2018

Bivariate Generalization of the Time-to-Event Conditional Reassessment Method with a Novel Adaptive Randomization Method

Donglin Yan
University of Kentucky, donglin.yan@uky.edu
Author ORCID Identifier:
https://orcid.org/0000-0002-5111-9437
Digital Object Identifier: https://doi.org/10.13023/ETD.2018.037

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Donglin Yan, Student

Dr. Emily V. Dressler, Major Professor

Dr. Steven Browning, Director of Graduate Studies

Bivariate Generalization of the Time-to-Event Conditional Reassessment Method with a Novel Adaptive Randomization Method

---

DISSERTATION

A dissertation submitted in partial
fulfillment of the requirements for
the degree of Doctor of Philosophy
in the College of Public Health at
the University of Kentucky

By

Donglin Yan

Lexington, Kentucky

Director: Dr. Emily V. Dressler, Associate Professor of Biostatistics
Lexington, Kentucky 2018

ABSTRACT OF DISSERTATION

Bivariate Generalization of the Time-to-Event Conditional Reassessment Method with a Novel Adaptive Randomization Method

Phase I clinical trials in oncology aim to evaluate the toxicity risk of new therapies and identify a safe but also effective dose for future studies. Traditional Phase I trials of chemotherapies focus on estimating the maximum tolerated dose (MTD). The rationale for finding the MTD is that better therapeutic effects are expected at higher dose levels as long as the risk of severe toxicity is acceptable. With the advent of a new generation of cancer treatments such as the molecularly targeted agents (MTAs) and immunotherapies, higher dose levels no longer guarantee increased therapeutic effects, and the focus has shifted to estimating the optimal biological dose (OBD). The OBD is a dose level with the highest biologic activity with acceptable toxicity. The search for OBD requires joint evaluation of toxicity and efficacy. Although several seamleass phase I/II designs have been published in recent years, there is not a consensus regarding an optimal design and further improvement is needed for some designs to be widely used in practice.

In this dissertation, we propose a modification to an existing seamless phase I/II design by Wages and Tait (2015) for locating the OBD based on binary outcomes, and extend it to time to event (TITE) endpoints. While the original design showed promising results, we hypothesized that performance could be improved by replacing the original adaptive randomization stage with a different randomization strategy. We proposed to calculate dose assigning probabilities by averaging all candidate models that fit the observed data reasonably well, as opposed to the original design that based all calculations on one best-fit model. We proposed three different strategies to select and average among candidate models, and simulations are used to compare the proposed strategies to the original design. Under most scenarios, one of the proposed strategies allocates more patients to the optimal dose while improving accuracy in selecting the final optimal dose without increasing the overall risk of toxicity.

We further extend this design to TITE endpoints to address a potential issue of delayed outcomes. The original design is most appropriate when both toxicity and efficacy outcomes can be observed shortly after the treatment, but delayed outcomes are common, especially for efficacy endpoints. The motivating example for this TITE extension is a Phase I/II study evaluating optimal dosing of all-trans retinoic acid

(ATRA) in combination with a fixed dose of daratumumab in the treatment of relapsed or refractory multiple myeloma. The toxicity endpoint is observed in one cycle of therapy (*i.e.*, 4 weeks) while the efficacy endpoint is assessed after 8 weeks of treatment. The difference in endpoint observation windows causes logistical challenges in conducting the trial, since it is not acceptable in practice to wait until both outcomes for each participant have been observed before sequentially assigning the dose of a newly eligible participant. The result would be a delay in treatment for patients and undesirably long trial duration. To address this issue, we generalize the time-to-event continual reassessment method (TITE-CRM) to bivariate outcomes with potentially non-monotonic dose-efficacy relationship. Simulation studies show that the proposed TITE design maintains similar probability in selecting the correct OBD comparing to the binary original design, but the number of patients treated at the OBD decreases as the rate of enrollment increases.

We also develop an R package for the proposed methods and document the R functions used in this research. The functions in this R package assist implementation of the proposed randomization strategy and design. The input and output format of these functions follow similar formatting of existing R packages such as "dfcrm" or "pocrm" to allow direct comparison of results. Input parameters include efficacy skeletons, prior distribution of any model parameters, escalation restrictions, design method, and observed data. Output includes recommended dose level for the next patient, MTD, estimated model parameters, and estimated probabilities of each set of skeletons. Simulation functions are included in this R package so that the proposed methods can be used to design a trial based on certain parameters and assess performance. Parameters of these scenarios include total sample size, true dose-toxicity relationship, true dose-efficacy relationship, patient recruit rate, delay in toxicity and efficacy responses.

KEYWORDS: Phase I/II trials, Dose Finding, Time to Event endpoints, R package, Continual Reassessment Method

Author's signature: Donglin Yan

Date: February 13, 2018

Bivariate Generalization of the Time-to-Event Conditional Reassessment Method with a Novel Adaptive Randomization Method

By
Donglin Yan

Director of Dissertation: Emily V. Dressler

Director of Graduate Studies: Steven Browning

Date: February 13, 2018

ACKNOWLEDGMENTS

I greatly appreciate all the help and guidance from my committee over the past three years. I would like to express the most important acknowledgment of gratitude to my mentor Dr. Emily Dressler, for without whom I couldn't have completed this dissertation. I am thankful for her patience and guidance. Throughout my PhD study, Dr. Dressler was always there to help me and guide me, for which I can't say enough thank you. I would like to thank Dr. Wages. This dissertation is a natural extension of his original design, and I appreciate the time and efforts he spent in helping me through this dissertation research. As a new researcher, I learned a lot from working with Dr. Dressler and Dr. Wages.

I would also like to thank Dr. Heather Bush for giving me guidance and inspirations since the beginning of my PhD study, and Dr. Erin Abner for being a great source of knowledge to me. Many of my foundation classes were taught by Dr. Abner, and those well designed classes have been beneficial to my research and career. I would like to thank Dr. Saeed for serving on my committee and providing clinical insights to this work, and Dr. Sylvie Garneau-Tsodikova for being my outside examiner and providing support during the defense process.

I express my gratitude to everyone in the College of Public Health and Biostatistics and Bioinformations Shared Resource Facility. I learned so much in the past three years, and made a lot of good friends. I would also like to thank parents and wife for their unconditional love and constant support. Getting a PhD is wonderful experience but there were difficult times. Your love is the brightest light in the darkest moments of life. Finally, my beloved son, Skyler, you are all the motivation I need to keep going.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1 Introduction

Phase I trials in oncology

Cancer remains a leading cause of death in the US. According to the National Cancer Institute (NCI) estimates, there will be about 1.7 million new cancer cases and 600,000 cancer related death in 2017 [1, 2]. The heavy burden of cancer on families, healthcare systems, and society demands rapid development of better therapies. Potential solutions to this public health issue include better cancer prevention programs and more advanced cancer treatments. Clinical trials are one of the most important steps in developing novel treatments against cancer. Not only cancer patients may benefit from the novel treatments developed through clinical trials, crucial information about the mechanism of cancer is also be obtained from clinical trials. Such information will also give new insights about cancer prevention.

Although the first recorded clinical trial dates back to the biblical descriptions in 500 BC., the study of clinical trial design is a very new field of research. The first controlled trial was conducted by Dr. James Lind in 1747. The concept of placebo effect arrived in the early 1800s and the first double blind controlled trial was conducted in 1943 [3, 4]. Statistics were not incorporated in clinical trials until late 1940s when the first randomized trial was proposed and implemented by Sir Bradford Hill [5].

Modern drug development involves multiple clinical trials which are conventionally divided into three Phases, Phase I, II, and III. Phase I is commonly known as the "first in human studies" as it is the first time a new drug is being tested on human subjects. Phase I trials are especially crucial in developing oncology drugs because of the inherent risks of cancer treatments [6]. Prior to phase I studies, extensive research must be conducted to provide rationale of potential therapeutic benefits.

Pharmacokinetics and pharmacodynamics (PKPD) models need to be established from non-human subjects. However, pre-clinical predictions are not always accurate and can only provide an estimate of the recommended dose range. Several doses or administration schedules may be proposed from pre-clinical studies, and the selection of optimal dosing schedule can only be obtained through trials on human subjects. In general, the goal of phase I trial is to evaluate the safety of a new drug on human subjects, establish the first human PKPD model, and select the most promising dose for further research.

The concept of phase I trials arose from the development of chemotherapies since the 1940s [7]. The primary goal of phase I trial for chemotherapy drugs is to find the maximum tolerated dose (MTD) beyond which the consequence of toxicity outweighs potential benefits. The therapeutic effects of chemotherapies are achieved by impairing the process of DNA replication or cell division. Chemotherapy drugs are classified as cytotoxic since they are toxic to all living cells. Because cancer cells are constantly dividing, they are more likely to be affected than normal cells. Therefore, for chemotherapies, the MTD is the most promising dose for therapeutic benefits.

The development of novel cancer treatments, such as molecularly targeted agents (MTAs) and immunotherapies, is changing the landscape of phase I trials. These treatments act through different mechanisms than traditional chemotherapies. MTAs are designed to target molecules on the pathway of cancer growth or self-repair. Immunotherapies enhance or assist the immune system to target cancer cells. These novel methods of cancer treatments are referred to as non-cytotoxic or cytostatic, since their therapeutic effects are not achieved through directly attacking living cells. For cytostatic agents, it is not always appropriate to assume that higher doses will increase the probability of a positive response. Therefore, the aim for trials of cytostatic agents is usually to identify the optimal biologic dose (OBD) which is defined as the dose with the highest probability of efficacy and acceptable toxicity. Dose-finding

trials aiming to select the OBD require novel dose-finding designs.

A well designed phase I trial not only provides a justified dosing schedule for future development, but also directly impacts the patients participating in those phase I studies. In modern clinical trials, patient safety and benefits deserve the highest priorities in the trial design. The population for Phase I cancer trials are patients who have often exhausted existing treatment options. The enrolled patients usually have advanced stage cancer, and possibly deteriorated health conditions due to exposure to prolonged treatments. Treating these patients either too aggressively or too conservatively raises ethical concerns. Therefore, phase I trials need to be carefully designed to minimize the risk and maximize potential benefits.

In the remaining sections of this chapter, we first introduce traditional early phase dose-finding designs, including rule-based designs and model based designs, as well as the expansion cohort approach. After the traditional designs, we review recently published literature on seamless phase I/II designs for finding OBD and discuss the rationale for the research in this dissertation.

1.1 Dose-finding trials in oncology

A dose-finding trial aims to select one dose level of a new agent from a set of I pre-defined doses $\mathcal{D} = \{d_1, d_2 \ldots, d_I\}$. If it is plausible to assume that efficacy monotonically increases with dose levels, then the goal is to find the MTD; otherwise, if it's possible for efficacy to plateau or decrease after an intermediate dose, the OBD need to be selected. Suppose toxicity and efficacy are observed through binary endpoints Y_j and Z_j, where j is the index of patients. Toxicity data are observed as the grade of adverse events (AEs) according to the Common Terminology Criteria for Adverse Events v4.0 (CTCAEv4) [8] and dichotomized into dose-limiting toxicity (DLT) or non-DLT. Generally, a grade 3 or 4 AE will be classified as DLT and lower grade AEs are non-DLT. Efficacy is observed through pre-defined measures such as

response evaluation criteria in solid tumors (RECIST), proportion of target molecule inhibition, or PD/PK measures.

$$Y_j = \begin{cases} 0, & \text{if no DLT observed} \\ 1, & \text{if DLT observed} \end{cases}, \text{ and } Z_j = \begin{cases} 0, & \text{if no efficacy response} \\ 1, & \text{if observe efficacy response} \end{cases}$$

The dose received by the j^{th} patient is denoted by $x_j \in \mathcal{D}$, and the length of time a patient has been treated by the testing drug is denoted by t_j. The data observed from the first n patients can be denoted by

$$\Omega_n = \{(x_j, t_j, y_j, z_j) | j = 1, 2, \ldots, n\}.$$

If the outcomes can be observed within a reasonable time frame, the follow-up time t_j may be ignored by the trial design. For example, the original CRM design only requires the toxicity data, $\Omega_n = \{(x_j, y_j)\}$, whereas time-to-event (TITE) CRM adds the time variable $\Omega_n = \{(x_j, y_j, t_j)\}$

Traditional dose-finding designs assume that patients treated at higher dose levels will have a larger probability of observing a toxicity response but also better chance of having therapeutic effects. Therefore, the goal for traditional designs is to find the MTD, defined as the dose level $i = 1, 2, \ldots, I$ with DLT probability closest to a pre-specified level ξ, that is

$$MTD = \arg\min_i |Pr(Y = 1|d_i) - \xi|.$$

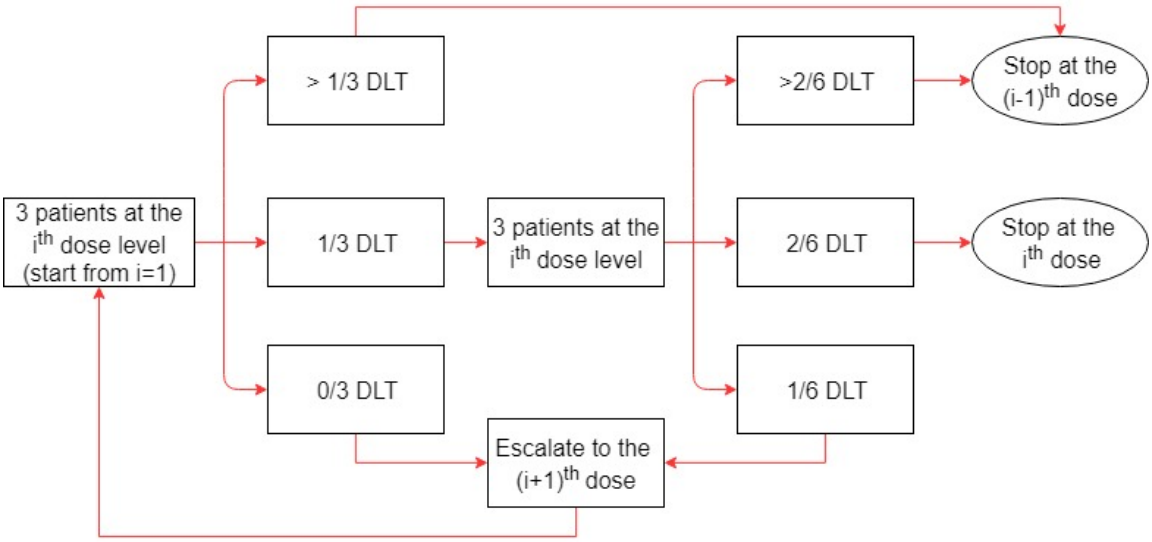
Conventionally, patients are enrolled and treated in cohorts of three, so $\xi = 33\%$ is often used in phase I oncology trials.

1.2 Rule-based designs

Rule-based designs determine dose level for the next cohort based on a pre-defined algorithm. The first cohort of patients are typically tested at the lowest dose level.

The dose level for the next cohort is determined by the toxicity outcomes of the current cohort. If the proportion of patients experience DLT is less than a pre-specified value, the next cohort will receive the next higher dose level; otherwise, the dose level will decrease or stay at the current level. The trial stops when either the quota for sample size is reached or MTD is concluded. Patients are conventionally enrolled in cohorts of 3, so these designs are also referred to as the '3+3' designs. As an example, figure (1.1) shows a standard '3+3' escalation design.

Figure 1.1: Standard '3+3' escalation design



There are several variations of the standard escalation design. For example, the up-and-down designs allow de-escalation based on the outcomes of the most recent cohort [9, 10]. Alternatively, a two-stage design can be used in order to treat less patients at sub-therapeutic doses. In a two-stage design, cohorts of one patient may be used for escalation, and the trial switches to different design after the first DLT is observed. Other designs such as the accelerated titration designs, biased coin design and its variations, and pharmacologically guided designs are also available for Phase I dose-finding trials [11–13]. Methods proposed in this dissertation are referred to as model-based designs rather than rule-based. Therefore, we will not introduce the

other rule-based designs in further details. A comprehensive review of dose escalation designs can be found in an article by LeTourneau [14].

1.3 Model-based designs

Model-based designs are built on underlying statistical models that describe the dose toxicity relationship. A major advantage of model-based designs is that all observed data are utilized, as opposed to rule-based designs where the dose assignment is determined by outcomes from the most current cohort.

CRM: The continual reassessment method (CRM) [15] is the first model-based design proposed for adoption in phase I trials. The CRM uses a one-parameter model to estimate the dose-toxicity curve. Note that the model only needs to reasonably well approximate the true dose-toxicity curve around the targeted toxicity level. Dose assignment by CRM is based on the estimated dose-toxicity curve rather than a set of pre-defined rules. After its original proposal, the CRM design gained popularity among biostatisticians and several CRM-related designs were proposed to handle more complex situations [16]. The original CRM design assumes a dose-toxicity model [15] $\pi_Y(x_j) = F(x_j;\theta)$, where π_Y denotes the probability of observing a DLT response from the j^{th} patient, and $x_j \in \mathcal{D}$ corresponds to the dose assigned to the j^{th} patient. Commonly used dose-toxicity models include the empirical model

$$F(x,\theta) = x^{exp(\theta)},$$

the logistic model

$$F(x,\theta) = \frac{exp(a_0+\theta x)}{1+exp(a_0+\theta x)},$$

and the hyperbolic tangent model

$$F(x,\theta) = \left(\frac{\tanh x + 1}{2}\right)^{\theta}.$$

The original CRM design was proposed using a Bayesian approach, by which the prior distribution $g(\theta)$ needs to be specified before the trial. For example, we can use

the empirical model and normal prior distribution, $\theta \sim N(\hat{\theta}_0, \sigma_\theta^2)$, where $\hat{\theta}_0$ and σ_θ^2 are prior mean and variance, respectively.

The CRM design requires a set of 'skeletons' to be specified prior to the trial. Denoted by $\mathbf{p} = \{p_1, p_2, \ldots, p_I\}$, 'skeletons' are our initial guesses of the probabilities of observing DLT at each dose level. The dose labels d_is are obtained by solving $p_i = F(d_i, \hat{\theta}_0)$.

For instance, suppose a study is designed to select a dose from 10, 20, 35, 50, 80 (mg/day). $p_3 = 40\%$ is the initially guessed DLT rate at the 3^{rd} dose level (35 mg/day). Using the empirical model $F(x;\theta) = x^{\exp(\theta)}$ with prior $\hat{\theta}_0 = 0$, we can obtain d_3 by solving $p_3 = F(d_3; \hat{\theta}_0) = d_3^{\exp(\hat{\theta}_0)} \Rightarrow d_3 = 0.4$. Let x_j denotes the dose that the j^{th} patient receives. For example, $x_2 = d_3$ means that the second enrolled patient is treated at the third dose level.

The posterior mean of θ after observing the outcomes from the first n patients can be estimated by

$$\hat{\theta}_n = \frac{\int \theta \mathcal{L}_n(\theta) g(\theta) d\theta}{\int \mathcal{L}_n(\theta) g(\theta) d\theta}, \tag{1.1}$$

where $\mathcal{L}_n(\theta)$ is the likelihood function:

$$\mathcal{L}_n(\theta) = \prod_{j=1}^{n} \{F(x_j, \theta)\}^{y_j} \{1 - F(x_j, \theta)\}^{1-y_j} \tag{1.2}$$

The original CRM design will treat the next patient at the current estimated MTD given all accrued patient data.

$$MTD = \arg\min_{x \in \mathcal{D}} |F(x, \hat{\theta}_n) - \xi|. \tag{1.3}$$

Certain dose-assigning restrictions may apply to override the estimated MTD to ensure patient safety. Commonly used restrictions include dose-skipping and coherence. Dose skipping is defined as assigning a patient to the i^{th} dose level before the $(i-1)^{th}$ dose level is tested. Coherence refers to avoiding dose escalation immediately after observing a DLT response from the current patient as well as de-escalation immediately after a non-DLT response. When properly calibrated [17], the CRM design can

be used in conjunction with algorithmic designs. For example, a trial may start with a rule-based design and switch to the CRM when the first DLT case is observed.

Previous studies have repeatedly shown that CRM designs have better accuracy in estimating the MTD and allocate more patients to the MTD than rule-based designs [18–20]. Rule-based designs often under-estimate the MTD while assigning a large portion of patients to sub-therapeutic dose levels. Despite these poor operating characteristics, '3+3' designs are still the most commonly utilized designs in phase I trials due to their long history of application and straightforward approaches. However, the use of model-based designs are gradually increasing [21, 22] as clinical investigators start realizing the statistical superiorities of model-based designs.

TITE-CRM: Traditional CRM is appropriate when the outcomes can be observed within a reasonable short follow-up interval since the design requires each patient to be completely followed before dose for the next patient can be assigned. When delays are expected for toxicity outcomes, the traditional CRM design may result in impractically long trial duration. Cheung and Chappell [23] proposed an extension to the CRM that incorporates partial information from patients who are still under observation for outcomes. This method is referred to as the time-to-event (TITE) CRM.

TITE-CRM incorporates partial follow-up information by using TITE outcomes, as opposed to traditional CRM that uses simple binary outcomes. Let $t_{j,n}$ denotes the follow-up time and $Y_{j,n}$ denotes the current outcome for the j^{th} patient by the time the $(n+1)^{th}$ patient is enrolled. Note that the value of $Y_{j,n}$ may change from 0 to 1 as n increases, indicating toxicity/efficacy effects manifest after certain time point. Toxicity is monitored by a weighted dose-toxicity model $\pi_Y(d_i, w) = G(d_i, \tau, \theta) = \tau F(\theta, d_i)$. The weight τ depends on the length of time a subject has been treated. If a toxicity response is observed, then the observation is given full weight $\tau_j = 1$; otherwise, the weight monotonically increases with the length of follow-up $t_{j,n}$. A commonly used

option is the linear weight function $\tau_j = t_{j,n}/T$. There are other options for weight function but this linear function is simple and performs well in simulations [23]. Using weighted power model, the likelihood function can be expressed as

$$\mathcal{L}^{\mathbf{w}}(\theta|\Omega_n) = \prod_{j=1}^{n} \{\tau_j(t_{j,n}, T) \cdot p_{x_j}^{\exp(\theta)}\}^{Y_{j,n}} \{1 - \tau_j(t_{j,n}, T) \cdot p_{x_j}^{\exp(\theta)}\}^{1-Y_{j,n}}.$$

Model estimation and dose assignment can be done similarly to the CRM. The posterior mean of model parameter can be calculated from the weighted likelihood function:

$$\hat{\theta}_n^{\mathbf{w}} = \frac{\int \theta \mathcal{L}_n^{\mathbf{w}}(\theta) g(\theta) d\theta}{\int \mathcal{L}_n^{\mathbf{w}}(\theta) g(\theta) d\theta}. \tag{1.4}$$

The MTD can be expressed as

$$MTD = \arg\min_{x \in \mathcal{D}} |F(x, \hat{\theta}_n^{\mathbf{w}}) - \xi|. \tag{1.5}$$

Other model-based designs: There are many other model-based designs available for Phase I oncology trials. For example, the escalation with overdose control (EWOC) [24] is essentially a modified CRM design with additional measures to reduce the chance of exposing patients to excessively high dose levels. Another example is the use of isotonic model [25] in Phase I designs. The isotonic model design differs from CRM designs by using non-parametric models to avoid misspecification of the dose-toxicity curve.

A recently proposed design combines advantages of rule-based and model-based designs. The Bayesian Optimal Interval (BOIN) design is easy to implement similar to the 3+3 design, but is more flexible for choosing the target toxicity rate and cohort size and yields a substantially better performance that is comparable to that of more complex model-based designs [26]. Unlike rule-based designs that estimate the MTD only using data from the most recent cohorts, the BOIN design uses isotonic regression to utilize data collected over the entire trial.

1.4 A new generation of cancer treatments and novel Phase I/II designs

Cytotoxic agents are drugs that achieve tumor shrinkage through non-selective attack of living cells, whereas cytostatic agents target tumor growth pathways without directly attacking cells. Because the mechanism of cytotoxic and cytostatic agents are fundamentally different, the paradigm of trial design needs to be adjusted. Traditionally, the dose-finding trials in Phase I aim to select the MTD, assuming the MTD is the most promising dose for therapeutic effect. For some cytostatic agents, higher doses will not increase and could possibly decrease efficacy after exceeding the intermediate dose [27].

Cytostatic agents are often used in combination with cytotoxic treatments in order to achieve optimal results. For example, angiogenesis is an important pathway in tumor growth, and upregulated VEGF/VEGFR expression is an established biomarker on this pathway. The VEGFR antibody DC101 in combination with vinblastine demonstrated full and sustained regression of large, established tumors in neuroblastoma xenograft models [28]. Several Phase III trials confirmed the effects of bevacizumab, an agent targeting VEGFA, when combined with platinum-based chemotherapy in patients with nonsmall cell lung cancer (NSCLC) or platinum-sensitive recurrent ovarian primary peritoneal, fallopian tube cancer [29–32].

Even though cytostatic agents are often less toxic than traditional chemotherapies, toxicity still needs to be closely monitored in early Phase trials. Severe adverse events are still possible and unpredictable before tested in patients. For example, unexpected severe toxicity effects in patients with MBC were observed when testing cetuximab (an effective EGFR antibody) in combination with cisplatin and topotecan [33, 34].

Therefore, trial designs involving cytostatic agents need to model monitor toxicity and efficacy at the same time, which are conventionally the objectives of two trial phases. Traditional designs may adopt the expansion cohort approach to collect additional efficacy data in phase I settings. Expansion cohorts are originally intended

to mitigate the limitation of small sample size in Phase I studies and verify the selected dose before moving on to Phase II. By enrolling and treating additional patients at the last recommended dose, more information regarding safety, efficacy, PD/PK can be obtained, and patient enrichment may be enhanced. The sample size of expansion cohort is not based on statistical considerations, and may vary from a few to dozens of patient. A review of 611 unique Phase I cancer trials found that about 25% of the trials included an expansion cohort, and trials were significantly more likely to use an expansion cohort if they were more recent, multi-center, or testing a non-cytotoxic agent [35]. However, this approach is not appropriate when dose-efficacy curves are not monotonic since the optimal dose may be under the MTD.

As an alternative to expansion cohort, a two-step approach can be adopted in which the MTD is located first and then search for OBD using efficacy endpoints. Hoering *et al.* suggested randomizing patients to three dose levels: the MTD, and the two nearby doses [36]. A similar approach by Yin *et al.* [37] is to identify toxicity upper bound of the searching range and determine OBD through dose de-escalation.

In recent years, there is a trend to integrate Phase I and Phase II trials into one seamless process. A seamless phase I/II trial can not only better address the issue of non-monotonic dose-efficacy curves, it also saves resources and accelerates the drug development process. Many new Phase I/II designs have been published [38–47]. A bivariate CRM design to incorporate efficacy was proposed by Braun. This design extended the original CRM to a bivariate CRM (bCRM) through a conditional probability model for efficacy and toxicity, but it was not designed to deal with non-monotonic dose-efficacy curves. Thall and Cook proposed an adaptive Bayesian dose-finding design based on a set of efficacy-toxicity trade off contours that partition the two-dimensional outcome probability domain to find the OBD. Bekele and Shen incorporates a continuous activity outcome with binary toxicity outcomes into the dose-finding design. Yin *et al.* jointly modeled toxicity and efficacy using a

bivariate binary model and considered the correlation between toxicity and efficacy. A design using continuation-ratio model was proposed by Zhang *et al.* to model toxicity and efficacy as a trinomial endpoint: no efficacy and no toxicity, efficacy without toxicity, and toxicity. Yuan and Yin proposed a design that jointly models toxicity and efficacy as time to event outcomes to address the issue of censoring at decision-making time. Wages and Tait proposed a seamless Phase I/II design that accounts for non-monotonic dose-efficacy relationships by fitting multiple efficacy models and using posterior model probabilities to select the best-fit efficacy model. This design combines features of the CRM design and order restricted inferences (Partial Order CRM) [48]. Riviere *et al.* employed a logistic model with a plateau parameter to capture the increasing-then-plateau feature of the dose-efficacy relationship. In addition, there are non-parametric approaches in the literature in order to avoid misspecification of the dose-toxicity curve. Gasparini and Eisele proposed to use a flexible product-of-beta prior in a curve-free model. Zang *et al.* introduced a non-parametric approach that uses the isotonic regression to estimate the OBD, and a semi-parametric approach by assuming a logistic model only around the current dose. A comprehensive overview of Bayesian designs for Phase I-II clinical trials is provided by Yuan, Nguyen, and Thall (2016) [49].

The research of seamless phase I/II designs is a recent topic and novel designs are still being developed. Although many designs have demonstrated better operating characteristics than traditional approaches, there is yet a universally accepted design that is widely used in practice. For a novel design to be frequently utilized in practice, it not only needs to demonstrate desirable statistical properties, but also needs to be easily understood by clinical investigators without backgrounds in statistics. CRM was the first model-based design and it finally begin to be accepted by the medical community. The method proposed by Wages and Tait [44] is essentially a design using multiple CRM-like models to account for different shapes of the dose-efficacy

curve. Therefore, it has the potential to be accepted by the medical community since the rationale behind this design can be easily understood by people who are familiar with CRM. Implementations of this design and its variations demonstrated the potentials [50, 51].

According to the method proposed by Wages and Tait [44], dose-toxicity and dose-efficacy relationship are modeled in parallel. Dose-toxicity is modeled using the traditional CRM from which a dose range with acceptable toxicity can be obtained. In order to account for non-monotonic dose-efficacy relationships, multiple working efficacy models are constructed based on a class of efficacy skeletons. Each of the working efficacy model is constructed in a way that is similar to the CRM. The best-fit working model is selected based on posterior model probabilities. The efficacy probabilities can be estimated from the selected working model and next doses are assigned based on the estimated efficacy probabilities. This method has demonstrated desirable operating characteristics through simulations.

However, there are a few potential improvements that can be incorporated into the original Wages and Tait design. First, the original Wages and Tait design constructs multiple CRM-like models to account for different shapes of the dose-efficacy curve, and uses a model selection criteria to choose a working efficacy model based on how well each model fits the observed data. When sample size is small, there is not enough data to conduct model selection. The original design proposed to randomize an arbitrary number of patients based on the current best-fit model. This is referred to as the adaptive randomization (AR) stage. We believe the AR stage could be improved because the number of patients to be randomized is arbitrary and the randomization probabilities are calculated from a working model that is selected based on small sample size. Second, as discussed in the original article, this design in its current form is most appropriate when both toxicity and efficacy endpoints can be observed in a similar time-frame. In practice, there is sometimes a delay in

the observation of endpoints, especially for efficacy. A TITE extension may be an effective solution to delayed outcomes. Third, in order for a method to be understood by other researchers and adopted in practice, there must be readily available software for simulations and applications.

In this dissertation, our overarching goal is to evaluate potential improvements and extensions to the seamless Phase I/II design proposed by Wages and Tait. In Chapter 2, we propose a redesign of the adaptive randomization stage used by the original design [44] and use simulations to show improvements. In Chapter 3, we extend the Wages and Tait design from binary endpoints to TITE endpoints. Simulations show that this extension may substantially reduce the total trial duration while maintaining approximately the same accuracy in selecting the OBD. In order to facilitate future research and promote practical application, we build and introduce an R package in Chapter 4 for the original Wages and Tait design as well as the modifications and extensions. Finally, in Chapter 5, we discuss the results and possible future directions of research.

Chapter 2 Improved Adaptive Randomization Strategies for a Seamless Phase I/II Dose-Finding Design

2.1 Abstract

In this chapter, we propose and evaluate three alternative randomization strategies to the adaptive randomization (AR) stage used in a seamless Phase I/II dose-finding design. The original design was proposed by Wages and Tait [44] for trials of molecularly targeted agents in cancer treatments, where dose-efficacy assumptions are not always monotonically increasing. Our goal is to improve the design's overall performance regarding the estimation of optimal dose as well as patients allocation to effective treatments. The proposed methods calculate randomization probabilities based on the likelihood of every candidate model as opposed to the original design which selects the best model and then randomizes doses based on estimations from the selected model. Unlike the original method, our proposed adaption does not require an arbitrarily specified sample size for the AR stage. Simulations are used to compare the proposed strategies and a final strategy is recommended. Under most scenarios, our recommended method allocates more patients to the optimal dose while improving accuracy in selecting the final optimal dose without increasing the overall risk of toxicity.

2.2 Introduction

Background

Historically, phase I trial designs have arisen from the development of chemotherapies starting in the 1940s [7]. A common objective of phase I trials in oncology is to find the maximum tolerated dose (MTD) beyond which the consequence of toxicity outweighs any potential benefits. For chemotherapies, the MTD happens to be the most promising dose for therapeutic benefits, because both odds of observing toxicity and efficacy responses increase with dose levels. This traditional approach aims to estimate the MTD of an agent, with subsequent phase II studies to evaluate efficacy at the estimated MTD. Currently, most trials in Phase I use algorithm-based designs, such as 3+3 escalation design [9, 52], step-wise design or A+B design [53] due to their simplicity and long history of implementation. Model-based designs such as the Continual Reassessment Method (CRM) [15] have shown superior statistical properties over algorithm-based designs [18], and the use of model-based designs in practice is increasing [21, 22]. After its original proposal, the CRM design gained popularity among biostatisticians and several CRM-related designs were proposed to handle more complex situations [16].

In recent years, the development of cytostatic cancer treatments such as molecularly targeted agents (MTAs) [54] have shifted this paradigm in oncology clinical trials. For MTAs, it is not always appropriate to assume that higher doses tend to result in higher probability of observing desired effects. Once an MTA reaches a certain concentration level, higher doses may not increase efficacy, and efficacy could possibly decrease after exceeding an intermediate dose [27]. In addition, MTAs are often used in combination with cytotoxic agents. Even though MTAs are considered safer than traditional chemotherapies, severe adverse events were observed in practice and can not be predicated before the trial [33,34]. Therefore, the primary goal of these Phase

I trials is to select the optimal biological dose (OBD) which is determined not only by toxicity but also efficacy. Examples of these joint outcomes include balancing toxicity with minimum effective blood concentration level of the agent, percent target inhibition of a biomarker, or minimum expression level of a molecular target (targeted biologic response) [55].

Locating the OBD

There is a trend in cancer treatment development to seamlessly integrate Phase I and Phase II in order to accelerate drug development process and reduce costs. For novel cancer treatments like MTAs, it becomes necessary to combine the objectives of conventional Phase I and II studies since the search for OBD requires toxicity and efficacy to be considered jointly. Several Phase I/II designs have been proposed to incorporate efficacy data into the early stage of drug development during the last decade, including a bivariate CRM design to incorporate efficacy [38], an adaptive Bayesian method for Phase I dose finding based on trade-offs between the probabilities of treatment efficacy and toxicity [39], a method used toxicity and efficacy odds ratios and accounted for the correlation between toxicity and efficacy [41], and a two-step Phase I/II trial sequence that identifies MTD using traditional methods in Phase I and then search for OBD around the MTD in the following Phase II trials [36] among others. In 2015, Wages and Tait proposed a seamless Phase I/II design that accounts for non-monotonic dose-efficacy relationships by fitting multiple efficacy models and using posterior model probabilities to select the best-fit efficacy model [44]. This design combines features of the CRM design and order restricted inferences [48]. This design sequentially updates safety and efficacy probabilities in order to allocate patients to the best guess of OBD from the accrued patient data.

The original Wages and Tait method models dose-toxicity and dose-efficacy relationship in parallel. Dose-toxicity is modeled using the traditional CRM from which

a dose range with acceptable toxicity can be obtained. In order to account for non-monotonic dose-efficacy relationships, multiple working dose-efficacy models are constructed based on a class of CRM skeletons. The best-fit model is selected based on posterior model probabilities and the best-fit model is used to estimate efficacy probabilities and assign dose for the next patients. Early in the trial, the observed sample size is too small to correctly select the best fit model, hence, the dose estimated from the selected model can not be relied on entirely. Some doses within the estimated acceptable toxicity range may have never been tested. Therefore, an AR stage is introduced to give every dose a chance to be tested. It also prevents the method from prematurely favoring a particular model. The original AR stage randomizes the first n_{AR} patients based on the probabilities of efficacy estimated from the selected dose-efficacy model. After the first n_{AR} patients, the trial enters a maximization stage, in which the patients are allocated to the most efficacious dose estimated from the selected model.

Motivation

In this chapter, we introduce a redesign to the adaptive randomization (AR) stage in the original design. In order to better illustrate our proposed revision, we will briefly describe the original design in the next section and then introduce our proposed adaptive randomization strategies.

We believe that the AR stage could be improved for several reasons. First, in the early stage of the trial, there are too few data to accurately select the best dose-efficacy model, therefore, the estimated efficacy probabilities are also not accurate. Second, it is likely that the original approach picks a different efficacy model with each new iteration, which leads to dramatic changes in the dose assigning probabilities. Third, the size of AR phase n_{AR} is arbitrarily selected, and a clear optimization of n_{AR} has yet to be identified. We propose a new method that replace the AR stage of

the original design by Wages and Tait with the aim of improving the designs overall performance. The method proposed below have several advantages. First, it does not require a best-fit model to be selected when sample size is small. Instead, we weight the suggested best dose from each candidate model and gradually reduce the number of candidate models based on the amount of data accumulated. Second, we do not need to arbitrarily specify a sample size n_{AR} for the AR stage. Also, it simplifies the design by merging several stages into one seamless process.

2.3 The original Wages and Tait dose-finding design

Consider a trial aimed at selecting the OBD of a new agent from a set of I pre-defined dose levels $\mathcal{D} = \{d_1, d_2, \ldots, d_I\}$. Toxicity and efficacy are observed through binary endpoints Y_j and Z_j, where $j = 1, 2, 3, \ldots, N$ is the index of participating patient ordered by sequence of enrollment.

$$Y_j = \begin{cases} 0, & \text{if not DLT} \\ 1, & \text{if DLT} \end{cases}, \text{and } Z_j = \begin{cases} 0, & \text{if no efficacy response} \\ 1, & \text{if observe efficacy response} \end{cases}.$$

The dose level to be administered on the j^{th} subject is denoted by $X_j \in \mathcal{D}$. After observing the responses from the first n subjects, the toxicity and efficacy data can be given in form of $\Omega_n = \{(x_1, y_1, z_1), (x_2, y_2, z_2), \ldots, (x_n, y_n, z_n)\}$

The probability of observing a DLT and efficacy response at each dose level is denoted by $\pi_Y(d_i)$ and $\pi_Z(d_i)$, respectively. We assume toxicity monotonically increases with dose level, and an acceptable safe dose range $\mathbb{A} = \{d_i | \pi_Y(d_i) < \xi\}$ can be obtained by the traditional CRM, where ξ is the maximum tolerated DLT rate. Conventionally, the threshold value ξ is set to be 33%. OBD is defined as the dose with the highest efficacy within the range that still assures safety

$$OBD = \arg\max_{d_i \in \mathbb{A}} \{\pi_E(d_i)\}.$$

Since the shape of the dose-efficacy curve is unknown, a class of $L = 2 \times I - 1$ working models can be constructed ($2I$ models have shapes that plateau or peak at each dose level minus one duplicate), according to the original design by Wages and Tait [44].

Let $\mathbf{Q}$ denotes a class of skeletons:

$$\mathbf{Q} = \begin{pmatrix} \mathbf{q_1} \\ \mathbf{q_2} \\ \vdots \\ \mathbf{q_L} \end{pmatrix} = \begin{pmatrix} q_{11} & q_{12} & q_{13} & \cdots & q_{1I} \\ q_{21} & q_{22} & q_{23} & \cdots & q_{2I} \\ & & \cdots & & \\ q_{L1} & q_{L2} & q_{L3} & \cdots & q_{LI} \end{pmatrix} \tag{2.1}$$

Each row $\mathbf{q}_\ell \in \mathbf{Q}$ is a set of skeletons, denoting a unique shape of the does-efficacy curves. Elements in of the matrix, $q_{\ell i} \in \mathbf{Q}$, are constants that represents our initial guesses of the probability of efficacy at dose level i, under the ℓ^{th} working model. For example, if there are three dose levels, we can construct $\mathbf{Q}$ as a 5 by 3 matrix:

$$\mathbf{Q} = \begin{pmatrix} 0.3 & 0.4 & \mathbf{0.5} \\ 0.4 & \mathbf{0.5} & 0.4 \\ \mathbf{0.5} & 0.4 & 0.3 \\ 0.4 & \mathbf{0.5} & \mathbf{0.5} \\ \mathbf{0.5} & \mathbf{0.5} & \mathbf{0.5} \end{pmatrix}$$

Using the ℓ^{th} row in $\mathbf{Q}$ and the empirical function, the ℓ^{th} efficacy model can be expressed as

$$\pi_E(d_i, \ell) = \Pr(z_j = 1 | d_i, \ell) \approx G_\ell(x_j; \beta_\ell) = q_{\ell i}^{\exp(\beta_\ell)}, \tag{2.2}$$

where $\hat{\pi}_E(d_i, \ell)$ is the estimated probability of observing a efficacy response on a patient tested at dose level d_i. Based on the observed efficacy data in Ω_n, the likelihood function can be expressed as

$$\mathcal{L}_\ell(\beta_\ell | \Omega_n) = \prod_{j=1}^{n} \{G_\ell(x_{j,\ell})\}^{z_j} \{1 - G_\ell(x_j, \beta_\ell)\}^{1 - z_j}. \tag{2.3}$$

The posterior density of β_ℓ is

$$P_\ell(\beta_\ell|\Omega_n) = \frac{\mathcal{L}_\ell(\beta_\ell|\Omega_n)h(\beta_\ell)}{\int_{\beta_\ell}\mathcal{L}_\ell(\beta_\ell|\Omega_n)h(\beta_\ell)d\beta_\ell}, \tag{2.4}$$

where $h(\beta_\ell)$ is the prior distribution of the model parameter β_ℓ.

In addition, the posterior model probabilities given the observed data can be established as

$$w(\ell|\Omega_n) = \frac{\tau_\ell \int \mathcal{L}_\ell(\beta_\ell|\Omega_n)h(\beta_\ell)d\beta_\ell}{\sum_{\ell=1}^{L}\tau_\ell \int \mathcal{L}_\ell(\beta_\ell|\Omega_n)h(\beta_\ell)d\beta_\ell}, \tag{2.5}$$

where τ_ℓ is the prior probability that the working model built on skeleton $\mathbf{q}_\ell$ is the best model to describe the dose-response relationship. For the rest of this chapter, we set $\tau_\ell = 1/L$, but τ_ℓ can be adjusted to incorporate any existing knowledge of the dose-efficacy relationship if available. Each time a new patient enters the trial, a single model ℓ^* with the largest $w(\ell|\Omega_n)$ is selected, so that

$$\ell^* = \arg\max\{w(\ell|\Omega_n)\},$$

from which the probability of efficacy response $\pi_E(d_i, \ell^*)$ is estimated.

Since $w(\ell|\Omega_n)$ is obtained from small samples, especially in early stage of the trial, we do not rely entirely on the $w(\ell|\Omega_n)$ in selecting the best fit model. The original design would randomize an arbitrary number of patients, n_{AR}, based on the estimated efficacy probability $\hat{\pi}_E(d_i, \ell^*)$ if the estimated risk of toxicity is acceptable. The original randomization probabilities are calculated as

$$R_i = \frac{\hat{\pi}_E(d_i, \ell^*)}{\sum_{d_i \in \mathbb{A}}\hat{\pi}_E(d_i, \ell^*)} \tag{2.6}$$

In the next section, we propose to randomize patients using $w(\ell|\Omega_n)$ to weight the implied best dose from each candidate model.

2.4 New adaptive randomization strategies

For $\ell = 1, 2, \ldots, L$, $\mathbf{q}_\ell \in \mathbf{Q}$ denotes the dose-efficacy skeleton over doses $1, 2, \ldots, I$. If a model was built using skeleton $\mathbf{q}_\ell$, then the dose level recommended by this model

can be expressed as $\mathcal{S}(\ell)$, that is

$$\mathcal{S}(\ell) = \min\{\arg\max_i(\hat{\pi}_E(d_i \in \mathbb{A}, \ell))\}. \tag{2.7}$$

The dose at level $\mathcal{S}(\ell)$ is expected to have the maximum chance of generating efficacy response if the shape of dose-efficacy model ℓ is similar to the true dose-efficacy relationship. The weight $w(\ell|\Omega_n)$ indicates the probability that the model built on skeleton $\mathbf{q}_\ell^{th}$ being the most consistent with the observed efficacy data. Therefore, we propose the following strategies to randomize the enrolled patient with probability R_i^*.

Strategy 1: We propose to randomize the first n_{AR} enrolled patients to the i^{th} dose level with probability R_i^* calculated as

$$R_i^* = \sum_{\ell=1}^{2I-1} w(\ell|\Omega_n)\mathbf{I}(d_i = \mathcal{S}(\ell)), \tag{2.8}$$

where $\mathbf{I}(\cdot)$ is an indicator function. Equation (2.8) considers the recommended dose from every candidate model and weights each recommendation by $w(\ell|\Omega_n)$, whereas the original design only considers the selected best-fit model. Similar to the original design, n_{AR} is arbitrarily selected and typically ranges from one third to one half of the total sample size [44].

Strategy 2: A potential flaw of strategy 1 is that some of the candidate skeletons and models are based on skeletons that have very different shapes than the true dose-efficacy curve. Therefore, in strategy 2, we only consider a subset of candidate models based on the posterior model probability $w(\ell|\Omega_n)$. By definition, $\sum_{\ell=1}^{L} w(\ell|\Omega_n) = 1$. As more data are observed, models that better represent the true does-efficacy are expected to have $w(\ell|\Omega_n) > 1/L$, indicating the data support some models while contradicting the others. Therefore, we only consider models with $w(\ell|\Omega_n) > 1/L$ and calculate randomization probabilities by

$$R_i^* = \frac{R_i^{**}}{\sum_{i=1}^{I} R_i^{**}}, \text{ where } R_i^{**} = \sum_{\ell=1}^{2I-1} w(\ell|\Omega_n)\mathbf{I}(d_i = \mathcal{S}(\ell) \text{ and } w(\ell|\Omega_n) \geq 1/L). \tag{2.9}$$

Models with $w(\ell|\Omega_n) < 1/L$ are temporarily excluded from the calculation of randomization probabilities, but all models will be reevaluated when additional data are observed. As n increases, we expect fewer models with $w(\ell|\Omega_n) > 1/L$. When n is sufficiently large, there is one and only one best-fit model satisfying $w(\ell|\Omega_n) > 1/L$. This strategy adaptively excludes less fit models and does not need n_{AR} to be specified. When maximum sample size is reached and there are more than one models satisfying $w(\ell|\Omega_n) > 1/L$, a best fit model $\ell^* = \arg\max_\ell\{w(\ell|\Omega_n)\}$ can be selected as the final dose-efficacy model, and the final dose selection is

$$\mathcal{S}(\ell^*) = \min\{\arg\max_i(\hat{\pi}_E(d_i, \ell^*)), \max(\mathbb{A})\} \tag{2.10}$$

Strategy 3: In this strategy, we reduce the number of models being considered based on the observed sample size.

$$R_i^* = \frac{R_i^{**}}{\sum_{i=1}^{I} R_i^{**}}, \text{ where } R_i^{**} = \sum_{\ell=1}^{2I-1} w(\ell|\Omega_n)\mathbf{I}(d_i = \mathcal{S}(\ell) \text{ and } w(\ell|\Omega_n) \geq w_{(L-L'+1)}). \tag{2.11}$$

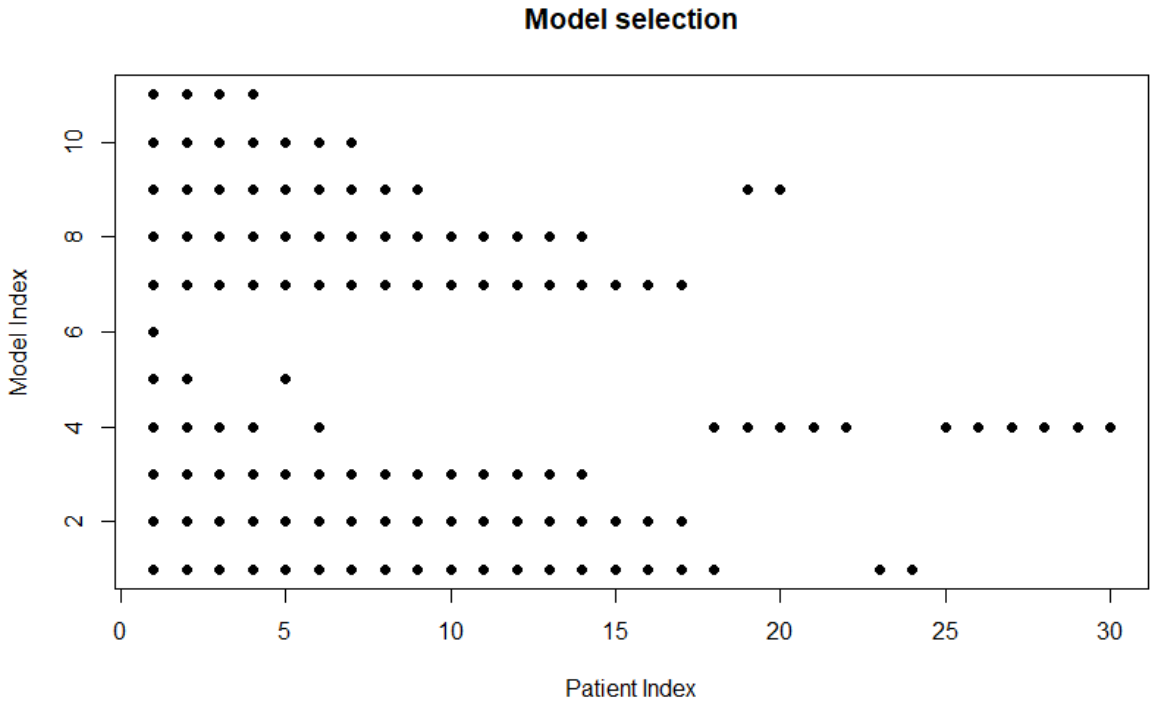
where $\mathbf{I}(\cdot)$ is an indicator function, and $w_{(1)} \leq w_{(2)} \leq \cdots \leq w_{(L)}$ denote the ordered posterior model probability $w(\ell|\Omega_n)$. Equation (2.11) considers the recommended dose from L' best-fit models and weights the recommendations of each model by $w(\ell|\Omega_n)$. When the suggested dose is outside the safety dose range, the highest dose in $\mathbb{A}$ will be used to ensure safety.

The number of models L' to be considered for patient randomization is gradually reduced from L to 1 based on the observed sample size n in relative to the planned maximum sample size N.

$$L' = \left\lceil \left(\frac{N-n}{N}\right)^{\delta} L \right\rceil, \tag{2.12}$$

where δ is a constant referred to as the drop rate parameter. When $\delta = 1$, one candidate model is excluded from calculating the randomization probability for each additional N/L patients. When $0 < \delta < 1$, models are dropped at a slower rate

Figure 2.1: An illustration of model selection process using a simulated trial of 30 patients. Drop rate $\delta = 2$. A solid dot indicates the corresponding candidate model is included in dose-assignment.



when sample size is small and faster when sample size is large, *vice versa* for when $\delta > 1$. When $L' = 1$, only the best fit model will be considered and the next enrolled patients will be randomized to the best dose recommended by the selected model with 100% probability. We illustrate how candidate models are eliminated using a simulated trial in Figure 2.1. Simulation studies are conducted to find appropriate values of δ.

2.5 Trial conduct

Dose-finding algorithm

Starting the Trial: There are several options to assign dose for the first patient. One may choose to start at the lowest dose level and use a simple escalation design

until the occurrence of DLT. Another option is to determine the acceptable range $\mathbb{A}$ from toxicity skeletons and calculate R_i^* with $w(\ell|\Omega_{n=0}) = 1/L$. For the rest of this chapter, we choose to start from the lowest dose under all circumstances.

Conducting the Trial: The trial conduct process varies slightly based on which randomization strategy is adopted. When strategy 1 is used, the first n_{AR} patients are randomized based on equation (2.8). This stage is referred to as the AR stage. The AR stage is followed by a maximization stage in which all patients are allocated to the estimated best dose from a selected efficacy model.

When strategy 2 or strategy 3 is used, the AR stage and maximization stage can be combined into one stage. Throughout the trial, all patients will be randomized based on equation (2.9) or (2.11). It also ensures that all patients are only randomized within the estimated safe range $\mathbb{A}$. In early stage of the trial, patients are randomized to a wide range of dose levels. However, the range of randomization is expected to be narrowed down as data accumulates.

Ending the Trial: The trial ends when a pre-defined maximum sample size N is reached. We can also terminate the trial early for safety when the acceptable range $\mathbb{A}$ is empty or for futility when the upper bound of an exact binomial confidence interval for efficacy is smaller than the standard treatment response rate at each dose level in the acceptable range. Since the number of observations on each dose level is small, conditions for early stopping rules are rarely met unless the true toxicity or efficacy is extremely different than initially expected. Early termination rules are included to avoid exposing patients to overly toxic or ineffective treatments.

Illustration

As an example, we describe a hypothetical Phase I/II dose-finding trial of the novel antiangiogenic peptide ATN-161. We choose to use this agent as an example because inverse U-shaped dose-response curve was observed in preclinical studies [56].

The possibility of non-monotonic dose-efficacy curve must be considered in designing human trials. ATN-161 is a fiveamino acid peptide that binds to several integrins, including $\alpha_5\beta_1$ and $\alpha_v\beta_3$, that play a role in angiogenesis and tumor progression [57]. It has been shown to inhibit tumor growth and metastasis and extend survival in multiple animal tumor models either when given as a single agent or when combined with chemotherapy [58, 59].

Suppose a phase I/II study of ATN-161 is conducted to select the OBD from a set of $I = 4$ dose levels, and the total sample size is $N = 30$. We assume monotonic dose-toxicity with toxicity skeleton set to be $\mathbf{p} = (0.1, 0.15, 0.2, 0.3)$. Since dose-efficacy could potentially be inverse U-shaped or plateau shaped, we construct a class of 7 working efficacy models $\mathbf{Q} = (\mathbf{Q_1}, \mathbf{Q_2})^T$ according to equation (4.1), with $\mathbf{Q_1}$ and $\mathbf{Q_2}$ respectively denoting peak and plateau shaped efficacy skeletons.

$$\mathbf{Q}_1 = \begin{pmatrix} 0.4 & 0.5 & 0.6 & 0.7 \\ 0.5 & 0.6 & 0.7 & 0.6 \\ 0.6 & 0.7 & 0.6 & 0.5 \\ 0.7 & 0.6 & 0.5 & 0.4 \end{pmatrix}, \text{and } \mathbf{Q}_2 = \begin{pmatrix} 0.5 & 0.6 & 0.7 & 0.7 \\ 0.6 & 0.7 & 0.7 & 0.7 \\ 0.7 & 0.7 & 0.7 & 0.7 \end{pmatrix}.$$

The trial is conducted using the algorithm described above, and suppose we observe the data summarized in Table 2.1 from the first 6 patients. The dose level for the next enrolled patient can be determined as following.

Table 2.1: Hypothetical data for demonstration

Dose Level	1	2	3	4
Number of Patients	1	1	2	2
Number of DLT Response	0	0	0	1
Number of Efficacy Response	0	0	1	1

Based on a standard CRM model, we estimate the probability of DLT is below 33% at dose level 4, so $\mathbb{A} = \{d_1, d_2, d_3, d_4\}$. Using equation (2.5), we calculate $w(\ell|\Omega_6)$ for each candidate model and display the results in Table 2.2. The original design will select model $\ell = 1$ as the current best fit efficacy model and randomize the next

patient solely based on efficacy probabilities estimated from this model. According to our proposed methods, we do not select the best fit model. Instead, we weight a subset of all candidate models by $w(\ell|\Omega_6)$.

Table 2.2: Numerical illustration of posterior model probabilities $w(\ell|\mathcal{D}_j)$ calculated using equation (2.5). Data source: Table 2.1.

k	1	2	3	4	5	6	7
$w(\ell\|\Omega_6)$	0.229	0.172	0.090	0.065	0.196	0.137	0.112

Dose level 4 is the suggested OBD by model $\ell = 1$; dose level 3 is suggested by model $\ell = 2$ and 5; dose level 2 is suggested by model $\ell = 3$ and 6; and dose level 1 is suggested by model $\ell = 4$ and 7. We weight the suggested OBD of each candidate model, and gradually reduce the number of models to be considered. Under strategy 1, all candidate models will be weighted. Under strategy 2, we only consider models with $w(\ell|\Omega_6) > 1/7$, that is, models $\ell = 3, 4, 6, 7$ will be excluded from calculating randomization probabilities. The number of models to be considered under strategy 3 depends on the drop rate parameter δ, the observed sample size n, and total sample size N. When the total sample size is $N = 30$ and the observed sample size is $n = 6$, we will consider the top five best fit models when $\delta = 2$, or the top four models when $\delta = 3$. For example, the probability of the next patient being randomized to the lowest dose d_1 is 0 when $\delta = 3$, since model $\ell = 4$ and $\ell = 7$ are excluded at this point. Randomization probabilities for each dose using each strategy are summarized in Table 2.3.

Table 2.3: Numerical illustration of dose-assigning probabilities calculated by the Original Wages and Tait method, and by the proposed method with $\delta = 2$ and $\delta = 3$.

Dose Level	1	2	3	4
Original Design	0.139	0.206	0.284	0.372
Strategy 1	0.176	0.227	0.368	0.229
Strategy 2	0	0	0.618	0.382
Strategy 3, $\delta = 2$	0.1323	0.162	0.435	0.271
Strategy 3, $\delta = 3$	0.000	0.187	0.501	0.312

.

2.6 Simulation studies

We conduct two sets of simulation studies: the tuning study and the operative characteristics study. In the tuning study, we evaluate the proposed method by varying the drop rate parameter δ, and decide on the most appropriate value. In the second set of simulations, we compare the proposed method to the original Wages and Tait design.

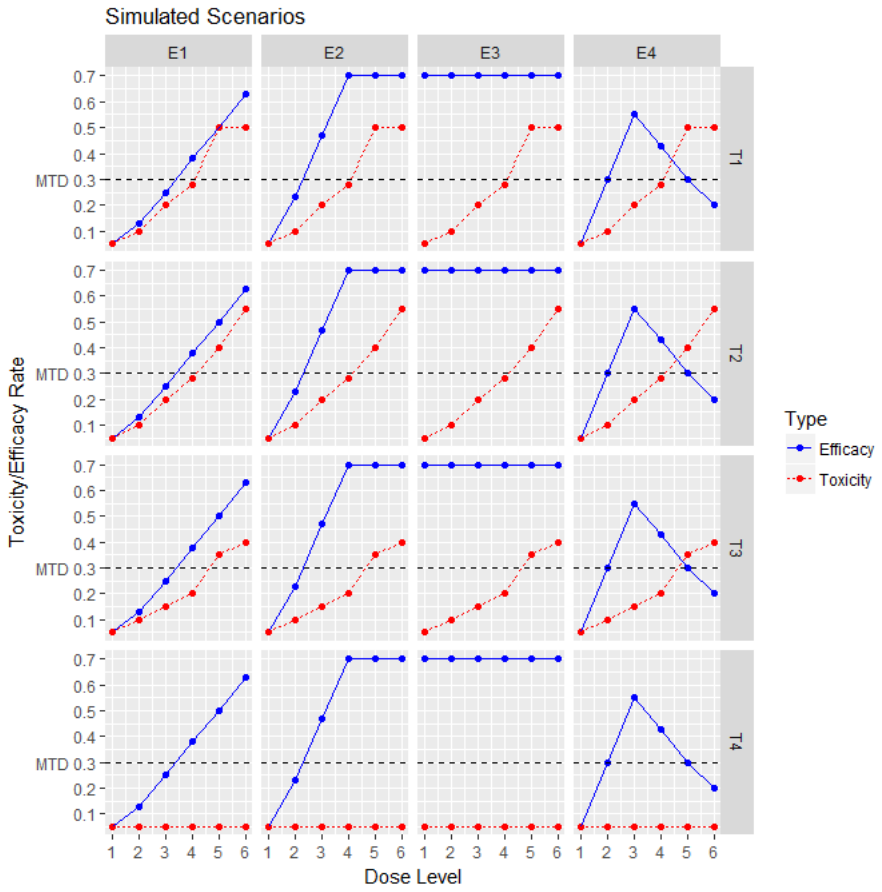
Simulation settings

Sixteen combinations of true dose-toxicity and dose-efficacy relationships are assumed in order to evaluate the performance of each design, which are summarized in Figure 2.2. Toxicity scenarios are denoted by T1 through T4 and efficacy scenarios are denoted by E1 through E4. E1-E4 are scenarios where the true dose-efficacy is increasing, plateaued, constant, or peaked. T1-T4 represent different toxicity profiles where the true MTD is at various dose levels. Scenarios with E1-E3 and T1-T4 were also used in simulation in the original Wages and Tait paper. We choose to use the exact same scenarios so that the results are directly comparable. This also validates our programs that utilize the original design. We include scenario E4 to assess the performance of each design when efficacy decreases after the optimal dose level.

Under each scenario, we simulate 1000 trials with each trial selecting a recommended dose from a set of $I = 6$ candidate doses. Using the definition used by the original study, a best dose is defined as the level that has maximum chance of efficacy while assuring safety, and a good dose is the level with 25% or higher chance of efficacy while assuring safety. We define safety as 33% or lower DLT rate. The performance of a design is evaluated by the probability of selecting good/best dose and the average number of patients treated at good/best dose.

For both methods, toxicity is modeled by traditional CRM using a power model with skeleton $\mathbf{p} = \{0.01, 0.08, 0.15, 0.22, 0.29, 0.36\}$. We justify the use of a single

Figure 2.2: True dose-toxicity and dose-efficacy curves in the simulation study. T1-T4 represent different toxicity levels. E1-E4 represent monotonic, plateaued, constant, and peaked dose-efficacy scenarios, respectively.



skeleton for all scenarios by robustness of the CRM method [60]. Efficacy is modeled with skeletons constructed using equation (4.1). The numerical values of efficacy skeletons are presented in Table 2.4. The same skeletons are used in all simulation scenarios for both methods to minimize potential bias in simulation studies.

Tuning δ for Strategy 3

Recall that in equation (2.12), the value of L' is controlled by the drop rate parameter δ. In this set of simulations, we compare the proposed design with difference choices of δ ranging from 0.5 to 4. In additional to different dose-toxicity and dose-efficacy scenarios, we also conduct simulations with different total sample size $N = 48, N =$

Table 2.4: Skeletons for efficacy models used in all simulation scenarios. There are 6 candidate dose levels, and 11 efficacy models are constructed to account for monotonic, plateaued or peaked dose-efficacy curves.

ℓ	Efficacy Model Skeletons
1	$\mathbf{q_1} = (0.60, 0.50, 0.40, 0.30, 0.20, 0.10)$
2	$\mathbf{q_2} = (0.50, 0.60, 0.50, 0.40, 0.30, 0.20)$
3	$\mathbf{q_3} = (0.40, 0.50, 0.60, 0.50, 0.40, 0.30)$
4	$\mathbf{q_4} = (0.30, 0.40, 0.50, 0.60, 0.50, 0.40)$
5	$\mathbf{q_5} = (0.20, 0.30, 0.40, 0.50, 0.60, 0.50)$
6	$\mathbf{q_6} = (0.10, 0.20, 0.30, 0.40, 0.50, 0.60)$
7	$\mathbf{q_7} = (0.20, 0.30, 0.40, 0.50, 0.60, 0.60)$
8	$\mathbf{q_8} = (0.30, 0.40, 0.50, 0.60, 0.60, 0.60)$
9	$\mathbf{q_9} = (0.40, 0.50, 0.60, 0.60, 0.60, 0.60)$
10	$\mathbf{q_{10}} = (0.50, 0.60, 0.60, 0.60, 0.60, 0.60)$
11	$\mathbf{q_{11}} = (0.60, 0.60, 0.60, 0.60, 0.60, 0.60)$

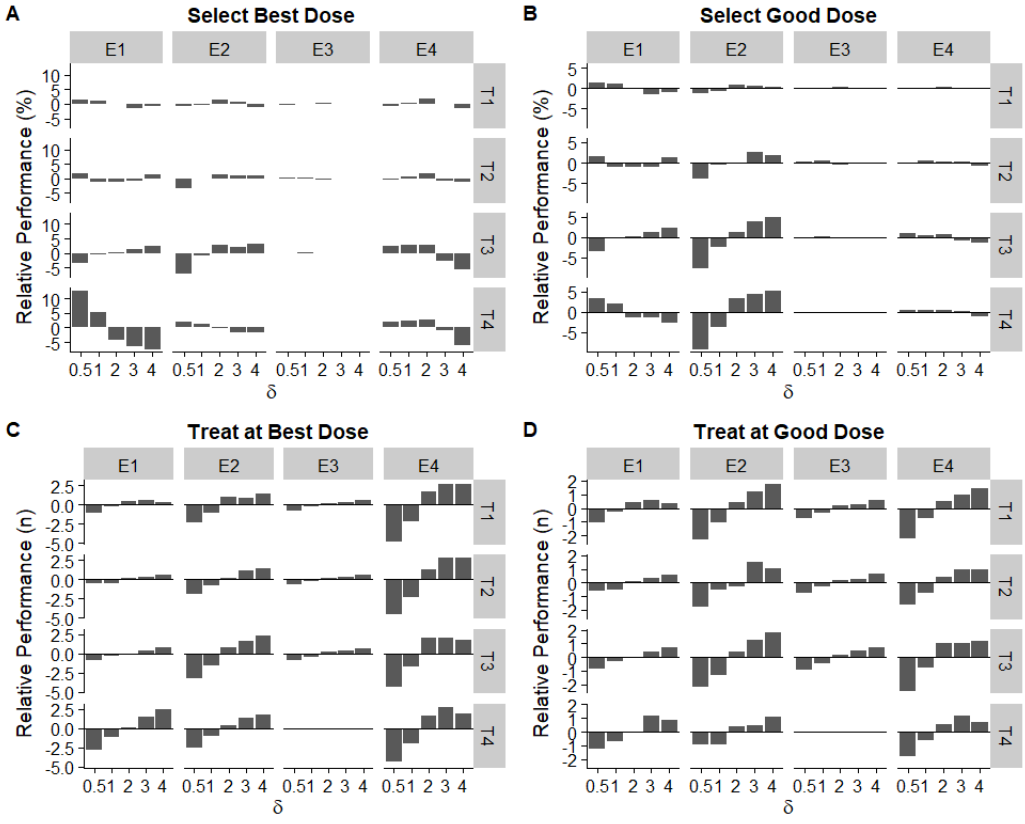
64 and $N = 80$. We only present the results with $N = 64$, as similar conclusions can be drawn for different values of N.

We summarize the results in Figure 2.3 which consists of four plots, A, B, C, and D, each representing a different criteria of performance. Full simulation results are provided in Appendix. Figure 2.3-A and 2.3-B respectively assess each value of δ by the chance of selecting best dose or a good dose. In Figure 2.3-C and 2.3-D, different values of δ are evaluated by the average number of patients allocated to best/good dose levels. Relative performance is calculated as the difference between a particular value of δ and the average performance of all five values of δs.

Figure 2.3-A and 2.3-B show that $\delta = 2$ and $\delta = 3$ perform better than the other values in most scenarios with E2 and E4. All methods perform similarly in scenarios with E3. Starting from $\delta = 4$, we begin to see decreases in accuracy of selecting the right dose in scenario E1T4. $\delta = 0.5$ and $\delta = 1$ perform exceptionally well when dose-efficacy is monotonically increasing and toxicity is constant at a very low level. However, their performance is relatively inferior in scenarios with E2. Figure 2.3-C and 2.3-D show that $\delta = 2$, and $\delta = 3$ allocate significantly more patients to either the best dose or a good dose throughout the trial under all twelve scenarios.

In conclusion, we find $\delta = 2$ and $\delta = 3$ have better performance under most scenarios except for R1T4. In E1T4, the performance of $\delta = 2$ and $\delta = 3$ are slightly inferior than $\delta \leq 1$. Therefore, we recommend $\delta = 2$ or $\delta = 3$ when the total sample size is between 48 and 80 which is a commonly used sample size in phase I/II studies. With $2 \leq \delta \leq 3$, it only takes a few of observations to exclude a model in the early stage and most patients are randomized based on several similar models.

Figure 2.3: Plots A, B, C, D respectively evaluate different values of δ in strategy 3 by the probability of selecting the best dose, probability of selecting a good dose, average number of patients treated at the best dose and average number of patients treated at a good dose. Relative performance is calculated as the difference in performance between a particular value of δ and the overall average performance of all δs. Higher number indicates better than performance.



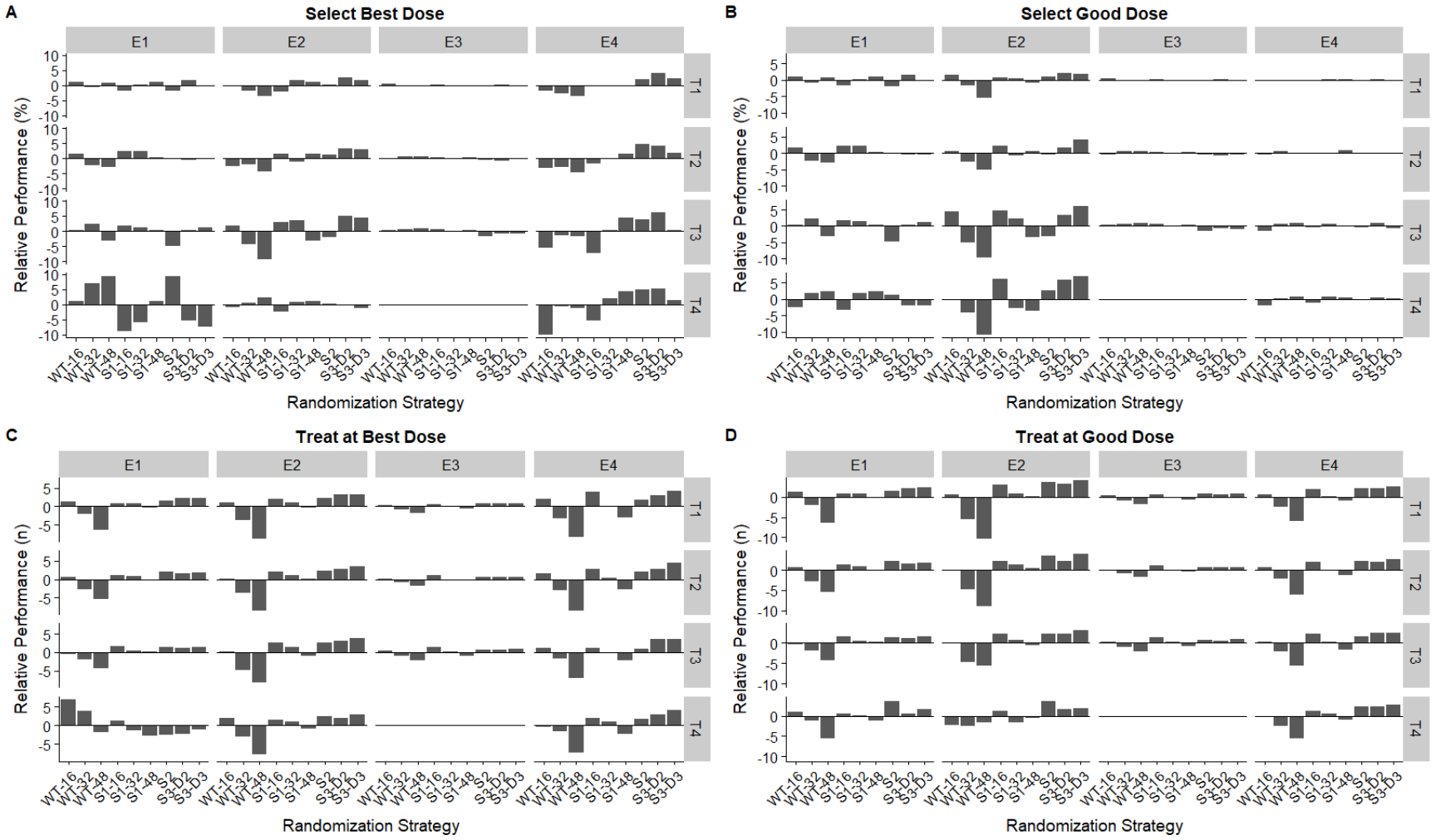
Operating characteristics

Another set of simulations compare the proposed randomization strategies to the original Wages and Tait design. The original Wages and Tait design chose to use $n_{AR} = 16$ as the sample size for the AR stage. We also include $n_{AR} = 32$ and $n_{AR} = 48$ for comparison. We denote the original designs by 'WT' followed by n_{AR}, that is, 'WT-16','WT-32', and 'WT-48'. Strategy 1 is simulated using $n_{AR} = 16, 32,$ and 48, which are denoted by 'S1-16', 'S1-32', and 'S1-48', respectively. Strategies 2 and 3 do not require n_{AR} to be specified. Strategy 3 uses $\delta = 2$ or $\delta = 3$ as suggested by the previous set of simulations, and we denote them by 'S3-D2' and 'S3-D3'. The total sample size for all designs is set to $N = 64$. Toxicity and efficacy outcomes are simulated as correlated variables with correlation parameter equals to 4.6 in order to match with settings of the original study. We also conducted simulations using other correlation parameter values, including independent toxicity and efficacy outcomes. For the sake of simplicity, we omit those simulations since the results are very similar.

Similar to the previous set of simulations, we report the results in Figure 2.4-A through Figure 2.4-D, and provide the full results in the Appendix. Strategy 1 with various n_{AR} occasionally performs better than the original method by a very small margin and sometimes performs worse. Hence, there is not evidence to recommend strategy 1. Strategies 2 and 3 outperform the original method by a higher probability in selecting best/good dose and a larger number of patients treated at the best/good dose in all but one scenario. In scenario E1T4, the original design with $n_{AR} = 48$ has better accuracy in selecting the best dose and allocates more patients to the best dose when $n_{AR} = 16$. However, we argue that the performance of strategies 2 and 3 is still acceptable. Our methods tend to select and allocate patients to be the best dose as well as the next lower dose. When the criteria is selecting/allocating to a good dose, Figure 2.4-B and 2.4-D show that the performance of the proposed design is very close to that of the original in E1T4.

Relative to strategy 2, strategy 3 is marginally more accurate in selecting the best/good dose, and both strategies have similar performance regarding treating patients at the best/good dose. Therefore, we recommend using strategy 3 with $\delta = 2$ or 3 as the method of randomization.

Figure 2.4: Plots A, B, C, D respectively evaluate different values of δ by the probability of selecting the best dose, probability of selecting a good dose, average number of patients treated at the best dose and average number of patients treated at a good dose. Relative performance is calculated as the difference in performance between a particular method and the overall average of all methods. Higher number indicates better than performance.



2.7 Discussion

In this chapter, we introduced and evaluated three alternative randomization strategies as modifications to a design proposed by Wages and Tait (2015). The original design requires a selection of the best-fit model even when the observed sample size is very small. In our final recommended randomization strategy, we eliminate the need to select the best model by evaluating and weighting every candidate model and gradually excluding unfit models as data accumulate. The revised design not only demonstrated better accuracy in selecting the OBD, but also allocated more patients to better treatment regimens. In addition, the original design includes two different stages while the proposed method simplifies it into one seamless process.

A limitation of the proposed design is the assumption that toxicity and efficacy data can be observed shortly after treatment. In real clinical practice, delayed responses are very common,as toxicity and may not be observed at the same time. Since time to event is not considered in this design, we can only enroll the next patient when both toxicity and efficacy outcomes are observed to update the models. Future studies can be conducted to include treatment cycles or length of treatment in order to shorten the total trial duration and make the design more applicable in real clinical settings with delayed responses.

Acknowledgment

This research was supported by the Biostatistics and Bioinformatics Shared Resource of the University of Kentucky Markey Cancer Center (P30CA177558), NIH National Center for Advancing Translational Sciences (UL1TR001998 and UL1TR00117), and the National Cancer Institute (K25CA181638).

Chapter 3 Bivariate Generalization of the Time-to-Event Continual Reassessment Method

3.1 Abstract

This chapter considers the problem of designing Phase I-II clinical trials with delayed outcomes. This design is motivated by is a Phase I-II study evaluating optimal dosing of all-trans retinoic acid (ATRA) in combination with a fixed dose of daratumumab in the treatment of relapsed or refractory multiple myeloma. The toxicity endpoint is observed in one cycle of therapy (*i.e.*, 4 weeks) while the efficacy endpoint is assessed after 8 weeks of treatment. The difference in endpoint observation windows causes logistical challenges in conducting the trial, since it is not practical to wait until both outcomes for each participant have been observed before sequentially assigning the dose of a newly eligible participant. In order to avoid delays in treatment for newly enrolled patients and to accelerate trial progress, we generalize the time-to-event continual reassessment method (TITE-CRM) to bivariate outcomes. Simulation studies are conducted to evaluate the proposed method, and we found that the proposed design substantially reduces the total trial duration with promising operating characteristics. However, the number of patients treated at the correct dose is affected by the rate of enrollment.

3.2 Introduction

Background

Historically, the primary objective of Phase I clinical trials is to identify the maximum tolerated dose (MTD) of the agent or agents being investigated. In a subsequent Phase II trial, the agent is evaluated for efficacy, often at the recommended dose (MTD). In oncology trials of chemotherapeutic agents, identification of the MTD is usually determined by considering dose-limiting toxicity (DLT) information only, with the assumption that the MTD is the highest dose that satisfies some safety requirement, so that it provides the most promising outlook for efficacy. In general, the design of Phase I trials is driven by the assumption of monotone increasing dose-toxicity and dose-efficacy relationships. Numerous Phase I designs have been proposed for identifying the MTD by studying the dose-limiting toxicity (DLT) information through the use of binary outcomes, including [52] [15] among many others.

By contrast, many biological agents are assumed safe overall and higher doses do not necessarily produce greater efficacious response. However, we must still monitor for the unexpected and account for safety. Conditional on safety, other endpoints may be the main summary endpoint that is used to determine which dose to carry forward. Examples include an early measure of efficacy (*i.e.* clinical response); pharmacokinetic/pharmacodynamics; biological targets (*i.e.* immune response). Dose-efficacy relationships may exhibit non-monotone increasing patterns, such as increasing at low doses and plateauing at higher levels, or peaking at an intermediate dose. For example molecules with anti-angiogenic activity often appear to exhibit hormesis, *i.e.,* bell-shaped dose-response curves [27]. A review of 24 Phase I targeted therapy trials show that patient receiving lower doses do not fare worse [61]. A lower dose than the MTD may exhibit as much activity as higher doses, and beyond this dose we are merely adding toxicity. If the dose-efficacy relationship is monotone increasing, the

MTD is the lowest safe dose providing the highest efficacy. In this case, we would want a dose-finding method to be able identify the MTD. However, if the dose-efficacy relationship plateaus at a dose lower than the MTD, we would want to recommend this lower dose. The goal of the trial shifts to identifying the optimal biologic dose (OBD), which is defined as the lowest dose with acceptable toxicity that maximizes efficacious response. In recent years, there have been several new methods proposed for locating the OBD in Phase I-II trials of biological agents, including [42], [38], [39], [44], among many others. A comprehensive overview of Bayesian designs for Phase I-II clinical trials is provided by Yuan, Nguyen, and Thall [49].

The design by Wages and Tait [44] is most appropriate when both binary toxicity and binary efficacy endpoints can be observed in a reasonably similar and short time-frame. A drawback of that method is that it requires efficacy and toxicity responses to be fully observed before it can be used to assign a dose to the next enrolled patient, Yin [37] and Yuan [43] note that in some practical situations, this may not be possible due to the fact that efficacy may occur much later than toxicity. For example, in a targeted agent or immuno-therapy trial, toxicity outcomes can often be observed in a relatively timely manner after treatment administration but efficacy is observed in a relatively longer time-frame. If this delay is expected, then the method outlined in Wages and Tait [44] is not optimal in its existing form because the trial would have to either pause before each patient is enrolled in order to fully observe the efficacy responses or to assign doses base on less efficacy data than toxicity data. If the delay is particularly long, then it will cause the duration of the trial to be much too long and wastes resources [43]. However, a modified approach can be applied to account for the delayed efficacy response. In this chapter, we outline a method for identifying the OBD using partial follow-up information that combines features of the continual reassessment method (CRM) [15], partial order CRM [48], and the time-to-event (TITE) CRM [23]. The rest of this chapter is organized as follows.

In the following sections, we first outline the statistical models and inference used in the proposed design, and describe the dose-finding algorithm. Then, we provide numerical results illustrating the operating characteristics of the design. We also provide a case study to demonstrate the design. Finally, we conclude with a discussion.

Locating the OBD

Consider a trial aimed at selecting the OBD of a new agent from a set of I pre-defined dose levels $\mathcal{D} = \{d_1, d_2, \ldots, d_I\}$. Toxicity and efficacy are observed through time dependent binary endpoints $Y_{j,n}$ and $Z_{j,n}$, where $j = 1, 2, 3, \ldots, N$ is the index of participating patients ordered by sequence of enrollment, and n denotes the number of patients currently being tested. Denote the toxicity probability at each dose level by $\pi_Y(d_i)$, and the efficacy probability at each dose level by $\pi_E(d_i)$. Based on the pre-specified target toxicity upper bound ξ, we want to exclude overly toxic doses, defining a set of acceptable (safe) doses as $\mathbb{A} = \{d_i | \pi_Y(d_i) < \xi\}$. For the rest of this chapter, we set $\xi = 33\%$. The primary objective of the study is to identify the OBD, defined the dose that maximizes efficacy, conditional on safety so that:

$$\text{OBD} = \arg\max_{d_i \in \mathbb{A}} \{\pi_E(d_i)\}. \tag{3.1}$$

The goal, both within and at the conclusion of the study, is to locate the OBD while some patients are still under observation for toxicity and efficacy responses.

A motivating example

This work is motivated by a Phase I-II clinical trial studying the optimal dosing of all-trans retinoic acid (ATRA) [62] in combination with a fixed dose of daratumumab [63] in the treatment of relapsed or refractory multiple myeloma. The trial was designed to find the OBD from among three dose levels of ATRA $\{15, 30, 45\,\text{mg/m}^2\}$ and a fixed

dose of daratumumab 16 mg/kg. The decision endpoints are dose-limiting toxicities (DLTs), based on protocol-specific adverse event definitions, in one cycle (*i.e.*, 28 days) of therapy, and early measures of efficacy, defined by at least a 25% reduction in M-protein in serum, 50% reduction in Bence-Jones proteinuria, or 25% reduction in plasmacytomas in patients with non-secretory disease after 8 weeks of treatment. To address the problem of delayed efficacy, we propose a time-to-event (TITE) extension to the method of Wages and Tait, which was only able to handle bivariate binary outcomes that were observable in a reasonably similar time frame. Other Phase I-II methods that are able to handle delayed outcomes include [43]; [64]; and [45].

3.3 Proposed time to event extension

We propose a TITE extension to the Wages and Tait [44] method to incorporate the use of partial follow-up information. Similar to the TITE CRM proposed by [23], the general idea is to incorporate all available information including the length of time a patient is on treatment into the model.

Let $t_{j,n}$ denotes the follow-up time for patient j by the time the $(n+1)^{th}$ patient is enrolled. The maximum follow-up time is denoted by T. Note that toxicity and efficacy outcomes may have different maximum follow-up time. The observed toxicity and efficacy response, $Y_{j,n}$ and $Z_{j,n}$ are dependent on $t_{j,n}$. Note that the value of $Y_{j,n}$ and $Z_{j,n}$ may change from 0 to 1 as n increases, indicating toxicity/efficacy effects manifest after certain time point. For each patient, we assume there are unobservable thresholds t_j^T and t_j^E, only beyond which a positive toxicity or efficacy response can be observed.

$$Y_{j,n} = \begin{cases} 0 \text{ (no DLT)} & \text{when } t_{j,n} < t_j^T \\ 1 \text{ (observe DLT)} & \text{when } t_{j,n} \geq t_j^T \end{cases}, Z_{j,n} = \begin{cases} 0 \text{ (no efficacy)} & \text{when } t_{j,n} < t_j^E \\ 1 \text{ (observe efficacy)} & \text{when } t_{j,n} \geq t_j^E \end{cases}$$

Dose level administered on the j^{th} subject is denoted by $X_j \in \mathcal{D}$. At the enrollment

of the $(n+1)^{th}$ patient, data observed from the first n^{th} patients can be expressed in form of $\Omega_n^T = \{(x_j, y_{j,n}, z_{j,n}, t_{j,n})\}$ for $j = 1, 2, \ldots, n$.

Models and inference for toxicity

Toxicity is modeled using TITE-CRM as proposed by Cheung and Chappell [23]. We only briefly describe the CRM and TITE-CRM here. The CRM is built on a given set of constants, called the skeletons $\mathbf{p} = (p_1, p_2, p_3, \ldots, p_I)$, with p_i representing our initial guesses of the toxicity risk at dose d_i. Toxicity is modeled by a parametric function $\pi_Y(d_i) = F(\theta, d_i)$. $F(\cdot)$ must be monotone increasing given any p_i. For example, the power model $F(\theta, d_i) = p_i^{\exp(\theta)}$, is a common choice.

When toxicity responses can be observed within a reasonable time-frame, the toxicity outcome for each subject j is simplified into a binary endpoint Y_j that is independent of time. In the case of a power model, the likelihood function given observed data Ω_n is

$$\mathcal{L}(\theta|\Omega_n) = \prod_{j=1}^{n} \{p_{x_j}^{\exp(\theta)}\}^{Y_j} \{1 - p_{x_j}^{\exp(\theta)}\}^{1-Y_j}.$$

$\hat{\theta}$ can be estimated from the posterior distribution or by maximizing the likelihood function. The probability of DLT at each dose level is estimated by $\hat{\pi}_Y(d_i) = F(\hat{\theta}, d_i)$.

TITE-CRM is an extension of the CRM by considering a weighted dose response model $\pi_Y(d_i, w) = G(d_i, \tau, \theta) = \tau F(\theta, d_i)$. The weight τ depends on the length of time a subject has been treated. If a toxicity response is observed, then the observation is given full weight $\tau_j = 1$; otherwise, define weight for subject j as $\tau_j = t_{j,n}/T$. There are other options for weight function, but for the purpose of this chapter, we choose to use this linear weight function. Using weighted power model, the likelihood function can be expressed as

$$\mathcal{L}^{\mathbf{w}}(\theta|\Omega_n) = \prod_{j=1}^{n} \{\tau_j(t_{j,n}, T) \cdot p_{x_j}^{\exp(\theta)}\}^{Y_{j,n}} \{1 - \tau_j(t_{j,n}, T) \cdot p_{x_j}^{\exp(\theta)}\}^{1-Y_{j,n}}.$$

Model estimation and dose assignment can be done similarly to the CRM.

Models and inference for efficacy

As in original design by Wages and Tait [44], efficacy is modeled by constructing multiple working models for efficacy and utilizes model selection to allow for the uncertainty in the dose-response curve. Both unimodal and plateau skeletons for a total of $L = 2I - 1$ models that are included in the set of possible models. For example, if $I = 6$, then we can construct a set of 11 working models as specified in equation (3.8). For a particular working skeleton $\ell, \ell = 1, 2, \ldots, 2I - 1$, there is a model

$$\pi_E(d_i) = \Pr(Z_j = 1|d_i) \approx G_\ell(d_i; \beta_\ell) = q_{i\ell}^{\exp(\beta_\ell)} \tag{3.2}$$

from a class of working does-efficacy models $G_\ell(d_i; \beta_\ell)$. Uncertainty in dose-response curve is taken into account through the posterior model probabilities of the efficacy models.

Given Ω_n^T, the weighted likelihood for efficacy model ℓ at the enrollment of the $n = 1^{th}$ patient can be expressed as

$$\mathcal{L}_\ell^{\mathbf{w}}(\beta_\ell|\Omega_n) = \prod_{j=1}^{n} \{\tau_j(t_{j,n}, T) \cdot q_{\ell x_j}^{\exp(\beta_\ell)}\}^{Z_{j,n}} \{1 - \tau_j(t_{j,n}, T) \cdot q_{\ell x_j}^{\exp(\beta_\ell)}\}^{1-Z_{j,n}}. \tag{3.3}$$

where the weight function $\tau^E(t_{j,n}, T)$ denotes the weight given to the outcome observed from the j^{th} patient by the time n patients have received the treatment. T is a constant denoting the observation window of the outcomes. $\tau(t_{j,n}, T)$ represents the proportion of follow-up time completed. We choose to use the linear weighting scheme, that is,

$$\tau(t_{j,n}, T) = \begin{cases} \arg\min(t_{j,n}/T, 1) & \text{if no response observed at } t_{j,n} \\ 1 & \text{if response observed at } t_{j,n} \end{cases}$$

$\tau^(t_{j,n}, T) = 1$ when $t_{j,n} \geq T$, denoting patient j has completed efficacy observation.

For the likelihood function of each candidate model ℓ, we proceed by evaluating the likelihood of the model with posterior model probability give by:

$$w(\ell|\Omega_n) = \frac{p(\ell)\int_{\beta_\ell}\mathcal{L}_\ell^{\mathbf{w}}(\beta_\ell|\Omega_n)h(\beta_\ell)d\beta_\ell}{\sum_{\ell=1}^{L} p(\ell)\int\mathcal{L}_\ell^{\mathbf{w}}(\beta_\ell|\Omega_n)h(\beta_\ell)d\beta_\ell}, \tag{3.4}$$

where $p(\ell)$ is the prior model probabilities. One option is to set $p(\ell) = 1/L$ for all ℓs so that every model is equally likely at the beginning of the trial, or we may set $p(\ell) = 0$ for some models if the drug mechanism indicates that such dose-efficacy relationships are not plausible. Each time a new patient enters the trial, all candidate models will be evaluated by their likelihood of representing the true dose-efficacy relationship. As proposed in the multidimensional CRM, we can choose model ℓ^* with the largest posterior probability such that $\ell^* = \arg\max\{w(\ell|\Omega_n)\}$ and estimate the probabilities of efficacy at each dose through

$$\hat{\pi}_E(d_i, \ell^*) = G_{\ell^*}(d_i, \hat{\beta}_{\ell^*}|\Omega_n),$$

where

$$\hat{\beta}_{\ell^*} = \frac{\int_{\beta_{\ell^*}}\beta\mathcal{L}_{\ell^*}^{\mathbf{w}}(\beta_{\ell^*}|\Omega_n)h(\beta_{\ell^*})d\beta_{\ell^*}}{\int_{\beta_{\ell^*}}\mathcal{L}_{\ell^*}^{\mathbf{w}}(\beta_{\ell^*}|\Omega_n)h(\beta_{\ell^*})d\beta_{\ell^*}}.$$

In the original Wages and Tait method [44], patients will be assigned to the dose with highest estimated efficacy probability. When sample size is very small, that is, when $n < n_{AR}$ with n_{AR} being a prespecified constant, patients will be randomized within the safe range $\mathbb{A}$ based on the estimated efficacy probabilities. Randomization of the first n_{AR} patients is referred to as the adaptive randomization stage.

In Chapter 2, we showed that an alternative randomization strategy tends to perform better regarding the accuracy of optimal dose selection and patients allocation to optimal doses. This alternative randomization strategy also eliminates the need of specifying n_{AR}. Therefore, for our proposed method, we choose to use the following randomization strategy.

The implied best dose from each candidate model can be given as

$$\mathcal{S}(\ell) = \min\{\arg\max_i(\pi_E(d_i \in \mathbb{A}, \ell))\}. \tag{3.5}$$

Therefore we can adaptively randomize patients based on the likelihood of each model and its corresponding best dose $S(\ell)$. The randomization probability is calculated as

$$R_i^* = \frac{R_i^{**}}{\sum_{i=1}^{I} R_i^{**}}, \text{ where } R_i^{**} = \sum_{i}^{2I-1} w(\ell|\Omega_n)\mathbf{I}(d_i = \mathcal{S}(\ell) \text{ and } w(\ell|\Omega_n) \geq w_{(L-L'+1)}). \tag{3.6}$$

where $\mathbf{I}(.)$ is an indicator function. $w_{(1)} \leq w_{(2)} \leq \cdots \leq w_{(L)}$ denote the ordered posterior model probability $w(\ell|\Omega_n)$. L is the total number of candidate models and L' is the number of models being considered in the calculation of randomization probabilities. Equation (3.6) considers the recommended dose from L' best-fit models and weights the recommendations of each model by $w(\ell|\Omega_n)$. Initially, there are $L = 2I - 1$ candidate models. Some models are mutually exclusive. Therefore, instead of considering all candidate models, we reduce the number of models being considered based on the observed sample size in relative to the total sample size N:

$$L' = \left\lceil \left(\frac{N-n}{N}\right)^{\delta} L \right\rceil, \tag{3.7}$$

where δ is a prespecified constant. As shown Chapter 2, we recommend to use $\delta = 2$ or $\delta = 3$ when total sample size is between 32 and 64. This allows the method to drop candidate models with fewer observations in the beginning stage of the trial, but more data are required to exclude a model towards the end.

Dose-finding algorithm

Starting the Trial: In order to get the trial underway, we will choose the efficacy skeleton with the largest prior probability, $p(\ell)$, among the orders being considered.

If several, or all, of the models have the same maximum prior probability, then there are several options to assign the dose for the first patient. One may choose to start at the lowest dose level or the prior-estimated MTD. Another option is to determine the acceptable range $\mathbb{A}$ from toxicity skeletons and the first patient will be randomized with probability R_i^* assuming $w(\ell|\Omega_{n=0}) = 1/L$. For the purpose of this study, we choose to start from the lowest dose.

Conducting the Trial: After the first patient, we update the data and re-fit the model each time another subject is enrolled into the study. Throughout the trial, all patients will be randomized based on equation (3.6). Our method ensures that all patients are only randomized within the estimated safe range $\mathbb{A}$. In early stage of the trial, patients are randomized to a wide range of dose levels. However, the range of randomization is expected to converge to the optimal dose as data accumulates.

Ending the Trial: The trial ends normally when a pre-defined maximum sample size N is reached. If the acceptable range $\mathbb{A}$ is empty during the trial, the study will be closed for safety. We may also close the trial early for futility or efficacy based on the exact binomial confidence interval for efficacy.

3.4 Simulation studies

Design specifications and model priors

In order to examine the operating characteristics of the proposed method, we conduct an extensive simulation study. There are 6 candidate doses from which the best dose is selected. We set the sample quota to be $N = 60$ and the maximum toxicity tolerance to be $\xi = 33\%$. Under each scenario, each method is simulated for 1000 iterations and the average results are compared by the accuracy of OBD selection, number of patients treated at the OBD and total trial duration.

We use the same priors for both of the methods. Toxicity prior skeleton is chosen as $(p_1, p_2, p_3, p_4, p_5, p_6) = (0.20, 0.30, 0.40, 0.50, 0.60)$. $L = 11$ possible efficacy models

are constructed as:

$$\mathbf{Q} = \begin{pmatrix} \mathbf{q_1} \\ \mathbf{q_2} \\ \mathbf{q_3} \\ \mathbf{q_4} \\ \mathbf{q_5} \\ \mathbf{q_6} \\ \mathbf{q_7} \\ \mathbf{q_8} \\ \mathbf{q_9} \\ \mathbf{q_{10}} \\ \mathbf{q_{11}} \end{pmatrix} = \begin{pmatrix} \mathbf{0.60}, 0.50, 0.40, 0.30, 0.20, 0.10 \\ 0.50, \mathbf{0.60}, 0.50, 0.40, 0.30, 0.20 \\ 0.40, 0.50, \mathbf{0.60}, 0.50, 0.40, 0.30 \\ 0.30, 0.40, 0.50, \mathbf{0.60}, 0.50, 0.40 \\ 0.20, 0.30, 0.40, 0.50, \mathbf{0.60}, 0.50 \\ 0.10, 0.20, 0.30, 0.40, 0.50, \mathbf{0.60} \\ 0.20, 0.30, 0.40, 0.50, \mathbf{0.60}, \mathbf{0.60} \\ 0.30, 0.40, 0.50, \mathbf{0.60}, \mathbf{0.60}, \mathbf{0.60} \\ 0.40, 0.50, \mathbf{0.60}, \mathbf{0.60}, \mathbf{0.60}, \mathbf{0.60} \\ 0.50, \mathbf{0.60}, \mathbf{0.60}, \mathbf{0.60}, \mathbf{0.60}, \mathbf{0.60} \\ \mathbf{0.60}, \mathbf{0.60}, \mathbf{0.60}, \mathbf{0.60}, \mathbf{0.60}, \mathbf{0.60} \end{pmatrix}. \tag{3.8}$$

$\mathbf{q_1}$ through $\mathbf{q_6}$ represent scenarios where the dose-efficacy peaked at dose d_1 through d_6, respectively; $\mathbf{q_7}$ through $\mathbf{q_{11}}$ are scenarios when dose-efficacy plateaus after the optimal dose. We assume no existing knowledge about the candidate models and set $h(\ell) = \frac{1}{11}$ for $\ell = 1, 2, \ldots, 11$.

TITE distributions

A couple of assumptions can be made on the arrival times of the patients in a clinical trial and the observation time of toxicity and efficacy responses. We assume the arrival of patients follows a Poisson distribution, with mean $\mu = 0.25, 1$ or 2 denoting the expected number of patients per week.

The time to event of toxicity or efficacy response is assumed to follow either a conditional uniform distribution or a Weibull distribution. For a conditional uniform distribution, the TITE would be randomly chosen from the interval $(0, T_{tox})$ or $(0, T_{eff})$ when a patient experiences a toxicity or efficacy event. We assume all toxicity and efficacy events will occur before T_{tox} and T_{eff}, respectively. An alterna-

tive assumption is the underlying TITE distribution follows a Weibull with a fixed shape parameter of 4 as in Cheung [23]. We take the same approach with the scale parameter, γ, as Cheung [23] and Braun [65] took and set it to make the cummulative distribution function at T equals to the probability of a response of the current dose level, that is,

$$F(x;\gamma) = \int_0^T \frac{4}{\gamma} x^3 \exp^{-(x/\gamma)^4} dx = \pi_E(d_i),$$

where $T = T_{tox}$ for toxicity responses and $T = T_{eff}$ for efficacy responses. We assume the maximum observation time for efficacy is $T_{eff} = 12$ weeks and separately simulate scenarios when toxicity can be observed in $T_{tox} = 4$ weeks and $T_{tox} = 12$ weeks.

Since we do not know the underlying distribution of the TITE variable, we simulate various truths for the TITE distribution. Note that 'U' denotes a conditionally Uniform distribution of the TITE variable; 'W' denotes a Weibull distribution.

Dose response scenarios

The dose-response scenarios are presented in Table 3.1. In scenario S1-S6, dose-toxicity is generally increasing but the overall toxicity rate is relatively low. The best dose occurs at level 1 through level 6 respectively. Scenario S7 through S9 represents scenarios with high toxicity rate and the true MTD is dose 4. Dose-efficacy in scenarios S7 through S9 is constant, increasing, and decreasing respectively.

Operating characteristics

Table 3.2 and Table 3.3 examine the operating characteristics of the proposed method regarding its ability to correctly select the OBD as well as the number of patients treated at the best dose. Table 3.2 summarizes results from simulations where toxicity outcomes can be observed within 4 weeks of treatment while efficacy effect takes 12 weeks to be observed; results in Table 3.3 are simulated assuming both toxicity and efficacy outcomes can be delayed up to 12 weeks.

Table 3.1: True probabilities of observing toxicity and efficacy responses at each dose levels.

Scenario		Dose level					
		1	2	3	4	5	6
S1	Toxicity	0.02	0.03	0.05	0.06	0.07	0.07
	Efficacy	0.40	0.25	0.20	0.15	0.10	0.05
S2	Toxicity	0.02	0.04	0.07	0.09	0.10	0.15
	Efficacy	0.15	0.45	0.30	0.25	0.16	0.10
S3	Toxicity	0.01	0.02	0.03	0.04	0.05	0.06
	Efficacy	0.10	0.15	0.35	0.18	0.12	0.07
S4	Toxicity	0.01	0.02	0.04	0.06	0.08	0.10
	Efficacy	0.05	0.15	0.30	0.45	0.35	0.30
S5	Toxicity	0.01	0.02	0.03	0.05	0.06	0.07
	Efficacy	0.15	0.25	0.33	0.47	0.60	0.40
S6	Toxicity	0.01	0.02	0.03	0.04	0.05	0.06
	Efficacy	0.05	0.15	0.30	0.35	0.40	0.50
S7	Toxicity	0.05	0.10	0.16	0.30	0.40	0.50
	Efficacy	0.30	0.30	0.30	0.30	0.30	0.30
S8	Toxicity	0.05	0.10	0.16	0.30	0.40	0.50
	Efficacy	0.10	0.20	0.30	0.40	0.50	0.60
S9	Toxicity	0.05	0.10	0.16	0.30	0.40	0.50
	Efficacy	0.50	0.40	0.30	0.20	0.10	0.05

Overall, the proposed method showed encouraging selection accuracy under various dose-response scenarios and different underlying simulation assumptions. The chance of correctly selecting the OBD generally ranges from 50% to 70%, except for scenario S6 and S7 under which the chance is only about 30%. The proposed method showed robustness against varying TITE distribution as well as the rate at which patients are enrolled. Whether the TITE follows a Weibull or conditional Uniform distribution, the chance of selecting the OBD remains the same. Similarly, the rate of patient arrival do not diminish the ability to select the OBD.

The distribution of patient allocation centers at the OBD with an average of 25 patients treated at the best dose. Note that even though the accuracy of dose selection is not affected by patient arrival rate, the distribution of patient allocation significantly depends on how fast patients are enrolled. The slower patients are enrolled, the more likely they are treated at the OBD, as more information become

available for dose assignment.

We further analyze scenario S6 and scenario S7 as results from these two scenarios are less desired than the others. Table 3.4 is a subset of the simulation results detailing the dose selection and patient allocation distribution for scenario S6 and S7. In table 3.4, the patient enrollment follows a Poisson distribution with rate = 0.25. Toxicity outcomes can be observed within 4 weeks. In scenario S6, both toxicity and efficacy are monotonically increasing with OBD being the highest dose. The results indicate that the proposed method is overly conservative and tends to select the highest 3 dose levels with similar probabilities. Patient allocation distribution centers at the next highest dose.

In scenario S7, all dose levels have the same efficacy probabilities while MTD is at the 4^{th} dose. The exact OBD is the first dose but dose 1-4 are also acceptable. Therefore, even though the probability of selecting OBD is only about 30%, most of the time, the selected dose is still acceptable. Similarly, the vast majority of the participating patients are treated at an efficacious dose level with less than 33% DLT probability.

An advantage of the proposed method is that the design significantly shortens the expected trial duration as dose assignment can be performed using partial information from patients who are still under observation. The trial duration shortens to approximately 5, 1.5 or 0.8 years when the rate of patient enrollment is respectively 0.25, 1, and 2 patients/week, as compared to 15 years if the next patient can only be enrolled when all current patients have completed follow up.

3.5 Application to motivating example

Recall the motivating example that studies the optimal dosing of ATRA in combination with a fixed dose of daratumuman in the treatment of relapsed or refractory multiple myeloma. The trial aims at selecting the optimal dose from three candidate

Table 3.2: Simulation results: toxicity and efficacy outcomes are observed for 4/12 weeks, respectively. 'U' and 'W' respectively denotes a conditionally Uniform distribution and Weibull distribution of toxicity or efficacy TITE. The enrollment of patients follow a Poisson process, with Rate=0.25, 1, or 2. The estimated probability of selecting the OBD is summarized in column 'Select', and the average number of patients treat at the OBD is under column 'Treat'.

Tox	Eff	Rate	Select	Treat	Tox	Eff	Rate	Select	Treat	Tox	Eff	Rate	Select	Treat
Scenario 1					Scenario 2					Scenario 3				
U	U	0.25	67.6	28.8	U	U	0.25	69.3	27.4	U	U	0.25	69.3	26.9
U	U	1	72.4	28.8	U	U	1	67.9	24.3	U	U	1	73.3	26.7
U	U	2	75.2	26.9	U	U	2	62.3	20.9	U	U	2	70.6	23.7
U	W	0.25	66	28.3	U	W	0.25	66.9	26	U	W	0.25	71.7	27.6
U	W	1	72.6	26.6	U	W	1	69.5	23.6	U	W	1	71.3	24.6
U	W	2	76.3	23.7	U	W	2	66.1	18.5	U	W	2	71.3	21
W	U	0.25	67.9	28.9	W	U	0.25	66.6	26.4	W	U	0.25	71.6	28
W	U	1	71.7	28.4	W	U	1	66.4	23.9	W	U	1	72	26.3
W	U	2	73.6	26.6	W	U	2	63.5	20.6	W	U	2	69.6	23.4
W	W	0.25	72	29.5	W	W	0.25	67.3	25.8	W	W	0.25	68.5	26.4
W	W	1	73.4	26.8	W	W	1	68.9	22.6	W	W	1	75.1	25.7
W	W	2	76.9	23.1	W	W	2	66.4	18.9	W	W	2	71.2	20.7
Scenario 4					Scenario 5					Scenario 6				
U	U	0.25	59.2	24	U	U	0.25	59.7	24.6	U	U	0.25	29.4	12.5
U	U	1	58.8	22.2	U	U	1	63.9	24.1	U	U	1	38	14.1
U	U	2	59.6	20.8	U	U	2	63.5	21.5	U	U	2	40.3	13.4
U	W	0.25	61.3	24.5	U	W	0.25	60.6	25	U	W	0.25	30.9	12.6
U	W	1	60	22.1	U	W	1	62	22.4	U	W	1	35.1	12.3
U	W	2	57	18.5	U	W	2	61.7	19.1	U	W	2	41	11.4
W	U	0.25	63.4	24.7	W	U	0.25	60.3	25.1	W	U	0.25	29.4	12.6
W	U	1	60.1	22.2	W	U	1	62.4	23.7	W	U	1	35.8	13.8
W	U	2	61.7	20.2	W	U	2	62.5	21.4	W	U	2	40.3	14.1
W	W	0.25	62.9	24.6	W	W	0.25	61.8	24.9	W	W	0.25	31	12.5
W	W	1	59	21.9	W	W	1	62.4	22.6	W	W	1	36.9	12.6
W	W	2	59.4	18.8	W	W	2	63	19.1	W	W	2	44.8	12
Scenario 7					Scenario 8					Scenario 9				
U	U	0.25	30.2	15.7	U	U	0.25	47.4	21.8	U	U	0.25	71.8	31.4
U	U	1	32.5	16.1	U	U	1	48.4	21.5	U	U	1	69.4	29.3
U	U	2	33.9	16	U	U	2	48.8	20	U	U	2	73.8	28.5
U	W	0.25	27.9	14.6	U	W	0.25	46.8	21.5	U	W	0.25	65.9	29.1
U	W	1	33.2	15	U	W	1	48.9	20.9	U	W	1	72.5	28
U	W	2	35.4	15.1	U	W	2	50.2	19.7	U	W	2	74.8	24.4
W	U	0.25	28.6	15	W	U	0.25	45.7	21.4	W	U	0.25	65.3	29.6
W	U	1	30.7	14.6	W	U	1	49.2	22.1	W	U	1	69.7	28.8
W	U	2	33	15	W	U	2	48.1	21.7	W	U	2	69.5	25.9
W	W	0.25	27.9	14.8	W	W	0.25	43.6	21	W	W	0.25	68.7	30.8
W	W	1	31.9	14.5	W	W	1	47.7	21.6	W	W	1	72.8	27.7
W	W	2	36.6	13.8	W	W	2	50.2	20.5	W	W	2	73.1	23.3

Table 3.3: Simulation results: toxicity and efficacy outcomes are both observed 12 weeks. 'U' and 'W' respectively denotes a conditionally Uniform distribution and Weibull distribution of toxicity or efficacy TITE. The enrollment of patients follow a Poisson process, with Rate=0.25, 1, or 2. The estimated probability of selecting the OBD is summarized in column 'Select', and the average number of patients treat at the OBD is under column 'Treat'.

Tox	Eff	Rate	Select	Treat	Tox	Eff	Rate	Select	Treat	Tox	Eff	Rate	Select	Treat
Scenario 1					Scenario 2					Scenario 3				
U	U	0.25	69.1	29.6	U	U	0.25	68.3	26.6	U	U	0.25	70.7	27.8
U	U	1	73.6	29.2	U	U	1	64.8	23.5	U	U	1	72.9	27.4
U	U	2	71.6	25.7	U	U	2	67.1	21.4	U	U	2	71.7	25.4
U	W	0.25	70.1	29.3	U	W	0.25	67.2	26.2	U	W	0.25	72.8	28.1
U	W	1	73.4	26.7	U	W	1	67.2	22.8	U	W	1	75.9	27.3
U	W	2	77.2	23.8	U	W	2	67.5	19	U	W	2	72.5	22.8
W	U	0.25	68	29.1	W	U	0.25	66.5	26	W	U	0.25	69.8	27.6
W	U	1	68.8	27.2	W	U	1	63.9	22.8	W	U	1	78.6	29
W	U	2	71.3	26.1	W	U	2	65.7	20.6	W	U	2	72.6	25.1
W	W	0.25	67.7	28.3	W	W	0.25	68.4	26.2	W	W	0.25	69.3	27.2
W	W	1	73.2	26.9	W	W	1	70.6	23.1	W	W	1	74.2	26.4
W	W	2	74.4	22.5	W	W	2	67.3	17.8	W	W	2	73.8	23
Scenario 4					Scenario 5					Scenario 6				
U	U	0.25	59.2	24.2	U	U	0.25	64.6	26.1	U	U	0.25	31.4	12.5
U	U	1	60.4	22.4	U	U	1	60	22.5	U	U	1	34.1	12.8
U	U	2	61.2	20.2	U	U	2	62.5	20.6	U	U	2	39.2	12.1
U	W	0.25	60.7	24.4	U	W	0.25	60.6	24.6	U	W	0.25	30.7	12.4
U	W	1	61.3	21.8	U	W	1	56.5	20.4	U	W	1	34.2	11
U	W	2	57.9	18.9	U	W	2	61.1	16.8	U	W	2	39.8	10.1
W	U	0.25	61.9	24.5	W	U	0.25	60.2	24.8	W	U	0.25	31.7	12.9
W	U	1	61.5	22.7	W	U	1	59.7	22.9	W	U	1	32.3	12.4
W	U	2	57.7	20.2	W	U	2	61.6	19.9	W	U	2	39	12.9
W	W	0.25	62.3	24.6	W	W	0.25	61.1	24.8	W	W	0.25	31.8	12.7
W	W	1	59.9	21.6	W	W	1	62	21.4	W	W	1	37.1	11.7
W	W	2	58.5	18.7	W	W	2	63.7	17.8	W	W	2	42.3	10.8
Scenario 7					Scenario 8					Scenario 9				
U	U	0.25	29.9	15.2	U	U	0.25	46.8	21.5	U	U	0.25	66.1	29.7
U	U	1	31.8	16.8	U	U	1	52.5	20.7	U	U	1	70.6	30
U	U	2	34.6	16.5	U	U	2	49	17.9	U	U	2	74.2	28.8
U	W	0.25	27.5	15.1	U	W	0.25	48.8	21.2	U	W	0.25	67.5	30
U	W	1	31.6	15.5	U	W	1	52.2	19.9	U	W	1	72.4	28.6
U	W	2	40.2	16.7	U	W	2	52.4	16.9	U	W	2	77.6	25.9
W	U	0.25	27.4	14.7	W	U	0.25	44.9	22	W	U	0.25	67.7	30.4
W	U	1	31.8	15.1	W	U	1	45.6	21.2	W	U	1	70.4	28.7
W	U	2	29.9	13.4	W	U	2	44.9	19	W	U	2	71.2	26.5
W	W	0.25	27.4	14.1	W	W	0.25	45.1	21.4	W	W	0.25	69.7	30.7
W	W	1	29.3	13.8	W	W	1	49.8	22	W	W	1	72.7	27.3
W	W	2	36	13.5	W	W	2	49.9	18.6	W	W	2	73	23.4

Table 3.4: Distribution of dose selection and patient allocation to each dose level after 1000 iterations of simulation. 'U' and 'W' respectively denotes a conditionally Uniform distribution and Weibull distribution of the toxicity and efficacy TITE variables. The enrollment of patients follows a Poisson process, with Rate=0.25. Maximum sample size is $N = 60$. The maximum tolerated DLT rate is set at $\xi = 33\%$. Dose-response scenarios are detailed in Table 3.1.

Scenario	Tox	Eff	Dose 1	Dose 2	Dose 3	Dose 4	Dose 5	Dose 6
Percentage of final selection								
6	U	W	0.5	2.4	20	22.1	24.3	30.7
6	U	U	0.9	2.4	17.5	24.1	23.7	31.4
6	W	U	0.3	2.8	17.6	22.8	24.8	31.7
6	W	W	0.6	2.7	17.9	22.6	24.4	31.8
Average number of patient treated								
6	U	W	2.5	4.6	11.5	13.6	15.4	12.4
6	U	U	2.6	4.3	10.8	14.0	15.9	12.5
6	W	U	2.4	4.4	11.06	13.3	15.9	12.9
6	W	W	2.5	4.3	10.9	13.9	15.7	12.7
Percentage of final selection								
7	U	W	27.5	27	26.9	17.2	1.4	0
7	U	U	29.9	24.4	28.7	16.2	0.8	0
7	W	U	27.4	25.6	30.6	15.5	0.9	0
7	W	W	27.4	24.4	30.4	16.3	1.5	0
Average number of patient treated								
7	U	W	15.1	14.9	16.4	11.0	2.3	0.3
7	U	U	15.5	15.0	17.0	10.2	2.4	0.3
7	W	U	14.7	14.6	17.0	10.8	2.5	0.3
7	W	W	14.1	14.5	17.2	10.9	3.0	0.4

dose levels of ATRA, $\{15, 30, 45\,\mathrm{mg/m^2}\}$, denoted by d_1, d_2, d_3. To design a study using the proposed method, we assume dose-toxicity is monotonic and choose toxicity prior skeleton $\mathbf{p} = (0.15, 0.25, 0.35)$. A full class of $L = 5$ efficacy skeletons can be constructed as

$$\mathbf{Q} = \begin{pmatrix} 0.4 & 0.3 & 0.2 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.3 & 0.4 \\ 0.3 & 0.4 & 0.4 \\ 0.4 & 0.4 & 0.4 \end{pmatrix}.$$

For this study, we excluded the first two skeletons because we do not expect

efficacy to decrease at higher doses. $\mathbf{Q}$ is reduced to

$$\mathbf{Q} = \begin{pmatrix} 0.2 & 0.3 & 0.4 \\ 0.3 & 0.4 & 0.4 \\ 0.4 & 0.4 & 0.4 \end{pmatrix}.$$

DLTs can be observed within 4 weeks of the treatment while efficacy effects may take as long as 8 weeks to be observed. If we waited until each patient is fully observed before we can enroll the next patient, a trial of 35 patients will be impractically long. However, if we can enroll a patient every two weeks, our proposed method can shorten the trial duration to 14 months by using partial information obtained from the patients who are still under observation.

For example, suppose the trial starts with the first patient receiving d_1, and no DLT or efficacy responses are observed when the second patient is enrolled 2 weeks later. We weight of toxicity outcome from the first patient is given by $\tau(t_{j,n}, T) = \arg\min(2/4, 1) = 0.5$ and the weight of efficacy outcome is $\tau(t_{j,n}, T) = \arg\min(2/8, 1) = 0.25$. Applying the standard TITE CRM, we can estimate the probability of DLT from the lowest dose to the highest dose, that is $(0.11, 0.20, 0.29)$.

According to equation (3.3), the likelihood function for the ℓ^{th} efficacy model can be expressed as

$$\begin{aligned} \mathcal{L}_\ell^{\mathbf{w}}(\beta_\ell|\Omega_1) &= \{\tau_j(t_{j,n}, T) \cdot q_{\ell x_j}^{\exp(\beta_\ell)}\}^0 \{1 - \tau_j(t_{j,n}, T) \cdot q_{\ell x_j}^{\exp(\beta_\ell)}\}^1 \\ &= 1 - 0.25 \cdot q_{\ell x_j}^{\exp(\beta_\ell)}. \end{aligned}$$

With the likelihood function, we can calculate the posterior model probabilities for each ℓ according to equation (3.4). For $\ell = 1$ through $\ell = 3$, $w(\ell|\Omega_{n=1}) = 0.339, 0.333,$ and 0.327, respectively. The posterior model probabilities are very close to each other because we only have one observation. According to equation (3.5), the implied best dose for each model $\ell = 1, 2, 3$ is dose level d_3, d_2, d_1, respectively.

The second patient will be randomized, and the dose assigning probability is calculated using equation (3.6). The probability of each dose is $0.327, 0.333,$ and

0.339. All three dose levels have positive probabilities to be assigned because the toxicity skeletons and observed data suggest that the estimated risk of DLT at the highest dose is smaller than the prespecified threshold $\xi = 33\%$.

For the second patient, all 3 models are included in the calculation of dose assigning probabilities since the sample is currently very small. As the sample size increases, we gradually reduce the number of models included in the calculation. The number of models used in probability calculation can be obtained from equation 3.7. This process continues until the maximum sample size is reached. The second patient will have approximately the same probability of receiving each of the candidate doses. This is intuitively appropriate as we currently do not have enough information to favor one dose over another and rely mostly on the model skeletons. As shown in the simulation study, the OBD tends to higher probabilities than the other doses.

For the motivating example, we run simulations and show that the proposed method has about 50% chance in selecting the OBD under various true dose-response scenarios with a sample size of 35. The chance of selecting a dose above the true MTD is less than 10%. We detailed the simulation scenarios and results in Table 3.5. Patient enrollment is assumed to follow a Poisson distribution with rate = 0.5 per week. Time to toxicity and time to efficacy responses are assumed to follow a conditional uniform distribution. Toxicity and efficacy responses are observed up to 4 and 8 weeks, respectively.

Table 3.5: Motivating example simulation scenarios and results

TITE Dist'n	True (toxicity, efficacy) probability		
Scenario 1	(0.05, 0.15)	(0.10, 0.30)	**(0.20,0.45)**
Tox-U, Eff-U	9.7	25.9	**64.4**
Tox-U, Eff-W	9.7	29.7	**60.9**
Tox-W, Eff-W	9.2	25.2	**65.6**
Tox-W, Eff-U	8.2	24.6	**67.2**
Scenario 2	(0.05, 0.15)	**(0.10, 0.35)**	0.20, 0.35)
Tox-U, Eff-U	10.2	**48.0**	41.8
Tox-U, Eff-W	10.6	**49.1**	40.3
Tox-W, Eff-W	9.9	**48.6**	41.5
Tox-W, Eff-U	10.2	**47.0**	42.8
Scenario 3	**(0.05, 0.30)**	(0.10, 0.30)	(0.20, 0.30)
Tox-U, Eff-U	**44.5**	30.9	24.3
Tox-U, Eff-W	**46.3**	29.3	24.4
Tox-W, Eff-W	**45.0**	29.9	25.1
Tox-W, Eff-U	**43.5**	30.6	25.9
Scenario 4	(0.15, 0.15)	**(0.30, 0.30)**	(0.40, 0.30)
Tox-U, Eff-U	35.5	**51.8**	12.7
Tox-U, Eff-W	36.2	**51.8**	12
Tox-W, Eff-W	36.5	**51.8**	11.7
Tox-W, Eff-U	39	**48.9**	12.1

3.6 Discussion

In this chapter, we answered the question posed in the Wages and Tait [44] article about extending the method to situations where binary toxicity and efficacy outcomes do not occur in a reasonably short time-frame. We outlined the a phase I/II method that accounts for varying degrees of delayed outcomes for both toxicity and efficacy of MTAs.

Through extensive simulations, we demonstrated the method's ability to accurately select the best dose while ensuring all available information is used in assigning doses to participating patients. The proposed design is most appropriate when non-monotone dose-efficacy relationship is expected. We also showed the robustness of the proposed method against various dose finding scenarios, dose-relationships, and patient recruit rates.

Even though faster patient enrollment will not diminish the accuracy of OBD selection, it will decrease the number of patients receiving the OBD. Therefore, when enrollment is very fast, the proposed method may raise ethical concerns. A solution is to temporarily pause enrollment to accumulate data from the existing patients. It requires further research to find and test appropriate enrollment restrictions.

Another potential direction for future study is early termination rules. Currently, the trial will continue until sample size is reached. Early terminations, either for futility or safety, will further reduce expected trial duration and save resources. In addition, inefficient/unsafe dose levels may be excluded from the study once enough patients have been assigned to it.

Chapter 4 An R Package for Seamless Phase I/II Adaptive Design Using Extensions of the Continual Reassessment Method

4.1 Abstract

With the emergence of cytostatic cancer treatments, the paradigm of dose-finding trials has been shifted. Traditional Phase I trials of cytotoxic agents aim to select the maximum tolerated dose (MTD) because cytotoxic agents non-selectively attack the process of cell division and better therapeutic effects can be expected at higher dose level. Cytostatic agents act through a different mechanism than cytotoxic agents and the dose-efficacy curve is not necessarily monotonic. Therefore, dose-finding clinical trials need to incorporate efficacy data in order to address the issue of non-monotonic dose-response relationships. When designing such dose-finding trials, we need to simultaneously protect trial participants from excessive toxicity and treat them at a therapeutically optimal dose level. Meanwhile, a design is only feasible if it can be completed within a reasonable time-frame. We proposed new seamleass phase I/II designs for cytostatic agents such as molecularly targeted agents and biologic therapies. In this chapter, we describe the usage and implementation of the R package **bpocrm**. The goal of our R package is to provide a tool for utilization and further investigation of the proposed designs. The **bpocrm** package contains functions to calculate dose-assigning probabilities given accrued data from currently enrolled patients, make final dose selections at the end of the trial, conduct simulation studies to evaluate operating characteristics such as probability of correct dose selection and number of patients treated at the best dose level, and generate plots to illustrate the dose-finding process. The package can be used for binary endpoints when both toxicity and efficacy can be observed within a reasonably short time-frame or for time-to-event endpoints when delayed responses are expected.

4.2 Introduction

A conventional Phase I oncology trial of chemotherapy agent aims at identifying the maximum tolerated dose (MTD). The MTD is a dose level beyond which the risk of severe toxicity outweighs potential therapeutic benefits. The MTD selected from Phase I will be further evaluated in a subsequent Phase II trial for efficacy. MTD is determined primarily through toxicity data without considering efficacy. This Phase I followed by Phase II approach assumes that the MTD is the highest dose with acceptable toxicity risk and it is also the most promising dose for efficacy. This assumption is not necessarily plausible for some cytostatic cancer drugs, such as molecularly targeted agents (MTAs) [54] and biologic therapies, where higher dose levels do not guarantee increased chance of efficacy. The drug therapeutic effects could possibly decrease or plateau after exceeding an intermediate dose while toxicity risk continues to increase. Therefore, the primary goal of this type of trials might shift to select an optimal biologic dose select the optimal biological dose (OBD) that is defined as the dose level with the highest probability of efficacy and acceptable risk of toxicity. A dose-finding design that aims to select the OBD needs to jointly model toxicity as well as efficacy outcomes.

Several designs have been developed for Phase I/II trials by jointly modeling toxicity and efficacy data, including a bivariate CRM design by Braun [38], an adaptive Bayesian method by Thall *et al.* that balances trade-offs between the probabilities of treatment efficacy and toxicity [39], a design that models the odds ratio of toxicity and efficacy by Yin *et al.* [41], and an approach proposed by Hoering that identifies MTD using traditional methods in Phase I and then search for OBD around the MTD in the following Phase II trials [36], and a design by Wages and Tait [44] that constructs and evaluates multiple CRM-like models to account for different possible dose-efficacy curves, among many others.

Even though many recently proposed seamless Phase I/II designs have demon-

strated desirable operating characteristics, such as high probability in selecting the correct dose, there is yet a design that is widely utilized in practice. It takes time for new designs to be accepted by the other researchers. More importantly, novel designs need to be accepted not only by statisticians, but also the medical research community. The CRM design was proposed in early 1990s and only start becoming a frequently utilized design until recent years [21, 22], even though its statistical superiorities have long been recognized by statisticians. The design by Wages and Tait is by nature an extension of the CRM design, which makes it easier to be accepted by people who have experience with the CRM. In addition, we proposed a redesign of the adaptive randomization stage used by the original Wages and Tait design in Chapter 2 to optimize its performance. We demonstrated through simulations that the proposed redesign allocates more patients to the OBD while slightly improve the probability of correct dose selection. We also extended the original design to incorporate TITE outcomes to address the issue of delayed outcomes in Chapter 3. The original design expects both toxicity and efficacy outcomes can be observed within a short follow-up period, but delayed endpoints are fairly common, especially for efficacy outcomes. The proposed TITE extension addressed this issue and substantially shorten the total trial duration.

However, there is not a software package available to implement and investigate this design and its extensions. In this chapter, we introduce an R package, **bpocrm** that provides tools for implementing the original Wages and Tait design as well as the redesigned adaptive randomization stage and TITE extensions. This package also includes functions to conduct simulation studies and generate plots to demonstrate the dose-finding process. The rest of this chapter is organized as follows: Section 4.3 describes the original design, redesigned adaptive randomization strategy, and the TITE extension; Section 4.4 demonstrates the main functions in this package with various examples; Section 4.5 summarizes the limitations and future development of

the design and R package.

4.3 Methods

Suppose a trial aims at recommending a dose for further investigation from I pre-defined dose levels $\mathcal{D} = \{d_1, d_2, \ldots, d_I\}$. When it is plausible to assume a monotonic dose-efficacy curve, the goal is usually to identify the MTD. For the design by Wages and Tait [44], it considers the possibility for efficacy to plateau or decrease after an intermediate dose level, and the goal is to find the OBD. Binary toxicity/efficacy endpoints, denoted by Y_j and Z_j, are used when outcomes can be observed within a reasonably short time-frame. The subscript $j = 1, 2, 3, \ldots, N$ is the index of participating patient ordered by sequence of enrollment, with N denoting the maximum sample size.

$$Y_j = \begin{cases} 0, & \text{if not DLT} \\ 1, & \text{if DLT} \end{cases}, \text{and } Z_j = \begin{cases} 0, & \text{if no efficacy response} \\ 1, & \text{if observe efficacy response} \end{cases}$$

Dose level administered on the j^{th} subject is denoted by $X_j \in \mathcal{D}$. After observing the response from the first n subjects, the toxicity and efficacy data can be given in form of $\Omega_j = \{(x_1, y_1, z_1), (x_2, y_2, z_2), \ldots, (x_n, y_n, z_n)\}$

When delayed responses are expected, TITE outcomes can be used to utilize partial information from patients who are still under observation and potentially shorten the total trial duration. Let $t_{j,n}$ denotes the follow-up time for patient j by the time the $(n+1)^{th}$ patient is enrolled. The maximum length of follow-up time, denoted by T, is the same for each patient. The observed toxicity and efficacy response, $Y_{j,n}$ and $Z_{j,n}$ are dependent on $t_{j,n}$. Note that the value of $Y_{j,n}$ and $Z_{j,n}$ may change from 0 to 1 as n increases, indicating toxicity/efficacy effects manifest after a certain time point. For each patient, we assume there are unobservable thresholds t_j^T

and t_j^E, only beyond which a positive toxicity or efficacy response can be observed.

$$Y_{j,n} = \begin{cases} 0 \text{ (no DLT)} & \text{when } t_{j,n} < t_j^T \\ 1 \text{ (observe DLT)} & \text{when } t_{j,n} \geq t_j^T \end{cases}, \text{ and}$$

$$Z_{j,n} = \begin{cases} 0 \text{ (no efficacy)} & \text{when } t_{j,n} < t_j^E \\ 1 \text{ (observe efficacy)} & \text{when } t_{j,n} \geq t_j^E \end{cases}.$$

By the time the $(n+1)^{th}$ patient is enrolled, data observed from the first n^{th} patients can be expressed in form of $\Omega_n^T = \{(x_j, y_{j,n}, z_{j,n}, t_{j,n})\}$ for $j = 1, 2, \ldots, n$.

The probability of observing a DLT and efficacy response at each dose level is denoted by $\pi_Y(d_i)$ and $\pi_Z(d_i)$, respectively.

Modeling Toxicity and efficacy

We assume toxicity monotonically increases with dose levels, and an acceptable dose range $\mathbb{A} = \{d_i | \hat{\pi}_Y(d_i) < \xi\}$ can be obtained by the traditional CRM [15, 66] or TITE CRM [23] depending on the outcome of interest and the time expected to observe it. Constant ξ is the maximum tolerated DLT rate to be specified by the users. We utilize the **dfcrm** package to run traditional CRM or TITE-CRM models. For the rest of this chapter, we set $\xi = 33\%$. OBD is defined as the dose with the highest efficacy within the range that still assures safety.

$$OBD = \arg\max_{d_i \in \mathbb{A}} \{\pi_E(d_i)\}.$$

Since the shape of dose-efficacy curve is unknown, the original Wages and Tait design [44] proposes to use a class of working models constructed from $L = 2 \times I - 1$ efficacy skeletons ($2I$ models will have shapes that plateau or peak at each dose level

minus one duplicate). Let $\mathbf{Q}$ denotes a class of skeletons:

$$\mathbf{Q} = \begin{pmatrix} \mathbf{q_1} \\ \mathbf{q_2} \\ \vdots \\ \mathbf{q_L} \end{pmatrix} = \begin{pmatrix} q_{11} & q_{12} & q_{13} & \cdots & q_{1I} \\ q_{21} & q_{22} & q_{23} & \cdots & q_{2I} \\ & & \cdots & & \\ q_{L1} & q_{L2} & q_{L3} & \cdots & q_{LI} \end{pmatrix} \tag{4.1}$$

Each row $\mathbf{q}_\ell \in \mathbf{Q}$ is a set of skeletons, denoting a unique shape of the does-efficacy curves. Note that some rows in $\mathbf{Q}$ may be removed if additional information about the dose-efficacy curve is available. For example, if there it is plausible to assume that efficacy will not decrease within the range of candidate doses, $L-1$ rows may be removed from $\mathbf{Q}$. Elements in of the matrix, $q_{\ell i} \in \mathbf{Q}$, are constants that represents our initial guesses of the probability of efficacy at dose level i, under the ℓ^{th} working model. For example, if there are three dose levels, we can construct $\mathbf{Q}$ as a 5 by 3 matrix:

$$\mathbf{Q} = \begin{pmatrix} 0.3 & 0.4 & \mathbf{0.5} \\ 0.4 & \mathbf{0.5} & 0.4 \\ \mathbf{0.5} & 0.4 & 0.3 \\ 0.4 & \mathbf{0.5} & \mathbf{0.5} \\ \mathbf{0.5} & \mathbf{0.5} & \mathbf{0.5} \end{pmatrix}$$

If there efficacy is unlikely to decrease, then $\mathbf{Q}$ can be reduced to

$$\mathbf{Q} = \begin{pmatrix} 0.3 & 0.4 & \mathbf{0.5} \\ 0.4 & \mathbf{0.5} & \mathbf{0.5} \\ \mathbf{0.5} & \mathbf{0.5} & \mathbf{0.5} \end{pmatrix}.$$

Based on each $\mathbf{q}_\ell$, a unique efficacy working model is constructed as model can be expressed as

$$\pi_E(\ell, d_i) = \Pr(z_j = 1 | \ell, d_i) \approx G_\ell(x_j; \beta_\ell) = q_{\ell i}^{exp(\beta_\ell)}, \tag{4.2}$$

where $\pi_E(d_i, \ell)$ is the estimated probability of observing a efficacy response on a patient tested at dose level d_i, and β_ℓ is the model parameter. Based on the binary

efficacy data in Ω_n, the likelihood function can be expressed as

$$\mathcal{L}_\ell(\beta_\ell|\Omega_n) = \prod_{j=1}^{n} \{q_{\ell x_j}^{\exp(\beta_\ell)}\}^{z_j} \{1 - q_{\ell x_j}^{\exp(\beta_\ell)}\}^{1-z_j}. \tag{4.3}$$

Given the TITE data, the likelihood function can be revised to

$$\mathcal{L}_\ell^{\mathbf{w}}(\beta_\ell|\Omega_n^T) = \prod_{j=1}^{n} \{\tau_j(t_{j,n}, T) \cdot q_{\ell x_j}^{\exp(\beta_\ell)}\}^{Z_{j,n}} \{1 - \tau_j(t_{j,n}, T) \cdot q_{\ell x_j}^{\exp(\beta_\ell)}\}^{1-Z_{j,n}}. \tag{4.4}$$

where the weight function $\tau^E(t_{j,n}, T)$ denotes the weight given to the outcome observed from the j^{th} patient by the time n patients have received the treatment. T is a constant denoting the observation window of the outcomes. $\tau(t_{j,n}, T)$ represents the proportion of follow-up time completed. For this version of R package, we choose to use the linear weighting scheme, that is,

$$\tau(t_{j,n}, T) = \begin{cases} \arg\min(t_{j,n}/T, 1) & \text{if no response observed at } t_{j,n} \\ 1 & \text{if response observed at } t_{j,n} \end{cases}$$

$\tau^(t_{j,n}, T) = 1$ when $t_{j,n} \geq T$, denoting patient j has completed efficacy observation.

The posterior probability distribution and expected value for model parameter β_ℓ can be expressed as follows:

$$\text{For binary endpints: } P_\ell(\beta_\ell|\Omega_n) = \frac{\mathcal{L}_\ell(\beta_\ell|\Omega_n)h(\beta_\ell)}{\int_{\beta_\ell} \mathcal{L}_\ell(\beta_\ell|\Omega_n)h(\beta_\ell)d\beta_\ell}, \tag{4.5}$$

$$\hat{\beta}_\ell = \frac{\int_{\beta_\ell} \beta_\ell \mathcal{L}_\ell(\beta_\ell|\Omega_n)h(\beta_\ell)d\beta_\ell}{\int_{\beta_\ell} \mathcal{L}_\ell(\beta_\ell|\Omega_n)h(\beta_\ell)d\beta_\ell} \tag{4.6}$$

$$\text{For binary endpints: } P_\ell(\beta_\ell|\Omega_n^T) = \frac{\mathcal{L}_\ell^{\mathbf{w}}(\beta_\ell|\Omega_n^T)h(\beta_\ell)}{\int_{\beta_\ell} \mathcal{L}_\ell^{\mathbf{w}}(\beta_\ell|\Omega_n^T)h(\beta_\ell)d\beta_\ell}, \tag{4.7}$$

$$\hat{\beta}_\ell = \frac{\int_{\beta_\ell} \beta_\ell \mathcal{L}_\ell^{\mathbf{w}}(\beta_\ell|\Omega_n^T)h(\beta_\ell)d\beta_\ell}{\int_{\beta_\ell} \mathcal{L}_\ell^{\mathbf{w}}(\beta_\ell|\Omega_n^T)h(\beta_\ell)d\beta_\ell} \tag{4.8}$$

In addition, the posterior model probabilities given the observed data can be estab-

lished as:

$$\text{For binary endpints: } w(\ell|\Omega_n) = \frac{\tau_\ell \int \mathcal{L}_\ell(\beta_\ell|\Omega_n)h(\beta_\ell)d\beta_\ell}{\sum_{\ell=1}^{L} \tau_\ell \int \mathcal{L}_\ell(\beta_\ell|\Omega_n)h(\beta_\ell)d\beta_\ell}, \tag{4.9}$$

$$\text{For TITE endpints: } w(\ell|\Omega_n^T) = \frac{p(\ell) \int_{\beta_\ell} \mathcal{L}_\ell^{\mathbf{w}}(\beta_\ell|\Omega_n^T)h(\beta_\ell)d\beta_\ell}{\sum_{\ell=1}^{L} p(\ell) \int \mathcal{L}_\ell^{\mathbf{w}}(\beta_\ell|\Omega_n^T)h(\beta_\ell)d\beta_\ell} \tag{4.10}$$

where τ_ℓ is the prior probability that the model built on skeleton q_ℓ is the best model to describe the dose-response relationship. We set $\tau_\ell = 1/L$, assuming no prior knowledge about the shape of dose-response relationship. Each time a new patient enters the trial, $w(\ell|\Omega_n)$ is updated and all candidate models are re-evaluated and compared against each other. Given any model ℓ probability of efficacy response can be estimated as

$$\hat{\pi}_E(d_i) \approx G_\ell(d_i; \hat{\beta}_\ell) = q_{i\ell}^{\exp(\hat{\beta}_\ell)}, \tag{4.11}$$

Adaptive randomization

Original Adaptive Randomization: Since $w(\ell|\Omega_n)$ is obtained from small samples, especially in early stage of the trial, we will not rely entirely on the $w(\ell|\Omega_n)$ in selecting the best fit model. The original design would randomize an arbitrary number of patients, n_{AR}, based on the estimated efficacy probability $\hat{\pi}(d_i, \ell^*)$ if the estimated risk of toxicity is acceptable. Specifically, when $n \leq n_{AR}$, we find

$$\ell^* = \begin{cases} \arg\max\{w(\ell|\Omega_n)\} & \text{for binary endpoints} \\ \arg\max\{w(\ell|\Omega_n^T)\} & \text{for TITE endpoints} \end{cases}$$

Based on the selected model ℓ^*, acceptable dose range $\mathbb{A}$, and estimated probability of efficacy $\pi_E(d_i, \ell^*)$, we can randomize the next patient to dose level i with probability

$$R_i = \frac{\hat{\pi}_E(d_i, \ell^*)}{\sum_{d_i \in \mathbb{A}} \hat{\pi}_E(d_i, \ell^*)} \tag{4.12}$$

Modified adaptive randomization: In Chapter 2, we proposed a different approach to randomize patients by using $w(\ell|\Omega_n)$ to weight the implied best dose from each candidate model. This modified randomization design tends to allocate more patients to the OBD and improve the chance of correctly selecting the OBD. In addition, we do not need to specify an arbitrary randomization sample size n_{AR}. Instead, this method is forced to converge into a final dose as sample size increases.

If a model was built using skeleton $\mathbf{q}_\ell$, then the dose level recommended by this model can be expressed as $\mathcal{S}(\ell)$, that is

$$\mathcal{S}(\ell) = \min\{\arg\max_i(\pi_E(d_i \in \mathbb{A}, \ell))\}. \tag{4.13}$$

The dose at level $\mathcal{S}(\ell)$ is expected to have the maximum chance of generating efficacy response if the shape of dose-efficacy model ℓ is closest to the true dose-efficacy relationship. $w(\ell|\Omega_n)$ indicates the likelihood of model ℓ being the most similar one to the true dose-efficacy curve given the observed data. Instead of selecting the best fit model-based on $w(\ell|\Omega_n)$, we use $w(\ell|\Omega_n)$ to weight the dose recommendation from each model. The dose-assigning probability can be calculated as

$$R_i^* = \frac{R_i^{**}}{\sum_{i=1}^{I} R_i^{**}}, \text{ where } R_i^{**} = \sum_{i}^{2I-1} w(\ell|\Omega_n)\mathbf{I}(d_i = \mathcal{S}(\ell) \text{ and } w(\ell|\Omega_n) \geq w_{(L-L'+1)}). \tag{4.14}$$

where $\mathbf{I}(.)$ is an indicator function. $w_{(1)} \leq w_{(2)} \leq \cdots \leq w_{(L)}$ denote the ordered posterior model probability $w(\ell|\Omega_n)$. Equation (4.14) considers the recommended dose from L' best-fit models and weights the recommendations of each model by $w(\ell|\Omega_n)$. When the suggested dose is outside the safety dose range, the highest dose in $\mathbb{A}$ will be used to ensure safety.

The number of models to be considered for patient randomization L' is gradually reduced from L to 1 based on the observed sample size.

$$L' = \left\lceil \left(\frac{N-n}{N}\right)^{\delta} L \right\rceil \tag{4.15}$$

δ is a constant referred to as the drop rate parameter. When $\delta = 1$, one candidate model is excluded from calculating the randomization probability for each additional N/L patients. When $0 < \delta < 1$, models are dropped at a slower rate when sample size is small and faster when sample size is large, vice versa for when $\delta > 1$. When $L' = 1$, only the best fit model will be considered and the next enrolled patients will be randomized to the best dose recommended by the selected model with 100% probability. We demonstrated in another study that $2 \leq \delta \leq 3$ gives the most desirable operating characteristics.

4.4 Usage of the package

Package functions

Package **bpocrm** depends on three libraries: **nnet**, **dfcrm**, and **binom** [67–69]. Package **nnet** was employed to find the maximum position in a vector. Package **dfcrm** was employed to utilize CRM and TITE CRM models for analyzing toxicity data. The **binom** package was employed for early stopping rules. The exact binomial confidence intervals for toxicity and efficacy probabilities are calculated at each dose level using the **binom** package.

Package **bpocrm** contains five functions: *get.skel*, *bpocrm.imp*, *bpocrmTITE.imp*, *bpocrm.sim*, and *bpocrmTITE.sim.* The *get.skel* function constructs a class of efficacy skeletons as shown in equation 4.1. Input options are summarized in table 4.1. In order to get generate a class of efficacy skeletons, the users need to specify the number of candidate dose levels, and maximum and minimum probability of efficacy within this dose range. A matrix of efficacy skeletons will be returned with each row representing a unique shape of the dose-efficacy curve. Note that the matrix generated by *get.skel* is not the only way to construct working efficacy skeletons. The users may construct their own efficacy skeletons based their specific situations. The option *pos.desc* determines whether efficacy can decrease after an intermediate dose level.

Table 4.1: Input options for function *get.skel*

Name	Role
max.eff	Presumed maximum probability of efficacy within the range of candidate doses.
min.eff	Presumed minimum probability of efficacy within the range of candidate doses.
ndose	Number of dose levels.
pos.desc	Whether or not efficacy can decrease after an intermediate dose

Functions *bpocrm.imp* and *bpocrmTITE.imp* have several arguments in common. Both functions calculate dose assigning probabilities given the observed data. The *bpocrm.imp* function is appropriate when the data is given in binary form, while the *bpocrmTITE.imp* function is designed for TITE outcomes. Input options for *bpocrm.imp* and *bpocrmTITE.imp* are summarized in Table 4.2 and Table 4.3, respectively. Both functions require accrued data to be recorded in vectors (*d.obs,* *y.obs*, and *z.obs*). Since we assume monotonic dose-toxicity relationship, the toxicity skeleton also needs to be a vector (*p.skel*). To account for different shapes of the dose-efficacy curve, efficacy skeletons (*q.skel*) need to be given as a matrix. Each row in matrix *q.skel* is an efficacy skeleton, representing a potential shape of the dose-efficacy curve. The users may write their own (*q.skel*) or use the *get.qskel* function to get a generic class of efficacy skeletons.

The *ar.strategy* option determines which adaptive randomization strategy will be used. When sample size is small, there is not enough data to select a working efficacy model, so patients are randomized. When ar.strategy="Original", the function will adopt the original design by Wages and Tait. Under this option, the users also need to specify the number of patients to be randomized through parameter (*n.ar*), and parameter (*drop.rate*) will be ignored. When ar.strategy ="Modified", adaptive randomization is based on equation (4.14). This option requires parameter (*drop.rate*)

to be provided. The (*drop.rate*) parameter is denoted by δ in equation 4.15. It controls how working candidate models are excluded from the calculation of dose assigning probabilities. Higher δ values will require less observations to exclude an additional candidate model during early stage of the trial. We recommend $\delta = 2$ (default) or $\delta = 3$ based on simulation studies. Parameter ($n.ar$) will be ignored when ar.strategy="Modified" as this method does not require an arbitrarily selected sample size for adaptive randomization. Instead, it gradually exclude lack-of-fit working models based on the number of observations.

The difference between *bpocrm.imp* and *bpocrmTITE.imp* is that the *bpocrmTITE.imp* function requires an additional vector ($t.obs$) to specify the length of follow-up for each patient. The *bpocrmTITE.imp* function also requires the maximum follow-up time for toxicity ($max.T.tox$) and efficacy ($max.T.eff$).

Table 4.2: Input options for function *bpocrm.imp*

Name	Role
d.obs	A vector of dose levels assigned to patients. The length of d.obs must be equal to y.obs and z.obs.
y.obs	Toxicity data: a vector of patient toxicity outcomes; 1 indicates efficacy, 0 otherwise. Length must equal to *d.obs*
z.obs	Efficacy data: a vector of patient toxicity outcomes; 1 indicates efficacy, 0 otherwise. Length must equal to *d.obs*
N	Total number of patients planned for this trial.
p.skel	Toxicity skeleton. A vector of values for the initial geusses of toxicity probabilities at each dose level. User may provide a vector with length equal to the number of dose levels, or use getprior() function from the dfcrm package.
q.skel	Efficacy skeletons. A matrix of efficacy values. Each row of the matrix is an efficacy skeleton. Each efficacy skeleton represents a potential dose-response curve. Function get.qskel() may be used to generate a generic matrix.
tul	Toxicity upper limit. Maximum probability of toxicity tolerated.
ell	Efficacy lower limit. Minimum probability of efficacy to continue the trial.
ar.strategy	Choose from "Original" or "Modified". "Original" is the strategy used in original Wages and Tait (2015), and "Modified" is based on equation (4.14).
n.ar	Number of patients to be randomized if *ar.strategy*="Original".
drop.rate	The δ parameter in equation 2.12 determining how working candidate models are excluded from the calculation of dose assigning probabilities. Higher δ values will require less data to exclude an additional candidate model during early stage of the trial. We recommend using $\delta = 2$ or $\delta = 3$ from simulation results.

Table 4.3: Input options for function *bpocrmTITE.imp*

Name	Role
...	Same arguments as function *bpocrm.imp*
t.obs	Length of follow-up time for each patient.
max.T.tox	Maximum follow-up time for toxicity outcomes.
max.T.eff	Maximum follow-up time for efficacy outcomes.

Functions *bpocrm.sim* and *bpocrmTITE.sim* can be used to conduct simulation studies, and the input options are summarized in Table 4.4 and Table 4.5. The input options for these two functions are very similar to function *bpocrm.imp* and *bpocrmTITE.imp*. They are only different in a couple of options. First, function *bpocrm.sim* and *bpocrmTITE.sim* require true probabilities of toxicity ($p0$) and efficacy ($q0$) at each dose level, as opposed to the implementation function *bpocrm.imp* and *bpocrmTITE.imp* that read accrued data. Vector ($p0$) and ($q0$) need to have length equal to the number of candidate dose levels. Second, the users need to specify how many iterations to be simulated through parameter (*ntrial*). Note that when $ntrial = 1$, one trial will be simulated and plots will be automatically generated to illustrate the dose-finding process. When $ntrial > 1$, summary statistics are returned to demonstrate the average performance of the method.

When using the *bpocrmTITE.sim* function, We assume patients enrollment follows a Poisson distribution and the users need to specify the rate parameter. In addition, the minimum time to observe toxicity and efficacy responses may follow a Weibull distribution or Uniform distribution. This is controlled through option (*tox.TITE*) and option (*eff.TITE*). Maximum follow-up time for toxicity and efficacy responses are controlled by (*max.T.tox*) and (*max.T.eff*) respectively.

Table 4.4: Input options for function *bpocrm.sim*

Name	Role
p0	True toxicity probabilities at each dose level.
q0	True efficacy probabilities at each dose level.
N	Total number of patients planned for each trial.
p.skel	Toxicity skeleton. A vector of values for the initial geusses of toxicity probabilities at each dose level. User may provide a vector with length equal to the number of dose levels, or use getprior() function from the dfcrm package.
q.skel	Efficacy skeletons matrix. Each row of the matrix is an efficacy skeleton, representing a potential shape of the dose-efficacy curve.
tul	Toxicity upper limit. Maximum probability of toxicity tolerated.
ell	Efficacy lower limit. Minimum probability of efficacy to continue the trial.
ar.strategy	Choose from "Original" or "Modified". "Original" is the strategy used in original Wages and Tait (2015), and "Modified" is based on equation (4.14).
n.ar	Number of patients to be randomized if *ar.strategy*="Original".
drop.rate	The δ parameter in equation (4.15) determining how working candidate models are excluded from the calculation of dose assigning probabilities. Higher δ values will require less data to exclude an additional candidate model during early stage of the trial. We recommend using $\delta = 2$ or $\delta = 3$ from simulation results.
start.set	Dose level for the first patients. If length greater than 1, a random dose will be selected.
ntrial	Number of trials to be simulated. When ntrial=1, a plot is automatically generated to illustrate the dose-finding process.

Table 4.5: Input options for function *bpocrmTITE.sim*

Name	Role
...	Same arguments as function *bpocrm.sim*
pt.dist	Patient recruitment rate. The expected number of patients to be enrolled per week.
tox.TITE	Distribution of toxicity minimum follow up time. Toxicity responses are only observable after the minimum follow up time. "U"= Uniform. "W"=Weibull.
eff.TITE	Distribution of efficacy minimum length of follow up."U"= Uniform. "W"=Weibull. Efficacy responses are only observable after the minimum follow up time.
max.T.tox	Maximum follow-up time for toxicity outcomes.
max.T.eff	Maximum follow-up time for efficacy outcomes.

Examples: assigning dose given observed data

Example 1: use the get.skel() function to generate a generic efficacy skeleton matrix.

Consider a dose-finding trial with $I = 6$ candidate dose levels, where dose-toxicity is assumed to be monotone but the shape of true dose-efficacy curve can be monotone, plateau, or peak. Toxicity can be modeled using CRM method with skeleton $\mathbf{p} = \{0.01, 0.08, 0.15, 0.20, 0.29, 0.36\}$. This toxicity skeleton is chosen arbitrarily for demonstration purposes. More functions on the CRM can be found in the **dfcrm** package.

In the get.qskel() function, the users specify the maximum/minimum probability of efficacy response (max.eff, min.eff) within the range of candidate doses and number of candidate doses (ndose). By default, the get.qskel() function will generate a matrix of efficacy skeletons that represent increasing, plateaued, or peaked dose-efficacy curves. The efficacy is unlikely to decrease after the intermediate dose, the user may use option pos.desc=FALSE.

q.skel=get.qskel(max.eff=0.6, min.eff=0.1, ndose=6) will generate the following efficacy skeletons.

$$\mathbf{Q} = \begin{pmatrix} 0.1 & 0.2 & 0.3 & 0.4 & 0.5 & \mathbf{0.6} \\ 0.2 & 0.3 & 0.4 & 0.5 & \mathbf{0.6} & 0.5 \\ 0.3 & 0.4 & 0.5 & \mathbf{0.6} & 0.5 & 0.4 \\ 0.4 & 0.5 & \mathbf{0.6} & 0.5 & 0.4 & 0.3 \\ 0.5 & \mathbf{0.6} & 0.5 & 0.4 & 0.3 & 0.2 \\ \mathbf{0.6} & 0.5 & 0.4 & 0.3 & 0.2 & 0.1 \\ 0.2 & 0.3 & 0.4 & 0.5 & \mathbf{0.6} & \mathbf{0.6} \\ 0.3 & 0.4 & 0.5 & \mathbf{0.6} & \mathbf{0.6} & \mathbf{0.6} \\ 0.4 & 0.5 & \mathbf{0.6} & \mathbf{0.6} & \mathbf{0.6} & \mathbf{0.6} \\ 0.5 & \mathbf{0.6} & \mathbf{0.6} & \mathbf{0.6} & \mathbf{0.6} & \mathbf{0.6} \\ \mathbf{0.6} & \mathbf{0.6} & \mathbf{0.6} & \mathbf{0.6} & \mathbf{0.6} & \mathbf{0.6} \end{pmatrix} \tag{4.16}$$

The users may also manually specify a efficacy skeletons to reflect prior knowledge of the dose-efficacy curve if such information is available.

Example 2: Consider a trial with maximum sample size $N = 64$ and $I = 6$ candidate doses. Toxicity and efficacy outcomes can be observed before the next patient is enrolled, and the data is in binary form. The hypothetical observed data is summarized in Table 4.6. We use the skeletons obtained in Example 1. Maximum acceptable toxicity rate is 33% (tul=0.33). The dose administrated, observed toxicity outcomes, and observed efficacy outcomes are denoted by the vector *d.obs, y.obs, z.obs*, respectively.

> d.obs=c(1, 1, 1, 2, 2, 2, 3, 3, 3, 4)

> y.obs=c(0, 0, 0, 0, 0, 0, 0, 0, 1, 1)

> z.obs=c(0, 0, 0, 0, 0, 0, 0, 1, 0, 1)

In this example, we use the original design by Wages and Tait with ar.strategy='Original', and n.ar=16.

> set.seed(580)

> bpocrm.imp(y.obs,z.obs,d.obs,N=64,p.skel=p.skel,q.skel=q.skel,tul=0.33, ell=0.04, ar.strategy="Original",n.ar=16)

Table 4.6: Hypothetical data for R-package examples

Patient number	1	2	3	4	5	6	7	8	9	10
Dose Level	1	1	1	2	2	2	3	3	3	4
y.obs (DLT=1)	0	0	0	0	0	0	0	0	1	1
z.obs (Efficacy)	0	0	0	0	0	0	0	1	0	1

Outputs from the above inputs are summarized into a list of R objects. The dose randomized to the next patient is in ($dose.assign). Because the current sample size is 10 and we used n.ar=16, this dose assignment is a result of adaptive randomization, and the probability each dose is in ($ar.prob). In this example, the next patient will be treated at the second dose level. The probability of receiving the first, second, and third dose is 15%, 33%, and 52%, respectively. The other dose levels currently have 0% probability because the estimated probability of toxicity is greater than 33%.

The estimated probability of efficacy is calculated in ($peff.hat). Each row in ($peff.hat) is the estimated probability of efficacy given a particular efficacy skeleton. In this example, there are 11 efficacy skeletons from which 11 candidate efficacy models are constructed. Therefore, there are 11 rows in ($peff.hat). Using the first row in matrix 4.16, we would estimate the probability of efficacy at each dose level to be 0.08, 0.16, 0.26, 0.36, 0.46, and 0.56, given the observed data. How well each candidate model fits the observed data is given in the posterior model probability ($postprob.eff).

The exact binomial confidence intervals for toxicity and efficacy at each dose level are used as early termination rules. Whether or not the trial should to stop for futility or safety are indicated by ($stopf) and ($stops). In this example, both ($stopf) and ($stops) equal to 0, so the trial would continue to enroll more patients. Note that to implement the modified adaptive randomization strategy, we need to change

the “ar.strategy” argument to “ar.strategy=Modified” and specify the “drop.rate” parameter.

$dose.assign

[1] 2

$ar.prob

[1] 0.1511697 0.3293778 0.5194525 0.0000000 0.0000000 0.0000000

$ptox.hat

[1] 0.04657312 0.07389260 0.28270437 0.36483626 0.43852408 0.50643940

$peff.hat

[1,] 0.07523605 0.1639289 0.2585277 0.3571782 0.45895543 0.56329579

[2,] 0.09617486 0.1734836 0.2636532 0.3647778 0.47558707 0.36477785

[3,] 0.10546731 0.1805253 0.2739004 0.3850571 0.27390042 0.18052526

[4,] 0.14138954 0.2276751 0.3360188 0.2276751 0.14138954 0.07650311

[5,] 0.25072569 0.3607698 0.2507257 0.1606142 0.09045426 0.04027012

[6,] 0.42203760 0.3101938 0.2128020 0.1309135 0.06600986 0.02047585

[7,] 0.09617486 0.1734836 0.2636532 0.3647778 0.47558707 0.47558707

[8,] 0.10546731 0.1805253 0.2739004 0.3850571 0.38505713 0.38505713

[9,] 0.12329303 0.2052681 0.3113183 0.3113183 0.31131829 0.31131829

[10,] 0.17262097 0.2740086 0.2740086 0.2740086 0.27400860 0.27400860

[11,] 0.25538505 0.2553850 0.2553850 0.2553850 0.25538505 0.25538505

$postprob.eff

[1] 0.171979595 0.151907937 0.140425321 0.057513786 0.014484311 0.006765511

0.151907937 0.140425321 0.088890872 0.044913352 0.030786058

$stopf

[1] 0

$stops

[1] 0

Example 3:

Consdier the same scenario as Example 2, except we incorporate the time of follow-up into the design. Toxicity is observed for up to 4 weeks and efficacy is followed up to 12 weeks. Suppose the first 6 patients have completed follow-up for both toxicity and efficacy and the last 4 patients have been treated observed for 5, 3, 2, and 1 weeks respectively. In addition, we use the ar.strategy="Modified" in this example. Under the modified adaptive randomization strategy, we do not specify how many patients are being randomized. Instead, candidate models are gradually eliminated, and the design would converge into a final dose. In this example, the next patient is assigned to the third dose level with 95% probability, while dose 2 still has 5% probability. Function input and results are listed below:

```
> bpocrmTITE.imp(d.obs,y.obs,z.obs,t.obs, p.skel,q.skel,tul,ell,N=64,
+ ar.strategy="Modified",n.ar=1, drop.rate=2,
+ max.T.tox=4, max.T.eff=12)
$dose.assign
[1] 3
$ar.prob
[1] 0.0000000 0.0469965 0.9530035 0.0000000 0.0000000 0.0000000
$ptox.hat
[1] 0.05104337 0.07987554 0.29358235 0.37599762 0.44946180 0.51684218
$peff.hat
[1,] 0.1052235 0.2072460 0.3080942 0.4081873 0.50772270 0.60681585
[2,] 0.1309417 0.2185305 0.3142908 0.4166258 0.52452473 0.41662576
[3,] 0.1436048 0.2283276 0.3271660 0.4389357 0.32716604 0.22832763
[4,] 0.1893360 0.2839540 0.3954236 0.2839540 0.18933598 0.11228211
[5,] 0.2945520 0.4062475 0.2945520 0.1987332 0.11966102 0.05853726
```

[6,] 0.4608129 0.3494873 0.2491433 0.1610483 0.08707241 0.03043070

[7,] 0.1309417 0.2185305 0.3142908 0.4166258 0.52452473 0.52452473

[8,] 0.1436048 0.2283276 0.3271660 0.4389357 0.43893567 0.43893567

[9,] 0.1677959 0.2591610 0.3696758 0.3696758 0.36967583 0.36967583

[10,] 0.2176739 0.3250769 0.3250769 0.3250769 0.32507686 0.32507686

[11,] 0.3027060 0.3027060 0.3027060 0.3027060 0.30270597 0.30270597

$postprob.eff

[1] 0.165591398 0.148057056 0.140455613 0.066733860 0.013561551 0.005811805

[7] 0.148057056 0.140455613 0.097269004 0.044709122 0.029297922

Examples of trial simulations

Example 4: In this example, we use the bpocrm.sim() function to simulate one trial in order to demonstrate the dose-finding process in a trial. Suppose the true probability of toxicity is $\mathbf{p_0}$ = c(0.05, 0.10, 0.20, 0.28, 0.50, 0.50) and probability of efficacy is $\mathbf{q_0}$ = c(0.05, 0.13, 0.25, 0.38, 0.50, 0.63). We use a total of 30 patients to find locate the OBD which is the 4^{th} dose level. The other options are similar to Example 2, and we have the following input for the bpocrm.sim() function:

> result= bpocrm.sim(p0,q0,p.skel=p.skel,q.skel=q.skel,

+ tul=0.33,ell=0.04,N=30,start.comb=1,

+ ar.strategy="Modified",n.ar=1,drop.rate=2,ntrial=1)

The output is a list containing the true probability of toxicity ($True.tox), true probability of efficacy ($True.eff), dose level assigned to each patient ($d.obs), observed toxicity ($y.obs) and efficacy ($z.obs) responses. In this example, because we set ntrial=1, plots are generated to demonstrate the dose-finding process. Specifically, in Figure 4.1, it shows the dose level (y-axis) received by each patient (x-axis). Ae shown in the plot, most patients are treated at the 4^{th} dose level, which is the true OBD. The second most frequently selected dose is the 3^{rd} dose level. The other

dose levels are only explored with a small number of patients.

If we set ar.strategy="Modified" as specified in this example, another plot is generated showing the how candidate models are eliminated (see Figure 4.2). A solid dot indicates that the corresponding model is included in the calculation of dose assigning probabilities. As shown in the plot, all models are initially included. By the 10^{th} patient, models based on the 5^{th} and 6^{th} rows in efficacy skeleton matrix are excluded. Note that this exclusion is only temporary, each model is re-evaluated as new data are observed. Towards the end of the trial, only one efficacy model is considered for dose-assigning, indicating a final dose-efficacy model is selected. The final model is based on the 3^{rd} row in the efficacy skeleton matrix.

Figure 4.1: An example of treated patient toxicity and efficacy outcomes during one trial scenario.

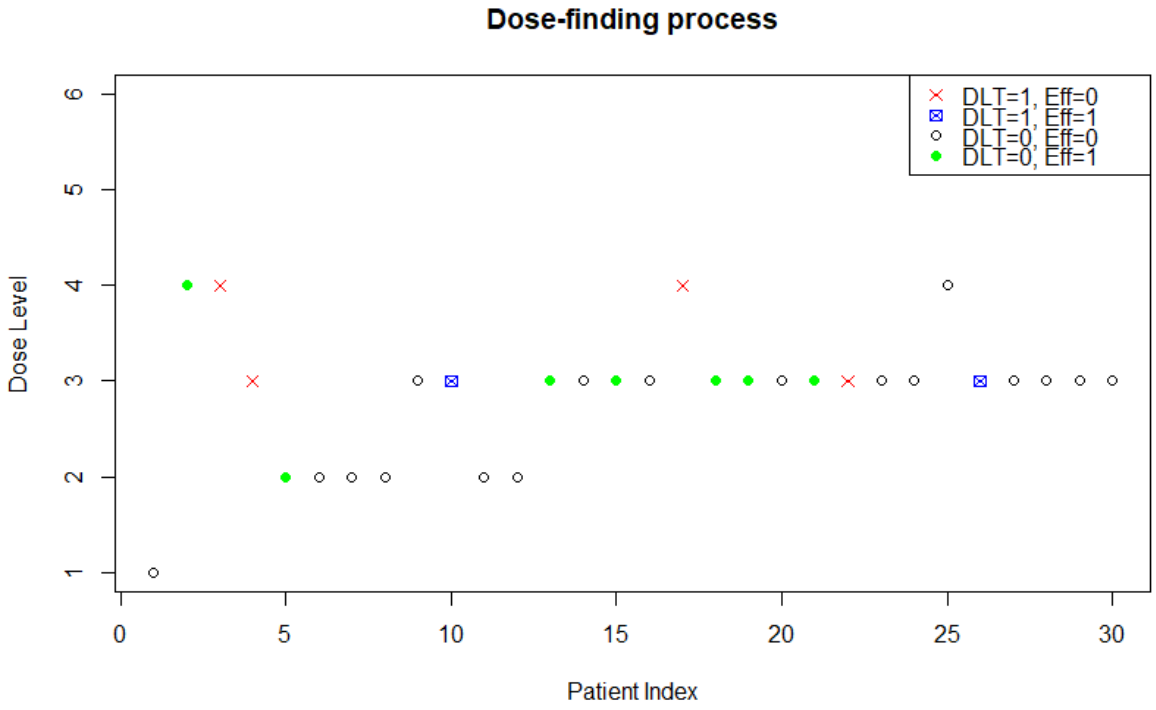
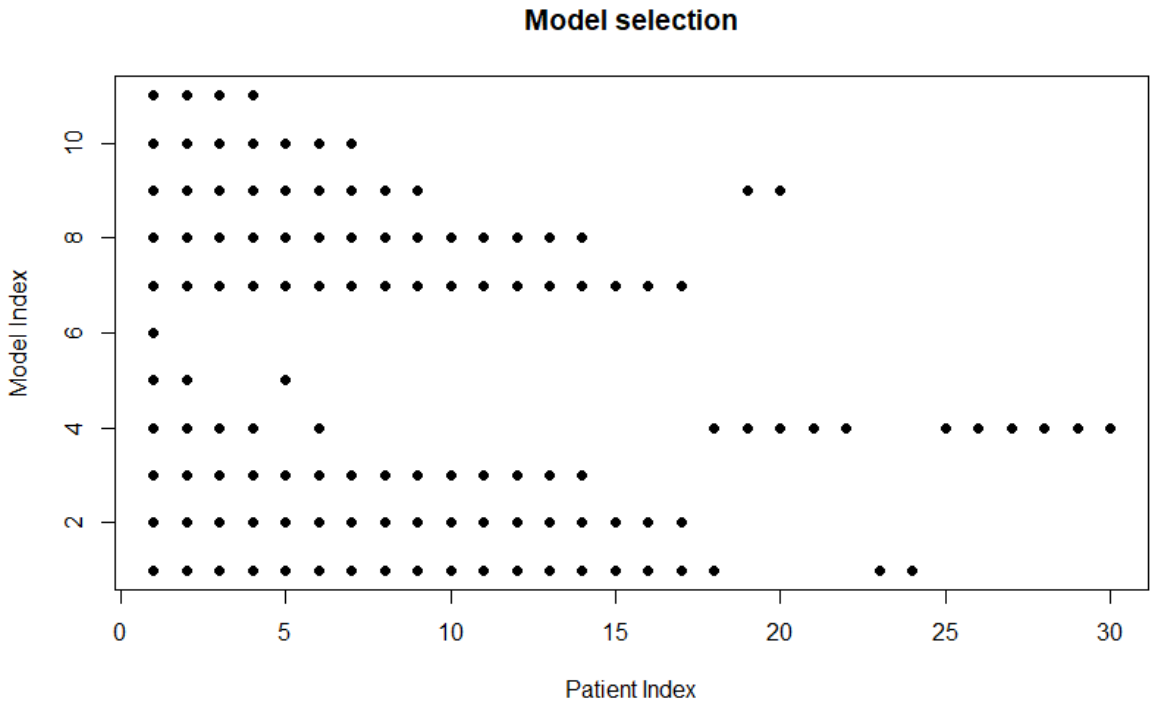


Figure 4.2: Illustration of the model selection process during a simulated trial in Example 4. A solid point indicates the corresponding model is included in the dose-assignment process.



Example 5: Similiar to example 4, except the data is observed in TITE format, and we repeat the simulation for 100 times. The time to toxicity or efficacy event follows a Uniform (tox.TITE="U") or Weibull (eff.TITE="W") distribution. In this example, we assume that toxicity is followed up to 4 weeks (max.T.tox=4) and efficacy is monitored up to 12 weeks (max.T.eff=12). In addition, the patient enrollment follows a Poisson distribution (pt.dist="P") with 2 patients per week on average (pt.rate=2).

```
> bpocrmTITE.sim(p0,q0,p.skel,q.skel,tul=0.33,ell=0.05,N=64, start.set=c(1),
+ ntrial=100, ar.strategy="Modified", n.ar=64 ,drop.rate=2,
+ max.T.tox=4,max.T.eff=12, pt.rate=2, pt.dist="P", tox.TITE="U", eff.TITE="W")
```

The result is a list containing true rate of toxicity and efficacy at each dose level ($True.tox and $True.eff), percent of final dose selection ($Selection.pct), average number of patients treated at each dose level ($Treat.avg), average total trial duration ($total.time), total number of toxicity/efficacy outcomes per trial ($n.tox, $n.eff).

In this example, after repeating the trial for 100 times, 44% of trials selected the true OBD (the 4^{th} dose, and an average of 18.5 patients were treated at the OBD. On average, 14.4 patients experienced DLT, and 17.6 patients showed efficacy responses. The total trial duration is about 44 weeks, assuming a 2 patients per week enrollment rate.

```
$True.tox
[1] 0.05 0.10 0.20 0.28 0.50 0.50
$True.eff
[1] 0.05 0.13 0.25 0.38 0.50 0.63
$Selection.pct
[1] 1 22 22 44 11 0
$Treat.avg
[1] 4.65 17.73 15.54 18.46 6.77 0.85
```

6 $total.time

[1] 43.93541

$n.tox

[1] 14.4

$n.eff

[1] 17.61

4.5 Discussion

In this chapter, we detailed the usage and implementation of the **bpocrm** package, corresponding to the several recently proposed methods and designs. This package provides functions for calculating dose-assigning probabilities, and conducting simulation studies using a recently proposed dose-finding Phase I/II clinical trial design method and its extensions. These proposed methods demonstrated desirable operating characteristics, and this package serves as a tool for utilizing the proposed designs. Meanwhile, this package is also useful for further development, comparison and evaluation of relevant Phase I/II designs.

There are several limitations to the current version of the **bpocrm** package. The early stopping rules used in the designs are based on exact binomial confidence intervals. Conditions for early terminations are rarely met unless the true probabilities are extreme. We notice the suggested dose level usually converges to a single dose during late stage of the trials and adding more patients are very unlikely to change the dose recommendation. Therefore, in future studies, we would like to investigate the use of convergence as an early stopping rule.

Another feature that we hope to develop in the future versions is a function to calculate dose recommendations for all possible outcome scenarios in the next 3-5 patients. This feature can make model-based designs more like rule-based designs, which makes it easier to be implemented in practice.

In summary, the functions included in this R package can be utilized to implement and simulate the seamless Phase I/II design proposed in the Wages and Tait (2015) paper. We also modified the original adaptive randomization stage to improve dose selection accuracy and allocate more patients to the best dose. By extending the original design to incorporate TITE outcomes, we addressed the potential issue of delayed outcomes and long follow-up time. The simulation functions in this package is a convenient tool for exploring the operating characteristics of the proposed methods under various scenarios. We plan to submit this package to CRAN shortly after the completion of this dissertation so that all functions and codes are freely available. All necessary functions are also available upon request.

Chapter 5 Summary and Discussion

5.1 Summary

Phase I dose-finding trials aim to select the most promising dose with an acceptable toxicity profile for future investigation. When testing cytostatic agents, such as protein inhibitor, angiogenesis inhibitor or receptor modulator, the shape of dose-efficacy curve is not necessarily monotonic. Therefore, the trial design should not only ensure safety, but also incorporate information from efficacy outcomes. In this dissertation, we completed three aims: 1, we redesigned the adaptive randomization stage used in the original Wages and Tait (2015) method, and used simulations to show that the proposed method improves the probability of selecting the correct dose, and allocates more patients to the correct dose during adaptive randomization stage; 2, We extended the Wages and Tait design to incorporate TITE outcomes. This extension solved a potential issue of delayed outcomes, which is frequently encountered in practice; 3, We introduced an R package to provide tools for implementing and simulating both the original design as well as the proposed extensions.

Aim 1: Re-designed adaptive randomization for the Wages and Tait method

We proposed and evaluated three alternative randomization strategies for a recently published seamless phase I/II adaptive design. The original design by Wages and Tait in 2015 was proposed for trials of molecularly targeted agents in cancer treatments. The proposed randomization strategies calculate randomization probabilities using the likelihood of every candidate model as opposed to the original design, which selects the best model and then randomize based on estimations from the selected model. Through simulations under various scenarios, we evaluated the proposed randomization strategies and compared them to the original design. The simulation

results showed that one of the proposed strategies allocates more patients to the OBD while maintaining approximately the same level of accuracy in selecting the optimal dose without increasing the overall risk of toxicity.

Aim 2: Bivariate Generalization of the TITE CRM

We developed a seamless phase I/II design for delayed toxicity and efficacy response, where dose-toxicity is strictly increasing but the shape of dose-efficacy curve can be increasing, peak, or plateau.

In this design, toxicity and efficacy are modeled in parallel and collected as TITE endpoints. Toxicity and efficacy data are collected as TITE endpoints. The primary goal is to optimize the number of patients allocated to the OBD at which the chance of observing efficacy response is maximized but the risk of DLT is maintained below a prespecified level. This design combined features of time-to-event (TITE) CRM and partial order (PO) CRM.

This design is an extension of an existing seamless design proposed by Wages and Tait [44]. By modeling TITE endpoints instead of binary, this design allows enrollment of new patients while some patients are still under observation. Toxicities are monitored by TITE-CRM, and efficacy is monitored by evaluating and weighting a class of efficacy working models. Since the dose-efficacy relationship is not necessarily monotonic, our design will construct multiple dose-efficacy models, with each model representing a unique dose-efficacy shape. The posterior probabilities of each model being the best-fit will be calculated. When there are not enough data to consistently and reliably select the best fit-model, patients will be randomized within an acceptable dose range. The length of time each patient has been followed through the trial is incorporated into the weighted likelihood functions.

Simulation studies show that the proposed design greatly shortened the total trial duration as compared to the original design while maintaining approximately the same

chance of correctly selecting the OBD in most scenarios. However, faster recruitment reduces the number of patients allocated to the OBD because dose assigning decisions are made with less information at each enrollment.

Aim 3: An R package

This R package serves as a tool for implementing and simulating the proposed designs. It also provides documentation to the functions used in dissertation. The functions in this package can be utilized by the biostistician to adaptively calculate the recommended dose for the next patient. There are also functions available to conduct or replicate simulation studies in order to further explore the characteristics of the proposed designs. This package also include functions to implement and simulate the original method proposed by Wages and Tait (2015) [44].

5.2 Discussion

Strengths

Dose-finding trials are generally evaluated by three operation characteristics: probability of correct final dose selection, proportion of patients treated at or near the optimal dose, and total trial duration.

One of the strengths from our research is that all proposed methods are evaluated by comprehensive simulation studies. Different combinations of true dose-toxicity and dose-efficacy curves are used to examine the performance of the designs. The re-designed adaptive randomization stage showed increased chance of correct OBD selection and more patients treated at the OBD under most scenarios. The only exception is that when the true dose-efficacy is monotonic increasing, the proposed adaptive randomization method is more conservative than the original design. The proposed method has a lower probability of selecting the OBD, but the probability of selecting the OBD and the next lower dose is the same. This indicates that the

final dose selected by the proposed method is likely to the OBD or at least very close to the OBD.

The proposed TITE extension of the Wages and Tait design answered a question posed in the original Wages and Tait (2015) article. The original design is most appropriate when both toxicity and efficacy outcomes can be observed within a reasonably short follow-up window, but delayed outcomes are frequently encountered in practice, especially for efficacy outcomes. When delays are expected in observation of outcomes, the original design would not be able to utilize all accrued information. The TITE extension proposed in our research is an effective solution to address this practical issue, and makes this design more applicable in practice. Simulation studies show that the chance of correct OBD selection is similar to the original design, but the total trial duration can be substantially reduced.

Another strength is that we developed the **bpocrm** R package to implement both our proposed methods and the original Wages and Tait design. This user-friendly tool makes these designs accessible for future implementation in practice. It also provides a good source of documentation for future research and further improvement.

Limitations and future directions

There are several limitations that we hope to addressed in future research.

For both of the proposed methods, even though there are early stopping rules in the proposed designs, conditions for early trial termination are rarely met in simulations. This is because the current early termination rules are based on exact binary confidence intervals which are very wide given we only have few patients at each dose level. Unless the true scenarios are extreme, these intervals are not very helpful for a typical dose-finding trial. We noticed that patients enrolled towards the end of trials are usually assigned to the same dose level, indicating a convergence in dose assignment. A potential direction for future research is to propose and implement early

termination rules that are based on convergence rather than confidence intervals.

Another limitation shared by both of the proposed methods is the absence of dose-skipping rules. Consequently, the dose assignment is not stable in the beginning stage of the trials. We believe that dose-skipping rules are likely to have only a slight impact of the design performance, but it is a necessary for the design to be implemented in practice. Dose-skipping rules and other practical modifications can be added to future versions of the R package.

In the TITE extension of the original design, even though the probability of selecting the OBD is not affected by enrollment rate, the number of patients allocated to the OBD will be substantially reduced if the time between each enrollment is short. A plausible explanation is that when patients are recruited fast, there is less information available to determine the best dose at each enrollment. On the other hand, if we halt enrollment until every current patient has finished follow-up, every dose assignment will be based on full information from the previous cohorts, but the trial duration will be impractically long. A direction for future research is to investigate how to control the enrollment rate. A potential solution is to halt enrollment when the number of patients under observation exceeds a certain threshold.

In conclusion, we redesigned the adaptive randomization stage for a seamless Phase I/II design proposed by Wages and Tait (2015) and used simulations to show that this modification improved the performance of the original design under most scenarios. We further expanded the Wages and Tait design to incorporate TITE outcomes to utilize partial information from patients who are still under observation. This extension addressed a practical issue of delayed outcomes. Finally, we introduced an R package to implement both the original design and the proposed extensions. The R package also provides tool for simulation studies.

The methods proposed in this research are valuable additions to the original design, and R package makes these methods accessible to other researchers.

Appendices

Table A1: Full simulation results of tuning the drop rate parameter δ (Chapter 2, section 2.5). A best dose is defined as the level that has maximum chance of efficacy while assuring safety, and a good dose is the level with 25% or higher chance of efficacy while assuring safety. Safety is defined as 33% or lower DLT rate.

Probability of Selecting A Best/Good Dose										
Scenarios	Delta=0.5		Delta=1		Delta=2		Delta=3		Delta=4	
	Best Dose	Good Dose	Best Dose	Good Dose	Best Dose	Good Dose	Best Dose	Good Dose	Best Dose	Good Dose
R1T1	62	62	62	62	61	61	59	59	60	60
R1T2	55	55	53	53	53	53	53	53	55	55
R1T3	36	36	39	39	40	40	41	41	42	42
R1T4	69	96	62	95	52	91	50	91	49	90
R2T1	67	92	68	92	70	94	69	93	67	93
R2T2	61	82	64	86	65	86	65	89	65	88
R2T3	50	56	56	61	59	65	59	67	60	68
R2T4	100	24	99	29	97	36	96	37	96	38
R3T1	99	99	99	99	100	100	99	99	99	99
R3T2	99	99	99	99	98	98	99	99	99	99
R3T3	96	96	96	96	96	96	96	96	96	96
R3T4	100	100	100	100	100	100	100	100	100	100
R4T1	75	99	75	99	77	100	75	99	74	99
R4T2	74	99	75	99	76	99	74	99	73	98
R4T3	72	98	72	98	72	98	66	96	64	96
R4T4	69	100	69	100	69	100	66	99	61	98
Average Number of Patient Treated at A Best/Good Dose										
Scenarios	Delta=0.5		Delta=1		Delta=2		Delta=3		Delta=4	
	Best Dose	Good Dose	Best Dose	Good Dose	Best Dose	Good Dose	Best Dose	Good Dose	Best Dose	Good Dose
R1T1	25	25	26	26	27	27	27	27	27	27
R1T2	23	23	23	23	24	24	24	24	24	24
R1T3	20	20	21	21	21	21	21	21	22	22
R1T4	18	50	20	50	21	51	22	52	23	52
R2T1	27	45	28	46	30	47	30	48	31	49
R2T2	26	42	27	44	28	44	29	46	29	45
R2T3	23	34	25	34	27	36	28	37	29	38
R2T4	51	22	52	22	54	23	55	23	55	24
R3T1	61	61	61	61	62	62	62	62	62	62
R3T2	60	60	61	61	61	61	61	61	62	62
R3T3	59	59	59	59	60	60	60	60	60	60
R3T4	64	64	64	64	64	64	64	64	64	64
R4T1	28	54	30	56	34	57	35	57	35	58
R4T2	28	54	30	55	34	56	35	57	35	57
R4T3	25	51	28	53	31	55	31	55	31	55
R4T4	23	56	26	57	29	58	30	59	30	59

Table A2: Full simulation results of Operating Characteristics (Chapter 2, section 2.5). A best dose is defined as the level that has maximum chance of efficacy while assuring safety, and a good dose is the level with 25% or higher chance of efficacy while assuring safety. Safety is defined as 33% or lower DLT rate.

Probability of Selecting A Best/Good Dose																		
Scenarios	WT-16		WT-32		WT-48		S1-16		S1-32		S1-48		S2		S3-D2		S3-D3	
	Best	Good	Best	Good	Best	Good	Best	Good	Best	Good	Best	Good	Best	Good	Best	Good	Best	Good
R1T1	60	60	59	59	60	60	58	58	59	59	60	60	57	57	61	61	59	59
R1T2	53	53	51	51	50	50	55	55	55	55	53	53	53	53	53	53	53	53
R1T3	42	42	42	42	36	36	41	41	41	41	40	40	35	35	40	40	41	41
R1T4	57	92	64	95	67	96	49	90	51	95	59	96	67	95	52	91	50	91
R2T1	68	93	65	90	64	86	65	92	69	92	68	91	67	93	70	94	69	93
R2T2	60	85	60	82	57	80	63	87	61	84	63	85	63	84	65	86	65	89
R2T3	56	65	50	56	45	51	57	66	58	63	51	58	53	58	59	65	59	67
R2T4	95	32	98	26	99	19	95	36	98	27	98	27	97	33	97	36	96	37
R3T1	99	99	100	100	99	99	100	100	99	99	99	99	99	99	100	100	99	99
R3T2	99	99	100	100	99	99	99	99	99	99	99	99	98	98	98	98	99	99
R3T3	97	97	97	97	98	98	97	97	97	97	97	97	95	95	96	96	96	96
R3T4	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
R4T1	68	99	70	99	69	99	73	99	73	99	73	99	75	99	77	100	75	99
R4T2	67	98	69	100	67	99	70	99	72	99	74	100	77	99	76	99	74	99
R4T3	60	96	64	98	64	98	58	97	66	98	70	97	70	97	72	98	66	96
R4T4	57	98	63	99	63	100	59	98	66	100	69	100	69	99	69	100	66	99
Average Number of Patient Treated at A Best/Good Dose																		
Scenarios	WT-16		WT-32		WT-48		S1-16		S1-32		S1-48		S2		S3: D-2		S3: D-3	
	Best	Good	Best	Good	Best	Good	Best	Good	Best	Good	Best	Good	Best	Good	Best	Good	Best	Good
R1T1	26	26	23	23	18	18	25	25	25	25	24	24	26	26	27	27	27	27
R1T2	22	22	20	20	17	17	24	24	23	23	22	22	25	25	24	24	24	24
R1T3	20	20	18	18	16	16	22	22	21	21	20	20	21	21	21	21	21	21
R1T4	30	52	27	50	21	45	24	51	22	51	21	49	21	54	21	51	22	52
R2T1	28	44	23	39	18	34	29	47	28	45	27	44	29	48	30	47	30	48
R2T2	26	42	22	37	17	33	27	44	26	43	25	42	28	45	28	44	29	46
R2T3	25	34	20	29	16	28	27	36	26	35	23	33	27	36	27	36	28	37
R2T4	53	20	49	19	44	20	53	23	53	20	51	21	54	25	54	23	55	23
R3T1	61	61	60	60	59	59	62	62	61	61	61	61	62	62	62	62	62	62
R3T2	61	61	60	60	59	59	62	62	61	61	60	60	62	62	61	61	61	61
R3T3	60	60	58	58	57	57	60	60	59	59	58	58	60	60	60	60	60	60
R3T4	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64
R4T1	33	56	28	52	23	49	35	57	31	55	28	54	33	57	34	57	35	57
R4T2	31	55	28	52	22	48	34	56	31	54	28	53	33	56	34	56	35	57
R4T3	28	53	26	50	21	47	29	55	28	53	26	51	29	54	31	55	31	55
R4T4	28	57	25	54	19	51	28	57	27	57	24	55	28	58	29	58	30	59

Package 'bpocrm'

October 26, 2017

Type Package

Title Bivariate Generalization of the Partial Order Continual Reassessment Method

Version 0.1.0

Date 2017-09-10

Author Donglin Yan
Nolan Wages
Emily Dressler

Maintainer Donglin yan <donglin.yan2@gmail.com>

Description Provide functions to implement and simulate the bivariate extention of the partial order continual reassessment method (BPO-CRM) for use in Phase I/II trials of molecularly targeted agents and immunotherapies, where the dose-efficacy relationship is not monotonic. Provide functions to implement and simulate time-to-event (TITE) extension of the bivariate POCRM design.

Depends R (>= 3.3.1), stats, graphics

Imports nnet,dfcrm,binom

License GPL (>= 2)

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1.9000

NeedsCompilation no

R topics documented:

bpocrm.imp *bpocrm.imp applies the Bivariate Partial Order CRM using observed data.*

Description

bpocrm.imp assigns dose to the next patient given the currently observed data. It also estimates posterior model probabilies, adaptive randomization probabilities as well as probabilities of toxicity and efficacy. If the trial is in adaptive randomization stage, the next patient will be randomized within a safety range (under estimated MTD); otherwise the estimated optimal biological dose (OBD) will be assigned. Tox/eff data are observed in binary form.

Usage

```
bpocrm.imp(y.obs, z.obs, d.obs, N, p.skel, q.skel, tul, ell, ar.strategy,
  n.ar = 1, drop.rate = 2)
```

Arguments

y.obs	A vector of patient toxicity outcomes; 1 indicates toxicity, 0 otherwise.
z.obs	A vector of patient efficacy outcomes; 1 indicates efficacy, 0 otherwise.
d.obs	A vector of dose levels assigned to patients. The length of d.obs must be equal to y.obs and z.obs.
N	Total number of patients to be enrolled for this trial.
p.skel	Toxicity skeleton. A vector of values for the initial geusses of toxicity probabilities at each dose level. User may provide a vector with length equal to the number of dose levels, or use getprior() function from the 'dfcrm' package.
q.skel	Efficacy skeletons. A matrix of efficacy values. Each row of the matrix is an efficacy skeleton. Each efficacy skeleton represents a potential dose-response curve. Use get.qskel() to generate a generic matrix.
tul	Toxicity upper limit. A constant between 0 and 1. The highest rate of toxicity tolerated.
ell	Efficacy lower limit. A constant between 0 and 1. The minimum rate of efficacy to continue the trial.
ar.strategy	"Original" or "Modified". "Original" is based on section 3.1 in Wages and Tait (2015) "Modified" is based on equation (12) and (13) in Yan et. al. (In review) If ar.strategy="Original", then n.ar need to be specified. If ar.strategy="Modified", then drop.rate need to be specified.
n.ar	Sample size for adaptive randomization stage. Required when using "Original" ar.strategy. This argument is ignored when ar.strategy="Modified".
drop.rate	Required when using "Modified" ar.strategy. This rate decides how fast we eliminate unfit efficacy models from randomization. drop.rate=2 is recommended. This argument is ignored when ar.strategy="Original".

Details

Based on the method proposed in Wages and Tait 2015. Modified adaptive randomization strategy (ar.strategy) is base on Yan et. al. (In review).

Value

A list of objects:

`dose.assign`	Dose assigned to the next enrolled patient.
`ptox.hat`	Estimated probability of toxicity at each dose.
`ar.prob`	Calculated adaptive randomization probability at each dose.
`peff.hat`	Values on each row are the estimated probability of efficacy response calculated from the corresponding efficacy skeleton in q.skel. For example, peff.hat[4,3] is the estimated efficacy probability at the third dose level, calculated from the 4th efficacy skeleton.
`postprob.eff`	Posteria probability of each efficacy model (efficacy skeleton).
`stopf`	0 or 1. 1=Early termination for futility.
`stops`	0 or 1. 1=Early termination for safety.

References

1. Wages, N. A., & Tait, C. (2015). Seamless Phase I/II Adaptive Design for Oncology Trials of Molecularly Targeted Agents. J Biopharm Stat, 25(5), 903–920. http://doi.org/10.1080/10543406.2014.920873
2. Yan, D., Wages, N., Dressler, V. (In review) Improved Adaptive Randomization Strategies for the Multidimensional CRM in Phase I/II Trials. Journal of Biopharmaceutical Statistics.

See Also

get.qskel, dfcrm::getprior( )

Examples

```
#Suppose a trial is aimed at selecting the OBD from a set of 6 dose levels.
#The maximum sample size is 64 and currrently we observed the following data.
y.obs=c(0,0,0,0,0,0,0,0,1)
z.obs=c(0,0,0,0,0,0,0,0,1,0)
d.obs=c(1,1,1,2,2,2,3,3,3)
#Specify toxicity and efficacy skeletons:
p.skel=c(0.01,0.02,0.15,0.22,0.29,0.36)
q.skel=get.qskel(max.eff=0.6, min.eff=0.1, ndose=6, pos.desc=TRUE)
#To implement the Wage and Tait 2015 method with original adaptive randomization design:
set.seed(580)
bpocrm.imp(y.obs,z.obs,d.obs,N=64,p.skel=p.skel,q.skel=q.skel,tul=0.33,ell=0.04,
          ar.strategy="Original",n.ar=16)
```

`bpocrm.sim` *bpocrm.sim Bivariate Partial Order CRM simulator.*

Description

Simulation function for BPOCRM. This function is used to generate simulation replicates of Phase I/II trials using Bivariate Partial Order CRM under a specified dose-toxicity and dose-efficacy configuration.

Usage

```
bpocrm.sim(p0, q0, p.skel, q.skel, tul, ell, N, start.comb, ar.strategy, n.ar,
  drop.rate, ntrial)
```

Arguments

p0	True toxicity probabilities. A vector of values between 0 and 1 with length equal to the number of candidate doses. Represents the probability of observing a dose-limiting toxicity from the lowest dose to the highest dose, respectively.
q0	True efficacy probabilities. A vector of values between 0 and 1 with length equal to the number of candidate doses. Represents the probability of observing an efficacy response from the lowest dose to the highest dose, respectively.
p.skel	Toxicity skeleton. A vector of values for the initial geusses of toxicity probabilities at each dose level.
q.skel	Efficacy skeletons. A matrix of efficacy values. Each row of the matrix is an efficacy skeleton. Each efficacy skeleton represents a potential dose-response curve. Use get.qskel() to generate a generic matrix.
tul	Toxicity upper limit. A constant between 0 and 1. The highest rate of toxicity tolerated.
ell	Efficacy lower limit. A constant between 0 and 1. The minimum rate of efficacy to continue the trial.
N	Number of patients to be enrolled for this trial.
start.comb	The dose level at which the trial starts.
ar.strategy	"Original" or "Modified". If ar.strategy="Original", then n.ar need to be specified. Refer to Wages and Tait (2015) If ar.strategy="Modified", then drop.rate need to be specified. Refer to Yan et. al. (In review)
n.ar	Sample size for adaptive randomization stage. Required when using "Original" ar.strategy. This argument is ignored when ar.strategy="Modified".
drop.rate	Required when using "Modified" ar.strategy. This rate decides how fast we eliminate unfit efficacy models from randomization. drop.rate=2 is recommended. This argument is ignored when ar.strategy="Original".
ntrial	Number of trials to be simulated. If ntrial=1, a detailed dose-finding process is returned, including dose assignment, toxicity and efficacy outcomtes of every patient. A plot is also generated to visualize the dose-finding process.

Value

A list of objects.

True.tox	True probability of toxicity at each dose level. Same as p0
True.eff	True probability of efficacy at each dose level. Same as q0
Selection.pct	Percent of each dose being selected as final dose.
Treat.avg	Number of patients treated at each dose level.
n.tox	Number of patients experienced DLT.
n.eff	Number of patients experienced efficacy.
ntrial	Number of trials simulated

Examples

```
#' #Assume p0 and q0 are the true probability of toxicity/efficacy at each dose level
set.seed(2017)
p0=c(0.05,0.10,0.20,0.28,0.50,0.50)
q0=c(0.05,0.13,0.25,0.38,0.50,0.63)
#p.skel and q.skel are the skeletons for dose-toxicity and dose-efficacy model
p.skel=c(0.01,0.02,0.15,0.22,0.29,0.36)
q.skel=get.qskel(max.eff=0.6, min.eff=0.1, ndose=6, pos.desc=TRUE)
#If we simulate one trial, a plot will be generated to show the dose-assignment of each patient
 result= bpocrm.sim(p0,q0,p.skel=p.skel,q.skel=q.skel,
           tul=0.33,ell=0.04,N=64,start.comb=1,
          ar.strategy="Modified",n.ar=4,drop.rate=2,ntrial=1)
 result

#We can simulate 1000 trials to examine the operating characteristics of the proposed design.
## Not run:  result= bpocrm.sim(p0,q0,p.skel=p.skel,q.skel=q.skel,
           tul=0.33,ell=0.04,N=64,start.comb=1,
          ar.strategy="Modified",n.ar=4,drop.rate=2,ntrial=1000)
## End(Not run)
```

`bpocrmTITE.imp` *bpocrm.imp.TITE applies the TITE extension of BPOCRM using the observed data.*

Description

bpocrmTITE.imp assigns dose to the next patient given the observed data. It also estimates posterior model probabilies, adaptive randomization probabilities as well as probabilities of toxicity and efficacy. If the trial is in adaptive randomization stage, the next patient will be randomized within a safety range (under estimated MTD); otherwise the estimated optimal dose will be assigned. Different from the bpocrm.imp() function, the observed data is weighted by the length of follow-up. Positive outcomes (1's) and negative outcomes (0's) with follow-up time longer than a prespecified limit are given full weight; otherwise, a fractional weight is calculated.

Usage

```
bpocrmTITE.imp(d.obs, y.obs, z.obs, t.obs, p.skel, q.skel, tul, ell, N,
  ar.strategy, n.ar = 1, drop.rate = 2, max.T.tox, max.T.eff)
```

Arguments

`d.obs`	A vector of dose levels assigned to patients. The length of d.obs must be equal to y.obs and z.obs.
`y.obs`	A vector of patient toxicity outcomes; 1 indicates toxicity, 0 otherwise.
`z.obs`	A vector of patient efficacy outcomes; 1 indicates efficacy, 0 otherwise.
`t.obs`	A vector of patient follow-up time.
`p.skel`	Toxicity skeleton. A vector of values for the initial geusses of toxicity probabilities at each dose level.
`q.skel`	Efficacy skeletons. A matrix of efficacy values. Each row of the matrix is an efficacy skeleton. Each efficacy skeleton represents a potential dose-response curve. Use get.qskel() to generate a generic matrix.

tul	Toxicity upper limit. A constant between 0 and 1. The highest rate of toxicity tolerated.
ell	Efficacy lower limit. A constant between 0 and 1. The minimum rate of efficacy to continue the trial.
N	Total number of patients to be enrolled for this trial.
ar.strategy	"Original" or "Modified". "Original" is based on section 3.1 in Wages and Tait (2015) "Modified" is based on equation (12) and (13) in Yan et. al. (In review) If ar.strategy="Original", then n.ar need to be specified. If ar.strategy="Modified", then drop.rate need to be specified.
n.ar	Sample size for adaptive randomization stage. Required when using "Original" ar.strategy. This argument is ignored when ar.strategy="Modified".
drop.rate	Required when using "Modified" ar.strategy. This rate decides how fast we eliminate unfit efficacy models from randomization. drop.rate=2 is recommended. This argument is ignored when ar.strategy="Original".
max.T.tox	Maximum follow up time for toxicity outcomes.
max.T.eff	Maximum follow up time for efficacy outcomes.

Details

Based on the method proposed in Yan et. al. (In review). This design is a TITE extension of the bivariate partial order CRM. Toxicity is modeled by TITE-CRM (Cheung 2000). Efficacy is modeled using a matrix of efficacy skeletons to incorporate non-monotone dose-efficacy. Toxicity and efficacy responses are weighted by the length of follow-up in order to reduce total trial duration.

Value

A list of objects:

dose.assign	Dose assigned to the next enrolled patient.
ptox.hat	Estimated probability of toxicity at each dose.
ar.prob	Calculated adaptive randomization probability at each dose.
peff.hat	A matrix. Values on each row are the estimated probability of efficacy response calculated from the corresponding efficacy skeleton in q.skel. For example, peff.hat[4,3] is the estimated efficacy probability at the third dose level, calculated from the 4th efficacy skeleton.
postprob.eff	Posteria probability of each efficacy model (efficacy skeleton).

References

Cheung, Y. K., & Chappell, R. (2000). Sequential Designs for Phase I Clinical Trials with Late-Onset Toxicities. Biometrics, 56(4), 1177–1182. http://doi.org/10.1111/j.0006-341X.2000.01177.x

Yan, D., Wages, N. A., Tait, C., Kindwall-Keller, T., & Dressler, E. V. (in review). Bivariate Generalization of the Time-to-Event Continual Re- assessment Method. Journal of the Royal Statistical Society. Series C: Applied Statistics Royal Statistical Society.

See Also

bpocrm

Examples

```
#Suppose a trial is aimed at selecting the OBD from a set of 6 dose levels.
#The maximum sample size is 30 and currrently we enrolled 9 patients.
#6 pts have completed follow up and 3 are still under observation.
#The maximum follow-up time for toxicity outcomes is 4 (weeks).
#The maximum follow-up time for efficacy outcomes is 12 (weeks.
d.obs=c(1,1,1,2,2,2,3,3,3)
y.obs=c(0,0,0,1,0,0,1,0,0) #observed tox data
z.obs=c(0,1,0,1,1,0,0,1,0) #observed eff data
t.obs=c(12,12,12,12,12,12,2,3,4) #follow up time for each patient
#Specify design parameters:
p.skel=c(0.01,0.02,0.15,0.22,0.29,0.36)
q.skel=get.qskel(max.eff=0.6, min.eff=0.1, ndose=6, pos.desc=TRUE)
tul=0.33;ell=0.05;N=64;ar.strategy ="Modified";n.ar=1; drop.rate=2;
#Implement the TITE bpocrm model.
bpocrmTITE.imp(d.obs,y.obs,z.obs,t.obs, p.skel,q.skel,tul,ell,N,
               ar.strategy="Modified",n.ar=1, drop.rate=2,
               max.T.tox=4, max.T.eff=12)
```

---

bpocrmTITE.sim — *bpocrm.sim.TITE simulates the time-to-event extension of Bivariate Partial Order CRM.*

Description

Simulation function for TITE extension of BPOCRM. This function is used to generate simulation replicates of Phase I/II trials using time-to-event extension of Bivariate Partial Order CRM under a specified dose-toxicity and dose-efficacy configuration.

Usage

```
bpocrmTITE.sim(p0, q0, p.skel, q.skel, tul, ell, N, start.set, ntrial,
  ar.strategy, n.ar, drop.rate, max.T.tox, max.T.eff, pt.rate = pt.rate,
  pt.dist = pt.dist, tox.TITE = tox.TITE, eff.TITE = eff.TITE)
```

Arguments

p0	True toxicity probabilities. A vector of values between 0 and 1 with length equal to the number of candidate doses. Represents the probability of observing a dose-limiting toxicity from the lowest dose to the highest dose, respectively.
q0	True efficacy probabilities. A vector of values between 0 and 1 with length equal to the number of candidate doses. Represents the probability of observing an efficacy response from the lowest dose to the highest dose, respectively.
p.skel	Toxicity skeleton. A vector of values for the initial geusses of toxicity probabilities at each dose level.
q.skel	Efficacy skeletons. A matrix of efficacy values. Each row of the matrix is an efficacy skeleton. Each efficacy skeleton represents a potential dose-response curve. Use get.qskel() to generate a generic matrix.
tul	Toxicity upper limit. A constant between 0 and 1. The highest rate of toxicity tolerated.

ell	Efficacy lower limit. A constant between 0 and 1. The minimum rate of efficacy to continue the trial.
N	Number of patients to be enrolled for this trial.
start.set	The dose level at which the trial starts. If length >1, a random start will be selected.
ntrial	Number of iterations to repeat bpocrm. If ntrial=1, a detailed dose-finding process is returned, including dose assignment, toxicity and efficacy outcomtes of every patient. A plot is also generated to visualize the dose-finding process.
ar.strategy	"Original" or "Modified". If ar.strategy="Original", then n.ar need to be specified. Refer to Wages and Tait (2015) If ar.strategy="Modified", then drop.rate need to be specified. Refer to Yan et. al. (In review)
n.ar	Sample size for adaptive randomization stage. Required when using "Original" ar.strategy. This argument is ignored when ar.strategy="Modified".
drop.rate	Required when using "Modified" ar.strategy. This rate decides how fast we eliminate unfit efficacy models from randomization. drop.rate=2 is recommended. This argument is ignored when ar.strategy="Original".
max.T.tox	Maximum follow up time for toxicity outcomes.
max.T.eff	Maximum follow up time for efficacy outcomes.
pt.rate	Patient recruitment rate. The expected number of patients to be recruited per unit of time. For example, If the unit of time is week, rate=0.5 means we expect to enroll 1 patient every 2 weeks.
pt.dist	Distribution of patient arrival time. "P"=Poisson distribution;"f"=fixed intervals.
tox.TITE	Distribution of toxicity minimum follow up time. Toxicity responses are only observable after the minimum follow up time. "U"= Uniform. "W"=Weibull.
eff.TITE	Distribution of efficacy minimum length of follow up. "U"= Uniform. "W"=Weibull. Efficacy responses are only observable after the minimum follow up time.

Value

A list of objects

True.tox	True probability of toxicity at each dose level. Same as p0
True.eff	True probability of efficacy at each dose level. Same as q0
Selection.pct	Percent of each dose being selected as final dose.
Treat.avg	Number of patients treated at each dose level.
n.tox	Number of patients experienced DLT.
n.eff	Number of patients experienced efficacy.
total.time	Average trial duration.

References

Cheung, Y. K., & Chappell, R. (2000). Sequential Designs for Phase I Clinical Trials with Late-Onset Toxicities. Biometrics, 56(4), 1177–1182. http://doi.org/10.1111/j.0006-341X.2000.01177.x

Yan, D., Wages, N. A., Tait, C., Kindwall-Keller, T., & Dressler, E. V. (in review). Bivariate Generalization of the Time-to-Event Continual Re- assessment Method. Journal of the Royal Statistical Society. Series C: Applied Statistics Royal Statistical Society.

Examples

```
#Assume p0 and q0 are the true probability of toxicity/efficacy at each dose level
set.seed(2017)
p0=c(0.05,0.10,0.20,0.28,0.50,0.50)
q0=c(0.05,0.13,0.25,0.38,0.50,0.63)
#p.skel and q.skel are the skeletons for dose-toxicity and dose-efficacy model
p.skel=c(0.01,0.02,0.15,0.22,0.29,0.36)
q.skel=get.qskel(max.eff=0.6, min.eff=0.1, ndose=6, pos.desc=TRUE)
#If we simulate one trial, a plot will be generated to show the dose-assignment of each patient
bpocrmTITE.sim(p0,q0,p.skel,q.skel,tul=0.33,ell=0.05,N=64, start.set=c(1,2,3),
ntrial=1, ar.strategy="Modified", n.ar=64 ,drop.rate=2,
max.T.tox=4,max.T.eff=12, pt.rate=2, pt.dist="P", tox.TITE="U", eff.TITE="W")

#We can simulate 1000 trials to examine the operating characteristics of the proposed design.
## Not run:
bpocrmTITE.sim(p0,q0,p.skel,q.skel,tul=0.33,ell=0.05,N=64, start.set=c(1,2,3),
 ntrial=1000, ar.strategy="Modified", n.ar=64 ,drop.rate=2,
 max.T.tox=4,max.T.eff=12, pt.rate=2, pt.dist="P", tox.TITE="U", eff.TITE="W")
## End(Not run)
```

---

get.qskel *Generates a generic dose-efficacy skeleton matrix*

Description

get.qskel generates a generic matrix of dose-efficacy skeletons.

Usage

```
get.qskel(max.eff = 0.6, min.eff = 0.1, ndose = 6, pos.desc)
```

Arguments

max.eff	Presumed maximum probability of efficacy at any dose level.
min.eff	Presumed minimum probability of efficacy at any dose level.
ndose	Number of candidate doses.
pos.desc	1 or 0. 1= dose-efficacy may decrease or plateau after the OBD; 0=dose-efficacy may plateau after OBD but will not decrease.

Details

get.qskel() generates a generic matrix of dose-efficacy skeletons. If efficacy may decrease or plateau after the Optimal Biological Dose (OBD), then there are 2I-1 efficacy skeletons, where I is the number of candidate doses.
If efficacy may plateau after OBD but will not decrease, then there are I efficacy skeletons.

Value

A matrix of efficacy skeletons. Each row of this matrix is an efficacy skeleton which is correspond to an efficacy model.

References

Yan, D., Wages, N. A., Tait, C., Kindwall-Keller, T., & Dressler, E. V. (in review). Bivariate Generalization of the Time-to-Event Continual Re- assessment Method. Journal of the Royal Statistical Society. Series C: Applied Statistics Royal Statistical Society.

Examples

```
#Example 1: Suppose we have 6 dose levels. The highest presumed efficacy
#probabiliy is 0.6, and the lowest is 0.1.
#Assuming the probability of efficacy may decrease after the OBD.
#The get.qskel() function will generate 11 different efficacy skeletons
#to incoporate potential dose efficacy curves.
get.qskel(max.eff=0.6, min.eff=0.1, ndose=6, pos.desc=TRUE)
```

Index

Bibliography

[1] National Center for Health Statistics, *Health, United States, 2016: With Chartbook on Long-term Trends in Health*. 2017.

[2] National Cancer Institute, "SEER Cancer Statistics Review," 2015.

[3] R. Collier, "Legumes, lemons and streptomycin: A short history of the clinical trial," *CMAJ : Canadian Medical Association Journal*, vol. 180, no. 1, pp. 23–24, 2009.

[4] A. Bhatt, "Evolution of Clinical Research: A History Before and Beyond James Lind," *Perspectives in Clinical Research*, vol. 1, no. 1, pp. 6–10, 2010.

[5] Streptomycin in Tuberculosis Trials Committee, "Streptomycin Treatment of Pulmonary Tuberculosis: A Medical Research Council Investigation," *British Medical Journal*, vol. 2, no. 4582, pp. 769–782, 1948.

[6] T. S. Lawrence, C. Runowicz, S. Turner, and J. L. Wade III, "Critical role of phase I clinical trials in cancer treatment," *Journal of Clinical Oncology*, vol. 15, no. 2, pp. 853–859, 1997.

[7] B. A. Chabner and T. G. Roberts, "Chemotherapy and the war on cancer," 2005.

[8] National Cancer Institute, "Common Terminology Criteria for Adverse Events (CTCAE). Version 4.0.," 2009.

[9] S. K. Carter, "Study design principles for the clinical evaluation of new drugs as developed by the chemotherapy programme of the National Cancer Institute," in *The Design of Clinical Trials in Cancer Therapy* (M. Staquet, ed.), 1973.

[10] W. J. Dixon and A. M. Mood, "A Method for Obtaining and Analyzing Sensitivity Data," *Journal of the American Statistical Association*, vol. 43, no. 241, p. 109, 1948.

[11] S. D. Durham, N. Flournoy, and W. F. Rosenberger, "A random walk rule for phase I clinical trials.," *Biometrics*, vol. 53, no. 2, pp. 745–760, 1997.

[12] A. Ivanova, A. Montazer-Haghighi, S. G. Mohanty, and S. D. Durham, "Improved up-and-down designs for phase I trials.," *Statistics in medicine*, vol. 22, no. 1, pp. 69–82, 2003.

[13] M. Stylianou and D. A. Follmann, "The Accelerated Biased Coin Up-and-Down Design in Phase I Trials," *Journal of Biopharmaceutical Statistics*, vol. 14, no. 1, pp. 249–260, 2004.

[14] C. Le Tourneau, J. J. Lee, and L. L. Siu, "Dose Escalation Methods in Phase I Cancer Clinical Trials," *JNCI Journal of the National Cancer Institute*, vol. 101, no. 10, pp. 708–720, 2009.

[15] J. O'Quigley, M. Pepe, and L. Fisher, "Continual reassessment method: a practical design for phase 1 clinical trials in cancer.," *Biometrics*, vol. 46, no. 1, pp. 33–48, 1990.

[16] J. O'Quigley and M. Conaway, "Continual Reassessment and Related Dose-Finding Designs.," *Statistical science : a review journal of the Institute of Mathematical Statistics*, vol. 25, no. 2, pp. 202–216, 2010.

[17] S. M. Lee and Y. K. Cheung, "Model Calibration in the Continual Reassessment Method," *Clinical trials (London, England)*, vol. 6, no. 3, pp. 227–238, 2009.

[18] E. L. Korn, D. Midthune, T. T. Chen, L. V. Rubinstein, M. C. Christian, and R. M. Simon, "A comparison of two phase I trial designs," *Statistics in medicine*, vol. 13, no. 18, pp. 1799–1806, 1994.

[19] E. Reiner, X. Paoletti, and J. O'Quigley, "Operating characteristics of the standard phase I clinical trial design," *Computational Statistics & Data Analysis*, vol. 30, no. 3, pp. 303–315, 1999.

[20] A. Iasonos, A. S. Wilton, E. R. Riedel, V. E. Seshan, and D. R. Spriggs, "A comprehensive comparison of the continual reassessment method to the standard 3 + 3 dose escalation scheme in Phase I dose-finding studies," *Clinical Trials*, vol. 5, no. 5, pp. 465–477, 2008.

[21] M.-K. Riviere, C. Le Tourneau, X. Paoletti, F. Dubois, and S. Zohar, "Designs of drug-combination phase I trials in oncology: a systematic review of the literature," *Annals of Oncology*, vol. 26, no. 4, pp. 669–674, 2015.

[22] C. Chiuzan, J. Shtaynberger, G. A. Manji, J. K. Duong, G. K. Schwartz, A. Ivanova, and S. M. Lee, "Dose-finding designs for trials of molecularly targeted agents and immunotherapies.," *Journal of biopharmaceutical statistics*, vol. 27, no. 3, pp. 477–494, 2017.

[23] Y. K. Cheung and R. Chappell, "Sequential Designs for Phase I Clinical Trials with LateOnset Toxicities," *Biometrics*, vol. 56, no. 4, pp. 1177–1182, 2000.

[24] J. Babb, A. Rogatko, and S. Zacks, "Cancer phase I clinical trials: Efficient dose escalation with overdose control," *Statistics in Medicine*, vol. 17, no. 10, pp. 1103–1120, 1998.

[25] D. H. Leung and Y. Wang, "Isotonic designs for phase I trials.," *Controlled clinical trials*, vol. 22, no. 2, pp. 126–138, 2001.

[26] Y. Yuan, K. R. Hess, S. G. Hilsenbeck, and M. R. Gilbert, "Bayesian Optimal Interval Design: A Simple and Well-Performing Design for Phase I Oncology Trials.," *Clinical cancer research : an official journal of the American Association for Cancer Research*, vol. 22, no. 17, pp. 4291–4301, 2016.

[27] A. R. Reynolds, "Potential Relevance of Bell-Shaped and U-Shaped Dose-Responses for the Therapeutic Targeting of Angiogenesis in Cancer," *Dose-Response*, vol. 8, no. 3, pp. 253–284, 2010.

[28] G. Klement, S. Baruchel, J. Rak, S. Man, K. Clark, D. J. Hicklin, P. Bohlen, and R. S. Kerbel, "Continuous low-dose therapy with vinblastine and VEGF receptor-2 antibody induces sustained tumor regression without overt toxicity.," *The Journal of clinical investigation*, vol. 105, no. 8, pp. R15–24, 2000.

[29] M. Reck, J. von Pawel, P. Zatloukal, R. Ramlau, V. Gorbounova, V. Hirsh, N. Leighl, J. Mezger, V. Archer, N. Moore, and C. Manegold, "Phase III trial of cisplatin plus gemcitabine with either placebo or bevacizumab as first-line therapy for nonsquamous non-small-cell lung cancer: AVAil.," *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, vol. 27, no. 8, pp. 1227–1234, 2009.

[30] M. Reck, J. von Pawel, P. Zatloukal, R. Ramlau, V. Gorbounova, V. Hirsh, N. Leighl, J. Mezger, V. Archer, N. Moore, and C. Manegold, "Overall survival with cisplatin-gemcitabine and bevacizumab or placebo as first-line therapy for nonsquamous non-small-cell lung cancer: results from a randomised phase III trial (AVAiL).," *Annals of oncology : official journal of the European Society for Medical Oncology*, vol. 21, no. 9, pp. 1804–1809, 2010.

[31] A. Sandler, R. Gray, M. C. Perry, J. Brahmer, J. H. Schiller, A. Dowlati, R. Lilenbaum, and D. H. Johnson, "Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer.," *The New England journal of medicine*, vol. 355, no. 24, pp. 2542–2550, 2006.

[32] C. Aghajanian, S. V. Blank, B. A. Goff, P. L. Judson, M. G. Teneriello, A. Husain, M. A. Sovak, J. Yi, and L. R. Nycum, "OCEANS: a randomized, double-blind, placebo-controlled phase III trial of chemotherapy with or without bevacizumab in patients with platinum-sensitive recurrent epithelial ovarian, primary peritoneal, or fallopian tube cancer.," *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, vol. 30, no. 17, pp. 2039–2045, 2012.

[33] J. E. Kurtz, A.-C. Hardy-Bessard, M. Deslandres, S. Lavau-Denes, R. Largillier, C. Roemer-Becuwe, B. Weber, C. Guillemet, D. Paraiso, and E. Pujade-Lauraine, "Cetuximab, topotecan and cisplatin for the treatment of advanced cervical cancer: A phase II GINECO trial.," *Gynecologic oncology*, vol. 113, no. 1, pp. 16–20, 2009.

[34] O. Engebraaten, H. Edvardsen, E. Lokkevik, B. Naume, V. Kristensen, L. Ottestad, and V. Natarajan, "Gefitinib in Combination with Weekly Docetaxel in Patients with Metastatic Breast Cancer Caused Unexpected Toxicity: Results from a Randomized Phase II Clinical Trial.," *ISRN oncology*, vol. 2012, p. 176789, 2012.

[35] A. Manji, I. Brana, E. Amir, G. Tomlinson, I. F. Tannock, P. L. Bedard, A. Oza, L. L. Siu, and A. R. Abdul Razak, "Evolution of clinical trial design in early drug development: Systematic review of expansion cohort use in single-agent phase i cancer trials," *Journal of Clinical Oncology*, vol. 31, no. 33, pp. 4260–4267, 2013.

[36] A. Hoering, A. Mitchell, M. LeBlanc, and J. Crowley, "Early phase trial design for assessing several dose levels for toxicity and efficacy for targeted agents.," *Clinical trials (London, England)*, vol. 10, no. 3, pp. 422–9, 2013.

[37] G. Yin, S. Zheng, and J. Xu, "Two-stage dose finding for cytostatic agents in phase I oncology trials," *Statistics in Medicine*, vol. 32, no. 4, pp. 644–660, 2013.

[38] T. M. Braun, "The bivariate continual reassessment method. extending the CRM to phase I trials of two competing outcomes.," *Controlled clinical trials*, vol. 23, no. 3, pp. 240–56, 2002.

[39] P. F. Thall and J. D. Cook, "Dose-finding based on efficacy-toxicity trade-offs.," *Biometrics*, vol. 60, no. 3, pp. 684–93, 2004.

[40] B. Nebiyou Bekele and Y. Shen, "A Bayesian Approach to Jointly Modeling Toxicity and Biomarker Expression in a Phase I/II DoseFinding Trial," *Biometrics*, vol. 61, no. 2, pp. 343–354, 2005.

[41] G. Yin, Y. Li, and Y. Ji, "Bayesian dose-finding in phase I/II clinical trials using toxicity and efficacy odds ratios.," *Biometrics*, vol. 62, no. 3, pp. 777–84, 2006.

[42] W. Zhang, D. J. Sargent, and S. Mandrekar, "An adaptive dose-finding design incorporating both toxicity and efficacy.," *Statistics in medicine*, vol. 25, no. 14, pp. 2365–2383, 2006.

[43] Y. Yuan and G. Yin, "Bayesian dose finding by jointly modelling toxicity and efficacy as time-to-event outcomes," *Journal of the Royal Statistical Society. Series C: Applied Statistics*, vol. 58, no. 5, pp. 719–736, 2009.

[44] N. A. Wages and C. Tait, "Seamless Phase I/II Adaptive Design for Oncology Trials of Molecularly Targeted Agents," *J Biopharm Stat*, vol. 25, no. 5, pp. 903–920, 2015.

[45] M.-K. Riviere, Y. Yuan, J.-H. Jourdan, F. Dubois, and S. Zohar, "Phase I/II dose-finding design for molecularly targeted agent: Plateau determination using adaptive randomization," *Statistical Methods in Medical Research*, p. 096228021663176, 2016.

[46] M. Gasparini and J. Eisele, "A curve-free method for phase I clinical trials.," *Biometrics*, vol. 56, no. 2, pp. 609–615, 2000.

[47] Y. Zang, J. J. Lee, and Y. Yuan, "Adaptive Designs for Identifying Optimal Biological Dose for Molecularly Targeted Agents," *Clinical trials (London, England)*, vol. 11, no. 3, pp. 319–327, 2014.

[48] N. A. Wages, M. R. Conaway, and J. O'Quigley, "Continual reassessment method for partial ordering," *Biometrics*, vol. 67, no. 4, pp. 1555–1563, 2011.

[49] Y. Yuan, H. Q. Nguyen, and P. F. Thall, *Bayesian designs for phase I-II clinical trials*. Boca Raton, FL: Chapman and Hall/CRC, first edition ed., 2016.

[50] N. A. Wages, C. A. Portell, M. E. Williams, M. R. Conaway, and G. R. Petroni, "Implementation of a model-based design in a phase 1b study of combined targeted agents," *Clinical Cancer Research*, 2017.

[51] N. A. Wages, C. L. J. Slingluff, and G. R. Petroni, "Statistical controversies in clinical research: early-phase adaptive design for combination immunotherapies.," *Annals of oncology : official journal of the European Society for Medical Oncology*, vol. 28, no. 4, pp. 696–701, 2017.

[52] B. Storer, "Design and analysis of phase I clinical trials," *Biometrics*, vol. 45, no. 3, pp. 925 – 937, 1989.

[53] A. Ivanova, "Escalation, group and A + B designs for dose-finding trials," *Statistics in Medicine*, vol. 25, no. 21, pp. 3668–3678, 2006.

[54] F. Li, C. Zhao, and L. Wang, "Molecular-targeted agents combination therapy for cancer: Developments and potentials," *International Journal of Cancer*, vol. 134, no. 6, pp. 1257–1269, 2014.

[55] S. J. Mandrekar, R. Qin, and D. J. Sargent, "Model-based phase I designs incorporating toxicity and efficacy for single and dual agent drug combinations: methods and challenges.," *Statistics in medicine*, vol. 29, no. 10, pp. 1077–83, 2010.

[56] F. Donate, G. C. Parry, Y. Shaked, H. Hensley, X. Guan, I. Beck, Z. Tel-Tsur, M. L. Plunkett, M. Manuia, D. E. Shaw, R. S. Kerbel, and A. P. Mazar, "Pharmacology of the novel antiangiogenic peptide ATN-161 (Ac-PHSCN-NH2): observation of a U-shaped dose-response curve in several preclinical models of angiogenesis and tumor growth.," *Clinical cancer research : an official journal of the American Association for Cancer Research*, vol. 14, no. 7, pp. 2137–2144, 2008.

[57] P. Khalili, A. Arakelian, G. Chen, M. L. Plunkett, I. Beck, G. C. Parry, F. Donate, D. E. Shaw, A. P. Mazar, and S. A. Rabbani, "A non-RGD-based integrin binding peptide (ATN-161) blocks breast cancer growth and metastasis in vivo.," *Molecular cancer therapeutics*, vol. 5, no. 9, pp. 2271–2280, 2006.

[58] K. L. van Golen, L. Bao, G. J. Brewer, K. J. Pienta, J. M. Kamradt, D. L. Livant, and S. D. Merajver, "Suppression of Tumor Recurrence and Metastasis by a Combination of the PHSCN Sequence and the Antiangiogenic Compound Tetrathiomolybdate in Prostate Carcinoma," *Neoplasia (New York, N.Y.)*, vol. 4, no. 5, pp. 373–379, 2002.

[59] O. Stoeltzing, W. Liu, N. Reinmuth, F. Fan, G. C. Parry, A. A. Parikh, M. F. McCarty, C. D. Bucana, A. P. Mazar, and L. M. Ellis, "Inhibition of integrin alpha5beta1 function with a small peptide (ATN-161) plus continuous 5-FU infusion reduces colorectal liver metastases and improves survival in mice.," *International journal of cancer*, vol. 104, no. 4, pp. 496–503, 2003.

[60] J. O'Quigley and S. Zohar, "Retrospective robustness of the continual reassessment method.," *Journal of biopharmaceutical statistics*, vol. 20, no. 5, pp. 1013–1025, 2010.

[61] R. K. Jain, J. J. Lee, D. Hong, M. Markman, J. Gong, A. Naing, J. Wheler, and R. Kurzrock, "Phase I oncology studies: evidence that in the era of targeted therapies patients on lower doses do not fare worse.," *Clinical cancer research : an official journal of the American Association for Cancer Research*, vol. 16, no. 4, pp. 1289–1297, 2010.

[62] T. Schenk, S. Stengel, and A. Zelent, "Unlocking the potential of retinoic acid in anticancer therapy.," *British journal of cancer*, vol. 111, no. 11, pp. 2039–2045, 2014.

[63] S. Lonial, B. M. Weiss, S. Z. Usmani, S. Singhal, A. Chari, N. J. Bahlis, A. Belch, A. Y. Krishnan, R. A. Vescio, M.-V. Mateos, A. Mazumder, R. Z. Orlowski, H. J. Sutherland, J. Blade, E. C. Scott, H. Feng, I. Khan, C. M. Uhlar, T. Ahmadi, and P. M. Voorhees, "Phase II study of daratumumab (DARA) monotherapy in patients with 3 lines of prior therapy or double refractory multiple myeloma (MM): 54767414MMY2002 (Sirius).," *Journal of Clinical Oncology*, vol. 33, no. 18_suppl, pp. LBA8512–LBA8512, 2015.

[64] B. Guo and Y. Yuan, "Bayesian Phase I/II Biomarker-based Dose Finding for Precision Medicine with Molecularly Targeted Agents," *Journal of the American Statistical Association*, vol. 112, no. 518, pp. 0–0, 2016.

[65] T. M. Braun, "Generalizing the TITE-CRM to adapt for early- and late-onset toxicities," *Statistics in Medicine*, vol. 25, no. 12, pp. 2071–2083, 2006.

[66] Y. Cheung, *Dose Finding by the Continual Reassessment Method*. 2011.

[67] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S*. New York: Springer, fourth ed., 2002. ISBN 0-387-95457-0.

[68] K. Cheung, *dfcrm: Dose-finding by the continual reassessment method*, 2013. R package version 0.2-2.

[69] S. Dorai-Raj, *binom: Binomial Confidence Intervals For Several Parameterizations*, 2014. R package version 1.1-1.

Vita

Education

PhD candidate Epidemiology and Biostatistics University of Kentucky Lexington, KY.	2014 - present
M.S. Statistics Kansas State University Manhattan, KS.	2012 - 2014
B.S. Mathematics Northern Arizona University Flagstaff, AZ.	2010 - 2012

Professional Experience

Graduate Research Assistant Markey Cancer Center University of Kentucky Lexington, KY.	2016 - present
Graduate Teaching Assistant College of Public Health University of Kentucky Lexington, KY.	2015 - 2016
Research Assistant Center for Health Service and Research College of Public Health University of Kentucky Lexington, KY.	2014 - 2015
Research Assistant Applied Statistics Lab Kansas State University Manhattan, KS.	2013 - 2014
Graduate Teaching Assistant Department of Statistics Kansas State University Manhattan, KS.	2012 - 2013

Award and Certificates

Travel Grant University of Kentucky	2016
SAS Advanced Programmer Certificate SAS institute	2014
Presentation Award Northern Arizona University	2012
Vesto M. Slipher Scholarship Northern Arizona University	2011

Publications

1. (In Revision) Yan, D., Wages, N. A., Tait, C., Kindwall-Keller, T., & Dressler, E. V. Bivariate Generalization of the Time-to-Event Continual Re- assessment Method. Journal of the Royal Statistical Society. Series C: Applied Statistics Royal Statistical Society.
2. (In Revision) Yan, D., Wages, N., Dressler, V. Improved Adaptive Randomization Strategies for the Multidimensional CRM in Phase I/II Trials. Journal of Biopharmaceutical Statistics.
3. Book Contribution: (2017). Dressler, E. V., & Yan, D. (2017). Model-Based Designs for Safety In Handbook of Methods for Designing, Monitoring, and Analyzing Dose Finding Trials. In J. OQuigley, A. Iasonos, & B. Bornkamp (Eds.). Chapman and Hall/CRC.
4. (In press) Patel, R., Maxwell, S., Yan, D., Dressler, E., Romond, E., Mathew A. (2017) Medical Oncologists Perception of Anti-Estrogen Therapy Benefit in Premenopausal Women with Hormone Receptor-Positive Early Breast Cancer. Annuals of Internal Medicine.
5. Hough, A. R., Nechols, J. R., McCornack, B. P., Margolies, D. C., Sandercock, B. K., Yan, D., Murray, L. (2016). The Effect of Temperature and Host Plant Resistance on Population Growth of the Soybean Aphid Biotype 1 (Hemiptera: Aphididae). Environmental Entomology, 1(1), nvw160. http://doi.org/10.1093/ee/nvw160
6. Fanucchi, L., Yan, D., & Conigliaro, R. L. (2016). Duly noted: Lessons from a two-site intervention to assess and improve the quality of clinical documentation in the electronic health record. Applied Clinical Informatics, 7(3), 653659. http://doi.org/10.4338/ACI-2016-02-CR-0025