SAPIENZA
UNIVERSITÀ DI ROMA

Some Contributions to Phase I and II Clinical Trials: Incorporating Patient Characteristics and Potential Time Trends into Designs and Analysis

Scuola di dottorato di Scienze Statistiche

Dottorato di Ricerca in Statistica Metodologica – XXX Ciclo

Candidate

Ilaria Domenicano
ID number 1185853

Thesis Advisors

Prof. Lorenzo Trippa
Prof. Steffen Ventz

A thesis submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Methodological Statistics

February 2018

Thesis defended on February 2018
in front of a Board of Examiners composed by:

Prof.ssa Cira Perna (chairman)

Prof. Nicola Sartori

Prof. Paolo Giudici

Some Contributions to Phase I and II Clinical Trials: Incorporating Patient Characteristics and Potential Time Trends into Designs and Analysis
Ph.D. thesis. Sapienza – University of Rome

This thesis has been typeset by LaTeX and the Sapthesis class.

Author's email: ilaria.domenicano@uniroma1.it

A mio fratello Giovanni,

perche' mentre io portavo avanti una tesi,

lui portava avanti la nostra famiglia.

Grazie.

Contents

Preface

This work summarizes the two years of research that I have conducted at Dana-Farber Cancer Institute(DFCI)/Harvard T.H. Chan School of Public Health, in Boston (MA, USA), where I collaborated with Lorenzo Trippa (Associate Professor at Harvard University and Dana Farber Cancer Institute) and Steffen Ventz (Assistant Professor at University of Rhode Island).

The thesis is divided in two main parts. The first part represents the main contribute of my research and on which I spent a dominant portion of my PhD period. In this part, called "Bayesian Uncertainty-Directed Dose Finding Designs", we introduce Bayesian uncertainty directed (BUD) designs for dose finding trials. This class of designs assigns patients to candidate dose levels with the aims of maximizing explicit information metrics at completion of the trial, while also avoiding the treatment of patients with toxic or ineffective dose levels during the trial. Explicit information metrics provide, at completion of the clinical trial, accuracy measures of the final selection of optimal or nearly optimal dose levels. The BUD approach utilizes the decision theoretic framework, and builds on utility functions that rank candidate dose levels. The utility of a dose combines the probabilities of toxicity events and the probability of a positive response to treatment. We discuss the application of BUD designs in three distinct settings; (i) dose finding studies for single agents, (ii) dose optimization for combination therapies of multiple agents, and (iii) precision medicine studies with biomarker measurements that allow dose optimization at the individual level. The proposed approach and the simulation scenarios used in evaluation of BUD designs are motivated by a Stereotactic Body

Radiation Therapy (SBRT) study in lung cancer at Dana Farber Cancer Institute.

The second part of the thesis, called "Inference in Adaptive Trials under Time Trends in the Patient Population", is a smaller project that we started only a few months ago, and thus many questions about the topic have not been investigated yet. The project addresses the problem of changes in the patient population over time during a clinical trial.

Standard analysis methods in clinical trials implicitly assume that the patient characteristics do not change over time, and the treatment effect remains constant during the study period. Since trials run for many years, this hypothesis may not hold and time trends in the patient population can constitute a potential source of bias in both estimation and testing of the treatment effects. This is especially important for trials using adaptive randomization, where the randomization probabilities change as a function of the outcome observed during the trial. Consider a randomized two-arm trial of total sample size N with a binary endpoint. The response probability for the first $N/2$ patients is 0.2 for the control arm and 0.5 for the experimental arm. Due to changes in patient population, the response probabilities changes to 0.4 and 0.7 for the remaining patients in the two arms respectively. With balanced randomization (BR), where patients are allocated to the arms with equal probabilities, the expectation of the estimated overall response probabilities are 0.3 and 0.6 for the two arms, and the difference is 0.3, which is constant before and after the change. However, if response adaptive randomization is employed and the randomization probability changes to 2:1 for experimental vs control for the last $N/2$ patients, the expectation of the estimated overall response probabilities are now (0.2N/4 + 0.4N/6)/(N/4 + N/6) = 0.28 and (0.5N/4 + 0.7N/3)/(N/4 + N/3) = 0.61 for the control and experimental arms with a difference of 0.33, which is inflated by 10%.

In this work, we propose a procedure which reduces the bias of treatment effect estimates and preserves the frequentist operating characteristics. We account for time trends by using Generalized Additive Models (GAMs) to estimate the treatment effect. We then use a parametric bootstrap to obtain valid inferences for treatment

effects. The testing procedure can be implemented for any adaptive design and any estimator of the treatment effect.

We apply our procedure to some well-known Response Adaptive Randomization (RAR) designs to evaluate the performance of the proposed method. For each design, we assess the estimation and testing capabilities of the method by simulating different time trends in both standard multi-arm clinical trials and platform trials.

Part I

Bayesian Uncertainty-Directed Dose Finding Designs

Chapter 1

Introduction

We introduce Bayesian uncertainty directed (BUD) designs for dose finding clinical trials. These designs build on joint Bayesian modeling of toxicity and efficacy endpoints. A BUD design assigns patients sequentially to candidate dose-levels of a single drug, or dose combinations of multiple drugs. The aim is to identify effective dose levels and, in some cases, to optimize the dose at the patient level accounting for relevant biomarkers or other individual characteristics. We define the optimal dose through a utility function that balances toxicities and treatment effects. BUD designs utilize metrics that quantify the accumulated information during the course of the trial. In particular, the information measures that we use are representative of the uncertainty level involved in selecting optimal treatments. Patients are sequentially assigned to different doses consistently with the primary aim of maximizing the accumulated information at completion of the study.

In Phase I trials treatments are often administered for the first time to humans. The goal is typically to identify a dose that is safe. In many disease areas it is common to assume that the clinical benefit and the probability of toxicities increase with the dose level. Phase I designs in oncology often estimate the maximum tolerated dose (MTD), which is the highest dose level with an acceptable probability of severe toxicity events. The 3+3 design of Storer [56] is the most frequently used design. The design is simple to use, but has well documented limitations; MTD estimates have

large variability [24], and the correct MTD is selected with low probability [58]. The continual reassessment method (CRM) of O'Quigley et al. [47] uses sequentially the accumulated data, estimates the dose-toxicity response curve, and assigns patients to dose levels that are close to the current MTD estimate. Extensions of the CRM have been discussed in the literature [34, 24, 46, 4].

Phase I designs that focus only on toxicities have been criticized, because they do not consider treatment efficacy and positive effects [79, 60, 42]. Both toxicity and efficacy endpoints can be used to guide dose selection: we mention the methodology developed by Gooley et al. [25] for bone marrow transplantation studies, and the study designs in [62, 10, 61, 45, 79, 42].

Recent dose finding designs include the estimation of personalized MTDs based on individual characteristics, for instance age and biomarkers-profiles [50, 48, 5, 65]. Methodological work has also been directed to the estimation of MTDs for combination therapies [64, 72, 78, 7]. Although the above designs have substantial differences in their goals and complexity levels, most of the designs that we mentioned seem to conform with the sequential assignment of patients either to the current estimate of the MTD or to the estimated optimal dose. These strategies can be suboptimal in accumulating information to accurately identify optimal dose levels at completion of the study. Indeed, as discussed in [66, 57] these greedy assignment rules often assign a substantial proportion of the enrolled patients to a single dose with efficacy and toxicity probabilities considerably different from the most effective does levels. The approach that we propose attempts to overcome this limitation.

Our study is motivated by a personalized dose finding trial for *Stereotactic Body Radiation Therapy* (SBRT) at our institute. The radiation dose to the tumor and surrounding organs are precisely measured and correlates with the probability of cancer control versus organ injury. However, the distribution of radiations across patients varies depending tumor size, and location. Patients will be assigned to dose levels after measurement of a relevant biomarker, the individual dose volume histogram (DVH).

The SBRT trial that we mentioned will estimate optimal radiation doses for patients with a broad spectrum of DVH curves. The individual optimal dose is defined through a utility that balance the probabilities of toxicities and positive treatment effects of candidate radiation intensity levels. For each enrolled patient, the BUD algorithm estimates (i) utilities of candidate doses, specific for the enrolled patient, and (ii) the potential gains of information on the relations that link efficacy and toxicities to the treatment dose. The gain of information coincides with the reduction of uncertainty on efficacy and toxicity regression curves determined by the assignment of the enrolled patient to a specific dose. Potential gains of information can vary substantially across dose levels, they tend to be more pronounced for dose levels that are unexplored and less for dose levels that have been repeatedly assigned. Patients assignment in BUD designs is sequentially driven both by utility estimates and by potential gains of information.

The next chapters are organized as follows. In Chapter 2 we describe the SBRT trial. Chapter 3 introduces a probability model that we use to define the BUD dose finding algorithms in Chapter 4. In Chapter 5 we discuss BUD methods for (i) a dose finding design with a single treatment, (ii) combination therapies, and (iii) personalized dose finding designs that utilize biomarkers to predict toxicities and treatment effects at various dose levels. A discussion is provided in Chapter 6.

Chapter 2

The SBRT trial

This work is motivated by a trial at our institution that seeks to define personalized Stereotactic Body Radiation Therapy (SBRT) doses for patients with lung cancer. SBRT is a novel radiation therapy technique that has been developed in the last decade. It is a relatively new treatment technique, which has become widely adopted for the treatment of lung cancer, but the exact dose tolerance of various organs to high dose radiation remain unclear. The trial has two primary endpoints; chest wall pain within 10 weeks of radiation $T \in \{0, 1\}$ and tumor shrinkage $E \in \{0, 1\}$ after 8 weeks of treatment. The trial aim is to estimate personalized optimal doses.

Conventional radiation is the standard of care for lung cancer patients when surgical resection can't be performed, with a local relapse rate of 55-70% and a 5-years survival probability of 15-30% [18]. SBRT can deliver significantly higher doses of radiation to the tumor, while controlling exposure of healthy surrounding tissues to radiation better than conventional therapy. Previous SBRT studies showed encouraging improvements in tumor response [55], but also an unexpected incidence of chest wall toxicities. Both Stephans et al. [55] and Dunlap et al. [18] indicate that these adverse events may be correlated with the administered dose and volume of chest wall exposed to radiations.

A tomography scan is used for radiation planning which entails delineating the tumor and each nearby organ. The software utilized to specify radiation plans

can be used to generate patient-specific DVH curves $h_{i,d}(\cdot)$ [35, 43]. Assuming a prescribed dose d will be administered to patient i, the quantity $h_{i,d}(d')$ indicates the volume of tissues that accumulate radiation intensities higher than d' Gy, for d' within a range of interest [17]. For instance $h_{i,d}(30) = 15\%$ indicates that 15% of the radiated chest wall will accumulate $30Gy$ or more. Patients treated with the same prescribed dose-level d of radiation can have very different DVH curves for organs such as the chest wall because of differences in the tumor location, chest volume and other factors. Ideally most of the high dose radiation should concentrate around the tumor tissue with only a small amount of chest wall exposed to radiation. The optimal dose $d_i^\star$ that we will define in Chapter 4 is patient specific and depends on the individual DVH curves $\{h_{i,d}; d \in \mathcal{D}\}$.

Chapter 3

Probability model

We consider a dose finding trial that assigns up to N patients to different treatment doses or combinations of treatments. For patient i the vector $(X_i, T_i, E_i) \in \mathcal{X} \times \{0,1\}^2$ includes binary toxicity and efficacy outcomes (T_i, E_i) and the covariate profile $X_i \in \mathcal{X} \subset \mathbb{R}^p$, $p \geq 1$, which consists of a set of patients' characteristics, for example age, and the assigned dose d. The outcome $T_i = 1$ indicates a toxicity event, chest wall toxicity in the SBRT trial. Symmetrically $E_i = 1$ corresponds to a positive response to the therapy, tumor shrinkage and control in the SBRT trial. Finally, $\Sigma_i = \{(X_j, T_j, E_j);\ j = 1, \ldots, i\}$ denotes the observed data for the first i patients.

We use a bivariate Probit model [13] for the treatment outcomes

$$Pr\left(E_i = 1, T_i = 1 | X_i = x, \beta\right) = \Phi\left(x'\beta^E, x'\beta^T; \rho\right), \tag{3.1}$$

with marginal probabilities $Pr\left(E_i = 1 | X_i = x, \beta\right) = \Phi\left(x'\beta^E, \infty; \rho\right)$ and $Pr(T_i = 1 | X_i = x, \beta) = \Phi\left(\infty, x'\beta^T; \rho\right)$. Here $\Phi(\cdot, \cdot; \rho)$ is the distribution function of a bivariate normal random vector with variances equal to one and correlation ρ. The BUD designs that we will study in this manuscript can be defined using alternative Bayesian models, and the methodological approach remains identical.

3.1 Treatments with a single agent

A dose finding trial for a single treatment considers J dose levels $\mathcal{D} = \{d_1, \cdots, d_J\}$. We initially assume that the covariate vectors X_i includes only the dose level, $X_i \in \{0,1\}^J$, and the j-th entry is equal to one if the patient is assigned to dose d_j. Therefore, if the patient receives dose d_j, then $Pr\left(E_i = 1, T_i = 1 | X_i, \beta\right) = \Phi\left(\beta_j^E, \beta_j^T; \rho\right)$, where (β_j^E, β_j^T) indicates the j-th entry of (β^E, β^T). The probabilities of toxicity and efficacy in most cases are monotone with respect to the dose level. In our Bayesian model the vectors β^T and β^E are *a priori* independent and we use a Gaussian prior with monotone mean functions μ^ℓ,$\ell = T, E$, for $\beta^\ell \sim N(\mu^\ell, \Psi^\ell)$, and the covariance values $\Psi^\ell_{j,j'}$ are set equal to $\lambda^\ell \times \min(d_j, d_{j'})$. In different words $\beta^\ell_{j+1} - \beta^\ell_j = \mu^\ell_{j+1} - \mu^\ell_j + \epsilon^\ell_{j+1}$ with $\epsilon^\ell_{j+1} \sim N(0, \lambda^\ell | d_{j+1} - d_j|)$.

We do not enforce monotonicity $\beta^\ell_j \leq \beta^\ell_{j+1}, \ell = T, E$, with prior probability one, but utilize increasing mean functions μ^ℓ_j. The prior distributions for μ^ℓ can range from linear functions, $\mu^\ell_j = \gamma^\ell_0 + \gamma^\ell_1 d_j, \gamma^\ell_1 \geq 0$, with a hyperprior on γ^ℓ, to independent densities for $(\mu^\ell_1, \mu^\ell_2 - \mu^\ell_1, \cdots, \mu^\ell_J - \mu^\ell_{J-1})$. For the correlation parameter ρ we use a mixture prior $\rho \sim \pi\delta_0(\rho) + (1-\pi)\text{Unif}(\rho, -1, 1)$, where δ_0 is the Dirac function and $\text{Unif}(\rho, -1, 1)$ indicates the uniform distribution on $[-1, 1]$.

3.2 Combination therapies

In Section 5.2 we discuss BUD designs for therapies with two agents, with dose levels $d \in \mathcal{D} = \mathcal{D}^1 \times \mathcal{D}^2$. Here $\mathcal{D}^r = \{d^r_1, \cdots, d^r_J\}$ indicates candidate doses for agent $r = 1, 2$. The covariate vector $X_i = (X^1_i, X^2_i, X^3_i)$ of the Probit model (3.1) includes, as in Section 3.1, vectors $X^r_i, r = 1, 2$, that indicate the assigned dose of treatment r, and a third component X^3_i to account for potential interactions. The interaction term can be defined by multiplication $X^3_i = (d^1(i) \times d^2(i))$, of the assigned doses $d^1(i)$ and $d^2(i)$ for patient i or other functions, we refer to [1] for a survey on modeling approaches and definitions of interaction terms for binary regressions.

We use independent Gaussian distributions for the prior probability of $\beta^\ell =$

$(\beta^{\ell,1}, \beta^{\ell,1}, \beta^{\ell,3})$, $\ell = T, E$. In particular $\beta^{\ell,r} \sim N(\mu^{\ell,r}, \Psi^{\ell,r})$, $r = 1, 2$, and also in this setting, as in Section 3.1, we use increasing linear functions $\mu^{\ell,r}$ and hyperpriors on the corresponding coefficients. We use the same covariance structure as earlier $\Psi^{\ell,r}_{j,j'} = \min(d_j, d_{j'}) \times \lambda^{\ell}_r$, $r = 1, 2$.

Recently Riviere et al. [52] studied several probability models for dose finding in combination therapy trials and showed with simulations that the Probit model is competitive and often superior to more complex regression models.

3.3 The use of biomarkers to optimize individual dose levels

In the SBRT trial that we described in Chapter 2, each patient is assigned to one of J dose levels $d \in \mathcal{D}$. Typically the administered dose d is not homogeneously distributed over the radiated volume of the chest wall [35, 43, 55, 18]. Depending on the size and location of the tumor, and the radiation angle, some tissues are exposed to more radiations than others.

Before treatment of patient i, a radiation oncology software quantifies for each dose $d \in \mathcal{D}$ a DVH curve $h_{i,d}(\cdot)$, where $h_{i,d}(z) \in [0, 1]$ indicates the volume of the chest wall that receives more than $z \geq 0$ Gy of radiation. The probability of a toxicity event increases with the volume exposed to high intensity of radiation. Therefore we use the patients' DVH to predict toxicity events.

We compute for each dose a summary, the generalized *equivalent uniform dose* (EUD) of the patient DVH curve, $Q_i(d) = (\sum_k z_k^a \times h_{i,d}(z_k))^{1/a}$. Here $\{z_k\}$ is a grid of equal spaced values, and the parameter a is chosen using the recommendations in [77]; for the chest wall we use $a = 5$ [77]. See [28] for an insightful discussion of EUDs in radiation oncology.

Ideally, the tumor should be radiated with high intensity and the surrounding chest wall should not be damaged by radiations.

For each candidate dose $d \in \mathcal{D}$ we compute the EUD $Q_i(d)$. High values of

$Q_i(d)$ tend to increase the risk of toxicity [77]. We include the EUDs in the binary regression model. Alternative summaries of the DHV curves, different from $Q_i(d)$, could be considered and utilized in the proposed BUD designs.

The probability model $Pr(E_i|X_i, \beta)$ for the efficacy endpoints E_i remains identical as in Section 3.1. For patient i, we define $X_i = (X_i^1, Q_i(d_1), \cdots, Q_i(d_J))$ and, if the patient is treated with dose d_j, then the individual probability of toxicity becomes $P(T_i = 1|X_i, \beta) = \Phi(\beta_j^T + \beta_j^Q \times Q_i(d_j))$.

To complete the Bayesian model we need a prior distribution on the parameters β^T and β^Q. One can combine the prior distribution of β^T in Section 3.1 and a joint Gaussian distribution for the biomarker effects $\beta^Q = (\beta_1^Q, \cdots, \beta_J^Q) \sim N(\mu^Q, \Psi^Q)$ with covariance $\Psi_{j,j'}^Q = \lambda^Q \exp\{-|d_j - d_{j'}|/\lambda_S\}$.

Chapter 4

BUD dose finding

The goal of a BUD design is the maximization of information at completion of the study. The design can be applied in various settings, including trials for a single treatment or a combination therapy, as well as studies that seek to optimize the dose level and treatment duration simultaneously [36]. For simplicity, we describe BUD designs for dose finding studies with a single treatment. Extensions to dose finding designs for combination therapies or trials that utilize individual biomarkers are described in Section 4.1. The BUD approach can be described in three steps:

U-Step (Utility of a Dose). We start by specifying a utility function $u(d)$ that ranks candidate dose levels $d \in \mathcal{D}$. Utility functions, as it is standard in decision theory [9], express preferences. We focus on maps $u(d) = u(p_{T,d}, p_{E,d})$ that are increasing in the probability of a positive response to treatment $p_{E,d}$ and decreasing in the probability of toxicity $p_{T,d}$. For simplicity, we use weighted combinations of these two probabilities

$$u(d) = w_1 \times (1 - p_{T,d}) + (1 - w_1) \times p_{E,d}, \tag{4.1}$$

where $w_1 \in [0, 1]$. In our Bayesian analysis, both $p_{T,d}$ and $p_{E,d}$ are random variables, jointly specified *a priori* and updated during the trial. It follows that also $u(d)$ is a random variable whose conditional distribution, given the data, changes during the

trial.

The Bayesian model (3.1) and the accumulated data Σ_i are used sequentially to estimate dose-specific utilities and to quantify uncertainty on these estimates. We use $d^\star \in \mathcal{D}$ to indicate the dose that maximizes the utility u among all candidate dose levels $\mathcal{D}$, i.e. we define the random variable $d^\star = \arg\max_{d\in\mathcal{D}} u(d)$.

I-Step (Information Measure). Efficacy and toxicity data from the first i patients enrolled in the trial allow one to estimate the optimal dose,

$$Pr(d^\star = d|\Sigma_i) = Pr\big(\cap_{d'\in\mathcal{D}} \{u(d) \geq u(d')\}|\Sigma_i\big), \text{ for } d \in \mathcal{D}. \tag{4.2}$$

The posterior of our Bayesian model (3.1) can be used to quantify the accumulated information. We select an information measure $\mathcal{I}(\Sigma_i)$. The function may take negative values. Large values of $\mathcal{I}(\Sigma_N)$ indicate that the trial allows to select dose levels of high utility with good accuracy. Various information measures can be used, we mention three examples:

(i) The (negative) posterior variance of the utility generated by $d^\star$, $\mathcal{I}_1(\Sigma_i) = -\text{Var}[u(d^\star)|\Sigma_i]$. A low posterior variance of $u(d^*)$ indicates low uncertainty about the potential benefits of the treatment.

(ii) We also consider the (negative) entropy of the posterior distribution (4.2) of the optimal dose, $\mathcal{I}_2(\Sigma_i) = \sum_{d\in\mathcal{D}} Pr(d^\star = d|\Sigma_i) \log Pr(d^\star = d|\Sigma_i)$. The measure takes its minimum $\mathcal{I}_2(\Sigma_i) = -\log|\mathcal{D}|$ when the posterior distribution of $d^\star$ is uniform, and $\mathcal{I}_2(\Sigma_i) = 0$ if $d^\star = d$ with posterior probability one for some dose $d \in \mathcal{D}$.

(iii) Consider the dose level $\widehat{d}_i^\star := \arg\min_{d\in\mathcal{D}} E([u(d) - u(d^\star)]^2|\Sigma_i]$ after observing the first i outcomes. If the total sample size of the trial was $N = i$, then we would select $\widehat{d}_i^\star$. The average squared distance between the utilities of $d^\star$ and $\widehat{d}_i^\star$, i.e. $\mathcal{I}_3(\Sigma_i) = -E\big([u(d^\star) - u(\widehat{d}_i^\star)]^2|\Sigma_i\big)$, is a measure of uncertainty. Values close to zero indicate a low risk of selecting a dose with utility substantially below the maximum $u(d^*)$.

A-Step (Dose Assignment Rule). We define the sequential dose assignment

algorithm. It assigns each patient to a dose $d \in \mathcal{D}$ by weighting two goals: (i) the approximate optimization of the accumulating information expressed by the metric $\mathcal{I}(\Sigma_i)$ (see I-Step) and (ii) the assignment of patients to doses with high expected utility and low-risk of toxicity events.

We use convex functions $Pr(\cdot|\Sigma_i) \rightarrow \widetilde{\mathcal{I}}[Pr(\cdot|\Sigma_i)]$ that translate the posterior $Pr(\cdot|\Sigma_i)$ into an information summary $\mathcal{I}(\Sigma_i) = \widetilde{\mathcal{I}}[Pr(\cdot|\Sigma_i)]$. We previously described three examples. Convexity implies that for any $0 \leq \alpha \leq 1$ and pair of distributions p_1 and p_2 on the parameter space of our Bayesian model

$$\widetilde{\mathcal{I}}(\alpha \times p_1 + (1-\alpha) \times p_2) \leq \widetilde{\mathcal{I}}(p_1) \times \alpha + \widetilde{\mathcal{I}}(p_2) \times (1-\alpha).$$

The use of convex information measures ensures, by Jensen's inequality, that the accumulated information will on average increase $E[\mathcal{I}(\Sigma_i)|X_i, \Sigma_{i-1}] \geq \mathcal{I}(\Sigma_{i-1})$ with each additional observation (X_i, E_i, T_i).

For each patient $1 \leq i \leq N$ we compute the summary $\mathcal{I}(\Sigma_{i-1})$. We then determine the expected increment in information $G_i(d)$ if the patient is assigned to dose $d_j \in \mathcal{D}$,

$$G_i(d) = E[\mathcal{I}(\Sigma_i)|X_i = d, \Sigma_{i-1}] - \mathcal{I}(\Sigma_{i-1}).$$

Here $X_i = d$ indicates the assignment of patient i to dose d. By convexity $G_i(d) \geq 0$, and large values of $G_i(d)$ correspond to large expected gains in information. Next we define scores that weight the gain in information $G_i(d)$ and the estimated utility of the dose

$$S_i(d) = w_2 \times G_i(d) + (1 - w_2) \times E[u(d)|\Sigma_{i-1}].$$

The parameter $w_2 \in [0, 1]$ weights the two aims of the dose finding algorithm. With $w_2 = 1$ the dose with the maximum expected increment in information has the highest score, while with $w_2 = 0$ the dose with the highest expected utility ranks first.

Lastly, we restrict the dose assignment of patient i to a dose in $\mathcal{A}_i \subset \mathcal{D}$ that

(i) has a predicted probability of toxicity $p_{T,d}$ below the pre-specifies threshold $p_{\max}$,

(ii) has an expected utility $E[u(d)|\Sigma_{i-1}]$ above the pre-specifies threshold $u_{\min}$,

(iii) and is similar to the last assigned dose level, that is, we only allow gradual escalation and deescalation of dose levels by at most ϵ units at each assignment.

Alternative toxicity criteria can be used to replace (i). For instance we may select all doses d such that $Pr(p_{T,d} < p_{\max}|\Sigma_i) > \Delta$. Here Δ represents a threshold for the posterior probability of $\{p_{T,d} < p_{\max}\}$. The last two constraints (ii) and (iii) in the definition of $\mathcal{A}_i$ may not be necessary for all dose finding trials, in such cases we set $(\epsilon, u_{\min}) = (\infty, 0)$. The constraints $\mathcal{A}_i$ are key components of dose finding trials to protect patients safety, and their definition can be tailored to the specific trial characteristics. As we describe in the next paragraph BUD designs are driven by the scores $S_i(d)$ and select the dose for patient i within $\mathcal{A}_i$.

If the set $\mathcal{A}_i$ is empty, then the trial is stopped because candidate dose levels are not safe or promising. Otherwise, patient i is assigned to a dose $d_j \in \mathcal{A}_i$ with probability

$$\frac{S_i(d)^c \times I(d \in \mathcal{A}_i)}{\sum_{d' \in \mathcal{A}_i} S_i(d')^c}. \tag{4.3}$$

where $c \geq 0$. With $c \approx \infty$ patients are assigned to the dose with the highest score within $\mathcal{A}_i$, whereas with $c = 0$ they are assigned with identical probabilities to any of the dose levels in $\mathcal{A}_i$. We can now summarize the Bayesian uncertainty directed dose finding design in `Algorithm` 2.

4.1 Personalize dose finding designs

The BUD algorithm that we introduced can be applied to trials with combination therapies, i.e. $D = \mathcal{D}^1 \times \mathcal{D}^2$, without any substantial change other than the Bayesian modeling of toxicity and efficacy endpoints. Additionally, in this Section, we describe

Algorithm 1 The Bayesian uncertainty directed (BUD) dose finding algorithm

for Each patient $1 \leq i \leq N$ **do**
  Compute the expected utility $E[u(d)|\Sigma_{i-1}]$ for each $d \in \mathcal{D}$.
  Compute the information increament $G_i(d)$ for each $d \in \mathcal{D}$.
  Determine the set of candidate doses $\mathcal{A}_i$ for patient i.
  if $\mathcal{A}_i = \emptyset$, **then** stop the trial
  else
    Set $S_i(d) = w_2 \times G_i(d) + (1 - w_2) \times E[u(d)|\Sigma_{i-1}]$ for each $d \in \mathcal{A}_i$.
    Randomize patient i to a dose $d \in \mathcal{A}_i$ with probability

$$\frac{S_i(d)^c}{\sum_{d' \in \mathcal{A}_i} S_i(d)^c}.$$

  end if
end for
Output: $\widehat{d^\star} = \arg\max E[u(d)|\Sigma_i]$.

how to use the design in personalized dose finding studies. The probability of dose related toxicity and efficacy at dose d depends on the patient covariates x. We indicate variations of the probabilities of toxicity and efficacy across values of these characteristics x by $p_{T,d}(x)$ and $p_{E,d}(x)$ for each dose $d \in \mathcal{D}$. Similar to Chapter 3 and the previous Section, the vector $X_i = (d, x)$ indicates the assigned dose d and other characteristics x, such as age. For instance, in Section 5.3, we use the EUD to predict the probability of toxicity for each candidate dose $d_j \in \mathcal{D}$, $p_{T,d_j}(x) = \Phi(\beta_j^T + \beta_j^Q \times x)$, when the EUD equals $Q_i(d_j) = x$.

The BUD approach now remains essentially identical and proceeds as follows:

U-Step. We use the same function $u(d, x) := u(p_{T,d}(x), p_{E,d}(x))$ as before in (4.1) to determine the utility of dose d for a patient with profile x. The personalized optimal dose for a patient with profile x is $d^\star(x) = \arg_{d \in \mathcal{D}} \max u(d, x)$.

I-Step. We use $\mathcal{I}(\Sigma_i, x)$ to indicate the accumulated information on dose levels with high utility for patients with profile x, for example $\mathcal{I}(\Sigma_i, x) = \sum_d Pr(d^\star(x) = d|\Sigma_i) \log Pr(d^\star(x) = d|\Sigma_i)$. The BUD algorithm seeks to maximize the average $\mathcal{I}(\Sigma_N) = E[\mathcal{I}(\Sigma_N, X)] = \int_{\mathcal{X}} \mathcal{I}(\Sigma_N, x) dPr(x)$ at completion of the trial. Here the information measure $\mathcal{I}(\Sigma_N, x)$ is averaged with respect to the distribution of the

patients covariates x. In our application, we use available data from the department of radiation oncology at our institution to estimate the distribution $Pr(x)$. This estimate is plugged into the BUD algorithm.

A-Step. For each patient i we compute the expected information gains

$$G_i(d,x) = E[\mathcal{I}(\Sigma_i)|X_i = (d,x), \Sigma_{i-1}] - \mathcal{I}(\Sigma_{i-1}),$$

$d \in \mathcal{D}$, and the score function $S_i(d,x) = w_2 \times G_i(d,x) + (1-w_2) \times E[u(d,x)|\Sigma_{i-1}]$. We then determine the set of candidate dose levels $\mathcal{A}_i(x)$ that (i) are safe $E[p_{T,d}(x)|\Sigma_{i-1}] \leq p_{\min}$ and (ii) have expected utility $E[u(d,x)|\Sigma_{i-1}]$ above a minimum threshold $u_{\min}$. If no such dose exists, $\mathcal{A}_i(x) = \emptyset$, the patient with profile x will not be enrolled in the trial because no safe and effective dose is available. Otherwise the randomization probabilities within $\mathcal{A}_i(x)$ remains proportional to $[S_i(d,x)]^c$, as in the previous Section.

When patient i with profile x is not treated because $\mathcal{A}_i(x)$ is empty, we determine if the trial should continued or not. We evaluate the proportion of patients that could be treated accordingly to the available data Σ_{i-1}. The relative proportion of this group of patients is $Pr(x : \mathcal{A}_i(x) \neq \emptyset)$; as we mentioned in our application the patients' profiles x distribution is estimated from historical data. If this proportion falls below 20% the trial is terminated.

Chapter 5

Simulation studies

We discuss the operating characteristics of BUD designs for dose finding studies. We initially investigated the sensitivity of the BUD design with respect to the tuning parameter w_2, which balances two competing goals. Recall that by selecting $w_2 = 0$ the design assigns each patient i to the dose d with highest expected utility, while at the opposite extreme $w_2 = 0$ maximizes the expected information gain $G_i(d)$.

We illustrate a simulation study for a single agent with $J = 5$ doses and probabilities of response and toxicity equal to $(0.1, 0.2, 0.25, 0.5, 0.54)$ and $(0.05, 0.07, 0.1, 0.15, 0.35)$. We set $w_1 = 0.55$ in the definition of the utility (4.1) and $u(d_j) = 0.56, 0.60, 0.60, 0.69$ and 0.60 for doses $j = 1, \cdots, 5$. We then generated 1,000 trials using $w_2 = 0, 0.1, \ldots, 0.9$ or 1. Column A of Figure 7.1, summarizes the operating characteristics for a trial with $N = 30$ patients. Different colors in Figure 7.1 correspond to different dose levels. Dose levels are ordered accordingly to their utility, from blue (high utility) to yellow (low utility). We used the information measure $\mathcal{I}_1(\Sigma_i) = -\text{Var}[u(d^\star)|\Sigma_i)$. The top row shows the proportion of simulations that selected dose $d \in \mathcal{D}$ at completion of the phase I study as the optimal dose for future stages of drug development. The bottom row shows the average number of patients treated with each dose during the study.

In our simulation the parameter w_2 has moderate effect on the operating characteristics, with relatively small variations for parameter values of $w_2 > 0.4$. Values of

$w_2 < 0.2$ lead to a large proportion of simulations that select a suboptimal dose with low utility and a large number of patients treated with a suboptimal dose. With $w_2 = 0$ BUD selects in 40.1% of the simulations the dose with the highest utility and assigns on average 6.1 patients to this dose. Values of w_2 above 0.4 lead to sufficient accumulation of information to select in most simulations a dose with high utility. For instance with $w_2 = 0.4, 0.7$ or 1 the design selects in 61.4%, 63.6% and 68.1% of all simulated trials the dose with the highest utility and assigns on average 11.2, 11.2 and 10.5 patients to this dose.

This result is in agreement with our simulations in Section 5.1. Designs like the bCRM of [10], which assign each patient to doses close to the current estimate of the optimal dose (similar to BUD with $w_2 = 0$), select in a relevant proportion of simulations a suboptimal dose and tend to assign fewer patients to the optimal arm than a BUD with $w_2 > 0.2$.

5.1 Dose finding trials for single-drug therapies

We consider BUD designs for a single treatment. We consider scenarios with a maximum sample size of $N = 30$ patients and $J = 5$ dose levels. For comparisons we selected the same simulation scenarios as in [42], which are described in the first column of Table 6.1. For each scenario, we simulated 1,000 trials using BUD designs and three alternative designs. The toxicity and efficacy outcomes have been generated from a bivariate Gumbel model as detailed in [44, 42] with outcome probabilities summarized in Table 6.1. To facilitate comparisons we assume that the outcomes are available immediately after treatment assignment. We will relax this assumption later in this manuscript and discuss the effects of delayed outcomes.

We consider three BUD designs, with utility $u(d)$ defined as in (4.1) with $w_1 = 0.55$ and set $c = 1$ in (4.3). Each BUD design adopts a different information measure using either the posterior variance of the utility $u(d^\star)$, the negative posterior entropy of $d^\star$ or the mean squared loss between $u(\widehat{d}_i^\star)$ and $u(d^\star)$. These are the

information measures discussed in Chapter 4. We indicate the three designs by BUD1, BUD2 and BUD3, respectively.

We compare the BUD design to the bivariate continuous reassessment method (bCRM) of Braun [10] and the design introduced in Liu and Johnson [42] (LJ). The bCRM computes, before the assignment of each patient i, the distances between the estimated probabilities of response and toxicity $(E[p_{E,d}|\Sigma_{i-1}], E[p_{T,d}|\Sigma_{i-1}])$ of dose $d \in \mathcal{D}$ and target probabilities (θ_E, θ_T) as described in Braun [10]. We indicate the distance by $r_i(d)$. The patient is then assigned to the dose that minimizes this distance. At the end of the trial the bCRM design recommends dose $\widehat{d}^\star = \arg\min_d r_{N+1}(d)$ for future stages of drug development. The LJ design defines the utility of a dose d as

$$u(d) = p_{E,d} - w_3 \times p_{T,d} - w_4 \times p_{T,d} \times I(p_{T,d} > \theta_T^\star),$$

where $\theta_T^\star$ is a toxicity threshold [42]. Before the enrollment of each patient i, the LJ design determines the estimate $\widehat{d}_i^\star$ equal to $\arg\max_{d\in\mathcal{D}} Pr(u(d^\star) = u(d)|\Sigma_i)$, and assigns the patient to either $\widehat{d}_i^\star$ or the two neighboring (above and below) doses with probability proportional to $Pr(u(d^\star) = u(d)|\Sigma_i)$. At the end of the trial dose $\widehat{d}_{N+1}^\star$ is recommended. We set $(w_3, w_4) = (0.33, 1.09)$ and $\theta_T^\star = 0.3$ in the LJ design, $(\theta_E, \theta_T) = (0.4, 0.2)$ in the bCRM. These values ensure, that the optimal dose levels under bCRM, LJ snd BUD designs are the same in each scenario.

The bCRM, LJ and BUD designs utilize the probability model described in Section 3.1. To facilitate comparisons we do not include early stopping. The variance parameters λ_ℓ, $\ell = T, E$, for $\beta_\ell \sim N(\mu^\ell, \Psi^\ell)$ equal 1.5, and mean function μ^ℓ is centered around $(-0.85, -0.5, -0.25, 0, 0.25)$ for $r = E$ and $(-1.65, -1.25, -0.85, -0.5, -0.4)$ for $r = T$.

Table 6.1 reports, for each scenario, the proportion of times that each dose was selected at the end of the trial as the optimal dose across 1000 simulations. Values in parenthesis indicate the average number of patients that have been treated with

dose $d \in \mathcal{D}$ across simulations.

In Scenario 1, all doses have a low probability of toxicity and dose d_5 has, with 0.43, a substantial larger probability of response to treatment compared all other dose levels. The BUD1, BUD2 and BUD3 recommend in 65%, 67% and 66% of all simulated trials dose five for future drug development, compared to 34% and 59% for the bCRM and the JL designs respectively. In Scenario 4, the third and the fourth dose d_4, d_5 have nearly identical utilities, 0.62 and 0.61 compared to values between 0.51 and 0.56 for the remaining doses. The bCRM and LJ designs select the first two doses d_1, d_2 substantially more frequently across simulations than the BUD designs. The bCRM selects dose one or two in 49% of all simulations, while LJ selects these doses in 36% of all simulations. In comparison the BUD designs select these two suboptimal doses substantially less frequently (13.8%, 12.4%, and 13.2% for BUD1, BUD2 and BUD3).

We show, in Column A of Figure 7.2, for each scenario, the expected number of responses to treatment and toxicity events in a hypothetical future cohort of 100 patients treated with the recommended dose $\widehat{d^\star}$. In scenario 1 the five designs lead on average between 90 and 92 out of 100 patients without toxicity events. The three BUD designs lead on average to 36 patients that respond to the treatment at dose $d^\star$, compared to 28 and 34 patients for the bCRM and LJ designs. The performance of the BUD designs compared to the bCRM and LJ design are similar in the remaining scenarios. The BUD2 design, based on the posterior entropy of $d^\star$, seems outperforms BUD1 and BUD3 in scenarios 1, 3 and 5.

We also investigated the operating characteristics of BUD designs when the outcomes are not available immediately after treatment assignment. The additional time between enrollment and outcome data reduces the information utilized by the dose-assignment algorithm. We repeated the simulations of BUD trials assuming an average enrollment rate of 24 patients per year. For each patient i, the treatment response E_i and the toxicity outcome T_i become available 12 weeks after treatment. Within this time window Bayesian stochastic imputation can be used to determine

the randomization probabilities. In the first scenario, BUD1, BUD2 and BUD3 select in 64.4%, 61.5% and 62.1% of all simulated BUD trials the dose with the highest utility, compared to 65.2%. 67.3% and 65.6% when the outcomes are immediately available after randomization, see Table 7.2. The operating characteristics with delayed outcomes are similar for the remaining three scenarios, the proportion of simulated trials that select the optimal dose decreases between 1% to 4 % for BUD1, 2% to 10% for BUD2, and 2% to 10% for BUD3.

5.2 Combination therapies

We can now consider a therapy that combines two agents. The first drug is administered at one of three different dose levels $\mathcal{D}_1 = \{d_1^1, d_2^1, d_3^1\}$, and the second has four dose levels $\mathcal{D}_2 = \{d_1^2, d_2^2, d_3^2, d_4^2\}$. Combination therapies require typically a larger sample size compared to single-agent trials [78]. We simulated trials with $N = 100$ patients.

We consider four simulation scenarios (see Table 7.3), and simulated 1,000 trials under each scenario. In scenarios one and three the combinations with the highest utility are (d_2^1, d_2^2) and (d_1^1, d_4^2). In scenario two all combinations have similar response and toxicity probabilities, and the combination (d_2^1, d_4^2) has slightly higher utility than the other combinations. In scenario 4, combinations (d_1^1, d_4^2) and (d_2^1, d_3^2) have both optimal utility, see supplementary Table 7.3.

For BUD simulations we used again the same three information measures $\mathcal{I}(\Sigma_i)$ as in Section 5.1. In all simulations the first patient is assigned to dose (d_1^1, d_1^2). The hyperparameters $\mu_1^{\ell,r}, \mu^{\ell,2} - \mu^{\ell,1}, \cdots$ for $\beta_\ell^r \sim N(\mu^{\ell,r}, \Psi^{l,r}), r = 1, 2$ have, for both $\ell = E, T$, means $(-0.5, 0.25, 0.25)$ for the drug $r = 1$ and $(-0.5, 0.25, 0.25, 0.25)$ for the second drug. We also specify $E[\mu_j^{\ell,3}] = 0$, $\lambda_\ell^1 = \lambda_\ell^2 = 1.5$ and $\lambda_\ell^3 = 1$ for $\ell = T, E$.

We compare BUD designs to three alternative designs; the Bayesian optimal interval (BOIN) design of Lin and Yin [40], the dose finding design DFcomb of Riviere et al. [51] and the dose finding design of Yin and Yuan [78] (YY design).

The first two designs use only toxicity endpoints to determine a combination dose d with toxicity profile close to a target toxicity level θ_T.

The YY design is a two stage design. Let θ_T and θ_E be upper-toxicity and lower-efficacy parameters. In the initial stage, N_1 patients are randomized to identify a set of candidate dose combinations such that $Pr(p_{T,d} < \theta_T|\Sigma_{N_1}) > \Delta_T$, where $\Delta_T \in [0, 1]$ is a predefined threshold. In the second stage N_2 patients are randomized with equal probabilities to the identified subset of doses d. If during the second stage the posterior $Pr(p_{E,d} > \theta_E|\Sigma_i)$ falls below the threshold Δ_E, then dose d is dropped. At the end of the study, among all dose levels that remained active the YY design selects the dose that maximizes $Pr(p_{E,d} > \theta_E|\Sigma_i)$.

The BOIN design assigns cohorts of patients sequentially to different dose levels. Given the current cohort ℓ at combination (d_i^1, d_j^2), the design escalates to either (d_{i+1}^1, d_j^2) or (d_i^1, d_{j+1}^2) if less than t_ℓ^1 toxicity events occurred. The definition of t_ℓ^1 is described in Yuan et al. [81]. If more than t_ℓ^2, but less than t_ℓ^3, toxicity events occurred, than the dose is de-escalated to either (d_{i-1}^1, d_j^2) or (d_i^1, d_{j-1}^2). Finally, if more than t_ℓ^3 events occur the current dose and all doses $(d_k^1, d_l^2), k \geq i, l \geq j$ are not considered anymore, and the combination (d_i^1, d_j^2) is reduced. When the design has to choose between (d_{i+1}^1, d_j^2) and (d_i^1, d_{j+1}^2), or between (d_{i-1}^1, d_j^2) and (d_i^1, d_{j-1}^2), the combination with the lower estimated toxicity is preferred [40]. At the end of the trial an isotonic regression model is fitted, and a targeted toxicity probability θ_T is used to select d such that $\widehat{p}_T(d) \approx \theta_T$.

The DFcomb design of Riviere et al. [51] has similar characteristics. Let Δ_E and Δ_D be thresholds for dose escalation and de-escalation, such that $\Delta_E + \Delta_D > 1$. If patient i was assigned to dose $d = (d_l^1, d_j^2)$, and $Pr(p_{T,d} < \theta_T|\Sigma_i)$ is larger than a threshold Δ_E, then the dose is increased to d' in $\{(d_{l+1}^1, d_j^2), (d_l^1, d_{j+1}^2), (d_{l+1}^1, d_{j-1}^2), (d_{l-1}^1, d_{j+1}^2)\}$; the combination with posterior mean $E[p_{T,d'}|\Sigma_i]$ closest to the target p_T^0 is preferred. If instead the posterior $Pr(p_{T,d} \geq \theta_T|\Sigma_i)$ is larger than Δ_D, then the dose is reduced to the d' in $\{(d_{l-1}^1, d_j^2), (d_{l-1}^1, d_{j+1}^2), (d_l^1, d_{j-1}^2), (d_{l+1}^1, d_{j-1}^2)\}$ which has an expected toxicity event closest to the target probability p_T^0. Finally, if

$Pr(p_{T,d} < \theta_T|\Sigma_i) < \Delta_E$ and $Pr(p_{T,d} \geq \theta_T|\Sigma_i) < \Delta_D$ then the next patient is assigned again to the current dose d. When the maximum sample size N is reached this design selects the dose d that maximizes $Pr(\theta_T - \delta \leq p_{T,d} \leq \theta_T + \delta|\Sigma_N)$, with $\delta > 0$.

For BOIN and DFcomb, we select design parameters that match the optimal dose combinations across designs, including the proposed BUD designs. Similarly for the YY design we selected $(N_1, N_2) = (25, 75)$ patients and tune (θ_T, θ_E) to obtain the same target dose across designs.

Figure 7.3 shows the number of times each dose combination is selected at the end of the trial. Doses are ordered from light to dark according to their utility. In Scenario 1, BUD1 and BUD2 select the optimal dose (dark blue) in 67.4 and 73.7% of all simulated trials, compared to 50.6%, 25.3% and 42.2% for BOIN, DFcomb, and YY, respectively. In Scenario 2, where most doses have similar utility, BUD designs select the optimal dose with frequencies at least 7.5%, 6% and 26% higher than the DFcomb, BOIN and YY designs. Similarly, in Scenario 3 BUD designs select the optimal dose more frequently than the three alternative designs. In Scenario 4, with two optimal dose combinations, BUD2 performs best and selects in 67% of the simulations one of the two optimal combinations compared to 56.3% and 47.3% for BUD1 and BUD3, 51.6% for BOIN, 24.6% for DFcomb and 18.8% for YY.

Column B of Figure 7.2 shows the average number of responders and toxicity events in a hypothetical cohort of 100 patients treated with the recommended dose. The recommended dose combination varies across simulations. In Scenario 1, 3, 4 BUD designs achieve a higher average number of treatment responses than the alternative designs, but with a slightly larger number of toxicity events. In Scenario 3, the expected number of responders for the BUD designs is 54.3, 56.0 and 48 compared to 30.2, 19.7 and 26.5 for the BOIN, DFcomb and the YY designs. The expected number of adverse events is between 30.2 and 33.3 for BUD designs, 27.7 BOIN, 22.4 for DFcomb and 31.1 for the YY design.

5.3 Dose finding with biomarkers

We consider the SBRT study of Chapter 2. It sequentially assigns patients to $J = 4$ dose levels. Figure 7.4 summarizes four scenarios that we considered. For instance, in the first scenario the probability of toxicity equals $(0.05, 0.07, 0.10, 0.15)$ for dose $d_1, \cdots, d_4$ when the EUD summary is close to zero and it increases to $(0.4, 0.42, 0.45, 0.46)$ when the EUD is equal to its maximum. The first column of figure 7.4 shows the marginal densities of the EUD summaries $Q_i(d_j), j = 1, \cdots, J$ across patients.

The mean functions μ^ℓ for the Gaussian prior $\beta_\ell \sim N(\mu^\ell, \Psi^\ell), \ell = E, T$, equals $\mu_j^\ell = \gamma_0^\ell + \gamma_1^\ell d_j$, with (truncated) normal prior for $\gamma_0^\ell \sim N(-1.5, 1)$ and $\gamma_1^\ell \sim N(1.5, 1)I(\gamma_1^\ell > 0)$. Similarly, for $\beta^Q = (\beta_1^Q, \cdots, \beta_J^Q) \sim N(\mu^Q, \Psi^Q)$ we use $\mu_j^Q = 1.5$, $j = 1, \cdots, J$, and set $(\lambda^T, \lambda^Q, \lambda_S) = (1, 1/2, d_J - d_1)$.

For each scenario we simulated 1,000 trials using a BUD design that utilizes EUD biomarkers for dose assignments. Similar to Sections 5.1 and 5.2, we consider the same three information measures, the posterior variance of the utility $u(d^\star, Q_i)$, the posterior entropy of the optimal dose $d^\star(Q_i)$, and the average mean squared difference between the utility of the optimal dose $d^\star(Q_i)$ and $\widehat{d}_{i-1}^\star(Q_i) = \arg\min_d E([u(d, Q_i) - u(d^\star(Q_i), Q_i)]^2 | \Sigma_{i-1}, Q_i)$. We indicate the corresponding BUD designs with BUD1-EUD, BUD2-EUD and BUD3-EUD. To quantify the advantage in using biomarkers we also simulated trials using BUD designs without incorporating biomarkers as discussed in Section 5.1.

Similar to previous examples in Sections 5.1 and 5.2, the last column of Figure 7.2 shows for a future group of 100 patients, each treated with the recommended dose, accordingly with the trial results, the expected number of efficacy and and toxicity events. The use of biomarkers' information tends to reduce the expected number of toxicity events, because patient with high EUD values and risk of toxicity are treated at lower dose levels. In scenario 1 BUD-EDU designs recommend personalized doses, that lead on average to 79.8, 80.5 or 81.0 patients without toxicities (BUD1-EDU,

BUD2-EDU and BUD3-EDU) and 17.6, 18.5 and 19 patients that respond to the treatment. In Scenario 4 where the probability of toxicity correlates substantially with the EUD summaries, the BUD-EDU designs reduce the expected number of toxicity events by approximately 50% to 30% compared to BUD designs without biomarkers; 59.4, 58.9, 58.2 for BUD1, BUD2 and BUD3 compared to 30.3, 39.9, 40 for BUD1-EDU, BUD2-EDU and BUD3-EDU, respectively.

Chapter 6

Discussion

We proposed and evaluated Bayesian uncertainty directed (BUD) dose finding designs. This class of designs utilize information metrics to achieve accuracy in the final selection of dose levels with optimal or nearly optimal clinical utility at completion of the trial. A BUD design assigns sequentially patients to candidate dose levels d accordingly to a score $S_i(d)$ that weights the expected information gain $G_i(d)$ and the estimated utility $E[u(d)|\Sigma_{i-1}]$. The approach allows the statistician to tailor the utility function u to the clinical context. Interestingly, BUD designs similar to CRM designs, with scores defined primarily by the expected utility $E[u(d)|\Sigma_i]$ (i.e $w_2 \approx 0$) lead on average to a substantial number of patients assigned to suboptimal doses. In contrast, score functions that account for the expected information increments $G_i(d)$ are more likely to identify dose levels with high utility, and tend to assign suboptimal doses in a lower fraction of the enrolled patients.

Various information measures can be used. We discussed three examples. The choice of the information measure can be guided by the operating characteristics of BUD designs, using simulations, under plausible scenarios. The BUD approach is applicable to a variety of dose finding problems, we discussed three applications; (i) a single agent dose finding trial, (ii) a dose finding trial to optimize a combination therapy of two agents, and (iii) a dose finding design that seeks to match individual biomarker profiles to effective and non-toxic radiation intensities.

We compared BUD designs to alternative designs using simulations. Across scenarios BUD designs have good performances in selecting dose levels with high clinical utility.

Chapter 7

Figures and Tables

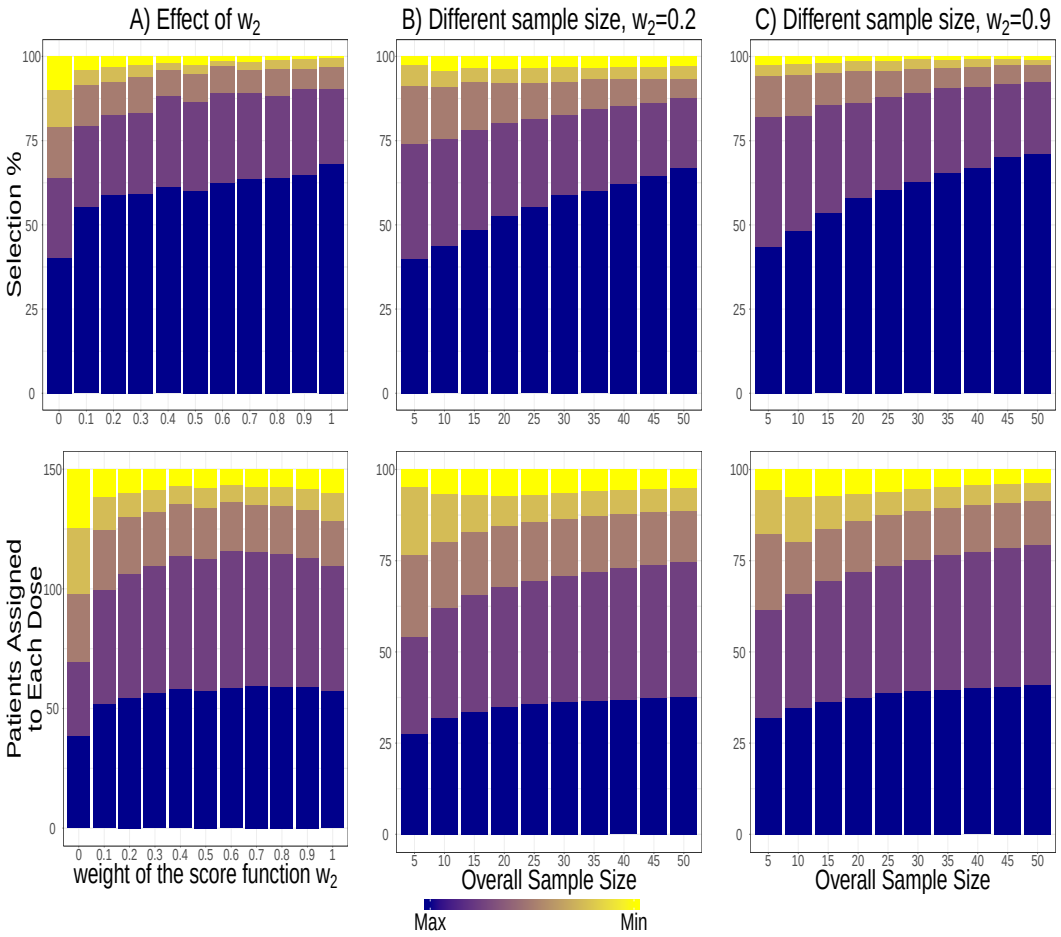

Figure 7.1. Proportion of times each dose was selected at completion of the study as the optimal dose across 1,000 simulations ($N = 30$) of a BUD trial (top row) and the average number of patients treated with each dose (bottom row). Column (A) shows the sensitivity of the BUD design to the information exploration weight $w_2 = 0, 0.1, \cdots, 0.9, 1$. Columns (B) and (C) show the performance of the BUD design for different overall sample sizes of the trials with either $w_2 = 0.2$ or $w_2 = 0.9$.

Dose	Design				
d, (p_E, p_T, u_d)	bCRM	LJ	$BUD1$	$BUD2$	$BUD3$
d_1, $(0.05, 0.02, 0.56)$	5.4 (3.1)	2.9 (2.9)	1.1 (1.7)	1.0 (2.0)	0.8 (2.3)
d_2, $(0.08, 0.05, 0.55)$	7.2 (3.1)	4.5 (2.7)	2.9 (1.9)	1.3 (2.7)	2.4 (3.3)
d_3, $(0.15, 0.07, 0.57)$	14.8 (5.2)	9.6 (4.3)	4.9 (3.8)	6.4 (4.1)	5.2 (4.5)
d_4, $(0.28, 0.10, 0.62)$	38.2 (9.5)	24.9 (8.6)	25.9 (11.6)	24.0 (11.3)	26.0 (10.8)
$\boldsymbol{d_5}$, $\mathbf{(0.43, 0.12, 0.68)}$	34.4 (8.8)	59.4 (12.0)	65.2 (10.7)	67.3 (9.7)	65.6 (8.9)
d_1, $(0.15, 0.10, 0.56)$	17.8 (06.5)	13.8 (5.9)	8.2 (3.2)	6.9 (3.1)	7.0 (3.2)
d_2, $(0.18, 0.12, 0.56)$	23.4 (07.1)	15.8 (6.7)	11.7 (6.6)	12.8 (5.9)	12.7 (6.0)
$\boldsymbol{d_3}$, $\mathbf{(0.38, 0.15, 0.63)}$	46.5 (11.6)	61.5 (11.3)	65.6 (7.6)	65.7 (7.2)	67.1 (7.0)
d_4, $(0.40, 0.36, 0.53)$	12.1 (03.4)	8.3 (4.7)	11.8 (8.4)	10.9 (8.7)	10.1 (8.7)
d_5, $(0.60, 0.65, 0.46)$	00.2 (01.2)	0.6 (1.2)	2.7 (4.1)	3.7 (4.8)	3.1 (4.9)
d_1, $(0.10, 0.05, 0.56)$	7.4 (3.8)	3.1 (3.5)	1.3 (1.8)	0.6 (1.8)	1.7 (2.5)
d_2, $(0.20, 0.07, 0.60)$	21.0 (6.1)	11.1 (4.1)	7.3 (2.4)	5.9 (2.3)	6.2 (3.6)
d_3, $(0.25, 0.10, 0.60)$	36.3 (9.8)	16.9 (6.8)	12.5 (6.5)	10.6 (5.6)	12.3 (6.3)
$\boldsymbol{d_4}$, $\mathbf{(0.50, 0.15, 0.69)}$	26.2 (6.4)	57.9 (10.4)	67.0 (10.3)	68.9 (10.7)	66.4 (9.6)
d_5, $0.54, 0.35, 0.60)$	9.1 (3.6)	11.0 (5.0)	11.9 (8.7)	14.0 (9.4)	13.4 (7.7)
d_1, $(0.18, 0.12, 0.56)$	25.9 (8.6)	17.0 (6.9)	4.4 (7.5)	4.8 (2.5)	6.5 (3.2)
d_2, $(0.23, 0.18, 0.55)$	22.9 (6.7)	18.9 (6.7)	9.4 (12.4)	7.6 (4.3)	7.7 (5.3)
$\boldsymbol{d_3}$, $\mathbf{(0.40, 0.20, 0.62)}$	31.5 (8.5)	39.5 (8.5)	41.4 (6.5)	40.9 (6.9)	40.7 (6.6)
d_4, $(0.44, 0.25, 0.61)$	16.8 (4.2)	23.1 (5.8)	38.9 (2.2)	40.8 (9.4)	37.4 (9.0)
d_5, $(0.48, 0.45, 0.51)$	2.9 (1.7)	1.5 (1.9)	5.9 (1.1)	5.9 (6.6)	7.7 (5.7)
d_1, $(0.02, 0.10, 0.49)$	6.0 (3.4)	4.4 (4.2)	1.2 (2.0)	1.7 (2.2)	0.6 (2.3)
d_3, $(0.10, 0.12, 0.49)$	20.5 (6.8)	10.3 (5.1)	2.7 (5.3)	2.7 (4.7)	2.6 (5.2)
$\boldsymbol{d_4}$, $\mathbf{(0.42, 0.15, 0.63)}$	56.8 (13.5)	66.5 (12.0)	68.3 (8.1)	69.3 (7.8)	64.7 (7.4)
d_4, $(0.45, 0.30, 0.61)$	15.6 (4.3)	18.4 (6.8)	26.3 (9.4)	25.0 (9.5)	30.6 (9.4)
d_5, $(0.50, 0.60, 0.52)$	1.1 (1.6)	0.4 (1.6)	1.5 (5.0)	1.3 (5.6)	1.5 (5.4)

Table 7.1. Proportion of times each dose was selected as the optimal dose across simulations, and the average number of patients treated with each dose (in parenthesis). We used 1,000 simulations of a trial with $N = 30$ patients using either the Bayesian uncertainty directed designs (BUD1, BUD2 and BUD3), the Bivariate Continual Reassessment Method (bCRM), the Liu and Johnson (LJ) Design. The optimal dose is shown in bold.

Dose d, (p_E, p_T, u_d) Treatment Outcomes are: available	Design $BUD1$	$BUD2$ immediately	$BUD3$	$BUD1$	$BUD2$ 12 weeks after enrollment	$BUD3$
d_1, $(0.05, 0.02, 0.56)$	1.1 (1.7)	1.0 (2.0)	0.8 (2.3)	1.3	0.7	1.0
d_2, $(0.08, 0.05, 0.55)$	2.9 (1.9)	1.3 (2.7)	2.4 (3.3)	2.1	2.4	2.6
d_3, $(0.15, 0.07, 0.57)$	4.9 (3.8)	6.4 (4.1)	5.2 (4.5)	7.6	10.9	8.0
d_4, $(0.28, 0.10, 0.62)$	25.9 (11.6)	24.0 (11.3)	26.0 (10.8)	24.4	24.5	26.3
$\boldsymbol{d_5}$, $(\mathbf{0.43, 0.12, 0.68})$	65.2 (10.7)	67.3 (9.7)	65.6 (8.9)	64.4	61.5	62.1
d_1, $(0.15, 0.10, 0.56)$	8.2 (3.2)	6.9 (3.1)	7.0 (3.2)	7.6	7.4	5.5
d_2, $(0.18, 0.12, 0.56)$	11.7 (6.6)	12.8 (5.9)	12.7 (6.0)	13.8	15.5	18.1
$\boldsymbol{d_3}$, $(\mathbf{0.38, 0.15, 0.63})$	65.6 (7.6)	65.7 (7.2)	67.1 (7.0)	61.5	56.5	57.5
d_4, $(0.40, 0.36, 0.53)$	11.8 (8.4)	10.9 (8.7)	10.1 (8.7)	12.3	15.7	13.5
d_5, $(0.60, 0.65, 0.46)$	2.7 (4.1)	3.7 (4.8)	3.1 (4.9)	4.8	4.9	5.4
d_1, $(0.10, 0.05, 0.56)$	1.3 (1.8)	0.6 (1.8)	1.7 (2.5)	1.2	0.6	0.6
d_2, $(0.20, 0.07, 0.60)$	7.3 (2.4)	5.9 (2.3)	6.2 (3.6)	6.0	5.2	8.5
d_3, $(0.25, 0.10, 0.60)$	12.5 (6.5)	10.6 (5.6)	12.3 (6.3)	12.8	10.5	13.0
$\boldsymbol{d_4}$, $(\mathbf{0.50, 0.15, 0.69})$	67.0 (10.3)	68.9 (10.7)	66.4 (9.6)	64.1	64.9	59.8
d_5, $0.54, 0.35, 0.60)$	11.9 (8.7)	14.0 (9.4)	13.4 (7.7)	15.9	18.8	18.1
d_1, $(0.18, 0.12, 0.56)$	4.4 (7.5)	4.8 (2.5)	6.5 (3.2)	5.5	5.2	4.4
d_2, $(0.23, 0.18, 0.55)$	9.4 (12.4)	7.6 (4.3)	7.7 (5.3)	9.6	6.9	10.4
$\boldsymbol{d_3}$, $(\mathbf{0.40, 0.20, 0.62})$	41.4 (6.5)	40.9 (6.9)	40.7 (6.6)	37.9	37.9	35.3
d_4, $(0.44, 0.25, 0.61)$	38.9 (2.2)	40.8 (9.4)	37.4 (9.0)	39.2	40.2	39.4
d_5, $(0.48, 0.45, 0.51)$	5.9 (1.1)	5.9 (6.6)	7.7 (5.7)	7.8	9.8	10.5
d_1, $(0.02, 0.10, 0.49)$	1.2 (2.0)	1.7 (2.2)	0.6 (2.3)	1.1	1.1	1.1
d_3, $(0.10, 0.12, 0.49)$	2.7 (5.3)	2.7 (4.7)	2.6 (5.2)	4.5	3.3	4.9
$\boldsymbol{d_4}$, $(\mathbf{0.42, 0.15, 0.63})$	68.3 (8.1)	69.3 (7.8)	64.7 (7.4)	64.5	59.1	61.8
d_4, $(0.45, 0.30, 0.61)$	26.3 (9.4)	25.0 (9.5)	30.6 (9.4)	27.7	34.5	29.6
d_5, $(0.50, 0.60, 0.52)$	1.5 (5.0)	1.3 (5.6)	1.5 (5.4)	2.2	2.0	2.6

Table 7.2. Proportion of times each dose was selected as the optimal dose across simulations, and the average number of patients treated with each dose (in parenthesis). Results are based on 1,000 simulations of a trial with overall sample size of $N = 30$ patients and an average enrollment of 24 patients per year. For each patient, the treatment response and the toxicity outcome become available 12 weeks after enrollment.

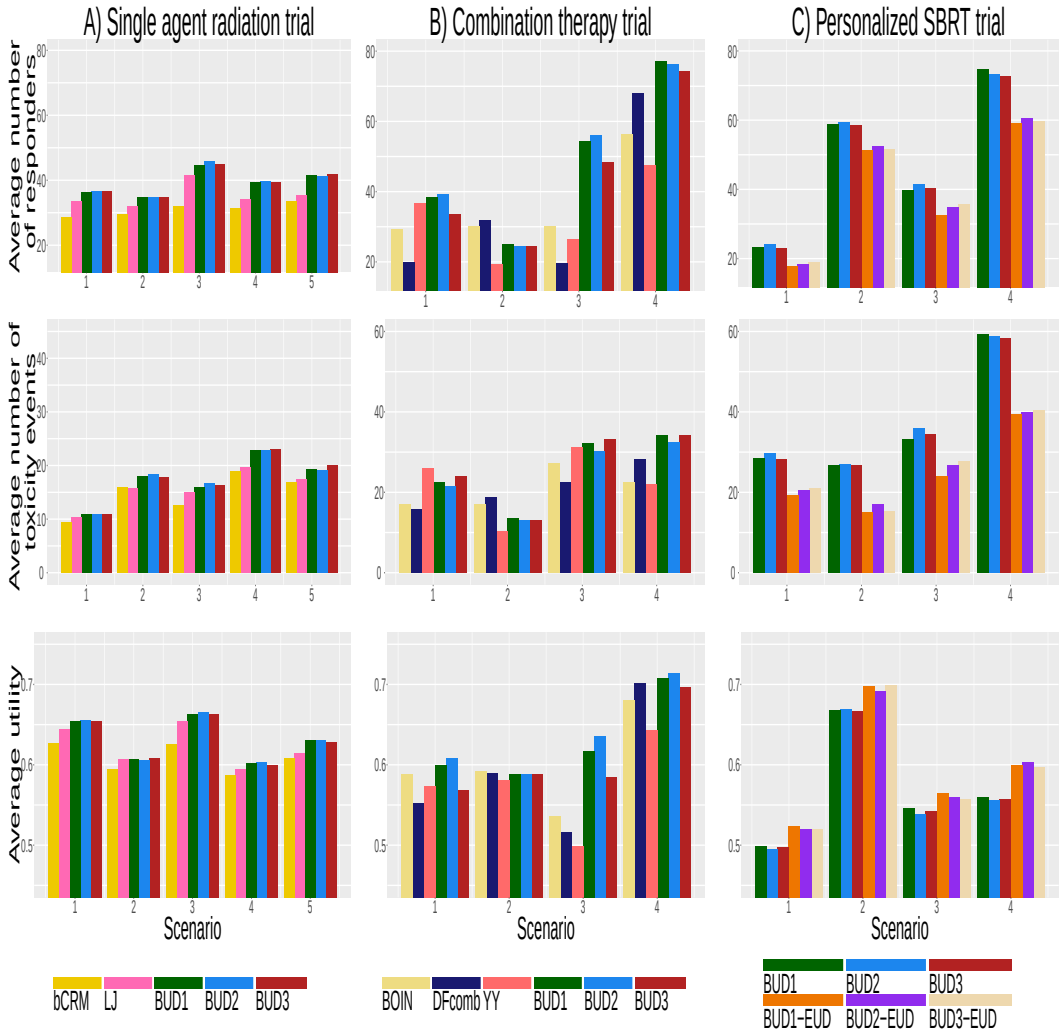

Figure 7.2. Expected number of patients in a hypothetical future cohort of 100 patients treated with the recommended dose $\widehat{d}^{\star}$ that respond to treatment (top row), the expected number of patients with a toxicity event (second row), and the average utility (bottom row). Results are based on 1,000 simulated single-agent trials (Column A), combination-therapy trials (Column B) and personalized dose finding trials (Column C), using either the Bayesian uncertainty directed designs (BUD1, BUD2 and BUD3), the Bayesian continual reassessment method (bCRM), the Liu and Johnson design (LJ), the BOIN design of [40], the DFcomb design of [51] or the YY design of [78].

Dose of agent 1	Dose of agent 2 d_1^2	d_2^2	d_3^2	d_4^2
			Scenario 1	
d_1^1	0.54 (0.10, 0.10)	0.53 (0.15, 0.15)	0.52 (0.38, 0.35)	0.45 (0.40, 0.50)
d_2	0.53 (0.15, 0.15)	**0.64** (0.42, 0.18)	0.53(0.45, 0.40)	0.48 (0.48, 0.52)
d_3^1	0.50 (0.44, 0.45)	0.46 (0.48, 0.55)	0.33 (0.50, 0.80)	0.30 (0.55, 0.90)
			Scenario 2	
d_1^1	0.58 (0.08, 0.01)	0.58 (0.10, 0.02)	0.56 (0.13, 0.08)	0.56 (0.15, 0.10)
d_2^1	0.56 (0.10, 0.05)	0.58 (0.18, 0.09)	0.59 (0.28, 0.15)	**0.60** (0.34, 0.18)
d_3^1	0.57 (0.15, 0.07)	0.59 (0.23, 0.10)	0.58 (0.30, 0.15)	0.57 (0.35, 0.18)
			Scenario 3	
d_1^1	0.50 (0.15, 0.20)	0.51 (0.20, 0.22)	0.52 (0.25, 0.25)	**0.65** (0.58, 0.28)
d_2^1	0.48 (0.22, 0.30)	0.42 (0.28, 0.45)	0.39 (0.30, 0.52)	0.49 (0.60, 0.60)
d_3^1	0.39 (0.26, 0.50)	0.35 (0.30, 0.60)	0.34 (0.33, 0.65)	0.42 (0.64,0.75)
			Scenario 4	
d_1^1	0.57 (0.15, 0.08)	0.57 (0.20, 0.12)	0.57 (0.23, 0.15)	**0.75** (0.68, 0.18)
d_2^1	0.62 (0.45, 0.22)	0.67 (0.60, 0.26)	**0.75** (0.80, 0.28)	0.65 (0.85, 0.45)
d_3^1	0.62 (0.65, 0.40)	0.63 (0.78, 0.48)	0.65 (0.85, 0.50)	0.59 (0.90, 0.65)

Table 7.3. Simulation scenarios for the combination therapy trial. The table shows the probabilities of toxicity $p_{T,d}$ and efficacy $p_{E,d}$, and the corresponding utility $u(d)$ for a combination treatment and candidate dose levels $d \in \mathcal{D}^1 \times \mathcal{D}^2$. Bold numbers correspond to the highest utilities.

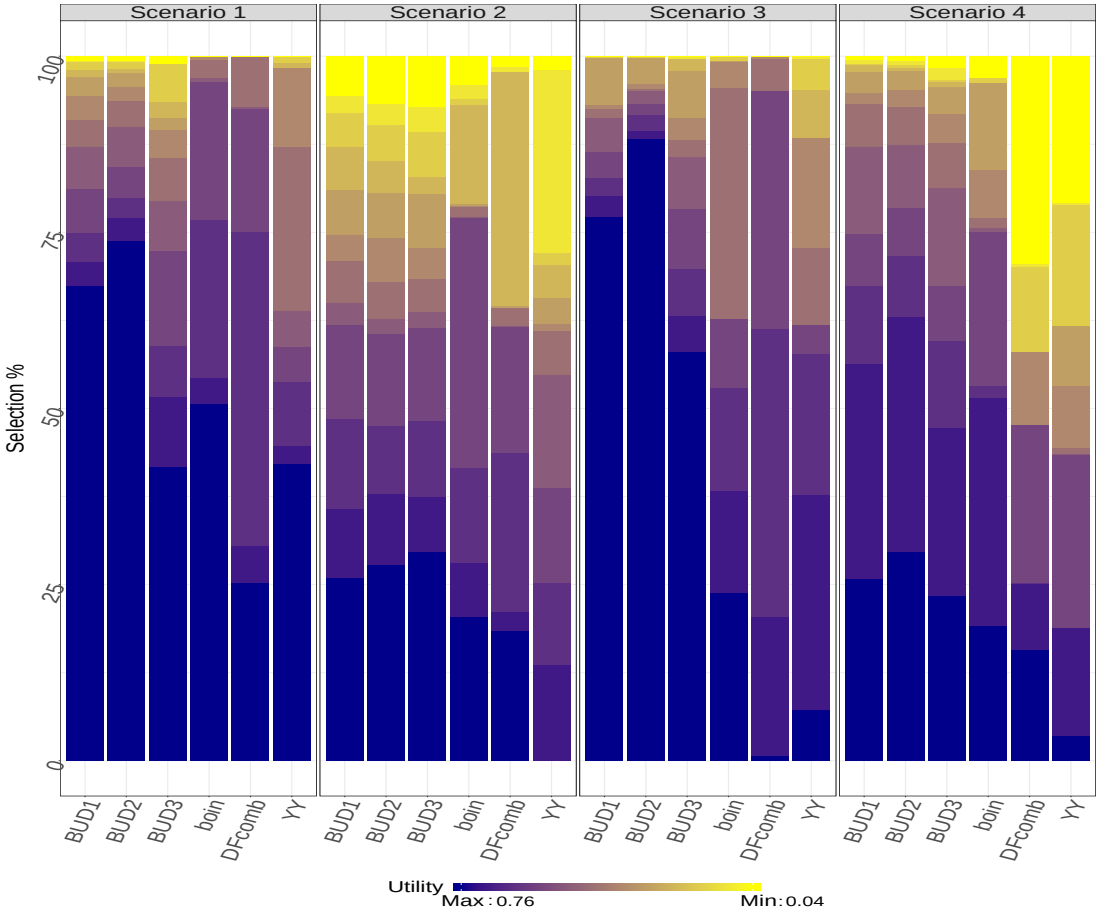

Figure 7.3. Proportion of simulations in which each combination dose $d = (d^1, d^2)$ was selected as the optimal dose at completion of the trial. Dose combinations d are ordered accordingly to their utility $u(d)$. Light colors correspond to combinations with low utility. BUD1, BUD2 and BUD3 correspond to three Bayesian information directed designs with different information measures. We compare BUD designs to three alternative designs indicated as BOIN design [40], DFcomb design [51] and YY design [78].

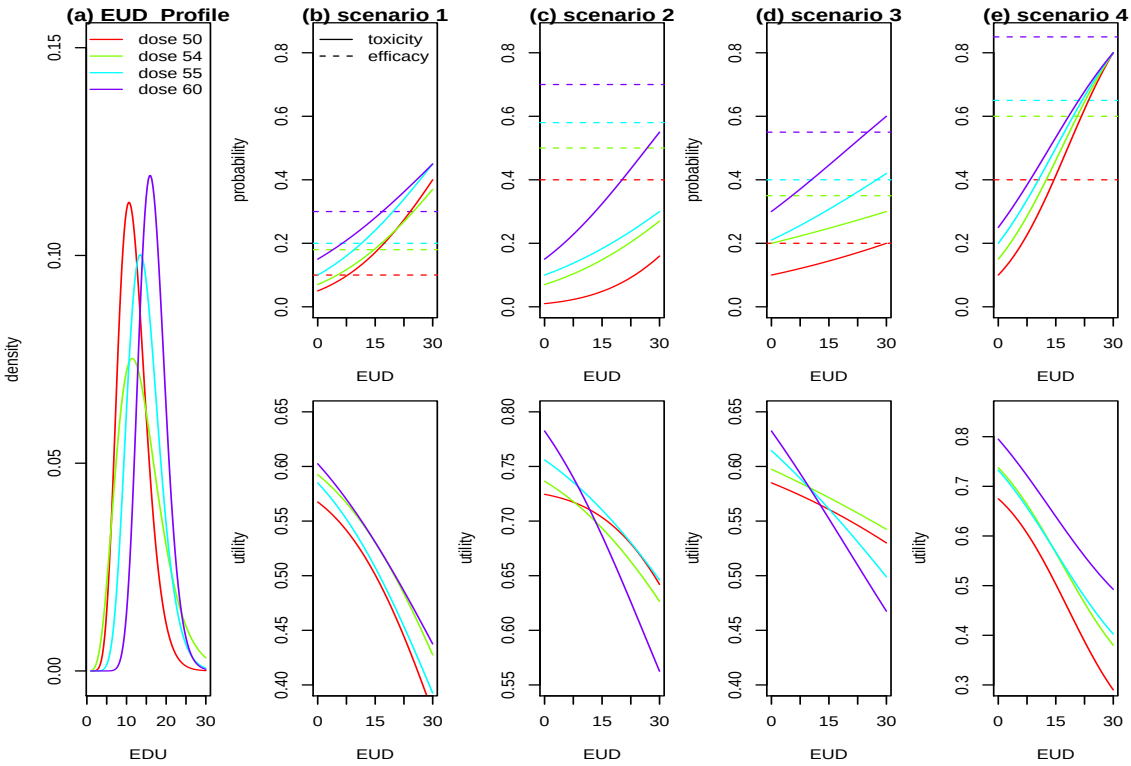

Figure 7.4. Simulation scenarios. Column (A) shows the marginal distribution of the EUD summaries. Columns (B) to (E) show the probability of efficacy (dashed lines) and toxicity (solid lines) across patients EUD values (first row) and the corresponding utility (second row) for each candidate dose levels.

Part II

Inference in Adaptive Trials under Time Trends in the Patient Population

Chapter 1

Introduction

Changes in patient population have always represented a substantial issue in clinical trials. Standard inference in clinical trials assumes that the recruited patients' characteristics don't change over the time, and the probability of the response to treatment remains constant during the study period. Since trials recruit patient for a long period, time trends in the patient population can constitute a potential source of bias in clinical trials. We investigate and assess the consequences of unknown changes in patients population for Response Adaptive Randomization (RAR) designs and propose procedures which correct treatment effect estimates for time trend and reduce inflation of the type I error rate. We incorporate time trends into the model outcome using Generalized Additive Models (GAMs). A parametric bootstrap is then used to account for potential time trend and estimate type I error rates. Through simulations studies under different time trends, we investigated the proposed methods in multi-arm clinical trials and platform trials.

Traditionally, clinical trials have been designed as two arm-study, where each experimental therapy is compared to the standard of care. However, when more than one promising treatment is available, conducting a multi-arm study is more efficient ([23], [73], [49], [15], [21]). Since the factors which affect the response to treatment are often unknown, randomization has been introduced as a method of assigning patients to treatments because it permits statistical inference without confounding

factors. Balanced randomization (BR) assigns patients to control and experimental arms at pre-specified ratio which is constant over the time. Doubts concerning its ethical implications have continued ([14]) and alternative approaches that balance potential benefits to study participants and future patients have been suggested in [82], [20],[53], [63], [37], [73] and others. Designs utilize response adaptive allocation procedures, which use the observed outcomes to assign patients to the most promising treatment with higher probability.

Recently, an new class of adaptive procedures known as platform trials has been introduced ([21], [30], [80], [70], [71]). A platform trial design allows the addition of new experimental arms to an ongoing trial. Standard multi-arm clinical trials require that all the therapies are at the same stage of development when the trial starts. This is often not possible. Since clinical trials enroll patients over many years, it is very likely that during this time a promising treatment emerges. The capability to add and to remove arms to ongoing trials represents a great advantage to save resources because no new trial has to be opened. Cancer studies as I-SPY 2, STAMPEDE, AML15 and AML16, the schizophrenia trial CATIE, and the international HIV trial 2NN are few examples of platform trials ([8], [29], [39], [11], [21], [68]). Nonetheless, since adaptive designs present poor frequentist operating characteristics compared to BR designs, the use of RAR approaches is criticized in many clinical settings ([33], [69], [59]).

The main drawback of employing adaptive treatment allocation arises from the potential presence of time trends in the prognostic mix of the patients accruing to the trial ([33]). Unlike the BR procedure which seeks to maintain balance at each point of the trial, and is less affected by fluctuations in potential outcomes that occur as sample enrollment proceeds ([26]), time trends are recognized as a probable source of bias in adaptive clinical trials. Many authors ([16], [76], [12],[75]) and in particular Altman and Royston [2] have underlined how adaptive randomization can lead to biased treatment effect estimate when the patient population changes dramatically during the trial. Despite the early stage in which this problem was

noticed, few methods have been proposed to deal with this problem ([31], [22], [41], [71]).

In this work, we investigate the effect of time trends on detecting the treatment effect when using data-adaptive allocation and propose a method to provide corrected estimation and inference under time trends. The patient population may change in different ways as a function of time, and the change is likely to be gradual instead of at a single time point. The outcome may be monotony improving if therapies for patients with poor outcome become available during the course of the trial. The change can also be cyclic, if seasonal changes such as the flu have an effect on patient outcome, such as in asthma trials. To capture potentially complex change in patient population, we suggest to use splines within the framework of a generalized additive model Generalized Additive Model (GAM) ([27]) to estimate treatment effects while accounting for time trends.

Because of the dependence among patients, standard Z statistics obtained from GAM are not valid and inference cannot be readily carried out. We propose a parametric bootstrap procedure for testing efficacy that extends the previous scheme proposed by Rosenberger and Hu [54]. The testing procedure can be used for any adaptive designs and estimators of the treatment effect. To show the general suitability and the efficiency of the method, we consider both frequentist and Bayesian RAR designs. We apply our procedure to the Randomized Play the Winner (PTW) algorithm by Li [38], to the Doubly Adaptive Biased Coin Design (DABCD) by Eisele [20] and to the Bayesian Adaptive Randomization (BAR) design by Thall and Wathen [63]. For all of the three RAR designs, we investigate the precision of the proposed estimator and study the related operating characteristics under different setting of trends.

The next chapter are organized as follows. Chapter 2 presents the response probability model and the testing procedure to handle the time trend. In Chapter 3 we describe briefly the outcome-adaptive multi-arm trial designs that we consider. Chapter 4 provides a simulation studies to evaluate the proposed testing algorithms.

We conclude with a discussion in Chapter 5.

Chapter 2

Method

We consider a Phase II clinical trial that assigned up to N patients to either A experimental arms or to the control. For each patient i, $C_i = a$ indicates that the $i-$th patient has been randomized to arm $a = 0, ..., A$, where $a = 0$ is the control arm. Let's denote also by $N'_a(i)$ the number of patient randomized to a before the i-th arrival and by $N_a(i)$ the number of observed outcome for arm a at the i-th enrollment. Let Y_i be a binary outcome that assumes 1 if the patient i responds positively to the therapy and 0 otherwise. Finally, we denote by $\Sigma_i = \{(N'_a(j), N_a(j), Y_j, C_j); j = 1, ..., i\}$ the data available at the $i-$th enrollment.

Assuming that the probability of the response to treatment is constant over the time, $Y_i|C_i = a$ has a Bernoulli distribution with response probability θ_a, for each $a = 0, ..., A$. The treatment effect of each experimental arm a can be tested as

$$H_a^0 : \theta_a - \theta_0 \leq 0 \quad \text{versus} \quad H_a^1 : \theta_a - \theta_0 > 0$$

In presence of time trends, assuming $i.i.d$ within an arm may lead to a biased estimate for the treatment effect and an inflated or deflated type I error rate in testing. We introduce a method to account for possible time trends when estimating the treatment effect and testing the global null hypothesis at the end of the trial.

2.1 Estimation of the treatment effect

In this Section, we specify a probability model where the response probabilities depend on both the assigned treatment and the enrollment time. Since true potential time trend is unknown, we adopt a flexible spline-based approach that is able to capture any type of smooth trend. The outcome probability of patient i enrolled at time t_i is modeled through a GAM ([27]) with a probit link function

$$P(Y_i = 1|\beta, C_i = a, t_i) = \begin{cases} \Phi(\beta_0 + f_\beta(t_i)) & a = 0 \\ \Phi(\beta_0 + \beta_a + f_\beta(t_i)) & a > 0 \end{cases} \tag{2.1}$$

where $f_\beta(\cdot)$ is a cubic smoothing spline with J knots ([19]). The model 2.1 assumes that the effect of time trend acts equally on patient in the control and experimental arms, and does not interact with the treatment effect. Thus, the treatment effect of the arm a is describe by the regression coefficient β_a, $\forall a > 0$.

The global null hypothesis is that there is no difference between the response rate of patient treated with therapy $C_i = a > 0$ and the control, e.g $H_a^0 : E[Y_i|\beta, C_i = a] - E[Y_i = 1|\beta,\ C_i = 0] \leq 0, \forall i$, or equivalently by considering

$$H_a^0 : \beta_a \leq 0 \quad \text{versus} \quad H_a^1 : \beta_a > 0.$$

Denoting by $\widehat{\beta}_a$ the estimator of β_a, we use the statistic $T_a = \widehat{\beta}_a / sd(\widehat{\beta}_a)$ to test the null hypothesis H_a^0 at significance level α, for each arm $a > 0$. Large positive values of the test statistics show evidence against the null hypothesis.

2.2 Testing procedure

We suggest to test the hypothesis discussed in Subsection 2.1 by adopting a bootstrap procedure similar to the algorithms in Rosenberger and Hu [54], Trippa et al. [67] and Ventz et al. [71]. We first discuss the procedure assuming no early stopping rule and extend the procedure for early stopping. The testing procedure starts from the

realization of a multi-arm trial $\mathcal{T}$ generated under a design d with potential time trend.

2.2.1 A bootstrap test for trials without early stopping rules

Given a trial $\mathcal{T}$ generated under a design d, we estimate the response probability model in 2.1 and compute a test statistics T_a for each arm a. For each arm a that we want to test, we generate B trials under design d with probability of response to the treatment for i-th patients equals to

$$P(Y_i = 1|C_i = a', t_i) = \begin{cases} \Phi(\widehat{\beta_0} + \widehat{f_\beta}(t_i)) & a' = a \\ \Phi(\widehat{\beta_0} + \widehat{\beta_{a'}} + \widehat{f_\beta}(t_i)) & a' \neq a \end{cases} \tag{2.2}$$

where $\widehat{\beta_0}, \widehat{\beta_{a'}}$ and $\widehat{f_\beta}(\cdot)$ denotes the estimated coefficients from the GAM model.

For each b, with $b = 1, ..., B$, we compute the test statistics $T_a^{*,b}$, and estimate the p-value as

$$\widehat{p(T_a)} = \frac{\sum_{b=1}^{B} I(T_a^{*,b} \geq T_a)}{B}$$

The null hypothesis is rejected at level α if $\widehat{p(T_a)} < \alpha$.

Algorithm 2 bootstrap hypothesis testing without stopping rules

1: **Input 1:** Trial $\mathcal{T}$ and trial design d
2: **Input 2:** Arm a and definition of test statistics T_a
3: Estimate the GAM response probability model in 2.1.
4: Compute the test statistic T_a
5: **for** b $= 1, ..., B$ **do**
6: Simulate a study under design d with response probabilities of patients
7: in 2.2
8: Compute $T_a^{*,b}$
9: **end for**
10: Estimate $\widehat{p(T_a)} = \frac{\sum_{b=1}^{B} I(T_a^{*,b} \geq T_a)}{B}$.
11: **Output**: Reject H_a^0 if $\widehat{p(T_a)} \leq \alpha$

2.2.2 A bootstrap test for trials with stopping rule for futility

We extend the procedure describe in the Subsection 2.2.1 to allow early stopping for futility. Exactly as in Ventz et al. [71], an experimental arm is dropped for futility after the enrollment of the $i-$th patient if the posterior probability of treatment effect is lower a threshold $f_{i,a}$. Here, we define $f_{i,a} = f \times (N_a(i)/n'_E)^g$, where $f \in [0,1]$, $g > 0$ and n'_E denotes a desired maximum number of patients in each experimental arm. The boundary, so defined, increases adaptively from 0 to 1 as the number of observed outcomes tends to the desired sample size for the arm.

We estimate the response probability model for a trial $\mathcal{T}$ generated under d. We compute the test statistics T_a for $a > 0$ and define $S_a = 1$ if arm a was not stopped early. For each arm a that we want to test and that was not stopped early during the trial $\mathcal{T}$, we simulate B trials under d with early stopping rule for futility by using the same response probabilities as in Subsection 2.2.1. Then, for each simulated trial b, we compute the test statistics $T_a^{*,b}$ and we set $S_a^{*,b}$ equals to zero if the arm a was dropped early for futility and equal to 1 otherwise.

Unlike before, we need to correct the procedure by accounting for the fact that the arm a may have been stopped early in some simulated trials. We estimate the probability that the arm a is not stopped early for futility as the empirical proportion of trials that have not been dropped over the total B trials. The corrected p-value is estimated as $\widehat{p(T_a)} = \widehat{P(T_a^*)} \times \widehat{P(S_a^*)}$ where

$$\widehat{P(T_a^*)} = \frac{\sum_{b:S_a^{*,b}=1} I(T_a^{*,b} \geq T_a)}{\sum_{b=1}^{B} I(S_a^{*,b} = 1)} \quad \text{and} \quad \widehat{P(S_a^*)} = \frac{\sum_{b=1}^{B} I(S_a^{*,b} = 1)}{B}.$$

If $\widehat{p(T_a)} \leq \alpha$, the null hypothesis is rejected at level α.

Algorithm 3 bootstrap hypothesis testing with stopping rule for futility

Input: Trial $\mathcal{T}$ and trial design d

Input 2: Arm a and a definition of test statistics T_a

if Arm a is not stopped early $\mathcal{T}$ **then**

  Estimate the GAM response probability model in 2.1.

  Compute the test statistic T_a

  for b $= 1, ..., B$ **do**

    Simulate a study under d with early stopping rule and
    with response probabilities of patients in 2.2

    if arm a is dropped **then** $S_a^{*,b} = 0$
    else $S_a^{*,b} = 1$
    end if
    if $S_a^{*,b} = 1$ **then** Compute $T_a^{*,b}$
    end if

  end for

  Estimate $\widehat{P(T_a^*)} = \frac{\sum_{b:S_a^{*,b}=1} I(T_a^{*,b} \geq T_a)}{\sum_{b=1}^{B} I(S_a^{*,b}=1)}$ and $\widehat{P(S_a^*)} = \frac{\sum_{b=1}^{B} I(S_a^{*,b}=1)}{B}$.

  Estimate the p-value as $\widehat{p(T_a)} = \widehat{P(T_a^*)} \times \widehat{P(S_a^*)}$

  Output: Reject H_a^0 if $\widehat{p(T_a)} \leq \alpha$

end if

Chapter 3

Response Adaptive Randomization designs

We apply the bootstrap procedure to several adaptive designs. In this Chapter we summarize the considered designs.

3.1 The Randomized Play the Winner

Several randomized extensions of Zelen's original determinist "play-the winner" ([82]) have been proposed in the past years ([74], [3], [53]). Wei [74] considers a randomized version of the "play-the winner" ([75]) for multi-arm clinical trials. Consider an urn with A different particles, one for each treatments. In the proposed multi-arm design, a success of the treatment a generates a particle of type a, and a failure on treatment a generates $1/(A-1)$ particles of other $A-1$ types. However, if one treatment performs extremely poorly, then it is unreasonable to add the same number of particles of that type to the urn ([6]). Li [38] proposed a scheme that only generates particles of the type of the success, without adding anything to the failures. Each patient i is randomized to arm a with probability proportional to the number of particles a in the urn after $i-1$ patients

$$P(C_i = a|\Sigma_i) \propto \frac{1 + \sum_{j=1}^{i-1} Y_j I(C_i = a)}{1 + N'_a(i)}$$

for $a = 1, ..., A$.

3.2 The Doubly Adaptive Biased Coin Design

The DABC, proposed by Eisele [20], is a RAR design where the assignment of the patients is driven by both the proportion of subjects allocated to each arm and a vector of target proportions $\{\rho_a\}$ that depends on response rate. For instance, the Neymann allocation has the target allocation is defined as $\rho_a \propto \sqrt{\theta_a(1-\theta_a)}$. Since θ_a is unknown, the target vector is replace by the quantity $\widehat{\rho_a}$, estimated from the accumulated data. The probability that patient i is assigned to treatment a is given by

$$P(C_i = a|\Sigma_i) \propto \widehat{\rho_a}(i) \times q_a(i)$$

where $q_a(i) = (\widehat{\rho_a} \times (i+1)/(N'_a(i)+1))^{\beta}$, with $\beta > 0$. So defined, $q_a(i)$ ensures that if the current proportion of subjects allocated to the arm a is smaller than the target, the randomization probability to arm a for the next patient is larger than $\widehat{\rho_a}(i)$.

3.3 The Bayesian Adaptive Randomization design

The aim of a BAR design is to treat patients during the trial with the more efficient treatment. After a burn-in period where patients are randomized with equal probability to the arms, BAR uses the data accumulated during the trial to increase the probability of assigning the patients to treatments that show being most promising. The outcome model can be completed by setting a prior $\theta_a \sim Beta(v_1, v_2)$ for each response probability of arm a. The BAR procedure assigns patient i to arm a with probability

$$p(C_i = a|\Sigma_i) \propto \begin{cases} p(\theta_a > \theta_0|\Sigma_i)^{h_\ell(i)} & \text{if } a \in \mathcal{A} \\ c(i)exp\left\{-b \times [N'_0(i) - max_{a\in\mathcal{A}} N'_a(i)]\right\} & \text{if } a = 0 \end{cases}$$

where $b > 0$, $c = \sum_{a\in\mathcal{A}} p(\theta_a > \theta_0|\Sigma_i)^{h_1(i)}/\mathcal{A}$ and the function $h(\cdot)$ is an increasing function of the sample size (Thall and Wathen [63]). At the beginning $h(\cdot)$ is equals to zero and the randomization is balanced among the arms. As the sample size increases, $h(\cdot)$ increases and the patients are allocated to the arm with higher probability of positive response.

3.4 Extensions to platform trials

We apply the bootstrap procedure also to platform trials. We consider the RAR designs presented in Section 3.1, 3.2 and 3.3 to allow the addition of new arms by adopting the approach of Ventz et al. [71].

Let $A_2, ..., A_K$ be the groups of experimental arms added to the ongoing trial and $M_2, ..., M_K$ the arrivals of the M_k-th patient, for $k = 2, ..., K$. When a group of new treatments is included to the trial, the sample size of the study is increased by n_k additional patients. The designs are modified by multiplying the randomization probabilities during the trial for a scaling function $q_k(i)$. For PTW and BAR, the assignment rules to an experimental arm are multiply by a Gompertz function define as

$$q_k(i) = r_0 + r_1 exp\left\{-exp(N'^{(k)}(i) - m_k\right\}$$

where $N'^{(k)}(i)$ is the amount of patients randomized to the k-th group of experimental arms and m_k, r_1, $r_0 > 0$ are tuning parameters. When a sufficient number of patients are allocated to treatments in group k, e.g. $N'^{(k)} > m_k$, we have that $q_k(i) \approx r_0$ and patients are randomized to treatments according to standard PTW or BAR. For DABC, when new treatments are added, the target is re-defined and to avoid extremely unbalanced, the allocation probabilities become proportional to

$max(\widehat{\rho_a}(i) \times q_a(i), w(i))$, where $w(i) \propto 1/(1 + \sum_{k;a \in A_k} I(M_k \leq i, N'_a(i) < n'_E))$ and n'_E represents a desired maximum number of patients in each experimental arm.

See [71] for a more careful discussion.

Chapter 4

Simulation studies

In this Chapter, we discuss the operating characteristics of the presented procedures in both standard multi-arm clinical trials and platform trials. We initially investigate the precision of the estimator of the treatment effect. We compare the estimate of the treatment effect obtained through the GAM model with the estimate achieved with no trend adjustment (NTA).

For each design discussed in Chapter 3 and for a BR design where patients are allocated to the arms with equal probabilities, we simulate 1000 trials and compute the treatment effect as $100 * \Delta_a$. For GAM, we consider $\Delta_a = \beta_a$, for each $a > 0$. For NTA, Δ_a is defined as the difference between the estimate response rate of arm a and the control on a probit scale. This definition guarantees a fair comparison between the two treatment effect estimates.

Afterwards, through a simulation study, we assess the impact of time trends on the type I error rates and on the power of the test. We compare the proposed hypothesis testing which accounts for time trends with the test where time trends are not considered. In the NTA approach, the operating characteristics are still computed following the bootstrap procedures previously discussed, but the test statistic considered is the standardized difference between the estimated response probability of an arm a and the control. In this case, response rates are estimated assuming *i.i.d.* variables.

We fix a nominal level at 0.05 and run 1000 trials for all the RAR designs explained in Chapter 3 and for a BR design where patients are equally likely to be assigned to any arm of the trial. The number of bootstrap is set equal to 1000 in all the simulations. For the multi-arm clinical trial, we compute the type I error rates and the power of the test by using the Algorithm 2. For the platform trial we investigate also the ability of the method to preserve the type I error level in presence of stopping rule for futility, both Algorithm 2 and Algorithm 3 are then implemented.

Figure 6.1 shows the four different trends of success rate of a treatment over time that we examine in all the investigations. Trend 0 corresponds to absence of time trend in the date; Trend 1 describes a situation in which the success rate has a monotone decreasing behavior that corresponds to a population's change towards more fragile patients. In Trend 2 the success rate is represented as a periodical function that describes a seasonal effect on the patient population (for example patients may be weaker during the winter). Finally, Trend 3 describes a success rate that increases monotonically and that describes a population's change towards more healthier patients.

4.1 Multi-arm Clinical Trial

We consider a Phase II multi-arm clinical trial with three experimental arms, all available at the beginning of the study, and a control arm. The true regression coefficients are set equal to (0.13, -0.5, 0, 0.55), which correspond to the vector of true outcome probabilities of (0.55, 0.35, 0.55, 0.75) in absence of trend. The trial enrolls up to 240 patients, with an accrual rate of six patients per month and the outcome of each patient becomes available after 8 weeks from the treatment. For the BR design, we specify the allocation ratio such that patients are equally randomized among the arms, e.g. $q_0/q_a = 1/(A+1)$ for each $a = 1, .., A$. For the PTW, we fix the particles to add to the urn when a success is observed equal to 1

for all the arms. For DABC design, we follow Hu and Zhang [32] and use a target allocation as $\rho_a(\theta) \propto \sqrt{\theta_a}$, for $a = 1, .., A$, and $\rho_0(\theta) \propto max_{a>0}\sqrt{\theta_a}$ for the control. We also fix $\beta = 3$. Finally, for BAR, we define $h(i) = \beta \times [N'_a(i)/N]^\gamma$ and set $(\beta, \gamma, b) = (3, 1.5, 0.5)$.

Initially, we examine the impact of time trends on estimates of treatment effect for the different randomized procedures. Panel A of Figure 6.2 shows the estimated treatment effect of Arm 3 across 1000 simulations for BAR.The true treatment effect is equal to 55% and is represented by by the black line. When the data are not affected by trend or the trends are monotonic, both NTA and GAM estimates result almost correct with a bias of about 1%. Under seasonal trend, instead, the NTA estimator is more biased than the proposed estimator, and the treatment effect estimates are 41.52% for NTA and 54.01% for GAM. We extend the study of the estimated treatment effect to the other designs. The results are reported in Table 6.1. The estimators perform almost equally in absence of trend also for the other designs. In presence of monotonic trends, the estimates of treatment effect obtained through the GAM model are slightly more accurate for BAR, PTW and DABC, while for the BR the NTA estimates are almost unbiased. A substantial improvement is given by adopting the proposed method when the data follows a periodical behavior. For all the designs, NTA estimator is less correct, while the estimate of the treatment effect obtained through the GAM model is accurate. The estimates reported by NTA are 13.48%, 12%, 15.62% and 10.46% lower than the true value for BAR, DABC, PTW and BR, respectively. The treatment effects estimated by the proposed method, instead, have a bias of 1% for BAR, 1.5% fro DABC, 3.2% for PTW and 1% for BR.

Next, we explore the frequentist operating characteristics of the testing procedure illustrate in Subsection 2.2. Results are summarized in Table 6.2. As expected, if the outcome probabilities don't change during the study, the type I error rates is better controlled by the NTA than the proposed method. When the population changes towards more fragile patients, the type I error rate without adjustment decreases to around 4% while the proposed method is more able to target the predefined nominal

value. When the patient population changes systematically, the type I error rate without any action for time trend increases to 6% for BAR and decreases to 4.5%, 2.2% and 3.9% for BR, PTW and DABC, respectively. Under same trend, the values obtained from our bootstrap procedure are 4.8% for BR, 5.6% for PTW, 6.0% for DABC and 5.0% for BAR. Finally, the power of the test obtained without trend adjustment tend to be similar to rates computed through the suggested model. An exception is Trend 2, where under BR and BAR the NTA present a higher power (73.8% and 85.6% against 65.5% and 78.9%).

4.2 Multi-arm Platform Trial

Platform trials are considered a direct extension of RAR procedures. For their capability of adding new experimental arms during the course of a study, the attention on this type of design is becoming higher over the time. Unfortunately, also changes in the recruited patients' characteristics are more likely to occur during a platform trial. In this Section, we illustrate the operating characteristics of the previous RAR schemes when they are extended to platform trials through the approaches proposed by Ventz et al. [71] and summarize in Section 3.4.

We consider a Phase II multi-arm clinical trial that wants to assign up to 150 patients to either one of two experimental arms or to a control arm. As before, accrual rate is 6 patients per months and the response to the therapy is observable after 8 weeks from the treatment. Two new experimental treatments become available approximately 16 months after the beginning of the trial ($M_2 = 102$ and $A_2 = 2$) and the overall sample size is increased by 100 additional patients. Following [71], we modify the power function $h(\cdot)$ as $h(\cdot) = \beta[N'^{(k)}/n_k]^\gamma$, for $k = 1, 2$ and fix $(\beta, \gamma, b) = (3, 1.5, 0.5)$. The tuning parameters for the Gompertz function are set as $(r_0, r_1) = (1, 3)$ and $(m_1, m_2) = (30, 30)$. The rest of the parameters remain unchanged with respect to Subsection 4.1 As before, we investigate as first the precision of the estimator of treatment effect. For that purpose, we consider a

scenario where the two arms added after the beginning of the trial are the less efficacious and the more promising. Thus, we set the true regression coefficients as $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4) = (0.13, 0, 0.25, 0, 0.55)$; if the time trend is null, the true outcome probabilities are $(\theta_0, \theta_1, \theta_2, \theta_3, \theta_4) = (0.55, 0.55, 0.65, 0.55, 0.75)$.

Figure 6.2, Panel B, represents the estimated treatment effect for Arm 4, when it is added after the assignment of 102 patients, across 1000 simulated BAR trials. The true treatment effect is showed by the black line and it is equal to 55%. According to Section 4.1, when the outcome probabilities are not subject to changes in the patient population, both estimates result correct. In Trend 1 and Trend 3, GAM estimates the treatment effect correctly as 55.18% and 56.5%, while the estimates obtained with NTA are not accurate, 40.24% and 67.74%, respectively. In Trend 2, the NTA approach shows a strong bias and the estimate of the treatment effect is 35% lower than the true value, against a bias of 1.26% for the GAM model. Table 6.3 reports the simulated results also for the other designs. The GAM model is unable to accurately estimate the treatment effects for the added arms under a PTW scheme. This results in extremely high bias for the GAM model. On the other hand, the bias reported by the NTA approach is much lower (around 10%). For the other designs, the two methods perform equally well in absence of trend. In presence of any time trend, the values obtained through the GAM model are generally more correct. The estimates of the treatment effect for all the arms have a bias between 0-3%. NTA, instead, reports a very a higher bias for the added arms in Trend 1, 2 and 3. In Trend 2, the treatment effect estimates of Arm 4 are between 30% and 44% lower than the true value. In Trend 1 and 3, the bias of Δ_4 estimates are between 12% and 17% and around 10%, respectively.

We consider a different scenario to better evaluate the ability to protect the type I error inflation of the proposed method. We set the following scenario $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4) = (0.13, 0.25, 0.55, 0, 0)$, and Arm 3 and Arm 4 become available after that 102 patients have been already enrolled into the study. The corresponding true outcome probabilities in absence of trend are $(\theta_0, \theta_1, \theta_2, \theta_3, \theta_4) =$

$(0.55, 0.65, 0.75, 0.55, 0.55)$. The corresponding type I error rates are computed across 1000 simulated trials and the values are reported in Table 6.4. In a platform trial the presence of any kind of trend has a very strong impact on the type I error rate. If no adjustment for trends in the patient population is considered, the type I error inflates drastically. When the probabilities of a response to treatments decrease (Trend 1), and not adjustment is made, the type I error rates increase to around 8% for BR and PTW, and to more than 20% for BAR, while is reduced to 1.8% for DABC. Under the same trend, the proposed method is able to preserve the type I error close to the nominal value for all the designs. When the population changed towards healthier patients (Trend 4), the effect of the presence of trend is stronger and rates for NTA increase more than 7%, 25%, 40%, 25% for BR, PTW, DABC and BAR, respectively, against the 4%, 4.9% and 5.3 % for BR, DABC and BAR obtained from our bootstrap procedure. Under seasonal trends, we observed an inflation of the type I error rates for PTW (23.0%) and BAR (12%) and a deflation for BR (1.4%) and DABC (12%) when no modification for time trend is provided. Rates sufficient close to the nominal level are obtained for BR, DBC and BAR for the GAM adjustment, but a reduction to 2.1 is observed for PTW.

Finally, we investigate the performance of the proposed method when stopping rule for futility is introduced to the procedure (see Subsection 2.2.2). We set the parameters of the threshold $f_{i,a}$ as $(f, g, n'_E) = (.3, 1.5, 50)$ for all the designs and we simulate 1000 trials for each design. As before, the method obtains rates near to desired nominal value and protect the type I error by an extreme inflation. Besides the PTW that performs very poorly in all the trends, the other designs target the predefined type I error. Results are showed in the last column of Table 6.4. In Trend 1, DABC and BAR presents a rate almost equal to 5% and then preserve better the type I error respect to BR design (about 6%). Under seasonal trend, BAR controls better the nominal level, while BR and DABC reports a rates around 4%. Lastly, in Trend 4, the type I error rate has an inflation only of 1% for BAR and BR, and is well-preserved for DABC.

Chapter 5

Discussion

We propose and evaluate procedures to handle possible changes in patients population's characteristics during a clinical trial. The procedures start by modeling the response probability to a treatment as function of the enrollment time of each patient. We adopt a semi-parametric model, which assumes a constant treatment effect over the course of the study and an equal time trend effect for each patient that is independent from allocation. Next, we suggest a general parametric bootstrap algorithm for testing efficacy and preserving the type I error inflation. The algorithm is feasible to any clinical trial design and any estimator. Through the manuscript, we discuss and compare four randomization schemes.

We applied the procedures to a standard multi-arm clinical trial and later we extend the comparison to multi-arm platform trials. The procedures were robust to changes in patient population and significant improvements were observed about the control of the type I error rate when experimental arms are added to an ongoing study.

This work represents one of the few contribute about this issue in literature, and we are aware that further researches are needed to asses the potential impact of time trend in multi-arm clinical trials. Different scenarios may be considered, and a more rigorous study of the power need to be conducted. Moreover, accounting directly for time trends by using alternative adaptive-allocation rules in an ongoing study or

exploring the effect of time trends on RAR procedures with biomarkers represent the next challenges to be investigated.

Chapter 6

Figures and Tables

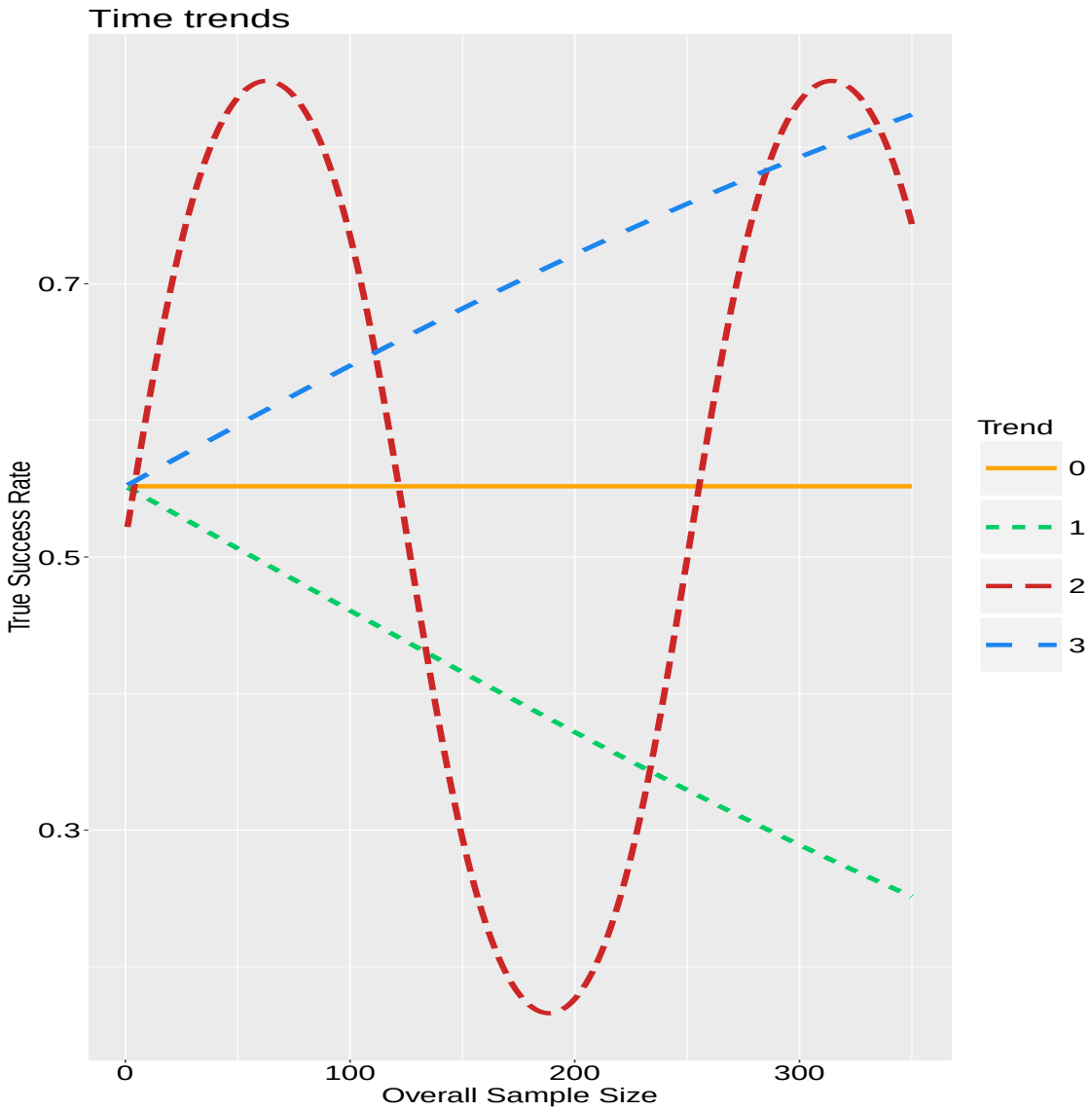

Figure 6.1. True success rate under different time trend assumptions during the trial. The true response probability at the beginning of the study is equal to 0.55. Trend 0: Success rate in absence of trend. Trend 1: Success rate decreases monotonically over the time. Trend 2: Success rate has a periodic behavior. Trend 3: Success rate increases monotonically over the time

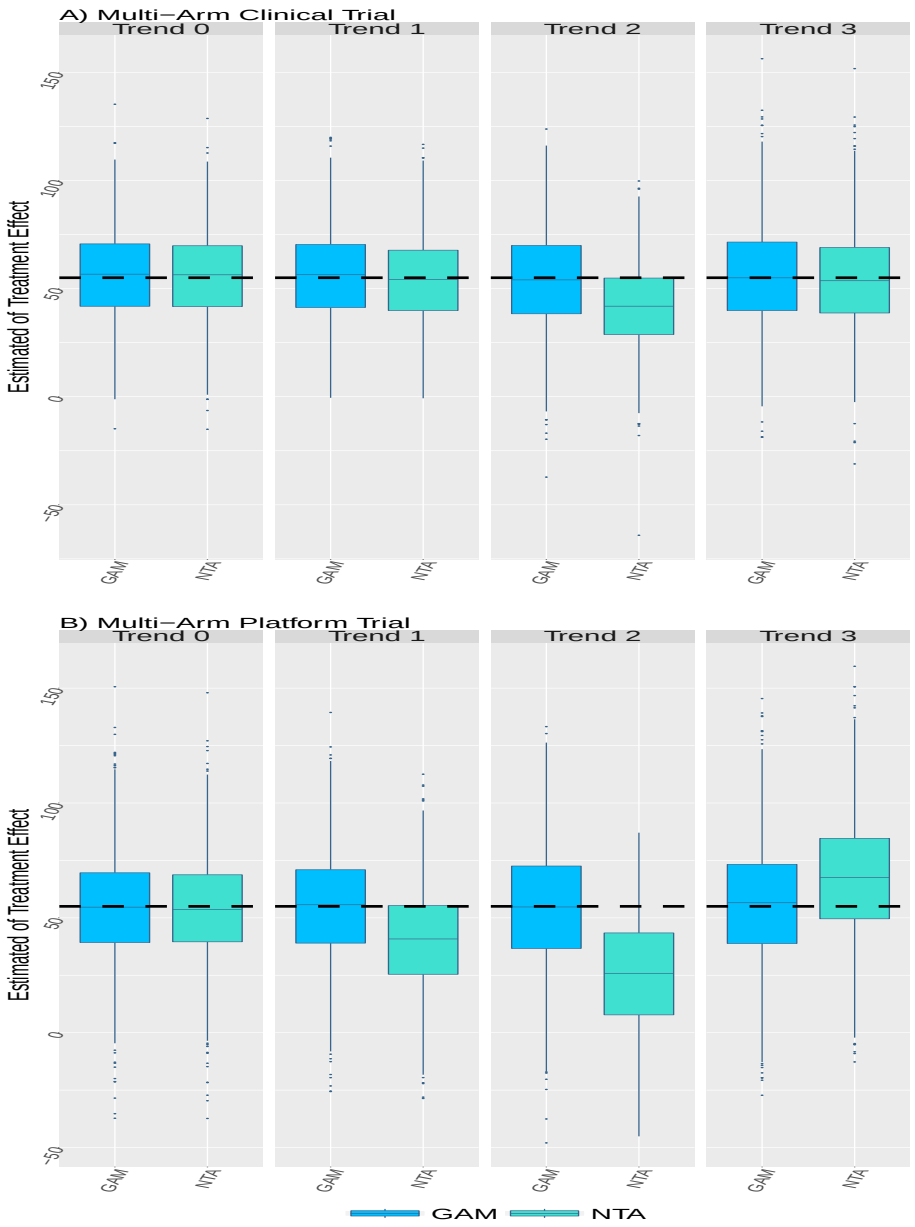

Figure 6.2. Treatment effect estimates as $100 * \Delta_a$, for $a > 0$. The black line corresponds to the true value of the treatment effect (55%). Results are based on 1000 simulated BAR trials using either the GAM model (GAM) and the No Trend Adjustment approach (NTA). Panel A: Treatment effect estimates of an initial experimental arm. The overall sample size is 240 patients. Panel B: Treatment effect estimates of an experimental arm added after 102 patients. The overall sample size is 250 patients. Different trends are considered: 0) Absence of time trend in the patients population during the trial. 1) The true outcome probabilities decrease monotonically for all treatment arms. 2) The true outcome probabilities follows a seasonal trend. 3) The true outcome probabilities increase monotonically for all treatment arms.

	NTA	GAM
Design	BR,PTW,DABC,BAR	BR,PTW,DABC,BAR
Scenario	100*($\Delta_1, \Delta_2, \Delta_3$)=(-50, 0 , 55)	
Arm	Trend 0	
1	-51.24, -52.20, -53.83, -50.90	-51.88, -52.98, -54.36, -51.7
2	1.72, 1.44, 0.89, 0.31	1.56, 1.43, 0.83, -0.06
3	55.83, 57.67, 54.35, 56	56.12, 58.30, 54.84, 56.40
	Trend 1	
1	-49.40, -47.79, -47.69, -40.99	-50.96, -60.18, -57.46, -53.54
2	2.08, 0.46, 3.00, 7.02	2.41, 0.35, 1.91, 0.01
3	55.0, 49.17, 54.85, 54.07	56.69, 56.16, 57.08, 56.25
	Trend 2	
1	-39.47, -32.00, -40.00, -19.87	-50.97, -50.84, -52.04, -50.83
2	2.82, 2.03, 0.35, 15.30	2.60, 3.07, 1.00, -0.28
3	44.54, 39.38, 43.00, 41.52	56.01, 58.27, 56.58, 54.01
	Trend 3	
1	-48.78, -56.39, -51.74, -59.49	-50.71, -53.07, -52.20, -50.91
2	2.37, 2.46, 1.60, -6.43	2.14, 2.46, 1.94, 0.42
3	55.35, 58.21, 57.01, 54.02	57.25, 56.69, 58.87, 55.50

Table 6.1. Treatment effect estimates as $100 * \Delta_a$, for $a > 0$, across 1000 simulated trials with all initial experimental arms and an overall sample size of 240 patients using either the GAM model (GAM) and the No Trend Adjustment approach (NTA). Different trend situations are explored: 0) Absence of time trend in the patients population during the trial. 1) The true outcome probabilities decrease monotonically for all treatment arms. 2) The true outcome probabilities follows a seasonal trend. 3) The true outcome probabilities increase monotonically for all treatment arms. For each trend, trials are run under balanced randomization (BR), play the winner design (PTW), doubly adaptive biased coin design (DABC) and Bayesian adaptive randomization (BAR).

	NTA	GAM
Design	BR, PTW, DABC, BAR	BR, PTW, DABC, BAR
Scenario	(Δ_2, Δ_3)=(0, 0.55)	
Arm	Trend 0	
2	5.0, 5.0, 4.0, 4.7	5.1, 5.1, 5.8, 5.4
3	75.0, 76.1, 75.0, 83.4	74.3, 78.2, 78.3, 83.0
	Trend 1	
2	4.1, 4.0, 4.3, 4.2	4.9, 4.1, 4.7, 5.5
3	76, 73.5, 79.5, 85.4	74.9, 77.6, 78.5, 84.9
	Trend 2	
2	4.5, 2.2, 3.9, 6.0	4.8, 5.6, 6.0, 5.0
3	73.8, 49.5, 60.0, 85.6	65.5, 70.4, 74.0, 78.9
	Trend 3	
2	4.0, 6.7, 5.2, 5.3	5.2, 6.1, 4.8, 5.9
3	71.3, 77.2, 69.0, 82.6	68.0, 80.7, 72.2, 75.0

Table 6.2. Empirical type I error rates and power based on 1000 simulated trials with all initial experimental treatments and an overall sample size of 240 patients, an accrual rate of 6 patients per month and a delay outcome of 8 weeks. Results are obtained under balanced randomization (BR), play the winner design (PTW), doubly adaptive biased coin design (DABC) and Bayesian adaptive randomization (BAR) by using either the GAM model (GAM) or the No Trend Adjustment (NTA) procedure. Different time trends are considered: 0) Absence of time trend in the patients population during the trial. 1) The true outcome probabilities decrease monotonically for all treatment arms. 2) The true outcome probabilities follows a seasonal trend. 3) The true outcome probabilities increase monotonically for all treatment arms.

	NTA	GAM
Design	BR,PTW,DABC,BAR	BR,PTW,DABC,BAR
Scenario	100*$(\Delta_1, \Delta_2, \Delta_3, \Delta_4)$=(0, 25, 0, 55)	
Arm	Trend 0	
1	-0.14, 0.08, 0.30, -3.20	-0.03, 0.09, 0.32, -3.71
2	24.42, 25.09, 23.88, 22.22	24.49, 25.33, 24.04, 22.05
3	2.56, 4.40, 1.86, -2.02	2.75, 24.08, 2.24, -2.71
4	56.06, 51.13,56.20, 53.82	56.47, 227.47, 57.22, 56.5
	Trend 1	
1	1.2, 0.41, 0.43, 5.50	1.39, 0.37, -1.41, -1.86
2	24.13, 21.90, 25.58, 23.65	25.20, 25.36, 23.84, 22.51
3	- 14.38, 11.13, -7.89, -5.73	3.47, 87.25, 0.83, -1.02
4	38.32, 65.19, 43.45, 40.24	57.41, 100.17, 54.76, 55.18
	Trend 2	
1	0.71, 0.54, -1.15, 7.33	1.17, 0.64, -0.98, -0.89
2	19.96, 18.13, 18.30, 17.93	24.37, 24.77, 23.13, 23.89
3	-32.69, 34.45, -19.75, -6.26	2.68, 87.25, 1.43, 1.04
4	11.14, 66.45, 21.09, 25.08	55.77, 219.51, 54.75, 53.74
	Trend 3	
1	-0.04, 0.13, -1.43, -8.53	-0.07, 0.28, 0.13, -1.39
2	27.73, 25.75, 23.28, 19.02	24.30, 24.67, 25.23, 22.48
3	19.35, -10.31, 13.01, 6.98	0.92, 143.97, 3.09, 0.97
4	73.32, 22.00, 66.01, 67.74	62.76, 552.94, 57.22, 56.50

Table 6.3. Treatment effect estimates as $100 * \Delta_a$, for $a > 0$, across 1000 simulated trials with four experimental treatments and an overall sample size of 250 patients using either the GAM model (GAM) and the No Trend Adjustment approach (NTA). The trial starts with two experimental arms and the control. Arms 3 and 4 are added at $M_2 = 102$. Different trend situations are explored: 0) Absence of time trend in the patients population during the trial. 1) The true outcome probabilities decrease monotonically for all treatment arms. 2) The true outcome probabilities follows a seasonal trend. 3) The true outcome probabilities increase monotonically for all treatment arms. For each trend, trials are run under balanced randomization (BR), play the winner design (PTW), doubly adaptive biased coin design (DABC) and Bayesian adaptive randomization (BAR).

Design Scenario	NTA BR, PTW, DABC, BAR	GAM BR, PTW, DABC, BAR $(\Delta_2, \Delta_3, \Delta_4) = (0.55, 0, 0)$	GAM with Stopping Rule BR, PTW, DABC, BAR
Arm		Trend 0	
2	64.1, 83.9, 71.2, 77.0	67.0, 79.9, 74.1, 82.7	66.0, 76.0, 71.0, 80.7
3	8.2, 17.2, 13.1, 26.1	4.2, 4.8, 6.1, 4.6	6.6, 4.0, 6.4, 4.2
4	7.5, 15.9, 13.4, 25.7	4.9, 4.8, 5.8, 5.7	6.2, 8.0, 4.6, 6.0
		Trend 1	
2	69.1, 75.8, 73.3, 79.6	67.8, 84.2, 76.1, 82.2	66.4, 80.0, 75.7, 79.1
3	8.2, 8.7, 1.8, 22.8	4.7, 5.2, 5.2, 4.6	6.4, 12.0, 5.0, 4.8
4	8.0, 8.5, 1.3, 25.3	5.1, 5.3, 5.0, 5.7	6.7, 8.0, 4.8, 4.9
		Trend 2	
2	69.0, 59.3, 51.7, 77.1	58.0, 79.1, 63.3, 70.7	58.3, 72.2, 64.1, 72.2
3	1.4, 23.0, 0.1, 12.3	4.2, 2.1, 4.3, 5.9	3.8, 3.3, 4.0, 5.4
4	0.9, 22.6, 0.1, 12.1	4.6, 2.1, 4.5, 5.8	4.1, 3.0, 4.1, 4.8
		Trend 3	
2	66.6, 77.9, 63.1, 78.5	62.5, 77.9, 66.6, 73.3	60.7, 72.7, 65.8, 72.7
3	7.9, 25.2, 48.4, 25.7	3.4, 0.1, 5.7, 3.7	5.9, 1.5, 5.5, 6.1
4	8.7, 25.5, 45.4, 24.1	4.4, 0.1, 4.9, 5.3	5.5, 1.5, 5.0, 6.1

Table 6.4. Empirical type I error rates and power across 1000 simulated trials with two initial experimental treatments (Arm 1 and 2) and two new arms (Arm 3 and 4) added at $M_2 = 102$. The overall sample size is 250 patients, with an accrual rate of 6 patients per month and a delay outcome of 8 weeks. Results are obtained under balanced randomization (BR), play the winner design (PTW), doubly adaptive biased coin design (DABC) and Bayesian adaptive randomization (BAR) by using either the GAM model (GAM) or the No Trend Adjustment (NTA) procedure. Different time trends are considered: 0) Absence of time trend in the patients population during the trial. 1) The true outcome probabilities decrease monotonically for all treatment arms. 2) The true outcome probabilities follows a seasonal trend. 3) The true outcome probabilities increase monotonically for all treatment arms.

Bibliography

[1] A. Agresti. Categorical Data Analysis. Wiley, 2002.

[2] D. G. Altman and J. P. Royston. The hidden effect of time. Statistics in medicine, 7(6):629–637, 1988.

[3] J. Andersen, D. Faries, and R. Ramura. A randomized play-the-winner design for multi-arm clinical trials. Communications in Statistics-Theory and Methods, 23(2):309–323, 1994.

[4] J. Babb, A. Rogatko, and S. Zacks. Cancer phase i clinical trials: efficient dose escalation with overdose control. Statistics in medicine, 17(10):1103–1120, 1998.

[5] J. S. Babb and A. Rogatko. Patient specific dosing in a cancer phase i clinical trial. Stat Med, 20(14):2079–2090, 2001.

[6] Z. Bai, F. Hu, and L. Shen. An adaptive design for multi-arm clinical trials. Journal of Multivariate Analysis, 81(1):1–18, 2002.

[7] S. Bailey, B. Neuenschwander, G. Laird, and M. Branson. A bayesian case study in oncology phase i combination dose-finding using logistic regression with covariates. Journal of Biopharm Stat, 19(3):469–484, 2009.

[8] A. Barker, C. Sigman, G. Kelloff, N. Hylton, D. Berry, and L. Esserman. I-spy 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. Clinical Pharmacology & Therapeutics, 86(1):97–100, 2009.

[9] J. O. Berger. Statistical decision theory and Bayesian analysis. Springer, 1985.

[10] T. M. Braun. The bivariate continual reassessment method: extending the crm to phase i trials of two competing outcomes. Controlled clinical trials, 23(3): 240–256, 2002.

[11] A. K. Burnett, N. H. Russell, R. K. Hills, A. E. Hunter, L. Kjeldsen, J. Yin, B. E. Gibson, K. Wheatley, and D. Milligan. Optimization of chemotherapy for younger patients with acute myeloid leukemia: results of the medical research council aml15 trial. Journal of clinical oncology, 31(27):3360–3368, 2013.

[12] D. P. Byar, R. M. Simon, W. T. Friedewald, J. J. Schlesselman, D. L. DeMets, J. H. Ellenberg, M. H. Gail, and J. H. Ware. Randomized clinical trials: perspectives on some recent ideas. New England Journal of Medicine, 295(2): 74–80, 1976.

[13] S. Chib and E. Greenberg. Analysis of multivariate probit models. Biometrika, 85(2):347–361, 1998.

[14] D. Clayton. Ethically optimised designs. British journal of clinical pharmacology, 13(4):469–480, 1982.

[15] D. R. Cohen, S. Todd, W. M. Gregory, and J. M. Brown. Adding a treatment arm to an ongoing clinical trial: a review of methodology and practice. Trials, 16(1):179, 2015.

[16] J. Cornfield, M. Halperin, and S. W. Greenhouse. An adaptive procedure for sequential clinical trials. Journal of the American Statistical Association, 64 (327):759–770, 1969.

[17] T. P. Coroller, R. H. Mak, J. H. Lewis, E. H. Baldini, A. B. Chen, Y. L. Colson, F. L. Hacker, G. Hermann, D. Kozono, E. Mannarino, et al. Low incidence of chest wall pain with a risk-adapted lung stereotactic body radiation therapy

approach using three or five fractions based on chest wall dosimetry. PloS one, 9(4):e94859, 2014.

[18] N. E. Dunlap, J. Cai, G. B. Biedermann, W. Yang, S. H. Benedict, K. Sheng, T. E. Schefter, B. D. Kavanagh, and J. M. Larner. Chest wall volume receiving> 30 gy predicts risk of severe pain and/or rib fracture after lung stereotactic body radiotherapy. International Journal of Radiation Oncology, 76(3):796–801, 2010.

[19] S. Durrleman and R. Simon. Flexible regression models with cubic splines. Statistics in medicine, 8(5):551–561, 1989.

[20] J. R. Eisele. The doubly adaptive biased coin design for sequential clinical trials. Journal of Statistical Planning and Inference, 38(2):249–261, 1994.

[21] J. J. Elm, Y. Y. Palesch, G. G. Koch, V. Hinson, B. Ravina, and W. Zhao. Flexible analytical methods for adding a treatment arm mid-study to an ongoing clinical trial. Journal of biopharmaceutical statistics, 22(4):758–772, 2012.

[22] H. Feng, J. Shao, and S.-C. Chow. Adaptive group sequential test for clinical trials with changing patient population. Journal of biopharmaceutical statistics, 17(6):1227–1238, 2007.

[23] B. Freidlin, E. L. Korn, R. Gray, and A. Martin. Multi-arm clinical trials of new agents: some design considerations. Clinical Cancer Research, 14(14): 4368–4371, 2008.

[24] S. N. Goodman, M. L. Zahurak, and S. Piantadosi. Some practical improvements in the continual reassessment method for phase i studies. Stat Med, 14(11): 1149–1161, 1995.

[25] T. A. Gooley, P. J. Martin, L. D. Fisher, and M. Pettinger. Simulation as a design tool for phase i/ii clinical trials: an example from bone marrow transplantation. Controlled clinical trials, 15(6):450–462, 1994.

[26] R. P. Guiteras, D. I. Levine, and T. H. Polley. The pursuit of balance in sequential randomized trials. Development Engineering, 1:12–25, 2016.

[27] T. Hastie and R. Tibshirani. Generalized additive models. Wiley Online Library, 1990.

[28] F. C. Henríquez, S. V. Castrillón, et al. A quality index for equivalent uniform dose. Journal of Medical Physics, 36(3):126, 2011.

[29] R. K. Hills and A. K. Burnett. Applicability of a ?pick a winner? trial design to acute myeloid leukemia. Blood, 118(9):2389–2394, 2011.

[30] B. P. Hobbs, N. Chen, and J. J. Lee. Controlled multi-arm platform design using predictive probability. Statistical methods in medical research, page 0962280215620696, 2016.

[31] F. Hu and W. F. Rosenberger. Analysis of time trends in adaptive designs with application to a neurophysiology experiment. Statistics in medicine, 19(15): 2067–2075, 2000.

[32] F. Hu and L.-X. Zhang. Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trials. Annals of Statistics, pages 268–301, 2004.

[33] E. L. Korn and B. Freidlin. Outcome-adaptive randomization: is it useful? Journal of Clinical Oncology, 29(6):771–776, 2010.

[34] 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. Stat Med, 13(18): 1799–1806, 1994.

[35] S. L. Kwa, J. C. Theuws, A. Wagenaar, E. M. Damen, L. J. Boersma, P. Baas, S. H. Muller, and J. V. Lebesque. Evaluation of two dose–volume histogram reduction models for the prediction of radiation pneumonitis. Radiotherapy and oncology, 48(1):61–69, 1998.

[36] J. Lee, P. F. Thall, Y. Ji, and P. Müller. Bayesian dose-finding in two treatment cycles based on the joint utility of efficacy and toxicity. JASA, 110(510):711–722, 2015.

[37] J. J. Lee, X. Gu, and S. Liu. Bayesian adaptive randomization designs for targeted agent development. Clinical Trials, 7(5):584–596, 2010.

[38] W. Li. Sequential designs for opposing failure functions. 1997.

[39] J. A. Lieberman, T. S. Stroup, J. P. McEvoy, M. S. Swartz, R. A. Rosenheck, D. O. Perkins, R. S. Keefe, S. M. Davis, C. E. Davis, B. D. Lebowitz, et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. N Engl j Med, 2005(353):1209–1223, 2005.

[40] R. Lin and G. Yin. Bayesian optimal interval design for dose finding in drug-combination trials. Statistical methods in medical research, page 0962280215594494.

[41] A. M. Lipsky and S. Greenland. Confounding due to changing background risk in adaptively randomized trials. Clinical Trials, 8(4):390–397, 2011.

[42] S. Liu and V. E. Johnson. A robust bayesian dose-finding design for phase i/ii clinical trials. Biostatistics, page kxv040, 2015.

[43] L. B. Marks, S. M. Bentzen, J. O. Deasy, J. D. Bradley, I. S. Vogelius, I. El Naqa, J. L. Hubbs, J. V. Lebesque, R. D. Timmerman, M. K. Martel, et al. Radiation dose–volume effects in the lung. International Journal of Radiation Oncology, 76(3):S70–S76, 2010.

[44] P. A. Murtaugh and L. D. Fisher. Bivariate binary models of efficacy and toxicity in dose-ranging trials. Communications in Statistics - Theory and Methods, 19 (6):2003–2020, 1990.

[45] B. Nebiyou Bekele and Y. Shen. A bayesian approach to jointly modeling

toxicity and biomarker expression in a phase i/ii dose-finding trial. Biometrics, 61(2):343–354, 2005.

[46] J. O'Quigley and L. Z. Shen. Continual reassessment method: a likelihood approach. Biometrics, pages 673–684, 1996.

[47] J. O'Quigley, M. Pepe, and L. Fisher. Continual reassessment method: a practical design for phase 1 clinical trials in cancer. Biometrics, pages 33–48, 1990.

[48] J. O'Quigley, L. Z. Shen, and A. Gamst. Two-sample continual reassessment method. Journal of Biopharm Stat, 9(1):17–44, 1999.

[49] M. K. Parmar, J. Carpenter, and M. R. Sydes. More multiarm randomised trials of superiority are needed. The Lancet, 384(9940):283, 2014.

[50] S. Piantadosi and G. Liu. Improved designs for dose escalation studies using pharmacokinetic measurements. Stat Med, 15(15):1605–1618, 1996.

[51] M.-K. Riviere, Y. Yuan, F. Dubois, and S. Zohar. A bayesian dose-finding design for drug combination clinical trials based on the logistic model. Pharm Stat, 13(4):247–257, 2014.

[52] M.-K. Riviere, F. Dubois, and S. Zohar. Competing designs for drug combination in phase i dose-finding clinical trials. Stat Med, 34(1):1–12, 2015.

[53] W. F. Rosenberger. Randomized play-the-winner clinical trials: review and recommendations. Controlled clinical trials, 20(4):328–342, 1999.

[54] W. F. Rosenberger and F. Hu. Bootstrap methods for adaptive designs. Statistics in medicine, 18(14):1757–1767, 1999.

[55] K. L. Stephans, T. Djemil, C. A. Reddy, S. M. Gajdos, M. Kolar, D. Mason, S. Murthy, T. W. Rice, P. Mazzone, M. Machuzak, et al. A comparison of two stereotactic body radiation fractionation schedules for medically inoperable

stage i non-small cell lung cancer: the cleveland clinic experience. Journal of Thoracic Oncology, 4(8):976–982, 2009.

[56] B. E. Storer. Design and analysis of phase i clinical trials. Biometrics, pages 925–937, 1989.

[57] R. S. Sutton and A. G. Barto. Introduction to reinforcement learning, volume 135. MIT Press Cambridge, 1998.

[58] P. Thall and S.-J. Lee. Practical model-based dose-finding in phase i clinical trials: Methods based on toxicity. International Journal of Gynecological Cancer, 13(3):251–261, 2003.

[59] P. Thall, P. Fox, and J. Wathen. Statistical controversies in clinical research: scientific and ethical problems with adaptive randomization in comparative clinical trials. Annals of Oncology, 26(8):1621–1628, 2015.

[60] P. F. Thall. Bayesian adaptive dose-finding based on efficacy and toxicity. J Statistical Research, 14:187–202, 2012.

[61] P. F. Thall and J. D. Cook. Dose-finding based on efficacy–toxicity trade-offs. Biometrics, 60(3):684–693, 2004.

[62] P. F. Thall and K. E. Russell. A strategy for dose-finding and safety monitoring based on efficacy and adverse outcomes in phase i/ii clinical trials. Biometrics, pages 251–264, 1998.

[63] P. F. Thall and J. K. Wathen. Practical bayesian adaptive randomisation in clinical trials. European Journal of Cancer, 43(5):859–866, 2007.

[64] P. F. Thall, R. E. Millikan, P. Mueller, and S.-J. Lee. Dose-finding with two agents in phase i oncology trials. Biometrics, 59(3):487–496, 2003.

[65] P. F. Thall, H. Q. Nguyen, and E. H. Estey. Patient-specific dose finding based on bivariate outcomes and covariates. Biometrics, 64(4):1126–1136, 2008.

[66] P. F. Thall, H. Q. Nguyen, S. Zohar, and P. Maton. Optimizing sedative dose in preterm infants undergoing treatment for respiratory distress syndrome. Journal of the American Statistical Association, 109(507):931–943, 2014.

[67] L. Trippa, E. Q. Lee, P. Y. Wen, T. T. Batchelor, T. Cloughesy, G. Parmigiani, and B. M. Alexander. Bayesian adaptive randomized trial design for patients with recurrent glioblastoma. Journal of Clinical Oncology, 30(26):3258–3263, 2012.

[68] F. van Leth, P. Phanuphak, K. Ruxrungtham, E. Baraldi, S. Miller, B. Gazzard, P. Cahn, U. Lalloo, I. Van Der Westhuizen, D. Malan, et al. Comparison of first-line antiretroviral therapy with regimens including nevirapine, efavirenz, or both drugs, plus stavudine and lamivudine: a randomised open-label trial, the 2nn study. The Lancet, 363(9417):1253–1263, 2004.

[69] S. Ventz and L. Trippa. Bayesian designs and the control of frequentist characteristics: a practical solution. Biometrics, 71(1):218–226, 2015.

[70] S. Ventz, B. M. Alexander, G. Parmigiani, R. D. Gelber, and L. Trippa. Designing clinical trials that accept new arms: An example in metastatic breast cancer. Journal of Clinical Oncology, pages JCO–2016, 2017.

[71] S. Ventz, M. Cellamare, G. Parmigiani, and L. Trippa. Adding experimental arms to platform clinical trials: randomization procedures and interim analyses. Biostatistics, (in press) DOI: 10.1093/biostatistics/kxx030:, 2017.

[72] K. Wang and A. Ivanova. Two-dimensional dose finding in discrete dose space. Biometrics, 61(1):217–222, 2005.

[73] J. Wason and L. Trippa. A comparison of bayesian adaptive randomization and multi-stage designs for multi-arm clinical trials. Statistics in medicine, 33(13): 2206–2221, 2014.

[74] L. Wei. The generalized polya's urn design for sequential medical trials. The Annals of Statistics, pages 291–296, 1979.

[75] L. Wei and S. Durham. The randomized play-the-winner rule in medical trials. Journal of the American Statistical Association, 73(364):840–843, 1978.

[76] M. C. Weinstein. Allocation of subjects in medical experiments. New England Journal of Medicine, 291(24):1278–1285, 1974.

[77] N. M. Woody, G. M. Videtic, K. L. Stephans, T. Djemil, Y. Kim, and P. Xia. Predicting chest wall pain from lung stereotactic body radiotherapy for different fractionation schemes. International Journal of Radiation Oncology, 83(1): 427–434, 2012.

[78] G. Yin and Y. Yuan. Bayesian dose finding in oncology for drug combinations by copula regression. JRSS: Series C, 58(2):211–224, 2009.

[79] G. Yin, Y. Li, and Y. Ji. Bayesian dose-finding in phase i/ii clinical trials using toxicity and efficacy odds ratios. Biometrics, 62(3):777–787, 2006.

[80] Y. Yuan, B. Guo, M. Munsell, K. Lu, and A. Jazaeri. Midas: a practical bayesian design for platform trials with molecularly targeted agents. Statistics in medicine, 35(22):3892–3906, 2016.

[81] 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, pages clincanres–0592, 2016.

[82] M. Zelen. Play the winner rule and the controlled clinical trial. Journal of the American Statistical Association, 64(325):131–146, 1969.

Thank you/Grazie

Esattamente come tre anni fa mi trovo qui a tirare le somme di un percorso intrapreso e concluso. Sebbene il prodotto finale di questi anni si concretizzi in una tesi, in realta' molto di piu' e' nascosto tra queste pagine. Gioie, frustrazioni, sforzi, esultanze, ma soprattutto... Persone. Alcune di esse erano gia' presenti tra le pagine delle mie precedenti tesi; altre si sono aggiunte nel corso di questi anni. In entrambi i casi, ci tengo a ringraziarle tutte. Ognuno a proprio modo, chi piu' e chi meno, ha contribuito a questa mia crescita. Mi avete preso per mano e mi avete accompagnato giorno dopo giorno. E quindi...

Thank you to the BCB and the OHS departments at the Dana Farber Cancer Institute for accepting me into their community. A special thanks to Luz and Lorraine for taking care of all my paperwork and always being ready to help me with procedures. A big thanks to Jean for always being very nice to me, I will never forget your kindness.

Ringrazio il Dipartimento di Scienze Statistiche all' Universita' di Roma, "La Sapienza". Sebbene abbia passato qui poco tempo, tutte le volte che sono tornata mi sono sempre sentita a casa.

Thanks to Lorenzo and Steffen for being my advisors and collaborating with me during these years. Our discussions helped me grow both as a statistician and as a person.

Grazie a tutti i miei colleghi dottorandi dell' Universita' di Roma, "La Sapienza",

per avermi tenuta sempre informata su cosa succedeva nel nostro dipartimento ed essere sempre stati disponibili ad aiutarmi quando ho avuto bisogno di qualcosa.

Thanks to Boyu, who I always called "Buio", for all your advice. I could not wish for a better officemate.

Grazie a Matteo, perche' in te ho trovato un supporto statistico e psicologico, ma soprattutto un grande amico.

Mille e piu' volte grazie ai miei genitori. Grazie per aver creduto ancora una volta nelle mie scelte. Grazie per avermi fatto sentire libera di poter prendere qualsiasi decisione. Grazie per non avermi mai fatto pesare il fatto che fossi andata lontano da voi. Grazie per aver colmato la vostra assenza fisica con tutte le attenzioni e cure che solo una madre e un padre possono dare. Grazie perche' nonostante la malinconia che ci accompagna ogni volta che ci salutiamo, leggo nei vostri occhi l' orgoglio di avermi come figlia e la felicita' di vedermi felice. Grazie per avermi insegnato che "amare qualcuno significa lasciarlo libero".

Un dolcissimo grazie a mio fratello Giovanni, in arte Rashid. E' grazie a te se ho potuto portare a termine questo percorso. Grazie per essere sempre stato pronto ad aiutarmi, per avermi rassicurato e per aver placato tutte le mie preoccupazioni. Ma soprattutto grazie per esserti preso cura della nostra famiglia in ogni momento, specialmente quando piu' ne avevamo bisogno. Grazie perche' in te, il "piccolo" della famiglia, ho ritrovato un grande uomo.

Grazie alle mie nonne, per avermi tenuto sempre stretta nei loro pensieri, e nel loro cuore.

Grazie alle mie zie, i miei zii e i miei cugini, da Nord a Sud, che con tutti i loro immancabili "Buongiorno" mi hanno fatto iniziare le mie giornate sempre con il sorriso.

Un ringraziamento speciale a mia cugina Roberta. Sei una pallina, ma riesci sempre a farmi ridere e a tirarmi su di morale. Grazie perche' so di poter contare su

di te, sempre.

Grazie mille al mio "Lukketto", perche' non importa dove siamo e per quanto tempo non ci vediamo, il nostro legame continua resistere. Grazie perche' sei una delle piu' grandi fortune che mi siano mai capitate.

Grazie mille alla mia "Palma" per esserci sempre stata. Grazie perche', esattamente come un acero, rappresenti conforto, comprensione, sostegno, lealta' e generosita'. Grazie perche' nonostante la mia intollerabilita', ritorni sempre. Ti prego, non ti stancare.

Thanks to Elif and Jason, and to all my wonderful friends from the MOE family for being more than just a fencing club. As someone suggested "you made me see the light, when everything was dark". It would not be the same without you guys!

Lastly, thanks to Ivan. I should thank you for many things: for helping me with my thesis, my emails, my grammar, my presentations and so on. And although I am really thankful for all of them, I have to thank you for something much more important. Thank you for believing in me more than I believed in myself. Thank you for being more happy for me than I was for myself. Thank you for always being inexplicably proud of me, no matter what. Thank you for always standing by my side and tolerating me during my bad days. Thank you for always letting me find your smile at the beginning and at the end of every day. Thank you for giving me the reason to come back and go ahead.

Grazie a tutti bella gente,

Thank you all beautiful people,

Ilaria