Phase I cancer clinical trial designs for cell therapies

Evan Mikal Bagley

Charlottesville, Virginia

M.S. University of Virginia, 2018

B.A. University of Virginia, 2017

A Dissertation presented 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 2023

Abstract

Cell therapies comprise one of the most important advances in oncology. One of the biggest challenges in the early development of cell therapies is to recommend a dose that is both feasible and safe to carry forward to middle development. The treatment involves extracting cells from a patient, expanding the cells, and infusing the cells back into the patient. Each dose level being studied is defined by the number of cells infused into the trial participant. A common problem arises when the assigned dose exceeds the ability of the patient's cell manufacturing process to generate that dose level, leading to a dose-feasibility issue for that patient. The primary design challenge is to efficiently use accumulated data from participants treated away from their assigned doses to efficiently allocate future trial participants and recommend a feasible maximum tolerated dose (FMTD) for all patients at the study conclusion. Currently, there are few available options for designing and implementing Phase I trials of cell therapies that can incorporate a dose-feasibility endpoint. Moreover, the application of these designs is limited to a traditional dose-finding framework where a single drug is evaluated, for a homogeneous patient population with toxicity outcomes determined in early cycles of therapy. This dissertation presents novel statistical methodology to implement three phase I designs for cell therapy trials. The first design simultaneously accounts for dose-feasibility and late-onset toxicities in a single drug dose-escalation framework. The second design simultaneously accounts for dose-feasibility and late-onset toxicities in a dual drug dose-escalation framework. The third design accounts for dose-feasibility and patient group heterogeneity. Each design was motivated by a proposed clinical trial or one that has been completed. If our designs are implemented we hope that they can improve clinical results and facilitate more efficient drug development and approval.

Acknowledgements

I would like to thank my Ph.D. advisor, Professor Nolan Wages. His support and guidance in over the past couple of years have meant so much to me. This dissertation would not have been possible without the many lessons I learned from him.

I would also like to thank my co-advisor, Professor Karen Kafadar. The classes I took with her gave me so much real world experience in problem solving. She always was such an important part of my graduate studies going all the way back to the smallest consulting class ever.

I must also give a special thanks to Professor Jeff Holt. His advising since I was an undergraduate has always put me along the right path and lead to my pursuit of Statistics. Without him, I do not think I would have found my love for Statistics.

Further acknowledgements are due to my other committee members Professor Dan Gingerich and Professor Jianhui Zhou. Thank you for reading my dissertation and providing vital feedback that will improve any implementation of my methodology in the future.

I also need to acknowledge my family and friends who have helped me so much along the way: Justin, Caroline, Claire, Mom, Dad, Nicky, Griffin, Eric, Noah, Taylor and many others. This was not possible without your support.

-For Rhonda, Janice, and Norman

Contents

1 Introduction **1**

1.1 Phase I cancer clinical trials 1

1.2 Phase I cell therapy trials 2

1.3 Outline of dissertation 6

2 Existing methodology for cell therapy trials **7**

2.1 The method of Thall, Sung and Choudhury 7

2.2 The method of Wages and Fadul 11

2.3 The method of Devlin, Iasonos and O'Quigley 13

2.4 Limitations of existing methodology 13

3 Impact of dose-feasibility on cell therapy trials when using the 3+3 **15**

3.1 Generating toxicity and feasibility scenarios 15

3.2 Simulation framework and methods 18

3.3 Simulation results 19

3.4 Discussion 23

4 Aim 1: Late-onset toxicity and dose-feasibility 24

4.1 Background on phase I trials with late-onset toxicity 24

4.2 Motivating example 25

4.3 Proposed model for toxicity 27

4.4 Proposed model for feasibility 30

4.5 Simulation Considerations 33

4.5.1 Toxicity curves 33

4.5.2 Feasibility curves 34

4.5.3 Stopping rule for toxicity 35

4.5.4 Weight functions and time to toxicity distributions 36

4.6 Trial conduct 38

4.7 Simulation Results 43

4.7.1 Simulation results across 6 toxicity/feasibility scenarios 43

4.7.2 Simulation results across 100 toxicity/feasibility scenarios 44

5 Aim 2: Drug combinations and dose-feasibility 47

5.1 Background on dose-finding for drug combinations 47

5.2 Motivating trial 2: Dual escalation of a cell therapy and an immunotherapy . 49

5.3 Proposed model for toxicity 51

5.4 Proposed model for feasibility 55
5.5 Simulation Considerations 57
5.5.1 Toxicity curves 57
5.5.2 Feasibility curves 58
5.5.3 Other considerations 59
5.6 Trial Conduct 59
5.7 Simulation Results 62

6 Aim 3: Patient heterogeneity and dose-feasibility 67

6.1 Background on dose finding with heterogeneous patient groups 67
6.2 Motivating trial 3: Dose finding for a cell therapy with 2 patient groups . . . 68
6.3 Proposed model for toxicity 71
6.4 Proposed model for feasibility 75
6.5 Trial Conduct 78
6.6 Simulation Results 81

7 Concluding remarks 86

A Additional simulation results for aim 1 89

B Additional simulation results for aim 2 94

C Large Sample Results 99

Chapter 1

Introduction

1.1 Phase I cancer clinical trials

Phase I cancer clinical trials are an important early step in the drug development process for novel cancer therapeutics. A phase I cancer trial's general purpose is to study the safety of different dose levels of the drug under evaluation. The fundamental assumption driving the design of these trials is that higher doses of the drug are associated with higher efficacy and toxicity rates. The primary objective is to determine the maximum tolerated dose (MTD), which is the highest dose level that does not exceed some pre-specified level of acceptable toxicity. The primary toxicity endpoint is dose-limiting toxicity (DLT), meaning we would not want to expose the patient to higher dose levels of the drug due to excessive, and possibly irreversible, toxic outcomes. Thus, it is in the interest of patients to escalate the dose throughout the trial in a well-designed manner while carefully considering the DLTs that could occur.

Generally, phase I cancer trials are non-randomized and adaptively allocate patients to one of a set of pre-specified discrete dose levels based on previous patients' DLT out-

comes. While trial allocation can be guided by numerous different algorithms or statistical models, a common goal is to treat as many patients as possible at and around the MTD. This objective makes accurate and efficient estimation of the MTD important from an ethical standpoint. Furthermore, accurate estimation of the MTD is crucial for the success of subsequent phases of clinical trials, which a drug must navigate for approval. Recently, researchers at UVA have shown that an efficient, well-designed phase I trial has a more significant impact on the cancer treatment's success than the sample size of the phase III trial [1]. This fact motivates the study, development, and use of novel phase I trial designs to improve the success rate of new anti-cancer treatments.

1.2 Phase I cell therapy trials

Cell therapies are relatively new in the broader context of anti-cancer drug development. From 2000 to 2019, the number of cell therapy trials have increased by approximately 3000% [2]. This surge in popularity motivates the need for novel phase I trial designs that address contemporary research questions currently being posed in early-phase cell therapy trials.

Unlike typical cancer treatments like chemotherapy, the manufacturing of cell therapy products presents unique trial design and conduct challenges. The primary challenge stems from the fact that the dose level is defined by the number of cells infused into the patient. Cell therapy treatment involves extracting cells from a patient, expanding the cells into the billions, and infusing the cancer-fighting cells back into the patient. For example, Table 1.1 shows the predefined dose levels for a phase I T cell therapy trial for treating Non-Hodgkins Lymphoma [3].

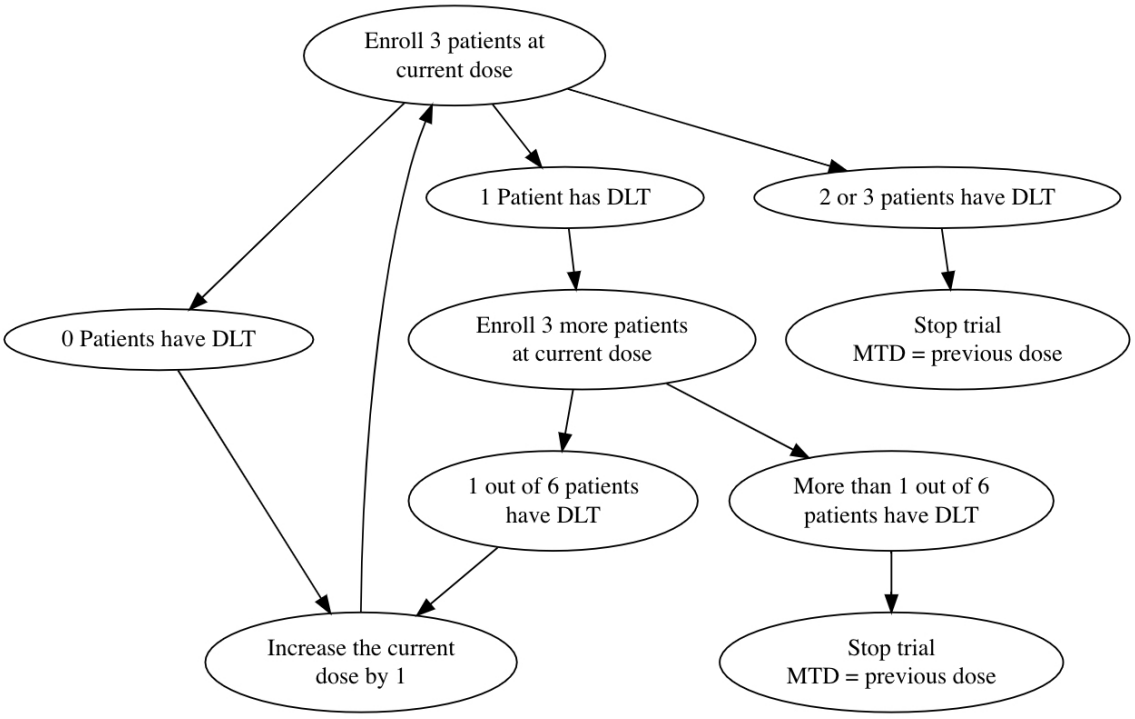
Table 1.1: Dose level definitions

Dose Level	Number of Cells $\times 10^9$
1	20
2	40
3	60
4	80

Table 1.1 provides the targeted dose levels for treating patients, but in practice, not all patients will be treatable at their assigned dose level. It is possible that the manufacturing process will not generate enough cells for a given patient to receive his or her assigned dose level. When this happens, a patient is considered not feasible to be treated at the allocated dose level and must be treated at a lower dose level based on the number of cells harvested. For example, suppose that, at some point during the trial, the recommended dose for an accrued patient is dose level 4 or 80 billion cells - but only 45 billion cells were grown for this patient after completion of the cell expansion. In this case, it would be infeasible for the patient to receive dose levels 3 and 4 since the 45 billion cells do not reach the number of cells that define the higher two dose levels. This example illustrates the nature of the problem in these types of studies: patients may have cell counts that are less than the dose that is assigned to them, preventing them from reaching their intended dose level. The main design challenge is how to efficiently use accumulated outcome data from patients that are treated away from their assigned dose to allocate future trial participants and recommend an MTD that is feasible at study conclusion.

Most phase I cancer trials, including cell therapies, use an algorithmic dose escalation framework known as the 3+3. The 3+3 algorithm has become the status quo for clinical trials and, according to some estimates, is utilized in 98.6% of phase I cancer trials [4]. The 3+3 algorithm is shown in Figure 1.

Figure 1.1: The 3+3 algorithm



The 3+3 algorithm is ubiquitous in phase I cancer trial design because it is the status-quo, and it is simple to implement, despite the fact that the 3+3 algorithm is outperformed by many model-based statistical designs, in terms of correctly recommending the MTD and allocating patients at and around the MTD. The statistical community has acknowledged that the 3+3 is inferior to model-based dose finding algorithms. We will not be focusing on the general issues with the 3+3 algorithm, they have already been explored in great detail [5–7], but we will consider design issues encountered in cell therapy trials that the 3+3 cannot handle. For instance, the 3+3 is incapable of handling dose assignment when a patient cannot receive treatment at their intended dose level due to insufficient cell expansion. If a patient's dose assignment does not fit into the algorithmic framework of the 3+3 then it is not clear how to incorporate it into the dose escalation process, and thus his or her data is excluded. For example, in a phase I CAR (chimeric antigen receptor) T-cell trial for treating acute lymphoblastic leukemia [8] that used the 3+3 the authors state, "Patients whose CAR T-cell product did not meet the dose to which they were assigned

did not inform dose escalation." This is relevant because phase I cancer trials usually do not have a large sample size, so discarding even a few data points unnecessarily can have a large effect on a trial's outcome. The issue of dose-feasibility is common to each of the cell therapy treatments that motivate this work, and will be a recurring dose finding problem that we address across a range of settings.

The FDA Center for Biologics Evaluation and Research (CBER) has issued a guidance document [9] to assist sponsors and investigators in designing early-phase clinical trials for cellular therapy products. This guidance provides current recommendations regarding clinical trials in which the primary objectives are the initial assessments of safety, tolerability, or feasibility of administration of investigational products. Under the early-phase objectives heading, the document states, "For cell therapy products, these early-phase trials often assess not only safety of specific dose regimens and route of administration, but also other issues, such as feasibility of administration... Therefore, sponsors might include design elements that could help foster further product development." Under the feasibility assessment heading it states, "In the case of cell therapy products, sponsors should consider designing early-phase trials to identify and characterize any technical or logistic issues with manufacturing and administering the product. Such issues may need to be addressed before proceeding further with product development." The FDA guidance is explicitly calling for early-phase trial designs that formally account for dose-feasibility.

Currently, only a few alternatives for the design and implementation of phase I cell therapy trials can handle the feasibility issue described above [10–12]. These designs were developed for trials where a single agent was under study for use in a homogeneous patient population where the DLT endpoint was observed in early cycles of therapy (28-days). As we will demonstrate in our motivating examples, contemporary dose-finding problems of late-onset DLT (several months post-treatment), drug combinations, and patient heterogeneity may be present in cell therapy trials, and the few cell therapy designs

that can handle dose-feasibility address none of these additional challenges. The goal of this dissertation is to develop designs for cell therapy trials motivated by real studies that currently have no statistical methodology to guide dose assignment and feasible MTD recommendation. We will describe the few cell therapy trial designs in the literature review and discuss how these designs are insufficient to handle the contemporary dose-finding challenges that have been raised by our motivating clinical trial proposals. Subsequently, we will provide novel statistical solutions to answer the research questions posed in each of the motivating studies.

1.3 Outline of dissertation

This dissertation will approach the problem of dose-feasibility in cell therapy trials from multiple angles. Chapter 2 describes the few existing statistical designs developed for cell therapy trials. These designs are valid methodology with good operating characteristics, but are useable only in a particular setting. Chapter 3 presents a simulation study of the 3+3 algorithm to motivate the need for further methodological work and implementation in early phase cell therapy trials. Chapter 4 provides the statistical framework and design for single-agent dose-escalation in cell therapy trials that have dose-feasibility issues in addition to late-onset toxicities. Chapter 5 provides statistical methodology for dual-drug dose-escalation in cell therapy trials with dose-feasibility issues and late-onset toxicities. Chapter 6 presents statistical methodology for a cell therapy trial with heterogeneous patient populations. Chapter 7 provides discussion and commentary on how our trial designs could have an impact on cancer drug development. Lastly, the appendices give additional simulation results.

Chapter 2

Existing methodology for cell therapy trials

While there is a wealth of literature devoted to the design of phase I cancer trials, only a few studies have explored the design of phase I cancer cell therapy trials that involve patients being treated away from their assigned dose. In this chapter, we present the methodologies employed in these studies, discuss their contributions to the broader base of phase I cancer trial literature, and provide insights into when and where they may be appropriately applied.

2.1 The method of Thall, Sung and Choudhury

Thall, Sung and Choudhury offered the first model-based design specifically tailored to phase I cell therapy trials [10]. The authors defined the notion of dose-feasibility in a mathematical context and propose methodology that allows outcomes of patients treated below their assigned dose level to inform trial conduct. Their design adaptively treats

patients based upon previous patients' DLT outcomes from a set of discrete dose levels $d_1 < d_2 < \cdots < d_J$ in a manner that is guided by independent toxicity and feasibility models.

The toxicity portion of the design is an extension of the continual reassessment method (CRM) [13]. The CRM guides dose assignment by estimating

$$p_j = (Y = 1|d_j) \qquad \text{for} \qquad j = 1, \ldots, J \tag{2.1}$$

where p_j is the true DLT probability at dose level j and Y is a binary DLT indicator. The CRM assumes a parametric model for p_j so that DLT data can be shared across dose levels. The authors implement a common form of the CRM known as the empiric model

$$p_j(q_j, \beta) = q_j^{\exp(\beta)} \qquad \text{for} \qquad j = 1, \ldots, J \tag{2.2}$$

where β is the parameter and q_j is a set of initial toxicity probabilities for each dose level which is commonly referred to as the skeleton. The CRM updates the estimates of p_j when a new patient enrolls in the trial and assigns them the current estimated MTD defined as

$$j^* = \underset{j \in \{1,\ldots,J\}}{\operatorname{argmin}} |\hat{p}_j - \theta^*| \tag{2.3}$$

where θ^* is the target DLT rate and $\hat{p}_j$ is the estimated DLT probability at dose level j. The estimate $\hat{p}_j$ is based upon the posterior mean of the model where the data is incorporated via the binomial likelihood

$$\mathcal{L}(\beta|n_j, y_j) = \prod_{j=1}^{J} (q_j^{\exp(\beta)})^{y_j} (1 - q_j^{\exp(\beta)})^{n_j - y_j} \tag{2.4}$$

where n_j is the number of patients treated at dose level j and y_j is the number of pa-

tients who experienced a DLT at dose level j. Thus, the posterior mean of the model is calculated as

$$\hat{p}_j = \int \frac{q_j^{\exp(\beta)} \mathcal{L}(\beta|n_j, y_j) f(\beta)}{\int \mathcal{L}(\beta|n_j, y_j) f(\beta) d\beta} d\beta. \tag{2.5}$$

where $f(\beta)$ is the prior distribution of β. The difference in how Thall, Sung and Choudhury implement the CRM is that if a patient is ineligible to receive the CRM assigned dose then they receive their highest feasible dose. This is relevant because this data would be ignored if the trial had used the 3+3 algorithm, but because the data can be included through the likelihood function it contributes to subsequent dose assignments and toxicity estimation.

To model feasibility Thall, Sung and Choudhury define L as the number of cells expanded for a particular patient. The number of cells will correspond to a given dose level and be recorded as the highest dose level that the patient could receive. They record this by equating the events

$$\{C = j\} \equiv \{d_j \leq L < d_{j+1}\} \qquad \text{for} \qquad j = 1, \ldots, J \tag{2.6}$$

where C is the patient's highest feasible dose level. Thus, the patient is feasible to receive any dose level up to and including dose level C. Note that it is possible that the patient may not have enough cells expanded to be treatable at any dose level and thus would be recorded as $\{C = 0\}$. The authors define the probability being able to receive dose level j as

$$\phi_j = P(C \geq j) \qquad \text{for} \qquad j = 1, \ldots, J \tag{2.7}$$

and the probability that dose level j is the highest feasible dose as

$$\pi_j = P(C = j) = \phi_j - \phi_{j+1} \tag{2.8}$$

with $\phi_0 = 1$ and $\phi_{(J+1)} = 0$. The probability model is given by first determining J fixed prior feasibility probabilities $r_1 < r_2 < \cdots < r_J$. With this, they assume the parametric model

$$\phi_j(r_j, \alpha) = r_j^{\exp(\alpha)} \tag{2.9}$$

where α is the unknown parameter. Determining which dose levels are feasible is done by comparing the feasibility model to a required minimum feasibility probability ϕ^*. Thus, dose level j is not feasible if

$$P(r_j^{\exp(\alpha)} < \phi^* | data) > p_{u,f} \tag{2.10}$$

where $p_{u,f}$ is some upper threshold. The only role that the feasibility model plays during the trial is that the trial will stop if the lowest dose level becomes infeasible, because patients are treated at the CRM recommended dose as long as it is feasible for them, even if it is considered not feasible for the patient population.

The final important contribution from this paper was altering the goal of the trial to estimate the feasible maximum tolerated dose (FMTD). This is defined at trial conclusion by estimating the MTD using the CRM and determining highest feasible dose (HFD) using (2.10). Formally, the FMTD is defined as

$$FMTD = \min(j^*, HFD). \tag{2.11}$$

Thus, at trial conclusion the recommended FMTD can be used to help inform dose assignments in subsequent phases of clinical trials.

2.2 The method of Wages and Fadul

Wages and Fadul gave the second model-based design for phase I cell therapy trials [11]. The authors model toxicity and feasibility independently with the goal of identifying the FMTD.

The authors extend a model based on isotonic regression [14] to model toxicity. This method assumes J independent beta-binomial models

$$y_j|p_j \sim \text{Binomial}(p_j) \text{ and } p_j \sim \text{Beta}(\alpha_j, \beta_j) \tag{2.12}$$

where n_j are the number of patients treated at dose level j, y_j is the number of DLTs at dose level j, and (α_j, β_j) are priors for the beta distribution. Then the posterior mean of p_j is calculated for all of the tried dose levels as

$$\hat{p}_j = \frac{y_j + \alpha_j}{n_j + \alpha_j + \beta_j}. \tag{2.13}$$

Once calculated, any estimates of $\hat{p}_j$ that violate the assumption of increasing DLT probability with the dose level are replaced with weighted averages of the offending estimates. This replacement procedure is carried out by the pooled adjacent violators algorithm (PAVA) [15], and it allows for information sharing across the dose levels. During the trial patients are assigned to the dose level with estimated DLT probability closest to the target DLT probability if they are feasible to receive it. If a patient is not feasible to receive the assigned dose level, they are treated at their HFD. This method also has the desirable feature of being able to incorporate data observed at unplanned dose levels. At trial conclusion the MTD is selected by choosing the dose with estimated DLT probability (adjusted using PAVA) closest to the target DLT rate.

The authors model the probability of feasibility ϕ_j at each dose level using J indepen-

dent beta-binomial models with z_j being the number of patients eligible to receive up to and including dose level j. Formally, the model is

$$z_j|\phi_j \sim \text{Binomial}(\phi_j) \qquad \text{and} \qquad \phi_j \sim \text{Beta}(\tau_j, \nu_j) \tag{2.14}$$

where τ_j and ν_j are the priors for the distribution of ϕ_j at each dose level. Then the authors calculate the posterior probability

$$\phi_j|z_j \sim \text{Beta}(\tau_j + z_j, \nu_j + m - z_j) \tag{2.15}$$

where m is the total number of patients who have been evaluated to determine their highest feasible dose level. With this posterior probability the authors establish a set of feasible doses

$$\mathcal{F} = \{d_j : P(\phi_j < \phi^*|z_j) \leq p_{u,f}\} \tag{2.16}$$

where $p_{u,f}$ is an upper probability cutoff for feasibility as in (2.10). During the trial if the lowest dose level leaves the set $\mathcal{F}$ the trial will stop early. When all patients have had their cells extracted and expanded the maximum dose in $\mathcal{F}$ is the HFD.

Taking the toxicity and feasibility models together their design attempts to treat patients at, or near, the MTD provided that it is feasible. If a patient is unable to receive the dose indicated by the toxicity model, then the patient is treated at his or her individual highest feasible dose. At the conclusion of the trial the FMTD is chosen in the same way as in (2.11).

2.3 The method of Devlin, Iasonos and O'Quigley

The third and final model-based design specifically tailored for phase I cell therapy trials was recently published by Devlin, Iasonos, and O'Quigley [12]. The authors take a different approach than either of the previous two cell therapy designs. Instead of defining the FMTD and treating patients from a predefined set of dose levels, they instead allow patients to be treated in between dose levels. They extend the CRM so that its likelihood can include and weight those assignments accordingly. They are essentially changing the dose-feasibility part of the problem and are more concerned with how to attribute toxicity at unplanned fractional dose levels. Their design is a much different way to think about dose-feasibility in the context of phase I cell therapy trials and has objectives that differ from the examples that motivate this dissertation. For this reason, we do not go into as much detail as the previous methods.

2.4 Limitations of existing methodology

It is remarkable that phase I cell therapy trials have been conducted for at least 20 years and these three papers make up the entire literature of model-based cell therapy designs. Yet, these three designs were intended for use in studies of single-agent, anti-cancer therapies with homogeneous patient populations where DLT outcomes are observable after a cycle of therapy (usually 28-day cycles). The growing popularity of novel cell therapy studies combined with the absence of statistically valid designs that can handle contemporary dose-finding problems when there are dose-feasibility concerns threatens to slow cell therapy drug development unnecessarily. Clinicians will undoubtedly fall back on variations of 3+3 algorithms for dose-escalation in the absence of no readily available designs that fit into their protocol. This dissertation aims to make designs available for

phase I cancer clinical cell therapy trials that account for DLTs that occur in later cycles of treatment, allow for the study of dose-combinations of multiple agents, and permit heterogeneous patient groups to be included in a single study.

Chapter 3

Impact of dose-feasibility on cell therapy trials when using the 3+3

We conducted a simulation study to explore the characteristics of 3+3 in cell therapy trials with dose-feasibility considerations [16]. We have included the parts of the study most relevant for this dissertation in this chapter, but the full study and results can be found in *Contemporary Clinical Trials Communications* at `https://doi.org/10.1016/j.conctc.2021.100877`.

3.1 Generating toxicity and feasibility scenarios

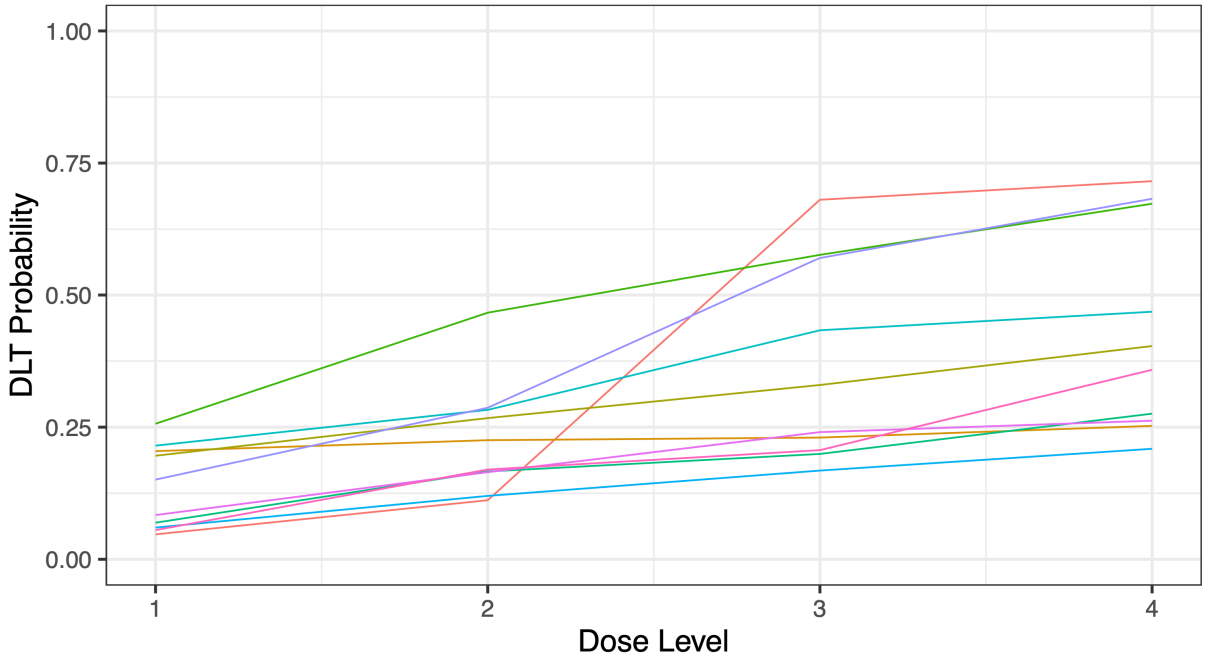
To set up the simulation, we assumed that the dose level and the probability of toxicity have a monotonic increasing relationship while the dose level and probability of feasibility have a monotonic decreasing relationship. We examined six hypothesized scenarios for DLT and feasibility probabilities over four possible dose levels. To conduct the simulations we generated toxicity curves using the algorithm proposed by Conaway and Petroni [1].

This algorithm generates S-shaped curves with the form

$$f(x) = \mu + \frac{\delta - \mu}{1 + e^{-\nu(logit(x))}} \tag{3.1}$$

where x takes on values from 0.001 to 0.999 in increments of 0.001, $\mu > 0$ and $\delta > \mu$. In this formulation μ and δ represent the minimum and maximum DLT rates respectively. Ten DLT probability curves are given in Figure 3.1.

Figure 3.1: 10 DLT probability curves



Simulating feasibility scenarios was trickier since there are no existing algorithms or methods to study dose-feasibility. Thus we propose the first class of dose-feasibility curves through the following algorithm.

Algorithm: Feasibility curve generation

1. Choose a value for the feasibility parameter b that is between 0 and 1. A larger value for b will correspond to a higher probability that a number of the dose levels will not

be feasible.

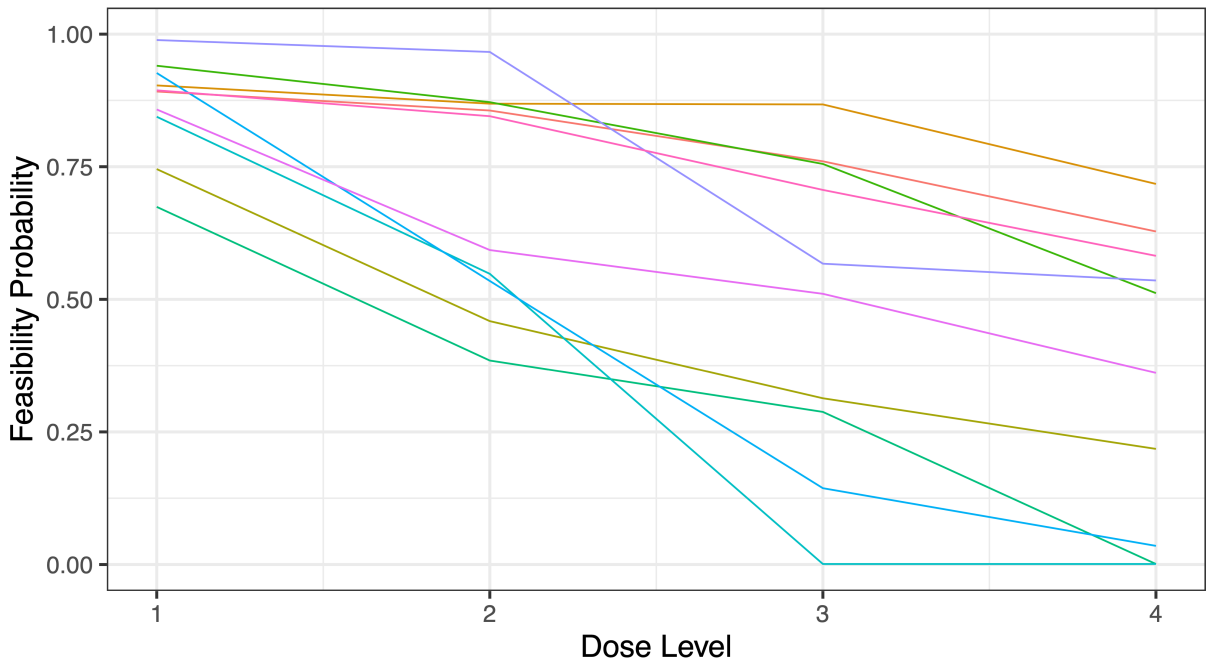
2. For each curve generate a uniform$(0, b)$ random variable and denote it τ.

3. For J dose levels we generate the probability: $P(\text{dose } j \text{ is feasible}) = \phi_j$ for $j = 1, \ldots, J$. Now let $\phi_1 = \text{uniform}(1 - \tau, 1)$ and $\phi_i = \text{ uniform}(\phi_{i-1} - \tau, \phi_{i-1})$.

4. For any $\phi_i < 0$ set $\phi_i = 0$.

To illustrate how the algorithm works, consider the following example of how one feasibility curve is randomly generated. Let the number of doses $J = 4$. According to step 1 set the feasibility parameter as $b = 0.70$. Then according to step 2 generate a uniform$(0, b)$ random variable $\tau = 0.36$. Thus, the feasibility probability at dose level 1 from step 3 is $\phi_1 \sim$ uniform$(0.64, 1) = 0.72$. We iterate to form the subsequent $\phi_i's$: $\phi_2 \sim$ uniform$(0.36, 0.72) = 0.49$, $\phi_3 \sim$ uniform$(.13, .49) = .26$, and $\phi_4 \sim$ uniform$(-0.10, 0.26) = -0.08$. According to step 4 we set $\phi_4 = 0$. Thus, the final feasibility scenario is (0.72, 0.49, 0.26, 0.00). Figure 3.2 displays 10 example feasibility curves generated by our algorithm.

Figure 3.2: 10 feasibility probability curves



3.2 Simulation framework and methods

Under each of the six scenarios, we simulated 10,000 trials using the 3+3 algorithm. Each trial accrued a maximum number of 24 participants while evaluating a maximum of 30 participants for feasibility. These settings mimic those considered in the simulation of a Phase I trial of activated T cells in combination with radiation therapy and temozolomide in newly diagnosed glioblastoma participants (NCT03344250) studied by Wages and Fadul [11]. Following the conduct of Lee et al. [8] and Lum et al. [17] in executing the 3+3, participants were enrolled sequentially in each trial and if their assigned dose was not feasible for them to receive, then they were treated at their highest feasible dose level. The participants that received an unplanned dose did not have their DLT outcome inform subsequent participants' dose assignments. If a participant was evaluated for feasibility and was not feasible for the lowest dose, they were not treated under the protocol.

We focus on the characteristics of the 3+3 design that are directly attributable to excluding data observed at unplanned dose levels. We report the following statistics:

1. Percentage of all participants treated at an unplanned dose that had a DLT excluded from use in future dose assignments or MTD recommendation.
2. Percentage of trials where there is at least one excluded DLT outcome.
3. Percentage of trials in which a dose lower than the recommended MTD had an estimated DLT rate higher than the estimated DLT rate at the recommended MTD.
4. Percentage of total DLT outcomes that were excluded in executing the design.
5. Average number of trial participants treated at an unplanned dose that had a DLT outcome excluded from use in future dose assignments and MTD recommendation.

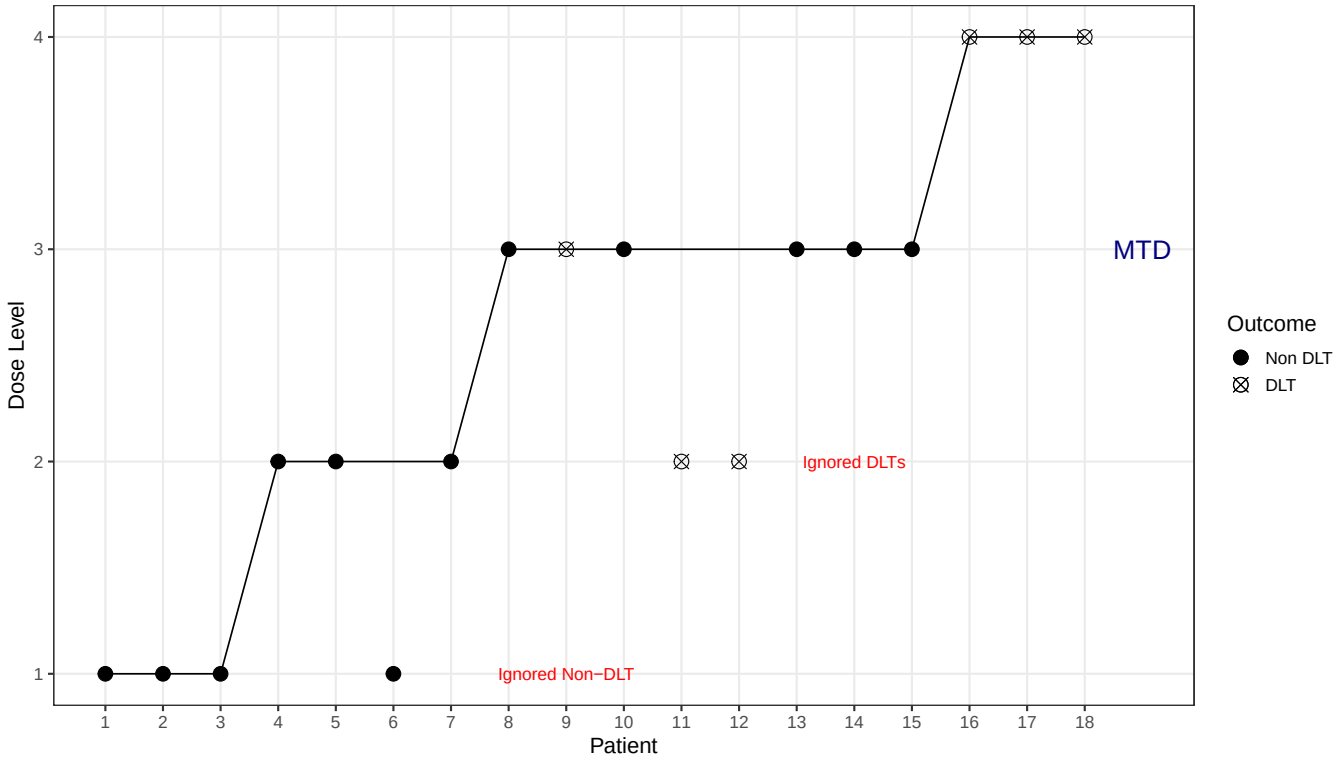
Statistic 1 provides a measure of participant safety. When a DLT occurs at an unplanned dose, it is important to ask whether a trial should continue to escalate. Statistic 2 indicates how prevalent undesirable behavior is occurring across all of the simulated trials in a large number of hypothesized scenarios. Statistic 3 illustrates that a fundamental assumption of dose-finding designs, that of an increasing relationship between dose and toxicity, can be violated if some safety outcomes are not used to inform dosing decisions. The possibility of choosing an MTD when there are lower doses with higher observed DLT rates is considerably problematic. Statistic 4 determines how many of the observed DLT outcomes are not factored into the trial design. Statistic 5 gives a per trial sense of how many participants are not having their outcome, regardless of the result, contribute to the design conduct.

3.3 Simulation results

In our simulations, we assume that, as in Lee et al. [8], participants that are unable to receive their assigned dose "did not inform dose escalation." Figure 3.3 illustrates a single simulated example trial that motivates our exploration. The trial begins treating participants at dose level 1, which all three participants are feasible to receive. No DLTs are observed in this initial cohort, and the design escalates to dose level 2. In the next cohort, the design accrues participants to dose level 2, but one participant in this cohort must receive dose level 1 as this was his or her highest feasible dose. While no participants had a DLT at any level, the design has effectively excluded the non-DLT result of participant 6 and escalates to dose level 3. In the first cohort treated at dose level 3, 1 of 3 participants have a DLT, so the design accrues 3 more participants to dose level 3. However, the next two participants are not feasible to receive dose level 3, so they must receive their highest feasible dose (dose level 2). Both of these participants have a DLT,

but the design cannot incorporate their outcomes in dosing decisions, so it enrolls three more participants at level 3, none of whom has a DLT. At this point in the trial, we have observed a DLT rate of 1 of 6 at dose level 3 and a higher rate of 2 of 5 at dose level 2. Yet since the design cannot use safety data observed at unplanned dose levels, the trial escalates to dose level 4. If the design could have incorporated the two DLTs observed at the unplanned dose level 2, then perhaps the trial would not have escalated to dose level 4. At dose level 4, all three participants have DLTs, so the trial stops and dose level 3 is chosen as the MTD according to the decision rules of the 3+3. Note that in this example trial dose level 3 is selected as the MTD despite dose level 2 having a larger observed DLT rate. This illustrates that the 3+3 may not select a logical MTD in the presence of dose-feasibility considerations.

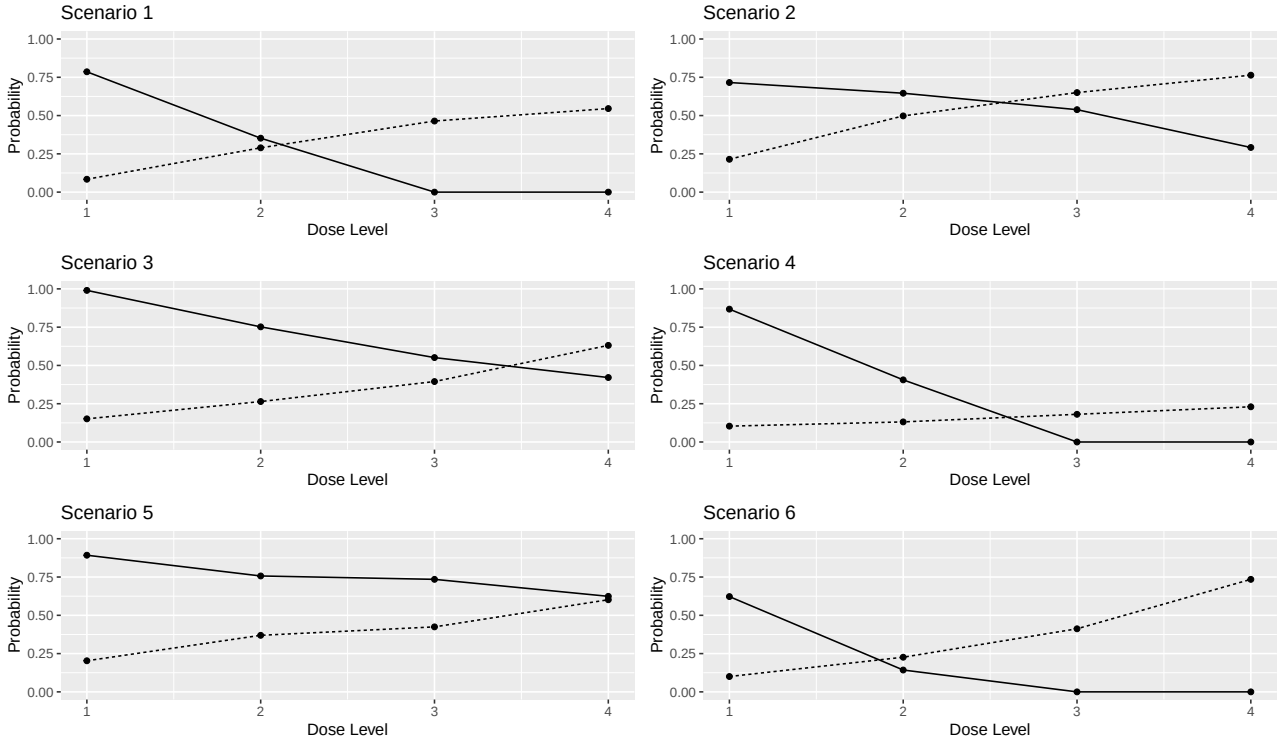
Figure 3.3: Example trial guided by the 3+3 when dose-feasibility issues arise



We explored the characteristics of the 3+3 algorithm over the six combinations of feasibility and toxicity probability curves given in Figure 3.4. These scenarios give us a

broad range of possible toxicity and feasibility situations to explore.

Figure 3.4: Six hypothetical feasibility and toxicity probability curves



The results for the six scenarios are given in Table 3.1. The percentage of DLTs at an unplanned dose translates to participants whose DLT was excluded from the dose assignment scheme. Under Scenario 2 this is especially problematic as it occurred in almost one third of all accrued participants. The percentage of trials where there is at least one DLT ignored is large under Scenario 4, which sees more than two thirds of the trials having this flaw. Scenarios 4 and 6 have a troubling percentage of trials that recommend an MTD when a lower dose has a higher estimated DLT rate than the recommended MTD. This sort of behavior is troublesome because it violates our assumptions about the monotonic relationship between toxicity and dose level. The fourth statistic is a general notion of how much of the toxicity information is being ignored in the dose allocation algorithm. Scenarios 1, 4, and 6 have a large number of excluded DLTs because the probability of feasibility and toxicity change quickly at the lower dose levels. The last statistic gives us an idea of how many participants do not have their DLT data included in the dose escala-

Table 3.1: Results for 6 scenarios under considerations

Statistics for measuring the impact of ignoring dose feasibility (standard errors in parentheses)	Scenario					
	1	2	3	4	5	6
% of patients treated at an unplanned dose that had a DLT excluded from use in future dose assignments or MTD rec.	13.20 (0.12)	28.97 (0.71)	19.49 (0.2)	11.15 (0.08)	21.42 (0.43)	10.41 (0.1)
% of trials where there is at least one excluded DLT outcome.	59.96 (0.53)	10.04 (0.3)	37.11 (0.48)	70.89 (0.47)	15.10 (0.37)	63.57 (0.55)
% of trials where a lower dose than the MTD rec had an estimated DLT rate exceed the estimated DLT rate at the MTD rec.	8.85 (0.28)	3.45 (0.15)	20.13 (0.48)	30.74 (0.53)	10.39 (0.33)	17.26 (0.38)
% of total DLT outcomes that were excluded in executing the design	47.98 (0.44)	5.05 (0.17)	19.69 (0.26)	64.23 (0.4)	6.89 (0.17)	53.11 (0.42)
Average number of trial participants whose DLT outcome did not inform future dose assignments or MTD rec.	10.11 (0.06)	0.47 (0.01)	3.19 (0.04)	13.45 (0.07)	0.89 (0.01)	10.50 (0.05)

tion decisions. Some of these numbers may seem small but consider that the maximum sample size is 24. Under Scenario 4, the value of 13.45 patients on average is equivalent to ignoring 56% of your available patients in the conduct of the trial. This data should be used in implementing a design that accounts for both safety and feasibility.

The primary take-away of our study is that when there are dose levels under consideration that are not feasible for some participants to receive, the 3+3 algorithm can have major deficiencies in safely accruing participants to the trial. This conclusion may carry over to other rule-based algorithms that only focus on the current dose to make allocation decisions, without borrowing information across dose levels.

3.4 Discussion

The inability to incorporate all toxicity information is ample evidence to avoid implementation of the 3+3 in phase I cell therapy trials. Several of our simulations produced trials where $>$50% of DLTs are ignored based solely on the inflexibility of algorithmic designs. For cell therapy trials, we recommend using a design that can incorporate patient toxicity data at unplanned dose levels. All of the available methods [10–12] use statistical models throughout the study to adaptively assign participants. The use of a model allows all available safety and feasibility data to be sequentially used in the trial conduct. No DLT outcomes are excluded when using these methods, which would result in a value of 0 for all five statistics used in this paper to quantify the impact of DLT exclusion. Conversely, since the 3+3 does not incorporate feasibility data into the design, leading to the exclusion of DLT data at unplanned dose levels, we recommend that a model-based design be used in designing Phase I cell therapy trials. In the following chapters we provide 3 different model-based designs that guide dose-finding in Phase I cancer trials while also addressing additional complexities.

Chapter 4

Aim 1: Late-onset toxicity and dose-feasibility

4.1 Background on phase I trials with late-onset toxicity

Many phase I cancer trials are designed and conducted with the expectation that possible treatment-related DLTs occur soon after treatment or within the first cycle of treatment (28-days). While this is an appropriate assumption for many anti-cancer therapeutics, there are many situations where late-onset toxicities could occur long after a patient's initial treatment. For example, consider a cancer trial [18] that administered Bortezomib for the treatment of mantle cell lymphoma. The treatment schedule was for each patient to receive up to 6 cycles of therapy, with each cycle lasting 21 days. Thus, it was possible to observe a treatment-related DLT outcome up to 126 days after the patient was initially accrued to the study. This lengthy DLT evaluation period made it impractical to wait to accrue a new patient to the study while the previous patient finished the whole treatment course. One solution could be to use methodology developed under the assumption that the DLT is observed early in treatment course. If there is potential for late-onset DLTs,

relevant toxicities are under-counted, and the recommended dose is weighted in favor of higher doses that appear safer than they actually are [19]. Consequently, recommended doses have a higher than anticipated toxicity rate. The FDA registers many novel therapies at doses different than those the Phase I trial identified [20], with many patients treated at overly toxic or sub-therapeutic levels throughout the development process. Designs can extend the DLT evaluation window to a longer period to capture late-onset DLTs in allocation decisions. However, this can lead to increased trial duration if completely observed DLT data for each participant is required before dosing decisions can be made.

There are many novel trial designs based on statistical methodology that have been proposed to utilize partially observed DLT information in an attempt to speed up trials without sacrificing accuracy [21–25]. Despite widespread availability of these methods most phase I cancer trials implement the 3+3 algorithm, which does not have the flexibility to properly account for late-onset DLT data in an efficient manner.

4.2 Motivating example

The motivating trial for this chapter was proposed by clinical investigators at the University of Virginia Comprehensive Cancer Center. The trial calls for the dose-escalation of Rituximab-based Bispecific activated T cells in combination with a fixed dose of Nivolumab, an FDA approved immunotherapy, for the treatment of high-grade B-cell lymphoma. Production of the cell therapy product involves extracting cells from the patient, activating the cells to fight the cancer, and expanding the cells to a very large number. Dose levels, given in Table 4.1, range from 20 to 80×10^9 T cells and the Nivolumab dose level remains fixed at 3mg/kg.

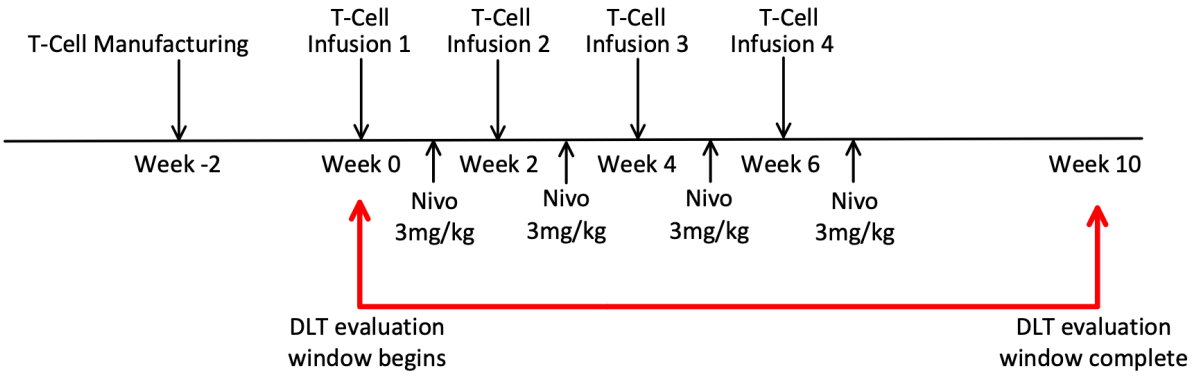
Patients undergo T-cell manufacturing and are administered the treatment over a 7-

Table 4.1: Dose level definitions for motivating trial

Dose Level	Number of T cells $\times 10^9$	Nivolumab
1	20	3 mg/kg
2	40	3 mg/kg
3	60	3 mg/kg
4	80	3 mg/kg

week period with the possibility of treatment-related DLTs occurring 10 weeks after the initial infusion. The patient is not considered fully followed for DLT until the end of the 10 weeks as they could experience a treatment-related DLT at any point in this interval. The treatment schedule is given in Figure 4.1. This trial presents two dose-finding challenges simultaneously: 1) some patients may not be feasible to receive their assigned dose level, and 2) late-onset toxicities threaten the speed at which the trial can enroll new patients. We have a 10-week DLT evaluation window and an expected accrual of two patients per month in the motivating trial. Consequently, we expect to accrue approximately five patients while waiting for a patient to complete the DLT evaluation window. In executing an existing design intended for cell therapy, the trial would need to suspend new patient enrollment until the current patient experiences a DLT or completes the DLT observation window. With a maximum sample size of 24 and an expected patient accrual rate of 5 patients per 10 weeks, the trial length could take as long as 5 years, with some patients having to await treatment for as much as 4 years. Furthermore, trial suspension slows down the drug development process and delays the approval of potentially successful cancer therapeutics to the detriment of those needing treatment. No trial designs that account for dose-feasibility and late-onset toxicity currently exist. We propose a design, that we recently published in *Journal of Biopharmaceutical Statistics* [26], that can execute this trial in an efficient manner that uses both late-onset toxicity and feasibility data.

Figure 4.1: Patient treatment schedule for motivating trial



4.3 Proposed model for toxicity

We propose a DLT probability model that adaptively assigns patients to one of a set of pre-specified dose levels $d_1 < d_2 < \cdots < d_J$ with the assumption that DLT probability increases with increasing dose level. With that assumption in mind we assume a single parameter Probit model for DLT probability

$$p_j(\beta, q_j) = \Phi\{\beta + \Phi^{-1}(q_j)\} \qquad \text{for} \qquad j = 1, \ldots, J \tag{4.1}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function, $\Phi^{-1}(\cdot)$ is the standard normal quantile function, β is the parameter, and q_j are pre-specified constants (termed the skeleton of the model) representing DLT probabilities at each dose level. We determine q_j using the method described by Lee and Cheung [27] because it has been shown to have robust characteristics across a large number of DLT probability scenarios. The toxicity model (4.1) belongs to a class of CRM dose-toxicity relations [28] $F(x, \beta) = \psi\{c(\beta)h(x)\}$ where β is a scalar, and the functions ψ, c and h are strictly monotone and known. The statistical properties of model (4.1) have yet to be evaluated in the Phase I literature. We also implement our design under an alternative model belonging to

this class of models , known as the complimentary log-log model. This model is given by

$$p_j(\beta, q_j) = 1 - (1 - q_j)^{\exp(\beta)} \qquad \text{for} \qquad j = 1, \ldots, J. \tag{4.2}$$

While it is generally believed that the choice of model with respect to the CRM will not greatly affect operating characteristics we still believe it is relevant to consider alternative models in evaluating novel designs.

In general, we assume that the doses being studied are to be administered over a pre-specified DLT evaluation window T, and that a DLT is defined as any adverse event meeting a protocol-specific DLT definition that occurs within the evaluation window. Any patient who experiences a DLT at any point during the evaluation window is considered to reach the endpoint and is taken "off study." By implementing the toxicity model in a manner that can account for partially observed DLT data, we can view DLT status as a function of time. Thus, the i^{th} patient's DLT status at time t is

$$y_{i,t} = \begin{cases} 1 & \text{if patient } i \text{ has DLT before time } t \\ 0 & \text{if patient } i \text{ has not had DLT before time } t \end{cases} \qquad \text{for} \qquad i = 1, \ldots, n. \tag{4.3}$$

The weighted binomial likelihood for models (4.1) and (4.2) respectively for the first n patients at time t is

$$\mathcal{L}_t(\beta | y_{i,t}, w_{i,t}) = \prod_{i=1}^{n} \left(w_{i,t} \Phi\{\beta + \Phi^{-1}(q_{i,j})\} \right)^{y_{i,t}} \left(1 - w_{i,t} \Phi\{\beta + \Phi^{-1}(q_{i,j})\} \right)^{1 - y_{i,t}} \tag{4.4}$$

and

$$\mathcal{L}_t(\beta|y_{i,t}, w_{i,t}) = \prod_{i=1}^{n} \left(w_{i,t}(1-(1-q_{i,j})^{\exp(\beta)})\right)^{y_{i,t}} \left(1-w_{i,t}(1-(1-q_{i,j})^{\exp(\beta)})\right)^{1-y_{i,t}} \quad (4.5)$$

where $q_{i,j}$ is the value of the skeleton corresponding to the dose patient i received, and $w_{i,t}$ represents how much of a patient's non-DLT is contributing to the model. Formally, $w_{i,t}$ is a weight function that describes how much information the i^{th} patient's DLT status is included in the model. Primarily, we implement a linear weight function defined for patient i as

$$w_{i,t} = \begin{cases} \frac{t}{T} & \text{for } t < T \\ 1 & \text{for } t \geq T \text{ or } y_{i,t} = 1 \end{cases} \quad (4.6)$$

where t is how long patient i has been on trial, and T is the length of the evaluation window. For example, suppose that a patient has been on trial for 8 weeks without any sign of DLT where the full evaluation window is 10 weeks. Although we have not observed the patient's outcome in the remaining 2 weeks of the evaluation window, an 8-week DLT-free follow-up is indicative of the safety of the dose that the patient received and can be considered as four-fifths of a non-DLT outcome. We explore and discuss an alternative form of the weight function in Section 4.5. The idea of using a weighted likelihood was first proposed in the time to event (TITE) CRM [21] framework, but has not been evaluated in the context of cell therapies with dose-feasibility considerations or with the functional forms (4.1) and (4.2).

Trial allocation is determined by treating newly enrolled patients at the working estimate of the MTD. Thus, if patient $n+1$ enrolls in the trial at time t, the estimated DLT

probability for each dose level is based on the posterior mean of the model

$$\hat{p}_j = \int \frac{\Phi\{\beta + \Phi^{-1}(q_j)\}\mathcal{L}_t(\beta|y_{i,t}, w_{i,t})f(\beta)}{\int \mathcal{L}_t(\beta|y_{i,t}, w_{i,t})f(\beta)d\beta} d\beta \qquad \text{for} \qquad j = 1, \ldots, J \tag{4.7}$$

where $f(\beta)$ is the prior distribution on β. We determine $f(\beta)$ using the algorithm in [29]. This prior, known as the least informative normal prior, assumes $\beta \sim N(0, \sigma_\beta^2)$ where the variance σ_β^2 is specified so that the prior probability of selecting any of the J dose levels is equal, implying that prior selection probability is uninformative. Then the current estimated MTD at time t is determined as

$$j_t^* = \underset{j \in \{1,\ldots,J\}}{\operatorname{argmin}} |\hat{p}_j - \theta^*| \tag{4.8}$$

where θ^* is the target DLT rate. The $n+1^{st}$ patient receives dose j_t^* if he or she is feasible to do so. If that patient's individual highest feasible dose (IHFD) is less than j_t^* then the patient receives j_t^*. Formally, a patient's IHFD has the same value as C as it was defined in (2.6). At trial conclusion the estimated MTD based on toxicity alone is determined by (4.8) after all patients have completed follow-up.

4.4 Proposed model for feasibility

Upon enrollment, patients have their cell therapy treatments manufactured. Each patient's IHFD is recorded as a count in the multinomial vector

$$\mathbf{X} = (X_0, X_1, \ldots, X_J) \sim \text{multinomial}(\pi_0, \pi_1, \ldots, \pi_J) \tag{4.9}$$

where $\sum_{j=0}^{J} \pi_j = 1$. In this formulation the count of patients with highest feasible dose j corresponds to the j^{th} coordinate of $\mathbf{X}$. For example, if a patient's IHFD is 3 they can

receive dose levels 1, 2 or 3 and their outcome is indexed in X_3. Note that X_0 is the number of patients who are not feasible to receive even the lowest dose level. If we let $\boldsymbol{\pi} \sim \text{Dirichlet}(\mathbf{a})$ where $\mathbf{a} = (a_0, a_1, \ldots, a_J)$ we can use the fact that the Dirichlet and multinomial form a conjugate family to determine the posterior distribution as

$$\boldsymbol{\pi}|\mathbf{X} \sim \text{Dirichlet}(\mathbf{a} + \mathbf{X}). \tag{4.10}$$

This type of model was posed by Thall, Sung and Choudhury [10] as a possible model alongside (2.9), but its statistical properties have never been fully evaluated. We specify the prior distribution for $\boldsymbol{\pi}$ in an uninformative manner (equal values for a_j). This yields robust results over many feasibility curves.

Consider that we are not necessarily interested in the probability that a given dose level is the highest feasible dose, rather we are interested in determining the probability that a given dose is feasible across the patient population. Thus, we propose the following definition that is similar to the definition of dose-feasibility given by Thall, Sung and Choudhury [10].

Definition: Dose level j is globally feasible if

$$P(\phi_j < \phi^*|\mathbf{X}) < p_{u,f} \tag{4.11}$$

where ϕ^* is the minimum required feasibility probability and $p_{u,f}$ is an upper probability cutoff. We can fit this into the Dirichlet-multinomial model by rewriting ϕ_j in terms of the π_j's:

$$\phi_j = P(C \geq j) \tag{4.12}$$

$$= P(C = j) + P(C = j+1) + \cdots + P(C = J) \tag{4.13}$$

$$\phi_j = \pi_j + \pi_{j+1} + \cdots + \pi_J \tag{4.14}$$

Then by the exchangeability property of the Dirichlet distribution we have:

$$(\pi_0, \pi_1, \ldots, \pi_j + \cdots + \pi_J | \mathbf{X}) = (\pi_0, \pi_1, \ldots, \phi_J | \mathbf{X}) \tag{4.15}$$

$$\sim \text{Dirichlet}(a_0 + X_0, a_1 + X_1, \ldots, \sum_{r=j}^{J} a_r + X_r) \tag{4.16}$$

then since each coordinate of the Dirichlet distribution is beta distributed we have

$$\phi_j | \mathbf{X} \sim \text{beta}(\sum_{r=j}^{J} a_r + X_r, \sum_{\ell=0}^{j-1} a_\ell + X_\ell) \qquad \text{for} \qquad j = 1, \ldots, J \tag{4.17}$$

which can be used with (4.11) to establish the set of globally feasible dose levels

$$\mathcal{F} = \{d_j : P(\phi_j < \phi^* | \mathbf{X}) < p_{u,f}\} \tag{4.18}$$

where the globally highest feasible dose (GHFD) is the maximum dose in this set. The trial proceeds by adaptively assigning patients to the current estimated MTD. If they are not feasible to receive their assigned dose, they are treated at their IHFD. During the trial if dose level 1 is not globally feasible the trial will stop. Otherwise, the toxicity model guides

dose assignments and trial conduct. At trial conclusion the FMTD is

$$\mathrm{FMTD} = \min(j^*, \mathrm{GHFD}) \tag{4.19}$$

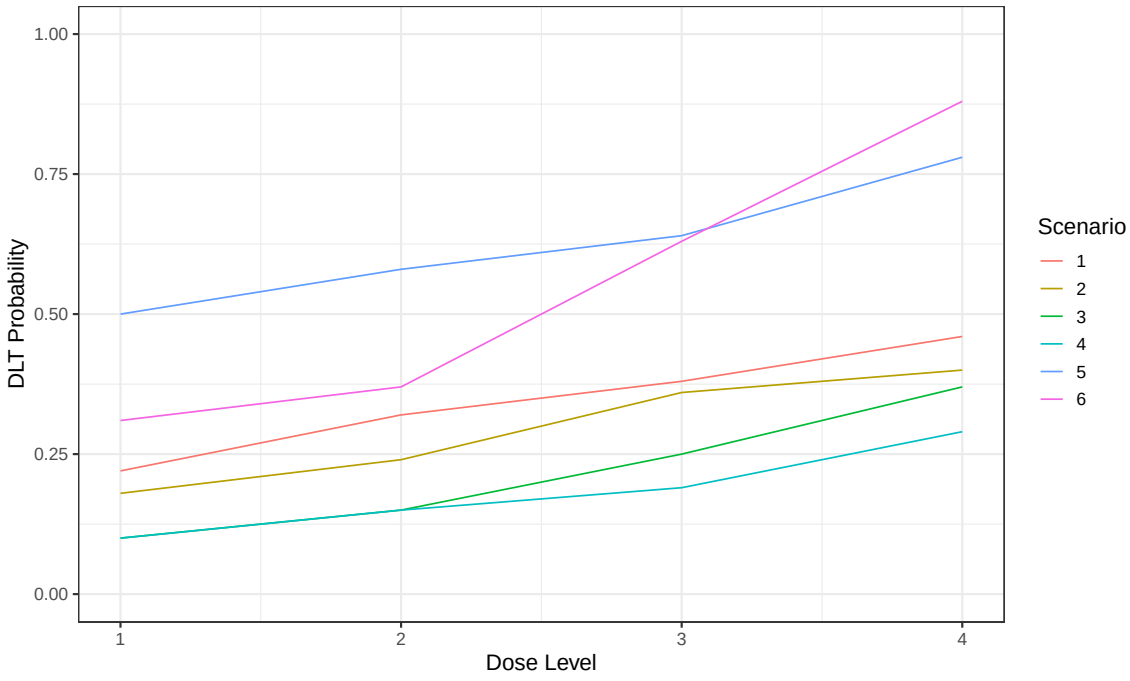
where j^* is the MTD recommended by the toxicity model once all patients are fully followed and GHFD is the maximum dose in (4.18). We also consider this dose to be the recommended phase II dose in subsequent trials.

4.5 Simulation Considerations

4.5.1 Toxicity curves

Since phase I clinical trials are adaptive it is difficult to use data from completed clinical trials to test a new design. The only way to evaluate the design using real data is to conduct an actual clinical trial. For this reason, new trial designs are usually evaluated via simulation. We generated dose-toxicity curves for use in our simulations by implementing the algorithm proposed by Conaway and Petroni [1] that we discussed in Chapter 3. The simulated DLT scenarios for this chapter are given in Figure 4.2.

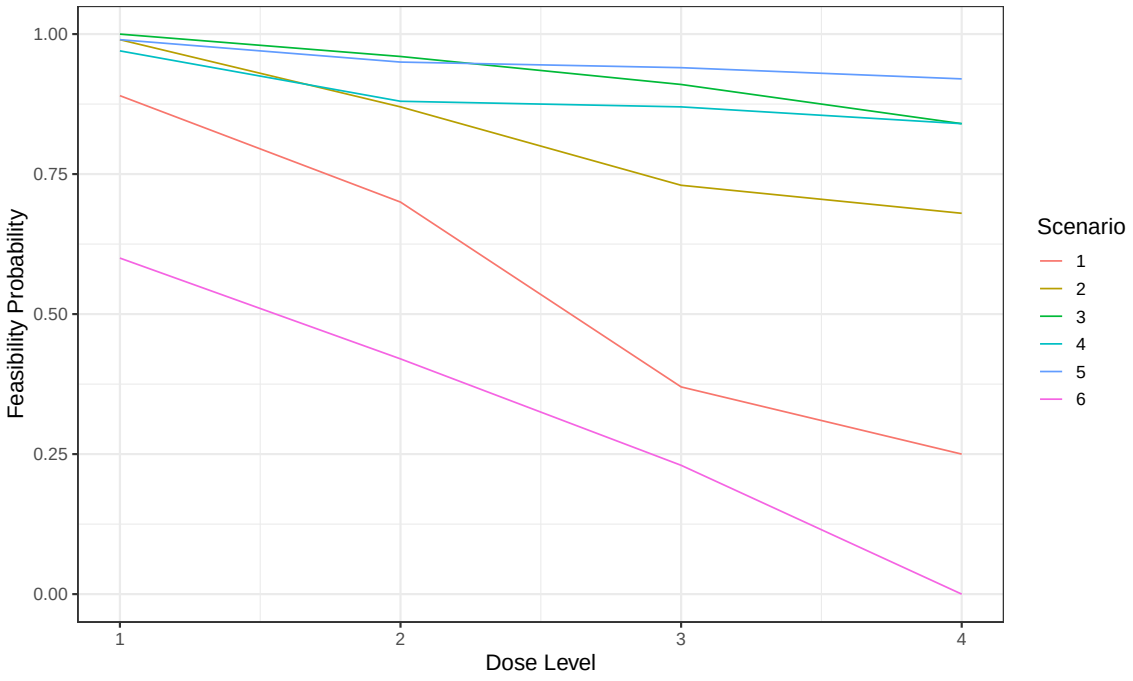
Figure 4.2: DLT probability scenarios for aim 1



4.5.2 Feasibility curves

We used the algorithm we proposed in Chapter 3 to generate the feasibility probabilities to evaluate our design under. The 6 feasibility scenarios to accompany the 6 toxicity scenarios are given in Figure 4.3.

Figure 4.3: Feasibility scenarios for aim 1



4.5.3 Stopping rule for toxicity

In traditional algorithmic designs rules are built into the trial to end patient accrual if excessive toxicity is observed. This is not the case with many model-based designs as they are mainly treating patients at the dose level with the estimated DLT probability closest to the target DLT probability. To prevent our design from continuing to treat patients if there is excessive toxicity, we will stop the trial if it is very likely that the DLT rate at the lowest level exceeds the target DLT rate. The toxicity rule is derived in the following manner:

$$p_{u,t} < P(\text{dose level 1 is too toxic}|\text{data}) \tag{4.20}$$

$$< P(\Phi\{\beta + \Phi^{-1}(q_1)\} > \theta^*|\text{data}) \tag{4.21}$$

$$< P(\beta + \Phi^{-1}(q_1) > \Phi^{-1}(\theta^*)|\text{data}) \tag{4.22}$$

$$p_{u,t} < P(\beta > \Phi^{-1}(\theta^*) - \Phi^{-1}(q_1)|\text{data}) \tag{4.23}$$

where $p_{u,t}$ is a pre-specified upper probability cutoff for toxicity. A value for (4.23) is easily computed by numerical integration.

4.5.4 Weight functions and time to toxicity distributions

Recall that in the likelihood function for our model we described that patients' weights represented their contributions to the likelihood. This weighting can be incorporated in practice in several ways. We defined a linear weight function

$$w_{i,t} = \begin{cases} \min(\frac{t}{T}, 1) & \text{if } Y_{i,t} = 0 \\ 1 & \text{if } Y_{i,t} = 1 \end{cases} \tag{4.24}$$

where t is how long patient i has been on the trial, T is the DLT window, and $Y_{i,t}$ is the i^{th} patient's toxicity status at time t. Note that we set a patient's weight to 1 when he or she experiences a DLT because at that point the patient is considered fully followed. The linear weight function is most appropriate if there is a belief that toxicity could occur at any point within the toxicity window with equal probability (often this belief is adopted due to lack of any information to the contrary). Cheung and Chappell [30], in addition to the linear weight function, consider an adaptive weight function where the shape of the

function changes based on the data accrued throughout the trial. Their adaptive weight function after n patients have been enrolled in the trial is

$$w_n(t; T, x_i) = \frac{\kappa(t)}{z+1} + \frac{1}{z+1}\left\{\frac{t - t_{(\kappa(t))}}{t_{\kappa(t)+1} - t_{\kappa(t)}}\right\} \tag{4.25}$$

where z the total number of DLTs, the time-to-toxicities $t_{(j)}s$ are ordered so that $0 = t_{(0)} < t_{(1)} \leq \cdots \leq t_{(z)} < t_{(z+1)} = T$ and $\kappa(t) = \max\{j \in [0, z] : t \geq t_{(j)}\}$. In the event that no DLTs have been observed, the adaptive weight function (4.25) simplifies to the linear weight function (4.24). The adaptive weight function adjusts the weights in a manner that depends on the timing of a DLT. For example, if three patients experience DLTs at week 1, 3 and 4 respectively a patients weight without a DLT at week 8 is 0.916 whereas in the linear weight function it would have been 8/10. This is adaptive because the toxicities in this example happened early so the risk of toxicity for a patient later in the trial is down-weighted. Despite the appeal to have a weight function that adapts to the data Cheung and Chappell found that the linear weight function (4.24) is sufficient and the use of an adaptive weight function (4.25) does not greatly improve the accuracy of the TITE-CRM. We found similar results when implementing the adaptive weight function. Simulation results for our methodology using the adaptive weight function are given in Appendix A.

A somewhat related concept to the weight function is the time to toxicity distribution. If a recruited patient in the trial experiences a DLT, the time that it happens becomes relevant. We simulated time to toxicities using uniform, Weibull and double exponential distributions. The distributions we considered are given in Figure 4.4.

The time to toxicity distribution used for simulations should reflect any known information about the drug under study. For example, if treatment-related DLTs are expected to occur only in the second half of the DLT window, then using the double exponential distribution to model time to toxicity may be more appropriate than using the uniform. Differences in

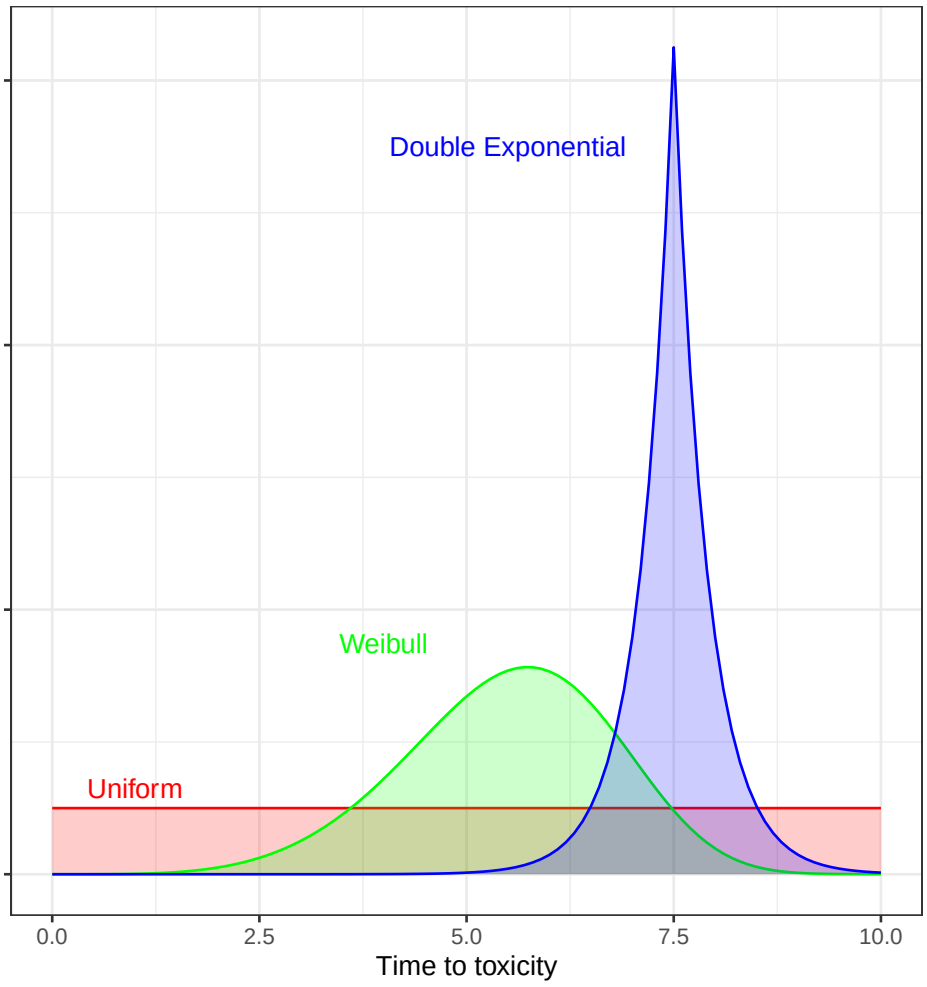


Figure 4.4: Time to toxicity distributions

simulation results, given in Appendix A, were not remarkable, but these are still relevant factors to consider when designing a phase I trial that could exhibit late-onset toxicities.

4.6 Trial conduct

Thus far, we have described the statistical models that make the calculations needed to conduct the trial in a safe and efficient manner. The following algorithm defines how all of the models work together in treating patients who enroll in the trial:

Trial conduct

1. Enroll a patient in the trial.
2. Determine the patient's IHFD. If IHFD$>$0, go to step 3. If IHFD=0, determine the GHFD. If the GHFD=0, stop the trial and set the FMTD=0. If the GHFD$>$0, go to step 1.
3. Determine the patients' feasibility and toxicity outcomes and update the data. If the toxicity stopping rule has been triggered, stop the trial and declare that no dose is the FMTD. If the toxicity stopping rule has not been triggered go to step 4.
4. Treat the patient at the current estimated MTD if their IHFD is greater than or equal to the estimated MTD. If they are not able to receive the current estimated MTD treat the patient at IHFD. If the trial is not fully enrolled go back to step 1.
5. When the trial is fully enrolled with N patients evaluated for toxicity, or the maximum number of M patients have been evaluated for feasibility stop the trial. Complete patient follow-ups and declare FMTD=min(GHFD,MTD).

To illustrate our design, we have included a simulated trial (Figure 4.5) to show how the design adaptively uses both partially and fully observed DLT data to allocate patients to their respective dose levels in a safe and efficient manner. We use the settings of the T-cell and Nivolumab phase I trial that motivated this work. The target DLT rate is $\theta^* = 0.25$ and the minimum feasibility probability is $\phi^* = 0.80$. Dose selection is guided by the trial conduct we have just described with an initial dose level of 1 and a dose-escalation restriction rule so that we do not skip untried dose levels. These restrictions are important for real world application, but will not affect simulation results a great deal. The maximum number of patients to undergo treatment is N = 24, and the maximum number of

patients to have cells expanded is M = 30. The DLT evaluation window is 10 weeks, patient time to toxicity is drawn from a continuous uniform$(0, 10)$ distribution and an average of five patients arrive every 10 weeks according to a Poisson process. The prior specifications for the skeleton are $(0.13, 0.25, 0.41, 0.59)$ determined by the algorithm described by Lee and Cheung [27]. We also set the prior distribution for β in model (4.1) to follow a Normal$(0, .74^2)$ distribution. This variance corresponds to the least informative normal prior termed by Lee and Cheung [29]. We chose these settings as they have been shown to have robust characteristics across many possible toxicity scenarios. To determine the prior parameters for the feasibility model, we assessed many specifications across 1000 feasibility scenarios generated by the algorithm proposed by Bagley and Wages [16]. The setting that was most robust across the large number of curves in selecting the correct GHFD was $a = (.2, .2, .2, .2, .2)$. This parameter choice was convenient because it represents a non-informative specification of prior dose-feasibility.

The first patient enrolls in the simulated trial (Figure 4.5) at dose level 1 and he or she does not have a DLT. When the second patient enrolls 4 weeks later, we have observed 40% of the first patient's non- DLT. Patient 2 is allocated to dose level 2 since that is the working estimate of the MTD according to the toxicity model. Patient 3 enrolls two weeks after patient 2. At this point, we have observed 60% of the first patient's non-DLT at dose level 1 and 20% of the second patient's non-DLT at dose level 2. The working estimate of the MTD is dose level 2 so patient 3 is also treated at dose level 2. Patient 4 enrolls in the trial when the working estimate of the MTD is still dose level 2, but their cell expansion results in them not being feasible to receive any of the dose levels, so they are not treated under the trial protocol. Despite patient 4 not being able to receive any dose level, the computed feasibility estimates indicate that the lowest dose level is globally feasible, so the trial continues to enroll patients. When patient 5 enrolls, the accumulated safety data amounts to 8/10 of a non-DLT at dose level 1 and 6/10 of a non-DLT at dose level 2. The working MTD is still dose level 2 at which patient 5 is allocated. Note that if we had

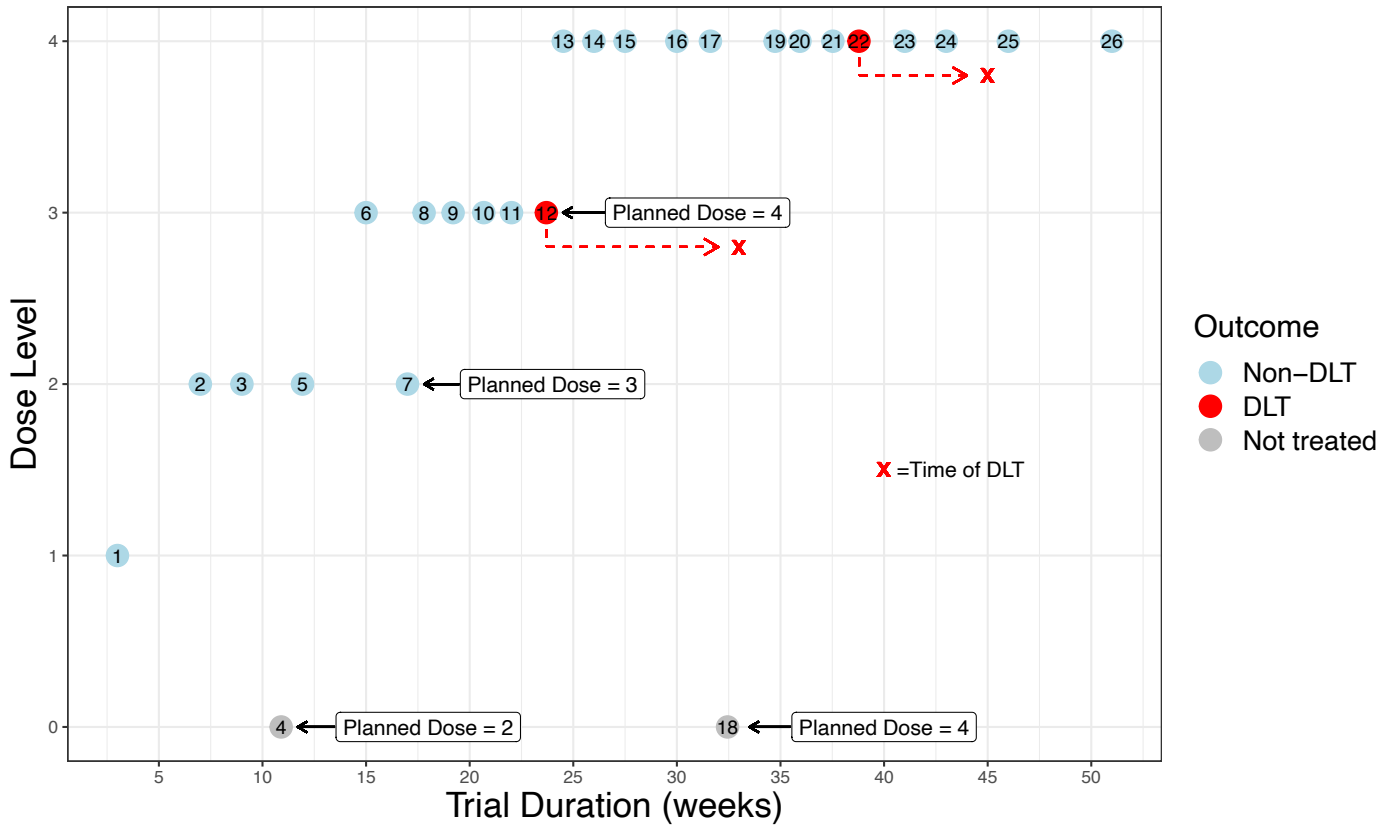


Figure 4.5: A simulated trial using the proposed time to event methodology

to suspend accrual until the full DLT evaluation window was completed we would have treated just one patient at this point, but because we are utilizing partially observed DLT data we have safely accrued 3 additional patients to the trial. When patient 6 enrolls in the trial we have observed 1 non-DLT at dose level 1 and the equivalent of 1.7 non-DLTs at dose level 2 even though none of the patients treated at this dose level have completed the DLT observation window. With this updated information the working estimate of the MTD increases to dose level 3, and patient 6 receives this dose. Patient 7 enrolls soon after, when the working estimate of the MTD is still dose level 3, but he or she is only feasible to receive up to dose level 2, so this patient are treated at dose level 2. Patients 8–11 enroll in quick succession and are all treated at the working estimate of the MTD, dose level 3. When patient 12 enrolls in the trial we have observed 1 non-DLT at dose level 1, 3.6 non-DLTs at dose level 2, and 2.1 non-DLTs at dose level 3. At this point the

estimated MTD is dose level 4, but patient 12 has an IHFD of 3 and thus is treated at that dose. Patient 12, despite being treated below the planned dose has a late-onset DLT, but this patient contributes to the toxicity model as a non-DLT until the DLT is observed 9 weeks after their enrollment in the trial. Since patient 12 is initially recorded as a non-DLT the working estimate of the MTD stays at dose level 4 and patients 13–17 are all treated at this dose level. Patient 18 is assigned to dose level 4 as well, but is excluded from the trial since the patient's IHFD is 0. The feasibility model is checked, and dose level 1 is globally feasible, so the trial continues. After patient 18 is unable to enroll in the trial, patient 12's late onset DLT is observed and is recorded as a fully observed DLT. When patient 19 enrolls we have observed 1 non-DLT at dose level 1, 4 non-DLTs at dose level 2, 5 non-DLTs and 1 DLT at dose level 3 and 3.6 non-DLTs at dose level 4. The estimated MTD at this point is dose level 4 and that is the estimated MTD for the rest of the trial. Patients 19–26 are treated at dose level 4 and one patient has a late-onset DLT. At trial conclusion the working estimate of the MTD is dose level 4 and the estimated GHFD according to the feasibility model is level dose 4. Thus, the estimated FMTD for the trial is dose level 4. Our design's ability to use partially observed DLT data caused the final trial duration to be 61 weeks. If we had to fully observe each patient's DLT data before enrolling a new patient, the trial could have taken as long as 240 weeks, nearly 4 times the amount of time it took to complete the trial using our design. While this is just a hypothetical illustration of how our design works, the possible benefits from the gain in trial efficiency are worth the design challenge of using partially observed DLT data.

4.7 Simulation Results

4.7.1 Simulation results across 6 toxicity/feasibility scenarios

We conducted a Monte-Carlo simulation to evaluate our trial design for six combinations of randomly generated dose-feasibility and dose-toxicity scenarios. We generated the dose–feasibility curves using the algorithm described by Bagley and Wages [16] as it provides curves that monotonically decrease with dose level, which is consistent with the set-up we have described. We generated the set of "s" shaped dose–toxicity curves using the algorithm described by Conaway and Petroni [1]. Under each scenario, we simulated 2000 trials using the same settings described in the previous section. We have also conducted simulation studies under a variety of other settings (i.e. time to toxicity distributions, weight functions, etc.), and included these in Appendix A. The generated curves are given, along with the operating characteristics of the design in Table 2. Table 2 displays the operating characteristics of our design, and a "complete data" form of our design that does not assign patients until all observations are fully observed. Recall that our design can use partially observed DLT data to allocate patients to the trial to make the trial more efficient. Trial efficiency is always preferred if it can be gained without undermining patient safety or reducing FMTD selection accuracy. Under scenario 1 the true FMTD is dose level 1 and each of the designs treat virtually the same number of patients at each dose level. The FMTD selection accuracy differs by only .3%, but the average trial duration for the more efficient design is about 200 weeks ($\sim$4 years) faster than the duration of the complete data design. Scenarios 2 and 3 see a little more variation in trial performance, but most differences are about one patient on average or a couple of percentage points in FMTD selection. This small difference compared with the almost 4-year reduction in trial duration is likely a beneficial trade-off. Scenario 4 has somewhat lower trial performance for both methods. This is likely due to the target DLT rate of

$\theta^* = 0.25$ splitting the true DLT rates at the surrounding dose levels. There is more of a difference in performance between the methods at dose level 4. The complete data design selects dose level 3, with an acceptable feasibility rate of 0.87 and a tolerable DLT rate of 0.19, in a higher percentage of trials than the proposed design. Scenario 5 is a situation where dose level 1 is too toxic and thus the correct decision is to stop the trial based on the toxicity stopping rule. The design that does not use partial information does about 5% better in correctly stopping the trial, treats three fewer patients, and the trial duration is reduced by about 2 years as opposed to four. Scenario 6 has true dose-feasibility probabilities below our prespecified minimum $\phi^* = 0.80$ and thus the right decision is to stop the trial because dose level 1 is not globally feasible. Both models have similar patient allocation, and FMTD selection under this scenario, and the gain in trial efficiency is about 60 weeks. In addition to the proposed design and the complete data design, we applied the method of Wages and Fadul [11] to the 6 scenarios using the web application located at `http://uvatrapps.uvadcos.io/wfdesign/`. This method is not a time-to-event method, but the goal of determining the FMTD is the same. Results are given in Figure 4.2. The results are quite similar to the version of our method that uses only complete data. This makes sense because the method proposed by Wages and Fadul [11] is not designed for late onset toxicities. The operating characteristics are similar to those of the proposed method with the exception of trial duration. This shows that our method has comparable operating characteristics to multiple complete data methods while also speeding up the trial.

4.7.2 Simulation results across 100 toxicity/feasibility scenarios

In addition to the six scenarios described in the previous section, we conducted a large-scale simulation over 100 possible scenarios. For these simulation results we kept the same simulation settings as in the previous section. Figure 4.6 shows the average dura-

Scenario		Dose 1	Dose 2	Dose 3	Dose 4	Tox Stop	Feas Stop	Average Duration (weeks)
1	(p_f, p_t)	**(0.89,0.22)**	(0.7,0.32)	(0.37,0.38)	(0.25,0.46)			
	Proposed	**69.1 [15.6]**	10.3 [5.3]	0.0 [1.2]	0.0 [0.3]	4.9	5.7	59.8
	Complete Data	**68.3 [14.7]**	19.5 [5.8]	0.0 [1.2]	0.0 [0.2]	7.0	5.3	254.6
	Wages Fadul	**66.0 [16.0]**	23.9 [6.1]	0.1 [0.3]	0.0 [0.1]	7.2	2.9	251.8
2	(p_f, p_t)	(0.99,0.18)	**(0.87,0.24)**	(0.73,0.36)	(0.68,0.4)			
	Proposed	33.8 [12.7]	**51.1 [6.5]**	11.3 [2.8]	1.5 [1.7]	2.3	0.0	57.2
	Complete Data	33.6 [11.7]	**50.6 [7.2]**	11.9 [3.3]	1.5 [1.5]	2.5	0.0	246.9
	Wages Fadul	29.2 [9.9]	**48.7 [8.8]**	16.2 [3.4]	3.5 [1.5]	2.4	0.0	237.6
3	(p_f, p_t)	(1,0.01)	(0.96,0.15)	**(0.91,0.25)**	(0.84,0.37)			
	Proposed	5.7 [8.1]	33.5 [6.1]	**49.6 [5.9]**	11.1 [3.9]	0.1	0.0	57.9
	Complete Data	5.6 [6.6]	32.4 [6.5]	**49.2 [6.9]**	12.6 [3.9]	0.3	0.0	248.7
	Wages Fadul	1.3 [2.9]	31.3 [8.0]	**48.1 [8.6]**	19.3 [4.6]	0.0	0.0	240.8
4	(p_f, p_t)	(0.97,0.1)	(0.88,0.15)	(0.87,0.19)	**(0.84,0.29)**			
	Proposed	8.6 [8.6]	27.2 [4.9]	37.0 [5.0]	**26.6 [5.3]**	0.2	0.5	58.4
	Complete Data	7.2 [6.8]	23.5 [4.8]	39.1 [6.1]	**29.9 [6.2]**	0.2	0.2	255.4
	Wages Fadul	5.2 [5.5]	18.6 [4.8]	39.1 [7.0]	**36.5 [6.7]**	0.4	0.1	246.8
5	(p_f, p_t)	(0.99,0.5)	(0.95,0.58)	(0.94,0.64)	(0.92,0.78)			
	Proposed	25.0 [14.4]	0.2 [1.0]	0.0 [0.4]	0.0 [0.3]	**74.7**	0.1	40.3
	Complete Data	16.3 [11.4]	0.1 [0.8]	0.0 [0.3]	0.0 [0.2]	**83.7**	0.1	130.1
	Wages Fadul	14.4 [10.9]	0.4 [1.4]	0.1 [0.4]	0.0 [0.1]	**85.2**	0.0	128.9
6	(p_f, p_t)	(0.6,0.31)	(0.42,0.37)	(0.23,0.63)	(0,0.88)			
	Proposed	5.1 [4.9]	0.0 [0.7]	0.0 [0.2]	0.0 [0.0]	2.0	**92.9**	23.6
	Complete Data	5.1 [4.6]	0.0 [0.8]	0.0 [0.2]	0.0 [0.0]	5.9	**89.1**	84.0
	Wages Fadul	6.2 [5.5]	0.0 [0.8]	0.0 [0.1]	0.0 [0.0]	8.4	**85.4**	106.9

Table 4.2: Operating characteristics and duration for the proposed design, the complete data design, and the design proposed by Wages and Fadul. FMTD selection percentages are given followed in brackets by the average number of patients treated at each dose. Tox stop and Feas stop refer to the stopping rule triggered in the case of early termination. $p_f = P(\text{dose level is feasible})$ and $p_t = P(\text{DLT at the given dose level})$.

tion of the two designs we have compared thus far. The proposed late onset method has a smaller average duration than the slower design across all 100 scenarios under consideration. This gain in trial efficiency is achieved without a significant decrease in trial accuracy. Across all 100 scenarios, the proposed design is about .5% less accurate at correctly identifying the true FMTD. The combination of increased trial efficiency without the loss of trial accuracy demonstrates how our method is fundamentally better suited for implementation in cell therapy trials with possibly late-onset DLT outcomes.

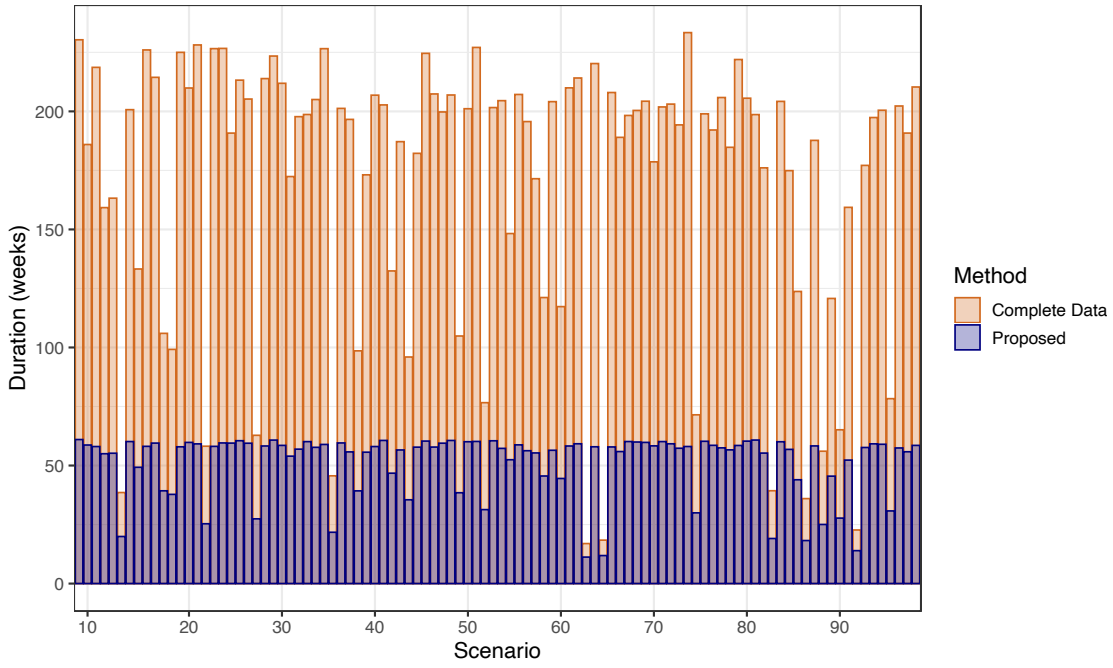


Figure 4.6: Average duration comparison for the proposed method and the method that can only use complete data over 100 toxicity feasibility scenarios

Chapter 5

Aim 2: Drug combinations and dose-feasibility

5.1 Background on dose-finding for drug combinations

Phase I cancer trials that consider multi-drug dose-escalation are of growing interest due to the potential positive response that a drug combination could provide. If this potential is large, then conducting a phase I dose-escalation trial of multiple agents can outweigh additional complications that accompany drug combination trials. Recall that the primary assumption for single agent dose-escalation is that DLT probability increases with the dose level. In principle, this assumption is still present for drug combination trials, but the full toxicity ordering of the dose levels may not be fully known. To illustrate this idea, consider the drug combination example given in Table 5.1.

For these dose level definitions, we may know only some of the orders between various dose levels with respect to toxicity. For example, if we were to escalate from d_1 to d_2 we would expect d_2 to be more toxic because we are holding Drug B fixed while escalating

Table 5.1: Example drug combination dose levels

		Drug A		
		20mg/kg	40mg/kg	60mg/kg
Drug B	50mg/kg	d_1	d_2	d_3
	100 mg/kg	d_4	d_5	d_6

Drug A from 20mg/kg to 40mg/kg. However, consider the escalation from d_2 to d_4. Drug B increases from 50mg/kg to 100mg/kg while Drug A decreases from 40mg/kg to 20mg/kg. Since we are increasing one drug dose while simultaneously decreasing the other, there is not a clear toxicity order between d_2 and d_4. Thus, we know only a partial order of the dose levels with respect to toxicity. To address this, it is beneficial to consider the possible simple toxicity orders of the drug combinations. Table 5.2 shows the simple orders of the dose levels defined in Table 5.1.

Table 5.2: Simple toxicity orders

Order	Less Toxic $\longrightarrow$ More toxic
1	$d_1 \to d_2 \to d_3 \to d_4 \to d_5 \to d_6$
2	$d_1 \to d_2 \to d_4 \to d_3 \to d_5 \to d_6$
3	$d_1 \to d_2 \to d_4 \to d_5 \to d_3 \to d_6$
4	$d_1 \to d_4 \to d_2 \to d_3 \to d_5 \to d_6$
5	$d_1 \to d_4 \to d_2 \to d_5 \to d_3 \to d_6$

Notice that in three of the orders d_2 is less toxic than d_4 while in the other two orders the reverse is true. Despite the uncertainty of the order between d_2 and d_4, we know some orders more fully. For example, we know that d_1 and d_6 are always the least and most toxic doses respectively since they are made up of the highest and lowest drug combinations. In this example, there are 5 possible dose-toxicity orders, but as the number of agents or dose levels under consideration increase the number of possible orderings will increase rapidly. One reason we have listed each of the simple orders is to illustrate and motivate the idea that if we can select and order then methodology developed for a single agent can be of use. We will discuss this at greater length when we introduce methodology for this aim.

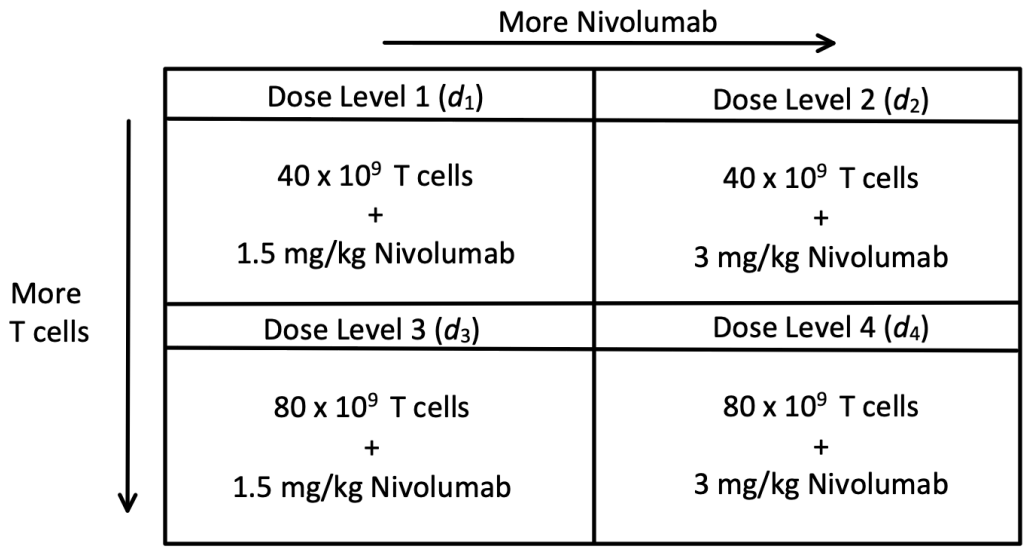
Despite these differences, the general objectives of a drug combination trial is similar to that of a single agent trial: Adaptively assign patients to their dose levels based on previous patients' DLT outcomes. Then at the conclusion of the trial estimate the maximum tolerated dose, or in this setting, estimate the maximum tolerated dose combination (MTDC). The identified MTDC should be associated with some pre-specified toxicity rate, and can then be used in later stages of the drug development process.

5.2 Motivating trial 2: Dual escalation of a cell therapy and an immunotherapy

The motivating trial for this aim was also proposed by oncologists at UVA's comprehensive cancer center. The trial calls for the treatment of high-grade B-cell lymphoma with Rituximab-based Bispecific antibody activated T cells in combination with Nivolumab. The T cell product is the cell therapy and Nivolumab is an anti-cancer immunotherapy that has been approved by the FDA in other drug combination settings. Similarly to the treatment discussed in the previous aim, we could expect dose-feasibility issues with respect to the cell therapy product where some patients might not be feasible to receive their assigned dose. The Nivolumab could also cause late-onset DLTs up to 10 weeks after initial treatment. The key difference for this aim is that we are considering multiple dose levels of both therapeutics as opposed to the previous assumption of a fixed dose of Nivolumab. The dose level definitions for this trial are given in Figure 5.1.

For this trial, the possible simple orders with respect to toxicity are $d_1 \to d_2 \to d_3 \to d_4$ and $d_1 \to d_3 \to d_2 \to d_4$, because from dose level d_2 to dose level d_3 we are increasing the T cell dose while simultaneously reducing the dose of the Nivolumab leaving the full toxicity order unknown. However, the order with respect to dose levels d_1 and d_4 are fully

Figure 5.1: Dose level definitions for motivating trial



known since they are combinations of the lowest and highest dose level of both drugs. Despite not knowing the full toxicity order of the dose levels, the feasibility order between all dose levels is fully known. Since the cell count for d_1 and d_2 are both 40×10^9 they are equally feasible. The same can be said for d_3 and d_4 since they both have 80×10^9 T cells. In this context, dose levels d_1 and d_2 are generally expected to be more feasible for a patient to receive than dose levels d_3 and d_4.

Drug combination trials are clearly more complex than single agent trials. We believe that the lack of statistically valid methodology tailored for this type of trial will result in oncologists using ad-hoc dose-escalation rules. According to a review [31] of published phase I drug combination trials, 88% used some form of 3+3 design and "all except one trial used a design developed for single-agent evaluation." This is despite the availability of many statistically valid trial designs proposed for dual escalation of drug combinations [14, 32–36].

Neither algorithmic designs, like the 3+3, nor statistical methods like the CRM, are suited to handle the issues of dose-feasibility, late-onset toxicity and partial toxicity or-

ders. The following sections of this chapter will provide the first phase I trial design for drug combination studies that simultaneously accounts for dose-feasibility considerations and late-onset toxicities. In this framework, our primary objective is to determine the feasible maximum tolerated dose combination (FMTDC), defined as the combination that is sufficiently safe that consists of a cell therapy dose that is feasible across the patient population. Patients are efficiently accrued to the trial and assigned to the working estimate of the MTDC while adaptively updating and accounting for the toxicity order of the dose combinations.

5.3 Proposed model for toxicity

We propose a DLT probability model that adaptively assigns patients to one of a set of prespecified drug combination dose levels where the DLT order is only partially known. To address the problem of partially ordered dose levels, we assume there are $m = 1, \ldots, M$ simple dose orders. Under this framework, we assume a single parameter toxicity model

$$p_{j,m}(\beta, q_{j,m}) = \Phi\{\beta + \Phi^{-1}(q_{j,m})\} \qquad \text{for} \qquad j = 1, \ldots, J \tag{5.1}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function, $\Phi^{-1}(\cdot)$ is the standard normal quantile function, β is the parameter, and $q_{j,m}$ are the skeleton values corresponding to a particular simple ordering m. To illustrate how this works consider the following example: suppose the values for the skeleton are (0.06, 0.14, 0.25, 0.38, 0.50, 0.61), then under the example in Table 5.2 the possible skeleton values for each simple order are given in Table 5.3.

	Skeleton Values					
Simple Order	q_1	q_2	q_3	q_4	q_5	q_6
$m=1$	0.06	0.14	0.25	0.38	0.50	0.61
$m=2$	0.06	0.14	0.38	0.25	0.50	0.61
$m=3$	0.06	0.14	0.38	0.50	0.25	0.61
$m=4$	0.06	0.38	0.14	0.25	0.50	0.61
$m=5$	0.06	0.38	0.14	0.50	0.25	0.61

Table 5.3: Skeleton values under different simple orders

For this aim, DLT outcome is still a function of time, so the i^{th} patient's DLT status at time t is

$$y_{i,t} = \begin{cases} 1 & \text{if patient } i \text{ has DLT before time } t \\ 0 & \text{if patient } i \text{ has not had DLT before time } t \end{cases} \quad \text{for} \quad i = 1, \ldots, n. \tag{5.2}$$

The Likelihood function is

$$\mathcal{L}_{t,m}(\beta | y_{i,t}, w_{i,t}) = \prod_{i=1}^{n} \left(w_{i,t} p_{j,m}(\beta, q_{i,j,m})\right)^{y_{i,t}} \left(1 - w_{i,t} p_{j,m}(\beta, q_{i,j,m})\right)^{1-y_{i,t}} \tag{5.3}$$

where m is the simple order under which the likelihood is evaluated, $p_{j,m}$ is defined as in (5.1), $q_{i,j,m}$ is the value of the skeleton under order m corresponding to the dose patient i received and $w_{i,t}$ determines the i^{th} patient's contribution to the likelihood function at time t.

The added complexity for this aim is that we need additional methodology to select the simple order $m = 1, \ldots, M$. Wages, Conaway and O'Quigley [32, 37, 38] suggest a procedure that selects a simple order by weighting each order using the likelihood function. Once an order is selected, estimation of the DLT probabilities reduces to the single agent methodology we have already developed. We utilize a similar strategy by weighting

each of the models by

$$\omega(m|\text{data}) = \frac{\omega(m)\int \mathcal{L}_{t,m}(\beta|y_{i,t}, w_{i,t})f(\beta)d\beta}{\sum_{m=1}^{M} \omega(m)\int \mathcal{L}_{t,m}(\beta|y_{i,t}, w_{i,t})f(\beta)d\beta} \qquad \text{for } m = 1, \ldots, M \qquad (5.4)$$

where $\omega(m)$ is the prior probability that order m is the correct order, $f(\beta)$ is the prior distribution for β and $\mathcal{L}_{t,m}(\cdot)$ is the likelihood function evaluated under order m at time t. Equation (5.4) is appealing because it is driven purely by the likelihood function and if the likelihood is large under a particular order with respect to the other orders, then the trial will proceed under that order. If there is no prior information on which order is most likely then a good prior for $\omega(m)$ is a discrete uniform distribution. However, if a clinical collaborator thinks a particular order is more likely this can easily be incorporated by changing the prior probabilities for $\omega(m)$. As patient data is accrued the order with the highest likelihood according to the data should be favored. To show how this works, consider the example patient data given in Table 5.4 where the dose levels correspond to Figure 5.1 and the two possible orders are $d_1 \to d_2 \to d_3 \to d_4$ and $d_1 \to d_3 \to d_2 \to d_4$. The observed DLT rates at d_2 and d_3 are $\frac{2}{3}$ and $\frac{1}{3}$ respectively which is evidence that d_2 is more toxic than d_3. The calculated values for Equation (5.4) for $m = 1$ and $m = 2$ are 0.39 and 0.61 respectively. This implies that $m = 2$ is more likely and patient 8 will have the assigned dose level determined under that simple order.

Patient #	1	2	3	4	5	6	7
Dose Received	d_1	d_2	d_2	d_2	d_3	d_3	d_3
DLT Status	0	1	1	0	0	0	1

Table 5.4: Example Trial to illustrate order selection

Once a simple order m is selected we can determine the estimated DLT probability at each dose level as

$$\hat{p}_{j,m} = \int \frac{\Phi\{\beta + \Phi^{-1}(q_{j,m})\}\mathcal{L}_{t,m}(\beta|y_{i,t}, w_{i,t})f(\beta)}{\int \mathcal{L}_{t,m}(\beta|y_{i,t}, w_{i,t})f(\beta)d\beta} d\beta \qquad \text{for} \qquad j = 1, \ldots, J \tag{5.5}$$

where $f(\beta)$ is the prior distribution on β. We determine $f(\beta)$ in the same manner as the toxicity model for the previous aim. The current estimated MTD at time t under order m is determined as

$$j^*_{t,m} = \underset{j \in \{1,\ldots,J\}}{\operatorname{argmin}} |\hat{p}_{j,m} - \theta^*| \tag{5.6}$$

where θ^* is the target DLT rate. If a patient is not feasible to receive their assigned dose they will receive their IHFD consisting of a dose combination that has an estimated DLT probability less than or equal to the MTD estimate. We will discuss this more formally in the following two sections. At the conclusion of the trial the estimated MTD, informed by all patient data, is determined by (5.6) under the final estimated simple order.

The added complexity of considering dose combinations will require, for reasons that will become clear in the following section, us to define a set of safe dose levels with respect to the estimated MTD. Let $\mathcal{S}_m$ be defined as

$$\mathcal{S}_m = \{d_j : \hat{p}_{j,m} \leq \hat{p}_{j^*_{t,m}}\} \tag{5.7}$$

where m is a particular simple order, and $\hat{p}_{j^*_{t,m}}$ is the DLT probability of the estimated MTD $j^*_{t,m}$ as defined in (5.6). Patients who are not feasible (cannot receive their assigned doses) will receive their IHFDs provided that they are safe. We discuss this set further in the following section.

5.4 Proposed model for feasibility

The problem of dose level feasibility for dual drug escalation trials where one of the drugs is a cell therapy is similar to that of the single-agent case. The key difference is that the feasibility will only change as the dose of the cell therapy changes. Let the doses of the cell therapy be indexed by $k = 1, \ldots, K$ as opposed to $j = 1, \ldots, J$. When a patient enrolls in the trial, the highest feasible cell therapy dose is recorded as a count in the multinomial vector

$$\mathbf{X} = (X_0, X_1, \ldots, X_K) \sim \text{multinomial}(\pi_0, \pi_1, \ldots, \pi_K) \tag{5.8}$$

where $\sum_{k=0}^{K} \pi_k = 1$. In this parameterization, the k^{th} coordinate of $\mathbf{X}$ is the count of patients with highest feasible cell dose k. X_0 is the number of patients not feasible to receive even the lowest dose of cell therapy and will not be enrolled in the trial. Note that cell dose k could correspond to multiple dose combinations. For the motivating example (Figure 5.1), dose levels d_1 and d_2, defined by 40×10^9 T cells, would correspond to $k = 1$ whereas dose levels d_3 and d_4, defined by 80×10^9 T cells, would correspond to $k = 2$. If we let $\boldsymbol{\pi} \sim$ Dirichlet($\mathbf{a}$) where $a = (a_0, a_1, \ldots, a_K)$ we can again use the result that the Dirichlet and multinomial distributions form a conjugate family and determine the posterior distribution as

$$\boldsymbol{\pi}|\mathbf{X} \sim \text{Dirichlet}(\mathbf{a} + \mathbf{X}). \tag{5.9}$$

This feasibility model is almost identical to the the feasibility model we have already introduced (4.10), except for the fact that the dimension is defined solely by the cell dose levels. Now, we define global feasibility for each dose level $j = 1, \ldots, J$ in terms of k.

Definition: Dose level j is globally feasible if its associated cell dose k satisfies

$$P(\phi_k < \phi^* | \mathbf{X}) < p_{u,f} \tag{5.10}$$

where ϕ^* is the minimum required feasibility probability, and $p_{u,f}$ is an upper probability cutoff. We can calculate these probabilities as

$$\phi_k | \mathbf{X} \sim \text{beta}(\sum_{r=k}^{K} a_r + X_r, \sum_{\ell=0}^{k-1} a_\ell + X_\ell) \qquad \text{for} \qquad k = 1, \ldots, K. \tag{5.11}$$

The global highest feasible cell dose is defined as

$$\hat{d_k} = \max_{k \in 1,\ldots,K} \{d_k : P(\phi_k < \phi^* | \mathbf{X}) < p_{u,f}\} \tag{5.12}$$

where $p_{u,f}$ is an upper probability cutoff for feasibility. Then define the function

$$\mu(d_i) = \text{ \# of cells defined by dose } d_i. \tag{5.13}$$

Putting all of this together, we can define the set of globally feasible dose levels as

$$\mathcal{F}_k = \{d_j : \mu(d_j) \leq \mu(\hat{d_k})\}. \tag{5.14}$$

The set defined by (5.14) contains all of the dose levels from $d_1, \ldots, d_J$ that are defined by a cell dose that is feasible for a large part of the patient population to receive. The possible sets globally feasible dose levels for the motivating trial are are $\mathcal{F}_2 = \{d_1, d_2, d_3, d_4\}$, $\mathcal{F}_1 = \{d_1, d_2\}$ and $\mathcal{F}_0 = \emptyset$. $\mathcal{F}_2$ is the situation where all levels of the cell therapy are feasible. $\mathcal{F}_1$ is when dose levels d_3 and d_4, defined by $80x10^9$ T cells, are not feasible according

to the data. $\mathcal{F}_0$ is the case where no dose levels are feasible. Note that there are fewer sets of feasible dose levels than total number of doses, so this should be reflected in the feasibility model.

Recall that one of the primary objectives for this type of trial is to determine the FMTDC. To do this, we first need to define the set of feasible and safe doses as $\Omega_{\mathcal{SF}}$ where

$$\Omega_{SF} = \{j : d_j \in \mathcal{S}_m \bigcap \mathcal{F}_k\} \tag{5.15}$$

where S_m is the set of safe doses under model m at trial conclusion as defined in (5.7), and $\mathcal{F}_k$ is the set of globally feasible dose levels as defined in (5.14). From this set, we define the FMTDC as

$$\text{FMTDC} = \underset{j \in \Omega_{SF}}{\operatorname{argmin}} |\hat{p}_{j,m} - \theta^*| \tag{5.16}$$

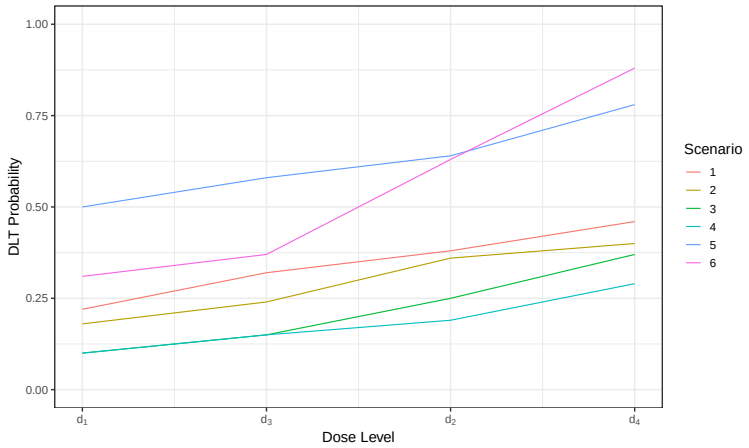
where $\hat{p}_{j,m}$ are the estimated DLT probabilities under model m, and θ^* is the target DLT probability. If the trial stops due to low feasibility or high toxicity we set the FMTDC to 0. The FMTDC will likely be the dose recommended for use in a phase II trial.

5.5 Simulation Considerations

5.5.1 Toxicity curves

We will utilize the same DLT probability curves as in the previous aim. The curves under $m = 1$ are given in Figure 4.2. The DLT probability curves under $m = 2$ are given in Figure 5.2. Notice that d_3 is less toxic than d_2.

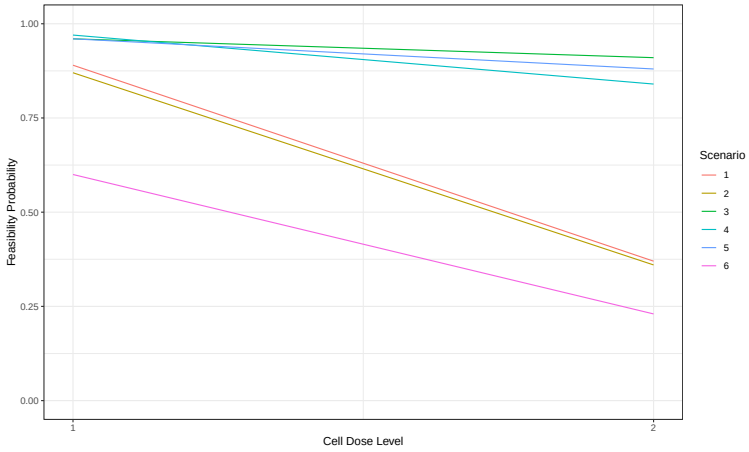
Figure 5.2: DLT probability scenarios under order m=2



5.5.2 Feasibility curves

Since there are fewer cell dose levels than in the previous aim we must use a different set of feasibility probabilities. The feasibility scenarios, generated by the algorithm we proposed in Chapter 3 are given in Figure 5.3. These scenarios look much simpler than the previous feasibility scenarios because there are just 2 levels of feasibility.

Figure 5.3: Feasibility scenarios for aim 2



5.5.3 Other considerations

For drug combinations, we have the same stopping rules as the previous aim. If the lowest combination is either too toxic or not feasible then the trial stops and no FMTDC is declared. We also have additional simulation results for the time to toxicity distribution, weight function, and DLT model choice in Appendix B. These simulation considerations are identical to that of the previous aim.

5.6 Trial Conduct

The feasibility and toxicity models described in the previous subsections guide trial conduct, determine which doses patients receive, and make final recommendations related to toxicity, dose order, and feasibility. The following formally defines how the trial operates.

Trial Conduct

1. Enroll a patient in the trial.
2. Determine the patient's IHFD for the cell therapy. If IHFD$>$0, go to step 3. If IHFD=0, determine the GHFD. If the GHFD=0, stop the trial and set the FMTD=0. If the GHFD$>$0, go to step 1.
3. Determine the patients' feasibility and toxicity outcomes and update the data. If the toxicity stopping rule has been triggered, stop the trial and declare that no dose is the FMTD. If the toxicity stopping rule has not been triggered go to step 4.
4. Set the order with the largest posterior probability to the current order breaking ties at random. Under this order, update the estimated DLT probabilities.

5. Treat the patient at the current estimated MTD if feasible. Otherwise, select from their feasible dose levels the dose with DLT probability closest to the target that is less than the estimated MTD's DLT probability. If the trial is not fully enrolled go to step 1. If the trial is fully enrolled go to step 6.

6. When the trial is fully enrolled with N patients evaluated for toxicity, or the maximum number of M patients have been evaluated for feasibility stop the trial. Complete patient follow-ups and determine the FMTDC as in (5.16).

Figure 5.4 gives a simulated trial to illustrate how our design adaptively uses patient DLT data to select the FMTDC from a set of dose combinations in a context where there could be late-onset toxicities. We utilize the settings of the motivating trial for this aim where we have a dual-escalation of T cells and Nivolumab. The dose levels we considered are defined in Figure 5.1. The target DLT rate is $\theta^* = 0.25$ and the minimum rate of dose-feasibility is $\phi^* = 0.80$. Dose selection and patient accrual is determined by the trial conduct we have just described. The maximum number of patients to be treated under the protocol is $N = 24$ with a maximum of $M = 30$ patients to have their cells expanded and evaluated for feasibility. The DLT evaluation window is 10 weeks, patient time to toxicity are simulated from a uniform(0,10) distribution and we expected 1 patient arrival per month, or approximately 2.5 patients per DLT evaluation window. The skeleton is set to $(0.13, 0.25, 0.41, 0.59)$ according to the algorithm given by Lee and Cheung [27]. We also use the least informative normal prior $\mathcal{N}(0, .74^2)$ described by Lee and Cheung [29]. For these priors to be useful in the context of a drug combination trial we need to select a simple order for dose levels with respect to toxicity. The two simple orders for $m = 1$ and $m = 2$ are $(d_1 \to d_2 \to d_3 \to d_4)$ and $(d_1 \to d_3 \to d_2 \to d_4)$ respectively. Prior to the trial, we have no indication whether $m = 1$ or $m = 2$ is correct so we set the prior probabilities as $\omega(m = 1) = \frac{1}{2}$ and $\omega(m = 2) = \frac{1}{2}$. As the trial progresses, we select the order with largest posterior probability as defined in (5.4) and estimate the DLT probabilities under

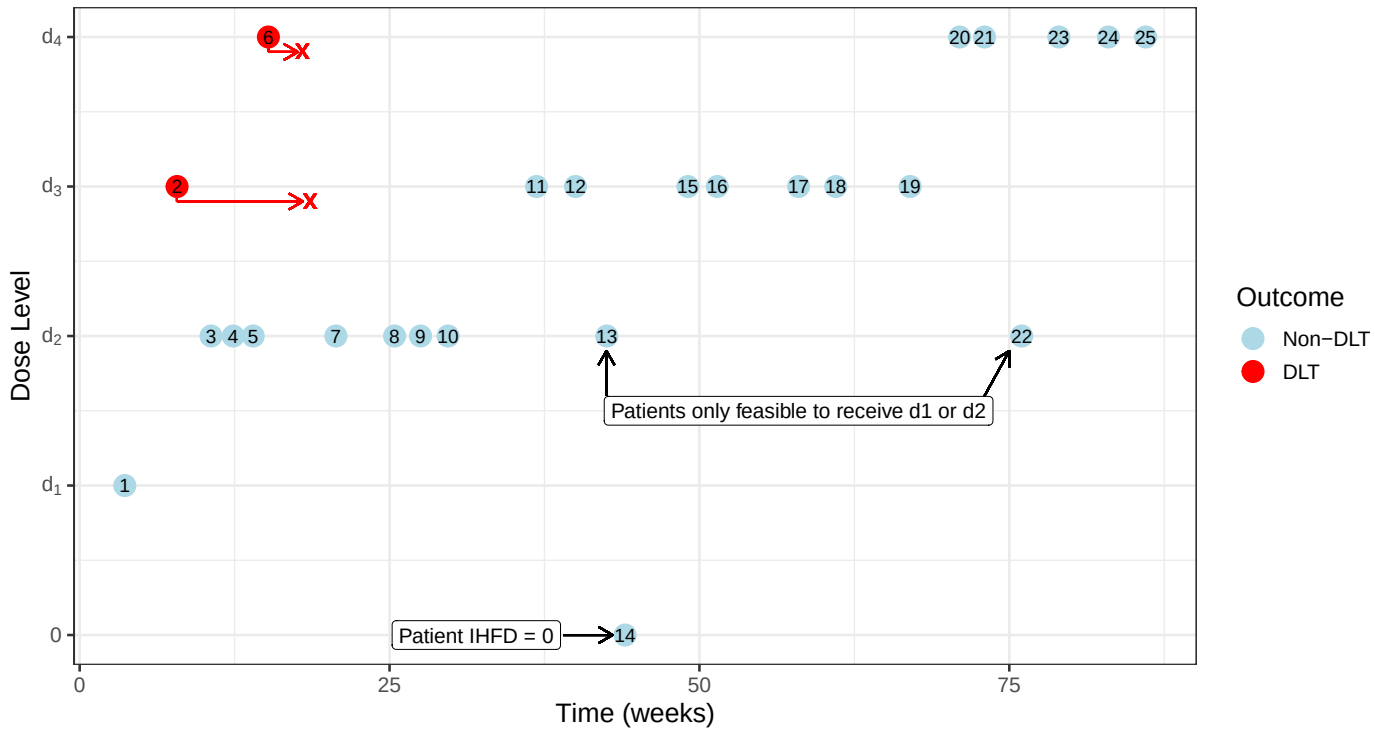


Figure 5.4: A simulated trial to demonstrate the proposed methodology

that under. For the feasibility model, we again used the non-informative prior $a = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, noting that we have three entries in this vector because we have 3 levels of feasibility: no dose levels, dose levels d_1 and d_2, or all dose levels.

The simulated trial begins with order $m = 1$ $(d_1 \rightarrow d_2 \rightarrow d_3 \rightarrow d_4)$ chosen at random from the two orders. Patient 1 is treated at dose level 1, since we still have the restriction on beginning the trial at this dose. By the time patient 2 enrolls in the trial, we still have no toxicity order information, so we select the order $m = 2$ $(d_1 \rightarrow d_3 \rightarrow d_2 \rightarrow d_4)$ at random, and the dose with estimated DLT probability closest to θ^* is dose level 3 so that is what patient 2 receives. When patient 3 enrolls in the trial patient 2 has not had their DLT yet, but the order with highest posterior probability is $m = 2$ and the estimated DLT probabilities for d_2 and d_3 are 0.30 and 0.15 respectively so patient 3 receives dose level 2. For the next two patients the order is $m = 2$ and they receive dose level 2. When patient 6 enrolls in the trial, dose level 4 has estimated DLT probability closest to the target so

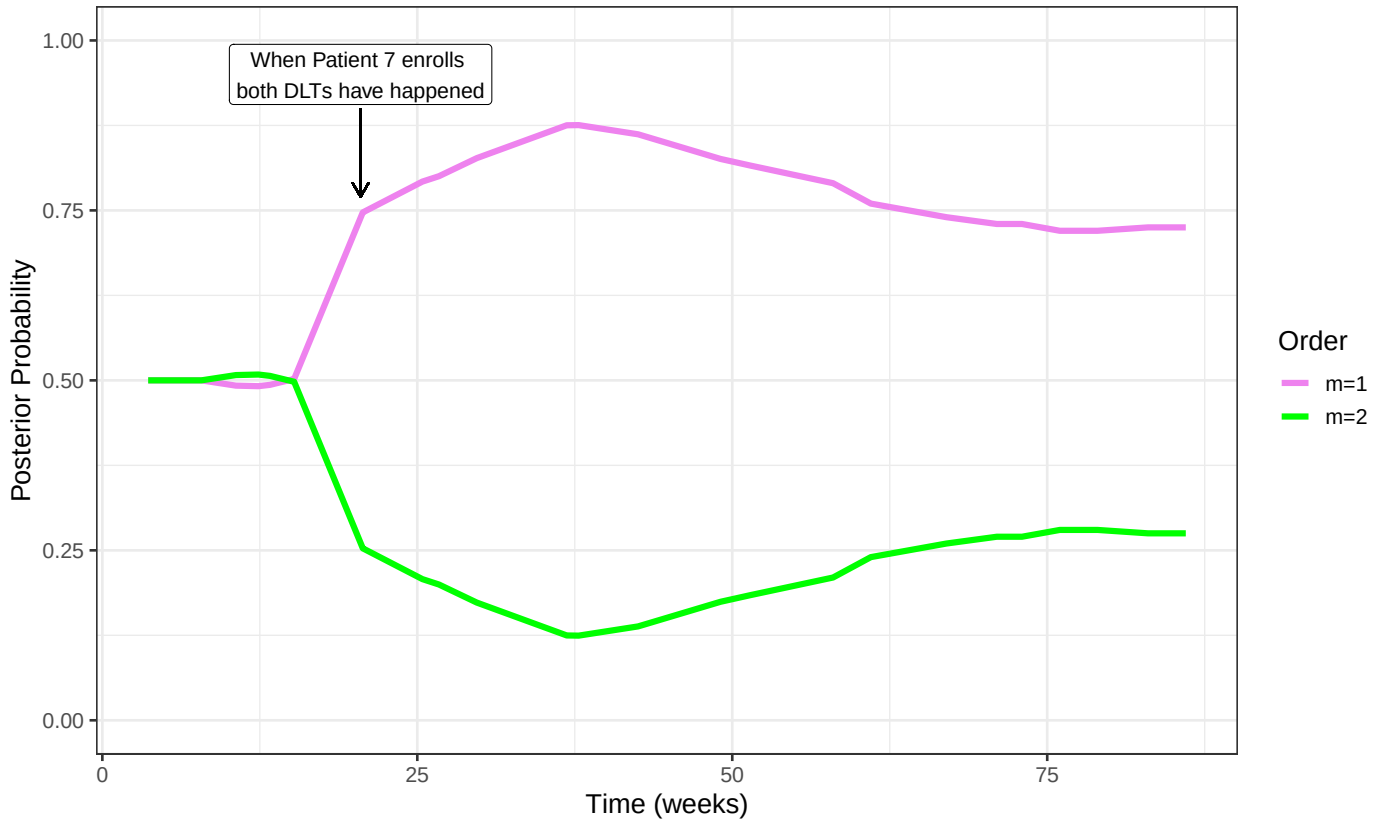
they receive that dose. By the time patient 7 enrolls in the trial, both patient 2 and patient 6 have had their DLTs and the posterior probability for each order are 0.75 and 0.25 for $m = 1$ and $m = 2$ respectively. Thus, for the rest of the trial, since there are no more DLTs, $m = 1$ has the highest posterior probability. Patients 7-10 receive dose level 2, but by the time patient 11 enrolls the estimated DLT probability for dose 3 is closest to the target DLT rate. Patients 13 and 14 are assigned to receive dose level 3, but they do not have enough cells expanded to be treated at this dose level. Patient 13 receives their highest feasible dose level (dose 2), and patient 14 is excluded from the trial as they did not have enough cells expanded to reach the lowest dose level. The trial proceeds with patients 15-19 allocated to dose level 3 since that is the estimated MTD. Patients 20-25 are assigned to receive dose level 4, but patient 22 receives dose level 2 due to feasibility issues. At the conclusion of the trial, dose 4 is selected as the FMTDC under order $m = 1$ since it has an estimated DLT probability that is closest to the target DLT rate, and it is feasible for a large part of the patient population.

A visualization of the posterior probabilities for each of the orders over the course of the trial are given in Figure 5.5. This representation of how we select the orders show the changes in posterior probability when a new patient enrolls, since that that is when we need to determine which order to treat that patient under. Since there are several DLTs towards the beginning of the trial that is when the largest changes in posterior probability occur.

5.7 Simulation Results

To evaluate our proposed methodology we conducted Monte-Carlo simulations across six pairs of randomly generated toxicity and feasibility curves. These simulation scenarios are are similar to the ones considered in the previous aim. A key difference is that the

Figure 5.5: Posterior probability of the two orders



feasibility curves have fewer levels since there are only 2 levels of dose-feasibility in this context. Recall that the cell dose levels for the motivating trial are 40×10^9 cells for d_1 and d_2, and 80×10^9 for d_3 and d_4. The toxicity curves are identical to the scenarios considered in the previous aim. Under each of the 6 scenarios we simulated 2000 trials with the settings given for the motivating trial. In addition to these results we conducted simulations under many other settings whose results are given in Appendix B. For the simulations given in Table 5.5 the FMTD distributions are promising. For scenarios 1 and 2 the feasibility probability for d_3 and d_4 is much lower than the minimum feasibility target of $\phi^* = 0.80$. This results in none of the trials selecting either of these dose levels as the FMTD. The DLT model does a good job at selecting the dose with DLT probability closest to the target DLT rate for both of these scenarios. For scenarios 3 and 4 all of the dose levels are feasible and in the case of scenario 3, the design successfully selects the true

FMTD 44.5% of the time. The model likely selects dose level d_2 as the FMTD 34.8% of the time due to incorrect final order recommendations. The performance of the method decreases in scenario 4, which is likely due to the closeness of the DLT probability of d_3 and d_4. Since both of these doses have DLT probability almost equidistant from the target DLT rate of $\theta^* = 0.25$ many of the trials incorrectly select d_3 as the FMTD. This is not the worst mistake as it is still pretty close to the target. The method described in aim 1 also had trouble with the DLT probabilities in scenario 4. Scenarios 5 and 6 have either too high DLT probability or too low feasibility probability for any dose to be the FMTD, so the correct choice is to stop the trial early, and not to choose an FMTD. The stopping rules perform reasonably well in scenario 5 and very well in scenario 6. The duration of the trial took approximately twice as long as the trials from aim 1, but this is not surprising since the expected accrual rate was about half of what it was in aim 1. Despite this longer duration, the trial would still progress much faster than a trial that cannot utilize both partially and full observed safety data.

Table 5.5: FMTD distributions for drug combinations ($m = 1$ is correct)

Scenario	Dose		p_t		p_f	FMTD %		None
1	d_1	d_2	0.22	0.32	0.89	**58.5**	36.8	4.7
	d_3	d_4	0.38	0.46	0.37	0.0	0.0	
2	d_1	d_2	0.18	0.24	0.87	34.8	**60.4**	4.8
	d_3	d_4	0.36	0.40	0.36	0.0	0.0	
3	d_1	d_2	0.10	0.15	0.96	3.0	34.8	0.1
	d_3	d_4	0.25	0.37	0.91	**44.5**	17.6	
4	d_1	d_2	0.11	0.15	0.97	2.5	33.5	0.1
	d_3	d_4	0.19	0.29	0.84	34.9	**29.0**	
5	d_1	d_2	0.50	0.58	0.96	32.2	0.8	**66.7**
	d_3	d_4	0.64	0.78	0.88	0.3	0.0	
6	d_1	d_2	0.31	0.37	0.60	6.9	1.5	**91.6**
	d_3	d_4	0.63	0.88	0.23	0.0	0.0	

A noticeable feature of the scenarios described in Table 5.5 is that the simple order of DLT probabilities correspond to $m = 1$ ($d_1 \rightarrow d_2 \rightarrow d_3 \rightarrow d_4$). While the method does not give preference to either order it is useful to consider what could happen when $m = 2$ ($d_1 \rightarrow d_2 \rightarrow d_3 \rightarrow d_4$) is the true simple order. Ideally, the method would choose simple order $m = 2$ more often, which it does, but the feasibility probabilities do not change. Simulation results for the case where $m = 2$ are given in Table 5.6. Interestingly, swapping the order actually helps FMTD selection in scenarios 1 and 2. In scenario 1, d_3 is not feasible so it is almost never chosen as the FMTD whereas d_2 is overly toxic, so 70.7% of the trials have d_1 as the FMTD. In scenario 2 the dose with DLT probability closest to the target is d_3, but again this dose is not feasible so 65.6% of the trials have d_1 as the correctly selected FMTD. Scenario 3 shows how the method switches order to select the select the FMTD. Since all dose levels are feasible we can see how the model selects d_2 more often under order $m = 2$ than $m = 1$. The reverse is true for d_3. Scenario 4 is still difficult for the method under order $m = 2$ since most of of the DLT probabilities are quite close. The FMTD distribution for scenarios 5 and 6 under $m = 2$ are about the same as they were under $m = 1$. This makes sense since the correct choice is to stop the trial and not select an FMTD.

Table 5.6: FMTD distributions for drug combinations ($m = 2$ is correct)

Scenario	Dose		p_t		p_f	FMTD %		None
1	d_1	d_2	0.22	0.38	0.89	**70.7**	23.7	5.7
	d_3	d_4	0.32	0.46	0.37	0.1	0.0	
2	d_1	d_2	0.18	0.36	0.87	**65.6**	29.0	5.6
	d_3	d_4	0.24	0.40	0.36	0.1	0.0	
3	d_1	d_2	0.10	0.25	0.96	3.7	**44.1**	0.1
	d_3	d_4	0.15	0.37	0.91	36.5	15.7	
4	d_1	d_2	0.11	0.19	0.97	2.7	39.5	0.1
	d_3	d_4	0.15	0.29	0.84	29.6	**28.3**	
5	d_1	d_2	0.50	0.64	0.96	32.0	0.6	**66.9**
	d_3	d_4	0.58	0.78	0.88	0.6	0.0	
6	d_1	d_2	0.31	0.63	0.60	6.8	0.1	**93.2**
	d_3	d_4	0.37	0.88	0.23	0.0	0.0	

Chapter 6

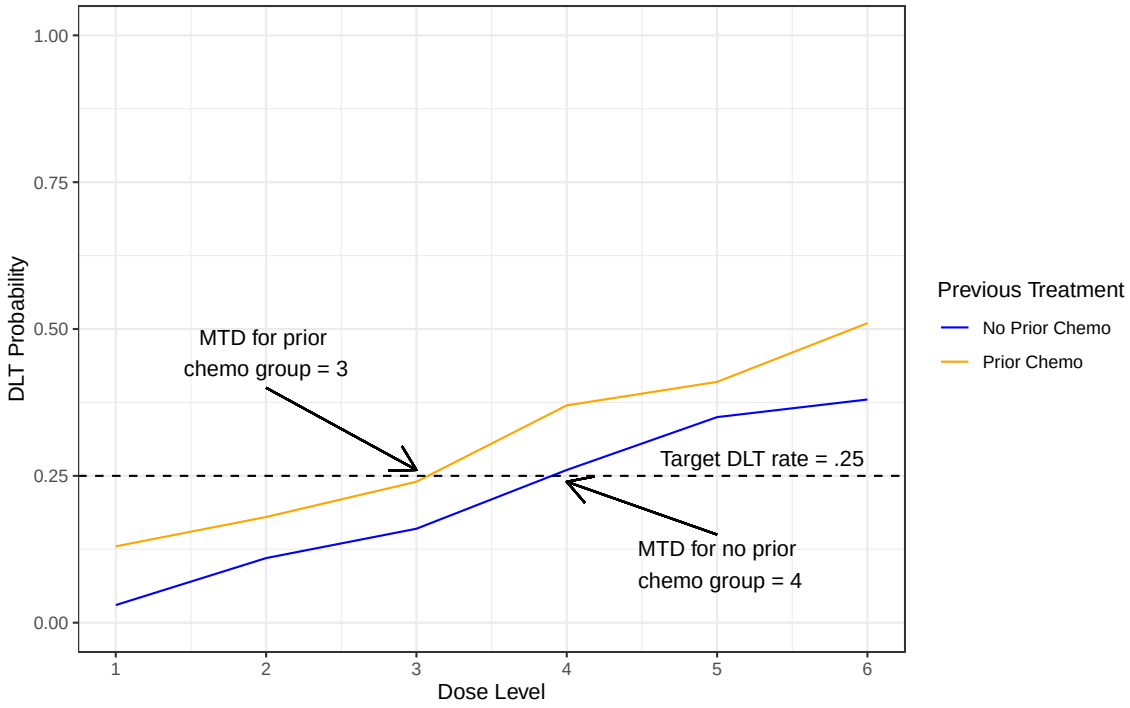
Aim 3: Patient heterogeneity and dose-feasibility

6.1 Background on dose finding with heterogeneous patient groups

Phase I cancer trials usually have inclusion criteria which limits the degree of patient heterogeneity that could complicate future phases of drug development. However, there can be certain situations where there might be certain groups of patients included in a trial that should be considered in the dose escalation process. For example, consider a trial where half of the patients have received prior chemotherapy treatment and half of them have not. It could be suspected that patients that have already received prior chemotherapy may be at a higher risk of having a DLT than patients who have not. Figure 6.1 shows how group specific DLT probabilities could result in group specific MTDs.

In this example, the prior chemotherapy group's MTD is dose level 3 whereas the group without prior chemotherapy has an MTD at dose level 4. In general, we would like to

Figure 6.1: Example Group Specific DLT Scenarios



be able to consider situations where group specific MTDs could be one or more doses away from each other. A common approach to handling patient heterogeneity is to run concurrent trials under the umbrella of a single trial. While that option is straightforward, it is often inefficient. Sharing data between the groups, if done in an appropriate manner, can lead to improvements in trial efficiency, trial accuracy, and patient safety.

6.2 Motivating trial 3: Dose finding for a cell therapy with 2 patient groups

The final motivating clinical trial for this dissertation is a phase I study for the treatment of acute lymphoblastic leukemia with Anti-CD19 CAR T cells [8]. This trial, unlike the other motivating examples, has already been completed and would have benefited greatly if there was methodology that accounted for the peculiarities of the study. The goal of

the trial, similar to previous examples, was to determine the FMTD of the treatment under study. The key difference in this trial is that patients who enrolled fell into one of two strata: prior allogenic stem cell transplant (SCT) or patients who have not received prior SCT. These strata are relevant from a clinical standpoint because it is believed that patients who have received prior SCT are more frail, and at more risk of DLT outcomes than patients who have not received prior SCT. The dose level definitions for this trial are given in Table 6.1.

Table 6.1: Dose Level Definitions

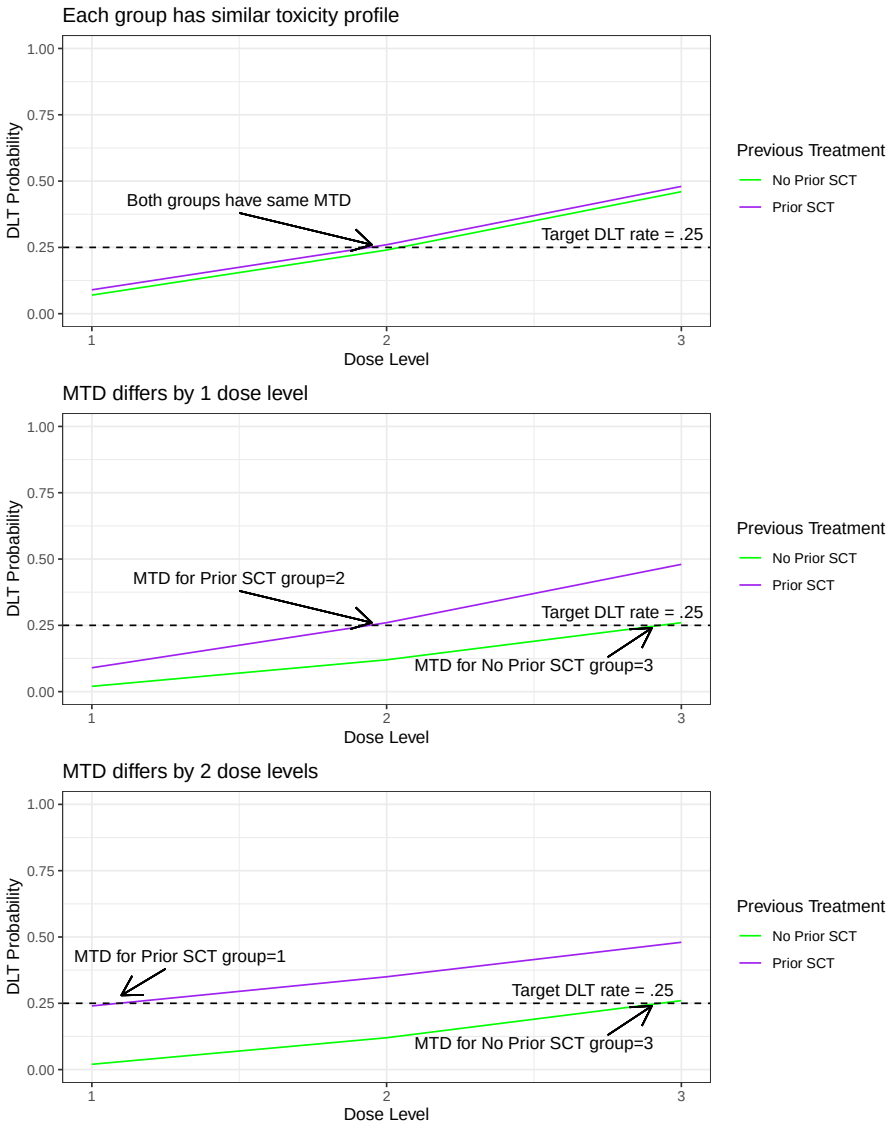
Dose Level	Number of T Cells/kg
1	1×10^6
2	3×10^6
3	1×10^7

Unlike the previous two motivating trials, Nivolumab was not part of the treatment regimen, so if there is a treatment-related DLT outcome it would be expected to occur soon after treatment. Since we do not expect late-onset toxicities we will not need to use partial DLT data to inform dosing decisions as patients are accrued to the trial. However, there are still feasibility considerations due to the cell therapy. In executing this trial, "Patients whose CAR T-cell product did not meet the dose to which they were assigned did not inform dose escalation but were assessed for toxicity and all other parts of the study [8]." This decision to exclude patient data in dose-escalation was likely due to the fact that the trial used the 3+3 to guide dose-escalation rendering it unable to utilize data observed at unplanned dose levels.

Since this trial has two patient groups to be evaluated across three dose levels we could expect the difference in MTD to vary by 0, 1 or 2 dose levels. An illustration of these three possibilities is given in Figure 6.2.

In the first scenario, both groups have similar DLT profiles so their MTDs are both dose level 2. In the second scenario, the group with prior SCT has an MTD of dose level 2 while

Figure 6.2: Three scenarios for DLT probability



the group without prior SCT has an MTD of dose level 3. In the third scenario, the group with prior SCT has an MTD of dose level 1, whereas the group without prior SCT has an MTD of dose level 3. While we are not asserting that any of these are true, we think it is important to use methodology that can account for a variety of situations. Clearly, treating the two groups of patients as a single homogeneous group could be inappropriate in situations where the DLT profiles are different, but at the same time having two concurrent trials could be detrimental to success if the DLT profiles are similar. Conducting concurrent trials can also result in an outcome where the group at more risk has a larger MTD

than the group at less risk. This type of outcome is also known as a reversal. Reversals have been studied [39], and can be controversial since they violate prior knowledge about underlying disease and the effects of previous treatments. Despite being common, reversals and other issues in phase I trials with patient groups can be resolved using statistical methodology that account for these phenomena.

Many statistical methods tailored to utilizing all possible data from multiple patient groups to inform trial progression have been developed [40–46]. These methods can advance phase I trials from both an efficiency standpoint, and an accuracy standpoint. Despite the availability of these statistical methods, they are not widely implemented. Furthermore, none of them account for the type of dose-feasibility issues that occur in cell therapy trials. Our methodology develops the first phase I design that can incorporate patient heterogeneity while accounting for dose-feasibility issues that could arise due to insufficient cell growth.

6.3 Proposed model for toxicity

As we have discussed in previous sections, a big advantage of developing our models within a CRM framework is that we have a great deal of flexibility. Further, many designs that account for patient heterogeneity in identifying group-specific MTDs were developed within this framework. Thus, our models and methodology will have a structure that is similar to the methodology developed in previous chapters. The two primary differences for this trial are that we have distinct patient groups and there is no late-onset toxicity to consider.

We propose a DLT probability model that adaptively assigns patients to one of a set of pre-specified dose levels where each patient belongs to one of two groups that could have

a certain order with respect to toxicity. To set this up, consider the dose level definitions given in Table 6.2.

Group	Cell Dose		
	1×10^6	3×10^6	1×10^7
1 - No Prior SCT	d_{11}	d_{12}	d_{13}
2 - Prior SCT	d_{21}	d_{22}	d_{23}

Table 6.2: Dose levels by patient group

These dose definitions have labels that identify the patient group and the dose they receive. For example, treatment at d_{12} means a patient from group 1 received dose level 2 whereas treatment at d_{22} means a patient from group 2 received dose level 2. While both of these patients received the same number of expanded CAR T cells their individual contributions to the likelihood function may not be equal even if they have the same DLT outcome.

To model and differentiate membership of a particular group, we consider a model selection approach similar to how we selected an order in our proposed design for drug combinations. If we consider the clinical belief that the group with prior SCT will be the same or more prone to a DLT outcome than the group without prior SCT, then we can consider several models that represent each possible shift in the MTD. From the CRM approach this can be accomplished with associating 2 skeletons, one for each group, with each dose level change in the MTD. These skeleton values/model choices are given in Table 6.3.

Model	Group	Cell Dose		
		1×10^6	3×10^6	1×10^7
$m = 1$	$g = 1$	0.06	0.13	**0.25**
	$g = 2$	0.13	**0.25**	0.41
$m = 2$	$g = 1$	0.06	0.13	**0.25**
	$g = 2$	**0.25**	0.41	0.59
$m = 3$	$g = 1$	0.13	**0.25**	0.41
	$g = 2$	0.13	**0.25**	0.41

Table 6.3: Skeleton values for each model choice. The target DLT rate is in bold.

It is important that each possible clinical scenario are covered by the models we choose from. Model choice $m = 1$ covers the situation where the MTD differs by one dose, $m = 2$ covers the situation that the MTD differs by two doses, and $m = 3$ covers the situation where the MTD in both groups is the same.

To model toxicity we propose a single parameter model

$$p_{j,m}(\beta, q_{j,m,g}) = \Phi\{\beta + \Phi^{-1}(q_{j,m,g})\} \qquad \text{for} \qquad j = 1, \ldots, J \tag{6.1}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function, $\Phi^{-1}(\cdot)$ is the standard normal quantile function, β is the parameter, and $q_{j,m}$ are the skeleton values corresponding to model m and group g. For this trial, a patient's DLT status is no longer a function of time due to the absence of late-onset toxicities. Thus, we denote the i^{th} patients DLT status as

$$y_{i,g} = \begin{cases} 1 & \text{if patient } i \text{ has a DLT} \\ 0 & \text{if patient } i \text{ dose not have a DLT} \end{cases} \qquad \text{for} \qquad i = 1, \ldots, n \tag{6.2}$$

where g is defined as

$$g = \begin{cases} 1 & \text{if patient has not had prior SCT.} \\ 2 & \text{if patient has had prior SCT.} \end{cases} \tag{6.3}$$

Under this formulation the likelihood function under a particular model selection m is

$$\mathcal{L}_m(\beta|y_{i,g}) = \prod_{i=1}^{n} \left(\Phi\{\beta + \Phi^{-1}(q_{j,m,g,i})\}\right)^{y_{i,g}} \left(1 - \Phi\{\beta + \Phi^{-1}(q_{j,m,g,i})\}\right)^{1-y_{i,g}}. \tag{6.4}$$

This likelihood function is similar to the likelihood function for drug combinations (5.3), with respect to the fact that it is evaluated under a particular model. The model choice for drug combinations is simple order, whereas the model choice for patient groups corresponds to a level shift in the MTD. This allows the likelihood function to incorporate data from multiple group specific skeletons.

To select a model we apply a similar approach to the method given in the drug combination section. Under each model, $m = 1, \ldots, M$, we compute

$$\omega(m|\text{data}) = \frac{\omega(m) \int \mathcal{L}_m(\beta|y_{i,g}) f(\beta) d\beta}{\sum_{m=1}^{M} \omega(m) \int \mathcal{L}_m(\beta|y_{i,g}) f(\beta) d\beta} \tag{6.5}$$

where $\omega(m)$ is the prior probability that model m is the correct model, $f(\beta)$ is the prior distribution for β, and $\mathcal{L}_m(\cdot)$ is the likelihood function evaluated under model m. This effectively weights the particular models based on the accumulating safety data. For the models given in Table 6.3, if we start to see data suggesting that the group MTDs are one dose level apart the weight for $m = 1$ will increase, if we start to see evidence that the group MTDs are two dose levels apart the weight for $m = 2$ will increase, and if we see evidence that the MTDs are the same dose level the weight for $m = 3$ will increase. When a patient enrolls in the trial they will be assigned a dose level under which ever model has the highest value of (6.5). Estimated DLT probabilities for groups 1 and 2 under model m by the posterior means

$$\hat{p}_{j,1} = \int \frac{\Phi\{\beta + \Phi^{-1}(q_{j,m,1})\} \mathcal{L}_m(\beta|y_{i,g}) f(\beta)}{\int \mathcal{L}_m(\beta|y_{i,g}) f(\beta) d\beta} d\beta \qquad \text{for} \qquad j = 1, \ldots, J \tag{6.6}$$

and

$$\hat{p}_{j,2} = \int \frac{\Phi\{\beta + \Phi^{-1}(q_{j,m,2})\}\mathcal{L}_m(\beta|y_{i,g})f(\beta)}{\int \mathcal{L}_m(\beta|y_{i,g})f(\beta)d\beta} d\beta \qquad \text{for} \qquad j = 1, \ldots, J \tag{6.7}$$

where $f(\beta)$ is the prior distribution on β. The patient is assigned to the current working estimate of the MTD defined as

$$j_1^* = \underset{j \in \{1,\ldots,J\}}{\operatorname{argmin}} |\hat{p}_{j,1} - \theta^*| \tag{6.8}$$

if the patient is a member of group 1 and

$$j_2^* = \underset{j \in \{1,\ldots,J\}}{\operatorname{argmin}} |\hat{p}_{j,2} - \theta^*| \tag{6.9}$$

if the patient is a member of group 2. The value θ^* is the target DLT rate for both groups. If a patient cannot receive the recommended dose the patient is treated at their IHFD. At the conclusion of the trial the final group specific MTD estimates are j_1^* and j_2^*.

6.4 Proposed model for feasibility

Modelling feasibility for this aim is very similar to the first model we proposed for feasibility, except that we need to record the feasibility data for each group separately. Unlike the known toxicity ordering of each group, there is not much information on how prior SCT could affect dose-feasibility. For this reason, we will implement the same model for each group, but keep them separate.

As patients enroll in the trial, their IHFD is recorded as a count a group specific multinomial vector. Patients from group 1 and group 2 will be denoted with those respective

subscripts. The vectors are

$$\mathbf{X}_1 = (X_{10}, X_{11}, \ldots, X_{1J}) \sim \text{Multinomial}(\pi_{10}, \pi_{11}, \ldots, \pi_{1J}) \tag{6.10}$$

and

$$\mathbf{X}_2 = (X_{20}, X_{21}, \ldots, X_{2J}) \sim \text{Multinomial}(\pi_{20}, \pi_{21}, \ldots, \pi_{2J}) \tag{6.11}$$

where $\sum_{j=10}^{1J} \pi_{1j} = 1$ and $\sum_{j=20}^{2J} \pi_{2j} = 1$. We can let $\boldsymbol{\pi}_1 \sim$ Dirichlet$(\mathbf{a}_1)$ and $\boldsymbol{\pi}_2 \sim$ Dirichlet$(\mathbf{a}_2)$ where $\mathbf{a}_1 = (a_{10}, a_{11}, \ldots, a_{1J})$ and $\mathbf{a}_2 = (a_{20}, a_{21}, \ldots, a_{2J})$. The respective posterior distributions are

$$\boldsymbol{\pi}_1 | \mathbf{X}_1 \sim \text{Dirichlet}(\mathbf{a}_1 + \mathbf{X}_1) \tag{6.12}$$

and

$$\boldsymbol{\pi}_2 | \mathbf{X}_2 \sim \text{Dirichlet}(\mathbf{a}_2 + \mathbf{X}_2). \tag{6.13}$$

These posterior distributions are almost identical to those presented in Section 4 for the single population case. The prior parameter vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ could be adjusted if the clinician or investigator believes that one group is more likely to have higher feasibility than the other.

We are interested in which doses are globally feasible with respect to each group. To this end we propose the following definitions:

Definition: Dose level j is globally feasible for group 1 if

$$P(\phi_j < \phi^* | \mathbf{X}_1) < p_{u,f} \tag{6.14}$$

Definition: Dose level j is globally feasible for group 2 if

$$P(\phi_j < \phi^* | \mathbf{X}_2) < p_{u,f} \tag{6.15}$$

where ϕ^* is the minimum required feasibility probability, and $p_{u,f}$ is an upper probability cutoff. Each of the probabilities follows the beta posteriors

$$\phi_j | \mathbf{X}_1 \sim \text{beta}(\sum_{r=j}^{J} a_{1r} + X_{1r}, \sum_{\ell=0}^{j-1} a_{1\ell} + X_{1\ell}) \qquad \text{for} \qquad j = 1, \ldots, J \tag{6.16}$$

for group 1, and

$$\phi_j | \mathbf{X}_2 \sim \text{beta}(\sum_{r=j}^{J} a_{1r} + X_{2r}, \sum_{\ell=0}^{j-1} a_{2\ell} + X_{2\ell}) \qquad \text{for} \qquad j = 1, \ldots, J \tag{6.17}$$

for group 2. We can use these posterior distributions to establish a set of globally feasible dose levels

$$\mathcal{F}_1 = \{d_j : P(\phi_j < \phi^* | \mathbf{X}_1) < p_{u,f}\} \tag{6.18}$$

for group 1, and

$$\mathcal{F}_2 = \{d_j : P(\phi_j < \phi^* | \mathbf{X}_2) < p_{u,f}\} \tag{6.19}$$

for group 2. The maximum dose in the first set is the global highest feasible dose for group 1 (GHFD1) and the maximum dose in the second set is the global highest feasible dose for group 2 (GHFD2).

At the conclusion of the trial group specific feasible maximum tolerated doses can be calculated as

$$\text{FMTD}_1 = \min(j_1^*, GHFD1) \tag{6.20}$$

for group 1, and

$$\text{FMTD}_2 = \min(j_2^*, GHFD2) \tag{6.21}$$

for group 2. The values j_1^* and j_2^* are the group specific MTDs determined in (6.8) and (6.9) respectively.

6.5 Trial Conduct

Trial conduct

The previous sections have described the models that define trial conduct for a cell therapy trial with multiple patient groups. The following formally describes how the trial progresses:

1. Enroll a patient in the trial.
2. Determine the patient's IHFD. If IHFD$>$0, go to step 3. If IHFD=0, determine the GHFD for both groups. If the GHFD1=0 and GHFD2=0, stop the trial and set both group specific MTDs to 0. If either GHFD1$>$0 or GHFD2$>$0, go to step 1.

3. Treat the patient at group specific MTD estimates under the current model. If the patient is not feasible to receive this dose treat them at their IHFD. Observe their DLT outcome and update the current model to the one with largest posterior probability. Update both group specific estimated MTDs, if the safety stopping rule is triggered stop the trial and set both group specific FMTDs to 0. If the trial is not fully enrolled go to step 1. If the trial is fully enrolled go to step 4.

4. When the trial is fully enrolled with N patients evaluated for toxicity, or the maximum number of M patients have been evaluated for feasibility stop the trial. Complete patient follow-ups and determine the FMTD_1 in (6.20) and FMTD_2 as in (6.21).

To demonstrate how our method works, consider the simulated example trial (Figure 6.3) that shows how the trial progresses while adaptively sharing DLT data between the patient groups. The design settings are set to the specifications of the motivating trial. The target DLT rate is $\theta^* = 0.25$ and the minimum feasibility probability is $\phi^* = 0.80$. Dose selection and trial progression is determined by the trial conduct we have just described. The maximum number of patients to undergo treatment is $N = 24$, and the maximum number of patients to have their cells expanded is $M = 30$. Each patient will come from either the group that has not had prior SCT or the group that has. Note that a key difference from the previous two aims is that the expected DLTs are not late-onset. This means that no patient's dose assignment is being informed by partial DLT data because that data is immediately obtainable. We determined the prior specifications for the skeleton by the algorithm described by Lee and Cheung [27]. These values, given in Table 6.3, allow for the MTD to vary between the groups by 0, 1, or 2 dose levels. Since prior to the trial we believe that the group that has had prior SCT to be more prone to DLTs, we set the prior model probabilities as $\omega(m = 1) = \frac{3}{7}$, $\omega(m = 2) = \frac{3}{7}$, and $\omega(m = 3) = \frac{1}{7}$. These priors down-weight the model where prior DLT probability between the groups is equal $(m = 3)$, but at the same time allows for it to be an option if the DLT data between the groups turns

out to be similar. For the feasibility model, we assumed a non informative prior for both groups such that $\mathbf{a}_1 = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ and $\mathbf{a}_2 = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ where $\mathbf{a}_1$ corresponds to the group that has not had prior SCT and $\mathbf{a}_2$ corresponds to the group that has had prior SCT.

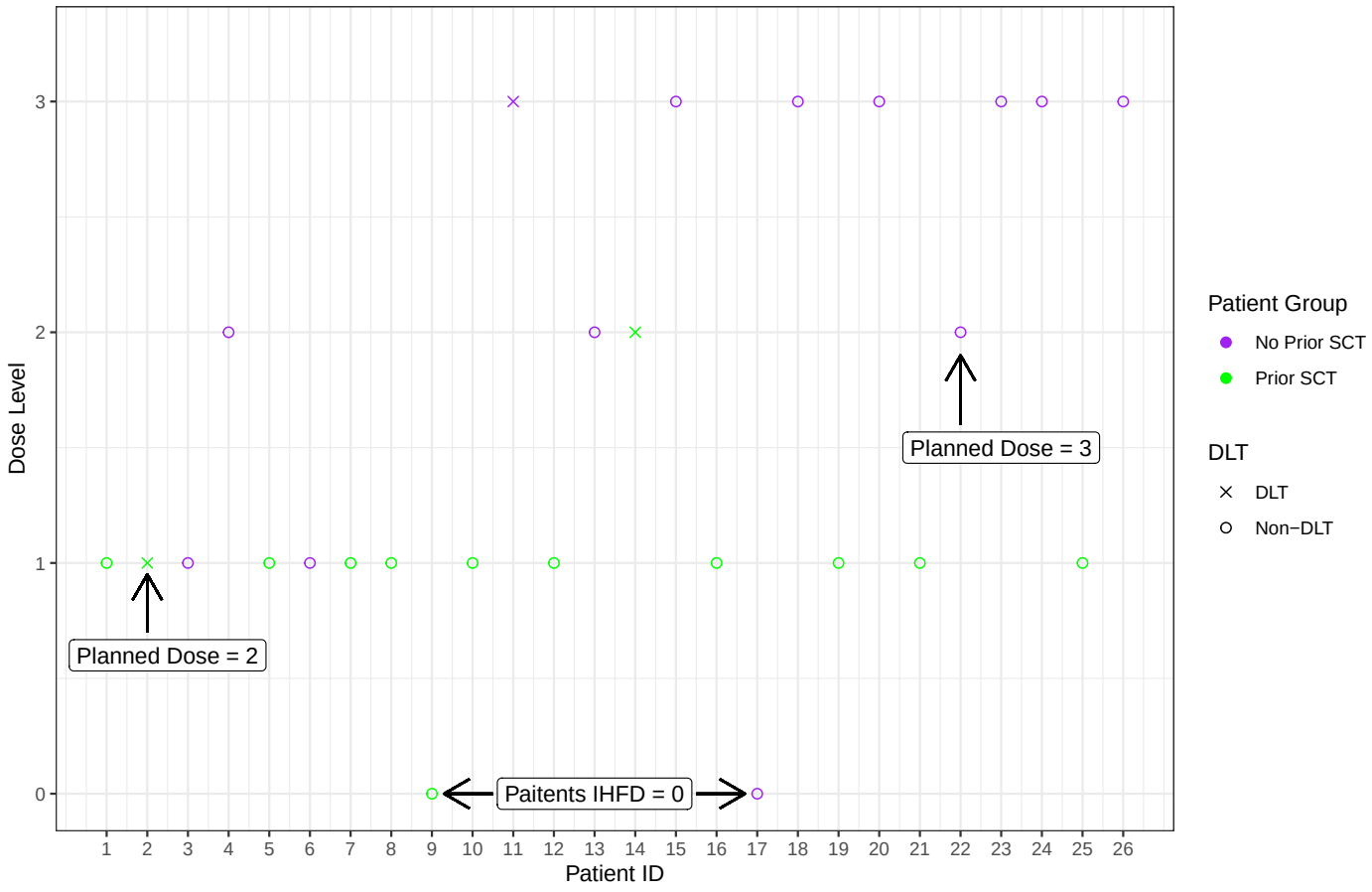


Figure 6.3: A simulated trial using the proposed methodology

To begin the trial, model $m = 2$ is randomly selected and the first patient, a member of group $g = 2$, is treated at dose level 1 and does not have a DLT. This result increases the posterior probability of model $m = 1$ and the trial proceeds under $m = 1$. The recommended dose for patient 2 is dose level 3, but their IHFD is dose level 1 which they receive and have a DLT. Now, the posterior model probabilities favor $m = 2$ and the trial proceeds under $m = 2$. The next two patients, both members of group $g = 1$, receive dose levels 1 and 2 respectively, followed by a non-DLT from a member of group $g = 2$. At this point in the trial, after 5 patients have been accrued, the target dose level for groups 1 and 2

are dose levels 1 and 3 respectively. Patient 6 is meant to receive this dose, but is only feasible to receive dose level 1. Patients 7-10, all members of group $g = 1$, receive dose level 1 except for patient 9 who is not able to receive any dose level due to insufficient cell expansion, so they are excluded from the trial. At this point, the posterior probability for model $m = 1$ is larger than $m = 2$ so the trial proceeds under $m = 1$. Patients 11-14 are treated under model $m = 1$, but the 2 DLTs result in model $m = 2$ having a higher posterior probability. For the rest of the trial, patients from group 1 and 2 are intended to receive dose levels 1 and 3 respectively. Patient 17 is excluded from the trial due to insufficient cell expansion, and patient 22 is treated at dose level 2 due to not being feasible to receive dose level 3. Both groups have high feasibility and the target MTDs at the conclusion of the trial from groups 1 and 2 are dose levels 3 and 1 respectively. So, the final estimates for the FMTDs are $\text{FMTD}_1 = 3$ and $\text{FMTD}_2 = 1$. The final choice of the model was $m = 2$ since it had the highest posterior probability.

6.6 Simulation Results

We evaluated our method through Monte-Carlo simulation across 12 combinations of dose-toxicity and dose-feasibility curves. Just like in the previous aims, we generated the feasibility curves using the algorithm given by Bagley and Wages [16] and we generated the toxicity curves using the algorithm proposed by Conaway and Petroni [1]. For each of the scenarios, we simulated 2000 trials using the settings for the motivating trial. The generated curves and FMTD distribution for each of the scenarios are given in Table 6.4.

The designs displayed in Table 6.4 are the method that we have just proposed and a version of our method that runs two concurrent separate trials for each group in parallel. The parallel implementation of our design is included because running two separate concurrent trials for each group is how many real cancer trials address the problem of

patient group heterogeneity. While we are not advocating for a parallel implementation of our design, it is useful as a comparison to our proposed design. For Scenarios 1-4 every dose level is feasible for each patient to receive so the FMTD is the same as the MTD determined solely by the toxicity curve. For scenarios 1,2 and 3 the proposed method correctly selects the FMTD in both groups a majority of the time. In general, the method that shares data performs better than the method that dose not. Scenario 2 has the parallel method performing slightly better, but the difference is not large. Scenario 4 represents the situation where both groups have identical feasibility and toxicity profiles. Overall, the method has poorer FMTD selection likely due to the down-weighting of $m = 3$ since that is the model that doesn't match our prior assumptions about the groups. We could adjust this to make the method more accurate under this scenario, but we have left it to show that there is still a certain degree of robustness in the event that model $m = 3$ is the correct selection. Scenarios 5-8 and 9-12 have the same toxicity profiles as scenarios 1-4, but the GHFD's are 2 and 1 respectively. The performance of the method is not hindered by the feasibility curve for scenarios 5-8. For scenarios 8-12 the FMTD accuracy is lowered in some cases where the model mistakes dose level 2 as being feasible. The feasibility probability at dose 1 and 2 are both somewhat close to the target minimum feasibility rate of $\theta^* = 0.80$ which can hurt the feasibility model's accuracy. The DLT model is not hindered by this, and MTD estimates are more accurate in these scenarios. This type of information could be assessed in subsequent phases of trials, and feasibility could be explored further in that setting.

Another important consideration in evaluating designs proposed for trials involving heterogeneous groups of patients is evaluating the possibility of reversals. Recall that a reversal is when the estimated FMTD/MTD violates the known toxicity order of the groups. For example, if the estimated FMTD for group 1 is dose 1 while the estimated FMTD for group 2 is dose 3, a reversal has occurred since the more sensitive group has a more toxic recommended dose. This does not mean our clinical assumptions about the toxi-

city order of each group are invalid, but that we have implemented a method that does not share data. The percentage of trials for each scenario that resulted in a reversal are given in Table 6.5. Notably, no trial guided by our proposed design results in a reversal. This is primarily because DLT data is being shared across groups via the likelihood function. However, this is a problem with the parallel trial implementation of our method. The prevalence of reversals range from 1.3% to 30.4% across all scenarios. Despite some of these percentages not being extremely high, the comparison with our method where reversals never happen should give investigators pause when considering methods that do not shared information between patient groups.

Table 6.4: FMTD Distribution for 12 scenarios. The FMTD is in bold.

Scenario		No Prior SCT ($g=1$)			Prior SCT ($g=2$)			
	Dose	1	2	3	1	2	3	None
	p_t	0.15	**0.26**	0.43	**0.25**	0.43	0.68	
1	p_f	0.98	**0.93**	0.84	**0.98**	0.93	0.84	
Proposed	FMTD%	19.0	**51.0**	26.4	**82.3**	13.8	0.1	3.6
Parallel	FMTD%	26.7	**41.3**	25.9	**73.0**	20.6	1.1	5.2
	p_t	0.07	0.11	**0.24**	**0.25**	0.43	0.68	
2	p_f	0.98	0.93	**0.84**	**0.98**	0.93	0.84	
Proposed	FMTD%	2.5	26.5	**70.6**	**73.4**	24.8	1.4	0.4
Parallel	FMTD%	3.9	21.6	**72.8**	**74.7**	22.0	1.6	1.7
	p_t	0.07	0.11	**0.24**	0.15	**0.26**	0.43	
3	p_f	0.98	0.93	**0.84**	0.98	**0.93**	0.84	
Proposed	FMTD%	0.9	16.2	**82.7**	34.3	**53.0**	12.5	0.2
Parallel	FMTD%	4.2	21.4	**73.8**	35.9	**44.2**	19.4	0.5
	p_t	0.15	**0.26**	0.43	0.15	**0.26**	0.43	
4	p_f	0.98	**0.93**	0.84	0.98	**0.93**	0.84	
Proposed	FMTD%	11.8	**47.6**	39.2	57.6	**36.9**	4.2	1.4
Parallel	FMTD%	31.1	**42.5**	24.8	33.6	**44.5**	20.4	1.6
	p_t	0.15	**0.26**	0.43	**0.25**	0.43	0.68	
5	p_f	0.94	**0.85**	0.60	**0.94**	0.85	0.60	
Proposed	FMTD%	20.0	**69.2**	8.0	**84.8**	12.3	0.0	2.9
Parallel	FMTD%	28.8	**57.8**	8.3	**72.8**	21.5	0.6	5.1
	p_t	0.07	**0.11**	0.24	**0.25**	0.43	0.68	
6	p_f	0.94	**0.85**	0.60	**0.94**	0.85	0.60	
Proposed	FMTD%	3.5	**73.4**	22.4	**74.3**	24.6	0.4	0.7
Parallel	FMTD%	5.2	**71.2**	22.0	**74.3**	23.4	0.7	1.6
	p_t	0.07	**0.11**	0.24	0.15	**0.26**	0.43	
7	p_f	0.94	**0.85**	0.60	0.94	**0.85**	0.60	
Proposed	FMTD%	1.4	**72.0**	26.2	35.1	**60.8**	3.9	0.2
Parallel	FMTD%	4.7	**72.5**	22.2	36.8	**57.7**	5.0	0.6
	p_t	0.15	**0.26**	0.43	0.15	**0.26**	0.43	
8	p_f	0.94	**0.85**	0.60	0.94	**0.85**	0.60	
Proposed	FMTD%	11.4	**73.8**	12.9	57.6	**39.1**	1.4	1.9
Parallel	FMTD%	30.9	**58.9**	8.1	35.2	**57.0**	5.7	2.2
	p_t	**0.15**	0.26	0.43	**0.25**	0.43	0.68	
9	p_f	**0.90**	0.66	0.41	**0.90**	0.66	0.41	
Proposed	FMTD%	**60.0**	35.2	0.1	**89.0**	6.2	0.0	4.7
Parallel	FMTD%	**66.2**	27.2	0.0	**82.4**	10.9	0.0	6.6
	p_t	**0.07**	0.11	0.24	**0.25**	0.43	0.68	
10	p_f	**0.90**	0.66	0.41	**0.90**	0.66	0.41	
Proposed	FMTD%	**55.1**	43.0	0.3	**85.3**	13.1	0.0	1.6
Parallel	FMTD%	**55.6**	41.5	0.3	**86.3**	13.2	0.0	2.5
	p_t	**0.07**	0.11	0.24	**0.15**	0.26	0.43	
11	p_f	**0.90**	0.66	0.41	**0.90**	0.66	0.41	
Proposed	FMTD%	**55.2**	42.5	0.2	**70.9**	27.0	0.0	2.1
Parallel	FMTD%	**57.3**	40.2	0.4	**71.2**	26.7	0.1	2.0
	p_t	**0.15**	0.26	0.43	**0.15**	0.26	0.43	
12	p_f	**0.90**	0.66	**0.41**	0.90	0.66	0.41	
Proposed	FMTD%	**58.5**	38.2	0.2	**77.9**	19.1	0.0	3.0
Parallel	FMTD%	**66.8**	29.3	0.0	**69.3**	26.9	0.0	3.9

Table 6.5: Percentage of trials with Reversals

Scenario	Proposed Method	Parallel Trials
1	0.0	7.4
2	0.0	1.3
3	0.0	7.0
4	0.0	30.4
5	0.0	7.0
6	0.0	1.8
7	0.0	6.4
8	0.0	27.6
9	0.0	8.0
10	0.0	2.1
11	0.0	8.5
12	0.0	28.7

Chapter 7

Concluding remarks

This dissertation proposed 3 novel designs to guide phase I cell therapy cancer clinical trials. The goal for each trial was to develop the statistical methodology needed to efficiently accrue patients and identify a safe and feasible dose level to carry forward to later stages of development. While we advocate for our designs and hope they can be used in a real world setting, methodology utilizing time-to-event framework is not without its drawbacks when it comes to implementation. Sharon et al. [47] described some limitations associated with designing and executing the TITE-CRM framework, many of which also apply to our designs from aims 1 and 2. Planning a trial using the TITE-CRM can be time-consuming relative to simple algorithmic rules, especially for a clinical trials unit trying to do so for the first time. Additional time and funding are needed to support the statistical team's effort in the planning phase. Throughout the study, strong communication between clinical, data management, and statistical units, as well as regular meetings with a safety review committee, may be required. Even though TITE-CRM designs intend to accelerate early-phase trials with lengthy DLT evaluation windows, having too many participants with pending DLT outcomes may be problematic. This complication can occur when accrual is too rapid relative to the DLT observation window and can lead to complex ethical and

safety considerations in assigning future participants or a pause in accrual and longer trial duration. Nonetheless, these administrative challenges can be addressed [48] and are far outweighed by the statistical benefits of adaptive designs. Additionally, a key assumption underlying the proposed design is that toxicity and feasibility endpoints are independent. If there are biological circumstances that, for example, cause a higher DLT probability if patients are only feasible at lower dose levels, then the independent toxicity and feasibility models that define our design may not be appropriate. Addressing non-independence of toxicity and feasibility could be an area of future research, but it is currently outside the scope of our design.

There is also an emerging viewpoint for some cell therapeutics where the classic assumptions of higher efficacy with higher dose level no longer apply. This shift in paradigm is not currently covered by the designs described in this dissertation, but could be an area of future work and extensions. Other future endeavors could involve extensions to more than two groups of patients and multi-cell product dose-escalation.

Another topic of interest for clinical trials in general is what happens with larger samples. While these phase I trials usually have very small sample sizes, approximately 10-30 patients, it is interesting to see how our designs perform if given a much larger sample size. The large sample results are given in Appendix C. Despite having better performance than the simulations we have already discussed they will likely never be practical.

We have described 3 efficient and accurate designs for phase I cell therapy trials where there could be patient heterogeneity or late-onset toxicities in a single or dual agent dose-escalation setting. There is a natural trade-off between speeding up the trial and adding risk due to using only partially observed DLT outcome data. Our TITE methods allow for this trade-off to be achieved in a manner that uses all possible safety and feasibility information so that patients do not need to wait too long to receive treatment. Our proposed modeling framework guides the trial so that an accurate FMTD can be de-

termined in a fast manner, allowing future clinical studies of the intervention to proceed. The methods described and investigated model formulations to update dose-toxicity and dose-feasibility probability estimates that have not been previously evaluated in the Phase I clinical trial design literature. New agents in oncology have a high failure rate (approximately 95%) in the drug development process [49]. Even the most well-designed, highly effective cancer therapy, with a well understood mechanism of action, cannot benefit patients unless that agent successfully navigates the development process [1]. We strongly believe that our methods can improve the success of drug development by reducing the number of infeasible treatments carried forward into middle development.

Appendix A

Additional simulation results for aim 1

For results in this appendix refer to Table B.1 to do determine the simulation settings for a particular result.

Table A.1: Scenario identification table for additional results for Aim 1.

Scenario code	Time to Toxicity Distribution	Weight function
a	uniform$(0, 10)$	Linear
b	uniform$(0, 10)$	Adaptive
c	Weibull$(5, 6)$	Linear
d	Weibull$(5, 6)$	Adaptive
e	Double-Exponential$(7.5, 5)$	Linear
f	Double-Exponential$(7.5, 5)$	Adaptive

Table A.2: FMTD selection percentage and duration statistics for Probit model.

Scenario	Dose 1	Dose 2	Dose 3	Dose 4	No Dose	Avg Dur.	Sd Dur.
1 (p_f,p_t)	**(0.89,0.22)**	(0.70,0.32)	(0.37,0.38)	(0.23,0.46)			
a	**69.1**	20.3	0.0	0.0	10.6	59.8	16.1
b	**68.0**	21.0	0.0	0.0	10.9	58.7	16.5
c	**69.9**	19.4	0.0	0.0	10.7	59.0	16.0
d	**67.2**	20.0	0.0	0.0	12.8	57.7	16.7
e	**65.3**	23.9	0.0	0.0	10.8	58.9	16.7
f	**68.2**	21.6	0.0	0.0	10.1	59.2	16.2
2 (p_f,p_t)	(0.99,0.18)	**(0.87,0.24)**	(0.73,0.36)	(0.68,0.40)			
a	33.8	**51.1**	11.3	1.5	2.3	57.2	10.7
b	35.0	**49.7**	10.9	1.5	2.9	57.2	10.7
c	33.9	**49.4**	12.8	1.5	2.4	57.3	10.1
d	33.1	**51.5**	11.9	0.9	2.6	57.8	10.7
e	25.0	**63.3**	9.0	0.7	2.0	57.3	10.4
f	31.1	**52.6**	12.5	1.5	2.2	57.8	10.7
3 (p_f,p_t)	(1.00,0.01)	(0.96,0.15)	**(0.91,0.25)**	(0.84,0.37)			
a	5.7	33.5	**49.6**	11.1	0.1	57.9	9.9
b	6.4	34.8	**47.5**	11.0	0.2	57.4	9.7
c	6.8	34.0	**47.7**	11.4	0.2	58.0	9.9
d	6.1	34.0	**49.2**	10.3	0.4	57.6	10.1
e	3.2	37.9	**51.0**	7.6	0.4	57.4	10.2
3f	4.9	31.6	**50.5**	12.8	0.3	57.7	9.8
4 (p_f,p_t)	(0.97,0.10)	(0.88,0.15)	(0.87,0.19)	**(0.84,0.29)**			
a	8.6	27.2	37.0	**26.6**	0.7	58.4	10.7
b	8.8	27.4	35.8	**27.5**	0.7	58.6	10.7
c	8.6	26.8	37.8	**26.6**	0.4	59.1	10.7
d	8.2	27.1	36.6	**27.4**	0.7	58.9	10.7
e	6.0	30.0	44.1	**19.6**	0.4	58.7	10.4
f	7.0	24.8	38.4	**29.2**	0.7	59.1	10.5
5 (p_f,p_t)	(0.99,0.50)	(0.95,0.58)	(0.94,0.64)	(0.92,0.78)			
a	25.0	0.2	0.0	0.0	**74.8**	40.3	14.8
b	20.4	0.1	0.0	0.0	**79.4**	36.7	15.3
c	24.9	0.2	0.0	0.0	**74.9**	40.4	14.2
d	20.0	0.1	0.0	0.0	**79.9**	37.2	15.6
e	25.7	0.6	0.0	0.0	**73.7**	39.9	15.5
f	28.3	0.3	0.0	0.0	**71.4**	41.9	14.1
6 (p_f,p_t)	(0.60,0.31)	(0.42,0.37)	(0.23,0.63)	(0.00,0.88)			
a	5.1	0.0	0.0	0.0	**94.9**	23.6	19.2
b	5.0	0.0	0.0	0.0	**95.0**	23.3	19.4
c	5.6	0.0	0.0	0.0	**94.4**	23.5	19.0
d	6.2	0.0	0.0	0.0	**93.8**	24.0	19.8
e	5.6	0.0	0.0	0.0	**94.4**	23.7	19.6
f	6.4	0.0	0.0	0.0	**93.6**	23.9	19.8

Table A.3: FMTD selection percentage and duration statistics for Comparison log log model.

Scenario	Dose 1	Dose 2	Dose 3	Dose 4	No Dose	Avg Dur	Sd Dur
1 (p_f,p_t)	**(0.89,0.22)**	(0.70,0.32)	(0.37,0.38)	(0.23,0.46)			
a	**67.6**	22.1	0.0	0.0	10.2	59.6	16.0
b	**65.3**	22.6	0.0	0.0	12.0	58.4	17.2
c	**67.9**	21.9	0.0	0.0	10.2	59.1	16.5
d	**65.3**	23.0	0.0	0.0	11.7	58.4	17.0
e	**67.0**	22.7	0.0	0.0	10.3	59.1	16.4
f	**66.3**	23.6	0.0	0.0	10.1	59.4	16.3
2 (p_f,p_t)	(0.99,0.18)	**(0.87,0.24)**	(0.73,0.36)	(0.68,0.40)			
a	27.4	**57.7**	11.5	0.9	2.6	57.5	10.8
b	30.2	**56.8**	10.5	0.6	1.9	57.6	10.9
c	27.5	**59.6**	10.8	0.4	1.6	57.5	10.7
d	29.1	**57.5**	10.2	0.5	2.5	57.4	11.1
e	28.8	**57.9**	10.3	0.5	2.5	57.7	11.0
f	23.5	**62.4**	11.6	0.6	2.0	57.7	10.6
3 (p_f,p_t)	(1.00,0.01)	(0.96,0.15)	**(0.91,0.25)**	(0.84,0.37)			
a	4.8	37.1	**50.4**	7.6	0.1	58.0	10.0
b	5.0	35.1	**51.7**	8.0	0.1	57.4	10.1
c	3.8	34.4	**52.4**	9.2	0.2	57.4	9.8
d	4.8	37.0	**49.5**	8.5	0.2	57.7	9.8
e	3.6	36.9	**52.0**	7.2	0.2	57.6	10.0
f	2.9	36.4	**52.3**	8.3	0.0	57.3	9.5
4 (p_f,p_t)	(0.97,0.10)	(0.88,0.15)	(0.87,0.19)	**(0.84,0.29)**			
a	5.9	31.7	43.0	**18.8**	0.7	59.2	10.8
b	6.9	31.4	41.2	**20.2**	0.2	59.0	10.3
c	6.4	29.6	43.9	**19.3**	0.8	58.4	10.5
d	6.1	29.5	43.0	**20.8**	0.7	59.1	10.4
e	5.1	30.4	44.1	**19.8**	0.5	58.6	10.4
f	4.6	27.7	45.2	**22.0**	0.5	58.7	10.3
5 (p_f,p_t)	(0.99,0.50)	(0.95,0.58)	(0.94,0.64)	(0.92,0.78)			
a	28.1	0.3	0.1	0.0	**71.4**	41.0	14.2
b	22.5	0.4	0.0	0.0	**77.1**	37.5	15.6
c	27.3	0.5	0.0	0.0	**72.2**	40.9	14.2
d	21.5	0.4	0.0	0.0	**78.0**	37.0	15.5
e	26.2	0.5	0.0	0.0	**73.2**	39.3	15.2
f	30.7	0.4	0.0	0.0	**68.8**	42.0	14.6
6 (p_f,p_t)	(0.60,0.31)	(0.42,0.37)	(0.23,0.63)	(0.00,0.88)			
a	5.8	0.0	0.0	0.0	**94.2**	23.8	19.6
b	5.4	0.0	0.0	0.0	**94.6**	22.7	19.3
c	6.8	0.0	0.0	0.0	**93.2**	23.9	20.1
d	5.7	0.0	0.0	0.0	**94.3**	22.7	19.5
e	6.0	0.0	0.0	0.0	**94.0**	24.1	19.4
f	5.7	0.0	0.0	0.0	**94.2**	24.4	19.8

Table A.4: Average patients treated at each dose level for the Probit model

Scenario	Dose 1	Dose 2	Dose 3	Dose 4
1 (p_f,p_t)	**(0.89,0.22)**	(0.70,0.32)	(0.37,0.38)	(0.23,0.46)
a	**15.6**	5.3	1.2	0.3
b	**15.9**	4.9	1.1	0.3
c	**15.0**	5.7	1.2	0.4
d	**15.7**	4.9	1.1	0.3
e	**12.8**	8.4	1.0	0.0
f	**14.0**	6.3	1.5	0.5
2 (p_f,p_t)	(0.99,0.18)	**(0.87,0.24)**	(0.73,0.36)	(0.68,0.40)
a	12.7	**6.5**	2.8	1.7
b	13.8	**5.7**	2.4	1.7
c	12.5	**6.3**	3.1	1.9
d	13.3	**6.0**	2.6	1.8
e	9.5	**10.8**	3.1	0.3
f	11.2	**6.7**	3.2	2.6
3 (p_f,p_t)	(1.00,0.01)	(0.96,0.15)	**(0.91,0.25)**	(0.84,0.37)
a	8.1	6.1	**5.9**	3.9
b	8.8	6.0	**5.5**	3.6
c	8.1	5.8	**5.8**	4.3
d	8.6	5.8	**5.5**	4.0
e	5.8	10.2	**6.7**	1.3
f	7.0	5.6	**6.2**	5.2
4 (p_f,p_t)	(0.97,0.10)	(0.88,0.15)	(0.87,0.19)	**(0.84,0.29)**
a	8.6	4.9	5.0	**5.3**
b	9.2	4.6	4.9	**5.2**
c	8.4	4.9	4.9	**5.8**
d	9.0	4.6	4.6	**5.7**
e	6.4	9.1	6.5	**1.8**
f	7.6	4.5	5.1	**6.7**
5 (p_f,p_t)	(0.99,0.50)	(0.95,0.58)	(0.94,0.64)	(0.92,0.78)
a	14.4	1.0	0.4	0.3
b	13.1	0.9	0.4	0.3
c	14.4	1.1	0.5	0.4
d	12.9	1.0	0.5	0.3
e	11.6	3.9	0.3	0.0
f	13.7	1.6	0.8	0.8
6 (p_f,p_t)	(0.60,0.31)	(0.42,0.37)	(0.23,0.63)	(0.00,0.88)
a	4.9	0.7	0.2	0.0
b	5.0	0.6	0.2	0.0
c	4.8	0.8	0.2	0.0
d	5.1	0.7	0.2	0.0
e	4.1	1.7	0.1	0.0
f	4.7	0.9	0.3	0.0

Table A.5: Average patients treated at each dose level for the Comparison log log model

Scenario	Dose 1	Dose 2	Dose 3	Dose 4
1 (p_f,p_t)	**(0.89,0.22)**	(0.70,0.32)	(0.37,0.38)	(0.23,0.46)
a	**13.4**	8.0	0.9	0.0
b	**14.1**	7.1	0.7	0.0
c	**13.1**	8.2	0.9	0.0
d	**13.7**	7.5	0.8	0.0
e	**13.0**	8.3	1.0	0.0
f	**11.9**	9.2	1.2	0.1
2 (p_f,p_t)	(0.99,0.18)	**(0.87,0.24)**	(0.73,0.36)	(0.68,0.40)
a	10.4	**10.1**	2.9	0.2
b	11.5	**9.6**	2.4	0.3
c	10.0	**10.5**	3.0	0.3
d	11.1	**9.8**	2.6	0.2
e	10.0	**10.4**	3.1	0.2
f	8.5	**11.2**	3.7	0.3
3 (p_f,p_t)	(1.00,0.01)	(0.96,0.15)	**(0.91,0.25)**	(0.84,0.37)
a	6.1	10.0	**6.7**	1.2
b	7.0	9.9	**6.1**	1.0
c	5.6	10.0	**7.1**	1.3
d	6.5	10.2	**6.2**	1.1
e	5.8	10.0	**6.9**	1.3
f	5.1	9.8	**7.5**	1.6
4 (p_f,p_t)	(0.97,0.10)	(0.88,0.15)	(0.87,0.19)	**(0.84,0.29)**
a	6.5	9.2	6.4	**1.8**
b	7.2	9.0	6.0	**1.6**
c	6.4	9.1	6.5	**1.9**
d	6.8	9.1	6.2	**1.8**
e	6.2	9.3	6.6	**1.9**
f	5.5	8.9	7.1	**2.4**
5 (p_f,p_t)	(0.99,0.50)	(0.95,0.58)	(0.94,0.64)	(0.92,0.78)
a	13.9	2.7	0.2	0.0
b	12.4	2.5	0.1	0.0
c	13.3	3.1	0.1	0.0
d	11.8	2.9	0.1	0.0
e	11.4	4.0	0.3	0.0
f	12.1	4.6	0.3	0.0
6 (p_f,p_t)	(0.60,0.31)	(0.42,0.37)	(0.23,0.63)	(0.00,0.88)
6a	4.5	1.4	0.1	0.0
6b	4.4	1.2	0.1	0.0
6c	4.5	1.5	0.1	0.0
6d	4.2	1.3	0.1	0.0
6e	4.2	1.6	0.1	0.0
6f	4.2	1.8	0.1	0.0

Appendix B

Additional simulation results for aim 2

For results in this appendix refer to Table B.1 to do determine the simulation settings for a particular result.

Table B.1: Scenario identification table for additional results for Aim 2.

Scenario code	Time to Toxicity Distribution	Weight function
a	uniform$(0, 10)$	Linear
b	uniform$(0, 10)$	Adaptive
c	Weibull$(5, 6)$	Linear
d	Weibull$(5, 6)$	Adaptive
e	Double-Exponential$(7.5, 5)$	Linear
f	Double-Exponential$(7.5, 5)$	Adaptive

Table B.2: FMTD selection %. Scenarios 1-3 for Probit model

Scenario	Dose		p_t		p_f	FMTD%		None
1 a	d_1	d_2	0.22	0.32	0.89	58.5	36.8	4.7
	d_3	d_4	0.38	0.46	0.37	0.0	0.0	
b	d_1	d_2	0.22	0.32	0.89	61.0	33.9	5.2
	d_3	d_4	0.38	0.46	0.37	0.0	0.0	
c	d_1	d_2	0.22	0.32	0.89	60.9	35.0	4.1
	d_3	d_4	0.38	0.46	0.37	0.0	0.0	
d	d_1	d_2	0.22	0.32	0.89	60.1	35.9	4.0
	d_3	d_4	0.38	0.46	0.37	0.0	0.0	
e	d_1	d_2	0.22	0.32	0.89	56.5	39.1	4.4
	d_3	d_4	0.38	0.46	0.37	0.0	0.0	
f	d_1	d_2	0.22	0.32	0.89	59.8	35.5	4.8
	d_3	d_4	0.38	0.46	0.37	0.0	0.0	
2 a	d_1	d_2	0.18	0.24	0.87	34.8	60.4	4.8
	d_3	d_4	0.36	0.40	0.36	0.0	0.0	
b	d_1	d_2	0.18	0.24	0.87	36.5	58.8	4.8
	d_3	d_4	0.36	0.40	0.36	0.0	0.0	
c	d_1	d_2	0.18	0.24	0.87	36.4	58.4	5.3
	d_3	d_4	0.36	0.40	0.36	0.0	0.0	
d	d_1	d_2	0.18	0.24	0.87	37.1	57.8	5.1
	d_3	d_4	0.36	0.40	0.36	0.0	0.0	
e	d_1	d_2	0.18	0.24	0.87	34.9	60.8	4.3
	d_3	d_4	0.36	0.40	0.36	0.0	0.0	
f	d_1	d_2	0.18	0.24	0.87	36.6	58.5	4.8
	d_3	d_4	0.36	0.40	0.36	0.00	0.00	
3 a	d_1	d_2	0.10	0.15	0.96	3.0	34.8	0.1
	d_3	d_4	0.25	0.37	0.91	44.5	17.6	
b	d_1	d_2	0.10	0.15	0.96	3.4	36.4	0.2
	d_3	d_4	0.25	0.37	0.91	44.0	15.9	
c	d_1	d_2	0.10	0.15	0.96	3.5	34.8	0.1
	d_3	d_4	0.25	0.37	0.91	45.0	16.5	
d	d_1	d_2	0.10	0.15	0.96	3.7	36.2	0.1
	d_3	d_4	0.25	0.37	0.91	43.9	16.0	
e	d_1	d_2	0.10	0.15	0.96	2.3	37.6	0.1
	d_3	d_4	0.25	0.37	0.91	44.1	15.8	
f	d_1	d_2	0.10	0.15	0.96	2.6	35.4	0.1
	d_3	d_4	0.25	0.37	0.91	46.4	15.4	

Table B.3: FMTD selection %. Scenarios 4-6 for Probit model

Scenario	Dose		p_t		p_f	FMTD%		None
4 a	d_1	d_2	0.11	0.16	0.97	2.5	33.5	0.1
	d_3	d_4	0.19	0.29	0.84	34.9	29.0	
b	d_1	d_2	0.11	0.16	0.97	2.6	33.1	0.0
	d_3	d_4	0.19	0.29	0.84	36.8	27.5	
c	d_1	d_2	0.11	0.16	0.97	2.0	34.2	0.1
	d_3	d_4	0.19	0.29	0.84	34.9	28.7	
d	d_1	d_2	0.11	0.16	0.97	3.2	34.0	0.1
	d_3	d_4	0.19	0.29	0.84	35.4	27.3	
e	d_1	d_2	0.11	0.16	0.97	2.8	33.8	0.0
	d_3	d_4	0.19	0.29	0.84	34.5	28.9	
f	d_1	d_2	0.11	0.16	0.97	3.0	35.6	0.1
	d_3	d_4	0.19	0.29	0.84	34.1	27.2	
5 a	d_1	d_2	0.50	0.58	0.96	32.2	0.8	66.7
	d_3	d_4	0.64	0.78	0.88	0.3	0.0	
b	d_1	d_2	0.50	0.58	0.96	30.6	0.8	68.4
	d_3	d_4	0.64	0.78	0.88	0.3	0.0	
c	d_1	d_2	0.50	0.58	0.96	32.0	0.6	67.0
	d_3	d_4	0.64	0.78	0.88	0.3	0.0	
d	d_1	d_2	0.50	0.58	0.96	30.3	0.8	68.3
	d_3	d_4	0.64	0.78	0.88	0.6	0.0	
e	d_1	d_2	0.50	0.58	0.96	39.2	0.6	59.9
	d_3	d_4	0.64	0.78	0.88	0.3	0.0	
f	d_1	d_2	0.50	0.58	0.96	35.4	0.7	63.6
	d_3	d_4	0.64	0.78	0.88	0.4	0.0	
6 a	d_1	d_2	0.31	0.37	0.60	6.9	1.5	91.6
	d_3	d_4	0.63	0.88	0.23	0.0	0.0	
b	d_1	d_2	0.31	0.37	0.60	6.3	1.8	91.9
	d_3	d_4	0.63	0.88	0.23	0.0	0.0	
c	d_1	d_2	0.31	0.37	0.60	7.0	1.5	91.6
	d_3	d_4	0.63	0.88	0.23	0.0	0.0	
d	d_1	d_2	0.31	0.37	0.60	6.3	1.6	92.1
	d_3	d_4	0.63	0.88	0.23	0.0	0.0	
e	d_1	d_2	0.31	0.37	0.60	6.8	2.0	91.3
	d_3	d_4	0.63	0.88	0.23	0.0	0.0	
f	d_1	d_2	0.31	0.37	0.60	5.8	1.8	92.4
	d_3	d_4	0.63	0.88	0.23	0.0	0.0	

Table B.4: FMTD selection %. Scenarios 1-3 for Comp log log model

Scenario	Dose		p_t		p_f	FMTD%		None
1 a	d_1	d_2	0.22	0.32	0.89	56.9	38.2	4.9
	d_3	d_4	0.38	0.46	0.37	0.0	0.0	
b	d_1	d_2	0.22	0.32	0.89	58.5	36.0	5.4
	d_3	d_4	0.38	0.46	0.37	0.0	0.0	
c	d_1	d_2	0.22	0.32	0.89	55.7	39.4	4.9
	d_3	d_4	0.38	0.46	0.37	0.0	0.0	
d	d_1	d_2	0.22	0.32	0.89	56.8	37.5	5.7
	d_3	d_4	0.38	0.46	0.37	0.0	0.0	
e	d_1	d_2	0.22	0.32	0.89	56.6	38.7	4.7
	d_3	d_4	0.38	0.46	0.37	0.0	0.0	
f	d_1	d_2	0.22	0.32	0.89	57.6	37.3	5.0
	d_3	d_4	0.38	0.46	0.37	0.0	0.0	
2 a	d_1	d_2	0.18	0.24	0.87	31.8	63.2	5.0
	d_3	d_4	0.36	0.40	0.36	0.0	0.0	
b	d_1	d_2	0.18	0.24	0.87	32.0	63.0	5.0
	d_3	d_4	0.36	0.40	0.36	0.0	0.0	
c	d_1	d_2	0.18	0.24	0.87	31.6	63.8	4.7
	d_3	d_4	0.36	0.40	0.36	0.0	0.0	
d	d_1	d_2	0.18	0.24	0.87	34.1	60.5	5.3
	d_3	d_4	0.36	0.40	0.36	0.0	0.0	
e	d_1	d_2	0.18	0.24	0.87	31.9	62.2	5.9
	d_3	d_4	0.36	0.40	0.36	0.0	0.0	
f	d_1	d_2	0.18	0.24	0.87	31.1	63.2	5.6
	d_3	d_4	0.36	0.40	0.36	0.0	0.0	
3 a	d_1	d_2	0.10	0.15	0.96	2.5	35.7	0.1
	d_3	d_4	0.25	0.37	0.91	44.5	17.1	
b	d_1	d_2	0.10	0.15	0.96	3.2	34.0	0.1
	d_3	d_4	0.25	0.37	0.91	44.9	17.7	
c	d_1	d_2	0.10	0.15	0.96	2.5	33.9	0.1
	d_3	d_4	0.25	0.37	0.91	44.5	18.9	
d	d_1	d_2	0.10	0.15	0.96	2.7	32.1	0.1
	d_3	d_4	0.25	0.37	0.91	46.7	18.4	
e	d_1	d_2	0.10	0.15	0.96	3.2	35.0	0.2
	d_3	d_4	0.25	0.37	0.91	42.5	19.0	
f	d_1	d_2	0.10	0.15	0.96	2.4	35.9	0.2
	d_3	d_4	0.25	0.37	0.91	42.8	18.7	

Table B.5: FMTD selection %. Scenarios 4-6 for Comp log log model

Scenario	Dose		p_t		p_f	FMTD%		None
4 a	d_1	d_2	0.11	0.16	0.97	3.0	32.9	0.0
	d_3	d_4	0.19	0.29	0.84	32.4	31.8	
b	d_1	d_2	0.11	0.16	0.97	2.6	31.4	0.1
	d_3	d_4	0.19	0.29	0.84	33.6	32.1	
c	d_1	d_2	0.11	0.16	0.97	2.2	31.1	0.1
	d_3	d_4	0.19	0.29	0.84	35.0	31.6	
d	d_1	d_2	0.11	0.16	0.97	2.2	31.8	0.1
	d_3	d_4	0.19	0.29	0.84	32.6	33.2	
e	d_1	d_2	0.11	0.16	0.97	1.9	30.0	0.0
	d_3	d_4	0.19	0.29	0.84	34.3	33.9	
f	d_1	d_2	0.11	0.16	0.97	2.2	34.1	0.1
	d_3	d_4	0.19	0.29	0.84	32.4	31.1	
5 a	d_1	d_2	0.50	0.58	0.96	35.3	0.6	63.5
	d_3	d_4	0.64	0.78	0.88	0.7	0.0	
b	d_1	d_2	0.50	0.58	0.96	30.5	0.8	68.3
	d_3	d_4	0.64	0.78	0.88	0.4	0.0	
c	d_1	d_2	0.50	0.58	0.96	36.0	1.4	62.0
	d_3	d_4	0.64	0.78	0.88	0.7	0.0	
d	d_1	d_2	0.50	0.58	0.96	34.5	0.9	64.2
	d_3	d_4	0.64	0.78	0.88	0.4	0.0	
e	d_1	d_2	0.50	0.58	0.96	40.4	1.0	58.4
	d_3	d_4	0.64	0.78	0.88	0.2	0.0	
f	d_1	d_2	0.50	0.58	0.96	36.1	1.0	62.3
	d_3	d_4	0.64	0.78	0.88	0.6	0.0	
6 a	d_1	d_2	0.31	0.37	0.60	6.2	2.0	91.8
	d_3	d_4	0.63	0.88	0.23	0.0	0.0	
b	d_1	d_2	0.31	0.37	0.60	5.8	2.0	92.2
	d_3	d_4	0.63	0.88	0.23	0.0	0.0	
c	d_1	d_2	0.31	0.37	0.60	6.3	2.1	91.6
	d_3	d_4	0.63	0.88	0.23	0.0	0.0	
d	d_1	d_2	0.31	0.37	0.60	5.7	1.8	92.6
	d_3	d_4	0.63	0.88	0.23	0.0	0.0	
e	d_1	d_2	0.31	0.37	0.60	6.6	1.9	91.6
	d_3	d_4	0.63	0.88	0.23	0.0	0.0	
f	d_1	d_2	0.31	0.37	0.60	6.3	1.9	91.8
	d_3	d_4	0.63	0.88	0.23	0.0	0.0	

Appendix C

Large Sample Results

Table C.1: Large sample FMTD distributions for aim 1 (1000 patients)

Scenario		Dose 1	Dose 2	Dose 3	Dose 4	None
1	(p_f, p_t)	**(0.89,0.22)**	(0.7,0.32)	(0.37,0.38)	(0.25,0.46)	
	FMTD%	**83.4**	0.0	0.0	0.0	16.6
2	(p_f, p_t)	(0.99,0.18)	**(0.87,0.24)**	(0.73,0.36)	(0.68,0.4)	
	FMTD%	8.7	**88.4**	0.0	0.0	2.9
3	(p_f, p_t)	(1,0.01)	(0.96,0.15)	**(0.91,0.25)**	(0.84,0.37)	
	FMTD%	0.1	3.5	**96.2**	0.0	0.2
4	(p_f, p_t)	(0.97,0.1)	(0.88,0.15)	(0.87,0.19)	**(0.84,0.29)**	
	FMTD%	0.1	5.8	21.9	**71.5**	0.7
5	(p_f, p_t)	(0.99,0.5)	(0.95,0.58)	(0.94,0.64)	(0.92,0.78)	
	FMTD%	0.0	0.0	0.0	0.0	**100**
6	(p_f, p_t)	(0.6,0.31)	(0.42,0.37)	(0.23,0.63)	(0,0.88)	
	FMTD%	0.0	0.0	0.0	0.0	**100**

Table C.2: Large sample FMTD distributions for aim 2 (1000 patients)

Scenario	Dose		p_t		p_f	FMTD %		None
1	d_1	d_2	0.22	0.32	0.89	**80.2**	17.0	2.8
	d_3	d_4	0.38	0.46	0.37	0.0	0.0	
2	d_1	d_2	0.18	0.24	0.87	21.3	**72.9**	5.8
	d_3	d_4	0.36	0.40	0.36	0.0	0.0	
3	d_1	d_2	0.10	0.15	0.96	0.0	1.4	0.1
	d_3	d_4	0.25	0.37	0.91	**93.4**	5.1	
4	d_1	d_2	0.11	0.15	0.97	0.0	1.6	0.2
	d_3	d_4	0.19	0.29	0.84	36.1	**62.1**	
5	d_1	d_2	0.50	0.58	0.96	0.0	0.0	**100**
	d_3	d_4	0.64	0.78	0.88	0.0	0.0	
6	d_1	d_2	0.31	0.37	0.60	0.0	0.0	**100**
	d_3	d_4	0.63	0.88	0.23	0.0	0.0	

Table C.3: Large sample FMTD Distribution for aim 3 (1000 patients)

Scenario		No Prior SCT ($g = 1$)			Prior SCT ($g = 2$)			
	Dose	1	2	3	1	2	3	None
	p_t	0.15	**0.26**	0.43	**0.25**	0.43	0.68	
1	p_f	0.98	**0.93**	0.84	**0.98**	0.93	0.84	
Proposed	FMTD%	2.4	**9.8**	0.7	**96.5**	0.4	0.0	3.1
	p_t	0.07	0.11	**0.24**	**0.25**	0.43	0.68	
2	p_f	0.98	0.93	**0.84**	**0.98**	0.93	0.84	
Proposed	FMTD%	0.0	0.2	**99.5**	**99.5**	0.2	0.0	0.3
	p_t	0.07	0.11	**0.24**	0.15	**0.26**	0.43	
3	p_f	0.98	0.93	**0.84**	0.98	**0.93**	0.84	
Proposed	FMTD%	0.0	0.0	**99.6**	7.8	**91.7**	0.1	0.4
	p_t	0.15	**0.26**	0.43	0.15	**0.26**	0.43	
4	p_f	0.98	**0.93**	0.84	0.98	**0.93**	0.84	
Proposed	FMTD%	0.1	**97.7**	0.7	5.8	**92.7**	0.0	1.5
	p_t	0.15	**0.26**	0.43	**0.25**	0.43	0.68	
5	p_f	0.94	**0.85**	0.60	**0.94**	0.85	0.60	
Proposed	FMTD%	3.6	**93.1**	0.0	**96.2**	0.5	0.0	3.3
	p_t	0.07	**0.11**	0.24	**0.25**	0.43	0.68	
6	p_f	0.94	**0.85**	0.60	**0.94**	0.85	0.60	
Proposed	FMTD%	0.0	**99.1**	0.0	**99.0**	0.1	0.0	0.9
	p_t	0.07	**0.11**	0.24	0.15	**0.26**	0.43	
7	p_f	0.94	**0.85**	0.60	0.94	**0.85**	0.60	
Proposed	FMTD%	0.0	**99.8**	0.0	9.7	**90.1**	0.0	0.2
	p_t	0.15	**0.26**	0.43	0.15	**0.26**	0.43	
8	p_f	0.94	**0.85**	0.60	0.94	**0.85**	0.60	
Proposed	FMTD%	0.3	**97.8**	0.0	8.8	**89.3**	0.0	1.9

Bibliography

1. Conaway, M. R. & Petroni, G. R. The Impact of Early-Phase Trial Design in the Drug Development Process. *Clinical Cancer Research* **25,** 819–27 (2019).
2. Yu, J. X., Upadhaya, S., Tatake, R., Barkalow, F. & Hubbard-Lucey, V. M. Cancer cell therapies: the clinical trial landscape. *Nat Rev Drug Discov* **19,** 583–84 (2020).
3. Lum, L. G., Thakur, A., Liu, Q. & Al-Khadhimi, Z. CD20-targeted T cells after stem cell transplantation for high risk and refractory non-Hodgkin's lymphoma. *Biol Blood Marrow Transplant* **19,** 925–33 (2013).
4. Rogatko, A., Schoeneck, D., Jonas, W. & Tighiouart, M. Translation of innovative designs into phase I trials. *J Clin Oncol* **25,** 4982–6 (2007).
5. North, B., Kocher, H. M. & Sasieni, P. A new pragmatic design for dose escalation in phase 1 clinical trials using an adaptive continual reassessment method. *BMC Cancer* **19** (2019).
6. Nie, L., Rubin, E. H., Mehrotra, N. & Pinheiro, J. Rendering the 3+3 Design to Rest: More Efficient Approaches to Oncology Dose-Finding Trial in the Era of Targeted Therapy. *Clin Cancer Res* **22,** 2623–9 (2016).
7. Paoletti, X., Ezzalfani, M. & Tourneau, C. L. Statistical Controversies in Clinical Research: requiem for the 3+3 design for phase I trials. *Ann Oncol* **26,** 1808–12 (2015).

8. Lee, D. W., Kochenderfer, J. N., Stevenson, M. S. & Cui, Y. T cells expressing CD19 chimeric antigen receptors for acute lymphoblastic leukaemia in children and young adults: a phase 1 dose-escalation trial. *Lancet* **385,** 517–28 (2015).

9. FDA. *Considerations for the Design of Early-Phase Clinical Trials of Cellular and Gene Therapy Products* `https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-design-early-phase-clinical-trials-cellular-and-gene-therapy-products` (2021).

10. Thall, P. F., Sung, H. G. & Choudhury, A. Dose-finding based on feasibility and toxicity in T-cell infusion trials. *Biometrics* **57,** 914–21 (2001).

11. Wages, N. A. & Fadul, C. E. Adaptive dose-finding based on safety and feasibility in early-phase clinical trials of adoptive cell immunotherapy. *Clinical Trials* **17,** 157–165 (2020).

12. Devlin, S. M., Iasonos, A. & O'Quigley, J. Phase I Clinical trials in adoptive T-cell therapies. *J R Stat Soc Ser C Appl Stat* **70,** 815–834 (2021).

13. O'Quigley, J., Pepe, M. & Fisher, L. Continual reassessment method: a practical design for phase 1 clinical trials in cancer. *Biometrics* **46,** 33–48 (1990).

14. Conaway, M. R., Dunbar, S. & Peddada, S. D. Designs for Single- or Multiple-Agent Phase I Trials. *Biometrics* **60,** 661–9 (2004).

15. Robertson, T., Wright, F. T. & Dykstra, R. *Order Restricted Statistical Inference* (Springer, 1998).

16. Bagley, E. M. & Wages, N. A. Impact of dose feasibility on the conduct of phase I trials of adoptive cell therapy. *Contemporary Clinical Trials Communications* **25,** 100877 (2022).

17. Lum, L. G., Thakur, A., Al-Khadhimi, Z. & Colvin, G. A. Targeted T cell Therapy in Stage IV Breast Cancer: A Phase I Clinical Trial. *Clin Cancer Res* **21,** 2305–2314 (2015).

18. Leonard, J. P., Furman, R. R., Cheung, Y. K. & Feldman, E. J. Phase I/II trial of bortezomib plus CHOP-Rituximab in diffuse large B cell (DLBCL) and mantle cell lymphoma (MCL). *Blood* **106,** 491 (2005).

19. Lee, S. M., Backenroth, D., Cheung, Y. K. & Hershman, D. L. Case Example of Dose Optimization Using Data From Bortezomib Dose-Finding Clinical Trials. *J Clin Oncol* **34,** 1395–1401 (2016).

20. Roda, D., Jimenez, B. & Banerji, U. Are Doses and Schedules of Small-Molecule Targeted Anticancer Drugs Recommended by Phase I Studies Realistic? *Clin Cancer Res* **22,** 2127–37 (2016).

21. Cheung, Y. K. & Chapell, R. Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics* **56,** 1177–82 (2000).

22. Yuan, Y., Lin, R., Li, D. & Warren, K. E. Time-to-Event Bayesian Optimal Interval Design to Accelerate Phase I Trials. *Clin Cancer Res* **24,** 4921–4930 (2018).

23. Yan, D., Tait, C., Wages, N. A., Kindwall-Keller, T. & Dressler, E. V. Generalization of the time-to-event continual reassessment method to bivariate outcomes. *Journal of Biopharmaceutical Statistics* **29,** 635–647 (2019).

24. Lin, R. & Yuan, Y. Time-to-event model-assisted designs for dose-finding trials with delayed toxicity. *Biostatistics* **21,** 807–24 (2020).

25. Wages, N. A., Braun, T. & O'Quigley, J. Isotonic design for phase I cancer clinical trials with late-onset toxicities. *Journal of Biopharmaceutical Statistics* **Epub ahead of print** (2023).

26. Bagley, E. M. & Wages, N. A. Dose-finding based on feasibility and late-onset toxicity in adoptive cell therapy trials. *Journal of Biopharmaceutical Statistics* **Epub Ahead of print** (2023).

27. Lee, S. M. & Cheung, Y. K. Model calibration in the continual reassessment method. *Clin Trials* **6,** 227–38 (2009).

28. Cheung, Y. K. *Dose Finding by the Continual Reassessment Method* (Chapman and Hall/CRC, 2011).

29. Lee, S. M. & Cheung, Y. K. Calibration of prior variance in the Bayesian Continual Reassessment Method. *Stat Med* **30,** 2081–89 (2011).

30. Cheung, Y. K. & Chapell, R. A simple technique to evaluate model sensitivity in the continual reassessment method. *Biometrics* **58,** 671–4 (2002).

31. Riviere, M. K., Tourneau, C. L., Paoletti, X., Dubois, F. & Zohar, S. Designs of drug-combination phase I trials in oncology: a systematic review of literature. *Annals of Oncology* **26,** 669–74 (2015).

32. Wages, N. A., Conaway, M. R. & O'Quigley, J. Continual Reassesment Method for Partial Ordering. *Biometrics* **67,** 1555–1563 (2011).

33. Lin, R. & Yin, G. Bayesian optimal interval design for dose finding in drug-combination trials. *Stat Methods Med Res* **26,** 2155–67 (2015).

34. Thall, P. F., Millikan, R. E., Mueller, P. & Lee, S.-J. Dose-finding with two agents in Phase I oncology trials. *Biometrics* **59,** 487–96 (2003).

35. Wang, K. & Ivanova, A. Two-dimensional dose finding in a discrete dose space. *Biometrics* **61,** 217–22 (2005).

36. Yuan, Y. & Yin, G. Sequential continual reassesment method for two-dimensional dose finding. *Stat Med* **27,** 5664–78 (2008).

37. Wages, N. A., Conaway, M. R. & O'Quigley, J. Dose-finding design for multi-drug combinations. *Clin Trials* **8,** 380–89 (2011).

38. Wages, N. A., Conaway, M. R. & O'Quigley, J. Using the time-to-event continual reassessment method in the presence of partial orders. *Stat Med* **32,** 131–141 (2013).

39. Horton, B. J., O'Quigley, J. & Conaway, M. R. Consequences of Performing Parallel Dose Finding Trials in Heterogenous Groups of Patients. *JNCI Cancer Spectrum* **3** (2018).

40. O'Quigley, J., Shen, L. Z. & Gamst, A. Two-sample continual reassesment method. *Journal of Biopharmaceutical Statistics* **9,** 17–44 (1999).

41. O'Quigley, J. & Paoletti, X. Continual reassessment method for ordered groups. *Biometrics* **59,** 430–40 (2003).

42. O'Quigley, J. Theoretical study of the continual reassessment method. *Journal of Statistical Planning and Inference* **136,** 1756–80 (2006).

43. O'Quigley, J. & Conaway, M. R. Continual Reassessment Method and Related Dose-Finding Designs. *Stat Sci* **25,** 202–16 (2010).

44. O'Quigley, J. & Iasonos, A. Bridging Solutions in Dose Finding Problems. *Stat Biopharm Res* **6,** 185–197 (2014).

45. Wages, N. A., Read, P. W. & Petroni, G. R. A Phase I/II adaptive design for heterogeneous groups with application to a stereotactic body radiation therapy trial A Phase I/II adaptive design for heterogenous group with application to stereotactic body radiation therapy. *Pharmaceutical Statistics* **14,** 302–10 (2015).

46. Horton, B. J., Wages, N. A. & Conaway, M. R. Shift models for dose-finding in partially ordered groups. *Clinical Trials* **16,** 32–40 (2019).

47. Sharon, E., Polley, M.-Y., Bernstein, M. B. & Ahmed, M. Immunotherapy and radiation therapy: considerations for successfully combining radiation into the paradigm of immuno-oncology drug development. *Radiat Res* **182,** 252–7 (2014).

48. Van Werkhoven, E., Hinsley, S., Frangou, E., Holmes, J. & de Haan, R. Practicalities in running early-phase trials using the time-to-event continual reassessment method (TiTE-CRM) for interventions with long toxicity periods using two radiotherapy oncology trials as examples. *BMC Med ResMethodol* **20** (2020).

49. Printz, C. Failure rate: Why many cancer drugs don’t receive FDA approval, and what can be done about it. *Cancer* **121,** 1529–30 (2015).