EARLY AND MIDDLE PHASE CLINICAL TRIAL DESIGNS FOR GROUPS

Connor John Celum

Clovis, California

Bachelor of Arts, University of California at Berkeley, 2020

Master of Science, University of Virginia, 2023

A Dissertation submitted to the Graduate Faculty

of the University of Virginia in Candidacy for the Degree of

Doctor of Philosophy

Department of Statistics

University of Virginia

May 2025

Dr. Mark Conaway

Dr. Karen Kafadar

Dr. Shan Yu

Dr. Ken Ono

Early and Middle Phase Clinical Trial Designs for Groups

Connor John Celum

(ABSTRACT)

This dissertation presents phase I and phase II trial designs for groups. In particular, groups completely or partially ordered by dose sensitivity are considered. Groups are completely or partially ordered when the groups can be completely or partially ordered by the probability of an adverse event for any given dose, respectively. A pair of phase I clinical trials are presented in this dissertation: the Quasi-CRM Shift method and the Group Averaged Bayesian Optimal Interval Design (GAB). The Quasi-CRM Shift method is the first design for partially ordered groups considering ordinal toxicity, allowing clinicians to control for the frequency and severity of adverse events during dose selection. GAB is the first model-assisted design for partially ordered groups, a class of designs marked by their simplicity. Simulation studies show that GAB performs as well as more complex model-based designs, demonstrating GAB provides clinicians with a simple design that performs well. Large sample properties show allocation under GAB tends to correct doses. Both the Quasi-CRM Shift and GAB demonstrate that utilizing the group ordering leads to increased accuracy in dose allocation during the trial and dose selection at the end of the trial. In addition to the phase I designs, a pair of phase II designs are presented in this dissertation. These designs consider a trial with multiple doses and two ordered groups. At the end of the trial, doses are determined to be acceptable or unacceptable. The first design is a single-stage design and maximizes power subject to a type I error constraint. The second design is a two-stage design with cutoffs determining if a dose

continues onto the second stage for a group. This design minimizes the number of unacceptable doses that continue onto the second stage while meeting power and type I error requirements. Through the development of these designs, clinicians are provided with multi-dose designs in the group framework, extending dose exploration and optimization into phase II.

Acknowledgments

First and foremost, I dedicate my dissertation to Jaskiran, my fiancée, for putting up with monologues on statistics and my dissertation over dinner, time she regrets she will never get back. Next, I thank my family John, Pamela, Carissa, Mathew, Parminder, and Gurvinder for their unwavering support and encouragement and my father's interest in my research papers, learning acronyms like DLT and MTD which this dissertation will later discuss. I thank Parminder and Gurvinder for assisting Jaskiran and I with the cross-country move from Fresno to Charlottesville. I thank my friends Isabella, Eleanor, Aydin, Hyunjae, Nathaniel, Rose, and Tyler for their friendship over these four years which sustained me socially. I thank my advisors Professor Mark Conaway and Dr. Karen Kafadar for their guidance and support throughout my time at UVA. Finally, I thank my committee members, Dr. Mark Conaway, Dr. Karen Karafar, Dr. Shan Yu, and Dr. Ken Ono for their time and helpful suggestions which have greatly improved the quality of this dissertation.

Contents

List of Figures **xi**

List of Tables **xiii**

1 Introduction **1**

1.1 Introduction 1

1.2 Phase I Trials 2

1.2.1 CRM 4

1.2.2 BOIN 7

1.3 Phase II Trials 11

1.3.1 MERIT 11

1.3.2 Simon's Two-Stage Design 13

1.4 Completely and Partially Ordered Groups 14

1.5 Outline 16

2 Quasi-CRM Shift **17**

2.1 Introduction 17

2.2 Groups and Ordinal Toxicity 19

2.3 Proposed Model and Allocation 22

2.3.1 ET MTD 25

2.3.2 DLT MTD 27

2.3.3 Allocation Rules 28

2.4 Simulations 29

2.4.1 Scenarios 29

2.4.2 Reversals 30

2.4.3 PCS and PCA 31

2.4.4 Sensitivity Analysis and Overdose Control 32

2.4.5 General Group Sizes and Orders 35

2.5 Conclusion 36

3 A Model-Assisted Design for Partially or Completely Ordered Groups 38

3.1 Introduction 38

3.2 GAB for Two Groups 39

3.2.1 BOIN in Parallel 39

3.2.2 Notation 40

3.2.3 Allocation Run-in 40

3.2.4 Allocation 41

3.2.5 Dose Selection 41

3.3 General GAB 43

3.3.1 Notation and Group Bundles 43

3.3.2 Allocation Run-in 45

3.3.3 Allocation 46

3.3.4 Dose Selection 48

3.4 Asymptotic Properties 51

3.5 Practical Considerations 53

3.6 Simulations 55

3.6.1 Generating Family of Curves 55

3.6.2 Results 57

3.6.3 Sensitivity Analysis 62

3.6.4 Convergence Rates 62

3.7 Discussion 67

4 Multi-Dose Phase Two Trial Designs for Two Ordered Groups 68

4.1 Introduction 68

4.2 Notation 69

4.3 Null and Alternative Regions 71

4.4 Traditional Trial 71

4.4.1 Finding the Optimal Cutoffs for the Traditional Design 73

4.5 Two-Stage Design 75

4.6 Simulations 78

4.7 Sensitivity Analysis 82

4.8 Discussion 87

5 Discussion and Future Areas of Research **89**

5.1 Summary of Methods Proposed 89

5.2 Future Research 91

Bibliography **94**

Appendices **102**

Appendix A Appendix for Chapter 2 **103**

Appendix B Appendix for Chapter 3 **105**

B.1 | Proofs 105

B.1.1 | Lemma 1 106

B.1.2 | Theorem 2 108

B.2 | MTD Configuration 112

B.3 | Generating Centers 114

B.4 | GAB Illustration 116

B.4.1 Two Group Illustrative Example 117

B.4.2 Three Group Illustrative Example 123

B.4.3 Dose Elimination Boundaries 131

B.5 | Sensitivity Analysis 131

Appendix C Appendix for Chapter 4 **140**

C.1 | Finding all Configurations 140

C.2 | Configurations 142

C.3 | Considering two groups with non-overlapping doses 142

List of Figures

1.1 Phases of a clinical trial. (MD Anderson Cancer Center n.d.) 3

1.2 3+3 flowchart as provided in G. Kim et al. 2018 4

1.3 Empirical model illustrated with a dotted line showing the DLT target, $\theta = 0.3$ 6

1.4 Decision boundaries from BOIN, as obtained from trialdesign. 10

2.1 DLT MTD and ET MTD Plot 21

3.1 GAB Allocation Procedure 42

3.2 GAB Run-In 46

3.3 General GAB Allocation Procedure 49

3.4 K-Means Centers for the Four-Dose Curves, providing "Representative" Curves. "Representatives" are provided for the two-group case and the three different three-group orders. 58

3.5 K-Means Centers for the Six-Dose Curves, providing "Representative" Curves. "Representatives" are provided for the two-group case and the three different three-group orders. 59

3.6 Percent Correct Selection (PCS) and probability (as a percentage) that the last patient is allocated to the MTD by Group and Method. . . . 64

3.7 Plot of Percent Correct Selection (PCS) and probability (as a percentage) that the last patient is allocated to the MTD by Group and Method. Plot is zoomed into sample sizes between 0 and 200. 65

3.8 Performance differences (between GAB and P-BOIN) for Percent Correct Selection (PCS) and probability (as a percentage) that the last patient is allocated to the MTD. The differences are provided for each group. 66

A.1 Four-dose ET and DLT curves, with the ET scores/DLT probabilities by dose level. 103

A.2 Six-dose ET and DLT curves, with the ET scores/DLT probabilities by dose level. 104

List of Tables

2.1 ET skeleton shift example for shift ($\Delta^S_{3,2} = 2, \Delta^S_{3,1} = 1$). 25

2.2 DLT skeleton shift example for shift ($\Delta^T_{3,2} = 2, \Delta^T_{3,1} = 1$). 25

2.3 Allocation rules for the proposed method with $d_g^{\max}$ denoting the highest observed dose in group g and $x_{n_g+1,g}$ denoting the MTD estimate for group g. 28

2.4 Reversal percentages for parallel Quasi-CRM trials for the four-dose scenarios. Standard Errors do not exceed 1.6%. 31

2.5 Reversal percentages for parallel Quasi-CRM trials for the six-dose scenarios. Standard Errors do not exceed 1.6%. 31

2.6 Percentage correct selection (PCS) and percentage correct allocation (PCA) by number of doses, number of patients, and method. Standard Errors do not 0.04%. 32

2.7 Average DLT rate and average (normalized) ET score for patients during the trial. Note that the normalized ET target is 0.313 and the DLT target is 0.33. 33

2.8 Overdose Statistics: comparing percentage of patients allocated to doses above the MTD and percentage of times a method selects a dose above the MTD at the conclusion of the trial. 33

2.9 Percentage correct selection (PCS) and percentage correct allocation (PCA) comparisons for simulations where ET Scores are randomly generated or group probabilities are unequal. Making comparisons to results with equal group rates and ET scores (0,.5,1,1.5). 34

2.10 Percentage correct selection (PCS) and percentage correct allocation (PCA) for the proposed method when varying SD and Halfwidth . . 35

3.1 Orders from Horton, O'Quigley, and M. R. Conaway 2019 converted to the bundle/group format in the "Bundles/Groups" column. 45

3.2 Dose elimination boundaries based on number of patients treated and number of DLTs. We require at least 3 patients to have been observed, use the prior Beta(0.5,0.5), and cutoff $\lambda = 0.975$. 54

3.3 Three Group Orders 55

3.4 Number of MTD configurations and curves by number of groups and group ordering. A MTD configuration is a set of possible group-specific MTDs under the group ordering. 56

3.5 Percentage correct selection (PCS) and Accuracy Index (AI) by the number of groups, scenario, number of doses, number of patients, and method. 61

3.6 Percentage correct allocation (PCA) and percentage of allocations to doses with DLT probability in the interval (λ_e, λ_d) (Int) by the number of groups, scenario, number of doses, number of patients, and method. 61

3.7 DLT Probabilities by group and dose for three different scenarios. . . 63

4.1 Cutoffs From MERIT (Yang et al. 2024)) 78

4.2 Type 1 Error, Power, cutoffs From MERIT for two doses 79

4.3 Cutoffs, Type 1 Error, and Power from the proposed traditional design 79

4.4 Cutoffs From MERIT with 4 doses (trialdesign) 80

4.5 Type 1 Error and Power From MERIT four doses 80

4.6 Type 1 Error, Power, cutoffs From MERIT four doses 80

4.7 Cutoffs and stats for two-stage design when maximum global type I error is $\alpha^\star = 0.2$ and minimum global power is $\beta^\star = 0.4$ 83

4.8 Cutoffs and stats for two-stage design when maximum global type I error is $\alpha^\star = 0.2$ and minimum global power is $\beta^\star = 0.5$ 83

4.9 Cutoffs and stats for two-stage design when maximum global type I error is $\alpha^\star = 0.2$ and minimum global power is $\beta^\star = 0.6$ 83

4.10 Cutoffs and stats for two-stage design when maximum global type I error is $\alpha^\star = 0.2$ and minimum global power is $\beta^\star = 0.7$ 84

4.11 Cutoffs and stats for two-stage design when maximum global type I error is $\alpha^\star = 0.3$ and minimum global power is $\beta^\star = 0.4$ 84

4.12 Cutoffs and stats for two-stage design when maximum global type I error is $\alpha^\star = 0.3$ and minimum global power is $\beta^\star = 0.5$ 84

4.13 Cutoffs and stats for two-stage design when maximum global type I error is $\alpha^\star = 0.3$ and minimum global power is $\beta^\star = 0.6$ 85

4.14 Cutoffs and stats for two-stage design when maximum global type I error is $\alpha^\star = 0.3$ and minimum global power is $\beta^\star = 0.7$ 85

4.15 Comparison with Different Interim Sample Sizes 85

4.16 Minimizing the number of patients at unacceptable doses 86

4.17 Optimal intern size n, by $N, \alpha^\star$, and $\beta^\star$ 86

4.18 Sensitivity Analysis from varying true odds ratio while using cutoffs when odds ratio is 2.0 and $\alpha^\star = 0.2$ 87

B.1 MTD configurations with two ordered groups and four doses 113

B.2 MTD configurations with three ordered groups and four doses under Order 1 (complete ordering) 113

B.3 MTD configurations with three ordered groups and four doses under Order 2 (Group 1 least sensitive) 114

B.4 MTD configurations with three ordered groups and four doses under Order 3 (Group 3 most sensitive) 115

B.5 Number of Observations by Group and Dose Level 121

B.6 Number of DLTs by Group and Dose Level 121

B.7 Smoothed DLT Proportions by Group and Dose Level 121

B.8 Weights for Bivariate Isotonic Regression by Group and Dose Level . 122

B.9 DLT Probability Estimates from Bivariate Isotonic Regression 122

B.10 Number of Observations by Group and Dose Level 128

B.11 Number of DLTs by Group and Dose Level 128

B.12 Smoothed DLT Proportions by Group and Dose Level 128

B.13 Number of Observations by Group and Dose Level 129

B.14 DLT Estimates by Group and Dose Level for Model 1 130

B.15 DLT Estimates by Group and Dose Level for Model 2 130

B.16 Dose elimination boundaries based on number of patients treated and number of DLTS. For these boundaries we require that more than 3 patients have been observed. Additionally, we use the prior Beta(0.5,0.5) and the cutoff $\lambda = 0.975$. 131

B.17 Selection Results from Sensitivity Analysis: Percentage correct selection (PCS) and Accuracy Index (AI) by the number of groups (G), scenario, number of doses (D), cohort size, number of cohorts, Group Membership Probabilities (Group Probs), and method. 134

B.18 Allocation Results from Sensitivity Analysis: Percentage correct allocation (PCA) and percentage of allocations to doses with DLT probability in the interval (λ_e, λ_d) (Int) by the number of groups (G), scenario, number of doses (D), cohort size, number of cohorts, Group Membership Probabilities (Group Probs), and method. 136

B.19 Early Termination Results from Sensitivity Analysis: Percentage of times each Group is removed from the trial early due to trial termination rules by the number of groups (G), scenario, number of doses (D), cohort size, number of cohorts, Group Membership Probabilities (Group Probs), and method. 139

B.20 Selection Results from Prior Sensitivity Analysis: Percentage correct selection (PCS) and Accuracy Index (AI) by the number of groups (G), scenario, number of doses (D), number of patients, and choice of Prior/choice of cutoff. 139

B.21 Allocation Results from Prior Sensitivity Analysis: Percentage correct allocation (PCA) and percentage of allocations to doses with DLT probability in the interval (λ_e, λ_d) (Int) by the number of groups (G), scenario, number of doses (D), number of patients, and choice of Prior/choice of cutoff. 139

C.1 Null Configurations 144

C.2 Alternative Configurations 145

C.3 Trial with non-overlapping doses 146

List of Abbreviations

AI: Accuracy Index

BOIN: Bayesian Optimal Interval Design

CRM: Continual Reassessment Method

DLT: Dose Limiting Toxicity

ET: Equivalent Toxicity

GAB-E: Group Averaged BOIN with Dose Elimination Rules

GAB: Group Averaged BOIN

Int: Percentage of allocations to doses with DLT probabilities in the BOIN interval

MTD: Maximum Tolerated Dose

OR-CRM: Order Restricted CRM

P-BOIN-E: Parallel BOIN Trials which implement Dose Elimination

P-BOIN: Parallel BOIN Trials

PCA: Percent of Correct Allocation

PCS: Percent of Correct Selection

Chapter 1

Introduction

1.1 Introduction

Oncology clinical trials typically have four phases. Phase I trials have the goals of safety and dose-finding for subsequent phases. Phase II trials measure drug efficacy and side effects. Phase III trials compare the new drug to the standard treatment. Phase IV trials proceed if a drug is approved and monitors long-term effects. Phase I and phase II designs will be the focus of this dissertation.

To provide an example of the first three phases of a clinical trial, consider trials for the treatment Ado-trastuzumab emtansine, also known as T-DM1. T-DM1 was an antibody-drug conjugate approved in 2014 for patients with "HER2-positive, unresectable, locally advanced, or metastatic breast cancer" who had previously undergone treatment with trastuzumab and a taxane (Dhillon 2014). First, a Phase I trial was conducted to find the dose for subsequent phases. This trial considered 0.3, 0.6, 1.2, 2.4, 3.6, and 4.8 mg/kg doses of T-DM1 every three weeks. The 3.6 mg/kg every three weeks dose was selected for subsequent phases (Krop, Beeram, et al. 2010). A Phase II study measured the efficacy of a 3.6 mg/kg dose of T-DMI every three weeks. The endpoint objective response rate (ORR) was used, where ORR is the percentage of patients with either a tumor shrinkage (known as a partial response) or disappearance of all signs of cancer (known as a complete response) (National Can-

cer Institute 2024). Based on an observed overall response rate of 34.5% (95% CI of 26.1% to 43.9%), the drug showed sufficient efficacy to warrant a phase III trial (Krop, P. LoRusso, et al. 2012). The T-DM1 efficacy was evaluated in the phase III TH3RESA study (Dhillon 2014). In the TH3RESA study, T-DM1 significantly prolonged the median progression-free survival (PFS) (by 2.15 months) compared to the treatment of the physician's choice, resulting in the approval of T-DM1 (Krop, S.-B. Kim, et al. 2014). Here, PFS is a time-to-event endpoint, looking at the time to disease progression.

In this chapter, we will introduce the phase I and phase II trial designs foundational to the designs presented in this dissertation. The phase I designs presented will be the Continual Reassessment Method (CRM) (O'Quigley, Pepe, and Fisher 1990) and the Bayesian Optimal Interval Design (BOIN) (Liu and Y. Yuan 2015). The phase II designs presented are the Multiple-dose Randomized Phase II Trial (MERIT) (Yang et al. 2024) and Simon's Two-Stage Design (Simon 1989).

After discussing phase I and phase II designs, trials with partially or completely ordered groups will be discussed. Examples of trials with ordered groups will be provided, along with an extensive literature review of designs for ordered groups. Finally, we will introduce the goals and outline of this dissertation.

1.2 Phase I Trials

Phase I trials have the goal of finding the highest allowable dose, known as the maximally tolerated dose (MTD). The MTD is the dose where the probability of a sufficiently adverse event, known as a dose-limiting toxicity (DLT), is closest to a target rate, called the target toxic rate, denoted by θ. In this paradigm, we assume

toxicity and efficacy have a monotonic relationship with the dose levels. The target toxicity rate can be understood as the toxicity rate deemed acceptable to achieve more efficacy.

Figure 1.1: Phases of a clinical trial. (MD Anderson Cancer Center n.d.)

Phase I trial designs are adaptive, meaning the responses from the previous patients are used for dose allocation to the next patient. At the end of the trial, a design estimates the MTD. From this, the two goals of a clinical trial design are accurate within-trial dose allocation and accurate end-of-trial dose selection. A brief description of the 3+3 design will be provided to overview the development of phase I trial designs.

The 3+3 design is the most commonly used design due to its simplicity. According to Paoletti, Ezzalfani, and Le Tourneau 2015, more than 95% of phase I trials use the 3+3 design. The 3+3 design, as provided in Figure 1.2, enrolls patients in cohorts of three and de-escalates, escalates, or stays at the same dose level based on the number of DLTs. Numerous papers, including M. R. Conaway and Petroni 2019 and Chiuzan and Dehbi 2024, highlighted the inferior dose selection and dose allocation of the 3+3 compared to the CRM (O'Quigley, Pepe, and Fisher 1990) and BOIN (Liu and Y. Yuan 2015) designs. This section will provide a literature review of the CRM and

BOIN as these designs form the foundations of the phase I designs presented in this dissertation.

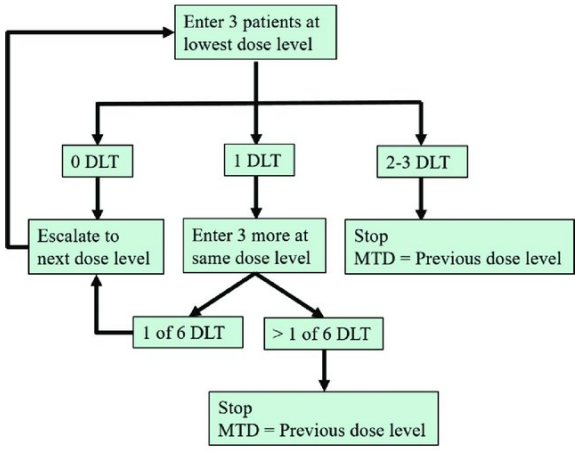


Figure 1.2: 3+3 flowchart as provided in G. Kim et al. 2018

1.2.1 CRM

Before describing the CRM and BOIN, notation will be provided, allowing these methods to be detailed. Let $d_1 < d_2 < \cdots < d_K$ denote the doses present in the trial. Let π_k denote the DLT probability for dose d_k. By the monotonic dose-toxicity relationship, we have $\pi_1 \leq \pi_2 \leq \cdots \leq \pi_K$. The DLT is the dose with DLT probability closest to the toxicity target θ. That is, the dose $d_{k'}$, where $k' = \text{argmin}_k|\pi_k - \theta|$.

The CRM is an effective model-based design ubiquitous in phase I design literature. This design estimates a dose-toxicity curve using prior beliefs on the dose-toxicity relationship and observed responses to update these beliefs. There are numerous variations of the CRM, including the one-stage Bayesian CRM, as presented in O'Quigley, Pepe, and Fisher 1990, and the two-stage Maximum Likelihood CRM, as presented in O'Quigley and Shen 1996. This dissertation considers the one-stage Bayesian CRM.

The CRM begins by assuming a model of the form

$$\pi_k = \psi(d_k, a),$$

where a is a parameter continually updated by the data. In the Bayesian CRM, we assign a prior distribution to a and update a using the posterior mean of a, which will be denoted as $\hat{a}$. As in Y. K. Cheung 2011, the normal prior, $a \sim N(0, \sigma^2)$, is considered. Details on the choice of prior variance, σ^2, can be explored in Y. K. Cheung 2011. The commonly used empiric dose-toxicity model will be utilized, so that

$$\psi(d_k, a) = p_k^{\exp(a)},$$

where the values $0 < p_1 < p_2 < \cdots < p_K < 1$ are prior values called the skeleton values. The skeleton values can be understood as initial guesses at the dose-toxicity relationship which are continually updated by the parameter a. Figure 1.3 illustrates how the empirical model estimates the dose-toxicity curve based on initial skeleton values and the parameter a. The R package "dfcrm" (K. Cheung 2019) provides skeleton values that perform well. O'Quigley and Zohar 2010 finds the CRM robust under reasonably spaced prior skeleton values and notes that two sets of skeletons with equally spaced skeleton values give equivalent models.

Next, skeleton values and observed responses are used to guide allocation. Suppose that n patients have been observed. For the k^{th} patient, let x_k denote the dose given and y_k indicate if a DLT was observed. Letting $\mathcal{D}$ denote the observed data, the

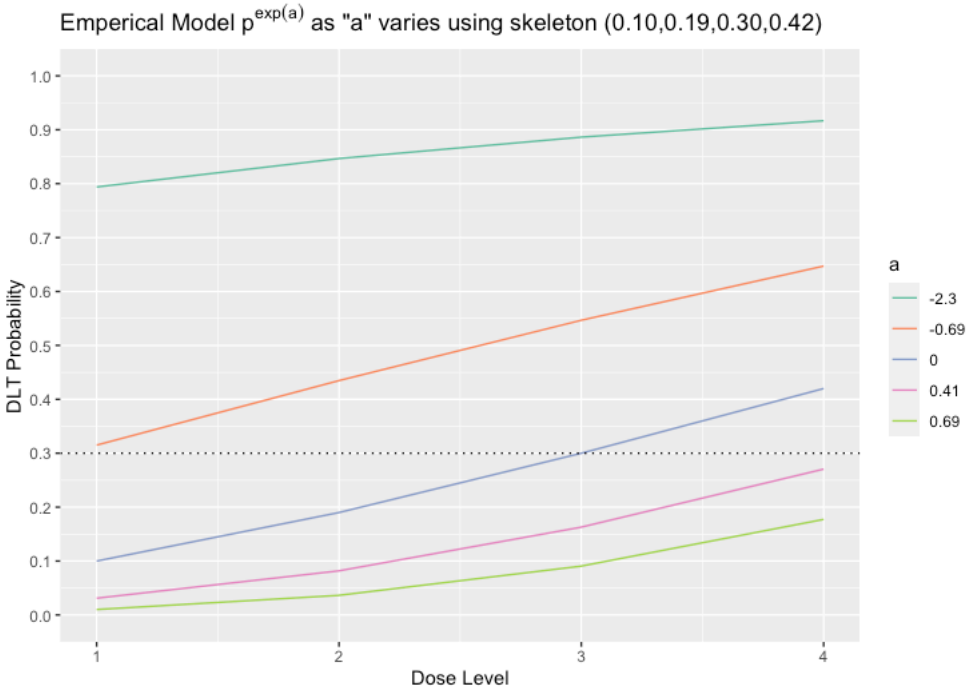


Figure 1.3: Empirical model illustrated with a dotted line showing the DLT target, $\theta = 0.3$

likelihood under parameter a is

$$L(a|\mathcal{D}) = \prod_{i=1}^{n} \psi(x_k, a)^{y_k}(1 - \psi(x_k, a))^{1-y_k}.$$

Let $g(a)$ denote the prior distribution of a. Then, the posterior mean of a is obtained as

$$\hat{a} = \frac{\int_{-\infty}^{\infty} a g(a) L(a|\mathcal{D}) da}{\int_{-\infty}^{\infty} g(a) L(a|\mathcal{D}) da}.$$

The next patient is assigned to the dose with the estimated DLT probability closest to the toxicity target θ. That is, the dose given to patient $n + 1$ is

$$x_{n+1} = \text{argmin}_{d_k} |\psi(d_k, \hat{a}) - \theta|.$$

Once all enrolled patients have been observed, the MTD is estimated as the dose that would have been assigned next.

1.2.2 BOIN

In this subsection, we describe BOIN, a widely used model-assisted design. This description of BOIN is adapted from Celum and M. Conaway 2024, and Liu and Y. Yuan 2015. BOIN is said to be model-assisted since the decision boundaries can be enumerated before the trial and does not require reestimation of parameters, as is the case in model-based designs, such as the CRM. The relative simplicity of the model-assisted designs is attractive to clinicians seeking designs with predefined decision boundaries. Statisticians debate whether model-assisted or model-based designs are superior, a debate this dissertation will not engage. For a comprehensive overview of model-assisted designs and a comparison between model-assisted and model-based designs favoring model-assisted designs, see Y. Yuan, Lee, and Hilsenbeck 2019. Conversely, for a comparison more favorable to model-based designs, see Horton, Wages, and M. R. Conaway 2017.

Different notation will be used to describe BOIN allocation. Let j denote the current dose level. Then, at dose level j, let y_j denote the number of DLTs, n_j denote the number of observations, and $\hat{\pi}_j = \frac{y_j}{n_j}$ denote the DLT rate. Using an interval (λ_e, λ_d), the current dose level is updated as follows:

1. If $\hat{\pi}_j \leq \lambda_e$ and j is not the highest dose level, the current dose level is escalated to $j + 1$

2. If $\hat{\pi}_j \geq \lambda_d$ and j is not the lowest dose level, the current dose level is de-escalated to $j - 1$

3. Otherwise, the current dose remains the same.

The interval-based allocation procedure is repeated until all patients have been observed. At the end of the trial, the pooled adjacent violators algorithm (pava) (Brunk et al. 1972) is used to estimate the MTD.

To obtain interval values, λ_e and λ_d, we consider toxicity thresholds θ_1 and θ_2. The threshold θ_1 is the highest toxicity probability deemed subtherapeutic, thus requiring dose escalation. The threshold θ_2 is the lowest toxicity probability deemed overly toxic, thus requiring dose de-escalation. Liu and Y. Yuan 2015 recommends the thresholds $\theta_1 = 0.6\theta$ and $\theta_2 = 1.4\theta$.

Let π_j denote the DLT probability at dose level j. Then three point hypotheses are considered:

$$H_{0j} : \pi_j = \theta,$$
$$H_{1j} : \pi_j = \theta_1,$$
$$H_{2j} : \pi_j = \theta_2$$

Prior probabilities are assigned to these hypotheses, letting $p(H_{ij})$ denote the prior probability for H_{ij}, where $i = 1, 2, 3$. Let $\mathcal{D}$, $\mathcal{R}$, and $\mathcal{E}$ denote the decisions to de-escalate, remain at the same dose, and escalate, respectively. The probability of

making an incorrect decision is given by

$$\begin{aligned}
\alpha(\lambda_e, \lambda_d) &= p(H_{0j})p(\mathcal{R}^c|H_{0j}) + p(H_{1j})p(\mathcal{E}^c|H_{1j}) + p(H_{2j})p(\mathcal{D}^c|H_{2j}) \\
&= p(H_{0j})p(y_j \leq n_j\lambda_e \text{ or } y_j \geq n_j\lambda_e|H_{0j}) + p(H_{1j})p(y_j > n_j\lambda_e|H_{1j}) \\
&+ p(H_{2j})p(y_j < n_j\lambda_d|H_{2j}) \\
&= p(H_{0j})\{Bin(n_j\lambda_e; n_j, \theta) + 1 - Bin(n_j\lambda_d - 1; n_j, \theta)\} \\
&+ p(H_{1j})\{1 - Bin(n_j\lambda_{1j}; n_j, \theta_1)\} + p(H_{2j})Bin(n_j\lambda_{2j} - 1; n_j, \theta_2)
\end{aligned}$$

In the equation above, $Bin(y, n, \theta)$ denotes the Binomial CDF evaluated at y with n trials and probability of success θ. Liu and Y. Yuan 2015 shows the optimal choices for λ_e and λ_d that minimize error decision $\alpha(\lambda_e, \lambda_d)$ are given by

$$\begin{aligned}
\lambda_e &= \frac{\log\left(\frac{1-\theta_1}{1-\theta}\right) + \frac{1}{n_j}\log\left(\frac{p(H_{1j})}{p(H_{0j})}\right)}{\log\left(\frac{\theta(1-\theta_1)}{\theta_1(1-\theta)}\right)}, \\
\lambda_d &= \frac{\log\left(\frac{1-\theta_2}{1-\theta}\right) + \frac{1}{n_j}\log\left(\frac{p(H_{0j})}{p(H_{2j})}\right)}{\log\left(\frac{\theta(1-\theta_2)}{\theta_2(1-\theta)}\right)}.
\end{aligned}$$

If we assume equally likely priors, so $p(H_{ij}) = \frac{1}{3}$, the boundaries simplify to

$$\begin{aligned}
\lambda_e &= \frac{\log\left(\frac{1-\theta_1}{1-\theta}\right)}{\log\left(\frac{\theta(1-\theta_1)}{\theta_1(1-\theta)}\right)}, \\
\lambda_d &= \frac{\log\left(\frac{1-\theta_2}{1-\theta}\right)}{\log\left(\frac{\theta(1-\theta_2)}{\theta_2(1-\theta)}\right)}.
\end{aligned}$$

Interval boundaries, (λ_e, λ_d), can be obtained using the function `get.boundary` from the R package "BOIN" (Yan, Zhang, et al. 2020), allowing for easy implementation.

Dose-elimination rules for eliminating doses demonstrating excessive toxicity can be

added for practical considerations. Let the DLT probability at dose level k, π_k, have a vague prior. For instance, we may consider $\pi_k \sim Beta(1,1) \equiv Unif(0,1)$. After observing a minimum number of patients at dose level j, dose levels j and above are eliminated from the trial if $P(\pi_j > \theta | n_j, y_j) > \lambda$. In the original BOIN paper, we require three patients be observed at a dose before considering dose elimination and the cutoff, λ, is set to 0.95. Additionally, the trial is halted if the lowest dose is eliminated. The site trialdesign provides tabulated decision boundaries, providing a user-friendly interface for clinicians. Figure 1.4 provides the decision boundaries when the target toxicity rate is $\theta = 0.3$ and $\lambda = 0.95$. In Chapter 3, we first present the Group Averaged BOIN (GAB) design without dose elimination and then discuss GAB with dose elimination.

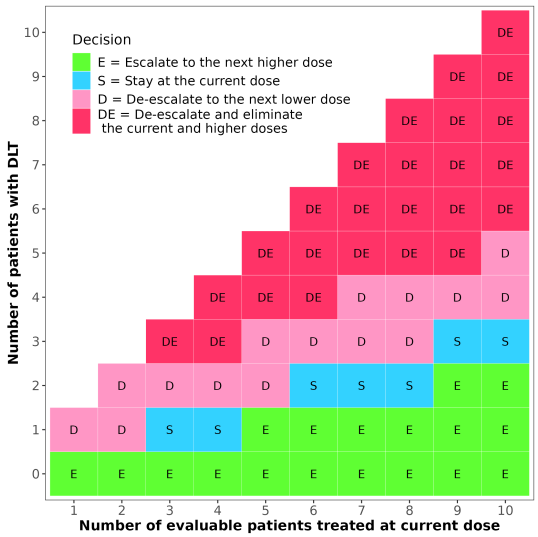


Figure 1.4: Decision boundaries from BOIN, as obtained from trialdesign.

1.3 Phase II Trials

Phase II trials are generally used as a preliminary test of efficacy before proceeding to a phase III trial. The MTD identified in phase I is often used in phase II. However, this MTD-centric approach is currently being challenged, and phase II oncology trials are seeing a paradigm shift. As will be discussed further in Section 4.1, the FDA is interested in seeing dose-randomization being incorporated into phase II trials (Yang et al. 2024; U.S. Food And Drug Administration 2024a).

In Chapter 4, we present two multi-dose designs for phase II trials with two ordered groups. This provides clinicians with phase II multi-dose trial designs for groups after completing a phase I trial with groups. The first design builds upon ideas from the multi-dose randomized trial (MERIT), found in Yang et al. 2024. The second design is a two-stage sequential design with similar objectives to the ubiquitous Simon's Two-Stage Design (Simon 1989). Literature reviews of MERIT and Simon's Two-Stage Design will be provided.

1.3.1 MERIT

MERIT is a simple to use design for a randomized multi-dose phase II clinical trials. Similar to BOIN, decision boundaries can be obtained on the website trialdesign, making this design accessible.

In this multi-dose trial, there are J doses under consideration, $d_1 < d_2 < \cdots < d_J$. Additionally, a total of n patients are observed at each dose, giving a total of $J \times n$ patients in the trial. For a dose to be acceptable, it must be sufficiently safe and sufficiently effective. Let θ_{T_0} denote the toxicity rate deemed overly toxic and θ_{T_1} denote

the toxicity rate deemed safe. Let θ_{E_0} denote the efficacy rate deemed insufficient and θ_{E_1} denote efficacy rate deemed sufficient. Let Y_T, Y_E denote the binary toxicity and efficacy outcomes. Additionally, let $\pi_{T,k} = P(Y_T = 1|d_k)$ and $\pi_{E,k} = P(Y_E = 1|d_k)$ denote the toxicity and efficacy probabilities at dose d_k, respectively. A dose is acceptable if $\pi_{E,k} > \theta_{E_0}$ and $\pi_{T,k} < \theta_{T_0}$.

After defining null and alternative toxicity and efficacy rates, the null hypothesis can be stated as

$$H_0 = \text{All doses are unacceptable}$$

and the alternative hypothesis can be stated as

$$H_1 = \text{At least one dose is acceptable.}$$

A dose can be deemed unacceptable in three different ways, being unsafe and ineffective $(\theta_{T_0}, \theta_{E_0})$, being unsafe and effective $(\theta_{T_0}, \theta_{E_1})$, and being safe and ineffective $(\theta_{T_1}, \theta_{E_0})$. Additionally, there are many different ways to have at least one acceptable dose. From this, H_0 and H_1 are composite hypotheses. Due to the composite nature of H_0 and H_1, we consider the global type one error, denoted as $\alpha^\star$, and the global power, denote as $\beta^\star$. Global type I error is the maximum type I error over all possible configurations (possibilities for parameters) in H_0 and global power is the minimum power over all configurations in H_1.

At the end of the trial, let $n_{T,k}$ and $n_{E,k}$ denote the number of DLTs and efficacious outcomes on dose d_k, respectively. Cutoffs m_T and m_E are defined so that m_T is the highest allowable number of DLTs and m_E is the minimum required efficacious responses. Dose d_k is deemed acceptable if $n_{T,k} \leq m_T$ and $n_{E,k} \geq m_E$. Merritt selects

the cutoffs to maximize global power under a global type I error constraint.

1.3.2 Simon's Two-Stage Design

This section presents Simon's Two-Stage Design (Simon 1989), a two-stage design for phase II trials to determine if a drug is sufficiently effective. The two-stage design presented in Chapter 4 will differ from Simon's Two-Stage Design as it considers both toxicity and efficacy endpoints, several doses, and two groups. However, we present Simon's Two-Stage Design to provide motivation for what will be called "Simon's Statistic" and to introduce two-stage designs.

Simon's design considers a phase II trial with a single agent. We are interested in accessing the efficacy of the agent in the trial to determine if the agent warrants a phase III trial. Let p_0 denote the unacceptable rate of efficacy and p_1 denote the acceptable rate of efficacy. From this, the null hypothesis is $H_0 : p \leq p_0$ and the alternative hypothesis is $H_1 : p \geq p_1$. Let n_1 denote the number of patients in the first stage of the trial and n_2 denote the number of patients in the second stage.

After observing the n_1 patients in the first stage, the trial continues onto the second stage if at least r_1 responses are observed. If a trial continues onto the second phase, the drug is deemed acceptable if the number of responses is at least r_2. Letting $EN(p_0)$ denote the expected number of patients observed under the null rate p_0, the goal of Simon's design is to find the optimal design parameters that minimize $EN(p_0)$, while meeting type I error and power constraints.

1.4 Completely and Partially Ordered Groups

In this subsection, we provide a literature review of completely and partially ordered groups. This literature review will cover previous trials with ordered groups and discuss the current literature on phase I trials for ordered groups. Before this dissertation, no phase II trials existed for ordered groups. This literature review is adapted from Celum, Horton, and M. Conaway 2024, and Celum and M. Conaway 2024.

Previous trials have stratified patients into heterogeneous groups. In these trials, clinicians had complete or partial prior knowledge of the relative sensitivity of the groups. If we have complete knowledge, the groups can be ordered by the probability of a DLT for any given dose. If we have partial knowledge, some but not all groups can be ordered by the probability of a DLT for any given dose.

Ramanathan et al. 2008; P. M. LoRusso et al. 2012; and Leal et al. 2011 provide examples of trials where groups are completely ordered. Ramanathan et al. 2008 and P. M. LoRusso et al. 2012 use liver health to stratify patients into four groups. Leal et al. 2011 uses renal to health stratify patients into five groups.

Innocenti et al. 2014 provides an example of a trial where groups are partially ordered. This trial was interested in finding the MTD for Irinotecan and used UGT1A1 genotype to create groups. It is known that patients with genotype $^*28/^*28$ are more sensitive than those with genotypes $^*1/^*1$ or $^*1/^*28$. Prior to the trial, it was not known if patients with genotype $^*1/^*1$ or genotype $^*1/^*28$ are more sensitive, creating a partial ordering.

Previous designs have been proposed for finding group-specific MTDs when there is a complete or partial ordering. Dose-finding designs for complete orderings include O'Quigley and Xavier Paoletti 2003; Zhilong Yuan and Rick Chappell 2004;

Ivanova and K. Wang 2006; O'Quigley and Iasonos 2014; Wages, Read, and Petroni 2015; and M. R. Conaway and Wages 2017. O'Quigley and Xavier Paoletti 2003, and O'Quigley and Iasonos 2014 add a "shift" parameter to modify the continual reassessment method (CRM). Zhilong Yuan and Rick Chappell 2004, and Ivanova and K. Wang 2006 use bivariate isotonic regression to locate group-specific doses. Wages, Read, and Petroni 2015 designed an adaptive phase I/II design that considers both efficacy and toxicity endpoints while stratifying patients into two ordered groups based on genetic and clinical factors. Muller et al. 2020 implements this design in a stereotactic body radiation therapy trial. M. R. Conaway and Wages 2017 consider a collection of possible DLT probability orderings, subject to the group orderings, and uses order-restricted methods from Hwang and Peddada 1994 to estimate DLT probabilities.

Dose-finding methods for partial orderings include M. R. Conaway 2017a; M. R. Conaway 2017b; Horton, Wages, and M. R. Conaway 2019; and Lin, Thall, and Y. Yuan 2020. M. R. Conaway 2017a first "smooths" the observed DLT proportions to agree with the partial ordering, then applies the CRM. M. R. Conaway 2017b first applies the CRM independently to each group to obtain initial estimates. Second, using the partial ordering and the order-restricted methods from Hwang and Peddada 1994, the initial estimates are adjusted. Horton, Wages, and M. R. Conaway 2019 considers all possible MTD configurations, called shifts, given the partial order. The CRM is applied to the shifts and the shift with the highest likelihood is used to estimate MTDs. Lin, Thall, and Y. Yuan 2020 designs a phase I/II trial for partially or completely ordered subgroups using both efficacy and toxicity outcomes while allowing outcomes to be delayed.

1.5 Outline

In this dissertation, we contribute to the existing literature on phase I trials for groups and create the first multi-dose phase II group designs. In particular, we create the first design for partially or completely ordered groups using ordinal toxicity and the first such model-assisted design.

This dissertation will be organized as follows. Chapter 2 will cover the Quasi-CRM Shift method, as adapted from Celum, Horton, and M. Conaway 2024, the first design for partially ordered groups using ordinal toxicity. Chapter 3 will cover the Group Averaged BOIN design, the first model-assisted design for partially or completely ordered groups, as adapted from Celum and M. Conaway 2024. Chapter 4 will cover two multi-dose phase II designs for ordered groups. Chapter 5 will conclude and discuss future research.

Chapter 2

Quasi-CRM Shift

2.1 Introduction

This chapter is adapted from Celum, Horton, and M. Conaway 2024.

This chapter proposes a trial design for estimating group-specific maximally tolerated doses (MTDs) when toxicity is ordinal. With an increased interest in personalized medicine, it is important to use group information if groups are present. Phase I trials often dichotomize toxicities as being dose-limiting toxicities (DLT) or not. This design uses both toxicity severity and DLTs to guide dose allocation and MTD estimation.

Toxicities have grades 0 to 5, corresponding to toxicities: none, mild, moderate, severe, life-threatening, or fatal (U.S. Department of Health and Human Services 2017); providing more information than binary toxicity. If a grade 5 toxicity occurs, a safety review is often required to resume the trial (Yuan, Chappell, and Bailey 2007). As grade 5 toxicities can cause a trial to be halted, grade 5 toxicities will not be considered. Depending on the protocol, DLTs are toxicities of grade 3 and above, or toxicities of grade 4 and above (Yuan, Chappell, and Bailey 2007). Using toxicity grades, we can both account for low-grade toxicities and differentiate a grade 3 DLT from a grade 4 DLT. Bekele and Thall 2004 designed a soft tissue sarcoma phase I trial that accounts for toxicity type. In this trial, clinicians were interested in

toxicity type as low-grade, non-DLTs inform us that a dose-limiting toxicity is likely to occur at a higher dose. Additionally, differentiating DLT severity is informative when a drug exhibits more severe toxicities, such as renal, as grade 4 renal toxicities are irreversible (Yuan, Chappell, and Bailey 2007; Pan et al. 2014).

Several dose-finding designs have been created for ordinal toxicity. C. Wang, T. T. Chen, and Tyan 2000 used toxicity grades to modify the CRM, making dose allocation more conservative after observing a grade 4 toxicity. Yuan, Chappell, and Bailey 2007 assigned toxicity scores, $(s_0, s_1, s_2, s_3, s_4)$, to weight the severity of each toxicity grade. The toxicity scores are called equivalent toxicity (ET) scores and are elicited from clinicians. These scores are normalized so the scores are in the unit interval. These normalized toxicity scores are plugged into a Bernoulli-likelihood. This likelihood is called the quasi-Bernoulli likelihood since the responses in the unit interval but are not Bernoulli (Papke and Wooldridge 1996). As the responses are evaluated in a Bernoulli likelihood, the CRM can be utilized. This is called the Quasi-CRM. Similar to Yuan, Chappell, and Bailey 2007, the designs in Van Meter, Garrett-Mayer, and Bandyopadhyay 2012; Pan et al. 2014; and O'Connell, Wages, and Garrett-Mayer 2023 estimate the MTD using graded toxicity. Van Meter, Garrett-Mayer, and Bandyopadhyay 2012 extends the CRM to the ordinal toxicity using the continuation ratio model. Pan et al. 2014 used the Quasi-CRM and Bayesian model selection to select the best model from a collection of models. O'Connell, Wages, and Garrett-Mayer 2023 found the MTD for drug combinations by combining the Quasi-CRM with CRM partial ordering methods from Wages, M. R. Conaway, and O'Quigley 2011.

This is the first design for locating group-specific MTDs using ordinal toxicity. To estimate group MTDs using ordinal toxicities, the shift method is combined with

the Quasi-CRM. Chapter 2 will proceed as follows. Section 2.2 will provide the framework for groups and ordinal toxicity. Section 2.3 will cover the proposed model and allocation procedures. Section 2.4 will cover the simulations. Section 2.5 will conclude and discuss future areas of research.

2.2 Groups and Ordinal Toxicity

Let $g = 1, 2, \ldots, G$ denote the group membership of a patient. Let $d_1 < d_2 < \cdots < d_K$ denote the doses in a trial. We use the partial ordering from the Intrinotecan trial (Innocenti et al. 2014) with $G = 3$. Group 1 will refer to the patients with genotype $^*1/^*1$, Group 2 will refer to the patients with genotype $^*1/^*28$, and Group 3 will refer to the patients with genotype $^*28/^*28$. Group 3 is the most sensitive group and there is not an a priori ordering between groups 1 and 2. Simulations will be conducted for $K = 4$ and $K = 6$ dose levels. As in Yuan, Chappell, and Bailey 2007, toxicities of grades 3 and 4 will be considered DLTs. Corresponding to the toxicity grades 0 to 4, there will be ET scores s_0 to s_4. We will use the ET scores from Yuan, Chappell, and Bailey 2007, as provided below. Note the DLT cutoff and toxicity scores depend on clinician preference and can differ from those presented in this chapter.

$$s_0 = s_1 = 0, s_2 = 0.5, s_3 = 1, s_4 = 1.5.$$

Yuan, Chappell, and Bailey 2007 elicited these scores from clinician preferences. For these clinicians, grade 1 toxicities are not concerning, two grade 2 toxicities are equivalent to a grade 3 toxicity, and a grade 2 toxicity plus grade 3 toxicity is equivalent to a grade 4 toxicity. Grade 3 toxicities are scored as 1 since these toxicities are the DLT cutoff. Then, from the previous relationship, grade 2 toxicities are scored as 0.5

and grade 4 toxicities as 1.5.

The ET score for the i^{th} patient in group g is denoted as s_{ig}, where s_{ig} is the ET score from the highest observed toxicity grade for this patient during the window of observation. The expected ET score for group g at dose d_k, is denoted as S_{gk}, where $S_{gk} = \mathbf{E}_g(s|d_k)$. Additionally, π_{gk} denotes the DLT probability for group g at dose d_k. We define θ_T as the DLT target and θ_S as the ET target. Similar to Yuan, Chappell, and Bailey 2007, we set the thresholds as $\theta_T = 0.33$ and $\theta_S = 0.47$. The ET target, $\theta_S = 0.47$, is obtained by considering the toxicity profile: 49% grades 0 or 1, 18% grade 2, 23% grade 3, and 10% grade 4, giving the expected ET score:

$$S = 0.49 \times 0 + 0.18 \times 0.5 + 0.23 \times 1 + 0.10 \times 1.5 = 0.47$$

The ET MTD for group g can be defined as the dose with ET score closest to the ET target; that is, d_k with $k = \text{argmin}_k |S_{gk} - \theta_S|$. The DLT MTD for group g can be defined as the dose with DLT probability closest to the DLT target; that is, d_k with $k = \text{argmin}_k |\pi_{gk} - \theta_T|$. The DLT MTD and ET MTD need not be the same. Simulations include examples when the DLT MTD is higher than the ET MTD and when the ET MTD is higher than the DLT MTD. In this chapter, we will define the MTD to be the minimum of the ET MTD and DLT MTD; thus, we control for the DLT probability and total toxicity profile of a dose. That is, for a group g, the group MTD is the dose d_k, where

$$k = \min \left\{ \underset{k}{\text{argmin}} \, |S_{gk} - \theta_S|, \underset{k}{\text{argmin}} \, |\pi_{gk} - \theta_T| \right\}. \tag{2.1}$$

From this definition, dose allocation and estimation will be at least as conservative as allocation and estimation using only DLT data. Suppose a group has the relationship

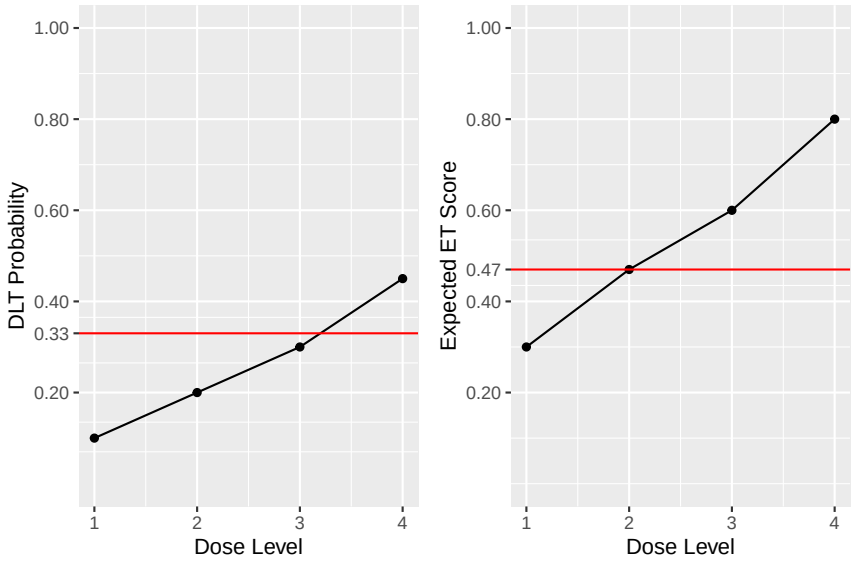


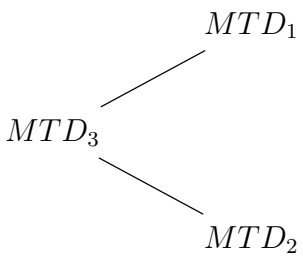
Figure 2.1: DLT MTD and ET MTD Plot

between dose, expected ET score, and DLT probability as provided in Figure 2.1. In Figure 2.1, horizontal red lines delineate the ET and DLT targets. From this, the group-specific DLT MTD is dose level 3 and the group-specific ET MTD is dose level 2, resulting in a group-specific MTD of dose level 2, this being the minimum of the two MTDs.

Now we normalize the ET scores, giving new scores between 0 and 1, allowing for the application of the Quasi-CRM: $s_i^\star = s_i/s_4$, $\theta_S^\star = \theta_S/s_4 = 0.31$, and $S_{gk}^\star = S_{gk}/s_4$. After normalization we have $s_i^\star, \theta_S^\star, S_{gk}^\star \in [0, 1]$.

Let MTD_g denote the MTD for group g. From the motivating example, $MTD_3 \leq MTD_1$ and $MTD_3 \leq MTD_2$ but it is not known if $MTD_1 \leq MTD_2$ or $MTD_2 \leq$

MTD_1. This relationship is illustrated in the diagram below.



A reversal occurs when the estimated MTDs contradict the known group ordering. From this relationship, reversals occur if $\widehat{MTD}_3 > \widehat{MTD}_2$ or $\widehat{MTD}_3 > \widehat{MTD}_1$. As discussed in Horton, O'Quigley, and M. R. Conaway 2019, reversals occur when trials are done independently for each group. Simulations will count the number of reversals that occur in independent Quasi-CRM trials.

The proposed method will make use of the shift method and Quasi-CRM. The shift method and CRM will be applied to the DLT data, and the shift method and Quasi-CRM will be applied to the ET data.

2.3 Proposed Model and Allocation

For the proposed method, we consider the possible shifts in DLT MTD and ET MTD for groups 3 and 1, and groups 2 and 1. These shifts will correspond to skeleton shifts, giving a collection of DLT and ET models. The best-fitting models will be used for allocation.

The skeleton is an initial guess at the DLT probabilities. Conceptually, the skeleton shift between, say, groups 1 and 3 is how many dose levels you need to shift the

skeleton of Group 3 up by in order to match the skeleton of Group 1. For instance, if the shift is one, the skeleton value for Group 3 at dose level one will be equal to the skeleton value for Group 1 at dose level two. For the given partial order, there will be $2 \times K - 1$ skeleton values for both the ET and DLT models. The value $2 \times K - 1$ comes from considering when the skeleton shift between the most and least sensitive group is $K - 1$. For this shift, the skeleton value for the most sensitive group, in this case, Group 3, at dose level 1 is equal to the skeleton value for the least sensitive group, in this case, groups 1 or 2, at dose level K. In this case, the most and least sensitive groups would only have one skeleton value in common, requiring $2 \times K - 1$ skeleton values. See Horton, Wages, and M. R. Conaway 2019 for details.

First, we consider a skeleton of size $2 \times K - 1$ that will be used for the ET responses,

$$0 \leq q_1 \leq q_2 \leq \cdots \leq q_{2\times K-1} \leq 1.$$

The ET shift skeleton was obtained using the "getprior" function from the R package "dfcrm" (K. Cheung 2019). Getprior has the arguments halfwidth, target, nu, levels, and model. Halfwidth controls the spread of the skeletons, with a larger halfwidth making the skeleton more spread out. Target is the target toxicity rate, nu is the prior MTD estimate, levels is the skeleton length, and "model" specifies the model being used. We use the empirical model, which is of the form $q^{\exp(a)}$.

In simulation studies, four and six-dose trials were considered. For the four-dose case, the function arguments were `getprior(halfwidth = 0.06, target = 0.31, nu = 3, nlevel = 7, model = "empiric")`. For the six-dose case, the function arguments were `getprior(halfwidth = 0.06, target = 0.31, nu = 4, nlevel = 11, model = "empiric")` (K. Cheung 2019). These parameters were chosen to

mirror the shift model in Horton, Wages, and M. R. Conaway 2019.

Second, we consider a skeleton of size $2 \times K - 1$ that will be used for the DLT responses,

$$0 \leq p_1 \leq p_2 \leq \cdots \leq p_{2\times K-1} \leq 1.$$

Similarly, the DLT shift skeleton was obtained using the "getprior" function. For the four-dose case, the function arguments were `getprior(halfwidth = 0.06, target = 0.33, nu = 3, nlevel = 7, model = "empiric")`. For the six-dose case, the function arguments were `getprior(halfwidth = 0.06, target = 0.33, nu = 4, nlevel = 11, model = "empiric")`.

Let Δ_{32}^S and Δ_{31}^S denote the ET skeleton shifts between groups 3 and 2 and between groups 3 and 1, respectively. Let Δ_{32}^T and Δ_{31}^T denote the DLT skeleton shifts between groups 3 and 2 and between groups 3 and 1, respectively. Then, $\Delta_{32}^S, \Delta_{31}^S, \Delta_{32}^T, \Delta_{31}^T \in \{0, 1, \ldots, K-1\}$. Corresponding to each pair of ET shifts, $(\Delta_{32}^S, \Delta_{31}^S)$, we get a Quasi-CRM model, giving $K \times K = M$ Quasi-CRM models. Similarly, corresponding to each pair of DLT shifts, $(\Delta_{32}^T, \Delta_{31}^T)$, we get a CRM model, giving $K \times K = M$ CRM models.

For an example of a model obtained from a shift, consider the ET model $(\Delta_{32}^S = 2, \Delta_{31}^S = 1)$, when $K = 4$. We first consider the larger shift, $\Delta_{32}^S = 2$. As the shift between groups 3 and 2 is 2, the skeleton for Group 3 starts with q_3. As the shift between groups 3 and 1 is 1, the skeleton for Group 1 should be one behind the skeleton for Group 3, so the starting skeleton value for Group 1 is q_2. A similar process would give the group skeletons for the DLT model $(\Delta_{32}^T = 2, \Delta_{31}^T = 1)$. Tables 2.1 and 2.2 provide the group skeletons corresponding to these models. For further

reference, Horton, Wages, and M. R. Conaway 2019 provides a table of the skeleton shifts corresponding to the 4×4 models for the four-dose scenario.

$(\Delta^S_{3,2} = 2, \Delta^S_{3,1} = 1)$	Dose 1	Dose 2	Dose 3	Dose 4
Group 3 skeleton	q_3	q_4	q_5	q_6
Group 2 skeleton	q_1	q_2	q_3	q_4
Group 1 skeleton	q_2	q_3	q_4	q_5

Table 2.1: ET skeleton shift example for shift $(\Delta^S_{3,2} = 2, \Delta^S_{3,1} = 1)$.

$(\Delta^T_{3,2} = 2, \Delta^T_{3,1} = 1)$	Dose 1	Dose 2	Dose 3	Dose 4
Group 3 skeleton	p_3	p_4	p_5	p_6
Group 2 skeleton	p_1	p_2	p_3	p_4
Group 1 skeleton	p_2	p_3	p_4	p_5

Table 2.2: DLT skeleton shift example for shift $(\Delta^T_{3,2} = 2, \Delta^T_{3,1} = 1)$.

MTD estimation will proceed as follows. First, the Quasi-CRM will select the model that fits the ET data the best. This model will be used to estimate group ET MTDs. Second, the CRM will select the model that fits the DLT data the best. This model will be used to estimate the group DLT MTDs. The group MTD estimates will be the minimum of the estimates for the group ET MTD and group DLT MTD. During the trial, the next patient is allocated to their respective group-MTD estimate, with the additional restriction of preventing doses from being skipped. Section 2.3.3 details the allocation rules.

2.3.1 ET MTD

For model m, the (normalized) ET score response for group g, given dose d_k, is modeled as

$$\mathbf{E}_{mg}(s^\star | d_k, a) = \phi_{mg}(d_k, a), \tag{2.2}$$

where

$$\phi_{mg}(d_k, a) = (q_{mgk})^{\exp(a)}\,.$$

We assume the parameter a has the prior $N(0, 1.34)$.

Let x_{ig} denote the dose for the i^{th} patient in group g, $s^{\star}_{ig}$ denote the (normalized) ET response for the i^{th} patient in group g, and n_g denote the number of observed patients in group g. For model, m, and ET data, $\mathcal{D}_S$, the Quasi-Bernoulli likelihood (Papke and Wooldridge 1996) is

$$L_m(a|\mathcal{D}_S) = \prod_{g=1}^{3}\prod_{i=1}^{n_g} \phi_{mg}(x_{ig}, a)^{s^{\star}_{ig}}(1 - \phi_{mg}(x_{ig}, a))^{1-s^{\star}_{ig}}. \tag{2.3}$$

Let $g(a)$ be the prior density function for a. Under model m, the posterior density for a is

$$g(a|\mathcal{D}_S, m) = \frac{g(a)L_m(a|\mathcal{D}_S)}{\int_{-\infty}^{\infty} g(a)L_m(a|\mathcal{D}_S)da}.$$

Now consider prior model probabilities for the ET models, $\{p_S(1), p_S(2), \ldots, p_S(M)\}$. In simulations, the models have equal prior likelihoods, so $p_S(m) = \frac{1}{M}$, for $m = 1, 2, \ldots, M$. Using the ET data, we obtain posterior model probabilities,

$$\begin{aligned} p_S(m|\mathcal{D}_S) &= \frac{p_S(m)p_S(\mathcal{D}_S|m)}{\sum_{m=1}^{M} p_S(m)p_S(\mathcal{D}_S|m))} \\ &= \frac{p_S(m)\int_{-\infty}^{\infty} g(a)L_m(a|\mathcal{D}_S)da}{\sum_{m=1}^{M} p_S(m)\int_{-\infty}^{\infty} g(a)L_m(a|\mathcal{D}_S)da}. \end{aligned}$$

Let m_S denote the model with the highest posterior probability, that is

$$m_S = \underset{m}{\operatorname{argmax}}\, p_S(m|\mathcal{D}_S).$$

Using model m_S, we obtain the posterior mean for a,

$$\hat{a}_{m_S} = \int_{-\infty}^{\infty} a g(a|\mathcal{D}_S, m_S) da.$$

Using the posterior mean, the expected ET score for group g is estimated as: $\phi_{m_S g}(d_k, \hat{a}_{m_S}) = (q_{m_S g k})^{\exp(\hat{a}_{m_S})}$. Thus, the ET MTD for group g is estimated as

$$x^S_{(n_g+1)g} = \underset{d_k}{\operatorname{argmin}} \left|\phi_{m_S g}(d_k, \hat{a}_{m_S}) - \theta^\star_S\right|.$$

2.3.2 DLT MTD

Using the DLT shift skeleton and the DLT data, we repeated the same procedure and obtain the estimates for the group-specific MTDs. We model the DLT response as

$$\mathbf{E}_{mg}(y|d_k, b) = \psi_{mg}(d_k, b), \tag{2.4}$$

where

$$\psi_{mg}(d_k, b) = (p_{mgk})^{\exp(b)}.$$

Similarly, a $N(0, 1.34)$ prior is used for b.

As in the previous section, we select the DLT model with highest posterior likelihood. Let m_T denote the DLT model with the highest posterior likelihood and $\hat{b}_T$ denote

the posterior mean for b under model m_T. The DLT MTD for group g is estimated as

$$x^T_{(n_g+1)g} = \underset{d_k}{\operatorname{argmin}} \left| \psi_{m_T g}(d_k, \hat{b}_{m_T}) - \theta_T \right|.$$

The estimated MTD for group g is the minimum of the two MTD estimates,

$$x_{(n_g+1)g} = \min \left\{ x^S_{(n_g+1)g}, x^T_{(n_g+1)g} \right\}.$$

2.3.3 Allocation Rules

Allocation is group-specific and does not allow doses to be skipped. If the next patient is in Group 3, the sensitive group, the patient will be allocated the minimum of the MTD estimate for Group 3, $x_{(n_3+1)3}$, and one dose higher than the highest dose observed in Group 3. If the next patient is in groups 1 or 2, the patient will be allocated the minimum of the group-specific MTD estimate, $x_{(n_g+1)g}$, and one dose higher than the overall highest observed dose. Table 2.3 provides the allocation rules, where $d_g^{\max}$ is the highest observed dose in group g, as denoted in Horton, Wages, and M. R. Conaway 2019. By constraining Group 3 by $d_3^{max} + 1$, and constraining groups 1 and 2 by $\max\{d_1^{\max}, d_2^{\max}, d_3^{\max}\} + 1$, allocation follows the known group ordering.

Group	Dose Allocation
1	$\min\left\{\max\{d_1^{\max}, d_2^{\max}, d_3^{\max}\} + 1, x_{(n_1+1)1}\right\}$
2	$\min\left\{\max\{d_1^{\max}, d_2^{\max}, d_3^{\max}\} + 1, x_{(n_2+1)2}\right\}$
3	$\min\left\{d_3^{\max} + 1, x_{(n_3+3)3}\right\}$

Table 2.3: Allocation rules for the proposed method with $d_g^{\max}$ denoting the highest observed dose in group g and $x_{n_g+1,g}$ denoting the MTD estimate for group g.

2.4 Simulations

The proposed method was compared to the shift method and the independent Quasi-CRM group trials method. These methods will be called the DLT Shift and the Independent Quasi-CRM. The empirical model, $p^{\exp(a)}$, was used for all methods, with $a \sim N(0, 1.34)$.

The DLT Shift method ignored the ET data and applied the shift method only to the DLT data. The DLT skeletons used in the proposed method were used for the DLT Shift method. The allocation rules in the proposed method that restricted dose skipping based on group membership were applied.

For each group, the Independent Quasi-CRM method applied the CRM to the DLT data to obtain the DLT MTD and the Quasi-CRM to the ET data to obtain the ET MTD. The estimated group MTD was the minimum of the two MTD estimates. In the independent group trials, dose escalation did not allow for the skipping of untried doses. The "getprior" function computed the skeleton values with "halfwidth" set to 0.06. In the four and six-dose scenarios, "nu" was set as 3 and 4, respectively. The reversal percentage was recorded for each scenario.

2.4.1 Scenarios

Simulations were run for trials with four and six-dose levels. There were 12 four-dose scenarios and 12 six-dose scenarios. Appendix A provides the plots of these scenarios. For the four-dose scenarios, trial sizes of 30, 45, and 72 were considered. For the six-dose scenarios, trial sizes of 45, 69, and 108 were considered. Since the six-dose scenarios have more dose options, the trial sizes were scaled accordingly.

Patient group membership was randomly generated with an equal probability that a patient belongs to a group while requiring each trial has at least one patient from each group. The larger trial sizes are justified since the expected number of patients in a group is a third of the trial size. For each scenario and trial size, 1000 trials were simulated. Dose allocation and dose selection percentages were recorded. This data and the R code used to generate this data is available in the online version of Celum, Horton, and M. Conaway 2024. For each combination of trial size and number of doses, the percentage of correct selection (PCS) and the percentage of correct allocation (PCA) was averaged over all groups and scenarios. For group-specific results, see sections 5 and 6 of supplementary materials in Celum, Horton, and M. Conaway 2024. The number of reversals was recorded for the Independent Quasi-CRM method. The proposed method does not allow for reversals and outperforms the Independent Quasi-CRM in PCS and PCA. Additionally, the proposed method uses the complete toxicity profile of a dose, accounting for low-grade and high-grade toxicities, improving the DLT Shift. Additional analyses for overdose control and model sensitivity were conducted. Analyses were run to test sensitivity to group proportions, choice of ET scores, choice of prior standard deviation, and choice of halfwidth. After running these analyses, we can conclude the proposed method is robust.

2.4.2 Reversals

For each scenario and trial size, the percentage of reversals was recorded for the Independent Quasi-CRM method. Note that the scenarios in the four-dose trials differ from the scenarios in the six-dose trials. With this in mind, one should only compare scenarios within a table and should not compare scenarios across the tables.

Looking at Tables 2.4 and 2.5, we see that the number of reversals decrease as the sample size increases. In the four-dose scenarios, the reversal percentage is highest in scenarios 3 and 7, when the MTD is the same for all groups. In the six-dose scenarios, the reversal percentage is highest in scenario 7, when groups 3 and 1 have the same MTD. These tables show that it is more likely for a reversal to occur when the MTD for Group 3 is closer to the MTDs for groups 1 and 2. These results are similar to the results in Horton, O'Quigley, and M. R. Conaway 2019 for the DLT MTD.

Number of Doses	Trial Size	1	2	3	4	5	6	7	8	9	10	11	12	AVG
4	30	28.4	39.8	49.2	23.8	1.4	33.2	51.6	25.4	20.4	27.4	28.4	25.3	29.5
4	45	25.3	37.5	43.7	20.7	0.1	31.2	51.0	20.9	16.5	25.2	25.2	19.7	26.4
4	72	18.6	30.8	42.3	11.7	0.0	29.7	43.7	13.1	9.8	17.9	22.5	14.8	21.2

Table 2.4: Reversal percentages for parallel Quasi-CRM trials for the four-dose scenarios. Standard Errors do not exceed 1.6%.

Number of Doses	Trial Size	1	2	3	4	5	6	7	8	9	10	11	12	AVG
6	45	28.6	20.0	4.7	21.0	3.2	21.7	32.2	12.7	18.0	26.1	6.5	20.0	17.9
6	69	22.6	14.5	1.0	16.1	1.7	14.3	31.0	8.4	12.2	18.9	5.0	14.6	13.4
6	108	16.8	9.0	0.2	10.6	0.3	9.7	24.6	4.2	8.1	13.3	1.6	8.7	8.9

Table 2.5: Reversal percentages for parallel Quasi-CRM trials for the six-dose scenarios. Standard Errors do not exceed 1.6%.

2.4.3 PCS and PCA

Table 2.6 provides the PCS and PCA, respectively. In this table, "Indep Quasi" is the Independent Quasi-CRM method. The proposed method performs the best in all trial size and dose number combinations. The DLT Shift method performs the worst since this method does not use graded toxicity. The proposed method outperforms the Independent Quasi-CRM method most when the trial sizes are small. When the trial size is 30, the expected number of patients in a group is 10, making it difficult to estimate the MTD for a specific group when performing trials independently. This

challenge highlights the proposed method's ability to use data from all 30 patients for estimation, effectively giving a larger sample size than that of the independent trials.

		PCS			PCA		
# Doses	# Patients	Proposed	Indep Quasi	DLT Shift	Proposed	Indep Quasi	DLT Shift
4	30	**46.9**	43.5	28.8	**39.9**	36.4	29.1
4	45	**51.8**	48.5	29.9	**42.7**	39.5	29.0
4	72	**57.4**	55.4	31.4	**47.1**	44.3	29.5
6	45	**44.2**	42.4	27.1	**37.0**	34.6	26.5
6	69	**49.3**	48.3	27.7	**40.4**	38.4	26.9
6	108	**55.6**	54.8	28.7	**44.7**	43.1	27.5

Table 2.6: Percentage correct selection (PCS) and percentage correct allocation (PCA) by number of doses, number of patients, and method. Standard Errors do not 0.04%.

2.4.4 Sensitivity Analysis and Overdose Control

Additional analysis was performed for overdose control, toxicity rates, and model sensitivity. Tables 2.7 and 2.8 provide statistics on toxicity rates and overdose control for all six combinations of trial size and number of doses. Table 2.7 provides the DLT percentages and average (normalized) ET scores for patients in the trial. Additionally, Table 2.8 provides the percentage of patients allocated to doses above the MTD and the percentage of selections above the MTD. From Table 2.7, the average observed ET score approaches the normalized ET target of 0.313 quicker in the proposed method than the Independent Quasi-CRM method. As the DLT Shift does not use toxicity grades, the average ET score for this method is above the ET target while the DLT rate approaches the DLT target of 0.33. It is important to note the primary goal in the trial is to allocate patients to the MTD, not minimize the number of DLTs and average ET score. Allocation results in Table 2.8 show the proposed method and Independent Quasi-CRM have similar rates of allocations above the MTD, with

the proposed method allocating slightly more patients to doses above the MTD when the trial size is smaller. This discrepancy can be attributed to the proposed method not requiring patients in groups 1 and 2 to start at the lowest dose if other groups have been observed. Selection results in Table 2.8 shows the proposed method selects doses above the MTD with the lowest frequency, highlighting the proposed method's ability to reduce overdose selection by sharing group information.

		Average ET ×100			DLT Rate ×100		
# Doses	# Patients	Proposed	Indep Quasi	DLT Shift	Proposed	Indep Quasi	DLT Shift
4	30	29.5	28.8	34.0	24.5	23.8	29.0
4	45	29.8	29.2	34.7	24.8	24.1	29.7
4	72	29.9	29.7	35.3	24.9	24.7	30.3
6	45	30.8	30.0	35.9	25.9	25.2	31.2
6	69	31.1	30.7	36.6	26.1	25.7	31.7
6	108	31.3	31.0	37.0	26.2	26.0	32.1

Table 2.7: Average DLT rate and average (normalized) ET score for patients during the trial. Note that the normalized ET target is 0.313 and the DLT target is 0.33.

		Percent Allocation Above MTD			Percent Selection Above MTD		
# Doses	# Patients	Proposed	Indep Quasi	DLT Shift	Proposed	Indep Quasi	DLT Shift
4	30	29.7	28.8	50.0	30.0	31.4	57.9
4	45	29.7	29.5	53.4	25.8	29.6	59.9
4	72	28.5	28.9	55.9	25.5	26.2	60.6
6	45	32.9	31.4	54.1	32.9	33.9	61.9
6	69	32.0	31.6	57.1	29.9	30.8	63.8
6	108	31.0	30.9	59.5	27.4	27.9	65.0

Table 2.8: Overdose Statistics: comparing percentage of patients allocated to doses above the MTD and percentage of times a method selects a dose above the MTD at the conclusion of the trial.

Tables 2.9 and 2.10 provide sensitivity analyses. For simplicity, these analyses were conducted for two combinations, four doses with 45 patients and six doses with 69 patients. Table 2.9 tests sensitivity to group membership probabilities being equal and sensitivity to the choice of ET scores. To test sensitivity to group probabilities, simulations were conducted with the probabilities 0.46, 0.41, and 0.13, for groups 1 through 3, respectively. These group probabilities were chosen to match the group

proportions in Innocenti et al. 2014. To test sensitivity to choice of ET score, a collection of 1000 ET scores were generated using the same procedure from the sensitivity analysis in Yuan, Chappell, and Bailey 2007. ET scores were generated by first fixing $s_0 = s_1 = 0$ and $s_3 = 1$, then sampling $s_2 \sim Unif(0.3, 0.7)$ and $s_4 \sim Unif(1.2, 1.8)$. When conducting 1,000 simulations for a given curve, each set of ET scores was used once, each time adjusting the method under consideration for the ET scores in use and computing statistics based on these scores. Table 2.9 shows the proposed method has superior performance, regardless of group probabilities or ET scores. The proposed method is the least affected by unequal group probabilities, with minimal change in PCS.

Table 2.10 provides a sensitivity analysis for the standard deviation of priors $a, b \sim N(0, \sigma^2)$, and the choice of halfwidth for the function `getprior` used to obtain skeletons. In simulations, we used $\sigma = \sqrt{1.34}$, and a halfwidth of 0.06. To test model sensitivity, we compared results when $\sigma = \sqrt{2}$ and results when halfwidth = 0.05. These results show the proposed method is not sensitive to the choice of standard deviation and halfwidth.

			PCS			PCA		
# Doses	# Patients	Change	Proposed	Indep Quasi	DLT Shift	Proposed	Indep Quasi	DLT Shift
4	45	None	**51.8**	48.5	29.9	**42.7**	39.5	29.0
4	45	Unequal Probabilities	**51.5**	47.5	31.0	**43.4**	40.5	31.8
4	45	Random Scores	**47.9**	45.8	31.0	**39.8**	37.1	29.1
6	69	None	**49.3**	48.3	27.7	**40.4**	38.4	26.9
6	69	Unequal Probabilities	**49.7**	47.4	29.2	**41.2**	39.0	29.0
6	69	Random Scores	**47.4**	46.1	29.2	**39.0**	36.7	27.3

Table 2.9: Percentage correct selection (PCS) and percentage correct allocation (PCA) comparisons for simulations where ET Scores are randomly generated or group probabilities are unequal. Making comparisons to results with equal group rates and ET scores (0,.5,1,1.5).

		PCS			PCA		
# Doses	# Patients	Regular	SD = $\sqrt{2}$	Halfwidth = 0.05	Regular	SD = $\sqrt{2}$	Halfwidth = 0.05
4	45	51.8	51.5	50.4	42.7	42.4	42.0
6	69	49.3	49.3	49.2	40.4	40.4	39.8

Table 2.10: Percentage correct selection (PCS) and percentage correct allocation (PCA) for the proposed method when varying SD and Halfwidth

2.4.5 General Group Sizes and Orders

This chapter considered $G = 3$ groups, with $MTD_1 \geq MTD_3$ and $MTD_2 \geq MTD_3$. The proposed method can extend to any partial or complete order with any number of groups. Horton, O'Quigley, and M. R. Conaway 2019 applies the DLT shift method to four different orderings with four groups. Since the DLT shift method can be generalized for any number of groups and any partial or complete ordering, the proposed method can be generalized likewise. We illustrate how to apply the proposed method to three completely ordered groups. For other group orderings, refer to Horton, O'Quigley, and M. R. Conaway 2019.

Consider the ordering: $MTD_1 \geq MTD_2 \geq MTD_3$. Then, we have the ET shift $\Delta^S_{21} \in \{0, 1, \ldots, K-1\}$. Now, if we are given the value of Δ^S_{21}, then $\Delta^S_{32} \in \{0, 1, \ldots, (K-1) - \Delta^S_{21}\}$. The ET shift Δ^S_{32} is bounded by $(K-1) - \Delta_{21}$, as $\Delta^S_{31} = \Delta^S_{32} + \Delta^S_{21}$, and the shift between groups 1 and 3, Δ^S_{31}, is bounded by $K-1$. Therefore, if the ET shift between groups 1 and 2 is $\Delta^s_{21} = k$, then there are $((K-1)-k)+1 = K-k$ options for Δ^s_{32}. From this, you get $K + (K-1) + (K-2) + \cdots + 1 + 0 = \frac{K(K+1)}{2}$ ET Models and, by following the same procedure with the DLT shifts, $\frac{K(K+1)}{2}$ DLT Models.

2.5 Conclusion

In this chapter, we proposed a phase-I trial design for locating group-specific MTDs when toxicity is ordinal. This is the first design that considers using both group information and ordered toxicity. The proposed method combines the Quasi-CRM and the shift method. In the motivating example, there are three groups that at are partially ordered by dose sensitivity. In this partial ordering, Group 3 is the most sensitive and the order between groups 1 and 2 is unknown. A reversal occurs if the estimated MTD for Group 3 is larger than the estimated MTD for either groups 1 or 2. In simulations, the proposed method was compared to Independent Quasi-CRM trials and the shift method. The proposed method avoided reversals and performed better than its competitors at allocating the group-specific MTDs to patients in the trial and recommending the group-specific MTDs at the end of the trial.

Future areas of research include extending the proposed method to account for patients having multiple toxicities and considering time-to-event toxicity. For instance, a patient could have been observed with both grade 2 and grade 3 toxicities. The proposed method only considers the maximum observed toxicity, in this case, the grade 3 toxicity. The proposed method could be extended to differentiate between different types of toxicities, for instance, renal and hematological toxicities, using methods from Monia Ezzalfani et al. 2013. In this paper, a "total toxicity profile" is computed for each patient as a sum of ET scores over neurological, renal, and hematological toxicities. While Monia Ezzalfani et al. 2013 used toxicity type to provide a more complete toxicity profile, this paper does not consider multiplicity of a single toxicity type, only considering the maximum toxicity grade from each toxicity type. For instance, the total toxicity profile would not differentiate a patient with

two neurological toxicities of grades 2 and 3, from a patient with two neurological toxicities of grade 3, only considering the maximum neurological toxicity (grade 3) from both patients. Extending the proposed method to consider the multiplicity of a single toxicity type is not straightforward. If multiple toxicities are considered for each toxicity type, the total toxicity profile is no longer bounded. In comparison, the total toxicity profile in Monia Ezzalfani et al. 2013 is bounded by the total toxicity profile from a patient with grade four toxicities across all toxicity types. Bounding the total toxicity profile is crucial since a bounded total toxicity profile can be normalized in the unit interval, allowing the Quasi-Bernoulli likelihood, and thus the Quasi-CRM, to be applied.

In the proposed method, as we use this highest observed toxicity in the window of observation, the last patient needs to be completely followed through this window before the next patient is assigned to a dose. This requirement can be lifted by weighting each observation by the amount of time observed over the window of observation, as in the time-to-event continual reassessment method (TITE-CRM) from Y. K. Cheung and Rick Chappell 2000.

Chapter 3

A Model-Assisted Design for Partially or Completely Ordered Groups

3.1 Introduction

This chapter is adapted from Celum and M. Conaway 2024.

In this chapter, we propose a phase I trial design for locating group-specific MTDs. Specifically, the proposed design is a model-assisted design for locating group-specific MTDs under a complete or partial ordering. While model-based designs have been proposed for estimating group-specific MTDs under a partial ordering, to the best of our knowledge, a model-assisted design has not be proposed for this situation. Model-assisted designs include the Bayesian optimal interval design (BOIN) (Liu and Y. Yuan 2015), cumulative cohorts design (Ivanova, Flournoy, and Chung 2007), and keyboard design (Yan, Mandrekar, and Y. Yuan 2017). Additionally, several extensions of BOIN have been developed, including BOIN for toxicity grades, called gBOIN (Mu et al. 2019); BOIN for a trade-off utility between toxicity and efficacy, called U-BOIN (Zhou, Lee, and Y. Yuan 2019); and BOIN for drug combinations (Lin and

Yin 2017). The proposed method will be called group averaged BOIN (GAB), as we will use group averaging to modify the BOIN recommendations. Model-assisted designs are often appealing to clinicians, as they are often simpler than model-based designs. Previous dose allocation methods for partially ordered groups require computing posterior probabilities or maximizing likelihoods. In comparison, allocation under the proposed method can be implemented using a spreadsheet and calculated by hand, providing clinicians with a method that is simple and easily understood. Additionally, the proposed method performs similarly to the model-based designs for partially ordered groups, thus, reducing the complexity with no performance cost. Finally, the proposed method has almost sure convergence properties, making this the first design for partially ordered groups with such results.

The rest of the chapter will be organized as follows. Section 3.2 covers the GAB for two ordered groups. Section 3.3 covers the GAB for any partial order. Section 3.4 covers the asymptotic properties of GAB. Section 3.5 covers dose-elimination rules. Section 3.6 covers the simulation studies, comparing GAB to OR-CRM and parallel BOIN trials. Section 3.7 discusses future areas of research and concludes.

3.2 GAB for Two Groups

3.2.1 BOIN in Parallel

A naive approach for dose-finding in groups is conducting parallel trials for each group using BOIN. This approach has the issue of possible reversals, when a more sensitive group is assigned a higher dose than a less sensitive group. Horton, O'Quigley, and M. R. Conaway 2019 investigates topic of reversals by computing the reversal per-

centage for parallel CRM and BOIN trials. By adapting BOIN, GAB avoids reversals and increases dose selection accuracy. Simulations studies show that by borrowing information across groups, GAB allocates patients to the MTD more often within the trial and selects the MTD more often at the end of the trial.

3.2.2 Notation

We will now consider GAB for two ordered groups and extend the notation in Section 1.2.2 by using $g = 1, 2$ to index the groups. For group g and dose level k, let y_{gk} denote the number of patients with a DLT, n_{gk} the total number of patients, $\hat{\pi}_{gk} = y_{gk}/n_{gk}$ the DLT rate, and π_{gk} the DLT probability. Group 1 is less sensitive than Group 2, which means that $\pi_{1k} \leq \pi_{2k}$, for all k. From this, at any dose, the probability of a DLT is higher for Group 2 than for Group 1. The MTD for group g is the dose $d_{k'}$, where $k' = \text{argmin}_k|\pi_{gk} - \theta|$.

3.2.3 Allocation Run-in

Allocation begins with a "run-in" stage until both groups have been observed. We make no assumptions about the flow of the patients based on group membership. If the first few patients are in Group 2, we proceed with standard BOIN allocation in Group 2 until we need to allocate to a Group 1 patient. When the first patient in Group 1 arrives, their assigned dose will be the current dose of Group 2. Conversely, if the first few patients are in Group 1, we proceed with the standard BOIN allocation in Group 1. When the first patient in Group 2 arrives, they will be assigned the lowest dose. The run-in procedure uses the known group ordering to increase the dose recommendation for Group 1 to the current dose for Group 2.

3.2.4 Allocation

After observing patients from both groups, GAB first applies BOIN and then, if necessary, adjusts the current dose recommendations using pooled averages, making dose recommendations agree with the group order. BOIN can cause the recommendation to disagree with the known order in two ways: first, if Group 2 escalates above Group 1, and second, if Group 1 de-escalates below Group 2.

If Group 2 escalates above Group 1, GAB takes a pooled average at the dose Group 2 escalated from. Note that the dose Group 2 escalated from is the current Group 1 recommendation. If the pooled average is at or below λ_e, Group 1 escalates to the current Group 2 recommendation; otherwise, Group 2 de-escalates to the current Group 1 recommendation.

If Group 1 de-escalates below Group 2, GAB takes a pooled average at the dose Group 1 de-escalated from. Note that the dose Group 1 de-escalated from is the current Group 2 recommendation. If the pooled average is below λ_d, Group 1 escalates to the current Group 2 recommendation; otherwise, Group 2 de-escalates to the current Group 1 recommendation. The algorithm is illustrated in Figure 3.1, where "Pooled" is pooled the average, g is the group index from the last patient, and j_g is the current recommendation for group g.

3.2.5 Dose Selection

After all patients have been observed, the group-specific MTDs are estimated using bivariate isotonic regression. Before conducting the bivariate isotonic regression

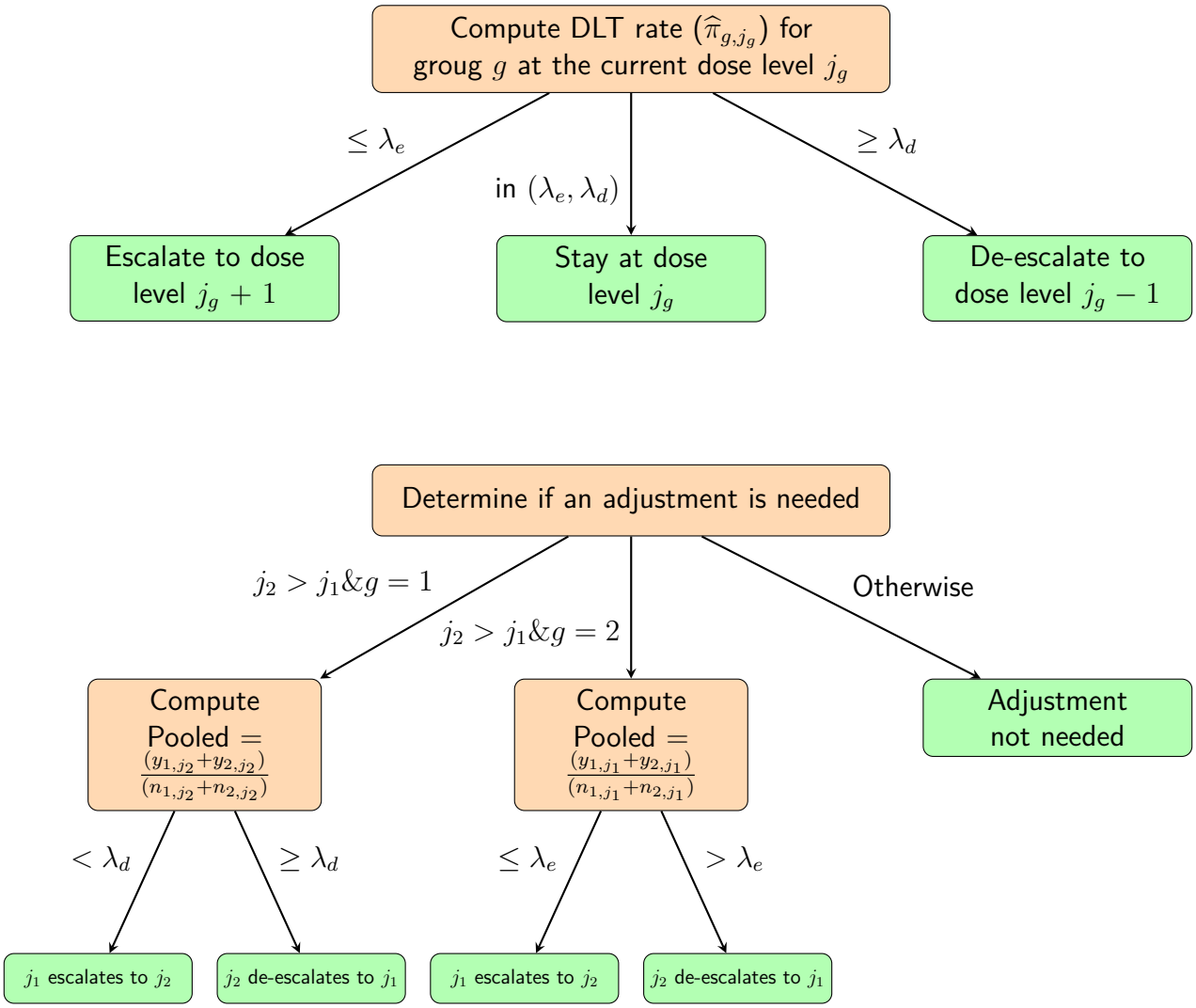


Figure 3.1: GAB Allocation Procedure

algorithm, the observed proportions are smoothed as

$$\hat{\pi}^s_{g,k} = \frac{y_{g,k} + \alpha_{g,k}}{n_{g,k} + \alpha_{g,k} + \beta_{g,k}},$$

where $\alpha_{g,k}, \beta_{g,k}$ are positive smoothing parameters. In simulations, we used the parameters $\alpha_{g,k} = \beta_{g,k} = 0.05$. Additionally, we used the weights $n_{gk} + 1$ for the pooled averages used in bivariate isotonic regression. The R package "BOIN" (Yan, Zhang,

et al. 2020) uses similar weights and smoothing parameters for the bivariate isotonic regression step in BOIN for drug combinations. For a detailed discussion on bivariate isotonic regression, see Dykstra and Robertson 1982, and to implement bivariate isotonic regression, use the R package "Iso" (Turner 2020).

Let $\tilde{\pi}_{gk}$ denote the estimates after conducting bivariate isotonic regression. Let A_g denote the admissible dose levels for group g, the doses that can be the MTD for group g. As Group 2 is the sensitive group, the admissible dose levels are those with Group 2 observations, that is, $A_2 = \{k : n_{2k} \neq 0\}$. As Group 1 is the less sensitive group, the admissible dose levels are those with an observation from either group, that is, $A_1 = \{k : n_{2k} \neq 0 \text{ or } n_{1k} \neq 0\}$. Let k'_g denote the MTD estimate for group g, where $k'_g = \text{argmin}_{k \in A_g} |\tilde{\pi}_{gk} - \theta|$. If there are ties for $\tilde{\pi}_{gk'_g}$, as can be the case in isotonic regression, if $\tilde{\pi}_{gk'_g} < \theta$, the largest such k'_g is selected, and if $\tilde{\pi}_{gk'_g} > \theta$, the smallest such k'_g is selected.

3.3 General GAB

3.3.1 Notation and Group Bundles

In this section, we discuss GAB for any partial or complete order. Before introducing generalized GAB, new notation will be introduced to generalize partial orders between groups, allowing us to generalize GAB and prove convergence properties. We will introduce "bundles" of groups to generalize all partial orders described in M. R. Conaway 2017b; M. R. Conaway 2017a; Horton, Wages, and M. R. Conaway 2019; and Horton, O'Quigley, and M. R. Conaway 2019.

We consider bundles of groups $b = 1, 2, \ldots B$, and within bundle b, we have groups

$g = 1, 2, \ldots, G_b$. The tuple (b, g) denotes group g in bundle b. Groups within the same bundle, say (b, g_1) and (b, g_2), do not have a prior ordering. That is, prior to the trial, it is not known if $\pi_{(b,g_1),k} \geq \pi_{(b,g_2),k}$ or $\pi_{(b,g_1),k} \leq \pi_{(b,g_2),k}$, where $\pi_{(b,g),k}$ denotes the DLT probability at dose level k for group g in bundle b. Across bundles, the ordering is known where groups from higher bundles are more sensitive. If $b_1 \leq b_2$, then for any $g_1 \in \{1, 2, \ldots, G_{b_1}\}$ and any $g_2 \in \{1, 2, \ldots, G_{b_2}\}$, we have $\pi_{(b_1,g_1),k} \leq \pi_{(b_2,g_2),k}$.

The grid below displays the layout of the bundles and groups where the different bundles are given by the rows. Note that rows do not need to be the same length as the amount of groups in each bundle can, and often do, differ.

$$\begin{array}{cccccc}
(1,1) & (1,2) & \cdots & \cdots & (1,G_1) & \\
(2,1) & (2,2) & \cdots & (2,G_2) & & \\
(3,1) & (3,2) & \cdots & \cdots & \cdots & (3,G_3) \\
\vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\
(B,1) & (B,2) & \cdots & \cdots & (B,G_B) &
\end{array}$$

We consider the partial orders in Horton, O'Quigley, and M. R. Conaway 2019 to demonstrate how these partial orders fit into the bundle framework. Horton, O'Quigley, and M. R. Conaway 2019 considered four orders: a complete order, loop order, twig order, and simple tree. Table 3.1 displays these orders, with the sensitivity level increasing from left to right, and converts these partial orders into the bundle framework in the "Bundles/Groups" column. To illustrate how the partial orders fit into the bundle framework, consider the loop ordering. As Group 1 is known to be the least sensitive group, Group 1 is the only group in Bundle 1. Groups 2 and 3 are known to be more sensitive than Group 1 and less sensitive than Group 4. Additionally, there is no known ordering between these groups; consequently, both groups are

in Bundle 2. Group 4 is known to be the most sensitive group; thus, Group 4 is in Bundle 3.

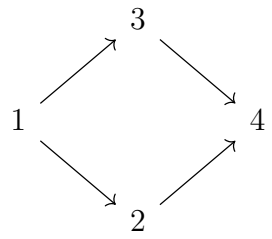
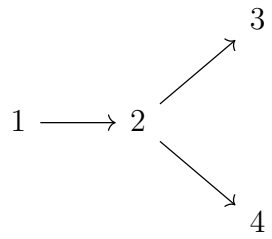
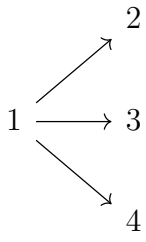
Ordering	Least Sensitive → Most Sensitive	Bundles/Groups
Complete	$1 \longrightarrow 2 \longrightarrow 3 \longrightarrow 4$	$1 = (1, 1)$ $2 = (2, 1)$ $3 = (3, 1)$ $4 = (4, 1)$
Loop		$1 = (1, 1)$ $2 = (2, 1) \quad 3 = (2, 2)$ $4 = (3, 1)$
Twig		$1 = (1, 1)$ $2 = (2, 1)$ $3 = (3, 1) \quad 4 = (4, 1)$
Simple Tree		$1 = (1, 1)$ $2 = (2, 1) \quad 3 = (2, 2) \quad 4 = (2, 3)$

Table 3.1: Orders from Horton, O'Quigley, and M. R. Conaway 2019 converted to the bundle/group format in the "Bundles/Groups" column.

3.3.2 Allocation Run-in

As in the two-group scenario, GAB begins allocation with a run-in stage. The first patient in group g and bundle b will be assigned to the lowest dose if there have not

been any observed patients in more sensitive groups; that is, groups in higher bundles. If there are observations from groups in higher bundles, the first patient in (b, g) will be assigned to the highest recommended dose for a group in a higher bundle. Suppose the next patient is the first patient from group g in bundle b and let $n_{(b',g')}$ denote the number of observed patients from group g' in bundle b'. Let $j_{(b',g')}$ denote the current recommendation for group g' in bundle b'. Figure 3.2 provides the run-in procedure.

if $n_{(b',g')} = 0, \forall b' > b, g' \in \{1, 2, \ldots, G_{b'}\}$ **then**
    $j_{(b,g)} \leftarrow 1$
else
    $j_{(b,g)} \leftarrow \max\{j_{(b',g')} : b' > b, g' \in \{1, 2, \ldots, G_{b'}\}, n_{(b',g')} \neq 0\}$
end if

Figure 3.2: GAB Run-In

To illustrate this procedure, consider the loop ordering in Table 3.1. Suppose we have observed patients in $(1, 1)$, $(2, 1)$, and $(3, 1)$, giving

$$j_{(1,1)} = 3$$
$$j_{(2,1)} = 3 \quad j_{(2,2)} = -$$
$$j_{(3,1)} = 2$$

, where "$-$" denotes that a group has not been observed. If the next patient is from $(2, 2)$, this patient would be assigned the dose level $j_{(3,1)} = 2$.

3.3.3 Allocation

After all groups have observations, we proceed using BOIN and pooled averages to adjust BOIN when necessary. Similar to the two-group case, there are two ways that dose recommendations can contradict the known ordering. The first is when a more sensitive group escalates above a less sensitive group, and the second is when a

less sensitive group de-escalates below a more sensitive group. Letting (b, g) be the bundle and group of the last observed patient, the first contradiction occurs when $j_{(b,g)} > j_{(b',g')}$, where $b > b'$, and the second when $j_{(b,g)} < j_{(b',g')}$, where $b < b'$.

When the first contradiction occurs, GAB takes a pooled average at the previous dose level for (b, g), this being $j_{(b,g)} - 1$, pooling over (b, g) and all the less sensitive groups currently at $j_{(b,g)} - 1$. For bundle and group $(b', g_{b'})$, and dose level k, let $y_{(b',g_{b'}),k}$ denote the number of DLTs and $n_{(b',g_{b'}),k}$ denote the number of observed patients. Thus,

$$\text{Pooled} = \frac{y_{(b,g),j_{(b,g)}-1} + \sum_{b'<b} \sum_{g_{b'}=1}^{G_{b'}} I\left(j_{(b',g_{b'})} = j_{(b,g)} - 1\right) y_{(b',g_{b'}),j_{(b,g)}-1}}{n_{(b,g),j_{(b,g)}-1} + \sum_{b'<b} \sum_{g_{b'}=1}^{G_{b'}} I\left(j_{(b',g_{b'})} = j_{(b,g)} - 1\right) n_{(b',g_{b'}),j_{(b,g)}-1}}. \quad (3.1)$$

Dose recommendations are adjusted as follows:

1. If Pooled $\leq \lambda_e$:
   (a) $j_{(b,g)}$ stays at $j_{(b,g)}$
   (b) For all $(b', g_{b'})$, with $b' < b$ and $j_{(b',g_{b'})} = j_{(b,g)} - 1$, escalate $j_{(b',g_{b'})}$ to $j_{(b,g)}$.
2. If Pooled $> \lambda_e$:
   (a) $j_{(b,g)}$ de-escalates to $j_{(b,g)} - 1$
   (b) Other dose recommendations remain the same.

When the second contradiction occurs, GAB takes a pooled average at the previous dose level for (b, g), this being $j_{(b,g)} + 1$, pooling over (b, g) and all the more sensitive

groups currently at $j_{(b,g)}+1$. Thus,

$$\text{Pooled} = \frac{y_{(b,g),j_{(b,g)}+1} + \sum_{b'>b} \sum_{g_{b'}=1}^{G_{b'}} I\left(j_{(b',g_{b'})} = j_{(b,g)}+1\right) y_{(b',g_{b'}),j_{(b,g)}+1}}{n_{(b,g),j_{(b,g)}+1} + \sum_{b'>b} \sum_{g_{b'}=1}^{G_{b'}} I\left(j_{(b',g_{b'})} = j_{(b,g)}+1\right) n_{(b',g_{b'}),j_{(b,g)}+1}}. \quad (3.2)$$

Dose recommendations are adjusted as follows:

1. If Pooled $< \lambda_d$:
   (a) $j_{(b,g)}$ escalates back to $j_{(b,g)}+1$
   (b) Other dose recommendations remain the same.
2. If Pooled $\geq \lambda_d$:
   (a) $j_{(b,g)}$ remains the same.
   (b) For all $(b', g_{b'})$, with $b' > b$ and $j_{(b',g_{b'})} = j_{(b,g)}+1$, de-escalate $j_{(b',g_{b'})}$ to $j_{(b,g)}$.

Pooled corrects dose recommendations by using the known group ordering to determine if we should de-escalate or escalate. The algorithm is illustrated in Figure 3.3, where (b, g) is the bundle and group index of the last patient and $j_{(b,g)}$ is the current recommendation for group g in bundle b.

3.3.4 Dose Selection

After observing all patients, DLT estimates are obtained by applying bivariate isotonic regression to the complete orders that are possible given the known partial order. Estimates with the maximum likelihood are used for dose selection.

As in the two-group scenario, observed proportions are smoothed, so

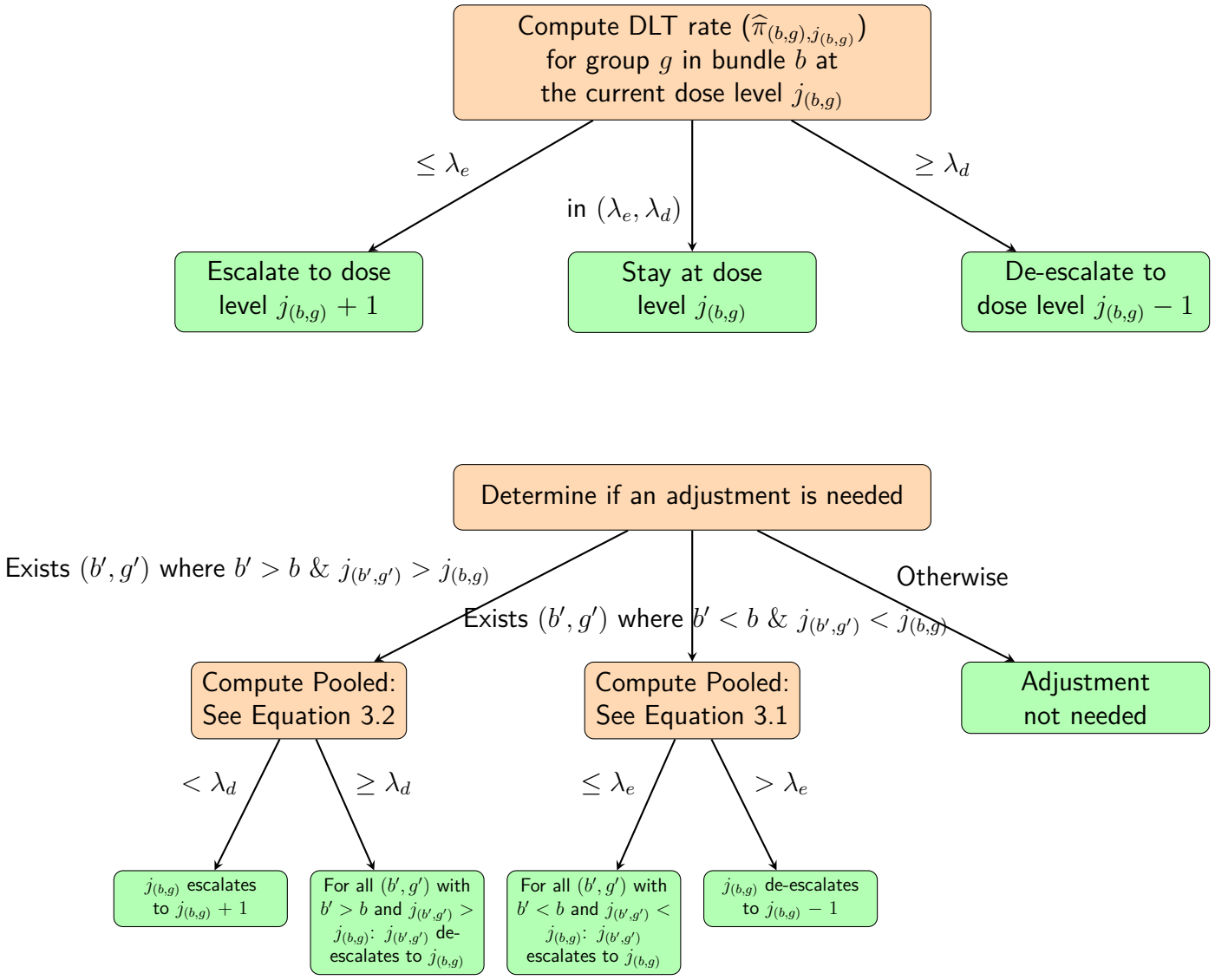


Figure 3.3: General GAB Allocation Procedure

$$\hat{\pi}^{s}_{(b,g),k} = \frac{y_{(b,g),k} + \alpha_{(b,g),k}}{n_{(b,g),k} + \alpha_{(b,g),k} + \beta_{(b,g),k}}.$$

We use the same smoothing parameters and weights used in the two-group scenario.

Now, we consider all possible complete orders from the known bundle structure.

All complete orders can be obtained by considering all combinations of permutations within each bundle. In general, as there are $G_1! \times G_2! \times \cdots \times G_b!$ different combinations of permutations, there are $M = G_1! \times G_2! \times \cdots \times G_b!$ different models. Let ρ_b be a permutation on $\{1, 2, \ldots, G_b\}$, permuting the groups within bundle b, and $\rho_b^{-1}(g_b)$ be the inverse image of $\rho_b(g_b)$. The M models can be obtained from all combinations, $(\rho_1, \rho_2, \ldots, \rho_B)$, returning a complete order

$$\pi_{(1,\rho_1^{-1}(1)),k} \leq \pi_{(1,\rho_1^{-1}(2)),k} \leq \cdots \leq \pi_{(1,\rho_1^{-1}(G_1)),k} \leq \pi_{(2,\rho_2^{-1}(1)),k} \leq \cdots \leq \pi_{(2,\rho_2^{-1}(G_2)),k} \leq \cdots \leq \pi_{(B,\rho_B^{-1}(1)),k} \leq \cdots \leq \pi_{(B,\rho_B^{-1}(G_B)),k}.$$

From the loop ordering, there are two combinations of permutations, $\{(1), (1,2), (1)\}$ and $\{(1), (2,1), (1)\}$, giving the complete orders $\pi_{(1,1),k} \leq \pi_{(2,1),k} \leq \pi_{(2,2),k} \leq \pi_{(3,1),k}$ and $\pi_{(1,1),k} \leq \pi_{(2,2),k} \leq \pi_{(2,1),k} \leq \pi_{(3,1),k}$. In practice, it is simple to find all possible models given the small number of groups.When considering the orders in Table 3.1, we notice the simple tree has the most models, totaling $1! \times 3! = 6$ models.

Now, given a complete order obtained from $(\rho_1, \rho_2, \ldots, \rho_B)$, we apply bivariate isotonic regression to the matrix order below.

$$\begin{array}{cccc}
\hat{\pi}^s_{(1,\rho_1^{-1}(1)),1} & \hat{\pi}^s_{(1,\rho_1^{-1}(1)),2} & \cdots & \hat{\pi}^s_{(1,\rho_1^{-1}(1)),K} \\
\vdots & \vdots & \ddots & \vdots \\
\hat{\pi}^s_{(1,\rho_1^{-1}(G_1)),1} & \hat{\pi}^s_{(1,\rho_1^{-1}(G_1)),2} & \cdots & \hat{\pi}^s_{(1,\rho_1^{-1}(G_1)),K} \\
\hat{\pi}^s_{(2,\rho_2^{-1}(1)),1} & \hat{\pi}^s_{(2,\rho_2^{-1}(1)),2} & \cdots & \hat{\pi}^s_{(2,\rho_2^{-1}(1)),K} \\
\vdots & \vdots & \ddots & \vdots \\
\hat{\pi}^s_{(2,\rho_2^{-1}(G_2)),1} & \hat{\pi}^s_{(2,\rho_2^{-1}(G_2)),2} & \cdots & \hat{\pi}^s_{(2,\rho_2^{-1}(G_2)),K} \\
\vdots & \vdots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \vdots \\
\hat{\pi}^s_{(B,\rho_B^{-1}(1)),1} & \hat{\pi}^s_{(B,\rho_B^{-1}(1)),2} & \cdots & \hat{\pi}^s_{(B,\rho_B^{-1}(1)),K} \\
\vdots & \vdots & \ddots & \vdots \\
\hat{\pi}^s_{(B,\rho_B^{-1}(G_B)),1} & \hat{\pi}^s_{(B,\rho_B^{-1}(G_B)),2} & \cdots & \hat{\pi}^s_{(B,\rho_B^{-1}(G_B)),K}
\end{array} \tag{3.3}$$

For the loop order, we apply bivariate isotonic regression to the matrices below.

$$\begin{matrix} \hat{\pi}^s_{(1,1),1} & \hat{\pi}^s_{(1,1),2} & \cdots & \hat{\pi}^s_{(1,1),K} \\ \hat{\pi}^s_{(2,1),1} & \hat{\pi}^s_{(2,1),2} & \cdots & \hat{\pi}^s_{(2,1),K} \\ \hat{\pi}^s_{(2,2),1} & \hat{\pi}^s_{(2,2),2} & \cdots & \hat{\pi}^s_{(2,2),K} \\ \hat{\pi}^s_{(3,1),1} & \hat{\pi}^s_{(3,1),2} & \cdots & \hat{\pi}^s_{(3,1),K} \end{matrix} \qquad \begin{matrix} \hat{\pi}^s_{(1,1),1} & \hat{\pi}^s_{(1,1),2} & \cdots & \hat{\pi}^s_{(1,1),K} \\ \hat{\pi}^s_{(2,2),1} & \hat{\pi}^s_{(2,2),2} & \cdots & \hat{\pi}^s_{(2,2),K} \\ \hat{\pi}^s_{(2,1),1} & \hat{\pi}^s_{(2,1),2} & \cdots & \hat{\pi}^s_{(2,1),K} \\ \hat{\pi}^s_{(3,1),1} & \hat{\pi}^s_{(3,1),2} & \cdots & \hat{\pi}^s_{(3,1),K} \end{matrix}$$

Let $m = 1, 2, \ldots, M$ denote the models obtained through the combinations of permutations and $\tilde{\pi}^m_{(b,g_b),k}$ denote the estimates obtained from bivariate isotonic regression on model m. We select the model that maximizes the likelihood, letting $m' = \text{argmax}_m \, L\left(\tilde{\pi}^m_{(b,g_b),k}\right)$, where $L\left(\tilde{\pi}^m_{(b,g_b),k}\right)$ is the likelihood evaluated over estimates $\tilde{\pi}^m_{(b,g_b),k}$.

Similar to the two-group scenario, the set of admissible dose levels for (b, g) is given by

$$A_{(b,g)} = \left\{k : n_{(b',g_{b'}),k} \neq 0, \text{ for some } (b', g_{b'}) \text{ where } b' > b, \text{ or } n_{(b,g),k} \neq 0\right\}. \qquad (3.4)$$

The MTD for (b, g) is estimated as the dose level $k'_{(b,g)}$, where $k'_{(b,g)} = \text{argmin}_{k \in A_{(b,g)}} |\tilde{\pi}^{m'}_{(b,g_b),k} - \theta|$. If there are ties for $\tilde{\pi}^{m'}_{(b,g_b),k'_{(b,g)}}$, if $\tilde{\pi}^{m'}_{(b,g_b),k'_{(b,g)}} < \theta$, the largest such $k'_{(b,g)}$ is selected, and if $\tilde{\pi}^{m'}_{(b,g_b),k'_{(b,g)}} > \theta$, the smallest such $k'_{(b,g)}$ is selected.

3.4 Asymptotic Properties

This section covers the asymptotic allocation properties of the GAB design. As allocation under BOIN has almost sure convergence properties, we would expect such

properties from GAB. Theorem 3.1 restates the convergence properties of BOIN (Liu and Y. Yuan 2015; Oron, Azriel, and Hoff 2011).

Theorem 3.1. *Dose allocation using BOIN converges almost surely to a dose level k, if $\pi_k \in (\lambda_e, \lambda_d)$, and k is the only dose level with $\pi_k \in [\lambda_e, \lambda_d]$. If we have $\lambda_d \leq \pi_1$, dose allocation will converge almost surely to dose level 1 and if we have $\lambda_e \geq \pi_K$, dose allocation will converge almost surely to dose level K. If we have no dose with DLT probability in (λ_e, λ_d) but $\theta \in [\pi_1, \pi_K]$, then we will have almost sure oscillation between the two doses straddling the interval. If there are multiple doses with $\pi_k \in (\lambda_e, \lambda_d)$, we have almost sure convergence to one of these doses.*

Lemma 3.2 is critical in proving almost sure convergence properties for GAB. This lemma states that doses visited infinitely often have observed DLT proportions converging to the true DLT probabilities. The proof of the lemma follows the same steps as Lemma 1 in Oron, Azriel, and Hoff 2011.

Lemma 3.2. *For all bundles and groups, (b, g), and doses d_k, $\hat{\pi}_{(b,g),k} \overset{a.s.}{\to} \pi_{(b,g),k}$, given $n_{(b,g),k} \to \infty$*

Theorem 3.3 states that GAB has the same asymptotic properties as original BOIN, making GAB the first dose-finding method for groups that can be shown to converge almost surely. This theorem claims if a group has a unique dose with DLT probability in $[\lambda_e, \lambda_d]$, eventually these patients will only be allocated to this dose. Additionally, if a group has several doses with DLT probability in $[\lambda_e, \lambda_d]$, eventually, these patients will only be allocated to one of these doses. As a result, GAB will eventually allocate patients to an acceptable dose. Appendix B.1 provides proofs for Lemma 3.2 and Theorem 3.3.

Theorem 3.3. *For any bundle and group, (b, g), GAB dose allocation converges almost surely to a dose level k, if $\pi_{(b,g),k} \in (\lambda_e, \lambda_d)$ and k is the only dose level with $\pi_{(b,g),k} \in [\lambda_e, \lambda_d]$. If we have $\lambda_d \leq \pi_{(b,g),1}$ dose allocation will converge almost surely to dose level 1 and if we have $\lambda_e \geq \pi_{(b,g),K}$ dose allocation will converge almost surely to dose level K. If we have no dose with DLT probability in (λ_e, λ_d) but $\theta \in [\pi_{(b,g),1}, \pi_{(b,g),K}]$, then we will have almost sure oscillation between the two doses straddling the interval. If there are multiple doses with $\pi_{(b,g),k} \in (\lambda_e, \lambda_d)$, we have almost sure convergence to one of these doses.*

3.5 Practical Considerations

Similar to the original BOIN paper, we adapt our method for practical considerations. In the package "BOIN" (Yan, Zhang, et al. 2020), dose elimination boundaries can be enumerated before the trial. To limit the number of patients assigned to overly toxic doses, GAB dose elimination rules pool toxicity rates and use dose elimination boundaries, as in "BOIN". Instead of using a uniform Beta(1,1) prior as in Liu and Y. Yuan 2015, we use Jeffery's prior, Beta(0.5,0.5), as in Lin and Yin 2017. This prior corresponds to the information from one subject. Additionally, instead of using the cutoff $\lambda = 0.95$, we use $\lambda = 0.975$. A higher cutoff is selected due to the between-group dependency with dose elimination. For instance, if dose d_k is eliminated for (b, g), then d_k would also be eliminated for all groups more sensitive than (b, g). That is, all (b', g'), where $b' > b$. A sensitivity analysis is provided in Appendix B.5, analyzing the results from using a uniform prior compared to Jeffery's prior, and using $\lambda = 0.95$ compared to $\lambda = 0.975$. This sensitivity analysis shows that the proposed method is robust.

Let (b, g) denote the most recently observed group, with $j_{(b,g)}$ denoting the dose level this group was observed at. Before eliminating $j_{(b,g)}$, we require that at least three patients have been observed at $j_{(b,g)}$, that is $n_{(b,g),j_{(b,g)}} \geq 3$. Then, assuming $\pi_{(b,g),j_{(b,g)}} \sim \text{Beta}(0.5, 0.5)$, check if $P(\pi_{(b,g),j_{(b,g)}} > \theta | n_{(b,g),j_{(b,g)}}, y_{(b,g),j_{(b,g)}}) > \lambda$. From the Beta-Binomial model, $\pi_{(b,g),j_{(b,g)}} | n_{(b,g),j_{(b,g)}}, y_{(b,g),j_{(b,g)}} \sim Beta(0.5 + y_{(b,g),j_{(b,g)}}, 0.5 + n_{(b,g),j_{(b,g)}} - y_{(b,g),j_{(b,g)}})$. If the posterior probability exceeds this threshold, we then consider the groups below (b, g). Let n_{Pooled} and y_{Pooled} denote the pooled number of observations and number of DLTs at $j_{(b,g)}$ from (b, g) and more sensitive groups. From this, $n_{Pooled} = n_{(b,g),j_{(b,g)}} + \sum_{b'>b} \sum_{g_{b'}=1}^{G_{b'}} n_{(b',g_{b'}),j_{(b,g)}}$ and $y_{Pooled} = y_{(b,g),j_{(b,g)}} + \sum_{b'>b} \sum_{g_{b'}=1}^{G_{b'}} y_{(b',g_{b'}),j_{(b,g)}}$. If $P(Beta(0.5 + y_{pooled}, n_{pooled} - y_{pooled} + 0.5) > \theta) > \lambda$, we eliminate dose level $j_{(b,g)}$ from consideration in group (b, g) and all groups (b', g'), where $b' > b$. If the lowest dose is eliminated from a group, this group is removed from the study. In this adaption, we also require all doses in $A_{(b,g)}$ to be doses that have not been eliminated. Before the trial, elimination boundaries can be tabulated. For these boundaries, see Table 3.2 below.

Number of Patients Treated	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Elimination Boundary	NA	NA	3	3	3	4	4	5	5	5	6	6	6	6	7

Table 3.2: Dose elimination boundaries based on number of patients treated and number of DLTs. We require at least 3 patients to have been observed, use the prior Beta(0.5,0.5), and cutoff $\lambda = 0.975$.

In the supplementary materials of the online version of Celum and M. Conaway 2024, the R script "Boundaries.R" provides a function for calculating dose elimination boundaries, assisting in implementation. Additionally, Appendix B.4 provides illustrative examples of trials conducted using the dose elimination rules. For the remainder of this chapter, GAB-E and GAB will refer to the proposed method with and without dose elimination, respectively.

3.6 Simulations

This section evaluates the performance of GAB and GAB-E relative to parallel BOIN trials and OR-CRM. We let P-BOIN-E and P-BOIN denote parallel BOIN trials with and without dose elimination rules, respectively. OR-CRM from M. R. Conaway 2017b is a well performing model-based method for partially ordered groups. Thus, our competitor methods are model-assisted designs not accounting for groups (P-BOIN and P-BOIN-E), and a model-based design accounting for groups (OR-CRM).

Comparisons were drawn using simulations with two groups and three groups. In the simulations with two groups, the groups were completely ordered. In the simulations with three groups, three different orders were considered, as provided in Table 3.3. In this table, we drop bundle/group notation for simplicity. The "Known" column states the known DLT probability ordering across groups, while the "Unknown" column states what DLT probability orderings are unknown across groups. For instance, in the second ordering, we know that Group 1 is the least sensitive group while we do not know the relative sensitivity of groups 2 and 3.

Ordering	Known	Unknown
1	$\pi_{1,k} \leq \pi_{2,k} \leq \pi_{3,k}$	Nothing
2	$\pi_{1,k} \leq \pi_{2,k}$ and $\pi_{1,k} \leq \pi_{3,k}$	$\pi_{3,k} \leq \pi_{2,k}$ or $\pi_{2,k} \leq \pi_{3,k}$
3	$\pi_{1,k} \leq \pi_{3,k}$ and $\pi_{2,k} \leq \pi_{3,k}$	$\pi_{1,k} \leq \pi_{2,k}$ or $\pi_{2,k} \leq \pi_{1,k}$

Table 3.3: Three Group Orders

3.6.1 Generating Family of Curves

Dose-toxicity curves were randomly generated to avoid selecting curves that favored GAB. These curves were randomly generated using a method similar to that in M. R.

Conaway 2017b; M. R. Conaway 2017a and two curves were generated for each possible MTD configuration. Four and six-dose trials were considered. Table 3.4 provides the number of MTD configurations by number of groups, group ordering, and number of doses. Additionally, Appendix B.2 provides complete tables of all four-dose MTD configurations. Next, we will demonstrate how to generate curves for the three-group scenarios.

Groups	Ordering	# Doses	# Configurations	# Curves
2	C	4	10	20
2	C	6	21	42
3	1	4	20	40
3	1	6	56	112
3	2,3	4	30	60
3	2,3	6	91	182

Table 3.4: Number of MTD configurations and curves by number of groups and group ordering. A MTD configuration is a set of possible group-specific MTDs under the group ordering.

Let $(\gamma_1, \gamma_2, \gamma_3)$ denote the configuration for which we are generating a curve. As in Liu and Y. Yuan 2015, we generate three centers "close" to the DLT target, θ, which we denote as $(c_1, c_2, c_3) = (\Phi(\varepsilon_1), \Phi(\varepsilon_2), \Phi(\varepsilon_3))$, where $\varepsilon \sim N(z(\theta), (0.05)^2)$, $\Phi()$ is standard normal CDF, and $z()$ is the inverse of the standard normal CDF. Next, centers are as sorted as $(c_{[1]}, c_{[2]}, c_{[3]})$. As detailed below, sorted centers are assigned to the group curves using the group ordering.

1. Order 1: $(\pi_{\gamma_1}, \pi_{\gamma_2}, \pi_{\gamma_3}) = (c_{[1]}, c_{[2]}, c_{[3]})$

2. Order 2:

   (a) If $\gamma_2 < \gamma_3$, then $(\pi_{1,\gamma_1}, \pi_{2,\gamma_2}, \pi_{3,\gamma_3}) = (c_{[1]}, c_{[3]}, c_{[2]})$.

   (b) If $\gamma_2 > \gamma_3$, then $(\pi_{1,\gamma_1}, \pi_{2,\gamma_2}, \pi_{3,\gamma_3}) = (c_{[1]}, c_{[2]}, c_{[3]})$.

(c) If $\gamma_2 = \gamma_3$, only generate two centers, and then $(\pi_{1,\gamma_1}, \pi_{2,\gamma_2}, \pi_{3,\gamma_3}) = (c_{[1]}, c_{[2]}, c_{[2]})$.

3. Order 3:

(a) If $\gamma_1 < \gamma_2$, then $(\pi_{1,\gamma_1}, \pi_{2,\gamma_2}, \pi_{3,\gamma_3}) = (c_{[2]}, c_{[1]}, c_{[3]})$.

(b) If $\gamma_1 > \gamma_2$, then $(\pi_{1,\gamma_1}, \pi_{2,\gamma_2}, \pi_{3,\gamma_3}) = (c_{[1]}, c_{[2]}, c_{[3]})$.

(c) If $\gamma_1 = \gamma_2$, only generate two centers, and then $(\pi_{1,\gamma_1}, \pi_{2,\gamma_2}, \pi_{3,\gamma_3}) = (c_{[1]}, c_{[1]}, c_{[2]})$.

Appendix B.3 illustrates how to generate centers. The algorithm for computing the rest of the probabilities follows the algorithm found in M. R. Conaway 2017a and M. R. Conaway 2017b. Three requirements are added to eliminate unreasonable curves. First, the lowest DLT probability is not below 0.01. Second, DLT probabilities within a group differ by at least 0.02 and 0.03 in the four-dose and six-dose scenarios, respectively. Third, DLT probabilities within a group do not differ by more than 0.15. Readers interested in generating curves, see the R script "Generate_Curves.R" in the online version of Celum and M. Conaway 2024.

Figures 3.4 and 3.5 plot a subset of representative curves, plotting six curves for each scenario (partial order). For each scenario, K-means clustering was applied to the collection of curves, using the cluster means as the representative curves in the plots. K-means clustering provides curves with varied shapes and differing group separation.

3.6.2 Results

For each scenario, we ran 1000 simulations for each curve and considered three sample sizes, with the sample size depending on the number of doses and patients. Group membership was randomly generated with equal group probabilities while adding the requirement that a trial has at least one patient from each group. The statistics

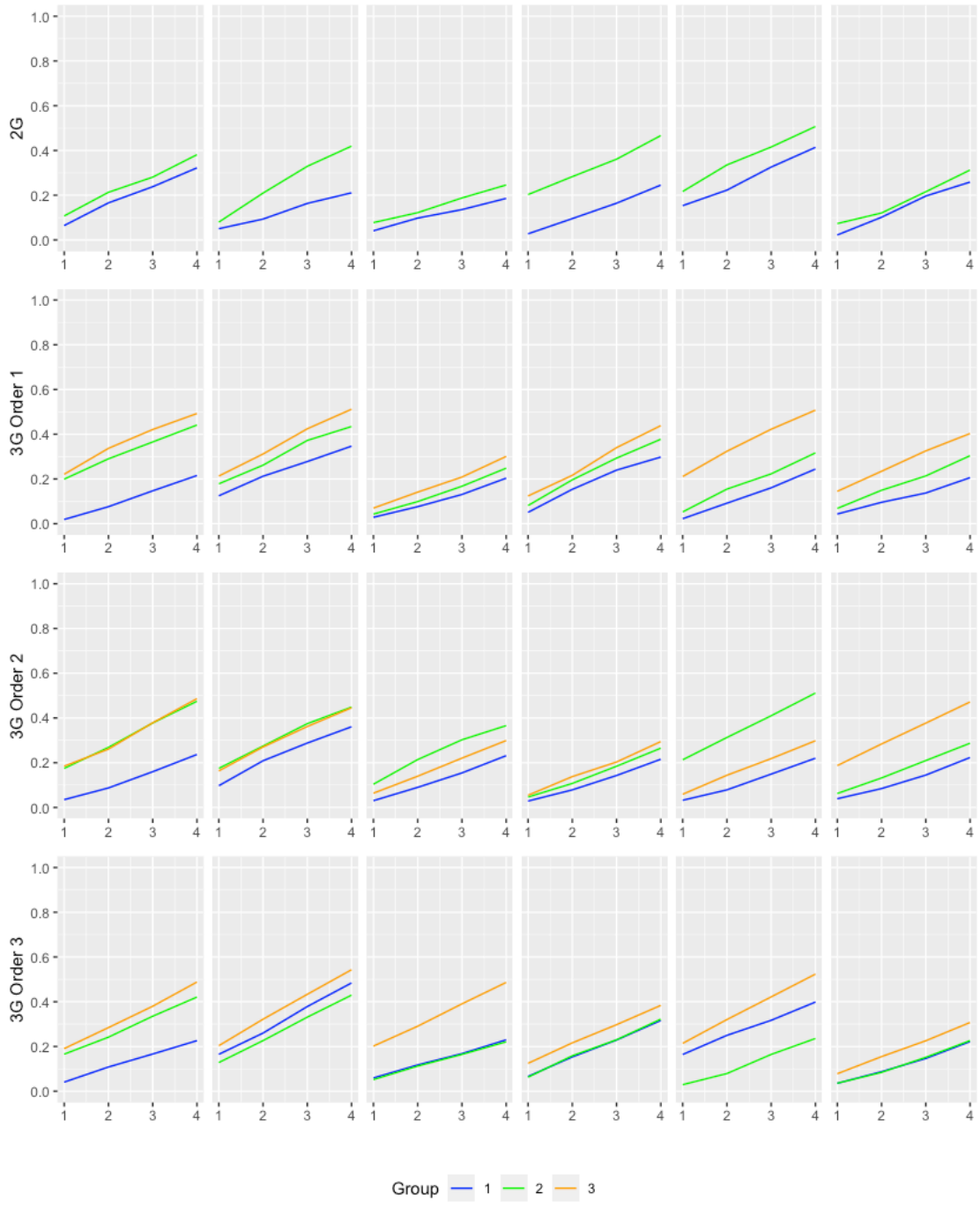


Figure 3.4: K-Means Centers for the Four-Dose Curves, providing "Representative" Curves. "Representatives" are provided for the two-group case and the three different three-group orders.

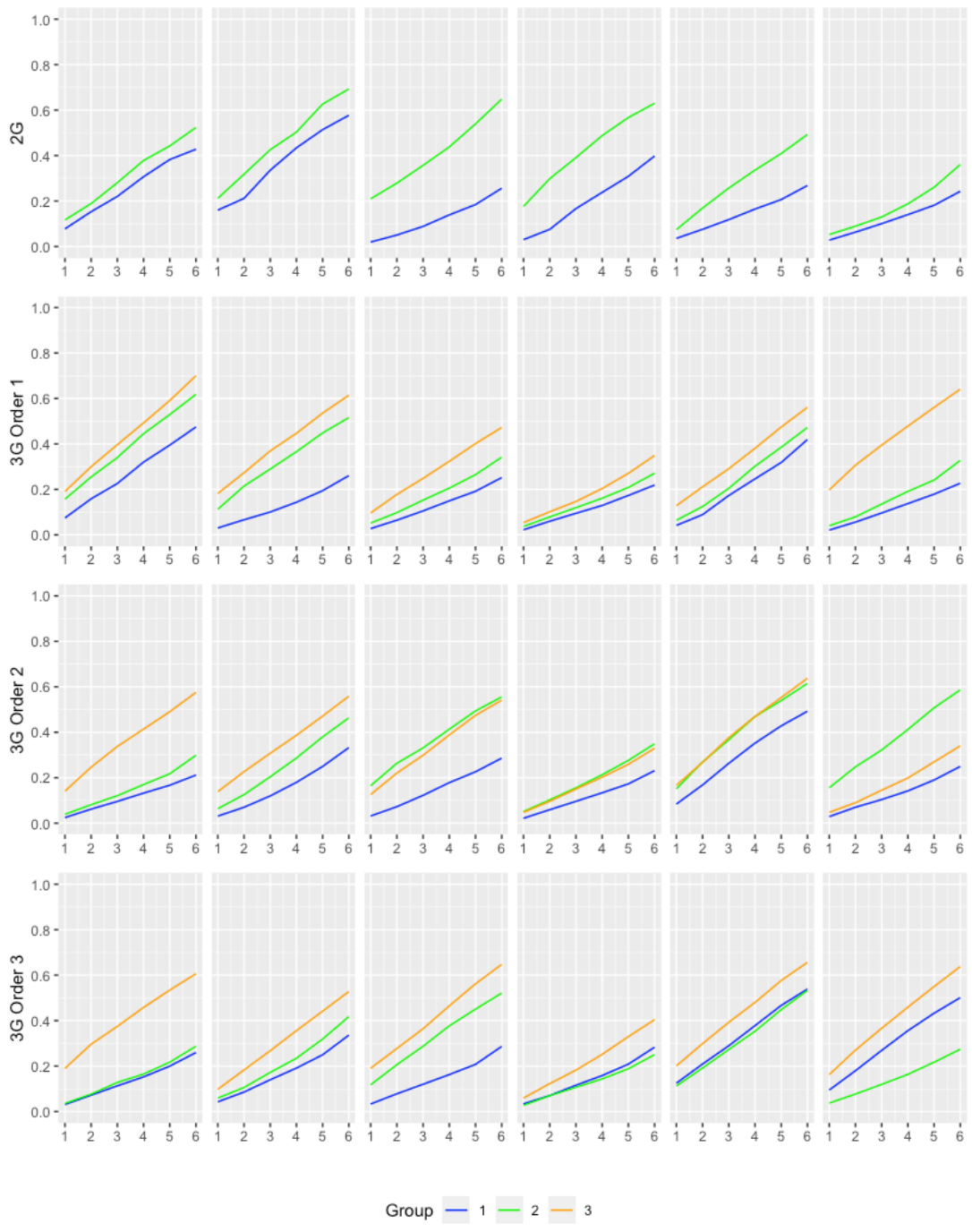


Figure 3.5: K-Means Centers for the Six-Dose Curves, providing "Representative" Curves. "Representatives" are provided for the two-group case and the three different three-group orders.

PCS and AI, and PCA and Int were calculated for end-of-trial selection and within-trial allocation, respectively. These statistics were averaged across the groups. The statistic Int is the percentage of patients allocated to doses with DLT probabilities in the interval (λ_e, λ_d), PCA/PCS is the percentage of times allocation/selection occurred at the MTD, and AI is the accuracy index, as given in Y. K. Cheung 2011, measuring accuracy using the entire selection distribution. The accuracy index is given by $\text{AI} = 1 - K\frac{\sum_{k=1}^{K} \rho_{g,k} \times \text{Probability of selecting dose level k}}{\sum_{k=1}^{K} \rho_{g,k}}$, where K denotes the number of doses under consideration and $\rho_{g,k} = |\pi_{g,k} - \theta|$ denotes the distance of the DLT probability at dose level k for group g from the DLT target. Tables 3.5 and 3.6 provide selection and allocation results, respectively. Both GAB and GAB-E significantly outperformed P-BOIN and P-BOIN-E in both allocation and selection, demonstrating the benefit of using pooled averages. GAB and OR-CRM performed similarly in end-of-trial selection as GAB chose the correct dose more often and OR-CRM had a higher AI more often. OR-CRM slightly outperformed GAB and GAB-E in allocation. There is less of a discrepancy between these methods for the Int metric since GAB and GAB-E allocate to doses within an interval, some of which may not be the MTD. We see that GAB outperformed GAB-E in dose selection, possibly due to GAB-E's ability to terminate trials early. Conversely, GAB-E outperformed GAB in dose allocation, possibly due to GAB-E eliminating overly toxic doses and narrowing down the available doses. It is important to note that trials resulting in an early termination did not contribute to the AI for GAB-E and P-BOIN-E. These trials could not be used to compute the AI since the AI uses the distance of the selected dose from the DLT target. Standard Errors are not included in the results below since, for each row of results, there are several curves with 1,000 simulations per curve, making the standard errors minuscule.

Selection													
				PCS					AI				
Groups	Scenario	Doses	Patients	GAB	GAB-E	P-BOIN	P-BOIN-E	OR-CRM	GAB	GAB-E	P-BOIN	P-BOIN-E	OR-CRM
2	C	4	30	**50**	47.8	45.9	44.2	49	**45.8**	44.3	39.3	38.4	43.8
2	C	4	50	**55.3**	52	52	49.6	55	**53.1**	50.8	48.3	46.8	52.5
2	C	4	70	59.7	55.7	56.7	52.8	**59.8**	58.1	55.8	54.3	51.3	**58.5**
2	C	6	40	**41.3**	40.1	38.6	37.1	41	52.1	51.9	46.6	46	**52.9**
2	C	6	60	**45.9**	44.3	43.6	42.1	45.6	58.1	57.4	53.8	53.3	**58.9**
2	C	6	80	**50.4**	47.3	48.1	45.7	49.6	62.6	60.9	59.2	57.9	**62.9**
3	1	4	40	52.2	50.4	46.6	44.8	**52.6**	**48.3**	47.5	37.8	36.7	48.1
3	1	4	60	56.8	54.2	51.4	49	**57.1**	**53.9**	52.5	45	43.6	**53.9**
3	1	4	80	**60.4**	57.2	55.3	52.6	60	**57.9**	56.3	50.2	48.8	57.2
3	1	6	60	**44.1**	42.6	39.6	37.9	42.6	**55.7**	54.9	47.4	46.5	**55.7**
3	1	6	80	**47.6**	45.5	43.2	41.3	45.8	**59.6**	58.8	52.5	51.5	59
3	1	6	100	**50.3**	47.6	46.4	43.8	48.3	**62.5**	61.2	56.4	55.2	61.9
3	2	4	40	**50.5**	48.8	46.4	44.7	50.3	**45.2**	44.4	38.5	37.4	44.9
3	2	4	60	54.8	52.4	51	48.9	**55.1**	51.3	50	45.6	44.5	**51.5**
3	2	4	80	58.6	55.1	55	51.9	**58.7**	55.9	53.7	50.8	49	**56**
3	2	6	60	**43**	41.5	39.8	38.1	41.3	**53.1**	52.3	47.2	46.3	**53.1**
3	2	6	80	**46.7**	44.5	43.5	41.6	45	**57.4**	56.4	52.4	51.4	**57.4**
3	2	6	100	**49.7**	47	46.8	44.3	47.7	**60.8**	59.3	56.4	55.2	60.4
3	3	4	40	49.5	47.6	45.5	43.6	**50**	**45.5**	44.8	38.6	37.4	**45.5**
3	3	4	60	53.9	51.2	50.2	47.8	**54.5**	51.4	50.2	45.8	44.4	**51.8**
3	3	4	80	57.1	54	53.7	50.6	**57.9**	55.4	54	50.6	48.8	**56.4**
3	3	6	60	**43.1**	41.4	39.9	38.2	42.6	53	52.4	47.6	47	**54.2**
3	3	6	80	**46.6**	44.5	43.7	41.6	46.1	57.3	56.5	52.9	52	**58.4**
3	3	6	100	**49.5**	47	46.8	44.2	48.9	60.6	59.6	56.8	55.7	**61.4**
Averages				**50.71**	48.32	47.07	44.9	50.19	54.8	53.6	48.9	47.71	**54.85**

Table 3.5: Percentage correct selection (PCS) and Accuracy Index (AI) by the number of groups, scenario, number of doses, number of patients, and method.

Allocation													
				PCA					Int				
Groups	Scenario	Doses	Patients	GAB	GAB-E	P-BOIN	P-BOIN-E	OR-CRM	GAB	GAB-E	P-BOIN	P-BOIN-E	OR-CRM
2	C	4	30	40.8	40.6	36.7	37.2	**41.2**	49.4	49.1	44.6	45.1	**49.6**
2	C	4	50	44.1	44.2	40.3	40.6	**46.1**	53	53.1	48.6	48.9	**54.9**
2	C	4	70	46.7	46.8	43.1	43.5	**49.5**	55.8	56	51.5	51.9	**58.6**
2	C	6	40	31.9	32.2	28.9	29.2	**32.4**	38.3	38.7	34.8	35.2	**39.2**
2	C	6	60	34.7	35.5	31.8	32.5	**36**	41.6	42.5	38.3	39.1	**43.5**
2	C	6	80	37.4	37.9	34.7	35.3	**38.9**	44.7	45.2	41.5	42.2	**47**
3	1	4	40	41.6	41.7	37.3	37.3	**44.1**	44	44.1	39.4	39.4	**46.6**
3	1	4	60	44.1	44.5	40.1	40.5	**47.7**	46.6	47	42.2	42.7	**50.3**
3	1	4	80	46.4	47.1	42.5	43.1	**50**	48.9	49.7	44.7	45.3	**52.7**
3	1	6	60	33	33.4	29.5	29.7	**33.9**	41.1	41.5	36.5	36.8	**42.5**
3	1	6	80	35.1	35.6	31.7	32.1	**36.3**	43.6	44.2	39.2	39.6	**45.4**
3	1	6	100	36.9	37.5	33.7	34.1	**38.4**	45.7	46.3	41.4	42	**47.9**
3	2	4	40	41.6	**41.7**	37	37.2	41.6	**44.7**	**44.7**	39.8	39.9	44.3
3	2	4	60	44.3	44.7	39.9	40.2	**45.3**	47.6	47.9	42.8	43.1	**48.2**
3	2	4	80	46.5	46.8	42.2	42.6	**48.2**	49.8	50.1	45.2	45.6	**51.2**
3	2	6	60	33	**33.4**	29.5	29.7	32.2	39.9	**40.3**	35.7	35.9	39
3	2	6	80	35.3	**35.6**	31.7	32.2	35	42.5	**42.9**	38.3	38.8	42.3
3	2	6	100	37.2	**37.6**	33.8	34.2	37.2	44.7	**45.2**	40.6	41.1	44.8
3	3	4	40	41.1	41.3	37	37.1	**42.2**	46.8	46.8	41.8	41.9	**47.5**
3	3	4	60	43.7	44	39.7	40.1	**45.6**	49.4	49.8	44.8	45.1	**51.2**
3	3	4	80	45.8	46.2	41.8	42.3	**48.2**	51.6	52.1	47	47.5	**54**
3	3	6	60	33.1	33.5	29.8	30.1	**33.6**	39.8	40.2	35.6	36	**40.4**
3	3	6	80	35.4	35.8	32	32.5	**36.3**	42.4	42.9	38.3	38.8	**43.5**
3	3	6	100	37.2	37.7	34	34.5	**38.5**	44.5	45.1	40.5	41.1	**46**
Averages				39.5	39.8	35.8	36.2	**40.77**	45.68	46.06	41.4	41.8	**47.11**

Table 3.6: Percentage correct allocation (PCA) and percentage of allocations to doses with DLT probability in the interval (λ_e, λ_d) (Int) by the number of groups, scenario, number of doses, number of patients, and method.

3.6.3 Sensitivity Analysis

The eligible patient population may consist of groups of unequal proportions. In this case, the probability the next patient belongs to a group differs across the groups, resulting in unequal group probabilities. Additionally, cohorts of size three may be used instead of enrolling one patient at a time. A sensitivity analysis analyzed how the proposed method performed when group probabilities varied and when patients enrolled in cohorts of three. Two-group simulations considered group probabilities 0.40 and 0.60, for groups 1 and 2, respectively. Three-group simulations considered group probabilities 0.20, 0.40, and 0.40, for groups 1, 2, and 3, respectively. For every single-cohort simulation, a three-cohort simulation was ran, with the number of cohorts equal to the ceiling of $\frac{\text{\# Patients in Single Cohort Simulation}}{3}$. GAB proved robust to differing group probabilities but performed worse in three-cohort trials compared to single-cohort trials. This reduction is not surprising since three-cohort trials escalate slower and try fewer doses. BOIN had a similar reduction in performance from single-patient cohorts to three-patient cohorts. Appendix B.5 provides these results.

3.6.4 Convergence Rates

This subsection considers how quickly the probability of correct dose section and allocation increases as a function of trial size. From Section 3.4, it is guaranteed that dose allocation tends to the correct dose as sample size increases, however, it is worthwhile to analyze how quickly performance improves relative to sample size. We suspect that if the true DLT probabilities are spread out and no doses have DLT probabilities near the boundary of the interval ($\lambda_e = 0.157, \lambda_d = 0.238$), then dose allocation and selection would tend to the correct dose quicker relative to the case

with a dose that has a DLT probability right outside of the interval. To analyze this, we considered three different scenarios: (1) when no doses are just the interval, (2) when one dose is right outside the interval, and (3) when two doses are right outside the interval.

Scenario 1				
	Dose			
Group	1	2	3	4
1	0.1	0.2	0.3	0.4
2	0.2	0.3	0.4	0.5
Scenario 2				
	Dose			
Group	1	2	3	4
1	0.15	0.2	0.3	0.4
2	0.2	0.25	0.4	0.5
Scenario 3				
	Dose			
Group	1	2	3	4
1	0.1	0.15	0.2	0.25
2	0.1	0.15	0.2	0.25

Table 3.7: DLT Probabilities by group and dose for three different scenarios.

Two statistics were considered in simulations: PCS and the probability (as a percentage) that the last patient in a group is allocated to their respective MTD, providing metrics on selection and allocation. Trial sizes went up to 1,000 patients per group. Figure 3.6 provides these results. Additionally, Figure 3.7 and Figure 3.8 provide plots zooming for sample sizes no more than 200 per group and demonstrating the differences in performance, respectively. As expected, the probability of correct allocation and selection converges to one quickest in scenario 1, followed by scenario 2, and scenario 3 converges the slowest. Additionally, we see the superior performance of GAB when compared to parallel-BOIN, with the GAB method converging quicker than parallel-BOIN trials.

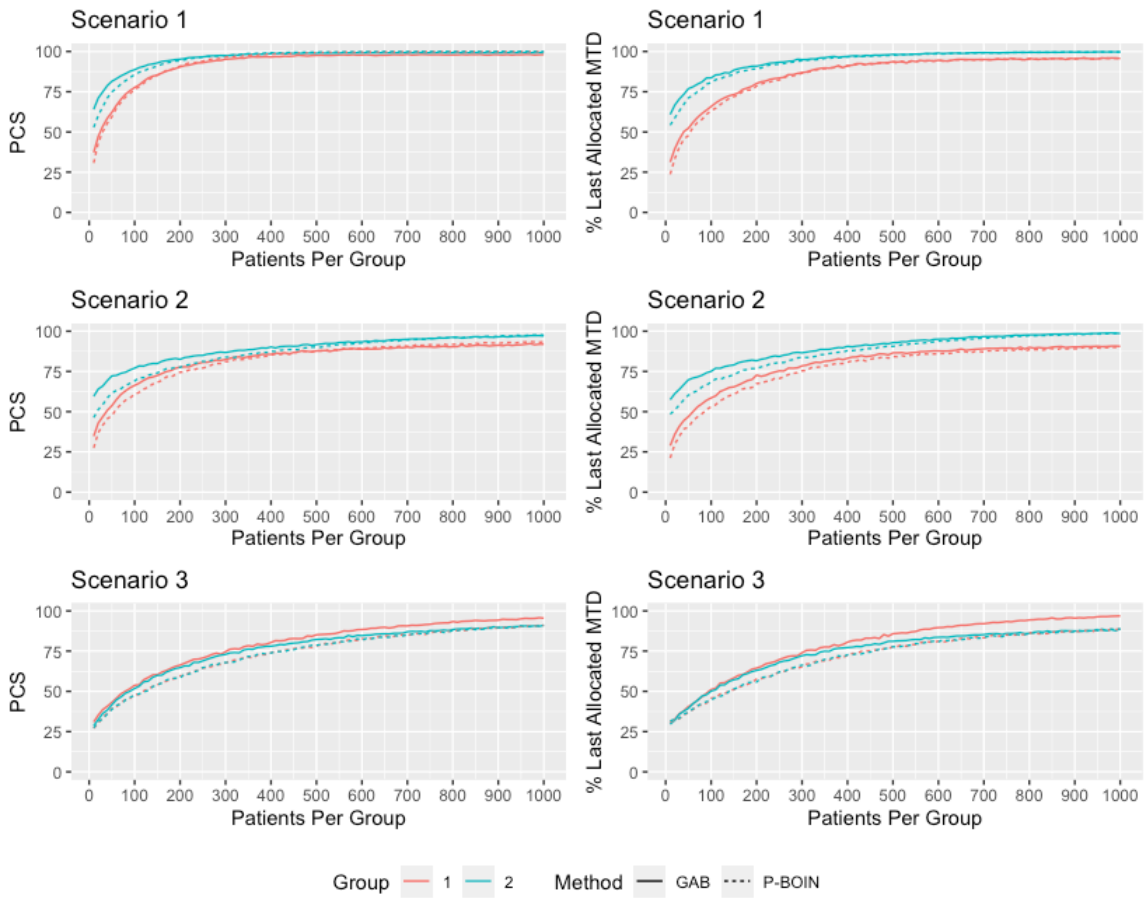


Figure 3.6: Percent Correct Selection (PCS) and probability (as a percentage) that the last patient is allocated to the MTD by Group and Method.

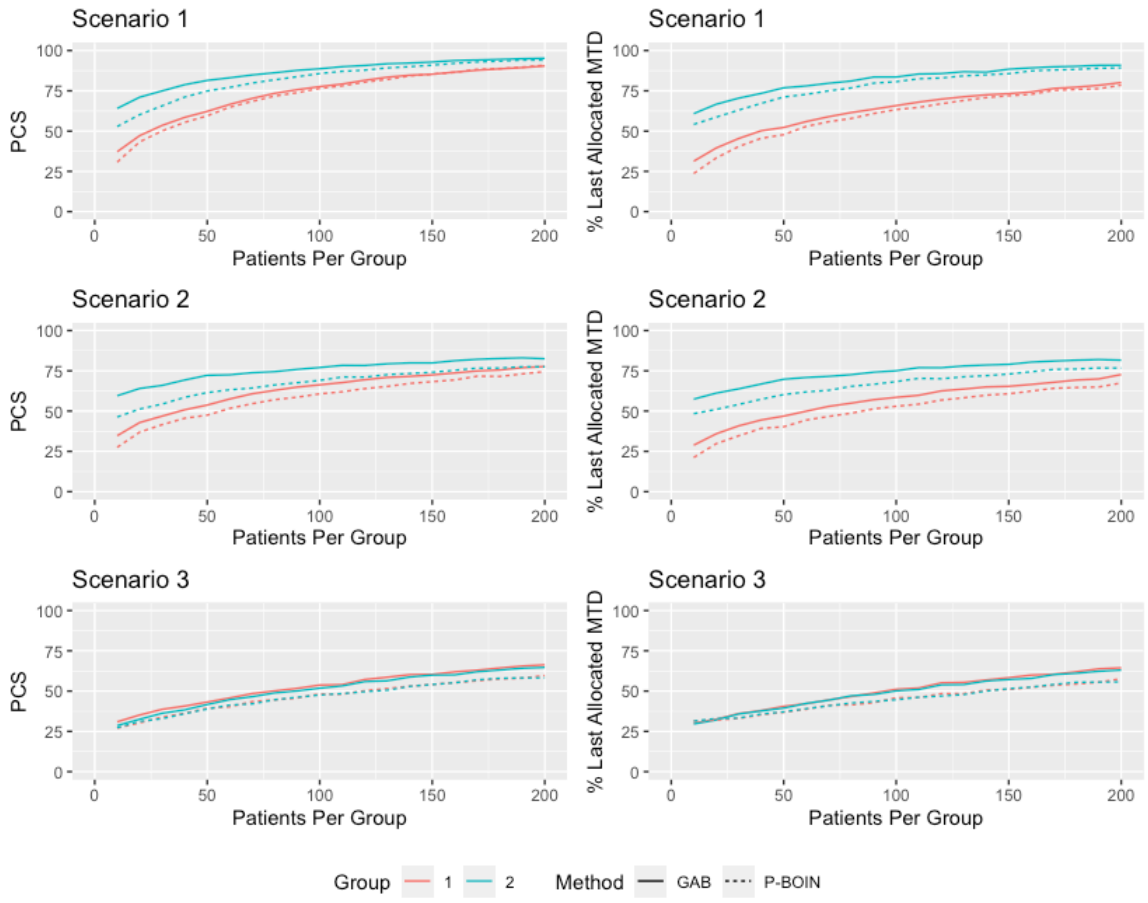


Figure 3.7: Plot of Percent Correct Selection (PCS) and probability (as a percentage) that the last patient is allocated to the MTD by Group and Method. Plot is zoomed into sample sizes between 0 and 200.

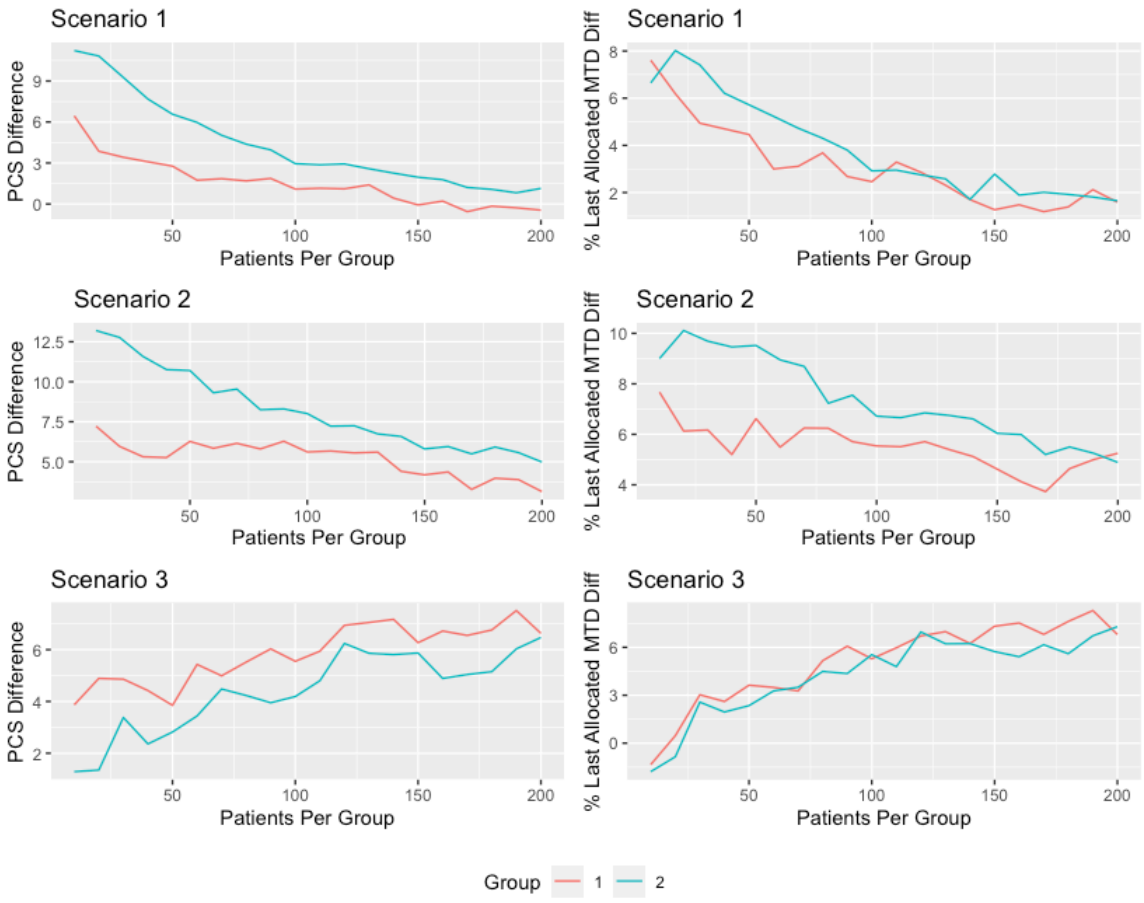


Figure 3.8: Performance differences (between GAB and P-BOIN) for Percent Correct Selection (PCS) and probability (as a percentage) that the last patient is allocated to the MTD. The differences are provided for each group.

3.7 Discussion

In this chapter, we proposed a model-assisted dose-finding design for partially or completely ordered groups. The proposed method adapts BOIN by taking group averages, giving the name "group averaged BOIN" (GAB). By taking group averages when necessary, GAB avoids reversals and outperforms parallel BOIN trials. Additionally, GAB performs similarly to OR-CRM, a model-based design in M. R. Conaway 2017b.

A future area of research is creating a model-assisted phase I/II design for partially ordered groups that considers toxicity and efficacy endpoints. Previous phase I/II designs for ordered groups include Wages, Read, and Petroni 2015 and Lin, Thall, and Y. Yuan 2020, both being model-based designs. Developing a model-assisted phase I/II design for partially ordered groups would provide clinicians with a simple design that incorporates efficacy endpoints.

The proposed method can be extended to phase I trials considering graded toxicity. Mu et al. 2019 provides an extension of BOIN for ordinal toxicity. Mu et al. 2019 modified BOIN by escalating, de-escalating, or staying at the same dose if the average ET score at the current dose is below, above, or within some interval around the toxicity target. Such rules could easily be extended to GAB.

Chapter 4

Multi-Dose Phase Two Trial Designs for Two Ordered Groups

4.1 Introduction

Phase II oncology paradigm is currently undergoing a monumental shift. Traditionally, a phase I trial would be conducted to identify the MTD. Subsequently, a phase II design would use the MTD from phase I. However, this paradigm is based on cytotoxic therapies and may not be well suited for new targeted therapies and immunotherapies as a higher dose may result in higher toxicity without any efficacious benefits (Shah et al. 2021; Papachristos et al. 2023). Recognizing the need for a paradigm shift, the FDA released "Project Optimus" (Murphy, Halford, and Symeonides 2023), an initiative for improving dose selection and optimization.

One recommendation in Project Optimus is randomization to a range of doses (U.S. Food And Drug Administration 2024b). The case of the drug Sotorasib, from the pharmaceutical company Amgen, provides an example of the FDA emphasizing the importance of dose-ranging. In phase I, patients received doses of 180, 360, 720, or 960 mg. The MTD was defined as the dose with 20% to 33% of patients experiencing DLTs. The 960 mg dose was identified as the MTD. After the phase I trial, accelerated approval was granted after significant signs of efficacy and due to the novelty of the

drug. However, as a postmarking requirement, the FDA required Amgen to compare the approved 960mg dose with a 240mg dose (Ratain, Tannock, and Lichter 2021; Moon 2022). This requirement aligns with the FDA's Oncology Center of Excellence recommendation to consider randomized multi-dose trials (U.S. Food And Drug Administration 2024a).

Yang et al. 2024 designed a multi-dose phase II clinical trial (MERIT) to determine which doses are safe and effective. In this chapter, we present a traditional (single-stage) and two-stage multi-dose phase II clinical trial design with patients stratified by group. By considering multi-dose trial designs with groups, we build upon the contributions of the MERIT design.

This chapter proposes a pair of multi-dose phase II designs for two ordered groups, the first such designs for this type of trial. First, we design a traditional trial that maximizes power while satisfying a type I error constraint. Second, we design a two-stage trial that minimizes "global Simon's statistic" while satisfying power and type I error requirements. The name "global Simon's statistic" comes from the ubiquitous Simon's Two-Stage Design, found in Simon 1989. This design minimizes the number of patients tired under the null hypothesis while satisfying power and type I error constraints. Our second design expands Simon's framework to a multi-dose, two-group design.

4.2 Notation

This section introduces the notation used to describe our designs. Let Y_E and Y_T denote the binary efficacy and DLT outcomes, respectively. For group g, we consider doses $d_{1,g} < d_{2,g} < \cdots < d_{K_g,g}$. Additionally, we require $d_{K_1,1} \geq d_{K_2,2}$ since the

highest dose for the less sensitive group should be higher than the highest dose for the more sensitive group. This chapter considers when the doses for groups 1 and 2 are the same, with doses d_1 and d_2. Although only two dose levels are considered, the methods presented in this chapter can be generalized when the number of doses exceeds two, and the doses are not the same in both groups. Appendix C.3 illustrates how the proposed methods can be modified if the two groups have different doses.

We denote the DLT and efficacy probabilities as $\pi_{T,g,k} = P(Y_T = 1|d_k, g)$ and $\pi_{E,g,k} = P(Y_E = 1|d_k, g)$, respectively.

Now, we consider the inequalities in DLT and ET probabilities that arise from the group and dose ordering. As a result of the monotonicity of toxicity and efficacy across doses, $\pi_{T,g,1} \leq \pi_{T,g,2}$ and $\pi_{E,g,1} \leq \pi_{E,g,2}$. As a result of the group ordering, $\pi_{T,1,k} \leq \pi_{T,2,k}$. Note that there is no a priori ordering for efficacy across groups. Now, consider the null and alternative parameters for toxicity and efficacy. Let $\theta_{T,0}$ denote the unacceptable toxicity rate and $\theta_{T,1}$ denote the acceptable toxicity rate, so $\theta_{T,0} > \theta_{T,1}$. Let $\theta_{E,0}$ denote the unacceptable efficacy rate and let $\theta_{E,1}$ denote the acceptable efficacy rate, so $\theta_{E,0} < \theta_{E,1}$. Dose d_k is deemed acceptable for group g if $\pi_{T,g,k} < \theta_{T,0}$ and $\pi_{E,g,k} > \theta_{E,0}$. A dose can be deemed unacceptable in three ways: having unacceptable toxicity and unacceptable efficacy, having unacceptable toxicity and acceptable efficacy, and having acceptable toxicity and unacceptable efficacy. The null hypothesis is that none of the doses are acceptable for any of the groups, while the alternative hypothesis is that at least one of the doses is acceptable for at least one of the groups. The null and alternative regions are high dimensional and are detailed in the next section.

4.3 Null and Alternative Regions

In this trial, there are two groups with multiple doses, making the null and alternative regions high-dimensional. Also contributing to the high dimensionality of the regions are the three different ways a dose can be unacceptable for a given group, as previously discussed in Section 4.2. First, we state the null hypothesis as

$$H_0 = \text{All doses are unacceptable for all groups}$$

and the alternative hypothesis as

$$H_1 = \text{At least one dose is acceptable for at least one group.}$$

Note that several configurations compose the null region. Appendix C.1 provides an algorithm for finding all possible configurations. Additionally, Appendix C.2 lists all configurations in the null and alternative space. The next section defines global type I error as the maximum type I error over all configurations in the null space and defines global power as the minimum power over all configurations in the alternative space.

4.4 Traditional Trial

This section proposes a traditional clinical trial design where each of the four arms (2 groups $\times$ 2 doses) are observed with N patients per arm, giving a total of $4 \times N$ patients. We create a design that maximizes power while constraining type 1 error. At the end of the trial, the design will accept or reject the null hypothesis that there

are no acceptable doses for either group. If rejected, the design determines which dose/group pairs are acceptable.

For dose d_k and group g, let $x_{E,k,g}$ and $x_{T,k,g}$ denote the number of efficacious outcomes and DLTs at the end of the trial. First, as $\pi_{T,k,g}$ increases across groups and dose levels, bivariate isotonic regression is applied to the matrix below.

$$\begin{pmatrix} x_{T,1,1} & x_{T,2,1} \\ x_{T,1,2} & x_{T,2,2} \end{pmatrix}$$

Let $\tilde{x}_{T,k,g}$ denote the "smoothed" value of $x_{T,k,g}$ obtained from bivariate isotonic regression. Additionally, values $\pi_{E,k,g}$ increase across dose levels. From this, we perform isotonic regression on $\{x_{E,1,1}, x_{E,2,1}\}$ and $\{x_{E,1,2}, x_{E,2,2}\}$. Let $\tilde{x}_{E,k,g}$ denote the "smoothed" value of $x_{E,k,g}$ obtained from isotonic regression. Recall the functions "biviso" and "pava" are available in the R package "Iso" (Turner 2020) for performing bivariate isotonic regression and isotonic regression. After obtaining smoothed toxicity and efficacy values, dose d_k is deemed acceptable for group g if $\tilde{x}_{T,k,g} \leq m_T$ and $\tilde{x}_{E,k,g} \geq m_E$. Here m_T is the maximum allowable number of DLTs and m_E is the minimum number of efficacious outcomes required. Our goal is to find the cutoffs, $m_E, m_T \in \{0, 1, 2, \ldots, N\}$, which maximizes power while not exceeding a type I error constraint.

First, we define global type I error and power similarly to Yang et al. 2024. Let C_0 denote all configurations in the null space and C_1 denote all configurations in the alternative space. For a given $c_0 \in C_0$, let $\alpha(c_0) = P(\text{Reject } H_0 | H_0(c_0))$ denote the type I error at the given configuration. Global type I error, denoted as $\alpha^\star$, is the

maximum type I error over all configurations in C_0,

$$\alpha^\star = \max_{c_0 \in C_0} \alpha(c_0).$$

Additionally, for a given $c_1 \in C_1$, power is defined as the probability H_0 is rejected and at least one of the doses deemed acceptable is actually acceptable. This definition, which originates from Yang et al. 2024, not only considers how often H_0 is correctly rejected but also how accurately doses are selected. For a given configuration in the alternative space, $c_1 \in C_1$, power is defined as

$$\beta(c_1) = P(\text{Reject } H_0 \text{ and at least one dose deemed acceptable is in fact acceptable} | H_1(c_1)).$$

Global power, denoted as $\beta^\star$, is the minimum power over all configurations in C_1,

$$\beta^\star = \min_{c_1 \in C_1} \beta(c_1).$$

After defining global power and type I error, a procedure to find the optimal cutoffs, m_E and m_T, will be provided.

4.4.1 Finding the Optimal Cutoffs for the Traditional Design

This section details the procedure for finding the optimal cutoffs for the traditional design. To obtain optimal cutoffs, for each configuration, samples will be drawn from multinomial distributions. To transform two binomial random variables ($X_{E,k,g} \sim Bin(N, \pi_{E,k,g})$ and $X_{T,k,g} \sim Bin(N, \pi_{T,k,g})$) into a single multinomial distribution, an association between $X_{E,k,g}$ and $X_{T,k,g}$ needs to be assumed. In this chapter, we assume an odds ratio of two. Section 4.7 determines the impact of misspecifying the

odds ratio. Now let $X_{k,g} \sim Multinomial(N, \pi_{k,g})$, denote the multinomial random variable obtained from $X_{E,k,g}$ and $X_{T,k,g}$. The vector of probabilities is given by $\pi_{k,g} = (\pi_{k,g,11}, \pi_{k,g,12}, \pi_{k,g,21}, \pi_{k,g,22})$, where

$$\pi_{k,g,11} = P(\text{Toxicity and Efficacy}|d_k, g)$$
$$\pi_{k,g,12} = P(\text{No Toxicity and Efficacy}|d_k, g)$$
$$\pi_{k,g,21} = P(\text{Toxicity and No Efficacy}|d_k, g)$$
$$\pi_{k,g,22} = P(\text{No Toxicity and No Efficacy}|d_k, g)$$

To relate the vector $\pi_{k,g}$ to the toxicity and efficacy probabilities, note that $\pi_{E,k,g} = \pi_{k,g,11} + \pi_{k,g,12}$ and $\pi_{T,,k,g} = \pi_{k,g,11} + \pi_{k,g,21}$. Using this relationship and setting the odds ratio to 2, the complete probability vector, $\pi_{k,g} = (\pi_{k,g,11}, \pi_{k,g,12}, \pi_{k,g,21}, \pi_{k,g,22})$, can be obtained. The odds ratio formula is provided below for reference.

$$\text{Odds Ratio} = \frac{(\pi_{k,g,11}/\pi_{k,g,21})}{(\pi_{k,g,12}/\pi_{k,g,22})}$$

After transforming binomial distributions into multinomial distributions, several trials are simulated to obtain the optimal cutoffs. Details are provided below.

1. Step 1: Approximate the distribution through samples:

   (a) Start by taking several samples from all possible configurations in the null space. That is, generate 1,000 samples for each configuration, c_0, in the null set of configurations, C_0. In each sample, for all groups, there are N patients on each dose, giving a total of $4 \times N$ patients.

(b) Next, take several samples from all possible configurations in the alternative space. Similarly, generate 1,000 samples for each configuration, c_1, in the alternative set of configurations, C_1.

2. Step 2: For all $m_E, m_T \in \{0, 1, \ldots, N\}$, the following procedure is conducted:

   (a) Determine the global type one error from cutoffs (m_E, m_T). The type one error at configuration $c_0 \in C_0$ using cutoffs (m_E, m_T) is denoted as $\alpha(m_E, m_T, c_0)$. Additionally, the global type one error using these cutoffs is denoted as $\alpha^\star(m_E, m_T)$, where $\alpha^\star(m_E, m_T) = \max_{c_0 \in C_0} \alpha(m_E, m_T, c_0)$.

   (b) Determine the global power using cutoffs (m_E, m_T). The power at configuration $c_1 \in C_1$ using cutoffs (m_E, m_T) is denoted as $\beta(m_E, m_T, c_1)$. Similarly, the global power using these cutoffs is denoted as $\beta^\star(m_E, m_T)$, where $\beta^\star(m_E, m_T) = \min_{c_1 \in C_1} \beta(m_E, m_T, c_1)$.

3. Step 3: Select (m_E, m_T) which maximizes $\beta^\star(m_E, m_T)$, out of all (m_E, m_T) satisfying $\alpha^\star(m_E, m_T) \leq \alpha^\star$.

4.5 Two-Stage Design

After providing a procedure to create a traditional design, a procedure for creating a two-stage multi-dose design for two groups is provided. This design is motivated by Simon's Two-Stage design (Simon 1989) and minimizes the number of patients observed at unacceptable doses in the second stage, subject to type I error and power constraints. The maximum allowable type I error is denoted as $\alpha^\star$ and the minimum required global power is denoted as $\beta^\star$. Let n denote the number of patients observed in each arm in the first stage. For a particular group, if a dose continues to the

second stage, we observe an additional $N - n$ patients on this dose. From this, for a given dose, if a group has observations from both stages, there will be a total of N patients. Let m_{E1} denote the minimum number of efficacious outcomes required for a group to continue enrolling patients on a dose in the second stage. Additionally, let m_{T1} denote the maximum number of DLTs allowed for a group to continue enrolling patients on a dose in the second stage.

We define "Simon's Statistic" at a configuration (can be in the null or alternative region), $c \in C$, as the expected number of unacceptable dose and group pairs, (d_k, g), that continue onto the second stage. Simon's Statistic at a given configuration, $c \in C$, is denoted as $s(c)$. Global Simon's statistics, denoted as $s^\star$, is the maximum Simon's statistic over all configurations,

$$s^\star = \max_{c \in C} s(c).$$

To find cutoffs $(m_{E1}, m_{T1}, m_E, m_T)$ that minimize $s^\star$, subject to global power constraint $\beta^\star$, and global type I error constraint $\alpha^\star$, we follow a similar procedure to that in Section 4.4.1. The details are provided below:

1. Step 1: Approximate the distribution through samples:

   (a) Start by taking several samples from all possible configurations in the null space. That is, generate 1,000 samples for each configuration, c_0, in the null set of configurations, C_0. In each sample, both stages are simulated. For each group and in each dose, there are n patients in the first stage and $N - n$ patients in the second stage.

   (b) Next, we take several samples from all possible configurations in the alternative space. Similarly, generate 1,000 samples for each configuration,

c_1, in the alternative set of configurations, C_1.

2. Step 2: For all $m_{E1}, m_{T1} \in \{0, 1, \ldots, n\}$ and all $(m_E, m_T) \in \{m_{E1}, \ldots, N\} \times \{m_{T1}, \ldots, N\}$, the following procedure is conducted:

   (a) Determine the global type one error from cutoffs $(m_{E1}, m_{T1}, m_E, m_T)$. The type one error at configuration $c_0 \in C_0$, using cutoffs $(m_{E1}, m_{T1}, , m_E, m_T)$, is denoted as $\alpha(m_{E1}, m_{T1}, , m_E, m_T, c_0)$. Additionally, the global type one error using these cutoffs is denoted as $\alpha^\star(m_{E1}, m_{T1}, m_E, m_T)$, where $\alpha^\star(m_{E1}, m_{T1}, m_E, m_T) = \max_{c_0 \in C_0} \alpha(m_{E1}, m_{T1}, m_E, m_T, c_0)$.

   (b) Determine the global power from cutoffs $(m_{E1}, m_{T1}, m_E, m_T)$. The power at configuration $c_1 \in C_1$, using cutoffs $(m_{E1}, m_{T1}, m_E, m_T)$, is denoted as $\beta(m_{E1}, m_{T1}, m_E, m_T, c_1)$. Similarly, the global power using these cutoffs is denoted as $\beta^\star(m_{E1}, m_{T1}, m_E, m_T)$, where $\beta^\star(m_{E1}, m_{T1}, m_E, m_T) = \min_{c_1 \in C_1} \beta(m_{E1}, m_{T1}, m_E, m_T, c_1)$.

   (c) Determine the global Simon's Statistic from cutoffs $(m_{E1}, m_{T1}, m_E, m_T)$. The Simon's Statistic at configuration $c \in C$ $(C = C_0 \cup C_1)$, using cutoffs $(m_{E1}, m_{T1}, , m_E, m_T)$, is denoted as $s(m_{E1}, m_{T1}, , m_E, m_T, c)$. Similarly, the global Simon's Statistic using these cutoffs is denoted as $s^\star(m_{E1}, m_{T1}, , m_E, m_T)$, where $s^\star(m_{E1}, m_{T1}, , m_E, m_T) = \max_{c \in C} s(m_{E1}, m_{T1}, , m_E, m_T, c)$.

3. Step 3: Select $(m_{E1}, m_{T1}, , m_E, m_T)$ that minimizes $s^\star(m_{E1}, m_{T1}, , m_E, m_T)$, out of all $(m_{E1}, m_{T1}, , m_E, m_T)$ satisfying $\alpha^\star(m_{E1}, m_{T1}, , m_E, m_T) \leq \alpha^\star$ and $\beta^\star(m_{E1}, m_{T1}, , m_E, m_T) \geq \beta^\star$.

4.6 Simulations

In this section, simulations are run to obtain the cutoffs from both the traditional and two-stage designs. Comparisons are made between MERIT and the traditional design, showing ad hoc modifications of MERIT do not translate to the group design, highlighting the necessity of the proposed traditional design. The null and alternative toxicity rates are $\theta_{T_0} = 0.4$ and $\theta_{T_1} = 0.2$. Additionally, the null and alternative efficacy rates are $\theta_{E_0} = 0.2$ and $\theta_{E_1} = 0.4$.

Initially, we compared the traditional design to MERIT for two doses, then drew comparisons using an ad hoc modification of MERIT. First, Yang et al. 2024 provides sample size calculations with $\alpha^\star = 0.1, 0.2, 0.3$ and $\beta^\star = 0.6, 0.7, 0.8$ for trials with 2 doses. As there are multiple comparisons in the trial (coming from the groups), these cutoffs will have an inflated global type I error if used in parallel for two groups. Our simulations will show the degree of type I error inflation. Before illustrating the type I error inflation, we provide the MERIT cutoffs in Table 4.1.

$\beta^\star$	$\alpha^\star = 0.1$			$\alpha^\star = 0.2$			$\alpha^\star = 0.3$		
	n	m_T	m_E	n	m_T	m_E	n	m_T	m_E
0.6	26	7	9	18	5	6	18	5	6
0.7	34	9	11	25	7	8	20	6	6
0.8	45	12	14	35	10	10	24	7	7

Table 4.1: Cutoffs From MERIT (Yang et al. 2024))

Using the nine combinations of n and $\alpha^\star$ from Table 4.1, we obtain the cutoffs for the traditional two-group design presented in this chapter. Note that the proposed traditional design is not presented as a sample size calculator but could be modified accordingly. As our design correctly calculates the type I error, it will be more conservative than two-dose MERIT, giving a lower power. Global type I error and global power results from two-dose MERIT are provided in Table 4.2. These statistics,

along with the decision boundaries, (m_T, m_E), for the two-group traditional design are provided in Table 4.3. These tables show that the traditional design presented in this chapter properly considers the high dimensional configuration of the null space and prevents the global type one error from exceeding the specified cutoff. Additionally, Table 4.2 shows it is improper to use the MERIT design for two groups as the type I error will be inflated.

$\alpha^\star = 0.1$					$\alpha^\star = 0.2$					$\alpha^\star = 0.3$				
n	m_T	m_E	$\alpha^\star$	$\beta^\star$	n	m_T	m_E	$\alpha^\star$	$\beta^\star$	n	m_T	m_E	$\alpha^\star$	$\beta^\star$
26	7	9	0.213	0.603	18	5	6	0.375	0.617	18	5	6	0.375	0.617
34	9	11	0.18	0.709	25	7	8	0.343	0.706	20	6	6	0.539	0.765
45	12	14	0.14	0.79	35	10	10	0.374	0.843	24	7	7	0.481	0.793

Table 4.2: Type 1 Error, Power, cutoffs From MERIT for two doses

$\alpha^\star = 0.1$					$\alpha^\star = 0.2$					$\alpha^\star = 0.3$				
n	m_T	m_E	$\alpha^\star$	$\beta^\star$	n	m_T	m_E	$\alpha^\star$	$\beta^\star$	n	m_T	m_E	$\alpha^\star$	$\beta^\star$
26	6	10	0.06	0.309	18	4	7	0.096	0.267	18	4	6	0.201	0.399
34	8	12	0.059	0.463	25	6	8	0.197	0.53	20	5	7	0.19	0.471
45	12	15	0.095	0.718	35	9	11	0.158	0.701	24	6	8	0.187	0.541

Table 4.3: Cutoffs, Type 1 Error, and Power from the proposed traditional design

An ad hoc attempt to reign in the type I error for MERIT in parallel is considering the four-dose MERIT cutoffs. The intuition behind this adjustment is there are essentially four arms in the trial with two groups and two doses, so one might consider the cutoffs from four-dose MERIT with the hope this adjustment would correct the inflated type I error. To obtain the cutoffs from the MERIT four-dose sample size calculator using $\alpha^\star = 0.1, 0.2, 0.3$ and $\beta^\star = 0.6, 0.7, 0.8$, the MERIT application was used from the website trialdesign. These cutoffs are provided in Table 4.4.

Similarly, we obtained the cutoffs for the proposed traditional design using these nine combinations of n and $\alpha^\star$. Global type I error and global power were obtained for both methods. Results from four-dose MERIT and the proposed traditional design

$\beta^\star$	$\alpha^\star = 0.1$			$\alpha^\star = 0.2$			$\alpha^\star = 0.3$		
	n	m_T	m_E	n	m_T	m_E	n	m_T	m_E
0.6	29	8	10	21	6	7	19	5	6
0.7	37	10	12	28	8	9	26	8	8
0.8	48	13	15	38	11	11	27	8	8

Table 4.4: Cutoffs From MERIT with 4 doses (trialdesign)

are provided in Table 4.5 and Table 4.6, respectively. These results show the ad hoc adjustment does not "fix" MERIT and, as a result, the design can not be used in parallel for groups. This adjustment of MERIT is not sufficient as the single-group four-dose structure is fundamentally different than the two-group and two-dose structure, with a lower dimensionality. Due to the differing structure, type I error is inflated using four-dose MERIT cutoffs.

$\alpha^\star = 0.1$			$\alpha^\star = 0.2$			$\alpha^\star = 0.3$		
n	$\alpha^\star$	$\beta^\star$	n	$\alpha^\star$	$\beta^\star$	n	$\alpha^\star$	$\beta^\star$
29	0.207	0.633	21	0.362	0.645	19	0.38	0.643
37	0.174	0.735	28	0.306	0.725	26	0.461	0.789
48	0.13	0.809	38	0.34	0.857	27	0.445	0.81

Table 4.5: Type 1 Error and Power From MERIT four doses

$\alpha^\star = 0.1$					$\alpha^\star = 0.2$					$\alpha^\star = 0.3$				
n	m_T	m_E	$\alpha^\star$	$\beta^\star$	n	m_T	m_E	$\alpha^\star$	$\beta^\star$	n	m_T	m_E	$\alpha^\star$	$\beta^\star$
29	8	10	0.207	0.633	21	6	7	0.362	0.645	19	5	6	0.38	0.643
37	10	12	0.174	0.735	28	8	9	0.306	0.725	26	8	8	0.461	0.789
48	13	15	0.13	0.809	38	11	11	0.34	0.857	27	8	8	0.445	0.81

Table 4.6: Type 1 Error, Power, cutoffs From MERIT four doses

Finally, we consider simulations using the two-stage design presented in this chapter. The two-stage design finds the optimal cutoffs, $(m_{E1}, m_{T1}, m_E, m_T)$, that minimizes the global Simon's Statistic, $s^\star$, while ensuring the global type I error is below $\alpha^\star$ and the global power is above $\beta^\star$. We consider when $\alpha^\star = 0.2, 0.3$ and $\beta^\star = 0.4, 0.5, 0.6, 0.7$. Additionally, for each pair $(\alpha^\star, \beta^\star)$, we consider six pairs of

(n, N), these pairs being $(7, 15), (10, 20), (15, 30), (20, 40)$ and $(30, 60)$. For several combinations of sample sizes (n, N), and global type I error and power requirements $(\alpha^\star, \beta^\star)$, there does not exist cutoffs $(m_{E1}, m_{T1}, m_E, m_T)$ satisfying type I error and power requirements. If this is the case, the row will be filled with "NA's". For these results, see Table 4.7 through Table 4.14.

Lastly, we considered the effects of intern sample size on the global Simon's Statistics. Simulations were considered with $N = 30$, while the interim sample size, n, varied from 12 to 18. In these simulations, the maximum allowable global type I error $\alpha^\star$ was 0.3, and the minimum required global power $\beta^\star$ was 0.6. Table 4.15 provides these simulations and shows $s^\star$ generally decreases as the interim sample size increases. It follows more stringent boundaries can be drawn if the intern sample size is larger since we have more confidence in our results with larger n. However, it is important to note that as n increases, more patients will be observed in the first stage, meaning more patients will be observed on unacceptable doses in the first stage. This illustrates the trade-off between having the interim stage later and having more patients observed on unacceptable doses in the first stage but having more confident decisions when removing doses, compared to having the interim stage earlier and having fewer patients observed on unacceptable doses in the first stage but having less confidence when removing doses.

Using the previous paragraph as motivation, note that at a configuration c, the number of patients expected to be observed at an unacceptable dose is equal to $N(c)\times n+s(c)\times(N-n)$, where $N(c)$ is the number of unacceptable doses under configuration c. From this, when looking at the maximum value of $N(c)\times n+s(c)\times(N-n)$ over all configurations, we consider $EN^\star = 4 \times n + s^\star \times (N - n)$, as $s^\star$ is maximized under configurations when all four doses are unacceptable. Table 4.16 shows when

$n = 14$, the number of patients at unacceptable doses is minimized. In addition, we explore the optimal interim size n for minimizing $EN^\star$, based on total trial size, type I error constraint, and power constraint. Table 4.17 provides these results for various choices of $N, \alpha^\star$, and $\beta^\star$. There does not seem to be a general relationship between the optimal interim size (n) and total trial size (N), so we recommend trying various values of n if attempting to minimize $EN^\star$.

4.7 Sensitivity Analysis

This section considers the impact of misspecifying the odds ratio, as defined as the ratio of 1) the odds of experiencing an efficacious outcome if a DLT occurred and 2) the odds of experiencing an efficacious outcome if a DLT did not occur. An odds ratio below 1 means the odds of experiencing an efficacy decrease given a DLT. An odds ratio of 1 means the odds of experiencing an efficacious outcome are the same, regardless of if a DLT occurred. Finally, an odds ratio greater than 1 signifies the odds of experiencing an efficacious outcome increase given a DLT.

Previously, optimal cutoffs were obtained under the assumption of an odds ratio of 2. This choice of odds ratio matches that of the sequential design in M. R. Conaway and Petroni 1995, which was selected based on results from previous trials. Similar to M. R. Conaway and Petroni 1995, we assess the impact of odds ratio misspecification on power and type I error. The optimal cutoffs for a traditional (single-stage) design that controls global type I error, $\alpha^\star$, at 0.2 with 25 patients per group and drug combination is $m_T = 7$ and $m_E = 9$. Table 4.18 shows the impact on power and type 1 error from using these cutoffs when the odds ratio differs from 2. From these results, underestimating the odds ratio leads to a more conservative design, while

$\alpha^\star = 0.2, \beta^\star = 0.4$								
n	N	m_{T_1}	m_{E_1}	m_T	m_E	s	$\alpha^\star$	$\beta^\star$
7	15	NA	NA	NA	NA	NA	NA	NA
10	20	4	3	5	7	2.021	0.168	0.406
15	30	4	5	7	9	0.455	0.105	0.417
20	40	6	7	10	13	0.398	0.056	0.426
25	50	7	9	13	16	0.211	0.051	0.417
30	60	7	10	14	18	0.101	0.017	0.422

Table 4.7: Cutoffs and stats for two-stage design when maximum global type I error is $\alpha^\star = 0.2$ and minimum global power is $\beta^\star = 0.4$

$\alpha^\star = 0.2, \beta^\star = 0.5$								
n	N	m_{T_1}	m_{E_1}	m_T	m_E	s	$\alpha^\star$	$\beta^\star$
7	15	NA	NA	NA	NA	NA	NA	NA
10	20	NA	NA	NA	NA	NA	NA	NA
15	30	5	5	8	9	0.886	0.158	0.542
20	40	5	6	10	11	0.448	0.147	0.525
25	50	7	8	12	15	0.349	0.053	0.548
30	60	8	10	14	18	0.175	0.023	0.513

Table 4.8: Cutoffs and stats for two-stage design when maximum global type I error is $\alpha^\star = 0.2$ and minimum global power is $\beta^\star = 0.5$

$\alpha^\star = 0.2, \beta^\star = 0.6$								
n	N	m_{T_1}	m_{E_1}	m_T	m_E	s	$\alpha^\star$	$\beta^\star$
7	15	NA	NA	NA	NA	NA	NA	NA
10	20	NA	NA	NA	NA	NA	NA	NA
15	30	5	4	8	9	1.287	0.195	0.623
20	40	6	6	10	11	0.652	0.169	0.656
25	50	7	8	14	15	0.349	0.099	0.601
30	60	9	10	15	18	0.309	0.04	0.608

Table 4.9: Cutoffs and stats for two-stage design when maximum global type I error is $\alpha^\star = 0.2$ and minimum global power is $\beta^\star = 0.6$

$\alpha^\star = 0.2, \beta^\star = 0.7$								
n	N	m_{T_1}	m_{E_1}	m_T	m_E	s	$\alpha^\star$	$\beta^\star$
7	15	NA	NA	NA	NA	NA	NA	NA
10	20	NA	NA	NA	NA	NA	NA	NA
15	30	NA	NA	NA	NA	NA	NA	NA
20	40	6	6	11	11	0.652	0.182	0.702
25	50	7	7	14	14	0.596	0.138	0.716
30	60	9	9	15	17	0.477	0.072	0.724

Table 4.10: Cutoffs and stats for two-stage design when maximum global type I error is $\alpha^\star = 0.2$ and minimum global power is $\beta^\star = 0.7$

$\alpha^\star = 0.3, \beta^\star = 0.4$								
n	N	m_{T_1}	m_{E_1}	m_T	m_E	s	$\alpha^\star$	$\beta^\star$
7	15	3	2	4	5	2.313	0.298	0.453
10	20	3	3	5	6	1.018	0.245	0.497
15	30	4	5	7	9	0.455	0.105	0.417
20	40	6	7	10	13	0.398	0.056	0.426
25	50	7	9	13	16	0.211	0.051	0.417
30	60	7	10	14	18	0.101	0.017	0.422

Table 4.11: Cutoffs and stats for two-stage design when maximum global type I error is $\alpha^\star = 0.3$ and minimum global power is $\beta^\star = 0.4$

$\alpha^\star = 0.3, \beta^\star = 0.5$								
n	N	m_{T_1}	m_{E_1}	m_T	m_E	s	$\alpha^\star$	$\beta^\star$
7	15	NA	NA	NA	NA	NA	NA	NA
10	20	4	3	5	6	2.021	0.266	0.543
15	30	5	5	8	9	0.886	0.158	0.542
20	40	5	6	10	11	0.448	0.147	0.525
25	50	7	8	12	15	0.349	0.053	0.548
30	60	8	10	14	18	0.175	0.023	0.513

Table 4.12: Cutoffs and stats for two-stage design when maximum global type I error is $\alpha^\star = 0.3$ and minimum global power is $\beta^\star = 0.5$

$\alpha^\star = 0.3, \beta^\star = 0.6$								
n	N	m_{T_1}	m_{E_1}	m_T	m_E	s	$\alpha^\star$	$\beta^\star$
7	15	NA	NA	NA	NA	NA	NA	NA
10	20	NA	NA	NA	NA	NA	NA	NA
15	30	5	4	8	9	1.287	0.195	0.623
20	40	6	6	10	11	0.652	0.169	0.656
25	50	7	8	14	15	0.349	0.099	0.601
30	60	9	10	15	18	0.309	0.04	0.608

Table 4.13: Cutoffs and stats for two-stage design when maximum global type I error is $\alpha^\star = 0.3$ and minimum global power is $\beta^\star = 0.6$

$\alpha^\star = 0.3, \beta^\star = 0.7$								
n	N	m_{T_1}	m_{E_1}	m_T	m_E	s	$\alpha^\star$	$\beta^\star$
7	15	NA	NA	NA	NA	NA	NA	NA
10	20	NA	NA	NA	NA	NA	NA	NA
15	30	7	3	8	9	3.271	0.246	0.703
20	40	6	6	11	11	0.652	0.182	0.702
25	50	7	7	13	13	0.596	0.208	0.724
30	60	9	9	15	17	0.477	0.072	0.724

Table 4.14: Cutoffs and stats for two-stage design when maximum global type I error is $\alpha^\star = 0.3$ and minimum global power is $\beta^\star = 0.7$

$\alpha^\star = 0.3, \beta^\star = 0.6$								
n	N	m_{T_1}	m_{E_1}	m_T	m_E	$s^\star$	$\alpha^\star$	$\beta^\star$
10	30	5	1	8	10	3.64	0.153	0.604
11	30	6	2	8	10	3.681	0.168	0.627
12	30	6	2	8	10	3.518	0.155	0.611
13	30	7	3	8	10	3.618	0.156	0.625
14	30	5	4	8	9	1.454	0.188	0.611
15	30	5	4	8	9	1.287	0.195	0.623
16	30	7	5	8	9	2.482	0.193	0.611
17	30	6	5	8	9	1.268	0.195	0.608
18	30	6	5	8	9	1.109	0.2	0.644
19	30	6	6	8	9	0.704	0.17	0.602
20	30	6	6	8	9	0.671	0.181	0.639

Table 4.15: Comparison with Different Interim Sample Sizes

$\alpha^\star = 0.3, \beta^\star = 0.6$			
n	N	$s^\star$	$EN^\star$
10	30	3.64	112.8
11	30	3.681	113.939
12	30	3.518	111.324
13	30	3.618	113.506
14	30	1.454	79.264
15	30	1.287	79.305
16	30	2.482	98.748
17	30	1.268	84.484
18	30	1.109	85.308
19	30	0.704	83.744
20	30	0.671	86.71

Table 4.16: Minimizing the number of patients at unacceptable doses

Two Stage Parameters			Optimal n and $EN^\star$	
N	$\alpha^\star$	$\beta^\star$	n	$EN^\star$
15	0.3	0.4	9	42.354
20	0.3	0.5	11	53.774
30	0.3	0.6	14	79.264
40	0.2	0.6	13	75.706

Table 4.17: Optimal intern size n, by $N, \alpha^\star$, and $\beta^\star$

overestimating the odds ratio leads to a more aggressive odds ratio. Overall, the impact on type 1 error is minimal, and the cutoff of $\alpha^\star = 0.2$ is not greatly exceeded when underestimating the odds ratio. These results mirror that of M. R. Conaway and Petroni 1995, and we recommend clinicians specify a range of possible odds ratios when designing a phase II trial.

	$n = 25, m_T = 7, m_E = 9$						
Odds ratio, θ	0.25	0.5	1.0	2.0	4.0	8.0	16.0
Global Type 1 Error, $\alpha^\star$	0.2552	0.2349	0.2161	0.1893	0.1658	0.1407	0.1125
Global Power, $\beta^\star$	0.5745	0.5687	0.5591	0.5491	0.5421	0.5351	0.527

Table 4.18: Sensitivity Analysis from varying true odds ratio while using cutoffs when odds ratio is 2.0 and $\alpha^\star = 0.2$

4.8 Discussion

This chapter proposed a pair of phase II designs for trials with multiple doses and two ordered groups. The first design maximizes power subject to a type I error requirement. The second design minimizes the number of patients on unacceptable doses, subject to power and type I error requirements.

The designs presented in this chapter are the first multiple doses designs for ordered groups, allowing a clinician to explore the dose-efficacy relationship in phase II, after conducting a phase I group trial. Exploring the dose-efficacy relationship in phase II aligns with the recommendation of Project Optimus to randomize patients to a range of doses in phase II. By introducing a multi-dose phase II trial for groups, clinicians can determine if a drug warrants a more expensive phase III trial and, if so, dial in the correct group-specific dose.

A future area of research includes creating a multi-dose group design that is more

flexible than the two-stage design presented in this chapter. The two-stage design presented in this chapter determines if a dose moves onto the second stage, a binary choice. However, one could consider allowing a dose to continue onto the second stage while allowing more or fewer patients to be administered the dose, based on the observations in the previous stage. One could conceive of creating such a design by assigning a prior distribution and using the posterior probability that a dose is acceptable for a given group to guide allocation in the second stage.

Chapter 5

Discussion and Future Areas of Research

5.1 Summary of Methods Proposed

In this dissertation, we discussed the different stages of clinical trials, provided examples of clinical trials with heterogeneous groups, and proposed phase I and phase II clinical trial designs. Phase I designs have the goal of locating the highest tolerable dose, known as the MTD. Phase II designs seek to determine if a drug is sufficiently effective and safe to warrant a more expensive phase III trial. Previous trials separated patients into heterogeneous groups completely or partially ordered by the probability of a DLT. In this chapter, we provide a brief recapitulation of each chapter and discuss future research topics.

Chapter 2, as adapted from Celum, Horton, and M. Conaway 2024, presented the Quasi-CRM Shift design, the first phase I group design using toxicity grades, which score the severity of an adverse event. Through the development of this design, clinicians can control for DLT severity and frequency, as well as control for adverse events that are not dose-limiting, when conducting a phase I group trial. Simulation studies showed this design significantly outperformed the ad-hoc method for using toxicity grades in a group trial, namely doing a parallel Quasi-CRM trial for each

group, this being the only competitor method before the creation of the Quasi-CRM Shift design. Additionally, performing parallel trials for heterogeneous groups is a fundamentally flawed method due to the problem of reversals (Horton, O'Quigley, and M. R. Conaway 2019), when the estimated MTD for a more sensitive group is higher than the estimated MTD of a less sensitive group.

Chapter 3, as adapted from Celum and M. Conaway 2024, presented the first model-assisted design for partially ordered groups (called GAB), developed the "bundles" framework to classify all partial orders in previous papers on partially ordered groups in phase I (Horton, Wages, and M. R. Conaway 2019; Horton, O'Quigley, and M. R. Conaway 2019; M. R. Conaway 2017a; M. R. Conaway 2017b), and proved convergence results. Before Celum and M. Conaway 2024, previous designs phase I designs for partially ordered groups utilized the CRM (O'Quigley, Pepe, and Fisher 1990), falling into the category of model-based designs, a class of designs requiring computational tools to update parameters after observing each patient. In contrast, GAB is a model-assisted design, meaning that decision boundaries can be laid out at the beginning of the trial, providing clinicians with a simple-to-use alternative in the group framework. GAB is based on BOIN (Liu and Y. Yuan 2015), and pools group information, when necessary, to keep allocation consistent with the group ordering. Simulation studies showed that GAB performs similarly to the more complex model-based design found in M. R. Conaway 2017b, thus providing simplicity without a loss in performance.

Chapter 4 presented the first multi-dose phase II group designs. In the dose-optimization initiative called Project Optimus (U.S. Food And Drug Administration 2024b), the FDA recommended dose randomization in a phase II trial to explore the dose-efficacy relationship which may plateau, as in the case of target therapies and immunother-

apies. To this end, Yang et al. 2024 designed the first multi-dose phase II design, determining which doses are sufficiently safe and effective. In Chapter 4, we extend the framework built by Yang et al. 2024 and design two multi-dose designs for groups, the first such designs, allowing clinicians to conduct a phase II group trial after a phase I group trial. The first design maximizes power while satisfying a type I error constraint. The second design minimizes the number of patients on unacceptable doses, while stratifying type I error and power requirements, similar to the objectives of Simon 1989.

5.2 Future Research

Each chapter discussed future areas of research, which will be briefly recapped, before discussing an additional future area of research. Chapter 2 discussed extending the Quasi-CRM Shift method to account for toxicity frequency as the model accounts only for the most severe toxicity observed from a patient, not how many toxic events a patient experienced. Such a modification would allow clinicians to account for persistent low-grade toxicities. Chapter 3 discussed creating the first model-assisted design for a phase I/II group trial, accounting for toxicity and efficacy. This design would be useful for clinicians preferring model-assisted designs and wanting to conduct a group trial that combines phases I and II, using both efficacy and toxicity. Chapter 4 discussed creating a more flexible two-stage design, allowing allocation in the second stage to be adaptive through using Bayesian methods, basing allocation on the probability a dose is acceptable.

Another future research topic is modifying model-based phase I group designs and GAB (a model-assisted design), to consider time to DLT. After such modifications

are made, simulations could compare the modified methods to see if the model-based or model-assisted designs perform better. In the phase I designs presented in this dissertation, patients are assigned doses one at a time, and the current patient is followed through before assigning a dose to the next patient. Depending on the protocol, patients may be followed for 3 months or longer. As a result, these phase I trials have a long duration. However, through modeling time to DLT, the CRM and BOIN, these designs being the basis for the partially ordered group designs, can be modified so a patient does not need to be followed through for the entire follow-up duration before assigning a dose to the next patient, substantially shortening the trial duration.

In the time-to-event continual reassessment method (TITE-CRM), as presented in Y. K. Cheung and Rick Chappell 2000, the likelihood function is adjusted by weighting each observation by the amount of time a patient is observed over the length of the follow-up window. For instance, if a patient has been observed for 2 months, while the follow-up window is 3 months, that patient's observation is weighted by 2/3 in the likelihood function. Y. K. Cheung and Rick Chappell 2000 noted through using the TITE-CRM, a trial taking up to 12 years using the CRM can be reduced to a trial duration of 2-4 years using the TITE-CRM. The CRM-based designs for partially ordered groups can be modified using the TITE-CRM, allowing for shortened trials.

The time-to-event Bayesian optimal interval design (TITE-BOIN), as presented in Y. Yuan, Lin, et al. 2018, modifies BOIN to allow the next patient to be allocated to a dose before the last patient is observed for the full follow-up time. Conceptually, BOIN is modified by imputing DLT rates based on the number of DLTs observed, trial duration, and length of time the patients have been followed up. It is assumed time to DLT follows a uniform distribution, while simulation studies show this assumption

does not need to be correct. The resulting method is still model-based and decision boundaries can be outlined before the trial. The TITE-BOIN and TITE-CRM identify the MTD a similar percentage of the time. One could conceive of a way to adapt GAB using the TITE-BOIN, allowing trials to be sped up. It would be interesting to compare the time-to-event modified GAB and a time-to-event modified model-based design to evaluate relative performance.

Bibliography

Brunk, HD et al. (1972). "Statistical inference under order restrictions (the theory and application of isotonic regression)". In: *International Statistical Review* 41, p. 395.

Dykstra, Richard L and Tim Robertson (1982). "An algorithm for isotonic regression for two or more independent variables". In: *The Annals of Statistics* 10.3, pp. 708–716.

Simon, Richard (1989). "Optimal two-stage designs for phase II clinical trials". In: *Controlled clinical trials* 10.1, pp. 1–10.

O'Quigley, John, Margaret Pepe, and Lloyd Fisher (1990). "Continual reassessment method: a practical design for phase 1 clinical trials in cancer". In: *Biometrics*, pp. 33–48.

Hwang, JT Gene and Shyamal Das Peddada (1994). "Confidence interval estimation subject to order restrictions". In: *The Annals of Statistics* 22.1, pp. 67–93.

Conaway, Mark R and Gina R Petroni (1995). "Bivariate sequential designs for phase II trials". In: *Biometrics*, pp. 656–664.

O'Quigley, John and Larry Z Shen (1996). "Continual reassessment method: a likelihood approach". In: *Biometrics*, pp. 673–684.

Papke, Leslie E and Jeffrey M Wooldridge (1996). "Econometric methods for fractional response variables with an application to 401 (k) plan participation rates". In: *Journal of applied econometrics* 11.6, pp. 619–632.

Shiryaev, Albert Nikolayevich (1996). "Probability". In: 2nd ed. Springer, pp. 519–520.

Cheung, Ying Kuen and Rick Chappell (2000). "Sequential designs for phase I clinical trials with late-onset toxicities". In: *Biometrics* 56.4, pp. 1177–1182.

Wang, Chinying, T Timothy Chen, and Irvin Tyan (2000). "Designs for phase i cancer clinical trials with differentiation of graded toxicity". In: *Communications in Statistics-Theory and Methods* 29.5-6, pp. 975–987.

O'Quigley, John and Xavier Paoletti (2003). "Continual reassessment method for ordered groups". In: *Biometrics* 59.2, pp. 430–440.

Bekele, B Nebiyou and Peter F Thall (2004). "Dose-finding based on multiple toxicities in a soft tissue sarcoma trial". In: *Journal of the American Statistical Association* 99.465, pp. 26–35.

Yuan, Zhilong and Rick Chappell (2004). "Isotonic designs for phase I cancer clinical trials with multiple risk groups". In: *Clinical Trials* 1.6, pp. 499–508.

Ivanova, Anastasia and Kai Wang (2006). "Bivariate isotonic design for dose-finding with ordered groups". In: *Statistics in medicine* 25.12, pp. 2018–2026.

Ivanova, Anastasia, Nancy Flournoy, and Yeonseung Chung (2007). "Cumulative cohort design for dose-finding". In: *Journal of Statistical Planning and Inference* 137.7, pp. 2316–2327.

Yuan, Z, R Chappell, and H Bailey (2007). "The continual reassessment method for multiple toxicity grades: a Bayesian quasi-likelihood approach". In: *Biometrics* 63.1, pp. 173–179.

Ramanathan, Ramesh K et al. (2008). "Phase I and Pharmacokinetic Study of Imatinib Mesylate in Patients With Advanced Malignancies and Varying Degrees of Liver Dysfunction: A Study by the National Cancer Institute Organ Dysfunction". In: *Journal of Clinical Oncology* 26.4, pp. 563–569.

Krop, Ian E, Muralidhar Beeram, et al. (2010). "Phase I study of trastuzumab-DM1, an HER2 antibody-drug conjugate, given every 3 weeks to patients with

HER2-positive metastatic breast cancer". In: *Journal of Clinical Oncology* 28.16, pp. 2698–2704.

O'Quigley, John and Sarah Zohar (2010). "Retrospective robustness of the continual reassessment method". In: *Journal of biopharmaceutical statistics* 20.5, pp. 1013–1025.

Cheung, Ying Kuen (2011). *Dose finding by the continual reassessment method*. CRC Press.

Leal, Ticiana B. et al. (2011). "Dose-escalating and pharmacological study of bortezomib in adult cancer patients with impaired renal function: a National Cancer Institute Organ Dysfunction Working Group Study". In: *Cancer Chemotherapy and Pharmacology*.

Oron, Assaf P, David Azriel, and Peter D Hoff (2011). "Dose-finding designs: the role of convergence properties". In: *The international journal of biostatistics* 7.1, p. 000010220215574679 1298.

Wages, Nolan A, Mark R Conaway, and John O'Quigley (2011). "Continual reassessment method for partial ordering". In: *Biometrics* 67.4, pp. 1555–1563.

Krop, Ian E, Patricia LoRusso, et al. (2012). "A phase II study of trastuzumab emtansine in patients with human epidermal growth factor receptor 2–positive metastatic breast cancer who were previously treated with trastuzumab, lapatinib, an anthracycline, a taxane, and capecitabine". In: *Journal of clinical oncology* 30.26, pp. 3234–3241.

LoRusso, Patricia M et al. (2012). "Pharmacokinetics and Safety of Bortezomib in Patients with Advanced Malignancies and Varying Degrees of Liver Dysfunction: Phase I NCI Organ Dysfunction Working Group Study NCI-6432Bortezomib Pharmacokinetics/Safety in Patients with Liver Dysfunction". In: *Clinical Cancer Research* 18.10, pp. 2954–2963.

Van Meter, Emily M, Elizabeth Garrett-Mayer, and Dipankar Bandyopadhyay (2012). "Dose-finding clinical trial design for ordinal toxicity grades using the continuation ratio model: an extension of the continual reassessment method". In: *Clinical trials* 9.3, pp. 303–313.

Ezzalfani, Monia et al. (2013). "Dose-finding designs using a novel quasi-continuous endpoint for multiple toxicities". In: *Statistics in medicine* 32.16, pp. 2728–2746.

Dhillon, Sohita (2014). "Trastuzumab emtansine: a review of its use in patients with HER2-positive advanced breast cancer previously treated with trastuzumab-based therapy". In: *Drugs* 74, pp. 675–686.

Innocenti, Federico et al. (2014). "Dose-finding and pharmacokinetic study to optimize the dosing of irinotecan according to the UGT1A1 genotype of patients with cancer". In: *Journal of clinical oncology* 32.22, p. 2328.

Krop, Ian E, Sung-Bae Kim, et al. (2014). "Trastuzumab emtansine versus treatment of physician's choice for pretreated HER2-positive advanced breast cancer (TH3RESA): a randomised, open-label, phase 3 trial". In: *The Lancet Oncology* 15.7, pp. 689–699.

O'Quigley, John and Alexia Iasonos (2014). "Bridging solutions in dose-finding problems". In: *Statistics in biopharmaceutical research* 6.2, pp. 185–197.

Pan, Haitao et al. (2014). "The continual reassessment method for multiple toxicity grades: A Bayesian model selection approach". In: *PloS one* 9.5, e98147.

Liu, Suyu and Ying Yuan (2015). "Bayesian optimal interval designs for phase I clinical trials". In: *Journal of the Royal Statistical Society: Series C: Applied Statistics*, pp. 507–523.

Paoletti, X, M Ezzalfani, and C Le Tourneau (2015). "Statistical controversies in clinical research: requiem for the 3+ 3 design for phase I trials". In: *Annals of Oncology* 26.9, pp. 1808–1812.

Wages, Nolan A, Paul W Read, and Gina R Petroni (2015). "A phase I/II adaptive design for heterogeneous groups with application to a stereotactic body radiation therapy trial". In: *Pharmaceutical statistics* 14.4, pp. 302–310.

Conaway, Mark R (2017a). "A design for phase I trials in completely or partially ordered groups". In: *Statistics in medicine* 36.15, pp. 2323–2332.

— (2017b). "Isotonic designs for phase I trials in partially ordered groups". In: *Clinical Trials* 14.5, pp. 491–498.

Conaway, Mark R and Nolan A Wages (2017). "Designs for phase I trials in ordered groups". In: *Statistics in medicine* 36.2, pp. 254–265.

Horton, Bethany Jablonski, Nolan A Wages, and Mark R Conaway (2017). "Performance of toxicity probability interval based designs in contrast to the continual reassessment method". In: *Statistics in medicine* 36.2, pp. 291–300.

Lin, Ruitao and Guosheng Yin (2017). "Bayesian optimal interval design for dose finding in drug-combination trials". In: *Statistical methods in medical research* 26.5, pp. 2155–2167.

U.S. Department of Health and Human Services (Nov. 2017). *Common Toxicity Criteria for Adverse Events*.

Yan, Fangrong, Sumithra J Mandrekar, and Ying Yuan (2017). "Keyboard: a novel Bayesian toxicity probability interval design for phase I clinical trials". In: *Clinical Cancer Research* 23.15, pp. 3994–4003.

Kim, Guowei et al. (2018). "Study protocol: phase 1 dose escalating study of Pressurized Intra-Peritoneal Aerosol Chemotherapy (PIPAC) with oxaliplatin in peritoneal metastasis". In: *Pleura and Peritoneum* 3.3.

Yuan, Ying, Ruitao Lin, et al. (2018). "Time-to-event Bayesian optimal interval design to accelerate phase I trials". In: *Clinical Cancer Research* 24.20, pp. 4921–4930.

Cheung, Ken (2019). *dfcrm: Dose-Finding by the Continual Reassessment Method*. R package version 0.2-2.1. URL: https://CRAN.R-project.org/package=dfcrm.

Conaway, Mark R and Gina R Petroni (2019). "The impact of early-phase trial design in the drug development process". In: *Clinical Cancer Research* 25.2, pp. 819–827.

Horton, Bethany Jablonski, John O'Quigley, and Mark R Conaway (2019). "Consequences of performing parallel dose finding trials in heterogeneous groups of patients". In: *JNCI Cancer Spectrum* 3.2, pkz013.

Horton, Bethany Jablonski, Nolan A Wages, and Mark R Conaway (2019). "Shift models for dose-finding in partially ordered groups". In: *Clinical Trials* 16.1, pp. 32–40.

Mu, Rongji et al. (2019). "gBOIN: a unified model-assisted phase I trial design accounting for toxicity grades, and binary or continuous end points". In: *Journal of the Royal Statistical Society Series C: Applied Statistics* 68.2, pp. 289–308.

Yuan, Ying, J Jack Lee, and Susan G Hilsenbeck (2019). "Model-assisted designs for early-phase clinical trials: simplicity meets superiority". In: *JCO Precision Oncology* 3, pp. 1–12.

Zhou, Yanhong, J Jack Lee, and Ying Yuan (2019). "A utility-based Bayesian optimal interval (U-BOIN) phase I/II design to identify the optimal biological dose for targeted and immune therapies". In: *Statistics in Medicine* 38.28, S5299–S5316.

Lin, Ruitao, Peter F Thall, and Ying Yuan (2020). "An adaptive trial design to optimize dose-schedule regimes with delayed outcomes". In: *Biometrics* 76.1, pp. 304–315.

Muller, Donald A et al. (2020). "STAT RAD: prospective dose escalation clinical trial of single fraction scan-plan-QA-treat stereotactic body radiation therapy for painful osseous metastases". In: *Practical Radiation Oncology* 10.6, e444–e451.

Turner, Rolf (2020). *Iso: Functions to Perform Isotonic Regression*. R package version 0.0-18.1. URL: https://CRAN.R-project.org/package=Iso.

Yan, Fangrong, Liangcai Zhang, et al. (2020). "BOIN: An R Package for Designing Single-Agent and Drug-Combination Dose-Finding Trials Using Bayesian Optimal Interval Designs". In: *Journal of Statistical Software* 94.13, pp. 1–32. DOI: 10.18637/jss.v094.i13.

Ratain, Mark J, Ian F Tannock, and Allen S Lichter (2021). "Dose optimization of sotorasib: is the US Food and Drug Administration sending a message?" In: *Journal of Clinical Oncology* 39.31, pp. 3423–3426.

Shah, Mirat et al. (2021). "The drug-dosing conundrum in oncology—when less is more". In: *New England Journal of Medicine* 385.16, pp. 1445–1447.

Moon, Hanlim (2022). "FDA initiatives to support dose optimization in oncology drug development: the less may be the better". In: *Translational and Clinical Pharmacology* 30.2, p. 71.

Murphy, Ravindhi, Sarah Halford, and Stefan Nicholas Symeonides (2023). "Project Optimus, an FDA initiative: considerations for cancer drug development internationally, from an academic perspective". In: *Frontiers in Oncology* 13, p. 1144056.

O'Connell, Nathaniel S, Nolan A Wages, and Elizabeth Garrett-Mayer (2023). "Quasi-partial order continual reassessment method: Applying toxicity scores to cancer dose-finding drug combination trials". In: *Contemporary Clinical Trials* 125, p. 107050.

Papachristos, Apostolos et al. (2023). "Dose Optimization in Oncology Drug Development: The Emerging Role of Pharmacogenomics, Pharmacokinetics, and Pharmacodynamics". In: *Cancers* 15.12, p. 3233.

Celum, Connor and Mark Conaway (2024). "A model-assisted design for partially or completely ordered groups". In: *Pharmaceutical Statistics*.

Celum, Connor, Bethany Jablonski Horton, and Mark Conaway (2024). "The quasi-CRM shift method for partially ordered groups". In: *Contemporary Clinical Trials* 136, p. 107400.

Chiuzan, Cody and Hakim-Moulay Dehbi (2024). "The 3+ 3 design in dose-finding studies with small sample sizes: Pitfalls and possible remedies". In: *Clinical Trials*, p. 17407745241240401.

National Cancer Institute (2024). *Comprehensive Cancer Information - NCI*.

U.S. Food And Drug Administration (2024a). *Optimizing the Dosage of Human Prescription Drugs and Biological Products for the Treatment of Oncologic Diseases*.

— (2024b). *Project Optimus: Reforming the dose optimization and dose selection paradigm in oncology*. URL: https://www.fda.gov/about-fda/oncology-center-excellence/project-optimus (visited on 04/25/2024).

Yang, Peng et al. (2024). "Design and sample size determination for multiple-dose randomized phase II trials for dose optimization". In: *Statistics in Medicine* 43.15, pp. 2972–2986.

MD Anderson Cancer Center (n.d.). *Phases of clinical trials*.

Appendices

Appendix A

Appendix for Chapter 2

Figure A.1: Four-dose ET and DLT curves, with the ET scores/DLT probabilities by dose level.

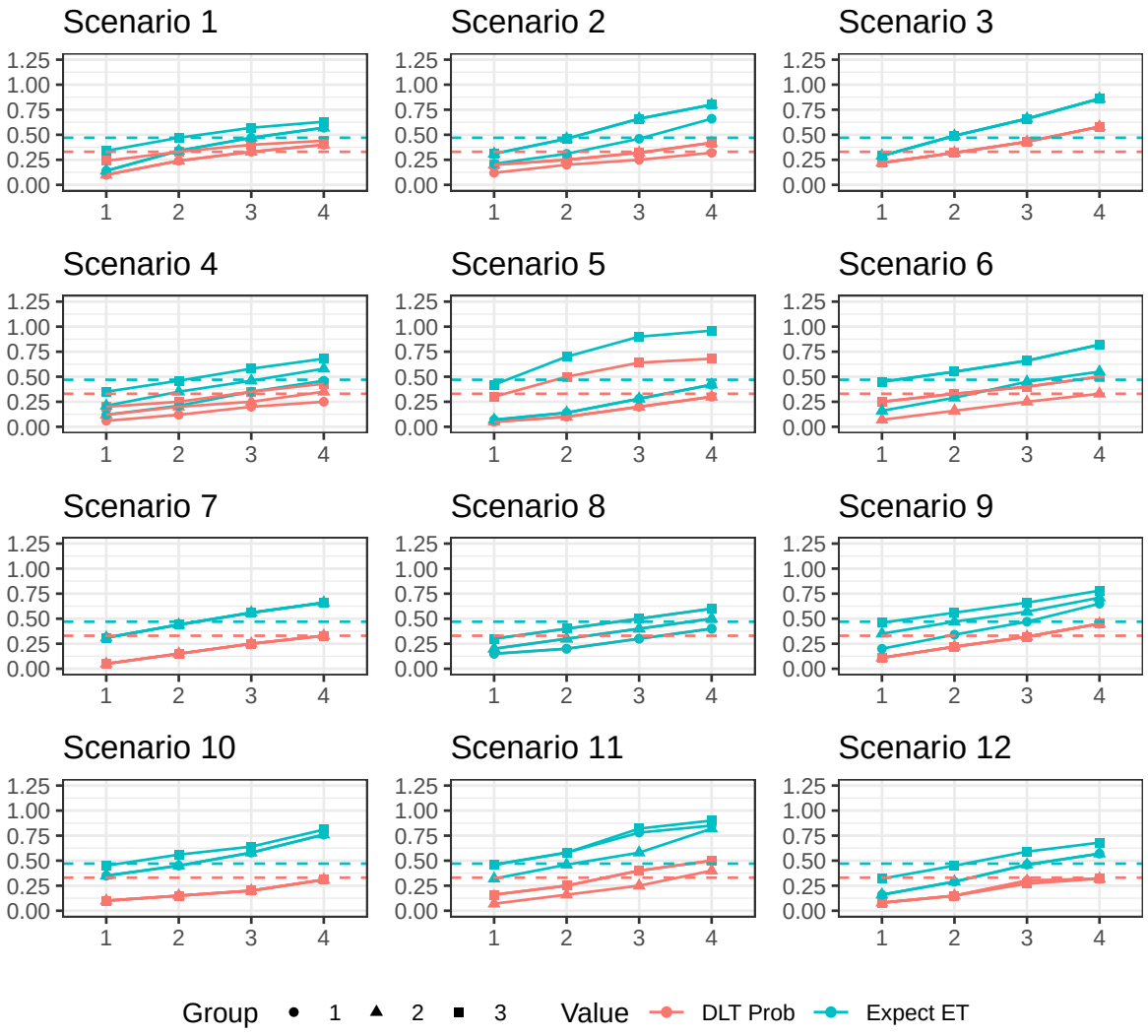
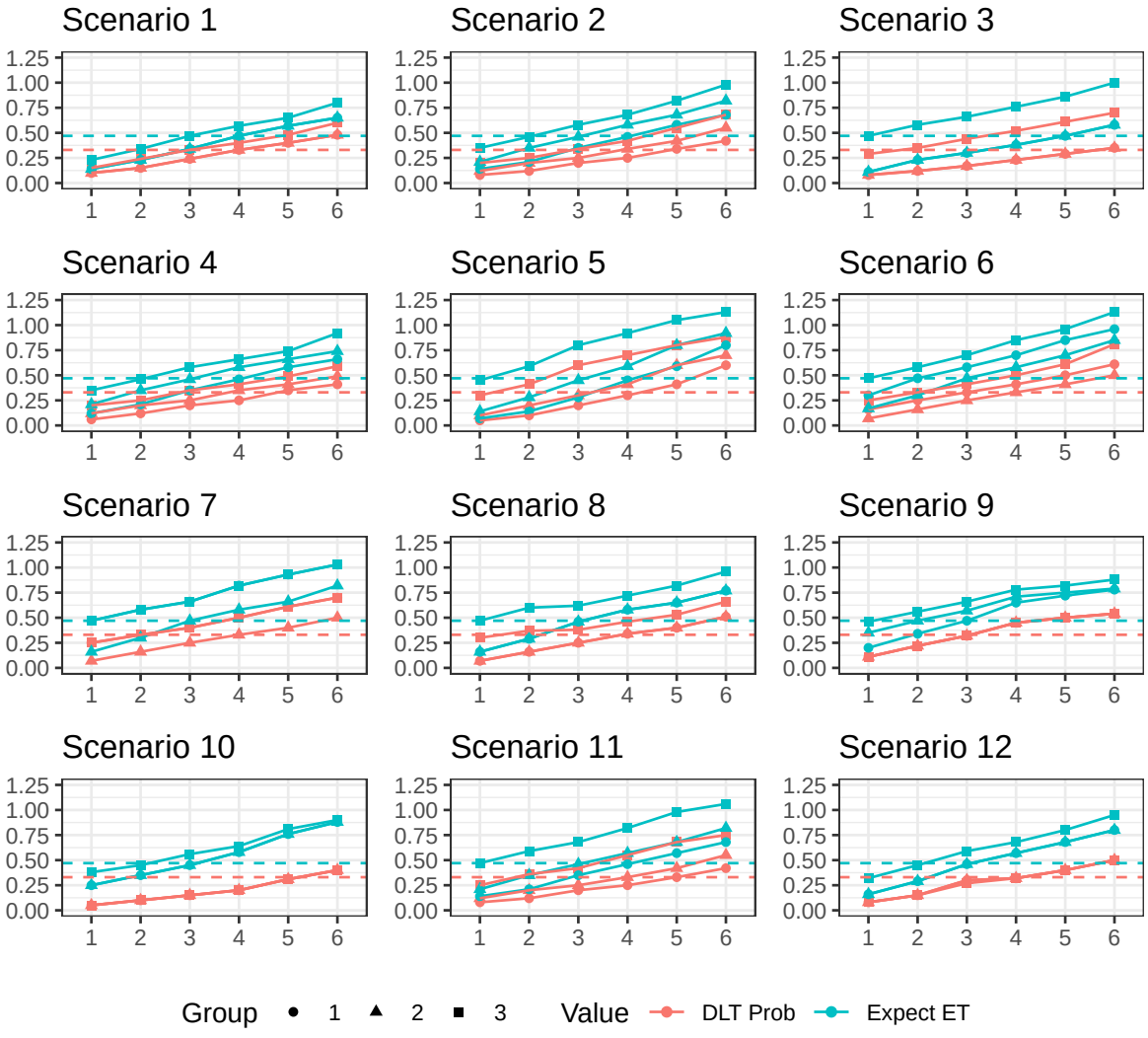


Figure A.2: Six-dose ET and DLT curves, with the ET scores/DLT probabilities by dose level.



Appendix B

Appendix for Chapter 3

B.1 | Proofs

The following proofs will use a similar notation to that in Oron, Azriel, and Hoff 2011. Let Ω consist of triples $(Y, (b, g), d_k)$, where Y is the binary DLT outcome, (b, g) is the bundle and group index, and d_k is the dose. For the i^{th} patient, let T_i denote the administered dose, (B_i, G_i) the group and bundle, and Y_i the DLT outcome. Note that T_i is a random variable since the responses from the previous patients determine the dose given to the i^{th} patient. We have that $Y(((b, g), d_k)) \sim Bern(\pi_{(b,g),k})$. Let Ω^{∞}_{GAB} denote all possible infinite sequences of triples under the design of GAB along with the requirement $n_{(b,g)} \to \infty$, for all (b, g).

We define the random set

$$S^{(b,g)} := \{k : n_{(b,g),k} \to \infty \text{ as } n \to \infty\}. \tag{B.1}$$

Here, $S^{(b,g)}$ is the random set that gives the dose levels that appear infinitely often for (b, g). It follows that $S^{(b,g)} \neq \emptyset$ since $n_{(b,g)} \to \infty$, for all (b, g).

B.1.1 | Lemma 1

First, decompose $\hat{\pi}_{(b,g),k}$ as

$$\hat{\pi}_{(b,g),k} = \pi_{(b,g),k} + \frac{1}{n_{(b,g),k}} \sum_{i=1}^{n} I[T_i = d_k \& (B_i, G_i) = (b, g)](\pi_{(b,g),k} - Y_i).$$

Then, define the random variable

$$M_n = \sum_{i=1}^{n} I[T_i = d_k \& (B_i, G_i) = (b, g)](\pi_{(b,g),k} - Y_i).$$

We can show M_n is a square integrable ($\mathbf{E}(M_n^2) < \infty$) martingale with filtration

$$\mathcal{F}_n = \sigma(Y_1, T_1, (B_1, G_1), \ldots, Y_n, T_n, (B_n, G_n)).$$

To show M_n is a martingale, we note

$$M_{n+1} = I[T_{n+1} = d_k \& (B_{n+1}, G_{n+1}) = (b, g)](\pi_{(b,g),k} - Y_{n+1}) + M_n,$$

giving

$$\mathbf{E}(M_{n+1}|\mathcal{F}_n) = \mathbf{E}(I[T_{n+1} = d_k \& (B_{n+1}, G_{n+1}) = (b, g)](\pi_{(b,g),k} - Y_{n+1})|\mathcal{F}_n) + M_n.$$

Given $\mathcal{F}_n$, the indicator found in the expectation is no longer random and, thus, can be pulled out of the expectation, giving

$$\mathbf{E}(M_{n+1}|\mathcal{F}_n) = (I[T_{n+1} = d_k \& (B_{n+1}, G_{n+1}) = (b, g)])\, \mathbf{E}((\pi_{(b,g),k} - Y_{n+1})|\mathcal{F}_n) + M_n.$$

If $I[T_{n+1} = d_k \& (B_{n+1}, G_{n+1}) = (b,g)] = 1$, then $Y_{n+1} \sim Bern(\pi_{(b,g),k})$, giving

$$(I[T_{n+1} = d_k \& (B_{n+1}, G_{n+1}) = (b,g)])\, \mathbf{E}((\pi_{(b,g),k} - Y_{n+1})|\mathcal{F}_n) = 0.$$

We have now shown $\mathbf{E}(M_{n+1}|\mathcal{F}_n) = M_n$, giving us that M_n is a martingale.

We will now find the quadratic variation of M_n, denote as $< M >_n$, and show $< M >_n \to \infty$, allowing us to apply the strong law of martingales (see Shiryaev 1996). By definition,

$$< M >_n = \sum_{j=1}^{n} \mathbf{E}((\Delta M_j)^2|\mathcal{F}_{j-1}),$$

where $\Delta M_j = M_j - M_{j-1}$. Expanding $(\Delta M_j)^2$ gives

$$\begin{aligned}(\Delta M_j)^2 &= \left(I[T_j = d_k \& (B_j, G_j) = (b,g)](\pi_{(b,g),k} - Y_j)\right)^2 \\ &= I[T_j = d_k \& (B_j, G_j) = (b,g)](\pi_{(b,g),k} - Y_j)^2.\end{aligned}$$

As before, given $\mathcal{F}_{j-1}$, we know the value of $I[T_j = d_k \& (B_j, G_j) = (b,g)]$, giving

$$\mathbf{E}\left((\Delta M_j)^2|\mathcal{F}_{j-1}\right) = I[T_j = d_k \& (B_j, G_j) = (b,g)]\mathbf{E}\left((\pi_{(b,g),k} - Y_j)^2|\mathcal{F}_{j-1}\right).$$

If $I\left[T_j = d_k \& (B_j, G_j) = (b,g)\right] = 0$, then $\mathbf{E}((\Delta M_j)^2|\mathcal{F}_{j-1}) = 0$. Conversely, if $I[T_j = d_k \& (B_j, G_j) = (b,g)] = 1$, then $Y_j \sim Bern(\pi_{(b,g),k})$ and $\mathbf{E}((\pi_{(b,g),k} - Y_j)^2|\mathcal{F}_{j-1}) = \pi_{(b,g),k}(1 - \pi_{(b,g),k})$. This gives

$$\mathbf{E}((\Delta M_j)^2|\mathcal{F}_{j-1}) = I[T_j = d_k \& (B_j, G_j) = (b,g)]\pi_{(b,g),k}(1 - \pi_{(b,g),k}),$$

implying

$$< M >_n = \pi_{(b,g),k}(1 - \pi_{(b,g),k})n_{(b,g),k} \propto n_{(b,g),k}$$

By our assumption, $n_{(b,g),k} \to \infty$ as $n \to \infty$, so $< M >_n \to \infty$ as $n \to \infty$. From the strong law of Martingales

$$\frac{M_n}{\pi_{(b,g),k}(1 - \pi_{(b,g),k})n_{(b,g),k}} = \frac{M_n}{< M_n >} \overset{a.s.}{\to} 0,$$

thus

$$\frac{1}{n_{(b,g),k}} \sum_{i=1}^{n} I[T_i = d_k \& G_i = g](\pi_{(b,g),k} - Y_i) = \frac{M_n}{n_{(b,g),k}} \overset{a.s.}{\to} 0.$$

This shows that $\hat{\pi}_{(b,g),k} \overset{a.s.}{\to} \pi_{(b,g),k}$, given $n_{(b,g),k} \to \infty$.

B.1.2 | Theorem 2

Remark B.1. The result will be proven when we have a unique k, for each (b, g), with $\pi_{(b,g),k} \in (\lambda_e, \lambda_d)$, and no other $\pi_{(b,g),k} \in [\lambda_e, \lambda_d]$. The logic for the proof in this unique case will extend to the other cases. For clarity, for (b, g), the unique dose level with DLT probability in (λ_e, λ_d) will be denoted as $k_{(b,g)}$.

Strong induction will be applied to the bundle index, first proving the result for bundle one, and then using induction to prove the result for the remaining bundles. First, note that $S^{(1,g)}$ is connected, that is, $S^{(1,g)}$ can only consist of consecutive dose levels. The reason we could not have, say, $S^{(1,g)} = \{1, 3\}$ is that after visiting level 3, $(1, g)$ must visit level 2 before visiting level 1, requiring $2 \in S^{(1,g)}$. From connectedness, $S^{(1,g)}$ can be listed as consecutive values, $S_1^{(1,g)}, \ldots, S_2^{(1,g)}$, where $S_1^{(1,g)}$ and $S_2^{(1,g)}$ are the minimum and maximum values, respectively, in $S^{(1,g)}$. Now consider any $g_1 \in \{1, 2, \ldots, G_1\}$ and partition Ω_{GAB}^{∞} into

events:

1. A: $S^{(1,g_1)} = k_{(1,g_1)}$
2. B: $S_1^{(1,g_1)} < S_2^{(1,g_1)}$ and $k_{(1,g_1)} \in S^{(1,g_1)}$
3. C: $k_{(1,g_1)} \notin S^{(1,g_1)}$

Showing allocation for $(1, g_1)$ converges almost surely to $k_{(1,g)}$ is equivalent to showing $P(A) = 1$. Thus, we will show $P(B) = P(C) = 0$. We partition C as

$$C = \{S_2^{(1,g_1)} < k_{(1,g_1)}\} \cup \{S_1^{(1,g_1)} > k_{(1,g_1)}\},$$

and first prove $P(\{S_2^{(1,g_1)} < k_{(1,g_1)}\} = 0$. Assume, without loss of generality, that $k_{(1,g_1)} \neq 1$, since otherwise $P(\{S_2^{(1,g_1)} < k_{(1,g_1)}\} = 0$. The event $\{S_2^{(1,g_1)} < k_{(1,g_1)}\}$ is further partitioned as

$$\{S_2^{(1,g_1)} < k_{(1,g_1)}\} = \bigcup_{k^\star : k^\star < k_{(1,g_1)}} \{S_2^{(1,g_1)} = k^\star\}.$$

From the partition above, it is sufficient to show $P(\{S_2^{(1,g_1)} = k^\star\}) = 0$, for any $k^\star < k_{(1,g_1)}$. To this end, consider any infinite sequence, $\omega \in \{S_2^{(1,g_1)} = k^\star\}$. As $k^\star \in S^{(1,g_1)}$, from Lemma 3.2, $\hat{\pi}_{(1,g_1),k^\star} \to \pi_{(1,g_1),k^\star}$, almost surely (a.s). Then, as $\pi_{(1,g_1),k^\star} < \lambda_e$, eventually it will always be the case $\hat{\pi}_{(1,g_1),k^\star} < \lambda_e$. As it is eventually always the case that $\hat{\pi}_{(1,g_1),k^\star} < \lambda_e$ and $(1, g_1)$ visits $k^\star$ infinitely often, $(1, g_1)$ must escalate from $k^\star$ infinitely often. After escalating from $k^\star$, the next patient in $(1, g_1)$ would be treated at $k^\star + 1$ or a higher dose, since other groups cannot bring down the recommended dose for $(1, g_1)$. This implies that an infinite number of $(1, g_1)$ patients are observed at doses higher than $k^\star$, contradicting $k^\star = S_2^{(1,g_1)}$. Thus, the only way for this to occur is if $\hat{\pi}_{(1,g_1),k} \not\to \pi_{(1,g_1),k}$, giving $P(\{S_2^{(1,g_1)} = k^\star\}) = 0$

Now we will show $P(\{S_1^{(1,g_1)} > k_{(1,g_1)}\}) = 0$. We can assume that $k_{(1,g_1)} \neq K$, otherwise

the result follows automatically. The event $\{S_1^{(1,g_1)} > k_{(1,g_1)}\}$ can be partitioned as

$$\{S_1^{(1,g_1)} > k_{(1,g_1)}\} = \bigcup_{k^\star : k^\star > k_{(1,g_1)}} \{S_1^{(1,g_1)} = k^\star\}.$$

It is sufficient to show $P(\{S_1^{(1,g_1)} = k^\star\}) = 0$, for any $k^\star > k_{(1,g_1)}$, and we consider any infinite sequence $\omega \in \{S_1^{(1,g_1)} = k^\star\}$. Since $k^\star \in S^{(1,g_1)}$, (a.s) we eventually always have $\hat{\pi}_{(1,g_1),k^\star} > \lambda_d$. At this point, whenever the last patient in $(1, g_1)$ is at $k^\star$, we will de-escalate if we do not have any (b, g_b) with $j_{(b,g_b)} = k^\star$, where $b > 1$. To simplify notation, for the remainder of the bundle 1 proof, whenever b appears, take $b > 1$. If there are a finite number to times when $j_{(b,g_b)} = k^\star$, for all b, $(1, g_1)$ will be observed at $k^\star - 1$ an infinite number of times, giving $k^\star - 1 \in S^{(1,g_1)}$, and contradicting $k^\star = S_1^{(1,g_1)}$.

Now assume there exists some, or several, (b, g_b) where $j_{(b,g_b)} = k^\star$ an infinite number of times. Begin by noting there could some (b', g'), $b' > 1$, when $j_{(b',g')} = k^\star$ an infinite number of times but only a finite number of patients in (b', g') are actually observed at $k^\star$. For this to occur, we would need a three-step cycle occurring infinitely often. First, a group from a lower bundle would de-escalate from $k^\star$, causing $j_{(b',g')} = k^\star - 1$. Second, a subsequent patient in (b', g') would be observed at, or below $k^\star - 1$. Third, the recommendation $j_{(b',g')}$ increases to $k^\star$ after an observation at $k^\star - 1$ from (b', g') or from some group (b'', g''), where $b'' > b'$. If the first case occurs infinitely often, $k^\star - 1 \in S^{(b',g')}$, thus $\hat{\pi}_{(b',g'),k^\star-1} \to \pi_{(b',g'),k^\star-1}$ (a.s). Then, as $\pi_{(b',g'),k^\star-1} > \pi_{(1,g_1),k^\star-1} > \lambda_e$, (a.s) eventually we always have $\hat{\pi}_{(b',g'),k^\star-1} > \lambda_e$, contradicting that for an infinite number of times $j_{(b',g')}$ increases to $k^\star$ after an observation from (b', g') at $k^\star - 1$. From this, the first case occurs with probability zero. Following the same steps, it can be shown that with probability zero $j_{(b',g')}$ increases to $k^\star$ infinitely many times after an observation from (b'', g''). Thus, the three-step cycle discussed occurs with probability zero.

Now assume that the groups, (b, g_b), with $j_{(b,g_b)} = k^\star$ infinitely often are observed at $k^\star$ infinitely often. From lemma 1, (a.s) eventually, we will always have Pooled $> \lambda_d$, implying

$j_{(1,g_1)} = k^\star - 1$ infinitely often. It is left to show that, with probability one, subsequent patients in $(1, g_1)$ will be observed at $k^\star - 1$. Suppose there are only a finite number of times when $(1, g_1)$ patients are observed at $k^\star - 1$ after $j_{(b,g_1)} = k^\star - 1$. For this to be the case, we would need some (b, g_b) escalating $j_{(1,g_1)}$ up from $k^\star - 1$ an infinite number of times, implying $k^\star - 1 \in S^{(b,g_b)}$. As $k^\star - 1 \in S^{(b,g_b)}$, (a.s) eventually we always have $\hat{\pi}_{(b,g_b),k^\star-1} > \lambda_e$, making it impossible for (b, g_b) to increase $j_{(1,g_1)}$ from $k^\star - 1$ to $k^\star$ an infinite number of times. Now, if there are an infinite number of patients observed at $k^\star - 1$ after $j_{(b,g_1)} = k^\star - 1$, we have $k^\star - 1 \in S^{(1,g_1)}$, contradicting $S_1^{(1,g_1)} = k^\star$. This shows that $P(S_1^{(1,g_1)} = k^\star) = 0$, for all $k^\star > k_{(1,g_1)}$. This concludes showing $P(C) = 0$

Next, we will show that $P(B) = 0$. If $k_{(1,g_1)} \in S^{(1,g_1)}$, (a.s) eventually we will always have $\hat{\pi}_{(1,g_1),k_{(1,g_1)}} \in (\lambda_e, \lambda_d)$ and the only way $(1, g_1)$ could move from $k_{(1,g_1)}$ is if a group from a higher bundle pushes the recommendation up. As $S^{(1,g_1)}$ is connected, we would then need $k_{(1,g_1)} + 1 \in S^{(1,g_1)}$. If this were the case, there would be some (b, g_b) escalating $j_{(1,g_1)}$ from $k_{(1,g_1)}$ to $k_{(1,g_1)} + 1$ an infinite number of times. As $b > 1$, $\pi_{(b',g'),k_{(1,g_1)}} > \pi_{(b,g),k_{(1,g_1)}}$. Additionally, as $k_{(b,g_b)} \in S^{(b,g_b)}$, (a.s) we eventually always have $\hat{\pi}_{(b,g_b),k_{(b,g_b)}} > \lambda_e$, contradicting that (b, g_b) moves $(1, g_1)$ from $k_{(1,g_1)}$ an infinite number of times. This shows that $P(B) = 0$.

We will now proceed using induction. Suppose for all $1 \leq l < b$, $P(S^{(l,g_l)} = k_{(l,g_l)}) = 1$, for all g_l. We will show that $P(S^{(b,g_b)} = k_{(b,g_b)}) = 1$.

Note sequences where $S^{(l,g_l)} = k_{(l,g_l)}$, for all $l < b$ and all g_l, $S^{(b,g_b)}$ is connected for all g_b. As a result, we can partition Ω_{avg}^{∞} as:

1. A: $S^{(b,g_b)} = k_{(b,g_b)}$ and $S^{(l,g_l)} = k_{(l,g_l)}$, for all $l < b$ and all g_l
2. B: $k_{(b,g_b)} \in S^{(b,g_b)}$, $S_1^{(b,g_b)} < S_2^{(b,g_b)}$ and $S^{(l,g_l)} = k_{(l,g_l)}$, for all $l < b$ and all g_l
3. C: $k_{(b,g_b)} \notin S^{(b,g_b)}$, and $S^{(l,g_l)} = k_{(l,g_l)}$, for all $l < b$ and all g_l
4. D: It is not the case that $S^{(l,g_l)} = k_{(l,g_l)}$, for all $l < b$ and all g_l

From our hypothesis, $P(D) = 0$, so we can assume $S^{(l,g_l)} = k_{(l,g_l)}$, for all $l < b$ and all g_l.

We will now show $P(C) = 0$. Suppose that we have $S_2^{(b,g_b)} = k^\star$, where $k^\star < k_{(b,g_b)}$. Eventually, we will always have $\hat{\pi}_{(b,g_b),k^\star} < \lambda_e$ (a.s) and all groups in bundles less than b will be a dose levels no lower than $k_{(b,g_b)}$. This will cause (b, g_b) to be observed at doses above $k^\star$ infinitely often, contradicting $k^\star = S_2^{(b,g_b)}$.

Suppose that we have $k^\star > k_{(b,g_b)}$, with $k^\star = S_1^{(b,g_b)}$. Then, (a.s) $\hat{\pi}_{(b,g_b),k^\star} \to \pi_{(b,g_b),k^\star} > \lambda_d$. The same logic from the bundle 1 proof will show that with probability 0 we will de-escalate only finitely many times. Thus, we have that $P(C) = 0$.

We now consider B and show $P(B) = 0$. As $k_{(b,g_b)} \in S^{(b,g_b)}$, (a.s) we will eventually always have $\hat{\pi}_{(b,g_b,k_{(b,g_b)}} \in (\lambda_e, \lambda_d)$. Eventually groups in lower bundles will not push (b, g_b) down from $k_{(b,g_b)}$ as such groups eventually stay at dose levels at or above $k_{(b,g_b)}$. Then, only way for (b, g_b) to change from $k_{(b,g_b)}$ is if a group from a higher bundle push it up. The same logic in the bundle one proof shows this happens with probability 0. Thus $P(B) = 0$, completing the proof.

B.2 | MTD Configuration

The tables in this section provide all possible configurations of MTDs amongst the groups in four dose trials. The MTD configurations are the possible group-specific MTDs under the known ordering. Table B.1 provides the MTD configurations when there are two completely ordered groups. Tables B.2, B.3, and B.4 provide the MTD configurations when there are three groups under orderings 1, 2, and 3, respectively.

MTD Configurations	
Group 1 MTD (γ_1)	Group 2 MTD (γ_2)
1	1
2	1
2	2
3	1
3	2
3	3
4	1
4	2
4	3
4	4

Table B.1: MTD configurations with two ordered groups and four doses

MTD Configurations		
Group 1 MTD (γ_1)	Group 2 MTD (γ_2)	Group 3 MTD (γ_3)
1	1	1
2	1	1
2	2	1
2	2	2
3	1	1
3	2	1
3	2	2
3	3	1
3	3	2
3	3	3
4	1	1
4	2	1
4	2	2
4	3	1
4	3	2
4	3	3
4	4	1
4	4	2
4	4	3
4	4	4

Table B.2: MTD configurations with three ordered groups and four doses under Order 1 (complete ordering)

MTD Configurations		
Group 1 MTD (γ_1)	Group 2 MTD (γ_2)	Group 3 MTD (γ_3)
1	1	1
2	1	1
2	2	1
2	2	2
2	1	2
3	1	1
3	2	1
3	2	2
3	3	1
3	3	2
3	3	3
3	1	2
3	1	3
3	2	3
4	1	1
4	2	1
4	2	2
4	3	1
4	3	2
4	3	3
4	4	1
4	4	2
4	4	3
4	4	4
4	1	2
4	1	3
4	2	3
4	1	4
4	2	4
4	3	4

Table B.3: MTD configurations with three ordered groups and four doses under Order 2 (Group 1 least sensitive)

B.3 | Generating Centers

In this section, we illustrate how centers are generated during the curve generation procedure. We will demonstrate how these curves are generated under the second three

MTD Configurations		
Group 1 MTD (γ_1)	Group 2 MTD (γ_2)	Group 3 MTD (γ_3)
1	1	1
1	2	1
1	3	1
1	4	1
2	1	1
2	2	1
2	2	2
2	3	1
2	3	2
2	4	1
2	4	2
3	1	1
3	2	1
3	2	2
3	3	1
3	3	2
3	3	3
3	4	1
3	4	2
3	4	3
4	1	1
4	2	1
4	2	2
4	3	1
4	3	2
4	3	3
4	4	1
4	4	2
4	4	3
4	4	4

Table B.4: MTD configurations with three ordered groups and four doses under Order 3 (Group 3 most sensitive)

group ordering with four dose levels. Suppose the we have the MTD configuration ($\gamma_1 = 4, \gamma_2 = 1, \gamma_3 = 3$). Let us generate three values $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$, where $\varepsilon \sim N(z(\theta), (0.05)^2) = N(z(0.2), (0.05)^2) = N(-0.8416212, 0.0025)$. We obtain $(\varepsilon_1 = -0.8273356, \varepsilon_2 = -0.8769697, \varepsilon_3 =$

-0.8000005) and consequently $(c_1 = \Phi(\varepsilon_1), c_2 = \Phi(\varepsilon_2), c_3 = \Phi(\varepsilon_3))$, where Φ is the standard normal CDF. From this, $(c_1 = 0.2040234, c_2 = 0.1902515, c_3 = 0.2118553)$. We now sort the centers as $(c_{[1]} = 0.1902515, c_{[2]} = 0.2040234, c_{[3]} = 0.2118553)$. As we are in order 2 with $\gamma_2 < \gamma_3$, then $(\pi_{1,\gamma_1}, \pi_{2,\gamma_2}, \pi_{3,\gamma_3}) = (c_{[1]}, c_{[3]}, c_{[2]})$. Thus, $(\pi_{1,4}, \pi_{2,1}, \pi_{3,3}) = (0.1902515, 0.2118553, 0.2040234)$

B.4 | GAB Illustration

These supplementary materials provide illustrative examples of trials conducted using GAB-E. By providing these examples, we aim to help clinicians understand the initial run-in period, allocation decisions, and MTD estimation. The first example considers two ordered groups with four dose levels while the second considers three partially ordered groups with six dose levels. As in the main paper, we consider the DLT target rate of $\theta = 0.2$, and the escalation and de-escalation boundaries of $\lambda_e = 0.157$ and $\lambda_d = 0.238$, respectively.

The patient responses and current dose recommendations are enumerated sequentially. The "Results" line states the group of the patient, the dose level received, and the patient outcome. Additionally, this line provides the updated current dose levels for each group. Comments are provided at several stages of the trials, stating how allocation decisions are determined using the rules from GAB-E. Table B.16 provides the dose elimination boundaries using the prior $Beta(0.5, 0.5)$, eliminating a dose if the probability a dose is excessively toxic is above 0.975. Note, this is the same as Table 3.2 in Chapter 3, and is provided below for convenience. We require at least three patients to be observed at a dose before eliminating it for a given group. After detailing the allocation procedure, we illustrate how group MTDs are estimated at the conclusion of the trial. For additional illustrative trials, see the R file Trial_Sims.R, which provides a function to generate worked trials.

B.4.1 Two Group Illustrative Example

First, we consider two ordered groups, where Group 1 is less sensitive than Group 2. The illustrative trial will have a total of 20 patients. After detailing dose allocation, we walk through MTD estimation.

Allocation

1. Results: Dose = 1, Group = 2, No DLT, $j_1 = NA$, $j_2 = 2$

   Comment: The first patient in Group 2 is assigned to dose level 1 and a DLT is not observed. Group 1 has not yet been observed, so we continue with standard BOIN for Group 2. At the current dose level, dose level 1, the DLT rate for Group 2 is $\hat{\pi}_{2,1} = 0/1 \leq 0.157$, thus Group 2 escalates to dose level 2, so $j_2 = 2$.

2. Results: Dose = 2, Group = 1, No DLT, $j_1 = 3$, $j_2 = 2$

   Comment: If Group 2 precedes Group 1, the first patient in Group 1 is assigned to the current Group 2 dose level, this being dose level 2. Thus, the first Group 1 patient is assigned to dose level 2. There is no DLT, so the current rate for Group 1 at dose level 2 is $\hat{\pi}_{1,2} = 0 \leq 0.157$. As a result, the current recommended dose level for Group 1 is 3, so $j_1 = 3$.

3. Results: Dose = 2, Group = 2, No DLT, $j_1 = 3$, $j_2 = 3$

   Comment: Group 2 is treated at dose level 2 and a DLT is not observed. At the current dose level, dose level 2, the DLT rate is $\hat{\pi}_{2,2} = 0/1 \leq 0.157$, thus Group 2 escalates to dose level 3, so $j_2 = 3$.

4. Results: Dose = 3, Group = 2, No DLT, $j_1 = 4$, $j_2 = 4$

Comment: Group 2 is treated at dose level 3 and a DLT is not observed. At the current dose level, dose level 3, the DLT rate is $\hat{\pi}_{2,3} = 0/1 \leq 0.157$, thus Group 2 escalates to dose level 4, so $j_2 = 4$. However, we now have $j_1 < j_2$, as $j_1 = 3$ and $j_2 = 4$. Thus, we consider a pooled average at the dose level Group 2 escalated from, this being dose level 3. We obtain $Pooled = (y_{1,3} + y_{2,3})/(n_{1,3} + n_{2,3}) = (0+0)/(0+1) = 0 \leq 0.157$, so Group 1 escalates to dose level 4, thus $j_1 = 4$.

5. Results: Dose = 4, Group = 2, DLT, $j_1 = 4$, $j_2 = 3$

   Comment: Group 2 is treated at dose level 4. There is a DLT, so the current DLT rate for Group 2 at dose level 4 is $\hat{\pi}_{2,4} = 1/1 \geq 0.238$. The recommendation for Group 2 decreases, so $j_2 = 3$

6. Results: Dose = 4, Group = 1, No DLT, $j_1 = 4$, $j_2 = 3$

   Comment: Group 1 is treated at dose level 4. There is no DLT, so the current DLT rate for Group 1 at dose level 4 is $\hat{\pi}_{1,4} = 0/1 \leq 0.157$. As dose level 4 is the highest dose level, Group 1 stays at dose level 4.

7. Results: Dose = 4, Group = 1, DLT, $j_1 = 3$, $j_2 = 3$

   Comment: Group 1 is treated at dose level 4. There is a DLT, so the current DLT rate for Group 1 at dose level 4 is $\hat{\pi}_{1,4} = 1/2 \geq 0.238$. As a result, Group 1 de-escalates a dose level, so $j_1 = 3$

8. Results: Dose = 3, Group = 1, No DLT, $j_1 = 4$, $j_2 = 3$

   Comment: Group 1 is treated at dose level 3. There is no DLT, so the current DLT rate for Group 1 at dose level 3 is $\hat{\pi}_{1,3} = 0/1 = 0 \leq 0.157$, so $j_1 = 4$

9. Results: Dose = 4, Group = 1, DLT, $j_1 = 3$, $j_2 = 3$

   Comment: Group 1 is treated at dose level 4. There is a DLT, so the current DLT rate for Group 1 at dose level 4 is $\hat{\pi}_{1,4} = 2/3 \geq 0.238$, so $j_1 = 3$

10. Results: Dose = 3, Group = 1, No DLT, $j_1 = 4$, $j_2 = 3$

    Comment: Group 1 is treated at dose level 3. There is no DLT, so the current DLT rate for Group 1 at dose level 3 is $\hat{\pi}_{1,4} = 0/2 = 0 \leq 0.157$, so $j_1 = 4$

11. Results: Dose = 4, Group = 1, DLT, $j_1 = 3$, $j_2 = 3$

    Comment: Group 1 is treated at dose level 4. There is a DLT, so the current DLT rate for Group 1 at dose level 4 is $\hat{\pi}_{1,4} = 3/4 \geq 0.238$, so $j_1 = 3$. Additionally, using Table B.16 and noting that $n_{1,4} = 4$ and $y_{1,4} = 3$, we consider eliminating dose level 4 from Group 1 since these values of $n_{1,4}$ and $y_{1,4}$ fall into the elimination boundary. However, we note that Group 2 is more sensitive, thus, if we eliminate dose level 4 from Group 1, we would also eliminate dose level 4 from Group 2. As a result, we consider $n_{Pooled} = n_{1,4} + n_{2,4} = 4 + 1 = 5$ and $y_{Pooled} = y_{1,4} + y_{2,4} = 3 + 1 = 4$. Then, using the Table B.16, we see that this combination of n_{Pooled} and y_{Pooled} fall into the elimination boundary, so dose level 4 is eliminated from both groups.

12. Results: Dose = 3, Group = 1, No DLT, $j_1 = 3$, $j_2 = 3$

    Comment: Group 1 is treated at dose level 3. There is no DLT, so the current DLT rate for Group 1 at dose level 3 is $\hat{\pi}_{1,4} = 0/3 = 0 \leq 0.157$. However, we do not escalate to dose 4 as this dose has been eliminated. As a result, Group 1 remains at dose level 3.

13. Results: Dose = 3, Group = 1, DLT, $j_1 = 3$, $j_2 = 3$

Comment: Group 1 is treated at dose level 3. There is a DLT, so the current DLT rate for Group 1 at dose level 3 is $\hat{\pi}_{1,4} = 1/4 = 0.25 \geq 0.238$. From this $j_1 = 2$. However, we now have $j_1 < j_2$. As a result, we consider a pooled average at the dose level group 1 deescalated from, that is, dose level 3. From this, $Pooled = (y_{1,3} + y_{2,3})/(n_{1,3} + n_{2,3}) = (1 + 0)/(1 + 4) = 1/5 = 0.2 < 0.238$. Thus, Group 1 returns to dose level 3 and $j_1 = 3$

14. Results: Dose = 3, Group = 2, DLT, $j_1 = 3$, $j_2 = 2$

    Comment: Group 2 is treated at dose level 3. There is a DLT, so the current DLT rate for Group 2 at dose level 3 is $\hat{\pi}_{2,3} = 1/2 \geq 0.238$, so $j_2 = 2$

15. Results: Dose = 3, Group = 1, No DLT, $j_1 = 3$, $j_2 = 2$

    Comment: Group 1 is treated at dose level 3. There is a no DLT, so the current DLT rate for Group 1 at dose level 3 is $\hat{\pi}_{1,4} = 1/5 = 0.2 \in (0.157, 0.0.238)$. As a result, Group 1 stays at dose level 3. Note that we would not consider increasing j_1 to dose level 4 since this dose level has been eliminated

16. Results: Dose = 2, Group = 2, No DLT, $j_1 = 3$, $j_2 = 3$

17. Results: Dose = 3, Group = 2, DLT, $j_1 = 3$, $j_2 = 2$

18. Results: Dose = 3, Group = 1, No DLT, $j_1 = 3, j_2 = 2$

19. Results: Dose = 2, Group = 2, No DLT, $j_1 = 3$, $j_2 = 3$

20. Results: Dose = 3, Group = 1, No DLT, $j_1 = 3, j_2 = 3$

MTD Estimation

Now that the trial has concluded, we use bivariate isotonic regression to obtain the final MTD estimates. First, we review the DLT observations, by group and dose level. The number of observations, $n_{g,k}$, by group and dose level, are provided in Table B.5.

$n_{g,k}$	Dose Level			
Group	1	2	3	4
1	0	1	7	4
2	1	3	3	1

Table B.5: Number of Observations by Group and Dose Level

Next, the number of DLTs, $y_{g,k}$, by group and dose level, are provided in Table B.6.

$y_{g,k}$	Dose Level			
Group	1	2	3	4
1	0	0	1	3
2	0	0	2	1

Table B.6: Number of DLTs by Group and Dose Level

Next, we obtained the smoothed DLT proportions, $\hat{\pi}^s_{g,k}$, which will later be input into the bivariate isotonic regression algorithm. The smoothed proportions are obtained as $\hat{\pi}^s_{g,k} = \frac{y_{g,k}+\alpha_{g,k}}{n_{g,k}+\alpha_{g,k}+\beta_{g,k}}$. In our analysis, we use smoothing parameters $\alpha_{g,k} = \beta_{g,k} = 0.05$, giving

$$\hat{\pi}^s_{g,k} = \frac{y_{g,k} + 0.05}{n_{g,k} + 0.1}.$$

Table B.7 provides the smoothed DLT proportions.

$\hat{\pi}^s_{g,k}$	Dose Level			
Group	1	2	3	4
1	0.5	0.04545455	0.1478873	0.7439024
2	0.04545455	0.01612903	0.6612903	0.9545455

Table B.7: Smoothed DLT Proportions by Group and Dose Level

The last input needed for bivariate isotonic regression is a matrix of weights. We use the weights $n_{g,k}+1$, making the weights proportional to the number of observations at a given dose level and group. Table B.8 provides these weights by group and dose level.

$n_{g,k}+1$	Dose Level			
Group	1	2	3	4
1	1	2	8	5
2	2	4	4	2

Table B.8: Weights for Bivariate Isotonic Regression by Group and Dose Level

Finally, using the smoothed proportions in Table B.7 and weights in Table B.8, we conduct bivariate isotonic regression. See Dykstra and Robertson 1982 for details on the bivariate isotonic regression algorithm. The function "biviso" function from the R package "Iso" is used to implement this algorithm (Turner 2020).

$\tilde{\pi}_{g,k}$	Dose Level			
Group	1	2	3	4
1	0.08292604	0.08292604	0.1478873	0.7439024
2	0.08292604	0.08292604	0.6612903	0.9545455

Table B.9: DLT Probability Estimates from Bivariate Isotonic Regression

Finally, we consider the admissible dose levels for each group. The admissible dose levels for Group 1, A_1, are the dose levels with observations from either group that have not been eliminated from Group 1. From this, $A_1 = \{1, 2, 3\}$. The admissible dose levels for Group 2, A_2, are the dose levels with observations from Group 2 that have not been eliminated from Group 2. Thus, $A_2 = \{1, 2, 3\}$.

The MTD estimate for Group 1 is admissible dose level with DLT estimate closest to the target 0.2. Mathematically, the MTD estimate for Group 1 is given by, k_1', where $k_1' = \text{argmin}_{k \in A_1} |\tilde{\pi}_{1,k} - 0.2| = 3$. Similarly, we obtain that the MTD estimate for Group 2 is k_2', where $k_2' = \text{argmin}_{k \in A_2} |\tilde{\pi}_{2,k} - 0.2| = 2$. At the end of the trial, dose level 2 is the MTD estimate for Group 2 and dose level 3 is the MTD estimate for Group 1.

B.4.2 Three Group Illustrative Example

Now we consider a worked example with three partially ordered groups. In this example, Group 1 is the least sensitive group while the relative ordering of Groups 2 and 3 is unknown. In the bundle framework, we have:

$$1 = (1,1)$$

$$2 = (2,1), 3 = (2,2)$$

The illustrative trial will have a total of 35 patients. For the rest of this example, we will use the bundle notation.

Allocation

1. Results: Dose = 1, Group = (1,1), No DLT, $j_{(1,1)} = 2, j_{(2,1)} = NA, j_{(2,2)} = NA$

   Comment: As we do not have any observations from either Groups $(2,1)$ or $(2,2)$, the first patient from Group $(1,1)$ is assigned to dose level 1. No DLT is observed, thus the DLT rate for Group $(1,1)$ at dose level 1 is $\hat{\pi}_{(1,1),1} = 0/1 \leq 0.157$. As a result, Group $(1,1)$ escalates to dose level 2, so $j_{(1,1)} = 2$.

2. Results: Dose = 1, Group = (2,2), No DLT, $j_{(1,1)} = 2, j_{(2,1)} = NA, j_{(2,2)} = 2$

   Comment: As Group $(2,2)$ is in the sensitive bundle, it cannot start at the current dose level for Group $(1,1)$, thus the first patient in Group $(2,2)$ is assigned to dose level 1. No DLT is observed, thus the DLT rate for Group $(2,2)$ at dose level 1 is $\hat{\pi}_{(2,2),1} = 0/1 \leq 0.157$. As a result, $(2,2)$ escalates to dose level 2, so $j_{(2,2)} = 2$.

3. Results: Dose = 2, Group = (1,1), No DLT, $j_{(1,1)} = 3, j_{(2,1)} = NA, j_{(2,2)} = 2$

4. Results: Dose = 3, Group = (1,1), No DLT, $j_{(1,1)} = 4, j_{(2,1)} = NA, j_{(2,2)} = 2$

5. Results: Dose = 1, Group = (2,1), No DLT, $j_{(1,1)} = 4, j_{(2,1)} = 2, j_{(2,2)} = 2$

   Comment: As Group $(2,1)$ is in the sensitive bundle, it starts at the lowest dose level, thus the first patient in Group $(2,1)$ is assigned to dose level 1. No DLT is observed, thus the DLT rate for Group $(2,1)$ at dose level 1 is $\hat{\pi}_{(2,1),1} = 0/1 \leq 0.157$. As a result, Group $(2,1)$ escalates to dose level 2, so $j_{(2,1)} = 2$.

6. Results: Dose = 2, Group = (2,2), No DLT, $j_{(1,1)} = 4, j_{(2,1)} = 2, j_{(2,2)} = 3$

   Comment: We observe Group $(2,2)$ at dose level 2. No DLT is observed, thus the DLT rate for Group $(2,2)$ at dose level 2 is $\hat{\pi}_{(2,2),2} = 0/1 \leq 0.157$. As a result, Group $(2,2)$ escalates to dose level 3, so $j_{(2,2)} = 3$. Note that $j_{(2,2)}$ is not constrained by $j_{(2,1)}$ as $j_{(2,2)}$ is only constrained by $j_{(1,1)}$, the current recommended dose for the less sensitive group.

7. Results: Dose = 2, Group = (2,1), No DLT, $j_{(1,1)} = 4, j_{(2,1)} = 3, j_{(2,2)} = 3$

8. Results: Dose = 4, Group = (1,1), No DLT, $j_{(1,1)} = 5, j_{(2,1)} = 3, j_{(2,2)} = 3$

9. Results: Dose = 3, Group = (2,1), No DLT, $j_{(1,1)} = 5, j_{(2,1)} = 4, j_{(2,2)} = 3$

10. Results: Dose = 4, Group = (2,1), DLT, $j_{(1,1)} = 5, j_{(2,1)} = 3, j_{(2,2)} = 3$

11. Results: Dose = 3, Group = (2,2), No DLT, $j_{(1,1)} = 5, j_{(2,1)} = 3, j_{(2,2)} = 4$

12. Results: Dose = 5, Group = (1,1), DLT, $j_{(1,1)} = 4, j_{(2,1)} = 3, j_{(2,2)} = 4$

13. Results: Dose = 4, Group = (2,2), No DLT, $j_{(1,1)} = 5, j_{(2,1)} = 3, j_{(2,2)} = 5$

    Comment: We observed Group $(2,2)$ at dose level 4. No DLT is observed and the current DLT rate for Group $(2,2)$ at dose level 4 is $\hat{\pi}_{(2,2),4} = 0/1 \leq 0.157$,

thus Group $(2,2)$ escalates to dose level 5. However, we now have $j_{(2,2)} = 5$ and $j_{(1,1)} = 4$, giving $j_{(1,1)} < j_{(2,2)}$. As a result, we consider a pooled average at the dose level Group $(2,2)$ escalated from, this being dose level 4. We have $Pooled = (y_{(1,1),4} + y_{(2,2),4})/(n_{(1,1),4} + n_{(2,2),4}) = (0+0)/(1+1) = 0 \leq 0.157$, thus Group $(1,1)$ also escalates to dose level 5, giving $j_{(1,1)} = 5$

14. Results: Dose = 5, Group = (1,1), DLT, $j_{(1,1)} = 4, j_{(2,1)} = 3, j_{(2,2)} = 4$

    Comment: Group $(1,1)$ is treated at dose level 5. A DLT is observed and the current DLT rate for Group $(1,1)$ at dose level 5 is $\hat{\pi}_{(1,1),5} = 2/2 = 1 \geq 0.238$, thus Group $(1,1)$ de-escalates to dose level 4. However, we now have $j_{(1,1)} < j_{(2,2)}$, so we consider a pooled average at the dose level Group $(1,1)$ de-escalated from, this being dose level 5. We have $Pooled = (y_{(1,1),5}+y_{(2,2),5})/(n_{(1,1),5}+n_{(2,2),5}) = (2+0)/(2+0) = 1 \geq 0.238$, so Group $(2,2)$ de-escalates to dose level 4.

15. Results: Dose = 4, Group = (1,1), No DLT, $j_{(1,1)} = 5, j_{(2,1)} = 3, j_{(2,2)} = 4$

16. Results: Dose = 5, Group = (1,1), No DLT, $j_{(1,1)} = 4, j_{(2,1)} = 3, j_{(2,2)} = 4$

17. Results: Dose = 3, Group = (2,1), No DLT, $j_{(1,1)} = 4, j_{(2,1)} = 4, j_{(2,2)} = 4$

18. Results: Dose = 4, Group = (1,1), No DLT, $j_{(1,1)} = 5, j_{(2,1)} = 4, j_{(2,2)} = 4$

19. Results: Dose = 4, Group = (2,2), No DLT, $j_{(1,1)} = 5, j_{(2,1)} = 4, j_{(2,2)} = 5$

20. Results: Dose = 5, Group = (2,2), No DLT, $j_{(1,1)} = 5, j_{(2,1)} = 4, j_{(2,2)} = 5$

    Comment: Group $(2,2)$ is treated at dose level 5. A DLT is not observed. Thus, the current DLT rate for Group $(2,2)$ at dose level 5 is $\hat{\pi}_{(2,2),5} = 0/1 = 0 \leq 0.157$, from this Group $(2,2)$ escalates dose levels so $j_{(2,2)} = 6$. However, as $j_{(1,1)} = 5$, we now have $j_{(1,1)} < j_{(2,2)}$. As a result, we consider a pooled average at the dose level Group $(2,2)$ escalated from, this being dose level 5. Then, $Pooled =$

$(y_{(1,1),5}+y_{(2,2),5})/(n_{(1,1),5}+n_{(2,2),5}) = (2+0)/(3+1) = 0.5$. As it is not the case that $Pooled \leq 0.157$, we return Group $(2,2)$ to dose level 5.

21. Results: Dose = 4, Group = (2,1), DLT, $j_{(1,1)} = 5, j_{(2,1)} = 3, j_{(2,2)} = 5$

22. Results: Dose = 3, Group = (2,1), DLT, $j_{(1,1)} = 5, j_{(2,1)} = 2, j_{(2,2)} = 5$

23. Results: Dose = 2, Group = (2,1), DLT, $j_{(1,1)} = 5, j_{(2,1)} = 1, j_{(2,2)} = 5$

24. Results: Dose = 5, Group = (2,2), DLT, $j_{(1,1)} = 5, j_{(2,1)} = 1, j_{(2,2)} = 4$

25. Results: Dose = 4, Group = (2,2), No DLT, $j_{(1,1)} = 5, j_{(2,1)} = 1, j_{(2,2)} = 5$

26. Results: Dose = 5, Group = (2,2), DLT, $j_{(1,1)} = 5, j_{(2,1)} = 1, j_{(2,2)} = 4$

27. Results: Dose = 1, Group = (2,1), No DLT, $j_{(1,1)} = 5, j_{(2,1)} = 2, j_{(2,2)} = 4$

28. Results: Dose = 4, Group = (2,2), No DLT, $j_{(1,1)} = 5, j_{(2,1)} = 2, j_{(2,2)} = 5$

29. Results: Dose = 5, Group = (1,1), DLT, $j_{(1,1)} = 4, j_{(2,1)} = 2, j_{(2,2)} = 4$

Comment: Group $(1,1)$ is treated at dose level 5 and a DLT is observed. The current DLT rate for Group $(1,1)$ at dose level 5 is $\hat{\pi}_{(1,1),5} = 3/4 \geq 0.257$, so Group $(1,1)$ de-escalates. However, as $j_{(1,1)} = 4$, we now have $j_{(1,1)} < j_{(2,2)}$, so we consider a pooled average at the dose level Group $(1,1)$ de-escalated from, this being dose level 5. Then, $Pooled = (y_{(1,1),5} + y_{(2,2),5})/(n_{(1,1),5} + n_{(2,2),5}) = (3+2)/(4+3) = 5/7$. As it is not the case that $Pooled \leq 0.257$, Group $(2,2)$ also de-escalates to dose level 4. Additionally, using Table B.16, noting $n_{(1,1),5} = 4$ and $y_{(1,1),5} = 3$, we consider eliminating dose level 5, and all higher dose levels, from Group $(1,1)$. However, as Group $(1,1)$ is the least sensitive group, if we eliminate dose level 5 from Group $(1,1)$, dose level 5 would be eliminated for both Groups $(2,1)$ and $(2,2)$. As a result, we consider $n_{Pooled} = n_{(1,1),5} + n_{(1,2),5} + n_{(2,1),5} = 4+0+3 = 7$ and $y_{Pooled} = y_{(1,1),5} + y_{(1,2),5} + y_{(2,1),5} = 3+0+2 = 5$. Using Table B.16, we see that this

combination of n_{Pooled} and y_{Pooled} falls into the elimination boundary. Thus, dose levels 5 and 6 are eliminated for all groups.

30. Results: Dose = 2, Group = (2,1), No DLT, $j_{(1,1)} = 4, j_{(2,1)} = 1, j_{(2,2)} = 4$

31. Results: Dose = 4, Group = (1,1), No DLT, $j_{(1,1)} = 4, j_{(2,1)} = 1, j_{(2,2)} = 4$

    Comment: Note, the although the current DLT rate for Group $(1, 1)$ at dose level 4 is $\hat{\pi}_{(1,1),4} = 0 \leq 0.157$, Group $(1, 1)$ does not escalate to dose level 5 as this dose level has been eliminated.

32. Results: Dose = 1, Group = (2,1), No DLT, $j_{(1,1)} = 4, j_{(2,1)} = 2, j_{(2,2)} = 4$

33. Results: Dose = 2, Group = (2,1), DLT, $j_{(1,1)} = 4, j_{(2,1)} = 1, j_{(2,2)} = 4$

34. Results: Dose = 1, Group = (2,1), No DLT, $j_{(1,1)} = 4, j_{(2,1)} = 2, j_{(2,2)} = 4$

35. Results: Dose = 2, Group = (2,1), DLT, $j_{(1,1)} = 4, j_{(2,1)} = 1, j_{(2,2)} = 4$

    We observed Group $(2, 1)$ at dose level 2 and a DLT was observed. The DLT rate for Group $(2, 1)$ at dose level 2 resulted in $j_{(2,1)}$ decreasing to dose level 1. Referencing Table B.16, we note that $n_{(2,1),2} = 5$ and $y_{(2,1),2} = 3$, resulting in dose level 2, and all higher dose levels, being eliminated for Group $(2, 1)$. Dose elimination for Group $(2, 1)$ does not affect dose elimination for other groups as Group $(2, 1)$ is in the sensitive bundle, as a result, we do not need to "pool" values before eliminating this dose level.

MTD Estimation

As the trial has concluded, DLT estimates will be obtained by applying bivariate isotonic regression to all possible complete orders given the known partial order. The DLT estimates with the highest likelihood will be utilized for dose selection.

First, we review the DLT observation, by group and dose level. The number of observation, $n_{(b,g),k}$, by Group and dose level are provided in Table B.10.

$n_{(b,g),k}$	Dose Level					
Group	1	2	3	4	5	6
(1,1)	1	1	1	4	4	0
(2,1)	4	5	3	2	0	0
(2,2)	1	1	1	4	3	0

Table B.10: Number of Observations by Group and Dose Level

Next, the number of DLTs, $y_{(b,g),k}$, by Group and dose level are provided in Table B.11.

$y_{(b,g),k}$	Dose Level					
Group	1	2	3	4	5	6
(1,1)	0	0	0	0	3	0
(2,1)	0	3	1	2	0	0
(2,2)	0	0	0	0	2	0

Table B.11: Number of DLTs by Group and Dose Level

Next, smoothed proportions, $\hat{\pi}^s_{(b,g),k}$, are obtained as

$$\hat{\pi}^s_{(b,g),k} = \frac{y_{(b,g),k} + 0.05}{n_{(b,g),k} + 0.1}.$$

Table B.12 provides the smoothed DLT proportions by group and dose level.

$\hat{\pi}^s_{(b,g),k}$	Dose Level					
Group	1	2	3	4	5	6
(1,1)	0.04545455	0.04545455	0.04545455	0.01219512	0.7439024	0.5
(2,1)	0.01219512	0.5980392	0.3387097	0.9761905	0.5	0.5
(2,2)	0.04545455	0.04545455	0.04545455	0.01219512	0.6612903	0.5

Table B.12: Smoothed DLT Proportions by Group and Dose Level

Similar to the two group example, we will use the weights $n_{(b,g),k} + 1$ for the bivariate isotonic regression algorithm. These weights are provided in Table B.13.

$n_{(b,g),k}+1$	Dose Level					
Group	1	2	3	4	5	6
(1,1)	2	2	2	5	5	1
(2,1)	5	6	4	3	1	1
(2,2)	2	2	2	5	4	1

Table B.13: Number of Observations by Group and Dose Level

In our example there are two bundles, the first with one group and the second with two groups. Thus, from permuting the two groups in bundle two, we obtain two complete orders:

$$\pi_{(1,1),k} \leq \pi_{(2,1),k} \leq \pi_{(2,2),k}$$

$$\pi_{(1,1),k} \leq \pi_{(2,2),k} \leq \pi_{(2,1),k}$$

Let $m = 1$ denote the model obtained from the first complete order and $m = 2$ denote the model obtained from the second complete order. We will then apply bivariate isotonic regression to regression to the two matrices below

$$\begin{array}{cccccc}
\hat{\pi}^s_{(1,1),1} & \hat{\pi}^s_{(1,1),2} & \hat{\pi}^s_{(1,1),3} & \hat{\pi}^s_{(1,1),4} & \hat{\pi}^s_{(1,1),5} & \hat{\pi}^s_{(1,1),6} \\
\hat{\pi}^s_{(2,1),1} & \hat{\pi}^s_{(2,1),2} & \hat{\pi}^s_{(2,1),3} & \hat{\pi}^s_{(2,1),4} & \hat{\pi}^s_{(2,1),5} & \hat{\pi}^s_{(2,1),6} \\
\hat{\pi}^s_{(2,2),1} & \hat{\pi}^s_{(2,2),2} & \hat{\pi}^s_{(2,2),3} & \hat{\pi}^s_{(2,2),4} & \hat{\pi}^s_{(2,2),5} & \hat{\pi}^s_{(2,2),6}
\end{array}$$

$$\begin{array}{cccccc}
\hat{\pi}^s_{(1,1),1} & \hat{\pi}^s_{(1,1),2} & \hat{\pi}^s_{(1,1),3} & \hat{\pi}^s_{(1,1),4} & \hat{\pi}^s_{(1,1),5} & \hat{\pi}^s_{(1,1),6} \\
\hat{\pi}^s_{(2,2),1} & \hat{\pi}^s_{(2,2),2} & \hat{\pi}^s_{(2,2),3} & \hat{\pi}^s_{(2,2),4} & \hat{\pi}^s_{(2,2),5} & \hat{\pi}^s_{(2,2),6} \\
\hat{\pi}^s_{(2,1),1} & \hat{\pi}^s_{(2,1),2} & \hat{\pi}^s_{(2,1),3} & \hat{\pi}^s_{(2,1),4} & \hat{\pi}^s_{(2,1),5} & \hat{\pi}^s_{(2,1),6},
\end{array}$$

that is,

$$\begin{matrix} 0.0454 & 0.0454 & 0.0454 & 0.0121 & 0.744 & 0.5 \\ 0.0121 & 0.598 & 0.339 & 0.976 & 0.5 & 0.5 \\ 0.0454 & 0.0454 & 0.0454 & 0.0121 & 0.611 & 0.5 \end{matrix}$$

$$\begin{matrix} 0.0454 & 0.0454 & 0.0454 & 0.0121 & 0.744 & 0.5 \\ 0.0454 & 0.0454 & 0.0454 & 0.0121 & 0.661 & 0.5 \\ 0.0121 & 0.598 & 0.339 & 0.976 & 0.5 & 0.5, \end{matrix}$$

From applying bivariate isotonic regression to the first set of matrices, we obtain the DLT estimates for model 1, as provided in Table B.14.

$\tilde{\pi}^{m=1}_{(b,g),k}$	Dose Level					
Group	1	2	3	4	5	6
(1,1)	0.02169781	0.02697709	0.02697709	0.02697709	0.6434364	0.6434364
(2,1)	0.02169781	0.36606373	0.36606373	0.37369339	0.6434364	0.6434364
(2,2)	0.04545455	0.36606373	0.36606373	0.37369339	0.6434364	0.6434364

Table B.14: DLT Estimates by Group and Dose Level for Model 1

From applying bivariate isotonic regression to the second set of matrices, we obtain the DLT estimates for model 2, as provided in Table B.15.

$\tilde{\pi}^{m=2}_{(b,g),k}$	Dose Level					
Group	1	2	3	4	5	6
(1,1)	0.02697709	0.02697709	0.02697709	0.02697709	0.6695157	0.6695157
(2,2)	0.02697709	0.02697709	0.02697709	0.02697709	0.6695157	0.6695157
(2,1)	0.02697708	0.49430740	0.49430740	0.78571430	0.7857143	0.7857143

Table B.15: DLT Estimates by Group and Dose Level for Model 2

We then select the model m', that maximizes the likelihood, so $m' = \text{argmax}_m L\left(\tilde{\pi}^m_{(b,g),k}\right)$

Here,

$$L\left(\tilde{\pi}^{m}_{(b,g),k}\right) = \prod_{k=1}^{6} \tilde{\pi}^{m}_{(1,1),k}{}^{y_{(1,1),k}}(1-\tilde{\pi}^{m}_{(1,1),k})^{n_{(1,1),k}-y_{(1,1),k}} \times \prod_{k=1}^{6} \tilde{\pi}^{m}_{(2,1),k}{}^{y_{(2,1),k}}(1-\tilde{\pi}^{m}_{(2,1),k})^{n_{(2,1),k}-y_{(2,1),k}} \times \prod_{k=1}^{6} \tilde{\pi}^{m}_{(2,2),k}{}^{y_{(2,2),k}}(1-\tilde{\pi}^{m}_{(2,2),k})^{n_{(2,2),k}-y_{(2,2),k}}$$

We obtain that $L\left(\tilde{\pi}^{m=1}_{(b,g),k}\right) = 2.54\times10^{-07}$ and $L\left(\tilde{\pi}^{m=2}_{(b,g),k}\right) = 2.16\times10^{-05}$, thus we use DLT estimates from model 2 for dose selection.

Now, let $A_{(b,g)}$ denote the admissible dose levels for Group (b,g). We obtain $A_{(1,1)} = \{1,2,3,4\}$, $A_{(2,1)} = \{1\}$, and $A_{(2,2)} = \{1,2,3,4\}$. We then select the admissible dose levels with DLT estimates closest to 0.2. Thus, the MTD estimate for Group $(1,1)$ is 4, the MTD estimate for Group $(2,1)$ is 1, and the MTD estimate for Group $(2,2)$ is 4.

B.4.3 Dose Elimination Boundaries

Number of Patients Treated	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Elimination Boundary	NA	NA	3	3	3	4	4	5	5	5	6	6	6	6	7

Table B.16: Dose elimination boundaries based on number of patients treated and number of DLTS. For these boundaries we require that more than 3 patients have been observed. Additionally, we use the prior Beta(0.5,0.5) and the cutoff $\lambda = 0.975$.

B.5 | Sensitivity Analysis

The following supplemental materials provide sensitivity analysis results. Tables B.17, B.18, and B.19 look at model sensitivity to cohort size and unequal groups. The Cohorts column states if the simulation considers single patient cohorts or cohorts of three. The Group

Probs column provides the probability that the next patient belongs to Group 1, Group 2, or Group 3. Table B.17 provides dose selection results, providing the statistics PCS and AI. Table B.18 provides dose allocation results, providing the statistics PCA and Int. Table B.19 provides early termination results, providing the percentage of trials in which a group is terminated from the trial. Overall, these results show the proposed method is robust.

In Section 3.5, we consider adding dose-elimination rules. To add dose-elimination rules, we assume $\pi \sim Beta(0.5, 0.5)$ and eliminate doses if the probability of excessive probability exceeds $\lambda = 0.975$. In this sensitivity analysis, we test sensitivity to the prior $Beta(0.5, 0.5)$ (called Jeffery's Prior) and the choice of the cutoff $\lambda = 0.975$. To test model sensitivity, we also consider the results from using a $Beta(1, 1)$ (equivalent to a Uniform) prior and using a cutoff of $\lambda = 0.95$. In these simulations, for simplicity, we consider trials with single-patient cohorts. Tables B.20 and B.21, provide selection and allocation results, respectively. These results show the proposed method is robust to the choice of prior.

Selection													
			Cohorts			**PCS**				**AI**			
G	Scenario	D	Size	Number	Group Probs	GAB	PB	GAB-E	PB-E	GAB	PB	GAB-E	PB-E
2	C	4	1	30	(.5,.5)	50	45.9	47.8	44.2	45.8	39.3	44.3	38.4
2	C	4	1	30	(.4,.6)	49.9	45.8	47.5	43.6	45.7	39	44.1	37.6
2	C	4	3	10	(.5,.5)	47.2	44.3	46.2	43	41.1	35	41	34.4
2	C	4	1	50	(.5,.5)	55.3	52	52	49.6	53.1	48.3	50.8	46.8
2	C	4	1	50	(.4,.6)	55.5	52	52.2	48.9	53.3	48.2	51.4	46.4
2	C	4	3	17	(.5,.5)	54.4	50.9	52.5	49.5	51.6	46.4	50.3	45.7
2	C	4	1	70	(.5,.5)	59.7	56.7	55.7	52.8	58.1	54.3	55.8	51.3
2	C	4	1	70	(.4,.6)	59.6	56.6	55.7	53.5	57.9	54	55.9	52.3
2	C	4	3	24	(.5,.5)	59.2	56.6	56.5	54.3	57.6	53.9	56.2	52.6
2	C	6	1	40	(.5,.5)	41.3	38.6	40.1	37.1	52.1	46.6	51.9	46
2	C	6	1	40	(.4,.6)	41.6	38.8	39.9	37.2	52.1	46.3	51.4	45.8
2	C	6	3	14	(.5,.5)	38.1	35.5	37.1	34.7	47.7	41.7	47.2	41.1
2	C	6	1	60	(.5,.5)	45.9	43.6	44.3	42.1	58.1	53.8	57.4	53.3
2	C	6	1	60	(.4,.6)	46.5	43.9	44.5	42.5	58.2	53.7	57.2	53.2
2	C	6	3	20	(.5,.5)	43.5	41.1	42	39.6	55.1	50.2	54.4	49.3
2	C	6	1	80	(.5,.5)	50.4	48.1	47.3	45.7	62.6	59.2	60.9	57.9

2	C	6	1	80	(.4,.6)	50.6	48.4	47.7	45.8	62.6	58.9	61	57.6
2	C	6	3	27	(.5,.5)	48.5	46.1	46.7	44.5	60.8	56.7	59.8	55.9
3	1	4	1	40	(.33,.33,.33)	52.2	46.6	50.4	44.8	48.3	37.8	47.5	36.7
3	1	4	1	40	(.2,.4,.4)	52.2	46.8	50.3	44.7	47.8	37.8	47.3	36.5
3	1	4	3	14	(.33,.33,.33)	50.1	44.3	48.9	43.3	46	35.4	45.3	34.7
3	1	4	1	60	(.33,.33,.33)	56.8	51.4	54.2	49	53.9	45	52.5	43.6
3	1	4	1	60	(.2,.4,.4)	56.5	51.4	54.2	48.6	53.4	44.7	52.5	43.5
3	1	4	3	20	(.33,.33,.33)	55.2	49.7	53.5	48.1	52.3	43	51.5	42.3
3	1	4	1	80	(.33,.33,.33)	60.4	55.3	57.2	52.6	57.9	50.2	56.3	48.8
3	1	4	1	80	(.2,.4,.4)	60.5	55.4	57.1	52.2	57.8	50.1	56.1	48.5
3	1	4	3	27	(.33,.33,.33)	59.4	54.1	57.2	52.2	57.2	49.1	56.1	48.3
3	1	6	1	60	(.33,.33,.33)	44.1	39.6	42.6	37.9	55.7	47.4	54.9	46.5
3	1	6	1	60	(.2,.4,.4)	44.3	39.5	42.6	37.5	55.3	46.2	54.6	45.5
3	1	6	3	20	(.33,.33,.33)	39.4	35	38.2	34.1	50.2	40.9	49.7	40.6
3	1	6	1	80	(.33,.33,.33)	47.6	43.2	45.5	41.3	59.6	52.5	58.8	51.5
3	1	6	1	80	(.2,.4,.4)	47.8	43	45.5	40.9	59.2	51.6	58.2	50.7
3	1	6	3	27	(.33,.33,.33)	44.1	39.6	42.5	38.2	56.2	48	55.5	47.5
3	1	6	1	100	(.33,.33,.33)	50.3	46.4	47.6	43.8	62.5	56.4	61.2	55.2
3	1	6	1	100	(.2,.4,.4)	50.5	46.2	47.7	43.7	62.2	55.6	61	54.8
3	1	6	3	34	(.33,.33,.33)	47.8	43.4	46	41.7	60.5	53.4	59.6	52.7
3	2	4	1	40	(.33,.33,.33)	50.5	46.4	48.8	44.7	45.2	38.5	44.4	37.4
3	2	4	1	40	(.2,.4,.4)	50.7	46.4	48.6	44.3	45.6	38.3	44.6	37.2
3	2	4	3	14	(.33,.33,.33)	47.8	43.5	46.6	42.4	41.1	33.8	40.4	33.3
3	2	4	1	60	(.33,.33,.33)	54.8	51	52.4	48.9	51.3	45.6	50	44.5
3	2	4	1	60	(.2,.4,.4)	55.2	51.1	52.4	48.4	51.6	45.4	50.3	44
3	2	4	3	20	(.33,.33,.33)	53	48.8	51.4	47.4	48.4	42.1	47.7	41.4
3	2	4	1	80	(.33,.33,.33)	58.6	55	55.1	51.9	55.9	50.8	53.7	49
3	2	4	1	80	(.2,.4,.4)	59	55	55.3	51.7	56.2	50.6	54.1	48.9
3	2	4	3	27	(.33,.33,.33)	57.3	53.5	55.4	51.7	54.2	48.8	53.3	47.8
3	2	6	1	60	(.33,.33,.33)	43	39.8	41.5	38.1	53.1	47.2	52.3	46.3
3	2	6	1	60	(.2,.4,.4)	43.6	39.6	41.7	37.9	53.3	46.2	52.4	45.5
3	2	6	3	20	(.33,.33,.33)	37.9	34.9	36.9	34	46.6	40.3	46.1	39.9
3	2	6	1	80	(.33,.33,.33)	46.7	43.5	44.5	41.6	57.4	52.4	56.4	51.4
3	2	6	1	80	(.2,.4,.4)	47.2	43.5	44.8	41.3	57.6	51.8	56.5	50.8
3	2	6	3	27	(.33,.33,.33)	42.9	39.7	41.4	38.2	53.3	47.7	52.6	47
3	2	6	1	100	(.33,.33,.33)	49.7	46.8	47	44.3	60.8	56.4	59.3	55.2
3	2	6	1	100	(.2,.4,.4)	50.2	46.7	47.4	44.1	60.9	55.8	59.5	54.6

3	2	6	3	34	(.33,.33,.33)	46.7	43.7	45	41.9	57.9	53.1	57.1	52.3
3	3	4	1	40	(.33,.33,.33)	49.5	45.5	47.6	43.6	45.5	38.6	44.8	37.4
3	3	4	1	40	(.2,.4,.4)	49	45.4	47.3	43.6	44.5	37.8	44	37
3	3	4	3	14	(.33,.33,.33)	48.2	44.6	46.7	43.3	41.4	34.8	40.6	34.1
3	3	4	1	60	(.33,.33,.33)	53.9	50.2	51.2	47.8	51.4	45.8	50.2	44.4
3	3	4	1	60	(.2,.4,.4)	53.5	50.1	50.7	47.4	50.6	44.9	49.3	43.8
3	3	4	3	20	(.33,.33,.33)	52.8	49.3	51	47.7	48.5	42.7	47.7	41.9
3	3	4	1	80	(.33,.33,.33)	57.1	53.7	54	50.6	55.4	50.6	54	48.8
3	3	4	1	80	(.2,.4,.4)	56.8	53.7	53.4	50.5	54.7	49.9	53	48.6
3	3	4	3	27	(.33,.33,.33)	56.8	53.7	54.5	51.5	54.1	49	53.1	48.2
3	3	6	1	60	(.33,.33,.33)	43.1	39.9	41.4	38.2	53	47.6	52.4	47
3	3	6	1	60	(.2,.4,.4)	42.7	39.6	41	38	51.9	46.4	51.3	45.7
3	3	6	3	20	(.33,.33,.33)	38.5	35.8	37.4	34.8	47.6	42.1	47.1	41.7
3	3	6	1	80	(.33,.33,.33)	46.6	43.7	44.5	41.6	57.3	52.9	56.5	52
3	3	6	1	80	(.2,.4,.4)	46.3	43.4	43.9	41.2	56.3	51.8	55.4	50.9
3	3	6	3	27	(.33,.33,.33)	43.2	40.3	41.6	38.9	53.8	49.1	53.2	48.5
3	3	6	1	100	(.33,.33,.33)	49.5	46.8	47	44.2	60.6	56.8	59.6	55.7
3	3	6	1	100	(.2,.4,.4)	49	46.5	46.3	43.9	59.4	55.7	58.3	54.8
3	3	6	3	34	(.33,.33,.33)	46.7	44.2	45	42.5	58.2	54.2	57.5	53.5

Table B.17: Selection Results from Sensitivity Analysis: Percentage correct selection (PCS) and Accuracy Index (AI) by the number of groups (G), scenario, number of doses (D), cohort size, number of cohorts, Group Membership Probabilities (Group Probs), and method.

Allocation													
			Cohorts			**PCA**				**Int**			
G	Scenario	D	Size	Number	Group Probs	GAB	PB	GAB-E	PB-E	GAB	PB	GAB-E	PB-E
2	C	4	1	30	(.5,.5)	40.8	36.7	40.6	37.2	49.4	44.6	49.1	45.1
2	C	4	1	30	(.4,.6)	40.6	36.4	40.5	36.5	49.2	44.1	49.1	44.3
2	C	4	3	10	(.5,.5)	37.6	34.5	37.6	34.5	44.8	40.7	44.9	40.8
2	C	4	1	50	(.5,.5)	44.1	40.3	44.2	40.6	53	48.6	53.1	48.9
2	C	4	1	50	(.4,.6)	43.9	40.1	44.2	40.4	52.9	48.3	53.1	48.6
2	C	4	3	17	(.5,.5)	42.4	39.3	42.6	39.6	50.5	46.6	50.7	46.8
2	C	4	1	70	(.5,.5)	46.7	43.1	46.8	43.5	55.8	51.5	56	51.9
2	C	4	1	70	(.4,.6)	46.6	43	47.1	43.7	55.8	51.4	56.3	52.1
2	C	4	3	24	(.5,.5)	45.5	42.5	45.7	42.8	54	50.2	54.4	50.5

2	C	6	1	40	(.5,.5)	31.9	28.9	32.2	29.2	38.3	34.8	38.7	35.2
2	C	6	1	40	(.4,.6)	32	28.6	32.2	28.9	38.4	34.4	38.7	34.7
2	C	6	3	14	(.5,.5)	28	25.7	28.2	25.9	33	29.9	33.1	30
2	C	6	1	60	(.5,.5)	34.7	31.8	35.5	32.5	41.6	38.3	42.5	39.1
2	C	6	1	60	(.4,.6)	34.9	31.8	35.6	32.6	41.8	38.1	42.5	39
2	C	6	3	20	(.5,.5)	31.6	29.2	31.6	29.2	37.4	34.2	37.4	34.2
2	C	6	1	80	(.5,.5)	37.4	34.7	37.9	35.3	44.7	41.5	45.2	42.2
2	C	6	1	80	(.4,.6)	37.5	34.5	38.1	35.3	44.8	41.2	45.3	42
2	C	6	3	27	(.5,.5)	34.8	32.2	34.9	32.4	41.2	37.9	41.2	38.1
3	1	4	1	40	(.33,.33,.33)	41.6	37.3	41.7	37.3	44	39.4	44.1	39.4
3	1	4	1	40	(.2,.4,.4)	41.4	36.2	41.9	36.4	43.7	38.2	44.2	38.4
3	1	4	3	14	(.33,.33,.33)	39.4	34.4	39.6	34.4	41.5	36.1	41.6	36.1
3	1	4	1	60	(.33,.33,.33)	44.1	40.1	44.5	40.5	46.6	42.2	47	42.7
3	1	4	1	60	(.2,.4,.4)	44.1	39.4	44.7	40	46.4	41.5	47.1	42.1
3	1	4	3	20	(.33,.33,.33)	42.9	37.9	42.8	37.9	45.1	39.9	45.1	39.8
3	1	4	1	80	(.33,.33,.33)	46.4	42.5	47.1	43.1	48.9	44.7	49.7	45.3
3	1	4	1	80	(.2,.4,.4)	46.4	42	47.2	42.6	48.8	44.2	49.7	44.8
3	1	4	3	27	(.33,.33,.33)	45.7	41	45.6	41	48.1	43.1	48.1	43.1
3	1	6	1	60	(.33,.33,.33)	33	29.5	33.4	29.7	41.1	36.5	41.5	36.8
3	1	6	1	60	(.2,.4,.4)	33.1	28.6	33.7	28.9	41.1	35.4	41.8	35.8
3	1	6	3	20	(.33,.33,.33)	29.2	25.7	29.2	25.7	35.4	30.4	35.4	30.4
3	1	6	1	80	(.33,.33,.33)	35.1	31.7	35.6	32.1	43.6	39.2	44.2	39.6
3	1	6	1	80	(.2,.4,.4)	35.3	31	36	31.5	43.7	38.2	44.4	38.8
3	1	6	3	27	(.33,.33,.33)	32.1	28.4	32.1	28.5	39.1	34.1	39.1	34.2
3	1	6	1	100	(.33,.33,.33)	36.9	33.7	37.5	34.1	45.7	41.4	46.3	42
3	1	6	1	100	(.2,.4,.4)	37.1	33.1	37.9	33.7	45.8	40.7	46.7	41.4
3	1	6	3	34	(.33,.33,.33)	34.3	30.7	34.5	30.8	42	37.1	42.2	37.2
3	2	4	1	40	(.33,.33,.33)	41.6	37	41.7	37.2	44.7	39.8	44.7	39.9
3	2	4	1	40	(.2,.4,.4)	41.6	35.7	41.8	36	44.8	38.4	45	38.7
3	2	4	3	14	(.33,.33,.33)	37.5	33.3	37.5	33.3	39.4	35.1	39.4	35.1
3	2	4	1	60	(.33,.33,.33)	44.3	39.9	44.7	40.2	47.6	42.8	47.9	43.1
3	2	4	1	60	(.2,.4,.4)	44.5	39.2	44.9	39.5	47.9	42.1	48.2	42.4
3	2	4	3	20	(.33,.33,.33)	41.2	36.9	41.2	37	43.5	39.1	43.5	39.1
3	2	4	1	80	(.33,.33,.33)	46.5	42.2	46.8	42.6	49.8	45.2	50.1	45.6
3	2	4	1	80	(.2,.4,.4)	46.8	41.7	47.1	42.2	50.2	44.7	50.5	45.2
3	2	4	3	27	(.33,.33,.33)	44.3	40.2	44.4	40.3	46.9	42.6	47	42.6
3	2	6	1	60	(.33,.33,.33)	33	29.5	33.4	29.7	39.9	35.7	40.3	35.9

3	2	6	1	60	(.2,.4,.4)	33.3	28.5	33.6	28.9	40.2	34.5	40.5	34.9
3	2	6	3	20	(.33,.33,.33)	27.8	25.3	27.9	25.4	32.6	29.3	32.6	29.4
3	2	6	1	80	(.33,.33,.33)	35.3	31.7	35.6	32.2	42.5	38.3	42.9	38.8
3	2	6	1	80	(.2,.4,.4)	35.6	31.1	36.1	31.6	42.8	37.4	43.3	38
3	2	6	3	27	(.33,.33,.33)	30.9	28.2	31	28.2	36.5	33	36.5	33
3	2	6	1	100	(.33,.33,.33)	37.2	33.8	37.6	34.2	44.7	40.6	45.2	41.1
3	2	6	1	100	(.2,.4,.4)	37.6	33.2	38.1	33.8	45	39.9	45.6	40.5
3	2	6	3	34	(.33,.33,.33)	33.4	30.5	33.5	30.6	39.6	35.9	39.7	36
3	3	4	1	40	(.33,.33,.33)	41.1	37	41.3	37.1	46.8	41.8	46.8	41.9
3	3	4	1	40	(.2,.4,.4)	40.5	36.3	40.9	36.7	46	41	46.3	41.3
3	3	4	3	14	(.33,.33,.33)	38.7	35.2	38.6	35.2	42.9	38.7	42.8	38.6
3	3	4	1	60	(.33,.33,.33)	43.7	39.7	44	40.1	49.4	44.8	49.8	45.1
3	3	4	1	60	(.2,.4,.4)	43.2	39.4	43.6	39.8	48.9	44.3	49.3	44.7
3	3	4	3	20	(.33,.33,.33)	41.9	38.5	41.8	38.4	46.6	42.4	46.5	42.4
3	3	4	1	80	(.33,.33,.33)	45.8	41.8	46.2	42.3	51.6	47	52.1	47.5
3	3	4	1	80	(.2,.4,.4)	45.3	41.6	45.8	42.2	51.1	46.7	51.7	47.3
3	3	4	3	27	(.33,.33,.33)	44.7	41.4	44.8	41.5	49.7	45.8	49.8	45.9
3	3	6	1	60	(.33,.33,.33)	33.1	29.8	33.5	30.1	39.8	35.6	40.2	36
3	3	6	1	60	(.2,.4,.4)	32.8	29.3	33.2	29.7	39.4	35	39.9	35.5
3	3	6	3	20	(.33,.33,.33)	29	26.5	29	26.6	33.5	30.2	33.5	30.3
3	3	6	1	80	(.33,.33,.33)	35.4	32	35.8	32.5	42.4	38.3	42.9	38.8
3	3	6	1	80	(.2,.4,.4)	35	31.7	35.5	32.1	42	37.7	42.5	38.2
3	3	6	3	27	(.33,.33,.33)	31.8	29.2	31.8	29.3	37.1	33.6	37.2	33.7
3	3	6	1	100	(.33,.33,.33)	37.2	34	37.7	34.5	44.5	40.5	45.1	41.1
3	3	6	1	100	(.2,.4,.4)	36.9	33.7	37.3	34.3	44.1	40	44.7	40.7
3	3	6	3	34	(.33,.33,.33)	34.1	31.4	34.2	31.7	40	36.5	40.1	36.7

Table B.18: Allocation Results from Sensitivity Analysis:
Percentage correct allocation (PCA) and percentage of allocations to doses with DLT probability in the interval (λ_e, λ_d) (Int) by the number of groups (G), scenario, number of doses (D), cohort size, number of cohorts, Group Membership Probabilities (Group Probs), and method.

Terminated Early													
			Cohorts			**Group 1**		**Group 2**		**Group 3**		**Average**	
G	Scenario	D	Size	Number	Group Probs	GAB-E	PB-E	GAB-E	PB-E	GAB-E	PB-E	GAB-E	PB-E
2	C	4	1	30	(.5,.5)	0.285	0.685	4.09	3.7	NA	NA	2.188	2.193

2	C	4	1	30	(.4,.6)	0.19	0.605	4.3	3.915	NA	NA	2.245	2.26
2	C	4	3	10	(.5,.5)	0.305	0.375	2.295	2.065	NA	NA	1.3	1.22
2	C	4	1	50	(.5,.5)	0.28	0.86	5.125	4.445	NA	NA	2.703	2.653
2	C	4	1	50	(.4,.6)	0.265	0.76	5.51	5.11	NA	NA	2.888	2.935
2	C	4	3	17	(.5,.5)	0.285	0.54	3.375	3.31	NA	NA	1.83	1.925
2	C	4	1	70	(.5,.5)	0.34	0.865	6.17	5.55	NA	NA	3.255	3.208
2	C	4	1	70	(.4,.6)	0.38	0.765	6.77	5.89	NA	NA	3.575	3.328
2	C	4	3	24	(.5,.5)	0.385	0.59	4.315	3.92	NA	NA	2.35	2.255
2	C	6	1	40	(.5,.5)	0.14	0.386	3.226	2.831	NA	NA	1.683	1.608
2	C	6	1	40	(.4,.6)	0.121	0.369	3.45	3.086	NA	NA	1.786	1.727
2	C	6	3	14	(.5,.5)	0.11	0.181	1.962	1.733	NA	NA	1.036	0.957
2	C	6	1	60	(.5,.5)	0.152	0.462	3.881	3.314	NA	NA	2.017	1.888
2	C	6	1	60	(.4,.6)	0.148	0.34	3.85	3.681	NA	NA	1.999	2.011
2	C	6	3	20	(.5,.5)	0.138	0.31	2.621	2.376	NA	NA	1.38	1.343
2	C	6	1	80	(.5,.5)	0.214	0.536	4.074	3.995	NA	NA	2.144	2.265
2	C	6	1	80	(.4,.6)	0.14	0.421	4.443	4.238	NA	NA	2.292	2.33
2	C	6	3	27	(.5,.5)	0.155	0.29	2.95	2.743	NA	NA	1.552	1.517
3	1	4	1	40	(.33,.33,.33)	0.075	0.403	0.75	1.498	5.128	4.26	1.984	2.053
3	1	4	1	40	(.2,.4,.4)	0.058	0.298	0.733	1.653	5.703	4.905	2.164	2.285
3	1	4	3	14	(.33,.33,.33)	0.095	0.213	0.685	0.888	3.343	2.553	1.374	1.218
3	1	4	1	60	(.33,.33,.33)	0.128	0.515	0.883	1.793	6.405	5.393	2.472	2.567
3	1	4	1	60	(.2,.4,.4)	0.06	0.37	0.943	1.978	7.053	6	2.685	2.783
3	1	4	3	20	(.33,.33,.33)	0.118	0.29	0.645	1.118	4.08	3.56	1.614	1.656
3	1	4	1	80	(.33,.33,.33)	0.123	0.475	1.098	2.108	7.61	6.253	2.943	2.945
3	1	4	1	80	(.2,.4,.4)	0.07	0.435	1.168	2.213	8.11	6.7	3.116	3.116
3	1	4	3	27	(.33,.33,.33)	0.128	0.385	0.88	1.278	5.103	4.183	2.037	1.948
3	1	6	1	60	(.33,.33,.33)	0.035	0.15	0.476	1.089	4.819	4.116	1.776	1.785
3	1	6	1	60	(.2,.4,.4)	0.021	0.122	0.52	1.197	5.127	4.541	1.889	1.954
3	1	6	3	20	(.33,.33,.33)	0.035	0.102	0.388	0.684	3.063	2.639	1.162	1.142
3	1	6	1	80	(.33,.33,.33)	0.037	0.183	0.579	1.149	5.639	4.75	2.085	2.027
3	1	6	1	80	(.2,.4,.4)	0.015	0.153	0.584	1.361	5.854	5.316	2.151	2.276
3	1	6	3	27	(.33,.33,.33)	0.029	0.128	0.478	0.748	3.821	3.472	1.443	1.449
3	1	6	1	100	(.33,.33,.33)	0.038	0.186	0.675	1.413	6.229	5.529	2.314	2.376
3	1	6	1	100	(.2,.4,.4)	0.019	0.158	0.739	1.429	6.787	6.021	2.515	2.536
3	1	6	3	34	(.33,.33,.33)	0.046	0.109	0.556	0.901	4.354	3.912	1.652	1.64
3	2	4	1	40	(.33,.33,.33)	0.053	0.262	2.917	2.752	2.87	2.665	1.947	1.893
3	2	4	1	40	(.2,.4,.4)	0.025	0.223	3.218	3.182	3.152	3.123	2.132	2.176

3	2	4	3	14	(.33,.33,.33)	0.055	0.135	1.822	1.693	1.723	1.732	1.2	1.187
3	2	4	1	60	(.33,.33,.33)	0.06	0.345	3.795	3.403	3.758	3.718	2.538	2.489
3	2	4	1	60	(.2,.4,.4)	0.033	0.243	3.793	3.763	4.055	3.813	2.627	2.607
3	2	4	3	20	(.33,.33,.33)	0.053	0.15	2.255	2.26	2.202	2.17	1.503	1.527
3	2	4	1	80	(.33,.33,.33)	0.063	0.333	4.303	4.243	4.215	4.012	2.861	2.863
3	2	4	1	80	(.2,.4,.4)	0.022	0.283	4.73	4.533	4.498	4.477	3.083	3.098
3	2	4	3	27	(.33,.33,.33)	0.087	0.213	2.863	2.855	2.942	2.765	1.964	1.944
3	2	6	1	60	(.33,.33,.33)	0.025	0.157	2.577	2.435	2.611	2.564	1.738	1.719
3	2	6	1	60	(.2,.4,.4)	0.009	0.121	2.688	2.684	2.958	2.865	1.885	1.89
3	2	6	3	20	(.33,.33,.33)	0.035	0.091	1.559	1.547	1.742	1.655	1.112	1.098
3	2	6	1	80	(.33,.33,.33)	0.036	0.152	2.938	2.758	3.055	2.982	2.01	1.964
3	2	6	1	80	(.2,.4,.4)	0.013	0.139	3.192	3.138	3.338	3.29	2.181	2.189
3	2	6	3	27	(.33,.33,.33)	0.04	0.096	1.926	1.948	2.146	2.074	1.371	1.373
3	2	6	1	100	(.33,.33,.33)	0.03	0.193	3.229	3.181	3.509	3.42	2.256	2.265
3	2	6	1	100	(.2,.4,.4)	0.015	0.152	3.408	3.395	3.79	3.692	2.405	2.413
3	2	6	3	34	(.33,.33,.33)	0.036	0.119	2.284	2.224	2.434	2.405	1.585	1.583
3	3	4	1	40	(.33,.33,.33)	0.325	1.105	0.397	1.048	5.69	4.747	2.137	2.3
3	3	4	1	40	(.2,.4,.4)	0.182	0.845	0.403	1.103	6.058	5.097	2.214	2.348
3	3	4	3	14	(.33,.33,.33)	0.27	0.668	0.332	0.542	3.47	2.81	1.357	1.34
3	3	4	1	60	(.33,.33,.33)	0.467	1.378	0.488	1.203	7.007	5.743	2.654	2.775
3	3	4	1	60	(.2,.4,.4)	0.237	1.018	0.495	1.315	7.405	6.48	2.712	2.938
3	3	4	3	20	(.33,.33,.33)	0.272	0.802	0.377	0.735	4.395	3.578	1.681	1.705
3	3	4	1	80	(.33,.33,.33)	0.495	1.408	0.545	1.418	8.203	6.725	3.081	3.184
3	3	4	1	80	(.2,.4,.4)	0.32	1.237	0.607	1.53	8.48	7.442	3.136	3.403
3	3	4	3	27	(.33,.33,.33)	0.365	0.97	0.503	0.915	5.547	4.71	2.138	2.198
3	3	6	1	60	(.33,.33,.33)	0.198	0.596	0.297	0.687	5.097	4.226	1.864	1.836
3	3	6	1	60	(.2,.4,.4)	0.112	0.507	0.302	0.736	5.33	4.663	1.914	1.968
3	3	6	3	20	(.33,.33,.33)	0.139	0.407	0.235	0.411	3.153	2.747	1.176	1.188
3	3	6	1	80	(.33,.33,.33)	0.201	0.735	0.328	0.824	5.782	4.842	2.104	2.134
3	3	6	1	80	(.2,.4,.4)	0.113	0.576	0.354	0.885	6.025	5.471	2.164	2.31
3	3	6	3	27	(.33,.33,.33)	0.159	0.431	0.286	0.513	3.887	3.425	1.444	1.457
3	3	6	1	100	(.33,.33,.33)	0.237	0.809	0.388	0.866	6.422	5.519	2.349	2.398
3	3	6	1	100	(.2,.4,.4)	0.146	0.625	0.405	0.921	6.829	6.051	2.46	2.532
3	3	6	3	34	(.33,.33,.33)	0.175	0.512	0.325	0.575	4.601	4.002	1.7	1.696

Table B.19: Early Termination Results from Sensitivity Analysis:
Percentage of times each Group is removed from the trial early due to trial termination rules by the number of groups (G), scenario, number of doses (D), cohort size, number of cohorts, Group Membership Probabilities (Group Probs), and method.

Selection									
				PCS			**AI**		
G	Scenario	D	Patients	Jeff,0.975	Jeff,0.95	Unif,0.975	Jeff,0.975	Jeff,0.95	Unif,0.975
2	C	6	60	44.3	41.4	43.8	57.4	54.8	57.2
3	1	6	80	45.5	42.4	45.0	58.8	56.3	58.4
3	2	6	80	44.5	41.5	44.2	56.4	53.8	56.2
3	3	6	80	44.5	41.6	44.1	56.5	54.6	56.4

Table B.20: Selection Results from Prior Sensitivity Analysis:
Percentage correct selection (PCS) and Accuracy Index (AI) by the number of groups (G), scenario, number of doses (D), number of patients, and choice of Prior/choice of cutoff.

Allocation									
				PCA			**Int**		
G	Scenario	D	Patients	Jeff,0.975	Jeff,0.95	Unif,0.975	Jeff,0.975	Jeff,0.95	Unif,0.975
2	C	6	60	35.5	35.2	35.4	42.5	42.0	42.4
3	1	6	80	35.6	35.6	35.7	44.2	44.1	44.2
3	2	6	80	35.6	35.4	35.7	42.9	42.6	43.0
3	3	6	80	35.8	35.7	35.8	42.9	42.7	42.9

Table B.21: Allocation Results from Prior Sensitivity Analysis:
Percentage correct allocation (PCA) and percentage of allocations to doses with DLT probability in the interval (λ_e, λ_d) (Int) by the number of groups (G), scenario, number of doses (D), number of patients, and choice of Prior/choice of cutoff.

Appendix C

Appendix for Chapter 4

C.1 | Finding all Configurations

To find all possible configurations, we consider the toxicity ordering amongst groups and doses, as well as the efficacy ordering amongst doses. We start with the toxicity and efficacy profile for Group 1 at dose level 1 and consider what that implies for the possible toxicity and efficacy profiles for Group 1 at dose level 2. Next, given the toxicity and efficacy profile for Group 1 at dose level 2, we consider the possible toxicity and efficacy profiles for Group 2 at dose level 2. Finally, given toxicity and efficacy profiles for Group 1 at dose level 1 and Group 2 at dose level 2, we consider the possible toxicity and efficacy profiles for Group 2 at dose level 1.

1. Given the toxicity and efficacy profile for Group 1 at dose level 1, then determine possible profiles for Group 1 at dose level 2.

   (a) $(\pi_{T11}, \pi_{E11}) = (\pi_{T1}, \pi_{E0}) \implies (\pi_{T12}, \pi_{E12}) \in \{(\pi_{T1}, \pi_{E0}), (\pi_{T1}, \pi_{E1}), (\pi_{T0}, \pi_{E0}), (\pi_{T0}, \pi_{E1})\}$

   (b) $(\pi_{T11}, \pi_{E11}) = (\pi_{T1}, \pi_{E1}) \implies (\pi_{T12}, \pi_{E12}) \in \{(\pi_{T1}, \pi_{E1}), (\pi_{T0}, \pi_{E1})\}$

   (c) $(\pi_{T11}, \pi_{E11}) = (\pi_{T0}, \pi_{E0}) \implies (\pi_{T12}, \pi_{E12}) \in \{(\pi_{T0}, \pi_{E0}), (\pi_{T0}, \pi_{E1})\}$

   (d) $(\pi_{T11}, \pi_{E11}) = (\pi_{T0}, \pi_{E1}) \implies (\pi_{T12}, \pi_{E12}) = (\pi_{T0}, \pi_{E1})$

2. Given the toxicity and efficacy profile for Group 1 at dose level 2, then determine the possible profiles for Group 2 at dose level 2.

   (a) $(\pi_{T12}, \pi_{E12}) = (\pi_{T1}, \pi_{E0}) \implies$
   $(\pi_{T22}, \pi_{E22}) \in \{(\pi_{T1}, \pi_{E0}), (\pi_{T1}, \pi_{E1}), (\pi_{T0}, \pi_{E0}), (\pi_{T0}, \pi_{E1})\}$

   (b) $(\pi_{T12}, \pi_{E12}) = (\pi_{T1}, \pi_{E1}) \implies$
   $(\pi_{T22}, \pi_{E22}) \in \{(\pi_{T1}, \pi_{E0}), (\pi_{T1}, \pi_{E1}), (\pi_{T0}, \pi_{E0}), (\pi_{T0}, \pi_{E1})\}$

   (c) $(\pi_{T12}, \pi_{E12}) = (\pi_{T0}, \pi_{E0}) \implies (\pi_{T22}, \pi_{E22}) \in \{(\pi_{T0}, \pi_{E0}), (\pi_{T0}, \pi_{E1})\}$

   (d) $(\pi_{T12}, \pi_{E12}) = (\pi_{T0}, \pi_{E1}) \implies (\pi_{T22}, \pi_{E22}) \in \{(\pi_{T0}, \pi_{E0}), (\pi_{T0}, \pi_{E1})\}$

3. Given the toxicity and efficacy profiles for Group 1 at dose level 1 and Group 2 at dose level 2, then determine the possible profiles for Group 2 at dose level 1.

   (a) $(\pi_{T22}, \pi_{E22}) = (\pi_{T1}, \pi_{E0}) \implies (\pi_{T21}, \pi_{E21}) = (\pi_{T1}, \pi_{E0})$

   (b) $(\pi_{T22}, \pi_{E22}) = (\pi_{T1}, \pi_{E1}) \implies (\pi_{T21}, \pi_{E21}) \in \{(\pi_{T1}, \pi_{E0}), (\pi_{T1}, \pi_{E1})\}$

   (c) $(\pi_{T22}, \pi_{E22}) = (\pi_{T0}, \pi_{E0})$ & $(\pi_{T11}, \pi_{E11}) \in \{(\pi_{T1}, \pi_{E0}), (\pi_{T1}, \pi_{E1})\} \implies$
   $(\pi_{T21}, \pi_{E21}) \in \{(\pi_{T1}, \pi_{E0}), (\pi_{T0}, \pi_{E0})\}$

   (d) $(\pi_{T22}, \pi_{E22}) = (\pi_{T0}, \pi_{E0})$ & $(\pi_{T11}, \pi_{E11}) \in \{(\pi_{T0}, \pi_{E0}), (\pi_{T0}, \pi_{E1})\} \implies$
   $(\pi_{T21}, \pi_{E21}) = (\pi_{T0}, \pi_{E0})$

   (e) $(\pi_{T22}, \pi_{E22}) = (\pi_{T0}, \pi_{E1})$ & $(\pi_{T11}, \pi_{E11}) \in \{(\pi_{T1}, \pi_{E0}), (\pi_{T1}, \pi_{E1})\} \implies$
   $(\pi_{T21}, \pi_{E21}) \in \{(\pi_{T1}, \pi_{E0}), (\pi_{T1}, \pi_{E1}), (\pi_{T0}, \pi_{E0}), (\pi_{T0}, \pi_{E1})\}$

   (f) $(\pi_{T22}, \pi_{E22}) = (\pi_{T0}, \pi_{E1})$ & $(\pi_{T11}, \pi_{E11}) \in \{(\pi_{T0}, \pi_{E0}), (\pi_{T0}, \pi_{E1})\} \implies$
   $(\pi_{T21}, \pi_{E21}) \in \{(\pi_{T0}, \pi_{E0}), (\pi_{T0}, \pi_{E1})\}$

C.2 | Configurations

To see all null and alternative configurations for the dose/group scenario presented in Chapter 4, see tables C.1 and C.2.

C.3 | Considering two groups with non-overlapping doses

This section discusses how to adapt the designs presented in Chapter 4 when there are two groups with different doses under consideration. In particular, a trial with three doses, d_1, d_2, d_3, and two groups $g = 1, 2$ is considered. In this trial, Group 1 is tried under doses d_2 and d_3, while Group 2 is tried under doses d_1 and d_2. This corresponds to a phase II trial resulting from a phase I groups trial where the MTD for Group 1 was d_3 and for Group 2 was d_2 with the phase II trial considering the group-specific MTDs and one dose lower. Table C.3 outlines the trial setup.

To design a traditional or two-stage design for this scenario, we need to find all possible configurations under this partial ordering. Once all possible configurations are enumerated, one repeats the steps outlined in sections 4.4 and 4.5, respectively.

Similar to Appendix C.1, we find all possible configurations using the known orderings. Steps to find all possible configurations will be outlined, after which, all configurations could be enumerated similar to Appendix C.2, a step we leave to a curious reader.

1. Given the toxicity and efficacy profile for Group 1 at dose level 2, then determine possible profiles for Group 1 at dose level 3.

(a) $(\pi_{T12}, \pi_{E12}) = (\pi_{T1}, \pi_{E0}) \implies (\pi_{T13}, \pi_{E13}) \in \{(\pi_{T1}, \pi_{E0}), (\pi_{T1}, \pi_{E1}), (\pi_{T0}, \pi_{E0}), (\pi_{T0}, \pi_{E1})\}$

(b) $(\pi_{T12}, \pi_{E12}) = (\pi_{T1}, \pi_{E1}) \implies (\pi_{T13}, \pi_{E13}) \in \{(\pi_{T1}, \pi_{E1}), (\pi_{T0}, \pi_{E1})\}$

(c) $(\pi_{T12}, \pi_{E12}) = (\pi_{T0}, \pi_{E0}) \implies (\pi_{T13}, \pi_{E13}) \in \{(\pi_{T0}, \pi_{E0}), (\pi_{T0}, \pi_{E1})\}$

(d) $(\pi_{T12}, \pi_{E12}) = (\pi_{T0}, \pi_{E1}) \implies (\pi_{T13}, \pi_{E13}) = (\pi_{T0}, \pi_{E1})$

2. Given the toxicity and efficacy profile for Group 1 at dose level 2, then determine the possible profiles for Group 2 at dose level 2.

(a) $(\pi_{T12}, \pi_{E12}) = (\pi_{T1}, \pi_{E0}) \implies (\pi_{T22}, \pi_{E22}) \in \{(\pi_{T1}, \pi_{E0}), (\pi_{T1}, \pi_{E1}), (\pi_{T0}, \pi_{E0}), (\pi_{T0}, \pi_{E1})\}$

(b) $(\pi_{T12}, \pi_{E12}) = (\pi_{T1}, \pi_{E1}) \implies (\pi_{T22}, \pi_{E22}) \in \{(\pi_{T1}, \pi_{E0}), (\pi_{T1}, \pi_{E1}), (\pi_{T0}, \pi_{E0}), (\pi_{T0}, \pi_{E1})\}$

(c) $(\pi_{T12}, \pi_{E12}) = (\pi_{T0}, \pi_{E0}) \implies (\pi_{T22}, \pi_{E22}) \in \{(\pi_{T0}, \pi_{E0}), (\pi_{T0}, \pi_{E1})\}$

(d) $(\pi_{T12}, \pi_{E12}) = (\pi_{T0}, \pi_{E1}) \implies (\pi_{T22}, \pi_{E22}) \in \{(\pi_{T0}, \pi_{E0}), (\pi_{T0}, \pi_{E1})\}$

3. Given the toxicity and efficacy profile for Group 2 at dose level 2, then determine the possible profiles for Group 2 at dose level 1.

(a) $(\pi_{T22}, \pi_{E22}) = (\pi_{T1}, \pi_{E0}) \implies (\pi_{T21}, \pi_{E21}) = (\pi_{T1}, \pi_{E0})$

(b) $(\pi_{T22}, \pi_{E22}) = (\pi_{T1}, \pi_{E1}) \implies (\pi_{T21}, \pi_{E21}) \in \{(\pi_{T1}, \pi_{E0}), (\pi_{T1}, \pi_{E1})\}$

(c) $(\pi_{T22}, \pi_{E22}) = (\pi_{T0}, \pi_{E0}) \implies (\pi_{T21}, \pi_{E21}) \in \{(\pi_{T0}, \pi_{E0}), (\pi_{T1}, \pi_{E0})\}$

(d) $(\pi_{T22}, \pi_{E22}) = (\pi_{T0}, \pi_{E1}) \implies (\pi_{T21}, \pi_{E21}) \in \{(\pi_{T0}, \pi_{E0}), (\pi_{T0}, \pi_{E1}), (\pi_{T1}, \pi_{E0}), (\pi_{T0}, \pi_{E0})\}$

Null Hypothesis Configurations								
Configuration		g	d_1	d_2		g	d_1	d_2
1	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,1}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,0}$
		2	$\theta_{T,1}$	$\theta_{T,1}$		2	$\theta_{E,0}$	$\theta_{E,0}$
2	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,1}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,0}$
		2	$\theta_{T,1}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,0}$
3	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,1}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,0}$
		2	$\theta_{T,0}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,0}$
4	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,1}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,0}$
		2	$\theta_{T,1}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,1}$
5	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,1}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,0}$
		2	$\theta_{T,0}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,1}$
6	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,1}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,0}$
		2	$\theta_{T,0}$	$\theta_{T,0}$		2	$\theta_{E,1}$	$\theta_{E,1}$
7	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,1}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,0}$
		2	$\theta_{T,0}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,0}$
8	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,0}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,0}$
		2	$\theta_{T,0}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,0}$
9	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,0}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,0}$
		2	$\theta_{T,1}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,1}$
10	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,0}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,0}$
		2	$\theta_{T,0}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,1}$
11	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,0}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,0}$
		2	$\theta_{T,0}$	$\theta_{T,0}$		2	$\theta_{E,1}$	$\theta_{E,1}$
12	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,0}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,1}$
		2	$\theta_{T,1}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,0}$
13	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,0}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,1}$
		2	$\theta_{T,0}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,0}$
14	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,0}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,1}$
		2	$\theta_{T,1}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,1}$
15	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,0}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,1}$
		2	$\theta_{T,0}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,1}$
16	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,0}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,1}$
		2	$\theta_{T,0}$	$\theta_{T,0}$		2	$\theta_{E,1}$	$\theta_{E,1}$
17	$\pi_{T,g,k}$	1	$\theta_{T,0}$	$\theta_{T,0}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,0}$
		2	$\theta_{T,0}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,0}$
18	$\pi_{T,g,k}$	1	$\theta_{T,0}$	$\theta_{T,0}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,0}$
		2	$\theta_{T,0}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,1}$
19	$\pi_{T,g,k}$	1	$\theta_{T,0}$	$\theta_{T,0}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,0}$
		2	$\theta_{T,0}$	$\theta_{T,0}$		2	$\theta_{E,1}$	$\theta_{E,1}$
20	$\pi_{T,g,k}$	1	$\theta_{T,0}$	$\theta_{T,0}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,1}$
		2	$\theta_{T,0}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,0}$
21	$\pi_{T,g,k}$	1	$\theta_{T,0}$	$\theta_{T,0}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,1}$
		2	$\theta_{T,0}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,1}$
22	$\pi_{T,g,k}$	1	$\theta_{T,0}$	$\theta_{T,0}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,1}$
		2	$\theta_{T,0}$	$\theta_{T,0}$		2	$\theta_{E,1}$	$\theta_{E,1}$
23	$\pi_{T,g,k}$	1	$\theta_{T,0}$	$\theta_{T,0}$	$\pi_{E,g,k}$	1	$\theta_{E,1}$	$\theta_{E,1}$
		2	$\theta_{T,0}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,0}$
24	$\pi_{T,g,k}$	1	$\theta_{T,0}$	$\theta_{T,0}$	$\pi_{E,g,k}$	1	$\theta_{E,1}$	$\theta_{E,1}$
		2	$\theta_{T,0}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,1}$
25	$\pi_{T,g,k}$	1	$\theta_{T,0}$	$\theta_{T,0}$	$\pi_{E,g,k}$	1	$\theta_{E,1}$	$\theta_{E,1}$
		2	$\theta_{T,0}$	$\theta_{T,0}$		2	$\theta_{E,1}$	$\theta_{E,1}$

Table C.1: Null Configurations

Alternative Hypothesis Configurations								
Configuration		g	d_1	d_2		g	d_1	d_2
1	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,1}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,0}$
		2	$\theta_{T,1}$	$\theta_{T,1}$		2	$\theta_{E,0}$	$\theta_{E,1}$
2	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,1}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,0}$
		2	$\theta_{T,1}$	$\theta_{T,1}$		2	$\theta_{E,1}$	$\theta_{E,1}$
3	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,1}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,0}$
		2	$\theta_{T,1}$	$\theta_{T,0}$		2	$\theta_{E,1}$	$\theta_{E,1}$
4	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,1}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,1}$
		2	$\theta_{T,1}$	$\theta_{T,1}$		2	$\theta_{E,0}$	$\theta_{E,0}$
5	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,1}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,1}$
		2	$\theta_{T,1}$	$\theta_{T,1}$		2	$\theta_{E,0}$	$\theta_{E,1}$
6	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,1}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,1}$
		2	$\theta_{T,1}$	$\theta_{T,1}$		2	$\theta_{E,1}$	$\theta_{E,1}$
7	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,1}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,1}$
		2	$\theta_{T,1}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,0}$
8	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,1}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,1}$
		2	$\theta_{T,0}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,0}$
9	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,1}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,1}$
		2	$\theta_{T,1}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,1}$
10	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,1}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,1}$
		2	$\theta_{T,1}$	$\theta_{T,0}$		2	$\theta_{E,1}$	$\theta_{E,1}$
11	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,1}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,1}$
		2	$\theta_{T,0}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,1}$
12	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,1}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,1}$
		2	$\theta_{T,0}$	$\theta_{T,0}$		2	$\theta_{E,1}$	$\theta_{E,1}$
13	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,0}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,0}$
		2	$\theta_{T,1}$	$\theta_{T,0}$		2	$\theta_{E,1}$	$\theta_{E,1}$
14	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,0}$	$\pi_{E,g,k}$	1	$\theta_{E,0}$	$\theta_{E,1}$
		2	$\theta_{T,1}$	$\theta_{T,0}$		2	$\theta_{E,1}$	$\theta_{E,1}$
15	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,1}$	$\pi_{E,g,k}$	1	$\theta_{E,1}$	$\theta_{E,1}$
		2	$\theta_{T,1}$	$\theta_{T,1}$		2	$\theta_{E,0}$	$\theta_{E,0}$
16	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,1}$	$\pi_{E,g,k}$	1	$\theta_{E,1}$	$\theta_{E,1}$
		2	$\theta_{T,1}$	$\theta_{T,1}$		2	$\theta_{E,0}$	$\theta_{E,1}$
17	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,1}$	$\pi_{E,g,k}$	1	$\theta_{E,1}$	$\theta_{E,1}$
		2	$\theta_{T,1}$	$\theta_{T,1}$		2	$\theta_{E,1}$	$\theta_{E,1}$
18	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,1}$	$\pi_{E,g,k}$	1	$\theta_{E,1}$	$\theta_{E,1}$
		2	$\theta_{T,1}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,0}$
19	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,1}$	$\pi_{E,g,k}$	1	$\theta_{E,1}$	$\theta_{E,1}$
		2	$\theta_{T,0}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,0}$
20	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,1}$	$\pi_{E,g,k}$	1	$\theta_{E,1}$	$\theta_{E,1}$
		2	$\theta_{T,1}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,1}$
21	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,1}$	$\pi_{E,g,k}$	1	$\theta_{E,1}$	$\theta_{E,1}$
		2	$\theta_{T,1}$	$\theta_{T,0}$		2	$\theta_{E,1}$	$\theta_{E,1}$
22	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,1}$	$\pi_{E,g,k}$	1	$\theta_{E,1}$	$\theta_{E,1}$
		2	$\theta_{T,0}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,1}$
23	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,1}$	$\pi_{E,g,k}$	1	$\theta_{E,1}$	$\theta_{E,1}$
		2	$\theta_{T,0}$	$\theta_{T,0}$		2	$\theta_{E,1}$	$\theta_{E,1}$
24	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,0}$	$\pi_{E,g,k}$	1	$\theta_{E,1}$	$\theta_{E,1}$
		2	$\theta_{T,1}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,0}$
25	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,0}$	$\pi_{E,g,k}$	1	$\theta_{E,1}$	$\theta_{E,1}$
		2	$\theta_{T,0}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,0}$
26	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,0}$	$\pi_{E,g,k}$	1	$\theta_{E,1}$	$\theta_{E,1}$
		2	$\theta_{T,1}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,1}$
27	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,0}$	$\pi_{E,g,k}$	1	$\theta_{E,1}$	$\theta_{E,1}$
		2	$\theta_{T,1}$	$\theta_{T,0}$		2	$\theta_{E,1}$	$\theta_{E,1}$
28	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,0}$	$\pi_{E,g,k}$	1	$\theta_{E,1}$	$\theta_{E,1}$
		2	$\theta_{T,0}$	$\theta_{T,0}$		2	$\theta_{E,0}$	$\theta_{E,1}$
29	$\pi_{T,g,k}$	1	$\theta_{T,1}$	$\theta_{T,0}$	$\pi_{E,g,k}$	1	$\theta_{E,1}$	$\theta_{E,1}$
		2	$\theta_{T,0}$	$\theta_{T,0}$		2	$\theta_{E,1}$	$\theta_{E,1}$

Table C.2: Alternative Configurations

Group	Doses Under Consideration.		
Group 1		d_2	d_3
Group 2	d_1	d_2	

Table C.3: Trial with non-overlapping doses