dose-toxicity curve. After each cohort is treated, these estimates are updated based on the accumulated DLT data for all dose levels, and the method assigns the next cohort to an "optimal" dose, the dose with a posterior mean estimated DLT probability closest to ϕ. (Zhou et al. 2018).

1.3 Modified Toxicity Probability Interval

While the 3+3 method is known to have poor performance, it is still widely used for its ease of understanding and implementation. Model-based designs like CRM may have better performance but can be substantially more complex. For this reason, the modified toxicity probability interval (mTPI) design was introduced to provide a method that is simpler to implement than CRM with better performance than the standard 3+3. The original TPI method relied on the specification of parameters in its model and the method is potentially vulnerable to misspecification of the model. Yuan et al. proposed the mTPI to provide a method that does not rely on the specification of any parameters beforehand. (Yuan et al. 2010)

mTPI makes decisions on when to escalate or de-escalate based on rules created using unit probability mass (UPM) calculations. For a given interval on the real line, UPM can be defined as the ratio of the probability of something falling within that interval and the length of the interval. For a given dose, a beta-binomial model is used to find the probability of DLT at that dose falling within an interval, and the intervals used are those for underdosing, proper dosing, and overdosing. The proper dosing interval for the original TPI method was initially provided by doses considered by physicians as close enough to the true target toxicity level that they would feel comfortable selecting these doses as the estimated MTD, and decisions on escalation or de-escalation required calculations based on prespecified parameters. Under the mTPI, after each cohort of patients on a given dose is observed, the toxicity probability for a given dose is updated based on new information, and the UPM values of each dose being below the lower bound of the target interval, within the target interval, and above the upper bound of the target interval are updated based on the total information available in the trial. When the UPM for underdosing is the maximum of the three, the study escalates to the next highest dose. Likewise, when the UPM for overdosing is the maximum, the study de-escalates to the next lowest dose, with the UPM for the proper dose being the maximum indicating that the study should continue to the next cohort at the current dose. This method utilizes a prespecified sample size, and at the exhaustion of that sample size, the MTD is selected based on which estimate of DLT probability for each dose level is closest to the target rate.

Yuan et. Al (2010), showed that using decision rules based on a beta-binomial model for determining when to escalate, de-escalate, or stay, a modified model can be used that is invariant to the specification of the target interval. The beta-binomial model follows a Bayesian paradigm, starting with a given beta prior distribution giving the distribution of a parameter p. This parameter p is then used as the parameter of a binomial model that represents the

distribution of the data. This combined posterior model follows a beta-binomial distribution, which is a special type of beta distribution where the beta distribution is a conjugate prior to the binomial data.

Stopping rules based on excessive toxicity are usually implemented. In this, the posterior probability of that dose level having a probability of toxicity given the data that exceeds the target level is calculated, and if it exceeds the excessive toxicity threshold, typically 0.95, then the dose level is excluded, or alternatively expressed as

$$P(p_i > p_T|\text{data}) > 0.95$$

If this occurs, that dose level and those above it are eliminated, and the study continues using the remaining dose levels. If all dose levels are eliminated due to excessive toxicity, then the trial is terminated, and no dose levels are recommended as MTD.

1.4 Bayesian Optimal Interval Design

Similarly to mTPI, Yuan et al. introduced the Bayesian optimal interval design (BOIN) in an effort to provide a more rigorous method than 3+3 that is simpler to implement or understand than CRM. (Yuan et al. 2016) This method takes the basic framework of the 3+3 and adapts a more flexible design, based on comparing the DLT probability at a given dose to prespecified boundaries. Let $\widehat{p_{cur}}$ represent the observed DLT rate at a given dose. Escalation and de-escalation occur if $\widehat{p_{cur}}$ falls outside the boundaries, with escalation when it is below the lower bound and de-escalation when it is above the upper bound. These boundaries are calculated such that if a given dose has a DLT occurrence rate at the upper or lower boundary, the target rate is the edge of a set confidence interval from the boundary so that whenever $\widehat{p_{cur}}$ falls within the boundaries it is reasonable that the true DLT probability at that dose is near the target rate. These boundaries and the escalation/de-escalation rules are set before the trial begins and can be modified during the study to account for deviations from planned cohort sizes.

Yan et al. proposed a similar excessive toxicity stopping rule for BOIN as that included in mTPI, where a dose is excluded once at least 3 patients have been enrolled at that dose if the posterior probability that the probability of toxicity exceeds the target level is greater than a threshold, also generally set at 0.95, or alternatively expressed as when $n \geq 3$

$$P(p_i > p_T|\text{data}) > 0.95$$

This stopping rule also eliminates a dose level and any higher levels, similarly to the one utilized in mTPI, and terminates the trial without recommending an MTD if all dose levels are eliminated due to excessive toxicity.

1.5 Expansion

Even with the more consistently accurate methods, the dose escalation stage does not always choose the true MTD in simulation studies, but it does often select a dose level that is within one level from the true MTD. Iasonos suggested using the expansion cohort to also reassess and/or confirm the MTD by dosing expansion cohort patients at the indicated MTD as well as one level above or below it. (Iasonos et al. 2016) Dose expansion cohorts can be incorporated into a study for a variety of reasons, including reassessing toxicity.

Therefore, in this paper we will explore simulation studies across multiple toxicity scenarios with each escalation dose-finding method, and apply one of 5 different expansion strategies to determine which has a better chance of selecting the correct dose, and how many toxicities occur in each one. These are discussed more in the methodology section.

CHAPTER 2

Methodology

2.1 3+3 Revisited

As was discussed earlier, the 3+3 method tends to be conservative in selecting a recommended dose for expansion or phase 2. To explore this, simulations were performed using true dose toxicity curves that represent possible curves that could be found in new drugs. In these simulations of the standard 3+3 method, using dose toxicity curves as shown in table 2.1.1, the correct dose was selected between 18% and 37% of the time, depending on the dose toxicity curve used. These simulations also showed that the dose level below the true correct dose was consistently selected as the recommended dose more often than the true correct dose.

Toxicity Dose Curves

	Dose 1	Dose 2	Dose 3	Dose 4	Dose 5
Scenario 1	0.0417	0.0833	0.1667	0.3333	0.6667
Scenario 2	0.1111	0.2222	0.3333	0.4444	0.5555
Scenario 3	0.3333	0.4714	0.5773	0.6667	0.7454
Scenario 4	0.0133	0.0533	0.1200	0.2133	0.3333
Scenario 5	0.0500	0.0500	0.3333	0.6000	0.6000

Table 2.1.1: Toxicity Dose Curves for Simulation

One might consider whether removing some of the safeguards to the 3+3 method could improve how often it selects the correct dose. The 3+3 method was modified to not stop for any reason until 30 patients had been enrolled, removing the stopping rules for too many toxicities at a given dose, 6 patients enrolled at a given dose, or escalation/de-escalation of dose at the outer limits of the available dose levels. This modified 3+3 method gave slightly improved results, ranging from 31% to 41% correct recommendations of dose level, but again this method consistently selected the dose level below the true correct dose more often than the actual true correct dose. Tables 2.1.2 and 2.1.3 show the exact recommendation values at each level for the standard and modified 3+3 simulations.

Standard 3+3 Recommended Dose Percentage

	No Dose	Dose 1	Dose 2	Dose 3	Dose 4	Dose 5
Dose Curve 1	0.0174	0.0681	0.2051	0.4003	0.2956	0.0135
Dose Curve 2	0.1147	0.3003	0.3361	0.1856	0.0557	0.0076
Dose Curve 3	0.5665	0.3461	0.0795	0.0070	0.0009	0.0000
Dose Curve 4	0.0022	0.0302	0.1250	0.2709	0.3289	0.2428
Dose Curve 5	0.0255	0.0249	0.5459	0.3713	0.0300	0.0024

Table 2.1.2: Standard 3+3 Recommended Dose Percentage

Modified 3+3 Recommended Dose Percentage

	No Dose	Dose 1	Dose 2	Dose 3	Dose 4	Dose 5
Dose Curve 1	0.0000	0.0021	0.0800	0.5521	0.3589	0.0069
Dose Curve 2	0.0035	0.1508	0.4223	0.3120	0.1032	0.0082
Dose Curve 3	0.4802	0.4093	0.0957	0.0138	0.0010	0.0000
Dose Curve 4	0.0000	0.0001	0.0136	0.1927	0.4797	0.3139
Dose Curve 5	0.0000	0.0016	0.5638	0.4088	0.0237	0.0021

Table 2.1.3: Modified 3+3 Recommended Dose Percentage

2.2 Simulation Methodology

For this simulation study, therefore, we will examine how these proposed modified expansion cohort methods proposed by Iasonos do in terms of toxicities and whether they provide insight into which dose is truly the MTD. The study will look at multiple methods of dose escalation, including 3+3, BOIN, MTPI, and CRM. It will then utilize five expansion cohort strategies to determine which of them offer improved determination of true MTD and what the change in toxicity is compared to simply extending the escalation cohort by the same number of patients. These strategies are as follows:

1. Standard Expansion – put the entire expansion cohort onto the recommended dose from escalation
2. Extended Escalation – simply increase the maximum escalation patient number by the same number as the expansion cohort(s) would be in total
3. Expansion and Dose Minus One – split the expansion cohort between the recommended dose and the dose level below it, putting the whole cohort on the recommended dose if it is the lowest

4. Expansion and Dose Plus One – split the expansion cohort between the recommended dose and the dose level above it, putting the whole cohort on the recommended dose if it is the highest
5. Expansion and Dose Plus and Minus One – split the expansion cohort between the recommended dose and the dose levels above and below it, with patients placed on the recommended dose if there is not a dose level higher or lower than the recommended one.

The simulation will be performed in RStudio under R version 4.1.2, using programs for each escalation method from the 'escalation' R package. There will be 30 patients enrolled for the escalation cohort, and 18 for the expansion cohort. The same five dose toxicity curves used for comparing escalation methods will also be used for this (Table 2.1.1). Each combination of dose toxicity curve, escalation method, and expansion method will be repeated over 10,000 simulations. Each scenario will be compared for the total number of toxicities, toxicities at each dose, and percent of toxicities at each dose, in escalation, expansion, and in total. Additionally, the accuracy of dose recommendation for each scenario will be compared after escalation and after expansion, and the change in accuracy provided by each expansion method from its corresponding escalation simulation will be evaluated. For trials where there is no recommended dose, meaning the lowest dose level was deemed too toxic, no expansion will be performed. The escalation methods have built-in techniques for determining the recommended dose. For the expansion portion of the study, after completing the escalation the additional expansion results will be simulated following each expansion method and starting with the recommended escalation dose for each simulation. Since the 3+3 method's code in the 'escalation' package is not easily adaptable to results that don't follow the 3+3 method's dosing structure, the recommended dose after expansion will be calculated using a simple comparison for which dose level has a simulated toxicity percentage closest to the target of 33% without going over after combining escalation and expansion results for each simulation. This method is not mathematically dissimilar to the normal dose selection process in 3+3, which essentially rules out dose levels once they have more toxicities than 1/3 of the total number of subjects on each dose level.

For the other three methods, the escalation and expansion results will be combined for each simulation and these results will be put back through their respective selection methods again, resulting in a recommended dose level after expansion using the same algorithm as the escalation recommended dose level.

In the escalation portion of the simulations, prespecified samples sizes and/or dose stopping rules at n=30 will be used for all escalation methods, adjusted to n=48 for the "extended escalation" expansion method. The target dose

toxicity level will be $p_T = 0.33$ in all escalation methods. For CRM, an uninformative prior dose toxicity curve will be used, that being a sequence from 0.1 to 0.5 increasing by 0.1 each level. and the adjusted method that does not skip dose levels will be utilized. The mTPI simulations will use interval bounds of $p_T \pm 0.05$, and an excessive toxicity threshold of 0.95. For BOIN, the same excessive toxicity threshold was used.

CHAPTER 3

Results

3.1 Escalation Simulation Results

After performing the simulations, comparisons across the escalation methods show that the 3+3 method consistently underestimates the true MTD across the different dose curves as previously mentinoed, selecting the dose below the true MTD as the recommended dose level. Even the modified 3+3 method consistently underestimates the true MTD though with slightly better performance than the standard 3+3 method. BOIN, CRM, and mTPI all consistently recommend the true MTD, with varying degrees of accuracy across different dose curves. CRM tends to be the most accurate, followed by mTPI and then BOIN. Figure 3.1.1 shows the simulated dose selection probabilities for each method and each dose toxicity curve.

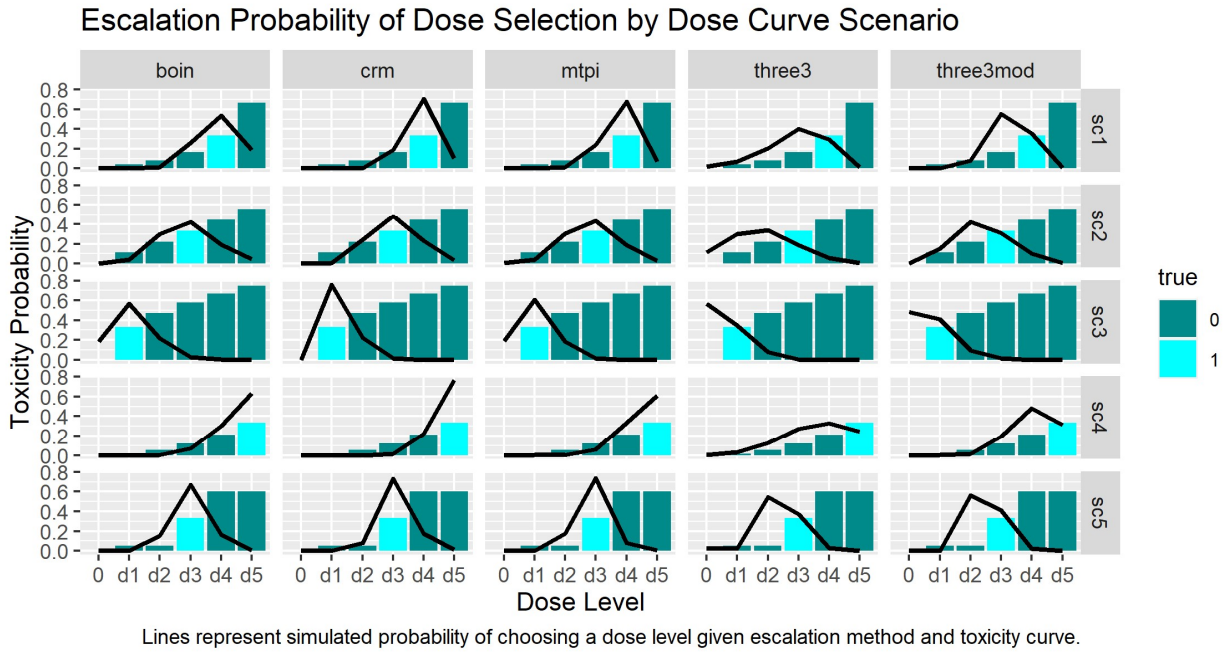

Figure 3.1.1: Escalation Probability of Dose Selection by Dose Curve across Escalation Methods

Figure 3.1.2 shows the percent of simulated standard escalation results that correctly recommended the true MTD for each curve, and their corresponding percent of simulated toxicities among subjects across all dose levels. In terms

of accuracy, 3+3 is the least accurate, with CRM, mTPI, and BOIN being significantly more accurate, though which is most accurate fluctuates depending on the dose toxicity curve. CRM tends to be more accurate than the other two. For all scenarios except Scenario 3, 3+3 had the lowest toxicity percent, with BOIN and mTPI generally having slightly higher toxicity rates, and CRM having the highest. This is all consistent with the literature regarding escalation methods.

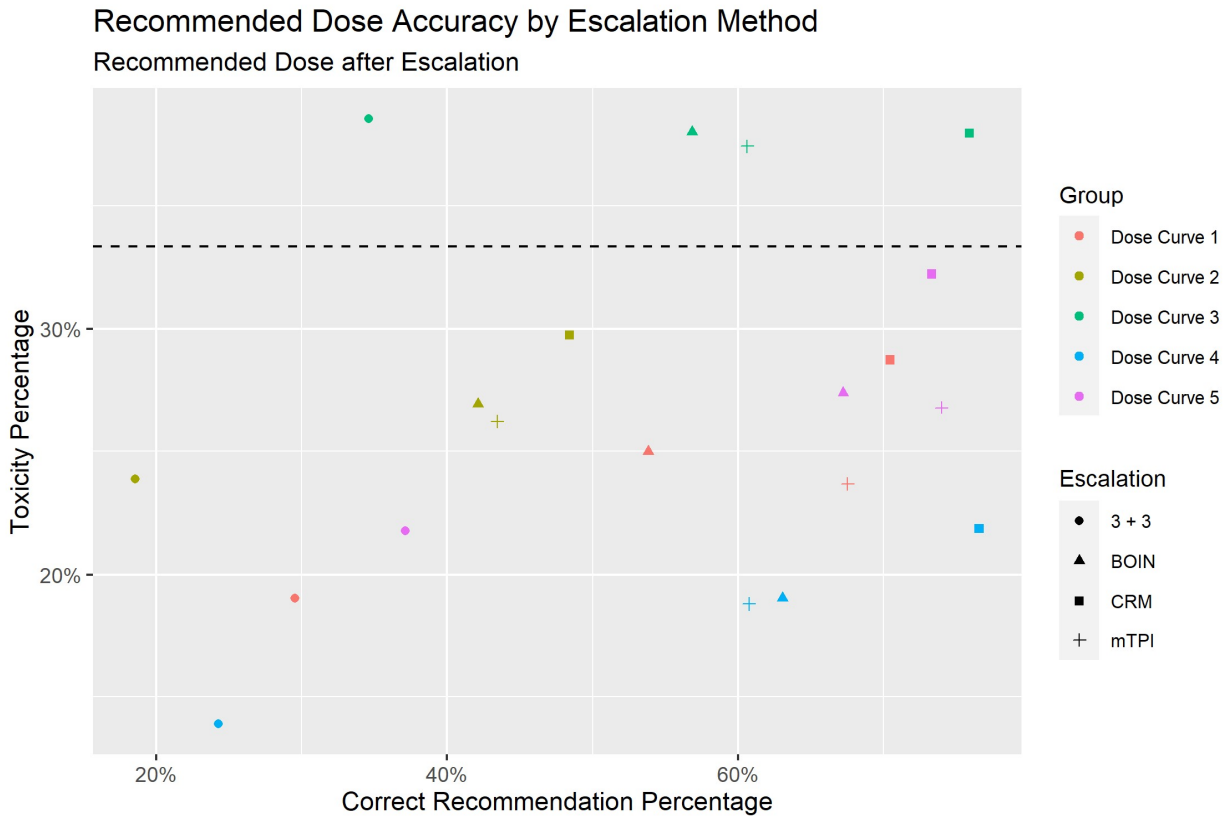

Figure 3.1.2: Recommended Dose Accuracy by Escalation Method

3.2 Expansion Simulation Results

After the expansion portion was included with the escalation portion of the simulation, similar results were shown as far as the accuracy of the escalation methods, with 3+3 being the least accurate and CRM tending to be the most. Of more interest is how the different expansion methods affected the accuracy of the final recommended dose. In particular, the "extended escalation" method outperformed the "regular expansion" method with 3+3 and CRM without substantially increasing toxicity, and the reverse is true for BOIN and mTPI, again without substantially

increasing toxicity. For 3+3, "extended escalation" gave a correct recommendation between approximately 19% and 48% of the time, while "regular expansion" ranged from around 14% to 28% (see Figure 3.2.1). With BOIN, "extended escalation" correctly recommended the true MTD in 51% to 72% of simulations, while "regular expansion" ranged from 54% to 73% correct recommendations. CRM had ranges from 69% to 85% and 69% to 83% for "extended escalation" vs. "regular expansion", and mTPI 48% to 76% and 51% to 77% for the same.

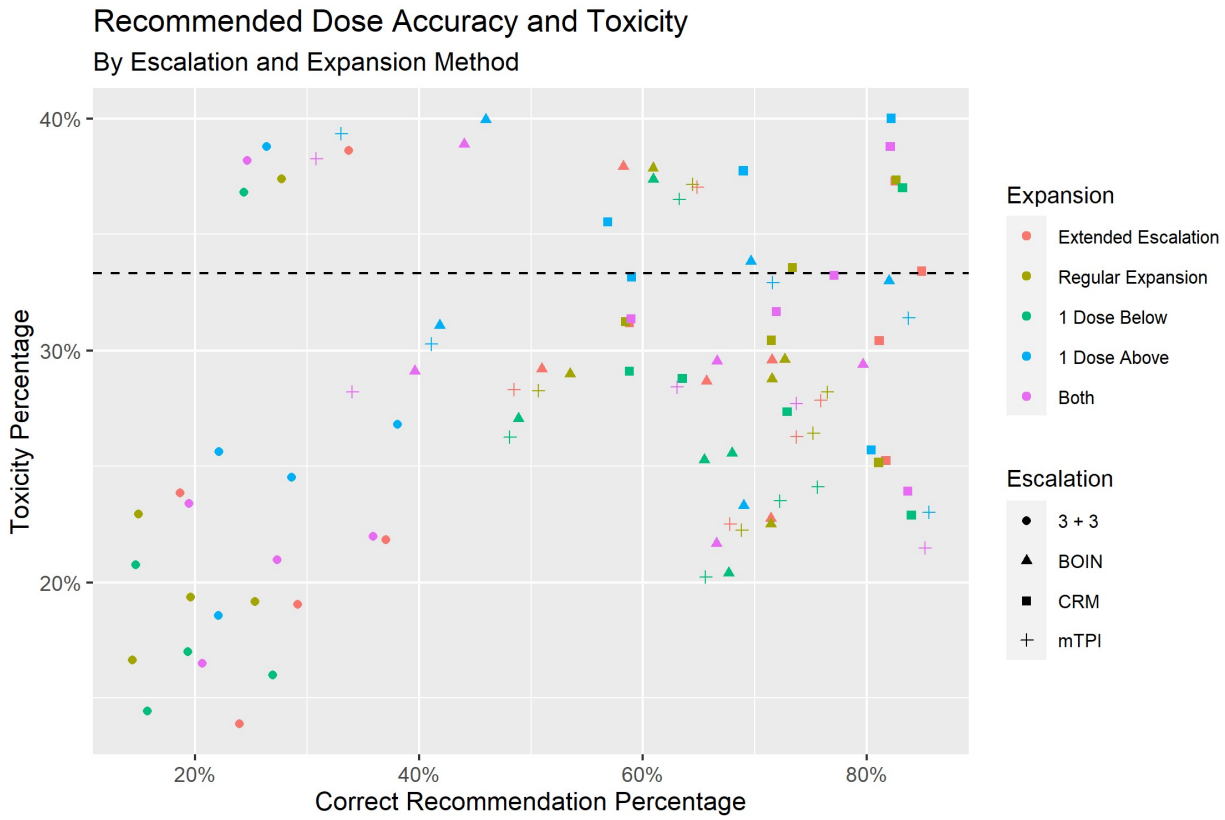

Figure 3.2.1: Recommended Dose Accuracy and Toxicity by Escalation and Expansion Method Combinations

There were not many consistent trends in accuracy by escalation and expansion method combinations for the proposed experimental expansion methods, either by escalation type or by true dose toxicity curve, as shown in Figure 3.2.2. The "one dose above" method did tend to outperform or equal the accuracy of the standard methods with 3+3 escalation, except in the case of dose curve scenario 3, where the true MTD is the lowest dose. No general trends existed for it in the other escalation methods. "One dose below" reduced accuracy in most escalation and dose curve combinations, though not for scenario 3 with escalation methods other than 3+3. The expansion method using the dose levels to either side of the recommended dose did not have any consistent trends outside of performing poorly with scenario 3.

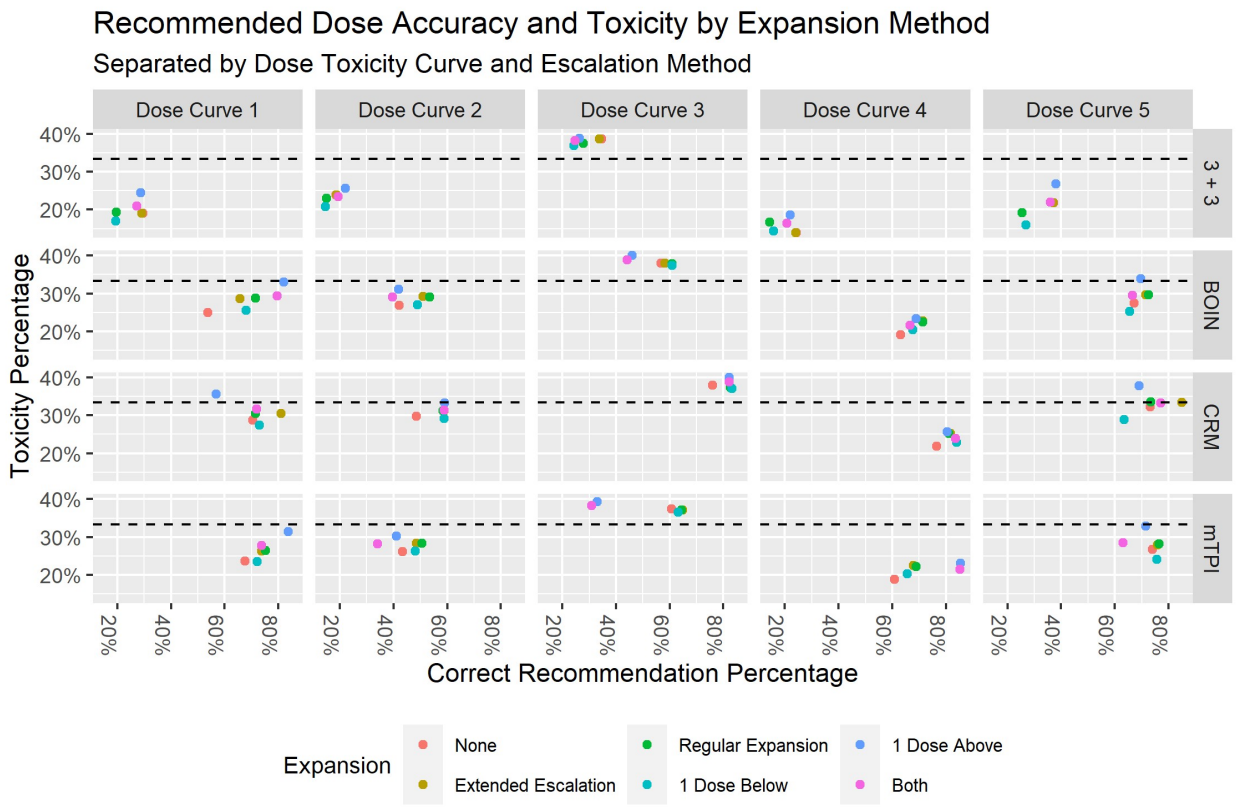

Figure 3.2.2: Recommended Dose Accuracy and Toxicity by Expansion, Separated by Dose Curve and Escalation Method

As expected, the "one dose below" and "one dose above" expansion methods resulted in fewer and more toxicities, respectively, when compared to the "no expansion", "extended escalation", and "regular expansion" methods across the board. The expansion method using "one dose on either side" of the recommended escalation dose did not usually result in substantially more toxicities, remaining close to the standard methods.

3.3 Change in Recommended Dose from Escalation to Expansion

The different expansion methods did result in changes in recommended MTD in each scenario as shown in Table 3.3.1, and Table 3.3.2 shows how many of these changes resulted in the expansion recommendation correcting an incorrect recommendation in escalation, while Table 3.3.3 shows the converse.

Percent of Changes in Recommended Dose from Escalation to Expansion

Group	Escalation	Regular Expansion	1 Dose Below	1 Dose Above	Both
Dose Curve 1	3 + 3	0.1322	0.1467	0.4173	0.3586
Dose Curve 1	BOIN	0.6557	0.6655	0.3449	0.3295
Dose Curve 1	CRM	0.4361	0.4429	0.5423	0.4311
Dose Curve 1	mTPI	0.1832	0.1634	0.3304	0.3690
Dose Curve 2	3 + 3	0.1547	0.1426	0.3715	0.3112
Dose Curve 2	BOIN	0.5710	0.5869	0.3654	0.3694
Dose Curve 2	CRM	0.3816	0.4037	0.3883	0.3920
Dose Curve 2	mTPI	0.1535	0.1331	0.5969	0.6298
Dose Curve 3	3 + 3	0.2028	0.1977	0.1942	0.1906
Dose Curve 3	BOIN	0.2955	0.2749	0.1356	0.1634
Dose Curve 3	CRM	0.1790	0.1825	0.1754	0.1820
Dose Curve 3	mTPI	0.0881	0.0708	0.3311	0.3803
Dose Curve 4	3 + 3	0.1239	0.1248	0.4781	0.4078
Dose Curve 4	BOIN	0.3920	0.3866	0.3321	0.3123
Dose Curve 4	CRM	0.2045	0.2115	0.1981	0.1941
Dose Curve 4	mTPI	0.1351	0.1102	0.3387	0.3266
Dose Curve 5	3 + 3	0.1730	0.1604	0.3051	0.2616
Dose Curve 5	BOIN	0.5979	0.6411	0.1307	0.1710
Dose Curve 5	CRM	0.4322	0.5268	0.4832	0.3712
Dose Curve 5	mTPI	0.1099	0.0984	0.2633	0.3282

Table 3.3.1: Percent of Changes in Recommended Dose in Escalation vs. Expansion

Percent of Correct Changes in Recommended Dose from Escalation to Expansion

Group	Escalation	Regular Expansion	1 Dose Below	1 Dose Above	Both
Dose Curve 1	3 + 3	0.0193	0.0137	0.1388	0.1092
Dose Curve 1	BOIN	0.8882	0.8502	0.6401	0.6107
Dose Curve 1	CRM	0.6746	0.7607	0.6013	0.7152
Dose Curve 1	mTPI	0.3964	0.3230	0.7036	0.5919
Dose Curve 2	3 + 3	0.0500	0.0327	0.1324	0.0944
Dose Curve 2	BOIN	0.5071	0.4731	0.1914	0.1774
Dose Curve 2	CRM	0.4261	0.4458	0.4353	0.4415
Dose Curve 2	mTPI	0.1697	0.1274	0.3814	0.3093
Dose Curve 3	3 + 3	0.1009	0.0625	0.0834	0.0565
Dose Curve 3	BOIN	0.3552	0.3245	0.0000	0.0000
Dose Curve 3	CRM	0.4764	0.4952	0.4619	0.4684
Dose Curve 3	mTPI	0.1305	0.0925	0.0000	0.0000
Dose Curve 4	3 + 3	0.0000	0.0000	0.0951	0.0747
Dose Curve 4	BOIN	0.5517	0.4839	0.4787	0.4222
Dose Curve 4	CRM	0.5058	0.5791	0.4795	0.5395
Dose Curve 4	mTPI	0.2392	0.1752	0.6880	0.6758
Dose Curve 5	3 + 3	0.0472	0.0318	0.2285	0.1759
Dose Curve 5	BOIN	0.9639	0.9377	0.2035	0.2041
Dose Curve 5	CRM	0.6867	0.7656	0.7730	0.7327
Dose Curve 5	mTPI	0.2600	0.2207	0.4131	0.3560

Table 3.3.2: Percent of Correct Changes in Recommended Dose from Escalation to Expansion

Percent of Incorrect Changes in Recommended Dose from Escalation to Expansion

Group	Escalation	Regular Expansion	1 Dose Below	1 Dose Above	Both
Dose Curve 1	3 + 3	0.3791	0.3724	0.3537	0.3239
Dose Curve 1	BOIN	0.4316	0.4717	0.0243	0.0501
Dose Curve 1	CRM	0.2686	0.2841	0.4451	0.2791
Dose Curve 1	mTPI	0.0730	0.0808	0.0927	0.1900
Dose Curve 2	3 + 3	0.4207	0.3415	0.3935	0.3751
Dose Curve 2	BOIN	0.4272	0.4902	0.2718	0.3059
Dose Curve 2	CRM	0.2496	0.2614	0.2546	0.2542
Dose Curve 2	mTPI	0.0743	0.0812	0.5528	0.6205
Dose Curve 3	3 + 3	0.3771	0.4007	0.3838	0.3947
Dose Curve 3	BOIN	0.1968	0.1852	0.1980	0.2264
Dose Curve 3	CRM	0.0611	0.0570	0.0594	0.0632
Dose Curve 3	mTPI	0.0374	0.0409	0.4619	0.5052
Dose Curve 4	3 + 3	0.4136	0.3508	0.3980	0.3671
Dose Curve 4	BOIN	0.1916	0.2088	0.1816	0.1941
Dose Curve 4	CRM	0.1008	0.0800	0.1010	0.0775
Dose Curve 4	mTPI	0.0304	0.0375	0.0378	0.0380
Dose Curve 5	3 + 3	0.3863	0.3432	0.3621	0.3277
Dose Curve 5	BOIN	0.3961	0.4838	0.0578	0.0977
Dose Curve 5	CRM	0.2498	0.4115	0.3407	0.2153
Dose Curve 5	mTPI	0.0471	0.0481	0.1745	0.2731

Table 3.3.3: Percent of Incorrect Changes in Recommended Dose from Escalation to Expansion

3.4 Net Change in Recommended Dose Accuracy

Figure 3.4.1 shows the net correct change percentage of each expansion method, or the percent of times that an incorrect escalation recommendation was corrected by expansion minus the percent of times that a correct escalation recommendation was changed to incorrect by expansion. This result represents how often an expansion method was shown to improve accuracy in the simulation assuming the escalation recommendation was just as likely to be correct as not correct. For 3+3, none of the expansion methods resulted in a net positive change in accuracy, though the “one dose above” and “one dose on either side” methods tended to be the most accurate. For BOIN, “regular expansion” always had a net positive result on change in accuracy and generally had the highest net change in accuracy. The other expansion methods fluctuated between positive and negative net changes. With CRM, each expansion method had a net positive change in accuracy, though which was most beneficial was not consistent. Only “regular expansion” and “one dose below” resulted in net positive accuracy changes for mTPI across every dose level, while “one dose above” and “one dose on either side” fluctuated from -50% to +60% net accuracy change.

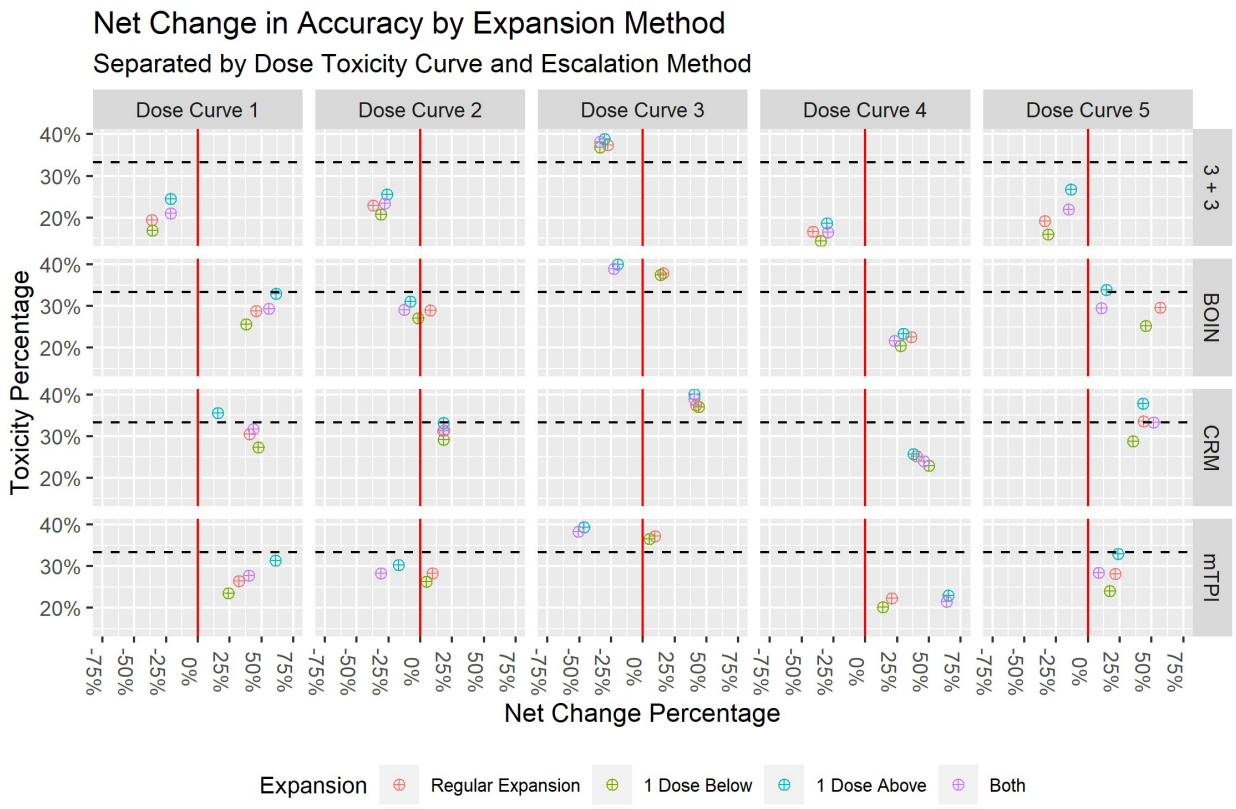

Figure 3.4.1: Net Change in Recommended Dose Accuracy by Expansion Method, Separated by Dose Curve and Escalation Method

CHAPTER 4

Conclusions

4.1 Discussion of Results

Based on these results, it is not immediately apparent that any of the proposed modified expansion methods can reliably provide improvements on accuracy in finding the MTD across all scenarios. Especially when utilizing BOIN or mTPI, the standard expansion method used performs similarly well or better than the modified methods across most dose curve scenarios evaluated. The "one dose above" and "one dose on either side" methods could have some value in use with the 3+3 escalation method compared to standard expansion, but if accurate MTD determination is the ultimate goal then based on these results any expansion at all does not seem to be the most effective method, as it results in worse accuracy than just the standard 3+3 escalation. There are multiple other reasons to include an expansion cohort in a study, but these results do not indicate that it improves accuracy in determining the MTD.

An interesting point is that with CRM, across all dose curve scenarios, the "one dose on either side" method outperformed standard expansion in ultimate accuracy and likelihood of improving accuracy compared to escalation, while not substantially increasing toxicity. These results indicate a correlation between using the "one dose on either side" expansion method and better accuracy in determining the true MTD when compared to using either standard expansion cohorts using CRM for escalation, or escalation only using CRM. This holds true across all dose level evaluated in the simulation and could benefit from future exploration to determine if this method continues to consistently outperform standard expansion alone.

4.2 Takeaway Recommendations

Some recommendations can be made following these results, especially in situations where the true dose toxicity curve is approximately known. If the accuracy of the recommended dose is the most important criterion for methodology selection, then CRM with the "one dose on either side" expansion method is the most effective across various dose curve options. Even though toxicity is the main concern in phase 1 trials, the poor accuracy of 3+3 and the relatively smaller increase in toxicity for the much larger increase in accuracy means any of the other escalation methods would be preferable when used in conjunction with either the standard expansion methods or the modified methods proposed in this paper. mTPI is the escalation method that achieves the best balance between accuracy and

toxicity most effectively in conjunction with standard expansion across all dose curve scenarios and performs well with the “one dose on either side” method in dose curve scenarios where the true MTD is likely to be among the higher doses tested. None of the modified expansion methods perform well compared to regular expansion when the true MTD is the lowest dose, as in dose curve three results.

Since many expansion cohorts also evaluate for efficacy, incorporating efficacy measures into the simulation could provide additional insight into the value of each method beyond mere MTD determination accuracy. Investigation into whether the alternate expansion methods improve efficacy evaluation could give justification for them even when their improvements on MTD determination are not significant.

References

He, W., Liu, J., Binkowitz, B., and Quan, H. (2006). *Stat. Med.*, 25(12):2027-2042

Iasonos, A., Wilton, A., Riedel, E., Seshan, V., and Spriggs, D. (2008) *Clinical Trials.* 5: 465-477

Iasonos, A., and O'Quigley, J. (2015). *Nat Rev Clin Oncol.,* 12(11): 626-628

Iasonos, A., and O'Quigley, J. (2016). *Stat Biopharm Res.,* 8(2): 161-170

O'Quigley, J., Pepe, M., and Fisher, L. (1990). *Biometrics*, 46(1): 33-48

Yan, F., Zhang, L., Zhou, Y., Pan, H., Liu, S., and Yuan, Y. (2020) *J. Stat. Soft.,* 94: 1-32

Yuan, J., Ping, L., Yisheng, L., and Bekele, B. (2010). *Clin Trials*., 7(6): 653-663

Yuan, Y., Hess, K., Hilsenbeck., S, and Gilbert, M. (2016). *Clin Cancer Res.,* 22(17): 4291-4301

Zhou, H., Yuan, Y., and Nie, L. (2018). *Clin Cancer Res.* 24(18): 4357-4364