# Covariate-Adjusted Response-Adaptive Designs for Clinical Trials

Ayon Mukherjee

Submitted in partial fulfillment of the requirements of the Degree of Doctor of Philosophy

School of Mathematical Sciences

Queen Mary University of London

# Statement of Originality

I, Ayon Mukherjee, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third partys copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature: Date:

# Acknowledgement

It has been a period of intense learning during my doctoral research, both in the scientific arena and on a personal level. Therefore, I would like to reflect on who have supported me throughout this period.

This research was supported by a Queen Mary Doctoral Scholarship. Foremost, I would like to express my sincere gratitude to my supervisor Dr. D.S. Coad for his continuous support during my Ph.D research. His patience and motivation helped me to achieve my goal. His guidance helped me in my research and also in writing this thesis. Dr. Coad consistently allowed this thesis to be my own work, but steered me in the right direction to enhance my ability to think individually.

Besides my supervisor, I would like to thank my assessor: Dr. L.I. Pettit and Dr. H. Maruri-Aguilar, for their insightful comments during the annual reviews and making me focus on improving my thesis writing skills. My sincere thanks also go to Dr. M. Witcombe and Ms K. Hale, for providing me with administrative support during my Ph. D study .

I also thank Professor W.F. Rosenberger and Dr. A Sverdlov who shared their ideas on my research topic during my visit to the USA, which was supported by Eileen Colyer Award. Their guidance and support played a significant role in me being able to present my work based on Chapter 4 of this thesis at the $39^{th}$ Annual conference of the International Society of Clinical Biostatistics (ISCB) Melbourne, Australia, in August 2018. My travel to the ISCB conference was supported by the Queen Mary Post-graduate Research Fund. Previous to this, the work in Chapter 2 of my thesis have also been presented at the $61^{st}$ World Statistics Congress of the International Statisticial Institute, in Marrakech, Morocco in July 2017 and the work in Chapter 3 of my thesis at the $28^{th}$ International Biometric Conference in Barcelona, Spain, in July

2018.

Each of the main chapters in this thesis are intended to be published in a journal. Chapter 2 will be submitted to the Journal of Statistical Planning and Inference. Next, the target journal for Chapter 3 is Statistics in Medicine. Finally, Chapter 4 is planned to be submitted in the Journal of Biopharmaceutical Statistics.

Last but not least, I would like to thank my family : my parents for supporting me spiritually throughout my life.

# Abstract

Covariate-adjusted response-adaptive (CARA) randomization designs use the available responses to skew treatment allocation proportions towards the better performing treatment, for a given patient's covariate profile. Such designs have previously been developed for two treatments assuming exponentially distributed survival responses. Designs are obtained here considering Weibull distributed survival responses. The optimal treatment allocation proportions are targeted using biased coin procedures. These sequentially estimated allocation proportions converge to the expected target values, which are functions of the Weibull regression coefficients. CARA designs for Weibull survival model are also studied using the distribution function of a Gumbel model as a link between patient information history and the present allocation. Results show that the proposed designs are a suitable alternative to balanced randomization designs according to their powers and type I error rates.

To make CARA designs more applicable, they are developed assuming proportional hazards of an event at a given time point. Optimal allocation proportions are derived that are targeted similarly. The estimates of these proportions converge to the expected target values, which are functions of the Cox regression coefficients. Other non-optimal CARA designs are also investigated which compete in different characteristics with the optimal CARA ones and the balanced designs. Simulation results show that the optimal CARA designs outperform the other designs based on their powers and type I error rates.

For the application of CARA designs in clinical trials with events due to multiple causes, they are developed assuming proportional sub-distribution hazards at a given time point. Optimal allocation proportions for the primary cause of interest are derived that are targeted similarly. The sequentially estimated allocation proportions for these

designs converge to the expected target values, which are functions of the Fine and Gray model coefficients. Simulation results reveal the need of a theoretical procedure for more complicated semi-parametric survival response models.

# Contents

**Statement of Originality** i

**Acknowledgement** ii

**Abstract** iv

**List of Figures** viii

**List of Tables** ix

**1 Introduction** 1

1.1 Introduction . . . . . 1

1.2 Journey of a test article during the process of a Clinical Trial . . . . . 2

1.3 Role of a Trial Statistician During the Process of Clinical Research . . 6

1.4 Adaptive Design in Clinical Trials . . . . . 9

1.5 Survival Responses . . . . . 13

1.6 Outline of the Thesis . . . . . 14

**2 Covariate-Adjusted Response-Adaptive Designs for Weibull Survival Models** 16

2.1 Introduction . . . . . 16

2.2 The Weibull Model . . . . . 19

2.3 Construction of the Likelihood . . . . . 22

2.4 The Proposed Allocation Designs . . . . . 26

2.4.1 CARA Design using the Optimal Allocation Approach . . . . . 28

2.4.2 Targeting the Derived Allocation Proportion . . . . . 31

2.4.3 The Effect of Delay in Survival Responses . . . . . 35
2.4.4 Using a Link Function to Develop a Suitable Alternative . . . . 35
2.5 Asymptotic Properties of CARA Designs . . . . . 39
2.6 Simulation Study . . . . . 42
2.6.1 Choice of Design Parameters . . . . . 42
2.6.2 Comparing CARA Designs for No Difference Between the Treatment Effects . . . . . 49
2.6.3 Comparing CARA Designs for Differences in Treatment Effects 52
2.6.4 Distribution of the Allocation Proportions . . . . . 58
2.7 Real-Life Example of the Proposed Methodology . . . . . 60
2.8 Critical Evaluation . . . . . 62
2.9 Discussion . . . . . 66

**3 Covariate-Adjusted Response-Adaptive Designs for Semi-Parametric Survival Models 69**

3.1 Introduction . . . . . 69
3.2 Background on Proportional Hazards . . . . . 73
3.3 Deriving the Parameter Estimates from Partial Likelihood . . . . . 79
3.4 The Proposed Semi-Parametric CARA Designs . . . . . 83
3.4.1 Deriving a Suitable Semi-Parametric Target Allocation Proportion 85
3.5 Simulation Results . . . . . 91
3.6 Re-designing a Real-Life Clinical Trial . . . . . 99
3.7 Discussion . . . . . 101

**4 Covariate-Adjusted Response-Adaptive Designs for Competing Risk Survival Models 105**

4.1 Introduction . . . . . 105
4.2 Regression Models for Competing Risk Problems . . . . . 108
4.3 Proposed CARA Designs Based on Competing Risk Models . . . . . 114
4.3.1 Semi-Parametric CARA Designs with Target for Competing Risks Models . . . . . 115
4.3.2 Rules to Target the Derived Allocation Proportions . . . . . 117

4.4 Simulation Results . . . . . 119
4.5 Re-designing a Clinical Trial Using the Derived Procedures . . . . . 126
4.6 Discussion . . . . . 128

**5 Discussion** **130**
5.1 Summary of the Main Conclusions . . . . . 130
5.2 Critical Evaluation . . . . . 138
5.3 Directions for Future Research . . . . . 140

**Appendices** **142**

**A Deriving the variance of the Weibull distribution parameter estimates** **143**

**B Derving the Analytical Form for $\epsilon_k(\mathbf{z}; \boldsymbol{\beta}_k, \gamma_k)$** **145**

**C Deriving the Asymptotic Variance of the Hazard Ratio** **148**

**D Deriving the Optimal Allocation Proportion for minimizing the Total Cumulative Hazard Function** **151**

**E Deriving the Optimal Allocation Proportion for minimizing the Overall trial size for Semi-parametric Survival Models** **155**

**F Deriving the Asymptotic Variance of the Cause Specific Cumulative Incidence Function** **159**

**G Deriving the Optimal Allocation Proportion for minimizing the Overall Cumulative Incidence Function for a given Cause** **161**

**Bibliography** **165**

# List of Figures

1.1 The Usual Process of Drug Development . . . . . 6

2.1 Generalized type I right censoring . . . . . 23
2.2 Effect Size usually skews CARA allocation proportions towards better treatment . . . . . 58
2.3 Rate of change of the Observed Allocation Proportions . . . . . 65

3.1 Effect Size for Semi-parametric models skews CARA allocation proportions 98
3.2 Type I error rates and Powers for the semi-parametric Competing Designs . . . . . 99

4.1 Scenario for competing risk . . . . . 108
4.2 Performance of the Competing Designs based on the primary cause in the Individual Trials . . . . . 125

# List of Tables


2.1 Values of model parameters . . . . . 43
2.2 List of the Competing Designs . . . . . 45
2.3 Performances of the competing designs in the case of the neutral model 50
2.4 Performances of the competing designs in the case of the positive model 53
2.5 Performances of the competing designs in the case of the negative model 56
2.6 CARA Designs Outperforms other Designs for treating individual patients 61
2.7 Behaviour of the different link function-based designs . . . . . 65

3.1 List of the competing designs . . . . . 92
3.2 Simulation Results of the competing designs for neutral model . . . . . 94
3.3 Performance of the competing designs when Treatment effects differ . . 96
3.4 Re-design of a survival trial in colorectal cancer . . . . . 100

4.1 Values of model parameters . . . . . 120
4.2 List of the Competing Designs . . . . . 121
4.3 Performance of the Competing Designs for Cause 1 when Treatment Effects are Similar . . . . . 122
4.4 Performance of the Competing Designs from Cause 1 when Treatment Effects Differ Significantly . . . . . 124
4.5 Comparison of the original study design with the derived CARA designs 127

# Chapter 1

# Introduction

## 1.1 Introduction

Clinical research is a branch of healthcare science that determines the safety and efficacy of new interventions intended for human use. It is different from clinical practice in the way that it collects evidence to establish a treatment, whereas in clinical practice established treatments are used. A clinical research refers to any test article from its inception in the lab to its introduction to the consumer market and beyond.

Clinical trials are experiments done in clinical research involving human participants, and are designed to answer specific questions about biomedical or behavioural interventions, including new treatments. A clinical trial study can be categorized into two broad categories:

**Interventional study**: This is a type of clinical study in which participants are assigned to groups that receive one or more interventions so that researchers can evaluate the effects of the interventions on biomedical or health-related outcomes.

**Non-Interventional study**: Non-interventional studies refer to studies where the medical products are prescribed in the usual manner in accordance with the terms of

the marketing authorisation. The assignment of the patient to a particular therapeutic strategy is not decided in advance by a trial protocol but falls within current practice and the prescription of the medicine is clearly separated from the decision to include the patient in the study. No additional diagnostic or monitoring procedures are applied to the patients and epidemiological methods are used for the analysis of collected data.

The main focus of the entire thesis would be on developing advanced strategies in order to aid healthcare industries design a Phase III interventional study where two competing treatments are being compared in terms of its efficacy on human patients' survival responses which are sequentially available during the course of the trial.

## 1.2 Journey of a test article during the process of a Clinical Trial

After the discovery of a test article or a molecule in the laboratory, it goes through different stages of experiments before being established as a final marketing product for human well-being. Once the molecule is identified in the laboratory, it is subjected to pre-clinical studies or animal studies where different aspects of the test article, including its safety, toxicity if applicable, and efficacy if possible at this early stage, are studied. Results obtained from the pre-clinical studies or other supporting evidence, are submitted in support of an Investigational New Drug (IND) application to the Food and Drug Administration (FDA) for review prior to conducting clinical studies that involve humans. A clinical research may require the approval of Institutional Review Board (IRB) or Research Ethics Board (REB) and possibly other institutional committee reviews, about prior submission of the research to the FDA. Clinical research review criteria will depend on which federal regulations the research is subject to and will depend on which regulations the institutions subscribe to, in addition to any more stringent criteria added by the institution possibly in response to state or local laws/policies or accreditation entity recommendations. This additional layer of review (IRB/REB in particular) is critical to the protection of human subjects especially when one considers

that often research subject to the FDA regulation for prior submission is allowed to proceed, by those same FDA regulations, 30 days after submission to the FDA unless specifically notified by the FDA not to initiate the study. Therefore ethics plays a very significant role behind the sucess of any clinical research.

After the pre-clinical approval, the molecule enters the clinical phases of the research. Clinical trials involving new drugs are commonly classified into five phases. Each phase of the drug approval process is treated as a separate clinical trial. The drug-development process will normally proceed through all the phases over many years. If the drug successfully passes through phases 0, 1, 2, and 3, it will usually be approved by the national regulatory authority for use in the general population. Each phase has a different purpose and helps scientists answer a different question in relation developing the new intervention. The phases of clinical trials are defined below. Each of the phases definitions is a functional one and the terms are not defined on a strict chronological basis.

**Phase 0:** Phase 0 is a recent designation for optional exploratory trials conducted in accordance with the United States Food and Drug Administration's (FDA) 2006 Guidance on Exploratory Investigational New Drug (IND) Studies.This is the FDA(2006) document often refered to by the clinicians and the biostatisticians. Phase 0 trials are also known as human microdosing studies and are designed to speed up the development of promising drugs by establishing very early on whether the drug behaves in human subjects as was expected from preclinical studies. Distinctive features of Phase 0 trials include the administration of single subtherapeutic doses of the study drug to a small number of subjects (10 to 15) to gather preliminary data on the drug's pharmacokinetics (what the body does to the drugs).

A Phase 0 study gives no information on safety or efficacy, being by definition a dose too low to cause any therapeutic effect. Drug development companies carry out Phase 0 studies to rank drug candidates in order to decide which has the best pharmacokinetic parameters in humans to take forward into further development. They enable go/no-go decisions to be based on relevant human models instead of relying on

sometimes inconsistent animal data.

**Phase I:** Formerly referred to as "first-in-man studies", this phase comprises of initial safety trials, side effects, best dose, and formulation method for a new medicine. Since 1990s this phase of clinical research is being referred to as "first-in-human studies". An attempt is made here to establish the dose range tolerated by volunteers for single and for multiple doses. In order to achieve the objectives, a small group of 20 to 100 healthy volunteers are recruited in this phase of clinical trials. These trials are often conducted in a clinical trial clinic, where the subject can be observed by full-time staff. The subject who receives the drug is usually observed until several half-lives of the drug have passed. Phase I trials are also sometimes conducted in severely ill patients (e.g., for patients with terminal cancer) or in less ill patients when pharmacokinetic issues are addressed (e.g. metabolism of a new antiepileptic medicine in stable epileptic patients whose microsomal liver enzymes have been induced by other antiepileptic medicines). Pharmacokinetic and pharmacodynamics trials are usually considered Phase I trials regardless of when they are conducted during a medicine's development. These trials are usually conducted in tightly controlled clinics called CPUs (Central Pharmacological Units), where participants receive 24-hour medical attention and oversight.

**Phase IIa:** This is also referred to as 'proof of concept' studies. It involves pilot clinical trials to evaluate efficacy ( and safety) in selected populations of patients with the disease or condition to be treated, diagnosed, or prevented. Objectives here may focus on dose-response, type of patient, frequency of dosing, or numerous other characteristics of safety and efficacy.

**Phase IIb:** This is also referred to as 'definite dose-finding' studies. These are well controlled trials to evaluate efficacy ( and safety) in patients with the disease or condition to be treated, diagnosed, or prevented. These clinical trials usually represent the most rigorous demonstration of a medicine's optimum dose at which shows biological activity with minimal side-effects . Sometimes referred to as pivotal trials.

**Phase IIIa:** Trials conducted after efficacy of the medicine is demonstrated, but prior to regulatory submission of a New Drug Application (NDA) or other dossier. These clinical trials are conducted in patient populations for which the medicine is eventually intended. Phase IIIa clinical trials generate additional data on both safety and efficacy in relatively large numbers of patients in both controlled and uncontrolled trials. Clinical trials are also conducted in special groups of patients (e.g., renal failure patients) , or under special conditions dictated by the nature of the medicine and disease. These trials often provide much of the information needed for the package insert and labeling of the medicine.

**Phase IIIb:** Clinical trials conducted after regulatory submission of an NDA or other dossier, but prior to the medicine's approval and launch. These trials may supplement earlier trials, complete earlier trials, or may be directed toward new types of trials (e.g., quality of life, marketing) or Phase IV evaluations. This is the period between submission and approval of a regulatory dossier for marketing authorization.

**Phase IV:** Studies or trials conducted after a medicine is marketed to provide additional details about the medicine's efficacy or safety profile. Different formulations, dosages, durations of treatment, medicine interactions, and other medicine comparisons may be evaluated. New age groups, races, and other types of patients can be studied. Detection and definition of previously unknown or inadequately quantified adverse reactions and related risk factors are an important aspect of many Phase IV studies. If a marketed medicine is to be evaluated for another (i.e., new) indication, then those clinical trials are considered Phase II clinical trials. The term post-marketing surveillance is frequently used to describe those clinical studies in Phase IV (i.e., the period following marketing) that are primarily observational or non-experimental in nature, to distinguish them from well controlled Phase IV clinical trials or marketing studies.

The usual process of drug development is depicted in Figure 1.1 .The entire process

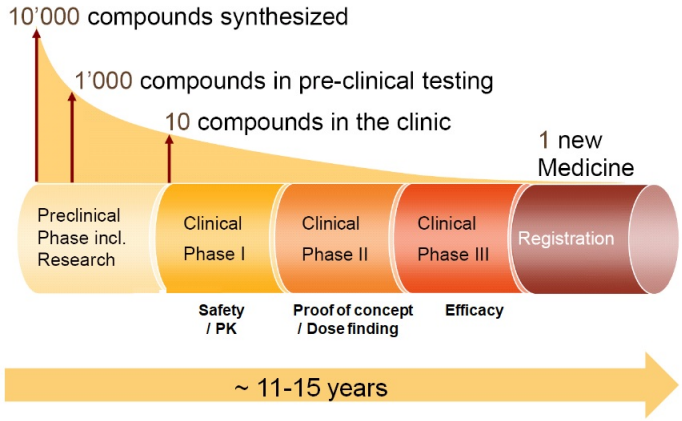


Figure 1.1: The Usual Process of Drug Development

of developing a drug from preclinical research to marketing can take several years and involves a huge amount of money. Therefore the intention of any clinical research is primarily on making the trial as ethical as possible without compromising much on the efficiancy of statistical evaluation. An ethical and efficient design can also attract more patients to paticipate in the clinical trial which can solve the problem of lack of patient participation in a clinical trial which the pharmaceutical industries often face mainly for late phase clinical trials. Therefore the role of a statistician becomes very significant in the whole process of the drug development.

## 1.3 Role of a Trial Statistician During the Process of Clinical Research

While performing a clinical research, the team involved in the process of drug development is called the clinical trial team (CTT). This team is responsible for the successful planning, conducting and execution of the trial. The team consists of many functions such as global trial leader, clinical managers, clinical scientist, data manager, trial statistician and statistical programmer. The statistical representative on the CTT is called the Trial Statistcian who would primarily be responsible for creating the Statistical Analysis Plan, protocol writing, concept sheet writing, inputs to data management

documents, preparing the Trial Design module which outlines the inclusion and exclusion criteria and also the visit schedule, as well as work on efficient dry run deliveries and writing of the final clinical study report. When a clinical objective is presented to a Trial Staitstician, he/she has to develop an appropriate trial design, decide on appropriate study endpoints and suggest a suitable analysis method to achieve the objectives. It is also the responsiility of the trial statistician to calculate appropriate sample size in the concept sheet to initiate the trial process. This is further documented in the sample size documentation. This is then further developed in the Protocol and then detailed in the Statistical Analysis Plan. For an interventional study, the Statistical Analysis Plan needs to be prepared before the First Patient First Visit (FPFV) date. Based on the Statistical Analysis Plan, the Trial Statistician then prepares the TFL shells which outlines all the Tables, Figures and Listing based on the requirement of a medical lead. The TFL shells play the pivotal role in the dry run delivery of the output and the final delivery of the efficacy or safety related outputs prior to the database lock, because the Trial Programmer programs the outputs based on the structure of the reports in the TFL shell prepared by the Trial Statistician. The TFL shells need to be finalized by the trial statistician no later than 6 months after the FPFV, whereas the dry run analysis outputs need to be finally reviewed by the trial statistician atleast 2 months prior to the Last Patient Last Visit (LPLV) date. In all these processes the design of the clinical trial plays a very important role as this is the process that generates the data which answers the primary and secondary objectives of a clinical research. While preparing all these regulatory documents, the Trial Statistician needs to maintain the Clinical Data Interchange Standards Consortium (CDISC) standards.

The CDISC is an open, multidisciplinary, neutral, non-profit standards developing organization (SDO) that has been working through productive, consensus-based collaborative teams, since its formation in the year 1997, to develop global standards and innovations to streamline medical research data and ensure a link with healthcare. The CDISC vision is “informing patient care and safety through higher quality medical research”. The CDISC suite of standards supports medical research of any type from protocol through analysis and reporting of results in the clinical study report.

In parallel to keeping the CDISC vision while preparing the regulatory documentation, the Trial Staitistician also tries to follow the statistical methods under research according to the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). The ICH is unique in bringing together the regulatory authorities and pharmaceutical industries to discuss scientific and technical aspects of drug registration. Since its inception in 1990, ICH has gradually evolved, to respond to the increasingly global face of drug development. ICH's mission is to achieve greater harmonisation worldwide to ensure that safe, effective, and high quality medicines are developed and registered in the most resource-efficient manner. The efficacy and safety of medicinal products should be demonstrated by clinical trials which follow the guidance in 'Good Clinical Practice: Consolidated Guideline' (ICH E6) adopted by the ICH, 1 May 1996. The role of statistics in clinical trial design and analysis is acknowledged as essential in that ICH guideline. The proliferation of statistical research in the area of clinical trials coupled with the critical role of clinical research in the drug approval process and health care in general necessitate a succinct document on statistical issues related to clinical trials. The ICH E9 document gives direction to sponsors in the design, conduct, analysis, and evaluation of clinical trials of an investigational product in the context of its overall clinical development. The document also assists scientific experts charged with preparing application summaries or assessing evidence of efficacy and safety, principally from clinical trials in later phases of development.

Recognizing the challenges for research and development and trends for productivity decline, in 2004 the US Food and Drug Administration (FDA) released the Critical Path Initiative (CPI) and in 2006 they released the Critical Path Opportunities Report (CPO) . These are two strategic documents that encourage innovation in drug development. One aspect of innovation is adaptive designs which are clinical trial designs that facilitate efficient learning from data in an ongoing trial and allow modification of certain aspects of the study according to pre-specified criteria to achieve some predetermined experimental objectives. This is where the application of sequential and adaptive randomization procedures has long been an area of active research. The FDA are now looking forward to a suitable adaptive randomization procedure which can be

fruitfully applied in real clinical trials as well as one which addresses the ethical objectives of a clinical study without compromising on the statistical efficiency of treatment comparison.

## 1.4 Adaptive Design in Clinical Trials

An adaptive design is defined as a clinical trial design that allows for prospectively planned modifications to one or more aspects of the design based on accumulating data from subjects in the trial. It can provide a variety of advantages over non-adaptive designs. These advantages arise from the fundamental property of clinical trials with an adaptive design: they allow the trial to adjust to information that was not available when the trial began. The specific nature of the advantages depends on the scientific context and types of adaptation considered, with potential advantages falling into the following major categories:

- **Statistical efficiency**: In some cases, an adaptive design can provide a greater chance to detect a true drug effect with greater statistical power than a comparable non-adaptive design.This is often true, for example, of group sequential designs and designs with adaptive modifications to the sample size. Alternatively, an adaptive design may provide the same statistical power with a smaller expected sample size or shorter expected calendar time than a comparable non-adaptive design.

- **Ethical considerations**:There are many ways in which an adaptive design can provide ethical advantages over a non-adaptive design. For example, the ability to stop a trial early if it becomes clear that the trial is unlikely to demonstrate effectiveness can reduce the number of patients exposed to the unnecessary risk of an ineffective investigational treatment and allow subjects the opportunity to explore more promising therapeutic alternatives.

- **Advantages in generalizability and improved understanding of drug effects**: An adaptive design can make it possible to answer broader questions than would normally be feasible with a non-adaptive design. For example, an adaptive enrichment design may make it possible to demonstrate effectiveness in either a given population of patients or a targeted subgroup of that population, where a non-adaptive alternative might require infeasibly large sample sizes. An adaptive design can also yield improved understanding of the effect of the experimental treatment. For example, a design with adaptive dose selection may yield better estimates of the dose-response relationship, which may also lead to more efficient subsequent trials.

- **Acceptability to stakeholders**: An adaptive design may be considered more acceptable to stakeholders than a comparable non-adaptive design because of the added flexibility. For example, sponsors may be more willing to commit to a trial that allows planned design modifications based on accumulating information. Patients may be more willing to enroll in trials that use response-adaptive randomization because these trials can increase the probability that subjects will be assigned to the more effective treatment.

**Example**: A clinical trial was conducted to evaluate Eliprodil for treatment of patients suffering from severe head injury (Bolland et al. 1998). The primary efficacy endpoint was a three category outcome defining the functional status of the patient after six months of treatment. There was considerable uncertainty at the design stage about the proportions of patients in the placebo control group who would be expected to experience each of the three different functional outcomes. An interim analysis was prespecified to update estimates of these proportions based on pooled, non-comparative data in order to potentially increase the sample size. This approach was chosen to avoid a trial with inadequate statistical power and therefore helped ensure that the trial would efficiently and reliably achieve its objective. The interim analysis ultimately led to a sample size increase from 400 to 450 patients.

Dragalin's (2006) classification of adaptive designs distinguishes four major types of adaptation:

- **Adaptive allocation rule**: change in the randomization procedure to modify the allocation proportion or the number of treatment arms.
- **Adaptive sampling rule**: change in the number of study subjects or change in study population.
- **Adaptive stopping rule**: early stopping due to efficacy, futility, or safety.
- **Adaptive decision rule**: change in the way decisions will be made about the trial (e.g., change of endpoint, change of test statistics, etc.).

The current thesis deals with *adaptive randomization designs*, i.e, the designs that fall in the first category of Dragalin's (2006) classification. A general adaptive randomization procedure is defined by specifying conditional randomization probabilities of treatment assignments as follows: Let $X_{(m+1)}$ be a treatment assignment indicator such that $X_{(m+1)} = (1, 0)$ according to the assignment of the $(m+1)^{th}$ patient to treatment $k = (A, B)$. Let $\mathcal{D}_m$ be the data structure that forms the basis for design adaptations. A general adaptive randomization procedure is defined by specifying conditional randomization probabilities of assignments of patients to a particular treatment arm eg: Arm $A$, as follows:

$$P_{(m+1),A} = P\{X_{(m+1)} = 1|\mathcal{D}_m\}, \tag{1.1}$$

Depending on the trial objectives, one can distinguish four types of adaptive randomization designs:

- **Restricted randomization**:

  This is the randomization procedure when $\mathcal{D}_m = \{\chi_1, ..., \chi_m\}$, the history of previous patients' treatment assignments. The goal is to prospectively balance treatment numbers in the trial.

- **Covariate-adaptive randomization**:

  This is the randomization procedure when $\mathcal{D}_m = \{(\chi_1, z_1), ..., (\chi_1, z_m), z_{m+1}\}$,the history of previous patients' treatment assignments and covariates, and the covariate vector of the current patient. The goal is to prospectively balance treatment assignments overall in the trial and across selected covariates.

- **Response-adaptive randomization**:

  This is the randomization procedure when $\mathcal{D}_m = \{(\chi_1, \upsilon_1), ..., (\chi_m, \upsilon_m)\}$, the history of previous patients treatment assignments and responses. The most common goal is to increase the chance for a patient to be assigned to a potentially better treatment. Other possible goals may include increasing estimation efficiency of the desired treatment effect or maximizing the power of a statistical test.

- **Covariate-adjusted response-adaptive randomization**

  This is the randomization procedure when $\mathcal{D}_m = \{(\chi_1, z_1, \upsilon_1), ..., (\chi_m, z_m, \upsilon_m), z_{m+1}\}$, the history of previous patients treatment assignments, responses and covariates, and the covariate vector of the current patient. The most common goal is to increase the chance for a patient to be assigned to a potentially better treatment given the patients covariate profile while maintaining the power of a statistical test.

The class of adaptive randomization procedures can be extended further by including adaptive designs with treatment selection for which randomization probabilities for some treatment arms can be set to 0 throughout the trial.

This thesis is dedicated towards discussing the development of Covariate-adjusted response-adaptive randomization designs when the response of patients to a particular treatment arm follow a survival model. Very limited discussion can be found in the literature about the development of Covariate-adjusted response-adaptive randomization procedures for survival trials. This is because considering survival responses does not allow the priviledge of handling independent and identically distributed observations.

## 1.5 Survival Responses

Survival responses are concerned with obsevations relating to times to some critical event, starting from some specific time origin. For example, oncology studies are often concerned with cancer survival from treatment to death. Here a major objective is to compare the effectiveness of treatments allowing for the effects of explanatory variables or factors. Random variables representing time to event outcomes do not always record a random sample of completed observations. For example, when the survival of patients following treatment for cancer is being measured, starting in each case with the date of treatment, some patients who live for a long time, do not have their completed survival times at the end of the study period. Omitting such observations from the analysis would clearly introduce a serious bias into the estimation of how long such patients do in fact survive. Such observations are said to be right-censored and their inclusion in the analysis in appropriate form is crucial.

There are three different forms of censoring: right-censoring, left-censoring and interval-censoring; truncation is a somewhat different phenomenon.

- **Right-Censoring** : This is the most frequently encountered type, at least in medical applications. Some subjects leave the study before the event occurs, so one only knows that their survival time $X$ lies in an interval $(t, \infty)$. If $X$ is the survival time, and $C$ is the time until the subject left the study, then in the case of right-censoring for the $i^{th}$ subject, $T_i = min(X_i; C_i)$ is only observed.

- **Left-Censoring** : This is much more rarely encountered. Sometimes some subjects have experienced the event before detailed observation commences. Thus their time to the event lies in the interval $[0, T)$. It is possible for both right and left censoring to occur in the same dataset.

- **Interval-Censoring** :Interval-censoring occurs when it is not clear when the event occurred. All that is known is that the time to event occurred within some interval $(t_1, t_2]$. This form of censoring often arises when observations are infrequent, and the event has occurred between two observation times.

- **Left-Truncation** : In left-truncation subjects enter the study at times after the origin for the events of interest and are followed up until the event of interest occurs or they are lost to followup. Subjects only contribute to the likelihood once they have entered the study, which will be at particular times greater than zero, even on their personal time origin bases. If they experience the event before their observation period starts they will not appear in the study.

- **Right-Truncation** : Right-truncation refers to studies where subjects are only observed if they experience the event. Studies based on death records are an example of this.

Throughout this thesis Covariate-adjusted response-adaptive (CARA) designs are developed for survival responses which are right-censored, assuming independence between survival times and censored times. Numerous real-life clinical trials, specially in the oncology theraputic area, deal with survival endpoints. However there has been very limited discussion about developing CARA designs for such trials. Sverdlov, Rosenberger and Ryzenik (2013) discussed development of CARA designs for exponential survival responses which are right-censored. Biswas, Bhattacharya and Park (2016) considered informative random censoring while developing CARA designs for exponential survival responses. An attempt has been made throughout this thesis to develop CARA designs further in order to enhance their applicability in real-life survival trials when observations are right-censored.

## 1.6 Outline of the Thesis

Chapter 2 of this thesis enhances the scope of applications of CARA designs beyond exponential survival responses, by considering Weibull distributed survival responses. In real clinical trials however the response of patients to a treatment infrequently follows a paramertic model, the scope of application of CARA designs is enhanced further in Chapter 3 by relaxing any parametric distributional assumption for survival responses

but considering the hazard of patients at a given time-point to be proportional. Often in clinical trials especially in the oncology therapeutic area, people are interested in survival endproints where events occur due to multiple causes. Trial protocols considering progression free survival as the primary endpoint often encounters such scenario. CARA designs are developed in Chapter 4 for such specific type of survival trials. This is followed in Chapter 5 by detailing the summary of the research findings in this thesis along with its critical evaluation and suggestion on few areas in which CARA designs can be developed further considering survival responses of patients who are arriving sequentially in the trial.

# Chapter 2

# Covariate-Adjusted Response-Adaptive Designs for Weibull Survival Models

## 2.1 Introduction

There is great interest in the possibility that clinical trials can be designed with adaptive features that may make the studies more efficient. An adaptively randomized clinical trial evaluates a treatment by observing patient responses on a prescribed schedule and modifies the parameters of the trial protocol in accordance with those observations. The adaptation process generally continues throughout the trial. Clinical trials are often designed with adaptive features to force balance in the sequential allocation of patients across two or more competing treatment arms. It is also used to force imbalance by allocating a greater number of study subjects to the better-performing treatment arm.

A clinical trial is a complex experiment on humans with multiple and often competing experimental objectives. Here, several treatments for a disease are compared with the purpose of obtaining information on their performance. Since human patients are involved, there is an ethical concern to treat as many patients as possible with the best

treatment. At the same time, there must be some allocation of patients to the worse treatment arm for making useful statistical inferences about treatment comparisons.

An increase in the number of patients receiving better treatments leads to sequential experiments in which data are analysed and new allocations are made in the light of the estimated parameters. However, the advocates of traditional balanced randomized designs argue that having a balanced allocation of patients across the treatment arms helps estimate treatment effects efficiently. Since clinical trials involve human patients, balanced allocation can be a serious problem as one would be more inclined to be treated with the better treatment and such balanced allocation leads to almost one half of the patients on the worse treatment arm. To balance these competing goals of ethics and statistical efficiency in a clinical trial, response-adaptive designs have been developed and used. A response-adaptive design uses the available response and treatment allocation histories to skew the treatment allocation probabilities in favour of the treatment arm found best at an interim stage in the trial. However, human patients are heterogeneous and therefore one needs to take into account such concomitant information when allocating a particular patient to a treatment arm.

Covariate-adjusted response-adaptive designs balance the competing goals of assigning a greater number of study subjects to the better treatment and achieving high statistical efficiency in estimating treatment effects. This is done in the presence of covariates, while maintaining randomness in treatment assignments. Investigators are often aware of important baseline covariates that may have a strong influence on patient responses and they may wish to adjust the randomization procedure for these covariates. Rosenberger and Sverdlov (2008) gave an overview of different techniques for handling covariates in the design of clinical trials and distinguished between two main approaches. These are covariate-adaptive randomization and covariate-adjusted response-adaptive randomization (CARA) procedures. CARA randomization is applicable to clinical trials where non-linear and heteroscedastic models determine the relationship between responses, treatments and covariates, and when multiple experimental objectives are pursued in the trial. The goal of a CARA procedure in a phaze III clinical trial may be to skew allocation in the direction of the better performing treatment arm, for a

given patient's covariate profile, while maintaining the power of a statistical test for treatment comparisons. These designs rely on correctly specified parametric models. Although the exponential model for survival responses has been previously considered by Sverdlov, Rosenberger and Ryzenik (2013) for developing CARA designs, to extend the scope of application of such designs in real-life clinical trials, the Weibull survival model has been considered here, which includes the exponential model as a special case. The memoryless property of the exponential survival model due to its constant hazard property limits its application in real-life clinical trials involving humans. Therefore to enhance the scope of the application of such designs a step further, methods are developed here considering survival responses following a Weibull model. The shape parameter for the Weibull survival model determines the shape of the density function as well as the hazard function of the responses. The exponential model is a special case when the shape parameter is unity which makes the hazard function of the time to an event a constant. A CARA design based on the Weibull model for the survival responses of the patients therefore extends the scope of application of such designs to scenarios beyond constant hazard. Extensive simulation study of the operating characteristics suggest that the proposed CARA procedure can certainly be considered as a suitable alternative to the traditional balanced randomization designs in survival trials, provided that sufficient responses are available during the interim stage of the trial to enable adaptations in the design.

The outline of this chapter is as follows. Section 2.2 explains the basic background relating to the Weibull survival model . The discussion in Section 2.3 focuses on the idea of obtaining parameter estimates using the maximum likelihood approach when survival times conform to a Weibull model and the right-censored times are independent of the event times. This is followed by Section 2.4 that proposes the various target allocation proportions to a treatment arm, for Weibull distriuted survival times and the CARA randomization procedures to achieve the derived allocation proportions. The derivation of the asymptotic properties of the CARA designs using a Taylor series expansion of the non-centrality parameter for the Wald test is detailed in Section 2.5. The findings of Section 2.4 have been validated using extensive simulations in Section 2.6. In Section 2.7, the applicability of the proposed CARA designs is further explained

through the results obtained from applying them to redesign a real-life clinical trial. A critical evaluation of some of the derived CARA designs for Weibull distributed survival responses is provided in Section 2.8. Section 2.9 concludes with a discussion of the overall findings and an outline of some future research in this direction.

## 2.2 The Weibull Model

The exponential model used by Sverdlov, Rosenberger and Ryeznik (2013) to develop CARA randomization procedures for survival trials is often referred to as a purely random failure pattern. It is useful for its mathematical simplicity, but the fact that the exponential distribution corresponds to a lack of memory model and that it has only one parameter make it less likely to fit data in practice.

Let $T$ be a non-negative random variable representing the time to a critical event. The lack of memory property of the exponential distribution means that $P(T \geqslant t + z|T \geqslant t) = P(T \geqslant z)$, which leads to its mathematical tractability but also reduces its applicability to many realistic applied situations. It is because of this distributional property that $E(T - t|T \geq t) = E(T) = \mu$; that is, the mean residual life or the expected remaining lifetime for a patient is constant at any point in the study. This shows that the time until the future occurrence of an event does not depend upon past history, and therefore this property is sometimes called the "no-aging" property. This property is also reflected in the exponential distribution's constant hazard rate which is independent of $t$. Although the exponential distribution has been historically very popular, its constant hazard rate appears to make it too restrictive in both health and industrial applications.

The Weibull distribution can be regarded as a generalization of the exponential, with an extra shape parameter, $\gamma > 0$ . Being the only parametric survival distribution which has both a proportional hazards representation and an accelerated life representation, it is the most popular distribution used in reliability and survival analysis. Let $x_i$ denote the treatment indicator for the $i^{th}$ patient such that $x_i = 1$ if the patient is assigned to treatment $A$, and $x_i = 0$ if the patient is assigned to treatment $B$. Associated with

patient $i = 1, 2, ..., n$ is a vector of baseline covariates $\mathbf{z}_i = (1, z_{1i}, ...., z_{pi})^T$. Let $\gamma_k$ be the shape parameter that determines the shape of the distribution of the responses for treatment $k$ as well as that of the hazard function, $\boldsymbol{\beta}_k = (\beta_{k0}, \beta_{k1}, ...., \beta_{kp})^T$ be the vector of unknown model parameters and $\mu_k(\mathbf{z}_i) = \exp(\boldsymbol{\beta}_k^T \mathbf{z}_i) > 0$, the scale parameter for treatment $k$ given the $i^{th}$ covariate $\mathbf{z}_i$. Conditional on $\mathbf{z}_i$, patient's survival time on treatment $k$, $T_{ik}$, follows a Weibull distribution with probability density function

$$f_k(t|\mathbf{z}_i; \boldsymbol{\beta}_k, \gamma_k) = \left\{ \frac{\gamma_k}{\mu_k(\mathbf{z}_i)} \right\} \left\{ \frac{t}{\mu_k(\mathbf{z}_i)} \right\}^{\gamma_k - 1} \exp \left[ - \left\{ \frac{t}{\mu_k(\mathbf{z}_i)} \right\}^{\gamma_k} \right] \tag{2.1}$$

for $t > 0$.

Let $S_k(t|\mathbf{z}_i) = \mathrm{P}_k(T \geq \mathrm{t} \ |\mathbf{z}_i)$ be the survivor function measuring the probability of a patient with covariate $\mathbf{z}_i$ at treatment $k$ surviving beyond a given time point $t$. The survivor function of the Weibull model is $S_k(t|\mathbf{z}_i; \boldsymbol{\beta}_k, \gamma_k) = \exp\left[-\{t/\mu_k(\mathbf{z}_i)\}\right]^{\gamma_k}$. Let $h_k(t|\mathbf{z}_i) = \frac{f_k(t|\mathbf{z}_i)}{S_k(t|\mathbf{z}_i)}$ be the hazard function representing the instantaneous failure rate of a subject with covariate $\mathbf{z}_i$ at time $t$. Therefore, the hazard function for the Weibull model can be expressed as $h_k(t|\mathbf{z}_i; \boldsymbol{\beta}_k, \gamma_k) = \{\gamma_k/\mu_k(\mathbf{z}_i)\}\{t/\mu_k(\mathbf{z}_i)\}^{\gamma_k - 1}$. It is also known that $S_k(t|\mathbf{z}_i; \boldsymbol{\beta}_k, \gamma_k) = \exp\{-\int_0^t h_k(u|\mathbf{z}_i; \boldsymbol{\beta}_k, \gamma_k) du\}$ where,

$$\int_0^t h_k(u|\mathbf{z}_i; \boldsymbol{\beta}_k, \gamma_k) du = H_k(t|\mathbf{z}_i; \boldsymbol{\beta}_k, \gamma_k) = -\log\{S_k(t|\mathbf{z}_i; \boldsymbol{\beta}_k, \gamma_k)\}$$

is called the cumulative hazard or the integrated hazard. The cumulative hazard for the Weibull distribution therefore is $H_k(t|\mathbf{z}_i; \boldsymbol{\beta}_k, \gamma_k) = \{t/\mu_k(\mathbf{z}_i)\}^{\gamma_k}$. The hazard function for a Weibull model is always monotonic, but can be increasing if $\gamma_k > 1$, decreasing if $0 < \gamma_k < 1$ or constant if $\gamma_k = 1$. Therefore, the Weibull distribution with shape parameter $\gamma = 1$ corresponds to the exponential distribution. The mean and the median survival times for the Weibull model are $\lambda_k(\mathbf{z}_i) = \mu_k(\mathbf{z}_i)\Gamma\{(\gamma_k + 1)/\gamma_k\}$ and $\phi_k(\mathbf{z}_i) = \{\log(2)\}^{1/\gamma_k}\mu_k(\mathbf{z}_i)$ respectively, where $\Gamma$ denotes the gamma function.

When comparing the survival experience of two arms, if an additional assumption is made about the shape parameter $\gamma_k$ being constant between the arms, then the hazard at any given time for an individual in one treatment group is proportional to the hazard at that time for an individual from the other treatment group. The hazard ratio $\psi$ of the two treatment groups satisfies

$$h_A(t|\mathbf{z}_i; \boldsymbol{\beta}_A, \gamma) = \psi h_B(t|\mathbf{z}_i; \boldsymbol{\beta}_B, \gamma). \tag{2.2}$$

Here, $\psi$ does not depend on $t$ and the responses for the two treatment groups follow Weibull distributions with the same shape parameter $\gamma$. The assumption of the response following a Weibull model with equal shape parameter can be checked by plotting the log cumulative hazard function marginalized over the covariate profile $\mathbf{z}_i$ of the patients, against the log time for both treatment arms. If the plot yields parallel lines then the assumption of a constant $\gamma$ between the treatment groups can be considered to have been satisfied. The intercept of the log cumulative hazard plot for patients receiving treatment $k$ against the log time is $-\gamma \log(\mu_k)$. If the survival times conform to a Weibull distribution with a constant shape parameter, the hazard function for patients at the two treatment arms will be

$$\psi = \frac{h_A(t|\boldsymbol{\beta}_A, \gamma)}{h_B(t|\boldsymbol{\beta}_B, \gamma)} = \left\{ \frac{\mu_B}{\mu_A} \right\}^{-\gamma} \tag{2.3}$$

and

$$\log(\psi) = -\gamma \log \left\{ \frac{\mu_B}{\mu_A} \right\}. \tag{2.4}$$

This shows that the hazard ratio between the treatment arms is constant if the responses of the patients to the treatments follow a Weibull distribution with the same shape parameter. It can also be seen from (2.4) that a crude estimate of the hazard ratio can be obtained by exponentiating the difference between the intercepts when the log cumulative hazard function for each treatment arm is plotted against log time. This model is called the Weibull proportional hazards model.

It is sometimes useful to work with the logarithm of the survival times. If $Y_{ik} = \log(T_{ik})$, where $T_{ik}$ follows a Weibull distribution, then $Y_{ik}$ has a Gumbel distribution with density function given by

$$f_k(y|\mathbf{z}_i; \boldsymbol{\beta}_k, \gamma_k) = \gamma_k \exp\left([y - \log\{\mu_k(\mathbf{z}_i)\}\gamma_k]\right) \exp(-e^{[y - \log\{\mu_k(\mathbf{z}_i)\}\gamma_k]}), \tag{2.5}$$

for $-\infty < y < \infty$. This distribution will frequently be referred to in this chapter while constructing the likelihood for the Weibull model and the CARA randomization procedure.

## 2.3 Construction of the Likelihood

In survival trials, observations are likely to be censored. Therefore in such trials one cannot observe a random sample of complete independent and identically distributed random variables. A critical assumption is that the survival times and the censoring times are independent. For construction of a likelihood function for censored data, one needs to carefully consider the information conveyed by each observation. Right censoring is the most frequently encountered type of censoring, which could happen when the patient followed does not experience the event or when the study is stopped before the event occurred. For example, when the data are measured from different start times, some patients who are recruited late to the study are likely to be alive at the time of analyzing the data. Therefore, the survival times for these patients would be unavailable. All that would be known about the survival times for these patients is that they exceed a certain time.When some subjects are right censored, all that is known is that their survival time lies in an interval $(t, \infty)$. Let $T_{ik}$ be the random variable representing the survival time for the $i^{th}$ patient with treatment $k$ and $C_{ik}$ be the one representing the censoring time for that patient.The survival times are assumed to be independently and identically distributed. For the $i^{th}$ patient, the observed outcome is a bivariate random vector $(\mathcal{T}_{ik},\delta_{ik})$, where $\mathcal{T}_{ik} = \min(T_{ik},C_{ik})$ and

$$\delta_{ik} = \begin{cases} 0 \text{ if } \mathcal{T}_{ik} \text{ is a right-censored time,} \\ 1 \text{ if an event occured at time } \mathcal{T}_{ik}. \end{cases} \tag{2.6}$$

When subjects join a study at different times and are all observed until a fixed moment in time, we have generalized type I right censoring. Here time is measured from a different origin for each subject, who has an individual specific fixed censoring time. Figure 2.1 shows this for five patients.

For right-censored observations, the event time is larger than the observed time. So the information is the survivor function evaluated at the on study time. It can be seen from Figure 2.1 above that each of the five patients in the trial enter the study at different times but are observed for a predetermined study period. The blank circles

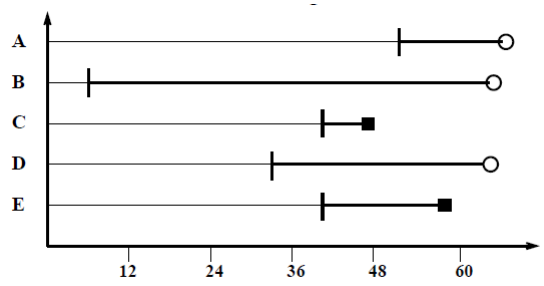


Figure 2.1: Generalized type I right censoring

in the figure represents censoring observations and the dark squares represent patients who have experienced the event. The patients who enter the study late are more likely to be censored. Censoring in such cases may occur if the patients did not experience the event until the end of the study period or if they have been lost to follow-up before the end of the study period. It can be seen from Figure 2.1 that patients A survive until the end of the study period and therefore their actual survival times $T_{ik}$ are not known. All that is known is that they survive beyond a certain time period. Also the time to event for patient B and patient D is censored as he/she has been lost to follow-up before the end of the study period. Therefore, $\mathcal{T}_{ik} = C_{ik}$ for patients A,B and D. However, for patients C and E, it can be seen that they have experienced the event and therefore their actual event times $\mathcal{T}_{ik} = T_{ik}$ are known.

In this chapter, a CARA randomization procedure has been developed for survival responses where the event times follow a Weibull model. Since patients arrive sequentially in the clinical trial and are observed until the end of the trial, the type of censoring considered here is generalized type I right censoring. With such a model, the outcome for the $i^{th}$ patient is given by $(t_{iA},\delta_{iA})$ if $x_i = 1$ or $(t_{iB},\delta_{iB})$ if $x_i = 0$. Any of the complications of the data such as censoring or truncation will affect the likelihood as this leads to a data which is not indenpendent and identically distributed. The likelihood may be written down by incorporating the exact event times through the density function $f_k(t_{ik}|\mathbf{z}_i;\boldsymbol{\beta}_k,\gamma_k)$ and the right censored observations through the survivor function $S_k(c_{ik}|\mathbf{z}_i;\boldsymbol{\beta}_k,\gamma_k)$. Here, $c_{ik}$ is a realization of the random variable $C_{ik}$ and censoring is assumed to be independent of the survival. Note that the functions are being adjusted for the effects $\boldsymbol{\beta}_k$ of the covariates $\mathbf{z}_i$ and also for the shape parameter $\gamma_k$ of the Weibull

model which scales the theoretical errors of the accelerated life model.

For the survival times from the patients on treatment $k$, the likelihood function is,

$$L_k = \prod_{\delta_{ik}=1} f_k(t_{ik}|\mathbf{z}_i;\boldsymbol{\beta}_k,\gamma_k) \prod_{\delta_{ik}=0} S_k(c_{ik}|\mathbf{z}_i;\boldsymbol{\beta}_k,\gamma_k) \tag{2.7}$$

Now, it can be seen that, for $\delta_{ik} = 0$,

$$P(\mathcal{T}_{ik} = c_{ik}, \delta_{ik} = 0) = P(\mathcal{T}_{ik} = c_{ik}|\delta_{ik} = 0)P(\delta_{ik} = 0).$$

Therefore, this can also be written as

$$P(\mathcal{T}_{ik} = c_{ik}, \delta_{ik} = 0) = P(\delta_{ik} = 0) = P(T_{ik} > c_{ik}),$$

which is the survivor function at time $c_{ik}$. Similarly for $\delta_{ik} = 1$,

$$P(\mathcal{T}_{ik} = t_{ik}, \delta_{ik} = 1) = P(\mathcal{T}_{ik} = t_{ik}|T_{ik} < c_{ik})P(T_{ik} < c_{ik}),$$

which can also be written as

$$\frac{f_k(t_{ik}|\mathbf{z}_i;\boldsymbol{\beta}_k,\gamma_k)}{1 - S_k(c_{ik}|\mathbf{z}_i;\boldsymbol{\beta}_k,\gamma_k)}\{1 - S_k(c_{ik}|\mathbf{z}_i;\boldsymbol{\beta}_k,\gamma_k)\} = f_k(t_{ik}|\mathbf{z}_i;\boldsymbol{\beta}_k,\gamma_k).$$

Thereofore, for a random sample of pairs ($t_{ik}$,$\delta_{ik}$) of $n_k$ patients, the likelihood function is given by

$$L_k = \prod_{i=1}^{n_k}\{f_k(t_{ik}|\mathbf{z}_i;\boldsymbol{\beta}_k,\gamma_k)\}^{\delta_{ik}}\{S_k(t_{ik}|\mathbf{z}_i;\boldsymbol{\beta}_k,\gamma_k)\}^{1-\delta_{ik}}. \tag{2.8}$$

It is well known from the distribution theory of survival analysis that $f_k(t_{ik}|\mathbf{z}_i;\boldsymbol{\beta}_k,\gamma_k) = S_k(t_{ik}|\mathbf{z}_i;\boldsymbol{\beta}_k,\gamma_k)h_k(t_{ik}|\mathbf{z}_i;\boldsymbol{\beta}_k,\gamma_k)$. Thus, (2.8) can be further simplified to be

$$L_k = \prod_{i=1}^{n_k}\{h_k(t_{ik}|\mathbf{z}_i;\boldsymbol{\beta}_k,\gamma_k)\}^{\delta_{ik}} S_k(t_{ik}|\mathbf{z}_i;\boldsymbol{\beta}_k,\gamma_k), \tag{2.9}$$

so that the combined likelihood function is

$$L = \prod_{k=A}^{B}\prod_{i=1}^{n}\{h_k(t_{ik}|\mathbf{z}_i;\boldsymbol{\beta}_k,\gamma_k)\}^{\delta_{ik}} S_k(t_{ik}|\mathbf{z}_i;\boldsymbol{\beta}_k,\gamma_k).$$

For a random sample of observations $(t_{ik}, \delta_{ik})$, from the Weibull distribution with independent right censoring, let $y_{ik} = \log(t_{ik})$. It was seen earlier that the density of $Y_{ik}$ is given by (2.5). Therefore, the likelihood function in this case is

$$L_k = \prod_{i=1}^{n_k}(\gamma_k\ e^{[y_{ik}-\log\{\mu_k(\mathbf{z}_i)\})\gamma_k]})^{\delta_{ik}} e^{-\exp[y_{ik}-\log\{\mu_k(\mathbf{z}_i)\}\gamma_k]}. \tag{2.10}$$

So the log-likelihood function here is given by

$$l_k = -\log(\gamma_k)\sum_{i=1}^{n_k}\delta_{ik} + \sum_{i=1}^{n_k}(\delta_{ik}[y_{ik} - \log\{\mu_k(\mathbf{z}_i)\}\gamma_k] - e^{[y_{ik}-\log\{\mu_k(\mathbf{z}_i)\}\gamma_k]}). \tag{2.11}$$

Maximising the log-likelihood in (2.11) separately for $1/\gamma_k$ and $\log\{\mu_k(\mathbf{z}_i)\}$, the maximum likelihood estimators for these model parameters can be obtained, which in turn give the maximum likelihood estimators of the regression parameters. This is because, for the Weibull accelerated life model, the linear predictor $\boldsymbol{\beta}_k^T\mathbf{z}_i$ is considered to be identical to $\log\{\mu_k(\mathbf{z}_i)\}$ and not $\mu_k(\mathbf{z}_i)$. Also the theroretical errors follow a two parameter Gumbel distribution that is scaled by the factor $1/\gamma_k$. The first partial derivatives of the log likelihood function (2.11) with respect to $\log\{\mu_k(\mathbf{z}_i)\}$ and $1/\gamma_k$ are

$$\frac{\partial l_k}{\partial \log\{\mu_k(\mathbf{z}_i))\}} = \sum_{i=1}^{n}\left(\delta_{ik} - e^{[y_{ik}-\log\{\mu_k(\mathbf{z}_i)\}\gamma_k]}\right)(-\gamma_k), \tag{2.12}$$

and,

$$\frac{\partial l_k}{\partial(1/\gamma_k)} = -\left(\sum_{i=1}^{n_k}\delta_{ik}\gamma_k + \sum_{i=1}^{n_k}(\delta_{ik} - e^{[y_{ik}-\log\{\mu_k(\mathbf{z}_i)\}\gamma_k]})\right)[y_{ik} - \log\{\mu_k(\mathbf{z}_i)\}]\gamma_k^2. \tag{2.13}$$

Equating the score functions in (2.12) and (2.13) to zero and solving for $1/\gamma_k$ and $\log\{\mu_k(\mathbf{z}_i)\}$ gives the maximum likelihood estimates for $1/\gamma_k$ and $\log\{\mu_k(\mathbf{z}_i)\}$. The variances of the estimated parameters can be obtained by calculating the second patial derivatives of the log-likelihood function (2.11) with respect to $1/\gamma_k$ and $\log\{\mu_k(\mathbf{z}_i)\}$ . The details are in Appendix A.

An important practical feature of Weibull accelerated life models is that they can be fitted to data using an algorithm which is a form of iteratively re-weighted least squares. The maximum likelihood estimate of $\boldsymbol{\beta}_k$ is obtained by this method. The maximum likelihood estimates of $\log\{\mu_k(\mathbf{z}_i)\}$ and $1/\gamma_k$ are also obtained by equating (2.12) and (2.13) to zero and numerically solving them using the method of iteratively re-weighted least squares. McCullagh and Nelder (1989) proved that this algorithm is equivalent to the Fisher scoring method and leads to the maximum likelihood estimates.

Given an initial trial estimate $\hat{\beta}_k$, the estimated linear predictor $\hat{\eta}_{ik} = \hat{\boldsymbol{\beta}}_k^T\mathbf{z}_i$ is calculated. This is used to obtain $\hat{\mu}_k(\mathbf{z}_i) = r^{-1}(\hat{\eta}_{ik})$, where $r(.)$ is the necessary one-to-one

continuous link function selected on the basis of the distribution of the responses. In the Weibull regression model, $r(.)$ is the natural log link function. These quantities are further used to calculate the working dependent variable

$$b_i = \hat{\eta}_{ik} + \{y_{ik} - \hat{\mu}_k(\mathbf{z}_i)\}\frac{\partial \hat{\eta}_{ik}}{\partial \hat{\mu}_k(\mathbf{z}_i)},$$

where the rightmost term is the derivative of the link function evaluated at the trial estimate. Next, the iterative weights are calculated. The $i^{th}$ weight is inversely proportional to the variance of the working dependent variable $b_i$, given the current estimates of the parameters. The iterative weights here are given by

$$\epsilon_{ik}(\boldsymbol{\beta}_k, \gamma_k, \mathbf{z}_i) = P(T_{ik} \leq C_i | \mathbf{z}_i; \boldsymbol{\beta}_k, \gamma_k), \tag{2.14}$$

for $i = 1, 2, ..., n_k$ , and $k = A, B$.

For the $i^{th}$ patient, (2.14) is the probability of observing an event before censoring conditional on $\boldsymbol{\beta}_k$, $\gamma_k$ and the covariates for that patient.This quantity will depend on the censoring mechanism in the trial. Finally, an improved estimate of $\boldsymbol{\beta}_k$ is obtained by regressing the working dependent variable $\mathbf{b}_i$ on the covariates $\mathbf{z}_i$ using the weights. This means that a weighted least squares estimate is approximated using

$$\hat{\boldsymbol{\beta}}_k = (Z_{n_k}^T W_k Z_{n_k})^{-1} Z_{n_k}^T W_k \mathbf{b}, \tag{2.15}$$

where $\mathbf{b}$ is a response vector with entries given by the working dependent variable $b_i$, $Z_{n_k}$ is the model matrix for the patients on treatment $k$, $W_k$ is an $n_k \times n_k$ diagonal matrix with the $i^{th}$ diagonal element as $\epsilon_{ik}(\mathbf{z}_i; \boldsymbol{\beta}_k, \gamma_k)$ given in (2.14) and $Z_{n_k}^T W_k Z_{n_k}$ is the approximate Fisher Information matrix for $\boldsymbol{\beta}_k$ which is scaled by $\gamma_k$. The procedure is repeated until successive estimates change by less than a specified small amount.

## 2.4 The Proposed Allocation Designs

The main goal of a CARA procedure is to use the accumulated data to skew the treatment allocation probabilities in favour of the treatment that is most efficacious

for a given patient. This is done for ethical considerations and also for enhancing the efficiency in estimating the treatment effects. This represents an advancement in the field of personalized medicine or theranostics, which is a medical model that separates people into different groups based on medical decisions, practices, interventions and/or products being tailored to the individual patient after taking into account the information related to their predicted response or risk of disease. The CARA randomization procedure is applicable when the responses of the patients to the treatment follow a non-linear and heteroscedastic model and when multiple experimental objectives are being pursued in the clinical trial. It relies on correctly specified parametric models.

The censoring scheme assumed throughout this chapter is a combination of uniform and generalized type I censoring. This has been described as administrative censoring in Latta (1981). Patient arrival times follow an independent uniform distribution on (0,$R$). The clinical trial has a recruitment period of length $R > 0$. The survival times of the patients follow a Weibull distribution with parameters $\mu_k(\mathbf{z}_i)$ and $\gamma_k$, and the right censored times, $C_{ik}$, are uniformly distributed over (0,$D$), where $D$ is the duration of the clinical trial. At time $D$, the subjects who have not experienced an event or have not yet been lost to follow-up are considered to be generalized right censored of type I. Such assumptions can be considered to be reasonable in a real-life clinical trial. In the subset of survival trials where the recruitment phase is long enough to accumulate a substantial amount of response data, the CARA randomization procedure is applicable.

The $\boldsymbol{\beta}_A$ and the $\boldsymbol{\beta}_B$ are population characteristics representing the covariate adjusted treatment effects of $A$ and $B$, respectively. During the initial phase of the trial, one uses some balanced randomization procedure to allocate the initial $2m_0$ patients equally among treatments $A$ and $B$, where $m_0$ is a positive integer. This ensures that at least $m_0$ patients are allocated to each treatment arm, and $m_0$ is chosen so that estimates of the parameters ($\boldsymbol{\beta}_A$ , $\boldsymbol{\beta}_B$) can be obtained from this initial sample. At stage $m = 2m_0$, one computes the maximum likelihood estimates ($\hat{\boldsymbol{\beta}}_{A,m}$, $\hat{\boldsymbol{\beta}}_{B,m}$) based on the responses of the first $2m_0$ patients, eliminating the effects of the prognostic factors. At stage $m \geq 2m_0 + 1$, when the $(m+1)^{\text{th}}$ patient enters the clinical trial with covariate vector $\mathbf{z}_{m+1}$, this patient is randomized to treatment $A$ with probability c($\hat{\boldsymbol{\beta}}_{A,m}$, $\hat{\boldsymbol{\beta}}_{B,m}$,$\hat{\gamma}$, $\mathbf{z}_{m+1}$) where

$0 \leq c(.) \leq 1$ is an allocation function which bridges the past allocation pattern, response histories and the covariate vectors of the $m$ patients through the sufficient statistics which the fittel model is a function of, to the $(m+1)^{\text{th}}$ allocation with the covariate vector $\mathbf{z}_{m+1}$. This allocation is chosen with the intention of skewing the treatment allocation probability in favour of the better treatment arm. This section discusses the method of derivation of this allocation function using two different approaches. One of these uses two different types of allocation function to target a derived optimal allocation proportion with the aim of achieving the ethical objective in a clinical trial while keeping the asymptotic variance of the treatment difference to be constant, and the other one uses a link function to skew the treatment allocation probabilities in favour of the better treatment arm after adjusting for the effects of the covariate profiles of the patients.

### 2.4.1 CARA Design using the Optimal Allocation Approach

One of the approaches to develop CARA randomization procedures is to derive the optimal allocation proportion for a model without covariates and use its covariate-adjusted version for the sequential allocation of patients (Zhang *et al.*, 2007). For a clinical trial, where the survival times of the patients are right censored and follow a Weibull distribution, the optimal allocation proportion can be found using the maximum likelihood approach presented in Section 2.3. Now, the variance of the logarithm of the estimated Weibull scale parameter is shown, in Appendix A to be

$$\sigma_k^2 = \text{var}[\log\{\hat{\mu}_k(\mathbf{z})\}] = \frac{G_k}{n_k \gamma_k^2}$$

for $k = A, B$, where $G_k$ is defined in (2.16) and $n_k$ denotes the number of patients on treatment $k$. Let $\varsigma_k = \gamma_k[Y_k - \log\{\mu_k(\mathbf{z})\}]$ and $\epsilon_k(\mathbf{z}; \boldsymbol{\beta}_k, \gamma_k)$ be the probability of an event, where $Y_k = \log(T_k)$ is the logarithm of the survival time on treatment $k$. Then

$$G_k = \frac{\epsilon_k(\mathbf{z}; \boldsymbol{\beta}_k, \gamma_k) + E(\varsigma_k^2 e^{\varsigma_k})}{\epsilon_k^2(\mathbf{z}; \boldsymbol{\beta}_k, \gamma_k) + \epsilon_k(\mathbf{z}; \boldsymbol{\beta}_k, \gamma_k) E(\varsigma_k^2 e^{\varsigma_k}) - \{E(\varsigma_k e^{\varsigma_k})\}^2} \tag{2.16}$$

where $\epsilon_k(\mathbf{z}; \boldsymbol{\beta}_k, \gamma_k) = P(T_k \leq C | \mathbf{z}; \boldsymbol{\beta}_k, \gamma_k)$. The derivation of a formal analytical form for $\epsilon_k(\mathbf{z}; \boldsymbol{\beta}_k, \gamma_k)$ is shown in Appendix B.

Zhang and Rosenberger (2007) derived an optimal allocation proportion by minimizing the average hazard in the trial subject to the constraint that the asymptotic variance of the difference in the estimated scale parameters is a constant. This would ensure that for any choice of the number of patients allocated to each treatment arm, the power for testing the significance of the difference in the treatment effects would remain fixed. For a survival trial with two treatment arms of overall duration $D$, if $\mu_k > 0$ is the scale parameter and $\gamma_k > 0$ is the shape parameter of a Weibull distribution, then the cumulative hazard at time $D$ will be $(D/\mu_k)^{\gamma_k}$. Therefore, minimizing the total cumulative hazard $n_A(D/\mu_A)^{\gamma_A} + n_B(D/\mu_B)^{\gamma_B}$ for the two treatments subject to the constraint that the asymptotic variance $G_A/(n_A\gamma_A^2)\} + G_B/(n_B\gamma_B^2)$ is constant, the optimal allocation proportion for treatment $A$ is given by

$$\varrho_{A0} = \frac{\gamma_B\sqrt{\{D/(\mu_B)\}^{\gamma_B}G_A}}{\gamma_B\sqrt{\{D/(\mu_B)\}^{\gamma_B}G_A} + \gamma_A\sqrt{\{D/(\mu_A)\}^{\gamma_A}G_B}}. \tag{2.17}$$

The corresponding covariate-adjusted target allocation proportion is therefore expressed as

$$\pi_{A_0}^W(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B, \gamma_A, \gamma_B, \mathbf{z}) = \frac{\gamma_B\sqrt{\{D/\mu_B(\mathbf{z})\}^{\gamma_B}G_A}}{\gamma_B\sqrt{\{D/\mu_B(\mathbf{z})\}^{\gamma_B}G_A} + \gamma_A\sqrt{\{D/\mu_A(\mathbf{z})\}^{\gamma_A}G_B}}. \tag{2.18}$$

Its derivation is detailed in Appendix D. An important feature of (2.18) is that, irrespective of the value of $\mathbf{z}$, the allocations are skewed towards the better treatment arm, but the degree of skewing is dependent on the parameters of the Weibull distribution.

Other functions could also be minimized to obtain different optimal allocation proportions. For instance, one could use the relationship between the expected survival time and the hazard in the exponential distribution case, and obtain an average hazard for the Weibull distribution. Maximizing the total expected survival time for the Weibull distribution, subject to a constraint on the asymptotic variance of the treatment difference similarly gives the optimal allocation for treatment $A$ as

$$\varrho_{A1} = \frac{\gamma_B\sqrt{\mu_A\Gamma\{(\gamma_A+1)/\gamma_A\}G_A}}{\gamma_B\sqrt{\mu_A\Gamma\{(\gamma_A+1)/\gamma_A\}G_A} + \gamma_A\sqrt{\mu_B\Gamma\{(\gamma_B+1)/\gamma_B\}G_B}}. \tag{2.19}$$

The corresponding covariate-adjusted target allocation proportion is therefore;

$$\pi_{A1}^W(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B, \gamma_A, \gamma_B, \mathbf{z}) = \frac{\gamma_B\sqrt{\mu_A(\mathbf{z})\Gamma\{(\gamma_A+1)/\gamma_A\}G_A}}{\gamma_B\sqrt{\mu_A(\mathbf{z})\Gamma\{(\gamma_A+1)/\gamma_A\}G_A} + \gamma_A\sqrt{\mu_B(\mathbf{z})\Gamma\{(\gamma_B+1)/\gamma_B\}G_B}}. \tag{2.20}$$

This approach does not take into account the fact that the hazard of a Weibull distribution depends on the survival times. It would be more appropriate when patients do not have a common follow-up time.

Apart from minimizing the total average hazard, one can consider other metrics to optimize and obtain a different allocation function when the survival times follow a Weibull distribution. For instance, minimizing the overall trial size subject to the asymptotic variance of the covariate-adjusted treatment difference remaining fixed, produces the Neyman allocation function. This allocation proportion is directly proportional to the standard deviation of the logarithm of the estimated scale parameter. The Neyman allocation function results in the maximum power of the CARA procedure for a given randomization function. The allocation proportions $\pi_{A2}^{W}(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B, \gamma_A, \gamma_B, \mathbf{z})$ and $\pi_{A3}^{W}(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B, \gamma_A, \gamma_B, \mathbf{z})$ given below are considered here as an alternative to (2.18) and (2.20). The first is the Neyman allocation that maximizes the power of the Wald test for treatment comparisons, for a given sample of size $n$. Minimizing the overall trial size subject to the asymptotic variance of the covariate-adjusted treatment difference remaining fixed, the Neyman allocation is obtained as,

$$\pi_{A2}^{W}(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B, \gamma_A, \gamma_B, \mathbf{z}) = \frac{\gamma_B \sqrt{G_A}}{\gamma_A \sqrt{G_B} + \gamma_B \sqrt{G_A}}. \tag{2.21}$$

For ethical considerations in CARA randomization procedures, it is desired to target an allocation proportion which is optimal in some sense. Biswas and Mandal (2004) proposed a procedure that results in an allocation which is a generalization of optimal allocation for normal responses. They generalized the binary optimal allocation for normal responses in terms of failures. Zhang and Rosenberger (2007) applied their approach in the case of exponentially distributed survival times to develop the response adaptive randomization (RAR) approach. This approach can also be applied to develop the CARA randomization procedure when the survival times conform to a Weibull distribution. The survivor function for Weibull distributed survival times can be obtained from (2.1). If the survival times are dichotomized based on some threshold constant $\varkappa$, that is, a survival time less than the threshold $\varkappa$ is considered a failure, and otherwise a success, then the function

$$n_A[1 - e^{-\{\varkappa/\mu_A(\mathbf{z})\}^{\gamma_A}}] + n_B[1 - e^{-\{\varkappa/\mu_B(\mathbf{z})\}^{\gamma_B}}],$$

can be minimized subject to the contraint on the asymptotic variance ;

$$\frac{G_A}{n_A\gamma_A^2} + \frac{G_B}{n_B\gamma_B^2} = \mathrm{k},$$

where $k > 0$ is a constant, to obtiain the allocation proportion

$$\pi_{A3}^{W}(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B, \gamma_A, \gamma_B, \mathbf{z}) = \frac{\gamma_B\sqrt{G_A[1 - e^{-\{\varkappa/\mu_B(\mathbf{z})\}^{\gamma_B}}]}}{\gamma_B\sqrt{G_A[1 - e^{-\{\varkappa/\mu_B(\mathbf{z})\}^{\gamma_B}}]} + \gamma_A\sqrt{G_B[1 - e^{-\{\varkappa/\mu_A(\mathbf{z})\}^{\gamma_A}}]}}. \tag{2.22}$$

As pointed out earlier, the CARA designs are desirable for ethical and efficiency reasons, without undermining the validity of the trial results and maintaining the randomized nature of the experiment. Rosenberger and Hu (2004) showed that, in the case of binary responses, using the doubly-adaptive biased coin design (DBCD) to randomize an incoming patient to a particular treatment arm, results in a very useful randomization procedure in terms of maintaining power while targeting any specific allocation proportion which considers only the responses and the treatment allocation history. Zhang and Rosenberger (2006) further established this in the same settings in the case of continuous outcomes using the normal responses as the special case. A suitable randomization procedure is therefore needed to target the derived CARA allocation proportions based on the Weibull accelerated life model.

### 2.4.2 Targeting the Derived Allocation Proportion

Various randomiation procedures can be used to target a specific derived allocation proportion. Each of the randomization procedures consists of a probability allocation function whose arguments approach the derived allocation proportion. A response-adaptive design is said to be efficient of the first-order if it attains a lower bound on the asymptotic variance of the observed allocation proportion. The DBCD procedure along with most of the other randomization procedures in the literature are not first order efficient. One of the exceptions is the drop-the-loser rule of Ivanova (2003). This rule considers balls of three types : type A, type B, and type 0. A ball is drawn at random. If it is type A or type B, then the corresponding treatment is assigned, and the patient's response is observed. If it is a suc-cess, then the ball is replaced and the

urn remains unchanged. If it is a failure, then the ball is not replaced. If a type 0 ball is drawn, then no subject is treated, and the ball is returned to the urn together with one ball of type A and one ball of type B. Even though this rule is asymptotically efficient of the first order, its applications are limited to urn allocation proportions and binary responses.

Hu, Zhang and He (2009) proposed a family of randomization procedures that attain the Cramer-Rao lower bound on the allocation variances for any allocation proportion but the allocation function for this procedure is discontinuous. The asymptotic theory for adaptive designs that relies on a Taylor series expansion for the allocation functions, is not applicable to non-differentiable cases. This family of efficient randomized-adaptive designs (ERADEs) can adapt to any desired allocation proportion and is easy to implement in practice for both discrete and continuous responses. Under certain mild conditions, the asymptotic normality and strong consistency of both the allocation proportions and the estimators of the population parameters have been obtained by Hu, Zhang and He (2009). However, their work ignores the fact that the patients in clinical trials are heterogeneous and therefore does not take into account the information related to the covariate profiles of the patients. Now, when the allocation function is disccrete, the commonly used techniques do not work anymore. The allocation probabilities of the ERADE randomization procedures are discrete functions. Introducing a stopping time of a martingale process as shown in Hu, Zhang and He (2009), one can overcome the difficulties arising out of discontinuity.

After the allocation of the two treatments to $m$ patients and observing their responses, let $N_A(m)$ and $N_B(m) = m - N_A(m)$ denote the numbers of patients assigned to each of the two treatments. When the $(m+1)^{th}$ patient enters the clinical trial with covariate vector $\mathbf{z}_{m+1}$, let $\hat{\pi}_m = \hat{\pi}_A^W(\hat{\boldsymbol{\beta}}_{A,m}, \hat{\boldsymbol{\beta}}_{B,m}, \hat{\gamma}_{Am}, \hat{\gamma}_{Bm}, \mathbf{z}_{m+1})$ represent the estimate of $\pi_A^W(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B, \gamma_A, \gamma_B, \mathbf{z})$ based on the responses observed from the $m$ patients, adjusted for the covariate $\mathbf{z}_{m+1}$ of the incoming patient. Using the covariate-adjusted ERADE (CAERADE) procedure, the $(m+1)^{th}$ patient can be assigned to treatment $A$ with probability $j_{m+1}\{N_A(m)/m, \hat{\pi}_m, \hat{\rho}_{Am}\}$. Let $\hat{\rho}_{Am} = \sum_{i=1}^{m}\{\hat{\pi}_A^W(\hat{\boldsymbol{\beta}}_{A,m}, \hat{\boldsymbol{\beta}}_{B,m}, \hat{\gamma}_{Am}, \hat{\gamma}_{Bm}, \mathbf{z}_i)\}/m$ be an estimate of the average target allocation for treatment $A$ based on the data for

the first $m$ patients. Therefore, the mathematical form of the allocation rule for the $(m+1)^{th}$ patient entering the clinical trial with covariate vector $\mathbf{z}_{m+1}$ to be assigned to treatment $A$ is

$$j_{m+1}\left\{\frac{N_A(m)}{m}, \hat{\pi}_m, \hat{\rho}_{Am}\right\} = \begin{cases} \alpha'\hat{\pi}_m & \text{if } \frac{N_A(m)}{m} > \hat{\rho}_{Am}, \\ \hat{\pi}_m & \text{if } \frac{N_A(m)}{m} = \hat{\rho}_{Am}, \\ 1-\alpha'(1-\hat{\pi}_m) & \text{if } \frac{N_A(m)}{m} < \hat{\rho}_{Am}, \end{cases} \tag{2.23}$$

where $0 \leq \alpha' < 1$ is a constant that reflects the degree of randomization. Hu, Zhang and He (2009) recommended a value of $\alpha'$ to be between 0.4 and 0.7. This gives a family of CARA designs that are fully randomized and also asymptotically efficient as it attains the Cramer-Rao lower bound on the asymptotic variance of the observed allocation proportion. The ERADE can be viewed as a generalization of Efron's biased coin design, an asymptotically efficient restricted randomizaton procedure, for any desired allocation function, which may depend on the unknown parameters. If the response distribution belongs to the exponential family, the CAERADE for any $\alpha' \in [0,1)$ is efficient of the first-order. When the survival outcomes conform to a Weibull distribution, the CAERADE also generates a first-order efficient allocation design for patients.

When the allocation probability function is a continuous and differentiable function of $\hat{\rho}_{Am}$, $\hat{\pi}_m$ and the current sample proportion, the asymptotic properties of adaptive designs are obtained using a Taylor series expansion. In such cases, the expected sample proportions cannot be efficiently approximated by $\hat{\rho}_{km}$. Therefore, the variances of the allocation proportions do not attain the Cramer-Rao lower bound.

The covariate-adjusted doubly-adaptive biased coin design (CADBCD) procedure (Zhang and Hu, 2009) can also be used to construct CARA randomization procedures. The CADBCD is a randomization procedure which is used to target the allocation proportions. This is a randomization procedure with low variability and follows in the path of Efron's biased coin design. It applies to cases where the desired allocation proportions are unknown, but estimated sequentially. The key component of

this procedure is an allocation function $j_{m+1}\{N_A(m)/m, \hat{\pi}_m, \hat{\rho}_m\}$, which is defined on [0,1]×[0,1]×[0,1]. If $N_A(m)$ and $N_B(m) = m$ - $N_A(m)$ represent the numbers of patients assigned to treatments $A$ and $B$, respectively, after $m$ allocations and $\hat{\pi}_m = \pi_A^W(\hat{\boldsymbol{\beta}}_{A,m}, \hat{\boldsymbol{\beta}}_{B,m}, \hat{\gamma}_{Am}, \gamma_{Bm}, \mathbf{z}_{m+1})$ denotes an estimate of the derived target allocation proportion based on the data from those $m$ patients, adjusted for the covariate $\mathbf{z}_{m+1}$, then, according to the CADBCD allocation rule, the probability of the $(m+1)^{th}$ patient with covariate vector $\mathbf{z}_{m+1}$ being assigned to treatment $A$ is given by

$$j_{m+1}\left\{\frac{N_A(m)}{m}, \hat{\pi}_m, \hat{\rho}_{Am}\right\} = \begin{cases} \frac{\hat{\pi}_m\{\hat{\rho}_{Am}/\frac{N_A(m)}{m}\}^\alpha}{\hat{\pi}_m\{\hat{\rho}_{Am}/\frac{N_A(m)}{m}\}^\alpha + (1-\hat{\pi}_m)[(1-\hat{\rho}_{Am})/\{1-\frac{N_A(m)}{m}\}]^\alpha} & \text{if } 0 < \frac{N_A(m)}{m} < 1, \\ 1 - \frac{N_A(m)}{m} & \text{if } \frac{N_A(m)}{m} = 0 \text{ or } 1. \end{cases} \tag{2.24}$$

Here $\alpha$ is a non-negative parameter controlling the degree of randomness of the CADBCD procedure. A value of $\alpha = 0$ corresponds to the procedure being most random and a value of $\alpha = \infty$ corresponds to it being most deterministic. The allocation function $j_{m+1}\{N_A(m)/m, \hat{\pi}_m, \hat{\rho}_m\}$ is strictly decreasing in $\{N_A(m)/m\}$ and strictly increasing in $(\hat{\pi}_m, \hat{\rho}_m)$ on [0,1]×[0,1].

The allocation function given in equation (2.24) is the probablity of an incoming patient, being assigned to treatment $A$. Following Theorem 3.1 of Zhang and Hu (2009), under mild conditions, $N_A(m)/m$ and $\hat{\pi}_m$ are strongly consistent and follow an asymptotic bivariate normal distribution with the asymptotic means being the expected value of the target alloation proportion. Also, $\sqrt{m}(\hat{\boldsymbol{\beta}}_{k,m} - \boldsymbol{\beta}_k)$ is asymptotically multivariate normal with zero mean vector and asymptotic covariance matrix $I_k^{-1}(\boldsymbol{\beta}_k, \gamma_k)$. However, since the survival times conform to a Weibull distribution and there is no closed form for the maximum likelihood estimators of the scale and shape parameters, an explicit asymptotic variance for the CADBCD procedure cannot be obtained. If some knowledge about the shape parameter is available, then the derivation of the asymptotic variance of the procedure can proceed as in the exponential case as described in Zhang and Rosenberger (2007). When $\alpha \to \infty$, the allocation function is the most

deterministic and thus achieves the Cramer-Rao lower bound in terms of its asymptotic variance.

### 2.4.3 The Effect of Delay in Survival Responses

Survival time or censoring time cannot be observed until an event or censoring has happened. Therefore, there is an inherent delay in measuring the survival responses which in turn delays the estimation process of the model parameters. Zhang and Rosenberger (2006) demonstrated by using extensive simulation that a moderate delay has a marginal effect on the asymptotic properties of response-adaptive randomization procedures. Zhang and Rosenberger (2007) provided a theoretical treatment for handling the problem of delay in survival responses. They showed that, if $\tau_m$ is the delay in the response of the $m^{th}$ patient and $\eta'_m$ his/her arrival time, then, under the assumption

$$P\{\tau_m > \eta'_{(n+m)} - \eta'_{(m)}\} = o(n^{-c}), \tag{2.25}$$

for some constant $c > 0$, the asymptotic results of the DBCD procedure still hold. Note that (2.25) implies that the probability that a patient will respond before $n$ additional patients arrive is of the order $n^{-c}$. The delay in response in this case is exactly the survival time or censoring time of the patient. The asymptotic properties of the CADBCD procedure when the survival times conform to a Weibull distribution, can be justified if the assumption in (2.25) holds.

### 2.4.4 Using a Link Function to Develop a Suitable Alternative

An alternative approach presented by Bandyopadhyay and Biswas (2001) used a suitable probit link function to develop adaptive designs in clinical trials which take into account the heterogeneity of patients due to the presence of concomitant information. In their approach, they assumed that the response from each patient to a treatment is a continuous random variable and follows a normal linear model but does not account for the covariate information of the incoming patient while doing the adaptation. The method provided is further developed in this chapter in the context of Weibull

distributed survival outcomes of patients which take the information of right censoring into account.

If $m$ patients have been enrolled in a clinical trial and $N_A(m)$ patients have been allocated to treatment $A$, the observed treatment allocation proportions for $m$ patients are given by $N_A(m)/m$ for treatment $A$ and $N_B(m)/m = 1-N_A(m)/m$ for treatment $B$. A suitable link function bridges the past history about treatment allocation, responses and the covariates of the patients to the $(m+1)^{th}$ allocation. Bandyopadhyay and Biswas (2001) suggested this to be a suitable cumulative distribution function $F(.)$ which is symmetric about zero. This means that $F(0) = 1/2$ and $F(-x) =$ 1- $F(x)$. Since they assumed that the responses of the patients conform to a normal linear model, a natural choice for $F$ was the probit link function $\Phi(.)$, which is the quantile function associated with the standard normal distribution. In the context of survival models, where the survival times of the patients to a treatment are right censored and conform to a Weibull distribution, there are other link functions which can address the ethical criteria of a clinical trial more effectively. An alternative link function considered here for the purpose of developing the CARA randomization procedure is

$$F(x) = g\left\{ \frac{x}{\widehat{SE}(X) + \widehat{HR}(\mathbf{z})} \right\},$$

where $g(.)$ is the cumulative distribution function of a random variable following the Gumbel distribution, $x$ represents the value of a random variable $X$,

$$\widehat{HR}(\mathbf{z}) = \{\hat{h}_A(t|\mathbf{z})/\hat{h}_B(t|\mathbf{z})\}$$

is the ratio of the estimated hazards of the two treatment groups at the time $t$ after adjusting for the effects of the other significant covariates and $\widehat{SE}(X)$ is the estimated standard error of $X$ . The quantity $\{\widehat{SE}(X) + \widehat{HR}(\mathbf{z})\}$ is used as a scaling factor to control the variability of the design. In this case, $\widehat{SE}(X)$ is based on the magnitude and the direction of the treatment difference, depending on the observed covariate values of the patients.

Bandyopadhyay and Biswas (2001), had used an arbitrary tuning parameter $\mathbf{T}$ to scale the estimated covariate-adjusted treatment difference and showed that the power of the design as well as its variability and the degree of skewness depends on the value

of $\mathbf{T}$ which is chosen by the experimenter. However, because of the arbitrariness of the user defined value of $\mathbf{T}$ on which the variability of the design solely depends on, this design has been widely criticised over the past decade, and, therefore, even though the design is ethically attractive, it has hardly been used in practice. To address this arbitrariness, Bandyopadhyay and Biswas (2001) stated a rule for selecting the value of this tuning parameter. They explained that one can start with a large value of $\mathbf{T}$ at the intial stage of the trial and then switch to progressively smaller values at later stages. The link function $g(.)$ is referred to as the glink function. Here, instead of using an arbitrary value of the tuning parameter $\mathbf{T}$, scaling the value $x$ by a more defined quantity, $\{\widehat{SE}(X) + \widehat{HR}(\mathbf{z})\}$, makes the design less arbitrary and also addresses the issue of statistical efficiency more consistently. Moreover, the Gumbel cumulative distribution function being light tailed and steeper compared to that of a standard normal distribution increases the chance of assigning more patients to the better-performing treatment thus far in the clinical trial.

At the beginning of the trial, one uses some restricted randomization procedure to randomize the initial $2m_0$ patients, where $m_0$ is a small positive integer, equally between treatment arms $A$ and $B$. This is performed in order to collect initial data at beginning of the trial to estimate the unknown model parameters. In any survival trial, outcomes are inherently delayed. Therefore, it will take a significant amount of time until the model parameters can be accurately estimated. Let $\Delta_{Am} = \log\{\mu_{A_{(m)}}(\mathbf{z})\} - \log\{\mu_{B_{(m)}}(\mathbf{z})\}$ represent the difference between the log transformed Weibull scale parameters for the two treatments upto the $m^{th}$ patient after adjusting for the effect of the covariates. . After $m$ allocations to the two treatment arms, the rule for each incoming patient is that they would be allocated to treatment $A$ with probability

$$\pi_{A4}^{W}\{\hat{\mu}_{A_{(m)}}(\mathbf{z}), \hat{\mu}_{B_{(m)}}(\mathbf{z}), \hat{\gamma}, \mathbf{z}\} = F(\hat{\Delta}_{Am}) = g\left\{\frac{\hat{\Delta}_{Am}}{\widehat{SE}(\hat{\Delta}_{Am}) + \widehat{HR}(\mathbf{z}_m)}\right\}, \tag{2.26}$$

and to treatment $B$ with probability

$$F[\hat{\Delta}_{Bm} = \log\{\hat{\mu}_{B_{(m)}}(\mathbf{z})\} - \log\{\hat{\mu}_{A_{(m)}}(\mathbf{z})\}] = g\left\{\frac{\hat{\Delta}_{Bm}}{\widehat{SE}(\hat{\Delta}_{Bm}) + \widehat{HR}(\mathbf{z}_m)}\right\}, \tag{2.27}$$

Unlike Bandyopadhyay and Biswas (2001) who used the cumulative distribution function of a standard normal distribution, the cumulative distribution function of

the Gumbel model is based on a location parameter and a non-zero scale parameter. The scale considered here for the glink function is $(1/\gamma)$ and the location parameter depends on the covariates of the incoming patients. The proportional hazard Weibull accelerated life model assumes equal shape parameter for the two treatment groups. Let $\hat{\gamma}_m$ be the common shape parameter which is estimated from the Weibull accelerated life model fitted based on the information related to the previous $m$ patients. Therefore $(1/\hat{\gamma}_m)$ is used here as the scale parameter for the glink function at every step of the adaptation process. It has been seen using simulated trial and error method that using $\{1/\hat{\mu}(\mathbf{z_{m+1}})\}$ as the location parameter for the glink function where $\mathbf{z_{m+1}}$ are the covariate information of the incoming patients, gives an ethical design.

This allocation rule favours the better-performing treatment at a given stage of the trial after accounting for the concomitant information of the previous patients as well as that of the incoming patient. If $\Delta = \log\{\mu_A(\mathbf{z})\} - \log\{\mu_B(\mathbf{z})\}$, using (2.26) and (2.27) the probability of allocating an incoming patient to treatment $A$ is being forced towards $F(\Delta)$, an increasing function of $\Delta$, provided that the estimators of the treatment effects are consistent.

Let $\mathcal{X}_m$, $\mathcal{Y}_m$ , $\mathcal{Z}_m$ denote, respectively, the sigma fields for the past treatment allocation history, the response history, the prognostic factors for the first $m$ patients and let $\mathbf{z_{m+1}}$ be the covariates of the incoming patient. Based on the glink, the conditional probability that the $(m+1)^{th}$ patient with covariate vector $\mathbf{z_{m+1}}$, will be assigned to treatment $A$ is given by

$$P(X_{(m+1)} = 1|\mathcal{X}_m, \mathcal{Y}_m, \mathcal{Z}_m, \mathbf{z_{m+1}}) = g\left\{ \frac{\hat{\Delta}_{Am}}{\widehat{SE}(\hat{\Delta}_{Am}) + \widehat{HR(\mathbf{z}_m)}} \right\}. \tag{2.28}$$

where, conditional on $\mathcal{X}_m$,

$$\hat{\Delta}_{Am} \xrightarrow{d} N(\Delta_{Am}, \sigma^2)$$

as $m \to \infty$. The variance $\sigma^2$ is calculated by inverting the Fisher information matrix obtained from the log likelihood function (2.11) as shown in Appendix A. Let $\zeta(m) =$ P $(X_{(m+1)}= 1)$. Then the sequence $\{\zeta(m) : m \geq 2m_0 + 1\}$ converges to $g\left\{ \frac{\Delta_{Am}}{HR(\mathbf{z}_m)} \right\}$. The limiting proportion of the allocation of patients to treatment $A$ is also $g\left\{ \frac{\Delta_{Am}}{HR(\mathbf{z}_m)} \right\}$ .

The glink function puts more weight on the available data to develop the CARA randomization procedure. The cumulative distribution function of a Gumbel random variable is light-tailed, and therefore gives more weight to the available data compared to its heavy-tailed counterparts like the Cauchy. The main advantage of using the glink function for developing the CARA randomization procedure is that it tends to allocate more patients to the better treatment arm as compared to the design based on the probit model. Thus, it fulfils the ethical considerations of adaptive designs more than that of Bandyopadhyay and Biswas (2001).

## 2.5 Asymptotic Properties of CARA Designs

An important aspect of the design of phase III clinical trials is the use of appropriate methods of randomization. The previous section gave a detailed explanation about various target allocation proportions and the discrete and continuous CARA randomization procedures targeting these allocation proportions in the design of phase III clinical trials. It would, however, be useful to compare the performance of these target allocations and the CARA randomization procedures to find out which randomization procedure targeting a specific allocation proportion outperforms the others. Hu and Rosenberger (2003) provided a theoretical template for the comparison of different response-adaptive randomization procedures and different target allocations in terms of power and expected failure rates when the response of the patients to a treatment is binary. Zhang and Rosenberger (2006) further developed this idea for continuous responses following a normal distribution. In this section a theoretical template is given for the comparison of different CARA designs and different target allocation proportions when the survival responses conform to a Weibull distribution.

The optimal allocation proportions are largely dependent on the choice of measure of difference between the treatment. In this chapter, the focus has been on the simple difference $[\Delta = \log\{\mu_A(\mathbf{z})\} - \log\{\mu_B(\mathbf{z})\}]$. The entire theory behind covariate-adjusted response-adaptive designs for phase III clinical trials rely on asymptotic approximation of the observed allocation proportion and the estimated target allocation proportion.

Asymptotically the Wald test, score (Rao) test, and the likelihood ratio test are equivalent in the sense that the probability limit of the test statistic for all the three tests are equal. Moreover in real clinical trials the Wald tests are frequently used for treatment comparison. Thus the Wald test is used here to test for the significance of this difference, $\Delta$, between the treatment effects.

Using a Taylor series expansion of the non-centrality parameter of the usual chi-squared test for testing the effect of the treatments, an explicit relationship can be derived between the target allocation proportion, the bias of the randomization procedure from the target and the variability induced by the randomization procedure for any allocation proportion. When there are random samples ($t_{ik}$,$\delta_{ik}$), ($k = A, B$) from Gumbel distributions with parameters $[\log\{\mu_k(\mathbf{z})\}, 1/\gamma_k]$, the Wald test statistic is given by

$$T_n = \frac{\log\{\hat{\mu}_A(\mathbf{z})\} - \log\{\hat{\mu}_B(\mathbf{z})\}}{\sqrt{\frac{G_A}{n_A\hat{\gamma}_A^2} + \frac{G_B}{n_B\hat{\gamma}_B^2}}}, \text{ where } T_n \xrightarrow{d} \text{N(0,1)},$$

as $n \to \infty$, for testing the hypothesis

$$H_0 : \Delta = 0$$

against,

$$H_A : \Delta \neq 0$$

For a design with $n_A$ and $n_B$ fixed and the times to events independent Weibull survival outcomes, $T_n^2$ is asymptotically chi-squared with one degree of freedom. Under the alternative hypothesis, power can be expressed as an increasing function of the non-centrality parameter of the chi-squared distribution for a fixed target allocation proportion $\pi$. Using the simple difference measure, the non-centrality parameter can be expressed as

$$\Lambda = \frac{[\log\{\mu_A(\mathbf{z})\} - \log\{\mu_B(\mathbf{z})\}]^2}{\frac{G_A}{n_A\gamma_A^2} + \frac{G_B}{n_B\gamma_B^2}},$$

which can be re-written as,

$$\frac{\Lambda(x)}{n} = \frac{[\log\{\mu_A(\mathbf{z})\} - \log\{\mu_B(\mathbf{z})\}]^2}{\frac{G_A}{(\pi+x)\gamma_A^2} + \frac{G_B}{(1-\pi-x)\gamma_B^2}},$$

where $x$ = ($n_A$/n) - $\pi$, and $\pi$ is the target allocation proportion on treatment $A$. Expanding $\frac{\Lambda(x)}{n}$ in a Taylor series about $\pi$ yields

$$\frac{\Lambda(n_a/n)}{n} = \frac{[\log\{\mu_A(\mathbf{z})\} - \log\{\mu_B(\mathbf{z})\}]^2}{\frac{G_A}{\pi\gamma_A^2} + \frac{G_B}{(1-\pi)\gamma_B^2}}$$

$$+[\log\{\mu_A(\mathbf{z})\} - \log\{\mu_B(\mathbf{z})\}]^2 \frac{(1-\pi)^2\frac{G_A}{\gamma_A^2} - \pi^2\frac{G_B}{\gamma_B^2}}{\{(1-\pi)\frac{G_A}{\gamma_A^2} + \pi\frac{G_B}{\gamma_B^2}\}^2}\left(\frac{n_A}{n} - \pi\right)$$

$$-[\log\{\mu_A(\mathbf{z})\} - \log\{\mu_B(\mathbf{z})\}]^2 \frac{(\frac{G_A}{\gamma_A^2}\frac{G_B}{\gamma_B^2})}{\{(1-\pi)\frac{G_A}{\gamma_A^2} + \pi\frac{G_B}{\gamma_B^2}\}^3}\left(\frac{n_A}{n} - \pi\right)^2 + o\left\{\left(\frac{n_A}{n} - \pi\right)^2\right\}$$

$$\text{or,} \quad \frac{\Lambda(n_a/n)}{n} = (I) + (II) + (III) + o\left\{\left(\frac{n_A}{n} - \pi\right)^2\right\} \tag{2.29}$$

The first term (I) is determined by $\pi$ and represents the non-centrality parameter for a fixed design. The Neyman allocation for patients with a given set of covariates $\mathbf{z}$ as in (2.21) maximizes this term. This term can be used to compare different target allocation proportions in terms of their powers. The second term (II) represents the bias of the actual allocation from the target allocation. With the design shifting to different sides from the target allocation proportion $\pi$, the non-centrality parameter will increase or decrease according to the coefficient

$$[\log\{\mu_A(\mathbf{z})\} - \log\{\mu_B(\mathbf{z})\}]^2 \frac{(1-\pi)^2\frac{G_A}{\gamma_A^2} - \pi^2\frac{G_B}{\gamma_B^2}}{\{(1-\pi)\frac{G_A}{\gamma_A^2} + \pi\frac{G_B}{\gamma_B^2}\}^2},$$

and this coefficient equals 0 if and only if $(1-\pi)^2\frac{G_A}{\gamma_A^2} - \pi^2\frac{G_B}{\gamma_B^2} = 0$ , that is

$$\pi = \pi_{A2}^W(\beta_A, \beta_B, \gamma_A, \gamma_B, \mathbf{z}) = \frac{\gamma_B\sqrt{G_A}}{\gamma_A\sqrt{G_B} + \gamma_B\sqrt{G_A}},$$

the Neyman allocation given in (2.21).

In a real-life scenario, especially in the field of personalized medicine, people may be interested to know the proportion of patients on a particular treatment for a given set of covariates $\mathbf{z}$. CARA randomization procedures involve, $N_{A|\mathbf{z}}(m)$, the number of patients with covariate $\mathbf{z}$ allocated to treatment $A$ after $m$ allocations. Given a covariate $\mathbf{z}$, the proportion of patients allocated to treatment $A$ is $N_{A|\mathbf{z}}(m)/N_{\mathbf{z}}(m)$,

$$\frac{\sum_{m=1}^n X_A(m) I_{\{\mathbf{z}_m = \mathbf{z}\}}}{\sum_{m=1}^n I_{\{\mathbf{z}_m = \mathbf{z}\}}} = \frac{N_{A|\mathbf{z}}(m)}{N_{\mathbf{z}}(m)},$$

where $N_{\mathbf{z}}(m)$ is the total number of patients with covariate $\mathbf{z}$ . Replacing $n_A/n$ and $n_B/n$ with $N_{A|\mathbf{z}}(m)/N_{\mathbf{z}}(m)$ and $N_{B|\mathbf{z}}(m)/N_{\mathbf{z}}(m)$ in (2.31), the test statistic $T_n^2$ still has an asymptotic chi-squared distribution. The specific procedures for which the asymptotic properties hold and the necessary conditions are well described in the response-adaptive randomization literature. The justification for the asymptotic chi-squared distribution of the test statistic can be deduced from Hu and Zhang (2004). A very general form of the asymptotic variance of the observed allocation proportion was given by Eisele and Woodroofe (1995) and Hu and Zhang (2004); actual computation depends on specific values of the target allocation proportion and the allocation rule, which is straightforwarward but tedious. The critical condition to ensure the chi-squared limit is that the covariate-adjusted allocation proportion on each treatment converges almost surely to a constant between 0 and 1 for the specific procedure. This substitution makes $\Lambda\{N_{A|\mathbf{z}}(m)/N_{\mathbf{z}}(m)\}$ a random non-centrality parameter, and therefore its expectation can be considered. For example, since $N_{A|\mathbf{z}}(m)/N_{\mathbf{z}}(m)$ is asymptotically unbiased for $\pi$, E[$\{N_{A|\mathbf{z}}(m)/N_{\mathbf{z}}(m)\}$ - $\pi$ ] $\rightarrow$ 0. Therefore, the average power of the test directly relates to the variance $E[\{N_{A|\mathbf{z}}(m)/N_{\mathbf{z}}(m)\} - \pi]^2$ of the CARA procedure. It is this explicit relationship that would be mostly used to evaluate the power performances of different CARA randomization procedures.

## 2.6 Simulation Study

### 2.6.1 Choice of Design Parameters

In the experimental setup to compare the different CARA randomization procedures, a two-arm survival trial with 400 patients has been considered. A patient's arrival time here is simulated from a uniform (0,365) distribution. The response time of a patient is added to the recruitment time of the patient and those whose outcomes have not been observed by specified time $D > R$ are said to be generalized type I right censored. The length of the recruitment period is considered to be $R = 365$ days and the overall trial duration is taken to be $D = 581.66$ days. The censoring time of the patients is

simulated from a uniform (0,581.66) distribution.

Following Rosenberger, Vidyashankar and Agarwal (2001), a covariate structure of three independent covariates has been generated. These are gender (Bernoulli, p = 0.5), age (Uniform(30,75)) and cholestrol level (Normal (200,400)). Treatment-covariate interactions are not considered during the simulation stage, as the covariate-adjusted treatment effects can be quite effective as compared to the individual treatment effects, even in the absence of such interaction terms. The survival time of a patient with covariate vector $\mathbf{z} = (1, z_1, z_2, z_3)^T$ in treatment group $k$ is simulated from the Weibull distribution with scale parameter $\mu_k(\mathbf{z})$ =exp($\boldsymbol{\beta}_k^T\mathbf{z}$) and shape parameter $\gamma_k = 2.07527$ when assessing the situation for a monotonic increassing hazard and $\gamma_k = 0.57527$ when assessing for a monotonic decreasing hazard. The values of shape parameters were chosen by trial and error, taking into account the three covariate structures with known distributions, Weibull distributed time to event outcomes, and to have approximate values of hazard ratio in the population as per the model parameters given in Table 2.1. The values of hazard ratio will vary across individual patients depending on their covariate values. Three choices of the treatment effects vector have been considered in this case, which are neutral effect of either treatment, positive effect of treatment $A$ and negative effect of treatment $A$. The effects of the corresponding covariates for the simulation model $\mu_k(\mathbf{z})$ =exp($\boldsymbol{\beta}_k^T\mathbf{z}$) are summarized in Table 2.1.

| **Model** | **Treatment** | **Covariate Effects** | | | |
|---|---|---|---|---|---|
| | | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\beta_3$ |
| Neutral | A | 1.896 | 0.810 | 0.038 | 0.001 |
| | B | 1.896 | 0.810 | 0.038 | 0.001 |
| Positive | A | 5.5042 | 0.810 | 0.038 | 0.001 |
| | B | 1.896 | 0.810 | 0.038 | 0.001 |
| Negative | A | -1.7112 | 0.810 | 0.038 | 0.001 |
| | B | 1.896 | 0.810 | 0.038 | 0.001 |

Table 2.1: Values of model parameters

In Table 2.1, the neutral treatment effect refers to the hypothetical experimental scenario where treatments $A$ and $B$ are equally effective. In the case of comparing a new treatment with a control, this scenario refers to the situation where the new treatment is as good as the existing control. The positive treatment effect refers to the hypothetical experimental scenario where treatment $A$ is more effective than treatment $B$, or the new treatment performs better than the control. The negative treatment effect refers to the hypothetical experimental scenario where treatment $B$ is more effective than treatment $A$, or in the case of comparing a new treatment with a control, this means that the new treatment is not as effective as the control. The procedure used here is a fully sequential one that recalculates the randomization probabilities after the arrival of each patient and there are 5000 such simulation runs. Since there are three predictive covariates in the model, the direction and magnitude of the treatment difference will vary for the patients, depending on their observed covariate values.

In order to compare the different competing designs, two response adaptive rules have also been considered for which the covariates are ignored at the design stage, but the final estimates of the treatment effects can be adjusted for all covariates. The competing randomization procedures and the corresponding design numbers are listed in Table 2.2.

In survival trials, the delay time for a patient is the patients' survival or censoring time. To facilitate CARA designs with delayed responses, it is required that, at the $i^{th}$ patient′s randomization time, only data from those patients who have responded before the $i^{th}$ patient's arrival are used in computing the randomization probability for the $i^{th}$ patient. In practice, the assumption of immediate responses is not feasible due to the inherent delay in time-to-event outcomes. For the implementation of the CARA and the response adaptive designs given in Table 2.2 above, initially $2m_0$ patients have been equally allocated to the two treatment arms using Efron's biased coin design. This is a restricted randomization procedure which allocates a patient to a treatment arm based on the history of the treatment assignment with the aim of achieving balance

| Design | Competing Randomization Procedures |
|---|---|
| I | Completely randomized design |
| II | Efron's biased coin design with p = 2/3 |
| III | Pocock and Simon design with p = 3/4 |
| IV | CARA CADBCD with (2.18) as the target |
| V | CARA CADBCD with (2.22) as the target |
| VI | CARA CADBCD with (2.21) as the target |
| VII | CARA CADBCD with (2.20) as the target |
| VIII | CARA CAERADE with (2.18) as the target |
| IX | CARA CAERADE with (2.22) as the target |
| X | CARA CAERADE with (2.21) as the target |
| XI | CARA CAERADE with (2.20) as the target |
| XII | Response Adaptive DBCD with (2.17) as the target |
| XIII | Response Adaptive ERADE with (2.17) as the target |
| XIV | CARA design based on the g-link function |

Table 2.2: List of the Competing Designs

without introducing biases which are usually avoided by using a completely randomized design. At a given stage, one computes the treatment imbalance and with probability 2/3 an incoming patient is assigned to the underrepresented treatment arm to reduce the overall imbalance. This design represents a class of biased coin procedures for the comparison of two treatments in which allocation of the treatment is determined probabilistically, but with a bias towards the underrepresented one. Permuted block design is another restricted randomization approach which is commonly used in practice and also achieves balance across treatment arms. This randomization scheme consists of a sequence of blocks such that each block contains a pre-specified number of treatment assignments in random order. The purpose of this is so that the randomization scheme is balanced at the completion of each block. Permuted blocks can be quite effective in achieving balanced designs but they suffer from the disadvantage that at certain points in the experiment, the experimenter knows for certain whether the next subject will be assigned as a treatment or as a control. For example, if the block size is

5, the probability is 1/6 that the experimenter will know for certain the assignment of units 8, 9 and 10, and 4/9 that he will know for certain the assignment of units 9 and 10. Efron's biased coin design is asymptotically the most efficient restricted randomization procedure and therefore enables one to achieve an allocation procedure with high power for treatment comparison. Here, $m_0$ is a positive number and using Sverdlov, Rosenberger and Ryeznik (2013) as a guidance, $2m_0$ is chosen to be 110 for the CARA designs and 80 for the response-adaptive designs, which are sufficiently large for accurate estimation of the model parameters. It must be noted that unlike the designs based exponential regression model, the designs here needs a longer time to start the adaptation process. This is because of having the number of patients to efficiently estimate models parameters which includes the extra shape parameter in this case.

At a given step, the Newton-Raphson method for fitting the Weibull regression model to the data may not converge and maximum likelihood estimate are not attainable. In that case, the treatment assignment for the patient is determined using an Efron's biased coin design. After the model parameters are estimated from the initial stage of the design, a randomization probability is calculated after each new patient who arrives sequentially into the trial. This randomization probability can be based on any one of the derived allocation functions. A pseudo random number generator (PRNG) is then used to draw a random number between 0 and 1. If the derived randomization probability is greater than or equal to this generated random number, the patient is assigned to treatment $A$ or else the patient is assigned to treatment $B$. The procedure described is repeated for the subjects entering the trial in the future. The randomization procedures which are being compared with the derived CARA designs are as follows :

- A completely randomized design, for which every patient is randomized to treatment $A$ or $B$ with probability 0.5.
- Efron's (1971) biased coin design in which allocation of the treatment is determined with a probability of 2/3 towards the underrepresented treatment.
- Pocock and Simon's (1975) covariate-adaptive randomization procedure. For its

implementation, all covariates must be categorical. Therefore, the continuous covariate age has been dichotomized according to age $< 53$ years and age $\geq 53$ years, whereas the covariate cholestrol level has been dichotomized to level $< 200$ and level $\geq 200$. For an incoming patient, one computes the treatment imbalance at each level of the patient's covariates and with probability 3/4 the patient is assigned to the treatment arm that reduces the overall covariate imbalance. If the imbalance is zero, the patient is randomized to treatment $A$ or $B$ with equal probability.

- A response-adaptive rule with (2.17) as its target, for which only the covariates are ignored at the design stage but the final estimates of the treatment effects can be adjusted for all covariates. The response-adaptive rule is implemented by means of the doubly adaptive biased coin design with $\alpha = 2$ (Hu and Zhang, 2004).

- A response-adaptive strategy with (2.17) as its target, which is implemented by means of the efficient randomized adaptive design with $\alpha' = 0.55$ (Hu, Zhang and He, 2009).

The observed allocation proportion $\{N_A(m)/m\}$ for the optimal designs at the $m^{th}$ stage of a clinical trial converges to its target allocation proportion $\pi$ at the rate of $n^{-1/2}$. The asymptotic results of the CADBCD and the CAERADE do not depend on its randomization parameter $\alpha$ and $\alpha'$ respectively. This is because the first order approximation of the allocation probability function do not depend on the randomization parameter. In practice, one need to choose a suitable value of the randomization parameter to implement the proposed designs. For the DBCD and the CADBCD designs in Table 2.2 above, it must be noted that following Zhang and Hu (2009), Hu and Zhang (2004) and Rosenberger and Hu (2004), the trade-off parameter for randomness is taken to be $\alpha = 2$. Similarly, Burman (1996) introduced the expected p-value deficiency to evaluate the performance of a particular design. Based on Burman's studies, Hu, Zhang and He (2009) recommended that, for appropriate implementation of the ERADE designs, it is reasonable to choose $\alpha'$ between 0.4 to 0.7. The parameter $\alpha'$ controls the degree of randomness of the design. When $\alpha'$ is smaller, the ERADE is

more deterministic and has a smaller variability. Here, $\alpha'$ for CAERADE and ERADE is chosen to be 0.55.

Similation of 5000 runs were considered for sensitivity analysis of the effect of $\alpha'$ on the variability of a CAERADE design. It has been seen that $\alpha'$ is related to the randomness of the design. When $\alpha'$ is smaller, the CAERADE is more determined and could have a smaller variance. Following Hu,Zhang and He (2009) a simulation study was conducted with $\alpha' = 0.125$, $\alpha' = 0.25$, $\alpha' = 0.50$, $\alpha' = 0.67$ and $\alpha' = 0.75$. Overall sample sizes of 50, 100 and 200 were considered. It has been found that the simulated results of $\alpha' = 0.125$ and $\alpha' = 0.25$ are very similar to the results of $\alpha' = 0.50$ in terms of obtaining the target allocation proportion and its variability. However, the CAERADE with $\alpha' = 0.75$ has slightly higher variability than others. Therefore, it is reasonable to choose $\alpha'$ in between 0.4 to 0.7 which agrees with the findings of Burman (1996) based on biased coin designs.

A similar simulation study was also conducted to assess the sensitivity of the randomization parameter $\alpha$ on the CADBCD. With $\alpha = 0$, the CADBCD becomes the adaptive randomized design proposed by Melfi, Page and Geraldes (2001). One disadvantage of the adaptive randomized design is that, for small experiments, the allocation at times could be far from the target proportion. This was shown by Efron (1971) with $\pi = 1/2$ . Such designs with $\alpha = 0$ also has the highest asymptotic variability and thus low power for treatment comparison. The CADBCD with $\alpha > 0$ always has smaller asymptotic variance than the adaptive randomized design. If $\alpha = 1$, the variance of CADBCD is half of the variance of the adaptive randomized design. Therefore, one can force a small-sized experiment to efficiently target a derived allocation proportion by choosing a value of $\alpha$ for the proposed designs. As $\alpha$ tends to $\infty$, the CADBCD procedure assigns the incoming patient to treatment $A$ with probability one if $N_A(m)/m < \pi$, and to treatment $B$ with probability 1 if $N_A(m)/m > \pi$. This procedure is entirely deterministic (except when $N_A(m)/m = \pi$). It turns out that the deterministic procedure has the smallest variability that can be attained by any procedure targeting the optimal allocation proportion. However, one loses the benefit of randomization. As $\alpha$ becomes smaller, we have more randomization, but also more variability. The

CADBCD with $\alpha = 2$ tends to have very good convergence for moderate sample sizes. Simulation results on the sensitivity analyses on the value of $\alpha$ show that $\alpha = 2$ is a good tradeoff that yields almost the same results as $\alpha = \infty$, but is slightly better than the adaptive randomized design. Results have also been simulated for $\alpha = 5$ and it was nearly identical to $\alpha = 2$. In every case, the CADBCD with $\alpha = 2$ works better than complete randomization, in terms of reducing the number of events in a clinical trial.

### 2.6.2 Comparing CARA Designs for No Difference Between the Treatment Effects

To visualize the performances of the various competing designs in Table 2.2 when the survival response of the patients follows a Weibull distribution, it is always useful to start with the simplest case, that is, when under the null hypothesis $H_0$ that there is no difference between the covariate-adjusted treatment effects is true. All simulations were carried out using the R statistical software. For each experimental procedure, a trial with $n = 400$ patients was simulated 5,000 times and the significance level of the Wald test for testing the difference between the covariate-adjusted treatment effects has been set to 0.05. Table 2.3 presents operating characteristics of the randomization designs in Table 2.2 in the case of the neutral model.

Table 2.3 consists of seven columns and it gives the summary results for the 14 different competing designs for the neutral model. The third column gives the observed allocation proportion of patients assigned to treatment $A$ and its standard error over 5000 simulation runs. The first column represents the two scenarios in which the shape parameter is less than or greater than 1, whereas the fourth and fifth columns provide the average numbers of patients categorized by their gender, allocated to each of the two treatments. The sixth column shows the average total number of events in a trial. The final column presents the type I error rate of the Wald test, which is used to test for the covariate-adjusted treatment difference. It is essential that the randomization pro-

| **Hazard** | **Design** | $\frac{N_A}{n}$ **(SE)** | $N_{AM}$-$N_{BM}$ | $N_{AF}$-$N_{BF}$ | **Event** | **Type I Error** |
|---|---|---|---|---|---|---|
| Decreasing | I | 0.50 (0.025) | 100-100 | 100-100 | 325 | 0.05 |
| | II | 0.50 (0.003) | 100-100 | 100-100 | 325 | 0.05 |
| | III | 0.50 (0.012) | 100-100 | 100-100 | 325 | 0.06 |
| | IV | 0.50 (0.038) | 100-100 | 100-100 | 325 | 0.03 |
| | V | 0.50 (0.033) | 100-100 | 100-100 | 325 | 0.03 |
| | VI | 0.50 (0.027) | 100-100 | 100-100 | 325 | 0.03 |
| | VII | 0.50 (0.025) | 100-100 | 100-100 | 325 | 0.04 |
| | VIII | 0.50 (0.026) | 100-100 | 100-100 | 325 | 0.03 |
| | IX | 0.50 (0.021) | 100-100 | 100-100 | 325 | 0.04 |
| | X | 0.50 ( 0.016) | 100-100 | 100-100 | 325 | 0.04 |
| | XI | 0.50 (0.006) | 100-100 | 100-100 | 325 | 0.04 |
| | XII | 0.50 (0.037) | 100-100 | 100-100 | 325 | 0.05 |
| | XIII | 0.50 (0.030) | 100-100 | 100-100 | 325 | 0.05 |
| | XIV | 0.40 (0.047) | 80-121 | 80-119 | 325 | 0.02 |
| Increasing | I | 0.50 (0.025) | 100-100 | 100-100 | 306 | 0.05 |
| | II | 0.50 (0.003) | 100-100 | 100-100 | 306 | 0.05 |
| | III | 0.50 (0.012) | 100-100 | 100-100 | 306 | 0.06 |
| | IV | 0.50 (0.038) | 100-100 | 100-100 | 306 | 0.04 |
| | V | 0.50 (0.034) | 100-100 | 100-100 | 306 | 0.04 |
| | VI | 0.50 (0.029) | 100-100 | 100-100 | 306 | 0.05 |
| | VII | 0.50 (0.026) | 100-100 | 100-100 | 306 | 0.05 |
| | VIII | 0.50 (0.033) | 100-100 | 100-100 | 306 | 0.05 |
| | IX | 0.50 (0.028) | 100-100 | 100-100 | 306 | 0.05 |
| | X | 0.50 ( 0.020) | 100-100 | 100-100 | 306 | 0.04 |
| | XI | 0.50 (0.009) | 100-100 | 100-100 | 306 | 0.05 |
| | XII | 0.50 (0.055) | 100-100 | 100-100 | 306 | 0.05 |
| | XIII | 0.50 (0.049) | 100-100 | 100-100 | 306 | 0.05 |
| | XIV | 0.40 (0.047) | 80-120 | 80-120 | 306 | 0.03 |

Table 2.3: Performances of the competing designs in the case of the neutral model

cedure maintains the nominal type I error rate and has high power to detect treatment differences.

Table 2.3 shows that under the neutral model, irrespective of the value of the shape parameter, almost all of the randomization designs result in an equal allocation of patients to treatments $A$ and $B$. The CARA design using the glink function on average allocates significantly more patients to treatment $B$ even though the difference between the covariate adjusted effects for the two treatments is zero. On the other hand, all of the CAERADE and the ERADE randomization procedures are less variable than the corresponding CADBCD and the DBCD procedures. This reflects the theoretical phenomenon explained earlier that the ERADE being the most efficient adaptive randomization design makes it the least variable, as the asymptotic variance of the observed allocation proportion attains the Cramer-Rao lower bound. The column in Table 2.3 showing the average total number of events in a trial suggests that, even after a slight under allocation in the CARA design based on the glink function, it does not result in any ethical gain or loss. This phenomenon of the CARA design with the glink function may be due to the fact that the cumulative distribution function of a Gumbel model is not symmetric about zero. All of the other competing designs result in an equal number of events on an average from 5000 different trials. Therefore when the difference between the covariate-adjusted effects for the two treatments is zero, the CARA designs ethically performs as well as the traditional balanced randomization procedures or the response-adaptive randomization procedures.

While maintaining the type I error rate of the Wald test for testing the covariate-adjusted treatment difference, it can be seen that Efron's biased coin design and the response-adaptive designs perform the best compared to the other traditional balanced randomization procedures when the value of the shape parameter is less than 1. When the value of the shape parameter is greater than 1, most of the competing designs have a type I error rate closer to the nominal value. The Pocock-Simon design, however, in both the cases gives a slightly inflated error rate. With $n = 400$ patients, most of the type I error rates for the designs when the shape parameter is less than 1, are slightly conservative. This is improved in the case when the shape parameter for the Weibull

model is greater than 1. This behaviour was further explored by simulation for larger sample sizes and other values of the shape parameter. The simulated type I error for adaptive designs continued to be conservative for shape parameter is less than 1. This may be because of the fact that the density function of the Weibull model is closer to normality when the hazard function is increasing compared to when the hazard function is decreasing. The CARA design with the glink function being a non-optimal design gives a very conservative type I error rate. The standard errors of the type I error rates are found to be between 0.002 and 0.004. All of the standard errors of the average number of events in a trial are found to be 8 in the case of an increasing hazard and 9 in the case of a decreasing hazard apart from the CARA design based on the glink function whose standard error for the events is 8.

### 2.6.3 Comparing CARA Designs for Differences in Treatment Effects

The usefulness of a CARA design is appreciated when the covariate-adjusted effects differ between the two treatment arms. Table 2.4 presents the operating characteristics of the competing randomization designs in Table 2.2 in the case of the positive model where the covariate-adjusted treatment effect has a positive impact on the survival experience of the patients having treatment $A$. Similar to Table 2.3, this also lists two scenarios according to the shape parameter of the Weibull model being less than 1 or greater than 1.

Unlike the neutral model, most of the designs apart from the traditional balanced randomization designs result in a skewed treatment allocation towards the better-performing treatment arm, which in this case is $A$. It can again be seen from column 3 of Table 2.4 that the CAERADE or the ERADE is less variable than the CADBCD or the DBCD respectively. If the variability of the randomization procedure is the only criterion for assessing the performance of a design, then the CAERADE proivide the

| **Hazard** | **Design** | $\frac{N_A}{n}$ **(SE)** | $N_{AM}$-$N_{BM}$ | $N_{AF}$-$N_{BF}$ | **Event** | **Power** |
|---|---|---|---|---|---|---|
| Decreasing | I | 0.50 (0.025) | 100-100 | 100-100 | 168 | 0.98 |
| | II | 0.50 (0.003) | 100-100 | 100-100 | 168 | 0.99 |
| | III | 0.50 (0.012) | 100-100 | 100-100 | 168 | 0.99 |
| | IV | 0.54 (0.035) | 106-94 | 111-89 | 166 | 0.95 |
| | V | 0.53 (0.031) | 104-97 | 106-93 | 166 | 0.94 |
| | VI | 0.39 (0.028) | 77-123 | 80-120 | 168 | 0.97 |
| | VII | 0.57 (0.027) | 113-87 | 113-87 | 165 | 0.96 |
| | VIII | 0.57 (0.029) | 107-93 | 121-79 | 166 | 0.97 |
| | IX | 0.53 (0.028) | 106-94 | 107-93 | 165 | 0.97 |
| | X | 0.38 (0.025) | 74-126 | 80-120 | 168 | 0.98 |
| | XI | 0.58 (0.020) | 116-84 | 116-84 | 166 | 0.98 |
| | XII | 0.55 (0.034) | 110-90 | 110-90 | 165 | 0.96 |
| | XIII | 0.55 (0.029) | 110-90 | 110-90 | 166 | 0.97 |
| | XIV | 0.68 (0.049) | 136-65 | 135-64 | 152 | 0.92 |
| Increasing | I | 0.50 (0.025) | 100-100 | 100-100 | 194 | 0.99 |
| | II | 0.50 (0.003) | 100-100 | 100-100 | 194 | 0.99 |
| | III | 0.50 (0.012) | 100-100 | 100-100 | 194 | 0.99 |
| | IV | 0.55 (0.027) | 111-91 | 108-90 | 192 | 0.96 |
| | V | 0.54 (0.027) | 109-88 | 111-92 | 192 | 0.96 |
| | VI | 0.44 (0.030) | 88-112 | 90-110 | 194 | 0.96 |
| | VII | 0.55 (0.028) | 110-90 | 110-90 | 192 | 0.96 |
| | VIII | 0.55 (0.024) | 110-90 | 110-90 | 192 | 0.99 |
| | IX | 0.53 (0.023) | 105-95 | 107-93 | 192 | 0.98 |
| | X | 0.44 (0.024) | 85-115 | 91-119 | 195 | 0.98 |
| | XI | 0.56 (0.020) | 113-87 | 113-87 | 192 | 0.97 |
| | XII | 0.54 (0.038) | 108-92 | 108-92 | 192 | 0.96 |
| | XIII | 0.53 (0.029) | 105-95 | 106-94 | 192 | 0.97 |
| | XIV | 0.65 (0.052) | 130-70 | 131-69 | 183 | 0.93 |

Table 2.4: Performances of the competing designs in the case of the positive model

most suitable design. It can be seen that almost all of the CARA designs result in a skewed allocation of patients to the better treatment arm, but the degree of skewness varies between the different designs. With $n$=400 patients, the powers of the balanced designs are the highest compared to the other ones. However, they result in more events, and are therefore ethically not as attractive as the CARA or the response-adaptive designs. The CARA design with the glink function provides the most skewed design towards the better treatment arm. This, in turn, results in considerably fewer events without compromising much on the power of the Wald test for testing the difference in the covariate-adjusted treatment effects. This design is not based on any formal optimization procedure, but sequentially takes into account the hazard ratio of the patients between the two treatment arms and uses it as a scaling factor . It has an ethical advantage over the other designs because the cumulative distribution function of a Gumbel model is a steep increasing function and the probability of allocation to $A$ is being forced towards $\mathrm{F}[\log\{\mu_{A_{(m)}}(\mathbf{z})\} - \log\{\mu_{B_{(m)}}(\mathbf{z})\}]$, a steep increasing function of $[\log\{\mu_{A_{(m)}}(\mathbf{z})\} - \log\{\mu_{B_{(m)}}(\mathbf{z})\}]$ provided the estimators are consistent. It can be seen that using the adaptive designs results in slightly fewer events compared to the traditional balanced randomization designs. The average number of events is quite similar across the competing designs. The standard error of these for all of the designs is 10 except for the CARA design based on the glink function whose standard error is 9 for the decreasing hazard scenario.

The final column of Table 2.4 shows the power of the Wald test. It can be seen that, when treatment $A$ has a more positive effect on the survival experience of the patients, using a CARA randomization procedure addresses the ethical criterion of a clinical trial of treating more patients wiith the better treatment without compromising much on the power of the Wald test, compared to that of the traditional balanced designs. In case both hazard scenarios, the response-adaptive designs and the CARA design with the glink function have the most variable power, whereas the balanced randomization designs have the least variable power. For the case of decreasing hazard, the variability of the power for design VIII is similar to that of the balanced randomization designs. In the case of increasing hazard the variability of the power for design V is similar to that of the response-adaptive procedures. All of the standard errors for the power ranged

from 0.001 to 0.004.

Unlike the case of the exponential survival responses, when comparing CARA and response adaptive designs with targets, the CARA designs perform slightly better when the Weibull distributed responses with an increasing hazard rate are delayed and the effect of treatment $A$ is better on the survival experience of the patients compared to that of treatment $B$. In contrast to the response adaptive designs, the CARA designs also cater for situations when there exists a need to allocate a treatment based on the individual's covariate profile. The CARA designs allocate more patients to the better treatment arm at each of the patient subgroup levels. It can also be seen that CARA designs targeting the Neyman allocation proportion results in more events compared to the other CARA designs. This is mainly because the Neyman allocation proportion does not account for any ethical criteria, but it minimizes the overall sample size for a fixed variance of the treatment difference.

Sometimes an experimenter can face situations where treatment $B$ performs better than treatment $A$, or, in situations when comparing a new treatment to a control, the control performs better than the new treatment. The operating characteristics in Table 2.5 show the performance of the competing designs in such a situation.

It can be seen from Table 2.5 that all of the competing designs are fairly powerful in the case of the negative model. Similar to the Positive model, both the CARA and the response adaptive designs result in slightly fewer events compared to the traditional balanced randomization procedures. The average number of events is quite similar across the competing designs. The standard error of the average number of events for almost all of the designs for the increasing hazard scenario is 6, apart from that of design I whose standard error is 7 and that of design XIV whose standard error is 8. For the decreasing hazard scenario, almost all of the designs have a standard error for the average number of events of 7 apart from those of design III and XI whose standard error is 6 and that of design XIV whose standard error is 8. The negative model also shows that, irrespective of the shape of the hazard function, using a CARA

| **Hazard** | **Design** | $\frac{N_A}{n}$ **(SE)** | $N_{AM}$-$N_{BM}$ | $N_{AF}$-$N_{BF}$ | **Event** | **Power** |
|---|---|---|---|---|---|---|
| Decreasing | I | 0.50 (0.025) | 100-100 | 100-100 | 361 | 0.99 |
| | II | 0.50 (0.003) | 100-100 | 100-100 | 361 | 0.99 |
| | III | 0.50 (0.012) | 100-100 | 100-100 | 360 | 0.99 |
| | IV | 0.44 (0.034) | 88-112 | 89-111 | 358 | 0.93 |
| | V | 0.46 (0.030) | 90-109 | 93-108 | 359 | 0.94 |
| | VI | 0.47 (0.028) | 94-106 | 95-105 | 360 | 0.95 |
| | VII | 0.49 (0.027) | 96-104 | 97-103 | 359 | 0.95 |
| | VIII | 0.46 (0.025) | 90-110 | 94-106 | 358 | 0.95 |
| | IX | 0.48 (0.018) | 92-106 | 99-103 | 359 | 0.96 |
| | X | 0.48 (0.014) | 94-106 | 99-101 | 360 | 0.96 |
| | XI | 0.49 (0.007) | 98-102 | 98-102 | 358 | 0.97 |
| | XII | 0.45 (0.033) | 89-111 | 89-111 | 359 | 0.93 |
| | XIII | 0.46 (0.024) | 92-108 | 92-108 | 359 | 0.94 |
| | XIV | 0.33 (0.029) | 67-135 | 66-132 | 328 | 0.90 |
| Increasing | I | 0.50 (0.025) | 100-100 | 100-100 | 351 | 0.99 |
| | II | 0.50 (0.003) | 100-100 | 100-100 | 351 | 0.99 |
| | III | 0.50 (0.012) | 100-100 | 100-100 | 351 | 0.99 |
| | IV | 0.44 (0.036) | 87-113 | 88-112 | 348 | 0.96 |
| | V | 0.45 (0.031) | 90-111 | 91-109 | 348 | 0.93 |
| | VI | 0.47 (0.029) | 93-117 | 94-116 | 348 | 0.98 |
| | VII | 0.48 (0.026) | 96-107 | 93-104 | 348 | 0.95 |
| | VIII | 0.46 (0.029) | 89-111 | 94-106 | 348 | 0.97 |
| | IX | 0.47 (0.023) | 90-110 | 96-114 | 348 | 0.96 |
| | X | 0.48 ( 0.016) | 93-107 | 99-101 | 349 | 0.99 |
| | XI | 0.49 (0.009) | 98-102 | 98-102 | 348 | 0.96 |
| | XII | 0.44 (0.034) | 89-111 | 89-111 | 348 | 0.94 |
| | XIII | 0.46 (0.027) | 92-108 | 91-109 | 348 | 0.94 |
| | XIV | 0.33 (0.029) | 65-135 | 65-135 | 315 | 0.90 |

Table 2.5: Performances of the competing designs in the case of the negative model

randomization procedure results in a skewing of the treatment allocation probabilities in favour of the better treament arm while maintaining a high statistical power of the Wald test. This further shows that even when the response of the patients follow a non-linear Weibull model, using a CARA randomization procedure is ethically more attractive compared to the traditional balanced randomization designs, while it also maintains high statistical efficiency in estimating treatment effects in the presence of covariates. The inverse proportionality between the average power of the Wald test for the covariate-adjusted treatment difference, and the variance of the CARA designs was established theoretically in section 2.5 using a Taylor series expansion of the random non-centrality paramenter of the test. Comparing the CARA designs, it can be seen that the CAERADE or ERADE designs a much less variable than the CACBCD or DBCD counterparts respectively. This is also true for the response-adaptive designs.

The variance of the CARA design based on the glink function also depends on the difference between the two treatment effects. It can be seen in Table 2.5 that this design has a relatively small variance, and thus a fairly high power of the Wald test for detecting the difference in the covariate-adjusted treatment effects. Since all of the CARA designs result in a skewing of the treatment allocation probabilities towards the better treatment and achieving high statistical efficiency in estimating treatment effects in the presence of covariates, they can be considered to be suitable alternatives to the traditional balanced designs. Similar to the positive model, for both increasing and decreasing hazard scenarios, the response-adaptive designs and the CARA design with the glink function have the most variable powers whereas the balanced randomization designs have the least variable power. For the case of decreasing hazard, the variability of the power for design X is similar to those of the balanced randomization designs. In the case of the increasing hazard, the variability of the power for design IV is similar to that of the response-adaptive procedures. All of the standard errors for the powers ranged from 0.001 to 0.004.

### 2.6.4 Distribution of the Allocation Proportions

Apart from observing the performances of the designs on an average level over 5000 simulation runs, it is useful sometimes to learn about the performance of the individual trials. The overall performance in the individual trials for the competing designs in Table 2.2 when the shape parameter of the Weibull model is less than 1 is shown in Figure 2.2.

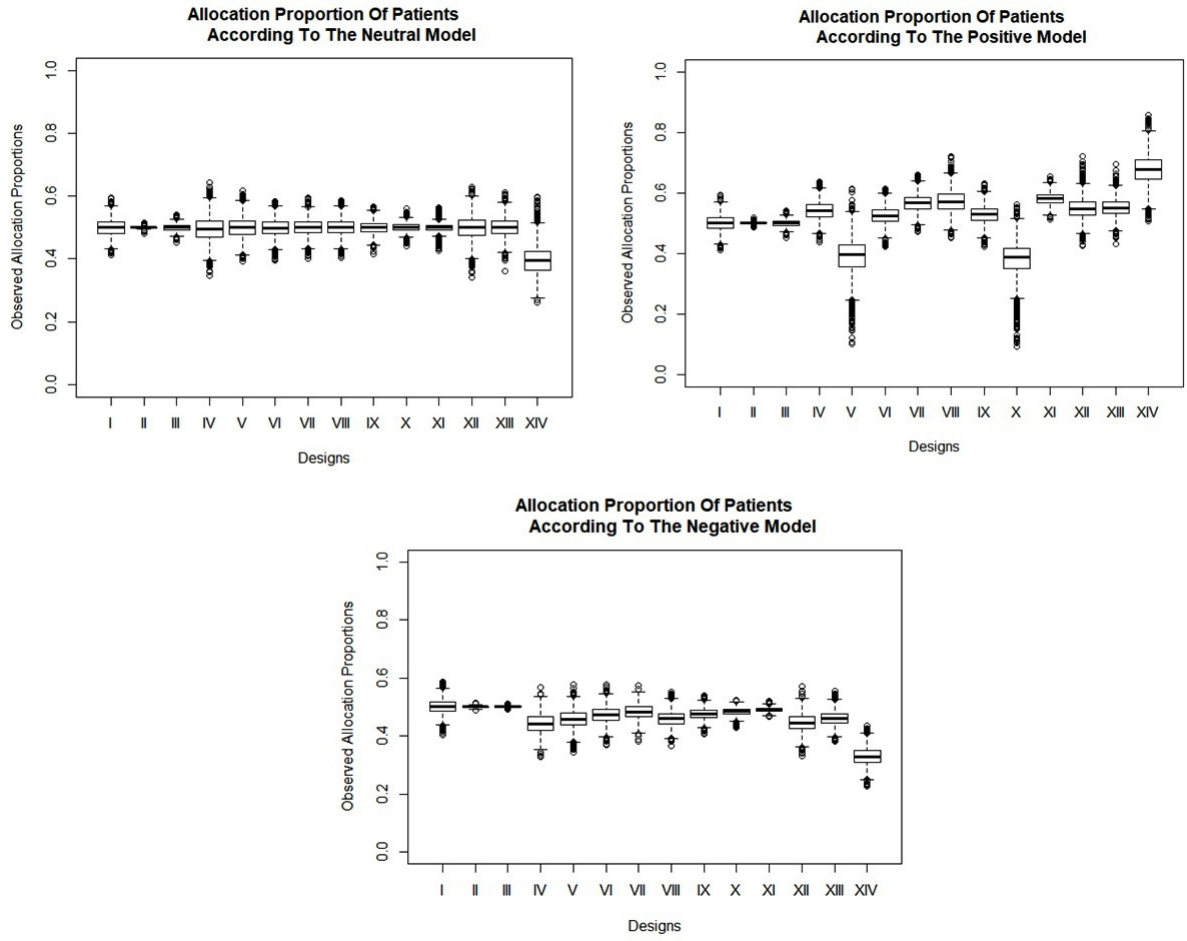


Figure 2.2: Effect Size usually skews CARA allocation proportions towards better treatment

The boxplots in Figure 2.2 depict the observed allocation proportions of the competing randomization procedures for $n = 400$ patients sequentially arriving in the trials. The distributions of the observed allocation proportions appear to be very close to a symmetric distribution, but with different means and with different variability. With

the response-adaptive procedures, the adaptive allocation started somewhat earlier in the trial as compared to the CARA designs, because the former estimate the main treatment effects only, and the latter involves estimation of the full vector of treatment effects. It can be seen that, for all three models, Efron's biased coin design and the Pocock and Simon covariate adaptive randomization procedure are least variable among the competing designs. However, they along with the completely randomized design allocate patients equally between the two treatment arms irrespective of their performance based on patient responses. Therefore, these traditional balanced designs suffer from the disadvantage of allocating more patients to the worse treatment arm during the course of the trial. On the other hand, the CARA as well as the response-adaptive designs skew the patient allocation on an average towards the better-performing treatment arm. The boxplots also confirm the finding that the CAERADE and the ERADE designs are more efficient than the corresponding CADBCD and the DBCD procedures respectively.

The simulation results in this section clearly suggest that, when the survival response of the patients to a treatment has a Weibull distribution, using the CARA designs would significantly skew the allocation probabilities away from balance, but the degree and the direction of the skewness may vary depending on the target allocation proportions that the CARA designs converge to. The variabilities of the designs may also vary. It has been seen that the CAERADE designs are the most efficient and that the CARA design with the glink function is the most ethical among all the other CARA designs. Therefore, when balancing the competing goals of statistical efficiency and of treating more patients with the better performing treatment, the CARA designs outperform the balanced randomization designs.

It was seen in Sverdlov, Rosenberger and Ryeznik (2013) that building the CARA designs assuming exponential regression model for the survival responses, the type I error rates of the Wald test for treatment comparison were slightly inflated. However it is seen from the simulation results here that using the designs of section 2.4 when the responses of the patients follow a Weibull regression model, gives slightly conservative type 1 error rates of the Wald test for treatment difference when the hazard is decreasing

over times. When the hazard rate is increasing over time the they are closer to the nominal value of 0.05 for most of the optimal allocation proportions. This may be because of the fact that when the hazard rate is increasing over time, the density function for the Weibull model is closer to the shape of normality as compared to the exponential density function. Therefore using the correct model for Weibull distributed survival responses helps controlling the type 1 error rate of the Wald test for treatment comparison which is a primary concern in any clinical trial.

This section dealt with simulation studies to validate the performance of the derived CARA designs for patients with Weibull distributed survival responses. It is therefore now worth applying these derived methodologies to a clinical trial in order to observe their performance in the trial. The next section gives an applied outlook to this detailed framework by using a real-life clinical trial.

## 2.7 Real-Life Example of the Proposed Methodology

To assess the performance of the derived methodologies in a real-life clinical trial, a survival trial in pulmonary adenocarcinoma has been re-designed. The study has been previously explored in Sverdlov, Rosenberger and Ryeznik (2013) in terms of CARA designs based on exponential models. Here, the study is re-designed based on the derived methodologies for Weibull distributed survival responses. In this phase 3 open-label study, during a 20-month period, 1217 adult patients from East Asia between the age 18 to 50 were randomly assigned in a 1:1 ratio between gefitinib (treatment $A$) and paclitaxel (treatment $B$). The study excluded all patients who were former smokers. The patients were followed up for a period of 12 months after the treatment phase. The primary objective of this trial was to test if treating patients with gefitinib would increase the time to relapse from pulmonary adenocarcinoma, as compared to those patients who are being treated with paclitaxel . The primary endpoint was progression-free survival which considered relapse from pulmonary adenocarcinoma to be the event of interest, whereas patients who do not experience the relapse until the end of the study

period or if they are lost to follow-up during the study or if they die to to other events are considered to be censored . Epidermal growth factor receptor (EGFR) gene mutation was considered to be one of the significant factor affecting the patients' response. The study showed that there was a significant interaction between the treatments and EGFR. Treatment $A$ was superior to treatment $B$ in the EGFR+ subgroup (hazard rate for progression 0.48) and inferior in the EGFR− subgroup (hazard rate for progression 2.85).

Patient survival times for treatment $k$ were simulated from an exponential distribution with mean $\exp(\beta_{k0} + \beta_{k1}z)$, for $k = A, B$ and $z = 0,1$. Following Sverdlov, Rosenberger and Ryeznik (2013), the parameters were chosen as follows: $R = 20$ months and $D = 26.5$ months; $z = 0$ (EGFR+) with probability 0.6 and $z = 1$ (EGFR−) with probability 0.4; $\beta_{A0} = 1.62$, $\beta_{A1} = 0.98$, $\beta_{B0} = 2.35$, and $\beta_{B1} = 0.80$. For appropriate implementation of the derived CARA designs, 315 patients were initially equally randomized to the two treatment arms using Efron's biased coin design before the adaptive randomization process started. The results for the performances of the designs are presented in Table 2.6.

| **Designs** | $\frac{N_A}{n}$ **(SE)** | $N_{A-}$-$N_{B-}$ | $N_{A+}$-$N_{B+}$ | **Events** | **Power** |
|---|---|---|---|---|---|
| I | 0.50 ( 0.014) | 243-243 | 365-365 | 1208 | 0.99 |
| II | 0.50 (0.001) | 244-244 | 365-365 | 1208 | 0.99 |
| IV | 0.46 (0.019) | 229-499 | 333-156 | 1196 | 0.98 |
| VIII | 0.45 (0.018) | 241-489 | 305-182 | 1197 | 0.99 |
| XII | 0.46 (0.017) | 335-358 | 229-295 | 1205 | 0.97 |
| XIII | 0.47 (0.016) | 320-356 | 251-290 | 1205 | 0.98 |

Table 2.6: CARA Designs Outperforms other Designs for treating individual patients

Table 2.6 summarizes the performances of some of the competing designs in Table 2.2. It can be seen that irrespective of the performance of the treatment arms, the completely randomized design and the Efron's biased coin design allocte patients equally between the two treatment arms, resulting in more events. They also do not take into

account the difference betweens the covariate-adjusted treatment effects within the levels of EGFR. Thus, they allocate equal numbers of patients within both the EGFR subgroups. Due to delayed responses, the convergence to the target allocation proportions was dampened. The average allocation proportions for both the CARA and the response-adaptive designs are similar. This is because of the different direction of the treatment effect in the EGFR+ and EGFR− subgroups. The overall allocation proportion is therefore close to 0.5 for both the CARA and the response-adaptive designs. However the CARA designs IV and VIII, unlike the response-adaptive designs XII and XIII, account for the difference in the direction of the treatment effect in the EGFR subgroups.

One of the primary reasons for developing CARA designs is that, in some clinical trials, the degree and direction of the treatment effect differ for patient subgroups within a treatment arm and the research design should account for such covariate-specific treatment effects. Therefore, within each of the EGFR subgroups, the CARA procedure allocates more patients to the better treatment arm and has, on average, fewer events than each of the response-adaptive randomization procedures and the two balanced designs. The response-adaptive designs also result in a skewed allocation towards treatment $B$, but the degree of skewing is similar across the EGFR subgroups. It can be seen that design VIII is less variable than design IV. Thus, if the sole criterion is to have an ethical design with minimum variability, the CAERADE design is the preferable one. The simulated type I error rates for the designs were between 0.03 and 0.05. All the randomization procedures have similar powers and the standard errors of the averge number of events were no more than 10.

## 2.8 Critical Evaluation

The main objective of the present thesis has been to explore the likely ways of widening the scope of applicability of CARA designs. The procedure of Bandyopadhyay and Biswas (2001) used probit link function for normally distributed responses to relate the past allocation, and covariate and response histories of the patients to the present

allocation. Here, the design was scaled by using an arbitrary tuning parameter to control the variability. This design has all of the major features of CARA randomization designs, except for the fact that, unlike the optimal allocation designs, it does not involve any adjustment for the covariates of the incoming patients. The concomitant information of the incoming patients can be crucial in many cases. For instance, if it relates to the gender of the patients and the patients are being treated with two competing treatments, if males and females react very differently to treatments $A$ and $B$, whether the next patient is male or female is an important element in making the treatment assignment to the incoming patient. It is because of this that the probit link based design cannot be categorized as a CARA design.

The design based on the glink function, however, is a CARA design, as its treatment allocation probabilities are sequentially modified based on the history of previous patients' treatment assignments, responses and covariates, and the covariates of the incoming patient. The scale parameter of the cumulative distribution function of the Gumbel model here is the reciprocal of the Weibull shape parameter obtained from the information about the previous allocation, covariate and response histories of the patients. On the other hand, the location parameter of this cumulative distribution function is the reciprocal of the covariate-adjusted scale parameter obtained from the Weibull regression accelerated life model based on the covariates of the incoming patient. Another important feature of the present CARA design based on the glink function is that, unlike the probit link based design of Bandyopadhyay and Biswas (2001), it does not rely on an arbitrary tuning parameter $\mathbf{T}$ as a scaling factor.

The rationale behind the inclusion of the glink based CARA design has been its close affinity to the other CARA designs in one very significant respect, namely that they are all ethically oriented and are aimed at skewing the patient allocation in favour of the better of the two competing treatments. Although the concept of skewing the treatment allocation arose from ethical considerations such designs are also conducive to minimization of outlay of resources in clinical trials. However, some care is needed from the experimenter before using the CARA design based on the glink function. Since the design is based on the cumulative distribution function of a Gumbel model which

is a light tailed, steeply increasing function, it is ethically very attractive in terms of treating a greater number of patients with the better treatment and achieving fewer events in the trial. Moreover, it gives more weight to the available data than the designs based on the probit link function or the Cauchy distribution function.

Figure 2.3 shows the probabilities of allocation to the better treatment arm for different estimates of the covariate adjusted treatment difference when treatment $A$ performs better treatment $B$ and the survival response follows a Weibull distribution. Here, 110 patients were initially equally randomized between the two treatment arms using Efron's biased coin design, and then adaptive allocation was used separately with each of the four link functions in Table 2.7. The scale parameter for each of these link functions is $\hat{\sigma}_m = 1/\hat{\gamma}_m$ and the location parameter is $\mu = 1/\hat{\mu}(\mathbf{z}_{m+1})$, where $\hat{\gamma}_m$ is the estimate of the Weibull model shape parameter from the previous $m$ allocations and $\hat{\mu}(\mathbf{z}_{m+1})$ is the estimate of the Weibull model scale parameter as a function of the concomitant information of the incoming patient. The covariates considered are the same as those in section 2.6 .

It can be seen from Figure 2.3 that using a light-tailed cumulative distribution function such as the glink leads to a faster increase in the allocation proportions with the covariate-adjusted treatment difference compared to a heavy-tailed counterpart such as the Cauchy distribution. An abrupt-tailed one like the uniform distribution would give over-optimistic skewed allocation because of its abrupt increase in the allocation proportion for every increase in the treatment difference. It can also be seen that the increase in the observed allocation proportions for the design based on the glink is slightly steeper than that for the probit link function and that the cumulative distribution function of the Gumbel model tends to assign slightly more patients to the better-performing treatment arm. The estimated rate of change of the allocation proportions with the covariate-adjusted treatment difference and the proportion of explained variability is summarized in Table 2.7:

The residual degrees of freedom when regressing the allocation proportions against

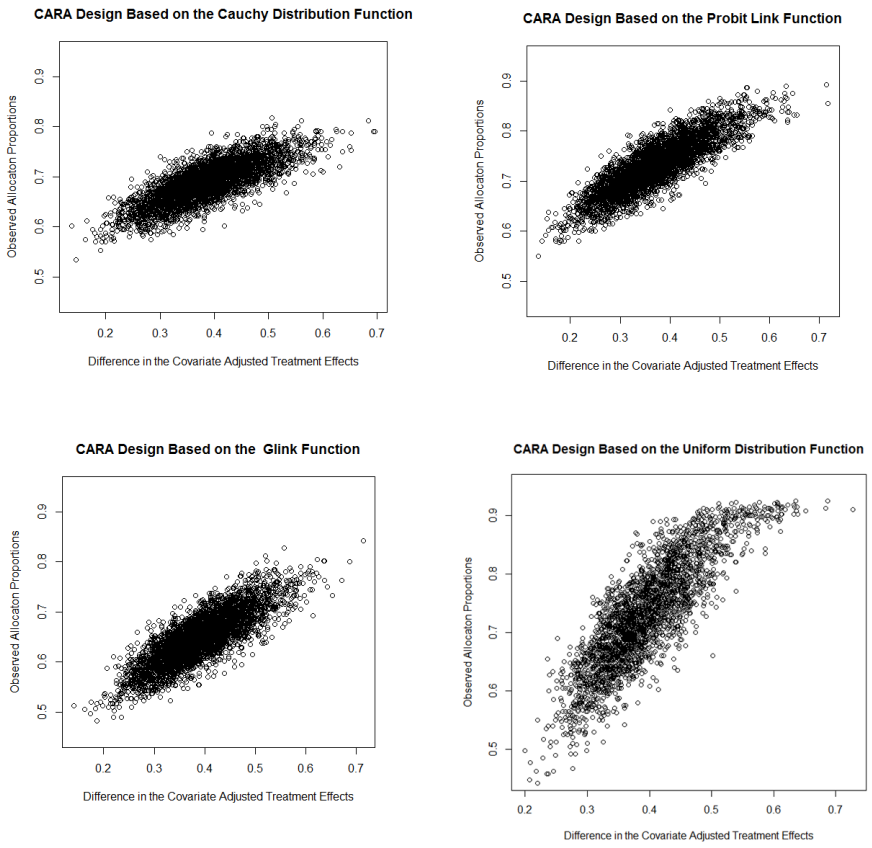


Figure 2.3: Rate of change of the Observed Allocation Proportions

| **Link** | **F(x)** | **Estimated Slope** | **Residual SE** | **$R^2$** |
|---|---|---|---|---|
| Cauchy | $\frac{1}{\pi}\arctan\left(\frac{x-\mu}{\sigma}\right)+\frac{1}{2}$ | 0.41 | 0.025 | 0.61 |
| Gumbel | $\exp\{-e^{(x-\mu)/\sigma}\}$ | 0.57 | 0.029 | 0.68 |
| Probit | $\frac{1}{2}\left\{1+erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right\}$ | 0.55 | 0.027 | 0.73 |
| Uniform | $\frac{x-\sigma}{\mu-\sigma}$ | 1.05 | 0.048 | 0.71 |

Table 2.7: Behaviour of the different link function-based designs

the covariate-adjusted treatment difference for each of the designs based on the link functions in Table 2.7 is 4998. Although the design based on the glink function tends to allocate more patients to the better treatment arm than those based on the other link functions, it has a serious drawback. In the initial stages, one should not rely heavily on the available data. If the first few observations happen to be outliers, more patients

might be allocated to the worse treatment arm than is desirable. It is therefore advisable that, to use the CARA design based on the glink function, the experimenter needs to choose a sufficiently large sample size for the interim stage of the trial during which the patients are balanced across the treatment arms using restricted randomization in order to calculate the maximum likelihood estimates to initiate the adaptation. As suggested by Bandyopadhyay and Biswas (2001), in practice, one may also decide to choose a design based on a heavy-tailed distribution in the initial stages and then switch to a light-tailed one. This is technically a good suggestion, but switching between two different designs for the same trial is less likely to be acceptable. Therefore, it is suggested that, if an experimenter wishes to conduct a clinical trial which is ethically very attractive, he might want to choose a sufficiently large sample size for the interim stage of the trial during which the patients are balanced across the treatment arms with the help of some balanced randomization procedure.

## 2.9 Discussion

The CARA designs here are based on two distinct approaches, the covariate-adjusted doubly adaptive biased coin design (Zhang and Hu 2009) and the covariate-adjusted version of the efficient randomized adaptive design (Hu, Zhang and He 2009) on the one hand, and the glink function. The approach based on the glink function bridges the past allocation and response histories and the present allocation pattern after allowing for the incorporation of prognostic factors. The glink is developed using the cumulative distribution function of a Gumbel model whose location parameter is the reciprocal of the scale parameter of the Weibull accelerated life model calculated from the covariate information of the incoming patient and its scale parameter is the reciprocal of the shape parameter calculated from the Weibull accelerated life model based on the information about the previous patients. The Gumbel model being asymmetric and light-tailed, gives more weight to the available data and tends to allocate more patients to the better treatment. Moreover, when the responses of the patients follow a Weibull distribution, the design based on the Gumbel model is more appropriate compared to

other continuous models because the theoretical errors in the Weibull accelerated life regression model follow a Gumbel distribution. The arbitrariness of choosing a value for the tuning parameter **T** present in the design based on the probit link function of Bandyopadhyay and Biswas (2001) is not present in the one based on the glink function. Scaling the estimated covariate-adjusted treatment difference by its standard error plus the estimated hazard ratio makes this design more applicable.

The operating characteristics of the proposed adaptive designs as well as the balanced randomization designs have been compared through simulation for a two-arm survival trial with three predictive covariates and right-censored data. It has been found that almost all of the proposed CARA designs generate skewed allocations towards the better treatment according to covariate-specific treatment effects, and thus result in fewer events in the trial, without compromising much on the statistical efficiency compared to the balanced randomization designs. The only exceptions to this are the CARA designs targeting the Neyman allocation proportion. This is because the objective function for the Neyman allocation does not address any ethical criteria. Its objective is to minimize the trial size. The degree of skewness also varies according to the background model that the design is based on. A slight delay in the response does affect the convergence of the CARA designs to their target allocation proportions. The skewness in the treatment allocation proportions in favour of the better treatment establishes the ethical gain of using the CARA designs compared to the traditional balanced randomization procedures. It has been established by simulation that such ethical gain is achieved most with the CARA design based on the glink function without compromising much on the power of the Wald test for the covariate-adjusted treatment difference. Using the correct model for adaptation when the responses of the patients follow a Weibull model helps to control the type 1 error rate which was slightly inflated while using the designs based on exponential accelerated life model.

A family of CARA designs has also been proposed here that are fully randomized and asymptotically efficient of the first order. The CAERADE can be regarded as a generalization of the Efron's biased coin design for any desired allocation proportion, which may depend on the unknown parameters. Delayed responses, which are very

common in the case of survival trials, present no logistical difficulty in their incorporation in the CAERADE. It has been established that, when the new treatment has a better effect on the survival experience of the patients than the existing control or when treatment $A$ performs better than treatment $B$, the CARA designs based on the Neyman allocation proportion assignsmore patients to the worse treatment and therefore are not ethically attactive compared to the other competing designs.

A relationship between the non-centrality parameter and the variance of the CARA designs has been arrived at using a Taylor series expansion of the non-centrality parameter of the Wald test for the difference between the covariate-adjusted treatment effects. It is known that the power is an increasing function of the non-centrality parameter of the Wald test and therefore it has been shown that the variance of the CARA designs is inversely proportional to the power of the Wald test for the difference between the treatment effects in the presence of covariates. The CAERADE being the asymptotically most efficient CARA design increases the power compared to the corresponding CADBCD. In the situation where efficiency is critically important, in theory, the CAERADE should be the best choice among all of the CARA randomization procedures. However, sometimes the CAERADE does not converge to the target allocation proportion as fast as the CADBCD does, although its finite-sample variances are always small. This is mainly because the allocation probabilities for the CAERADE are not stable. The allocation function being discrete, they always jump from one value to another. A continuous allocation function like the CADBCD can make the allocation probabilities stable and speed up the convergence of the sample allocation proportions. Since in a clinical trial the subjects are human beings, it would be insensitive for a statistician not to take into serious consideration the need for minimization as far as possible of the proportion of patients receiving the inferior treatment. Ethical considerations would therefore appear to be quite unavoidable. Despite the marginal loss of statistical power of the Wald test, a CARA design would irrefutably outperform any traditional balanced randomization design in terms of ethical considerations, while achieving reasonably high efficiency in estimating covariate adjusted treatment differences.

# Chapter 3

# Covariate-Adjusted Response-Adaptive Designs for Semi-Parametric Survival Models

## 3.1 Introduction

Clinical trials are designed to answer specific questions about biomedical interventions, including new treatments and known interventions that justify further study and comparison. Treatment comparisons for a particular disease are often performed in a clinical trial to obtain information about the efficacy of the competing treatments. However, the involvement of human patients gives rise to an ethical concern of treating as many as possible with the best treatment found so far during the course of the experiment. Patients arrive sequentially and are assigned to one of the competing treatments. In order to make use of this sequential arrival of patients, designs are developed in stages, after each of which a decision is made. Adaptive allocation schemes are sequential designs in which the method of allocation of treatments to patients is modified based on the results obtained in the previous stage until a particular treatment is declared to be a clear winner over the others. There has long been an interest in developing methods that use the accrued information in the course of a clinical trial.

Quite often there exists a trial protocol and a statistical analysis plan (SAP) which is set before the clinical trial begins. The protocol as well as the SAP pre-specifies the adaptation schedule and procedure. The adaptation process generally continues throughout the trial as prescribed in the trial protocol and the SAP. The aim of an adaptive trial is to more quickly and cost-effectively identify drugs or devices that have a therapeutic effect, and to zero in on patient populations for whom the drug is appropriate.

Over the past several decades, there has been an enormous amount of work on developing adaptive designs in clinical trials. Traditionally, interest mainly lay in balancing the patient allocation between the treatment arms while incorporating randomization in the method of allocation. The motivation behind this was to develop a method of experimentation which would retain the maximum power for testing the difference between the treatment effects. However, forcing a sequential experiment to be balanced leads to the problem of its results incurring several forms of bias. These includes selection and accidental bias. Kalish and Harrington (1988) found optimal designs for the special case when two treatments are available. They investigate empirically the loss of efficiency when equal numbers of patients are allocated to each treatment. Efron (1971), as a remedy to this problem, introduced biased coin designs for comparing two treatments in which allocation of the treatment is determined probabilistically, but with a bias towards the underrepresented treatment. Such a method of experimentation which decides on the next treatment allocation based on the information about the allocation history only is widely classified as restricted randomization.

One property of human patients not shared by inbred laboratory animals is that they are heterogeneous, that is, they can differ greatly in their responses to treatment. Such heterogeneity can be accounted for during the course of the clinical trial by considering the covariate information of the incoming patients. One of the disadvantages of Efron's scheme is that it does not include balance over covariates or prognostic factors which may affect the response of the patient to the treatment. This has led to the development of a method called the covariate-adaptive randomization where the current patient is randomized to a treatment arm based on the history of previous treatment assignments, the covariate vectors of past patients and the current patient's covariate vector. The

goal of covariate-adaptive randomization is to adaptively balance the covariate profiles of patients randomized to each of the competing treatments.

Clinical trials involve experimentation on human patients, there is a large regulatory presence in the running of trials. Therefore, it is quite often that, from ethical considerations, the goal of an adaptive design may be to allocate a large number of patients to the treatment performing better thus far in the trial. Designs that adapt to the responses of the previous allocated patients as well as the previous allocation history are termed to be response-adaptive designs.

In contrast, for CARA designs, the treatment allocation probabilities are sequentially modified based on the history of previous patients' treatment assignments, responses and covariates, and the covariates of the incoming patient. Information related to the covariate of the incoming patient is crucial in many cases when deciding on the treatment allocation for this patient. For instance, if males and females react very differently to the two treatment arms, then the information about the incoming patient being a male or a female is an important element in the assignment of a treatment to that patient.

Most of the CARA designs developed so far have dealt with clinical trials where the response of the patients to a particular treatment is considered to be a binary random variable. Hu, Zu and Hu (2015) developed unified family of CARA designs that balances the efficiency and ethical criterias of a clinical trial, using a tuning parameter as the power of the D-optimality criterion. Thier proposed family unifies several well-known randomization methods such as the covariate-adjusted doubly-adaptive biased coin design as given in (2.26) and the optimal biased coin design of Atkinson (1982). There has also been some work carried out with survival outcomes following a certain parametric model. Sverdlov, Rosenberger and Ryeznik (2013) used the exponential parametric model to develop CARA designs for survival trials with administrative right censoring for patients who are sequentially arriving in the trial. Biswas, Bhattacharya and Park (2016) introduced a class of covariate-adjusted response-adaptive designs for phase III clinical trials when the treatment response follows a parametric survival model and there is random censoring. They developed optimal allocation designs for parametric

survival responses based on the Koziol-Green (1976) model of random censoring, where the survivor function of the censoring variable is a positive power of that of the lifetime variable, and hence allows for the risk of experiencing the event of interest to depend on the censoring mechanism. Such a kind of random censoring is referred to as informative censoring.

While using a parametric form for the survival responses results in robust parameter estimates from the fitted statistical models, it is rare in real-life clinical trials that the survival responses conform closely to a certain parametric model. Therefore, an attempt is made in this chapter to develop a CARA design whch is more applicable in real-life situations, and, at the same time, is based on methods which yield sufficiently robust parameter estimates from the fitted statistical model. The CARA designs developed are based on the lighter assumption that the hazard functions of the patients at any given time point are proportional and time independent. It is because of this reason that the approach is termed semi-parametric. To handle such situations, Fidalgo and Lopez (2014) developed a partial information matrix and compared their derived optimal designs with the optimal designs based on the full likelihood information, and Konstantinou, Biedermann and Kimber (2015) obtained optimal designs by finding a closed-form expression for the asymptotic covariance matrix for the Cox model. These do not fit the definition of a CARA design which incorporates the covariate information of the incoming patient. In this chapter, various CARA designs are developed based on a semi-paramtric approach and the performance of the designs is validated by simulation.

The outline of this chapter is as follows. Section 3.2 explains the background material relating to the Cox proportional hazards model. The method of obtaining parameter estimates for the model is discussed in Section 3.3. This is followed by Section 3.4 that proposes the various CARA randomization procedures for a survival trial without any parametric assumptions on the survival responses, which are then validated in Section 3.5 using extensive simulation. The results obtained from applying the proposed CARA designs to re-design a real-life clinical trial are detailed in Section 3.6. Section 3.7 concludes with a discussion and an outline of some future research in this direction.

## 3.2 Background on Proportional Hazards

As mentioned earlier, Sverdlov, Rosenberger and Ryeznik (2013) used the exponential parametric model to develop CARA randomization procedures for survival trials. These have been generalized in Chapter 2 to the Weibull case citing the limitations of the applicability of the exponential model in real-life clinical trials. The methods discussed in this chapter extend the applicability of CARA designs even further by encompassing situations where the designs are suitable for survival responses conforming to any distribution, provided that the hazards of the event considered at any given time point are proportional and time-independent for any two patients in the trial.

In a medical context, the hazard rate is also known as the force of mortality and it represents a continuous version of a death rate per unit time. It is always convenient in survival analysis to describe the distribution of the survival responses in various different but inter-related ways. For a continuously distributed survival time $T$, let $f(t)$ be the density function and $F(t)$ be the distribution function. In addition, let:

- the survivor function be $S(t) = \mathrm{P}(T \geq \mathrm{t})$, where $S(t)$ gives the probability for a patient to survive beyond a given time point $t$;
- the hazard function be $h(t) = \frac{f(t)}{S(t)}$, where $h(t)$ can be interpreted as the instantaneous failure rate;
- the risk function be $h(t)\delta t$, which gives the risk of an event in the time interval $[t, t+\delta t)$, given survival up to time $t$.

The survivor function can also be written as

$$S(t) = e^{-\int_0^t h(u)du} = e^{-H(t)}, \tag{3.1}$$

where $H(t)$ is known as the integrated hazard or the cumulative hazard. It can therefore be seen that, for the distribution of $T$ to be proper, that is, for its density to integrate to one, $H(t) \to \infty$ as $t \to \infty$. If this is not true, the implication is that the individual may never die, though in some contexts this may not be an unreasonable approximating

assumption. For example, when measuring progression-free survival, children may be cured of a childhood tumour and live indefinitely in relation to the time scale of the study. Normally, statisticians would want to insist on the distribution of $T$ being proper.

The hazard function gives the event rate at a given time $t$, conditional on having survived to time $t$. The actual value of the hazard function is not usually of any practical importance. It is mainly used to calculate risk ratios or to compare the risks at different time points in a clinical trial. It is also useful for comparing death or failure rates over time. The hazard function can be used as a means of identifying an appropriate parametric model for the data or ruling out models that are not appropriate. For example, the exponential distribution has a constant hazard function. Therefore, if there is knowledge about the systems that they do not age with time, then the exponential distribution can be used to model the survival times. If the hazard rate is increasing, the risk of death or failure is also increasing with time because the ratio of the hazard rates will be the same as the ratio of the risk functions. This ratio is known as the risk ratio because over a small interval around time $t$ when comparing two treatments $A$ and $B$, we have

$$\frac{h_A(t)}{h_B(t)} = \frac{h_A(t)\delta t}{h_B(t)\delta t}. \tag{3.2}$$

The concept of a risk ratio is used extensively in survival modelling when emphasis lies in comparing different groups. This gives rise to the Lehmann family.

This family, also known as the proportional hazards family, is an important family of distributions in modelling survival times. If $\psi_k$ is an arbitrary constant with respect to time at treatment $k$, the form of the Lehmann-family can be generated by;

$$S_k(t;\psi_k) = \{S(t)\}^{\psi_k}, f_k(t;\psi_k) = \psi_k\{S(t)\}^{\psi_k - 1} f(t),\ h_k(t;\psi_k) = \psi_k h(t) \tag{3.3}$$

for $k = A, B$. It can be used to model the log hazard and is the basis for the important proportional hazards model, where the covariates act additively on the logarithm of the hazard function. In such cases, $\psi_k$ is a function of the model covariates. The exponential distribution and the Weibull distribution with constant shape parameter belong to the Lehmann family. From (2.3) it can be seen that, for the exponential model, the hazard

ratio $\psi_k$ can be obtained by taking the ratio of the covariate-adjusted means for the two treatment arms. For Weibull distributed survival responses with constant shape parameter $\gamma$, the value of $\psi_k$ is estimated as in (2.3). In both these cases, $\psi_k$ is time independent and is just a function of the model covariates.

Cox (1972) had used this concept to provide a semi-parametric approach to model time to event data where the survival experiences of patients in different groups can be compared after adjusting for the effects of other variables which have an effect on the patients' responses. Unlike the accelerated life models which assume a particular parametric distribution for the survival time of the patients, the Cox proportional hazards model does not make any strong assumption about the functional form of the survival times but a lighter assumption about the hazard ratio between two individuals at a particular time point being constant. Since the model makes no assumption about the functional form of the survival time distribution, the parameter estimates are not based on the probability of the observed outcomes given the parameter values. Instead of attempting to construct a full likelihood, Cox (1972) considered the conditional probability that, given that exactly one individual in the risk set $R_i$ with covariate vector $\mathbf{z}_m$ dies at time $t_i$, it is the $m^{th}$ individual that does so. Associated with patient $m = 1, 2, ...., n$ is a vector of baseline covariates $\mathbf{z}_m = (z_{1m}, ....., z_{pm})^T$, the vector of unknown model parameters $\boldsymbol{\beta_k} = (\beta_{k1}, ...., \beta_{kp})^T$ for treatment $k = A, B$, and a risk set $R_i$, which is defined as the set of individuals still at risk at time $t_i$, where $t_i$ is the $i^{th}$ ordered event time.

Let the hazard for the $m^{th}$ individual in the trial with treatment $k$ and covariate vector $\mathbf{z}_m$ be $h_k(t|\mathbf{z}_m) = h(t|\mathbf{z} = 0)\ e^{\boldsymbol{\beta}_k^T \mathbf{z}_m}$, where $h(t|\mathbf{z} = 0)$ denotes the baseline hazard function. Throughout this chapter, it is assumed that the survival responses follow a continuous time model, so that only one event occurs at any one time. Therefore, the conditional probability is given by

$$P(\text{ individual } m \text{ dies in}[t_i, t_i + \delta t]|\text{one death at } t_i) = \frac{h_k(t_i|\mathbf{z}_m)\delta t}{\sum_{l \in R_i} h_k(t_i|\mathbf{z}_l)\delta t}$$

$$\text{or, } P(\text{ individual } m \text{ dies in}[t_i, t_i + \delta t]|\text{one death at } t_i) = \frac{h(t_i|\mathbf{z} = 0)e^{\boldsymbol{\beta}_k^T \mathbf{z}_m}}{\sum_{l \in R_i} h(t_i|z = 0)e^{\boldsymbol{\beta}_k^T \mathbf{z_l}}},$$

which yields

$$P(\text{ individual } m \text{ dies in}[t_i, t_i + \delta t]|\text{one death at } t_i) = \frac{e^{\boldsymbol{\beta}_k^T \mathbf{z}_m}}{\sum_{l \in R_i} e^{\boldsymbol{\beta}_k^T \mathbf{z}_l}}.$$

Thus, the baseline hazard cancels out from the expression. This is the essence of the analysis to evaluate the conditional probability, the hazard at the event times $t_i$ only needs to be considered. The product of these conditional probabilities over all of the ordered event times $t_i$ is termed the partial likelihood, and is given by

$$PL = \prod_{k=A}^{B} \prod_{i=1}^{n_k} \frac{e^{\boldsymbol{\beta}_k^T \mathbf{z}_{\mathbf{m(i)}}}}{\sum_{l \in R_i} e^{\boldsymbol{\beta}_k^T \mathbf{z}_{\mathbf{l}}}}. \tag{3.4}$$

It can be seen from (3.4) that the individual times $t_i$ do not appear in the expression for the partial likelihood. This can be justified by the argument that, in the absence of a parametric form for the hazard, there is no information about its value between successive $t_i$: it could quite possibly be zero. It follows that the partial likelihood is a function of only the ranks of the times and it would be unchanged if the time scale were transformed by any monotonic transformation. The partial likelihood can be thought of as the joint density function of the subjects' ranks in terms of event order if there were no censoring and no tied event times. This means that the functional form of the baseline hazard function is not required. The censoring times do not enter the expression for $PL$ except to the extent that they help to determine the risk set. This is also reasonable: knowing where in an interval an individual was censored conveys no information about the hazard, provided that there exists independent censoring.

This intuitive justification, first proposed by Cox (1972) disguises the fact that $PL$ is not actually a likelihood. Indeed he originally described $PL$ as a conditional likelihood. Subsequent discussion led to a changed perception of the nature of the expression, which is closer to being a marginal likelihood. Cox (1975) was able to recast his method of estimation through what he called partial likelihood.

Ignoring the times at which the events occur leads to a loss of information. Therefore, using partial likelihood for estimation of parameters leads to a little loss of information because it suppresses the actual event times, even though they are known. Cox (1975) studied the properties of partial likelihood and showed that it does indeed

have asymptotic properties that justify treating it as if it were a likelihood. He showed that, even with the loss of information compared with a fully parametric analysis, the partial likelihood acts in a similar manner to the likelihood and has all of the usual properties.This approach is based on sound inferential principles and rigorous proofs showing the consistency and asymptotic normality of the partial likelihood estimator. Tsiatis (1981) demonstrated these large-sample properties. Anderson and Gill (1982) simplified and generalized these results through the use of counting processes. The amount of information lost by ignoring the actual event times is less than what one might expect if the Cox proportional hazards model fits the data well.

Another useful distinction between the accelerated life model and the Cox proportional hazards model is that the intercept is non-identifiable in the latter. The effect of the covariates in a proportional hazards model is to increase or decrease the hazard function by a constant proportion relative to the baseline hazard function. If two treatments are being compared, from (3.3), putting $\log(\psi_k) = \boldsymbol{\beta}_k^T \mathbf{z}_m$, $h_k(t;\psi_k) = h_k(t|\mathbf{z}_m)$ and $h(t) = h(t|\mathbf{z}=0)$ the logarithm of the hazard function for treatment $k$ at any time can be modelled as

$$\log\{h_k(t|\mathbf{z}_m)\} = \log\{h(t|\mathbf{z}=0)\} + \boldsymbol{\beta}_k^T \mathbf{z}_m, \quad h_k(t|\mathbf{z}_m) = e^{\boldsymbol{\beta}_k^T \mathbf{z}_m} h(t|\mathbf{z}=0). \tag{3.5}$$

Therefore, using this model leads to the estimation of the relative risk between subjects and not the absolute risk when the model parameters are estimated.

Although the values of the partial likelihood parameter estimates are interpretable by themselves, the overall survival behaviour of the model cannot be understood without knowing the baseline hazard function. One way to understand the baseline hazard function is to specify it. However, in practice clinicians seldom specify a parametric form for the baseline hazard function, because they seldom have even the faintest idea as to what it might look like. The Cox proportional hazards model offers a neat way to overcome this problem. When calculating the survival probabilities, one would estimate the cumulative baseline hazard using an adaptation of the Nelson estimator attributed to Breslow(1972).

$$\hat{H}_k(t|\mathbf{z}=0) = \sum_{t_i \le t} \frac{d_{ik}}{\sum_{l \in R_i} e^{\boldsymbol{\beta}_k^T \mathbf{z}_l}}, \tag{3.6}$$

where $d_{ik}$ represents the total number of events on treatment $k$ at time $t_i$, for $k = A, B$. This is precisely what is expected from a discrete distribution with $\hat{h}_k(t|\mathbf{z} = 0) = \sum_{t_i \leq t} \frac{1}{\sum_{l \in R_i} e^{\boldsymbol{\beta}_k^T \mathbf{z}_l}}$. From this, the baseline survivor function can be estimated by $\hat{S}_k(t|\mathbf{z} = 0) = e^{-\hat{H}_k(t|\mathbf{z}=0)}$.

Once the baseline survivor function is estimated, a survival curve can be constructed as follows with covariates $\mathbf{z}$:

$$\hat{S}_k(t|\mathbf{z}) = \hat{S}_k(t|z = 0)^{\exp(\boldsymbol{\beta}_k^T \mathbf{z})}. \tag{3.7}$$

In some sense, the discrete estimates $\hat{h}_k(t|\mathbf{z} = 0)$ can be thought of as similar to the maximum likelihood estimate from the full likelihood, provided that it is assumed that the hazard distribution is discrete. To estimate the average survival for a group of subjects, either the individual survival estimates can be averaged or the survival for a subject can be calculated using (3.7) with average covariates.

In this chapter, a survival trial is considered where patients enter the trial sequentially and must be immediately randomized to either of the treatment arms. In survival trials, it is not always possible to observe a random sample of completed observations. This is because the observations are often censored in real-life clinical trials. It is assumed that the survival time $T_{ik}$ is subject to an independent censoring time $C_{ik}$ , and that the observed response on treatment $k$ is a bivariate random vector ($\mathcal{T}_{ik}$, $\delta_{ik}$ ), where $\mathcal{T}_{ik} = \min(T_{ik}, C_{ik})$ and

$$\delta_{ik} = \begin{cases} 0 \text{ if } \mathcal{T}_{ik} \text{ is a right-censored time,} \\ 1 \text{ if an event occured at time } \mathcal{T}_{ik}. \end{cases} \tag{3.8}$$

With such a model, the $i^{th}$ patient's observed response with treatment $k = A, B$ is represented by ($t_{iA}$, $\delta_{iA}$ ) when $x_i = 1$ and ($t_{iB}$, $\delta_{iB}$ ) when $x_i = 0$. Throughout this chapter, the censoring scheme is assumed to be generalized type I right ceonsoring. The trial has a limited recruitment period of length $R > 0$ and the trial duration is fixed at $D > R$. At time $D$, patients who have not died or have not already been censored are considered to be generalized type I right censored. It is possible to facilitate CARA randomization only if the recruitment phase is relatively long and the number of accumulating survival responses during the recruitment phase is substantial. The

derivation of the parameter estimates using the partial likelihood function is considered in the next section.

## 3.3 Deriving the Parameter Estimates from Partial Likelihood

The partial likelihood function is a joint density function for the ranks of the patients in terms of the event order. It is formed by taking a product of the conditional probabilities that the $m^{th}$ patient, experiences the event in the interval $[t_i, t_i + \delta t)$, given that there is only one event at time $t_i$, over all of the event times. Note that the censoring times do not contribute towards the partial likelihood, as the values of probabilities are 1. Therefore, the partial likelihood for treatment $k$ expression can be written as

$$PL_k = \prod_{i=1}^{n_k} \left\{ \frac{e^{\boldsymbol{\beta}_k^T \mathbf{z}_{m(i)}}}{\sum_{l \in R_i} e^{\boldsymbol{\beta}_k^T \mathbf{z}_l}} \right\}^{\delta_{ik}}. \tag{3.9}$$

The essence of using the partial likelihood is that this function depends only on $\beta_k$, and is free of the baseline hazard function $h(t|\mathbf{z} = 0)$. Cox (1975) suggested treating $PL_k$ as a regular likelihood function and making inferences about $\beta_k$ accordingly. This means that the partial likelihood function can be maximized to obtain maximum partial likelihood estimate (MPLE) of $\boldsymbol{\beta}_k$, and also the negative partial second derivative of the log partial likelihood function can be used as the observed information matrix for calculating the variance of the MPLE.

Survival trials are likely to involve censored observations. A critical assumption made here is that the survival times and the censoring times are independent. The generalized type I right censoring is considered here because the patients arrive sequentially in the clinical trial and are observed until the end of the study. When subjects join a study at different times and are all observed until a fixed time, generalized type I right censoring. Here, time is measured from a different origin for each subject. This is shown in Figure 2.1.

Taking the logarithm of both sides of (3.9) yields

$$l_k(\boldsymbol{\beta}_\mathbf{k}) = \log(\mathrm{PL}_k) = \sum_{i=1}^{n_k} \delta_{ik}\{\boldsymbol{\beta}_k^T \mathbf{z}_{m(i)} - \log(\sum_{l \in R_i} e^{\boldsymbol{\beta}_k^T \mathbf{z}_l})\}. \tag{3.10}$$

This function can be maximized over $\boldsymbol{\beta}_k$ to produce the maximum partial likelihood estimates of the model parameters. Therefore, the partial score function is;

$$\nabla l_k(\boldsymbol{\beta}_\mathbf{k}) = \sum_{i=1}^{n_k} \delta_{ik} \left\{ \mathbf{z}_{m(i)} - \frac{\sum_{l \in R_i} \mathbf{z}_l e^{\boldsymbol{\beta}_k^T \mathbf{z}_l}}{(\sum_{l \in R_i} e^{\boldsymbol{\beta}_k^T \mathbf{z}_l})} \right\}. \tag{3.11}$$

The Hessian matrix of the partial log-likelihood is given by

$$\nabla^2 l_k(\boldsymbol{\beta}_\mathbf{k}) = -\sum_{i=1}^{n_k} \delta_{ik} \left\{ \frac{\sum_{l \in R_i} \mathbf{z}_l \mathbf{z}_l^T e^{\boldsymbol{\beta}_k^T \mathbf{z}_l}}{\sum_{l \in R_i} e^{\boldsymbol{\beta}_k^T \mathbf{z}_l}} - \frac{(\sum_{l \in R_i} \mathbf{z}_l e^{\boldsymbol{\beta}_k^T \mathbf{z}_l})(\sum_{l \in R_i} \mathbf{z}_l e^{\boldsymbol{\beta}_k^T \mathbf{z}_l})^T}{(\sum_{l \in R_i} e^{\boldsymbol{\beta}_k^T \mathbf{z}_l})^2} \right\}. \tag{3.12}$$

Using the score function and the Hessian matrix, the partial likelihood function can be maximized with the help of the Newton-Raphson algorithm. The negative inverse of the Hessian matrix, evaluated at $\hat{\boldsymbol{\beta}}_k$, the partial likelihood estimate of $\boldsymbol{\beta}_k$, can be used as an approximate covariance matrix for the estimate, and used to produce approximate standard errors for the estimated regression coefficients in the Cox proportional hazards model.

Let

$$\bar{\mathbf{z}}(t_i, \boldsymbol{\beta}_k) = \frac{\sum_{l \in R_i} \mathbf{z}_l e^{\boldsymbol{\beta}_k^T \mathbf{z}_l}}{\sum_{l \in R_i} e^{\boldsymbol{\beta}_k^T \mathbf{z}_l}} = \sum_{l \in R_i} \mathbf{z}_l w_l,$$

where $w_l = e^{\boldsymbol{\beta}_k^T \mathbf{z}_l} / \sum_{l \in R_i} e^{\boldsymbol{\beta}_k^T \mathbf{z}_l}$ is the weight that is proportional to the hazard of the patient experiencing the event. Therefore, $\bar{\mathbf{z}}(t, \boldsymbol{\beta}_k)$ can be interpreted as the weighted average of the covariate vectors among those individuals still at risk at time $t$ with weights $w_l$. Thus (3.12) can be written as

$$\nabla^2 l_k(\boldsymbol{\beta}_k) = -\sum_{i=1}^{n_k} \delta_{ik} [\sum_{k \in R_i} \mathbf{z}_l \mathbf{z}_l^T w_l - \bar{\mathbf{z}}(t_i, \boldsymbol{\beta}_k)\{(\bar{\mathbf{z}}(t_i, \boldsymbol{\beta}_k)\}^T], \tag{3.13}$$

which can also be written as

$$\nabla^2 l_k(\boldsymbol{\beta}_k) = -\sum_{i=1}^{n_k} \delta_{ik} [\sum_{l \in R_i} \{\mathbf{z}_l - \bar{\mathbf{z}}(t_i, \boldsymbol{\beta}_k)\}\{\mathbf{z}_l - \bar{\mathbf{z}}(t_i, \boldsymbol{\beta}_k)\}^T w_l]. \tag{3.14}$$

The quantity $V(t_i, \boldsymbol{\beta}_k) = \sum_{l \in R_i} \{\mathbf{z}_l - \bar{\mathbf{z}}(t_i, \boldsymbol{\beta}_k)\}\{\mathbf{z}_l - \bar{\mathbf{z}}(t_i, \boldsymbol{\beta}_k)\}^T w_l$ can be interpreted as the weighted covariance matrix of the covariates among those individuals still at risk

at time $t$. If $\{V_{ss}(t_i, \boldsymbol{\beta}_k)\}_{s=1}^{p}$ are the diagonal entries of this covariance matrix then $V_{ss}(t_i, \boldsymbol{\beta}_k) \geq 0$ and $\frac{\partial^2 l_k}{\partial \beta_{ks}^2} \leq 0$. Therefore, the log partial likelihood function has a unique maximizer which can be obtained by equating the score function to zero and solving for $\boldsymbol{\beta}_k$.

Moreover,

$$V_{st}(t_i, \boldsymbol{\beta}_k) = \left[ \frac{\sum_{l \in R_i} \mathbf{z}_{sl} \mathbf{z}_{tl} e^{\boldsymbol{\beta}_k^T \mathbf{z}_l}}{\sum_{l \in R_i} e^{\boldsymbol{\beta}_k^T \mathbf{z}_l}} - \left\{ \frac{(\sum_{l \in R_i} \mathbf{z}_{sl} e^{\boldsymbol{\beta}_k^T \mathbf{z}_l})(\sum_{l \in R_i} \mathbf{z}_{tl} e^{\boldsymbol{\beta}_k^T \mathbf{z}_l})}{(\sum_{l \in R_i} e^{\boldsymbol{\beta}_k^T \mathbf{z}_l})^2} \right\} \right]$$

can be interpreted as the weighted sample covariance between the $s^{th}$ and the $t^{th}$ elements of the covariate vector among those individuals at risk at time $t_i$. So the weighted covariance matrix of the covariates among those individuals at risk at time $t_i$, which consists of $\{V_{ss}(t_i, \boldsymbol{\beta}_k)\}_{s=1}^{p}$ as its diagonal entries and $\{V_{st}(t_i, \boldsymbol{\beta}_k)\}_{s,p=1}^{p}$ as its off-diagonal entries, is positive definite. Hence the observed information matrix for $\beta_k$, given by $J_k(\boldsymbol{\beta}_k) = \sum_{i=1}^{n_k} \delta_{ik} V(t_{(i)}, \beta_k)$, is also positive definite. The Hessian matrix - $J_k(\beta_k)$ is negative definite, which implies that the log partial likelihood is a concave function of $\boldsymbol{\beta}_k$ and hence has a unique maximum which can be obtained by setting $\nabla \mathbf{l}_k = 0$ . The maximizing estimate $\hat{\boldsymbol{\beta}}_k$ defines the MPLE of $\boldsymbol{\beta}_k$.

Using the martingale central limit theorem, Cox (1975) showed that

$$(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k) \xrightarrow{d} N_p\{0, J_k^{-1}(\boldsymbol{\beta}_k)\}.$$

In practice, since $\beta_k$ is unknown, $\hat{\beta}_k$ is substituted for $\beta_k$ in $J^{-1}(\boldsymbol{\beta}_k)$ and $J^{-1}(\hat{\boldsymbol{\beta}}_k)$ is used as the estimated covariance matrix of $\hat{\boldsymbol{\beta}}_k$. Since $J(\boldsymbol{\beta}_k)$ is positive definite, its unique inverse exists and is also positive definite.

To test the difference between the covariate-adjusted treatment effects, the Cox proportional hazard model uses the score function in (3.11). Let $\mathrm{u}_k(\beta_{ks}) = \frac{\partial l_k}{\partial \beta_{ks}}$ and let $\mathrm{U}_k(\boldsymbol{\beta}_k) = \nabla \mathbf{l}_k$ be the score function for the $k^{th}$ treatment arm. If $\beta_{ks}$ measures the difference between the covariate adjusted treatment effects, then, under the null hypothesis $H_0$: $\beta_{ks} = 0$, we have $\mathrm{u}_k(0) \xrightarrow{d} \mathrm{N}\{0, J_{kss}(0)\}$. Equivalently,

$$\left\{ \frac{\mathrm{u}_k(0)}{\sqrt{J_{kss}(0)}} \right\}^2 \xrightarrow{d} \chi_1^2.$$

Under the global null hypothesis $H_0 : \boldsymbol{\beta}_k = \mathbf{0}$, the score function $\mathrm{U}_k(0)$ converges to a multivariate normality and $\mathrm{U}_k(\mathbf{0}) \xrightarrow{d} N_p(\mathbf{0},\, J_k(\mathbf{0}))$. Equivalently,

$$[\{\mathrm{U}_k(\mathbf{0})\}J_k^{-1}(\mathbf{0})\{\mathrm{U}_k(\mathbf{0})\}^T] \xrightarrow{d} \chi_p^2.$$

Here, $\mathrm{u}_k(0)$ can be regarded as the difference between the sum over the survival times of the observed number of events from treatment $k$ and the expected number of events under the null hypothesis, which is also the numerator of the log-rank test. However, unlike the score function, the log rank test does not adjust for the effects of other variables having an effect on the survival experience of the patients.

The partial likelihood ratio test can also be performed to test for the difference between the covariate adjusted treatment effects. As in the ordinary likelihood theory, the null hypothesis can be regarded as $H_0$: $\boldsymbol{\beta}_k = \boldsymbol{\xi}_k$, where $\boldsymbol{\xi}_k$ is an arbitrary known constant vector of dimension $p$ for treatment $k$. If $H_0$ is true, then $\hat{\boldsymbol{\beta}}_k$, should be close to $\boldsymbol{\xi}_k$. Hence, $l_k(\hat{\boldsymbol{\beta}}_k)$ should be close to $l_k(\boldsymbol{\xi}_k)$ . Since $l_k(\hat{\boldsymbol{\beta}}_k)$ - $l_k(\boldsymbol{\xi}_k)$ is always non-negative, $H_0$ should be rejected if this difference is large. The partial likelihood ratio test uses the fact that, under $H_0$,

$$2\{l_k(\hat{\boldsymbol{\beta}}_k) - l_k(\boldsymbol{\xi}_k)\} \xrightarrow{d} \chi_p^2.$$

Therefore, for a given level of significance $\alpha$, $H_0$: $\boldsymbol{\beta}_k = \boldsymbol{\xi}_k$ is rejected if 2 $\{l_k(\hat{\boldsymbol{\beta}}_k)$ - $l_k(\boldsymbol{\xi}_k)\} \geq \chi_{p,\alpha}^2$, where $\chi_{p,\alpha}^2$ is the value of $\chi_p^2$ distribution such that $\mathrm{P}(\chi_p^2 > \chi_{p,\alpha}^2) = \alpha$.

Taylor expanding $l_k(\boldsymbol{\xi}_k)$ about $\hat{\boldsymbol{\beta}}_k$ gives

$$l_k(\boldsymbol{\xi}_k) \approx l_k(\hat{\boldsymbol{\beta}}_k) + \nabla l_k(\hat{\boldsymbol{\beta}}_k)(\boldsymbol{\xi}_k - \hat{\boldsymbol{\beta}}_k) + \tfrac{1}{2!}(\boldsymbol{\xi}_k - \hat{\boldsymbol{\beta}}_k)^T\nabla^2 l_k(\hat{\boldsymbol{\beta}}_k)(\boldsymbol{\xi}_k - \hat{\boldsymbol{\beta}}_k)$$

Since $\hat{\boldsymbol{\beta}}_k$ maximizes $l_k(\boldsymbol{\beta}_k)$ $\mathrm{U}_k(\hat{\boldsymbol{\beta}}_k) = \nabla l_k(\hat{\boldsymbol{\beta}}_k) = 0$ and $\nabla^2 l_k(\hat{\boldsymbol{\beta}}_k)$ = - $J_k(\hat{\boldsymbol{\beta}}_k)$ we have,

$$2\{l_k(\hat{\boldsymbol{\beta}}_k) - l(\boldsymbol{\xi}_k)\} \approx (\boldsymbol{\xi}_k - \hat{\boldsymbol{\beta}}_k)^T J_k(\hat{\boldsymbol{\beta}}_k)(\boldsymbol{\xi}_k - \hat{\boldsymbol{\beta}}_k).$$

Moreover, using the martingale central limit theorem, it can be shown that

$$(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\xi}_k) \xrightarrow{d} N_p\{\mathbf{0}, J_k^{-1}(\hat{\boldsymbol{\beta}}_k)\}.$$

Therefore under $H_0 : \beta_k = \boldsymbol{\xi}_k$,

$$2\{l_k(\hat{\boldsymbol{\beta}}_k) - l_k(\boldsymbol{\xi}_k)\} = (\hat{\boldsymbol{\beta}}_k - \boldsymbol{\xi}_k)^T J_k(\hat{\boldsymbol{\beta}}_k)(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\xi}_k) \xrightarrow{d} \chi_p^2.$$

The score test is the most powerful test when the true value of $\boldsymbol{\beta}_k$ is close to $\boldsymbol{\xi}_k$. The main advantage of the score test is that it does not require an estimate of the information matrix under the alternative hypothesis or unconstrained maximum partial likelihood. This constitutes a potential advantage in comparison to other tests, such as the Wald test and the partial likelihood ratio test, and makes testing feasible when the unconstrained MPLE is a boundary point in the parameter space. The score test based on the partial likhelihood function and the Wald test are used in the following section to develop CARA designs based on Cox propotional hazard model for comparing two treatment arms.

## 3.4 The Proposed Semi-Parametric CARA Designs

Clinical trials are complex experiments on humans with multiple and often competing expetimental objectives. Some of these objectives include maximizing the power to detect clinically relevent differences in treatment outcomes, maximizing the patient's personal experience while being treated in the trial and making the trial economically more effective. An optimal solution for allocating patients to the competing treatment arms is a function of the unknown parameters. Unlike inbred laboratory animals, the humans involved in clinical trials are heterogeneous. Therefore, the objective of maximizing the patient's personal experience while being treated in the trial needs to be satisfied after taking such heterogeneity into account. A CARA design achieves most of these objectives in a clinical trial after taking the between-patient heterogeneity into account.

CARA randomization is applicable when the responses of the patients to the treatments follow a non-linear and heteroscedastic model, and when multiple experimental objectives are being pursued in the clinical trial. In this chapter, it is assumed that the survival responses of the patients follow a semi-parametric survival model which in this case is a Cox proportional hazard model. This means that there is no assumption made about the underlying theoretical distribution of the responses. However it is as-

sumed that the hazard of an event at any time in the clinical trial for a patient in one treatment group is proportional to the hazard of the event at that particular point of time in the trial for another patient in the opposite treatment group. Proportionality here means that the ratio of the hazard functions for patients on the two treatment arms at a given point of time in the trial is not dependent on time. The censoring scheme assumed throughout is a generalized type I right censoring scheme, where the recruitment period is of length $R > 0$ and $D$ is the overall duration of the clinical trial. At time $D$, the subjects who have not experienced the event or have not already been right censored are considered to be generalized right censored of type I. When the recruitment phase of the survival trial is long enough to accumulate a substantial number of responses, CARA randomization is applicable. To allocate an incoming patient to a particular treatment arm, apart from the covariate profile of this patient, this type of randomization relies heavily on the response history, treatment allocation history and the history of the covariate profiles of the patients.

When comparing two treatments, let $\boldsymbol{\beta}_A$ and $\boldsymbol{\beta}_B$ be the covariate-adjusted treatment effects of treatments $A$ and $B$, respectively. During the initial phase of the trial, some restricted randomization procedure is used to randomize the initial $2m_0$ patients equally among the two treatment arms, where $m_0$ is a positive integer. This ensures that at least $m_0$ patients are allocated to each treatment arm, and that estimates of the parameters $(\boldsymbol{\beta}_A,\ \boldsymbol{\beta}_B)$ can be obtained from the observed responses of this initial sample. At a general stage, one computes the partial likelihood estimates $(\hat{\boldsymbol{\beta}}_{A,m},\ \hat{\boldsymbol{\beta}}_{B,m})$ of the model parameters $(\boldsymbol{\beta}_A\ ,\ \boldsymbol{\beta}_B)$. When the $(m+1)^{th}$ patient enters the trial with covariate vector $\mathbf{z}_{m+1}$, this patient is randomized to treatment $A$ with probability $\mathrm{c}(\hat{\boldsymbol{\beta}}_{A,m},\ \hat{\boldsymbol{\beta}}_{B,m},\mathbf{z}_{m+1})$, where $0 \leq c(.) \leq 1$ is an allocation function which bridges the past allocation pattern, response histories and covariate vectors of the $m$ patients to the $(m+1)^{th}$ allocation with the covariate vector $\mathbf{z}_{m+1}$. This allocation function skews the treatment allocation probability in favour of the better treatment arm found thus far in the course of the clinical trial, without compromising much on the power of the statistical test to detect any covariate adjusted treatment differences.

### 3.4.1 Deriving a Suitable Semi-Parametric Target Allocation Proportion

Treating the baseline hazard as arbitrary makes the design more dependent on the observed data compared to the designs based on parametric models. Such a design therefore increases its applicability in real-life clinical trials. Let $\epsilon_k(\mathbf{z};\boldsymbol{\beta}_k)$ be the probability of an event before censoring for a patient on treatment $k$ and with covariate vector $\mathbf{z}$. Then, $\epsilon_k(\mathbf{z};\boldsymbol{\beta}_k) = \mathrm{P}(T_k \leq C_i \ |\mathbf{z};\ \boldsymbol{\beta}_k)$. This probability can be estimated by the proportion of patients facing events on each treatment.

One way to meet most of the multiple experimental objectives in a clinical trial is to minimize the overall hazard for a patient with a given covariate subject to the constraint of keeping the asymptotic variance of the difference between the estimated hazard functions for the two treatment groups, constant. This is achieved by minimizing

$$n_A h_A(t|\mathbf{z}) + n_B h_B(t|\mathbf{z}),$$

$$\text{subject to } \ \mathbf{z}^T\{\mathbf{z}^T J_A^{-1}(\boldsymbol{\beta}_A)\mathbf{z}\}ze^{2\boldsymbol{\beta}_A^T\mathbf{z}} + \mathbf{z}^T\{\mathbf{z}^T J_B^{-1}(\boldsymbol{\beta}_B)\mathbf{z}\}\mathbf{z}e^{2\boldsymbol{\beta}_B^T\mathbf{z}} = \mathbf{k} > 0,$$

where $\mathbf{k}>0$ is a constant and $J_k(\boldsymbol{\beta}_k)$ is the observed information matrix of the Cox regression coefficients $\boldsymbol{\beta}_k$ for treatment $k$. If $Va_k\{\widehat{h}_k(t|\mathbf{z})\} = \mathbf{z}^T\{\mathbf{z}^T\bar{V}^{-1}(t,\boldsymbol{\beta}_k)\mathbf{z}\}\mathbf{z}e^{2\boldsymbol{\beta}_k^T\mathbf{z}}$ , the optimal allocation proportion for treatment $A$ is given by

$$\pi_{A1}^S(\boldsymbol{\beta}_A,\boldsymbol{\beta}_B,\mathbf{z}) = \frac{\sqrt{\epsilon_B(\mathbf{z};\boldsymbol{\beta}_B)h_B(t|\mathbf{z})Va_A\{\widehat{h}_A(t|\mathbf{z})\}}}{\sqrt{\epsilon_B(\mathbf{z};\boldsymbol{\beta}_B)h_B(t|\mathbf{z})Va_A\{\widehat{h}_A(t|\mathbf{z})\}} + \sqrt{\epsilon_A(\mathbf{z};\boldsymbol{\beta}_A)h_A(t|\mathbf{z})Va_B\{\widehat{h}_B(t|\mathbf{z})\}}}. \tag{3.15}$$

One can use other metrics of treatment difference and obtain different optimal allocations. For instance, minimizing the overall sample size subject to the constraint of keeping the asymptotic variance of the difference between the estimated hazard functions for the two treatment groups, constant, leads to the Neymann allocation given by

$$\pi_{A2}^S(\boldsymbol{\beta}_A,\boldsymbol{\beta}_B,\mathbf{z}) = \frac{\sqrt{\epsilon_B(\mathbf{z};\boldsymbol{\beta}_B)Va_A\{\widehat{h}_A(t|\mathbf{z})\}}}{\sqrt{\epsilon_B(\mathbf{z};\boldsymbol{\beta}_B)Va_A\{\widehat{h}_A(t|\mathbf{z})\}} + \sqrt{\epsilon_A(\mathbf{z};\boldsymbol{\beta}_A)Va_B\{\widehat{h}_B(t|\mathbf{z})\}}}. \tag{3.16}$$

The derivation of the allocation proportion in (3.16) is provided in Appendix E. The variance of the estimated hazard for a particular treatment arm has been derived by applying the delta method. This derivation is given in Appendix C.

Apart from using the formal optimization procedure, one can make use of the cumulative distribution function of a normal model to obtain an allocation proportion. This idea was first introduced by Bandyopadhayay and Biswas (2001), where they used a probit model as a link function to bridge the past history to the present allocation pattern for patients whose responses follow a normal linear model. However, that design did not fit the definition of a CARA design, as it did not take into account the covariate information of the incoming patient. Moreover, the main criticism which the design faced was of the arbitrariness of the user defined tuning parameter used to control the variability of the design. In the results obtained in the previous chapter it was seen that the increase in the allocation proportions with every increase in the treatment difference did not significantly differ between the probit link function and the glink function. Moreover unlike the Weibull accelerated failure time model whose theoretical errors follow a Gumbel distribution, the baseline hazard function for the Cox proportional hazard model is a nuisance parameter whose paramteric distribution is not estimable. A similar design to Bandyopadhayay and Biswas (2001) is therefore now developed for patients whose survival responses follow a semi-parametric model and uses the partial likelihood estimators instead of the least squares estimators.

Let $\Phi(.)$ be the probit link function, which is the cumulative distribution function of a normal model with mean zero and standard deviation determined by the covariates of the incoming patient. Let $\hat{\beta}_{A_{(m)}}$ and $\hat{\beta}_{B_{(m)}}$ be the partial likelihood estimators of the effects of the two treatments upto the $m^{th}$ stage of the clinical trial. Using $\Phi(.)$ it is intended to seek for a suitable cumulative distribution function $F(.)$ which can be used as a link function such that after $m$ allocations, the $(m+1)^{th}$ patient is allocated to treatment $A$ with probability $F[\log\{h_A(t|\mathbf{z}_m)\}-\log\{h_B(t|\mathbf{z}_m)\}]$ and to treatment $B$ with probability 1- $F[\log\{h_A(t|\mathbf{z}_m)\} - \log\{h_B(t|\mathbf{z}_m)\}] = \Phi[\log\{h_B(t|\mathbf{z}_m)\} - \log\{h_A(t|\mathbf{z}_m)\}]$. The probit link function being an increasing function of the treatment effect difference make this allocation procedure favour the treatment performing better at each stage of

the clinical trial. Let $\widehat{HR(\mathbf{z}_m)}$ represent the estimated hazard ratio of the two treaments at the $m^{th}$ stage of the clinical trial. The allocation function for treatment $A$ for the $(m+1)^{th}$ patient can therefore be written as :

$$\pi^S_{A3}(\hat{\beta}_{A_{(m)}}, \hat{\beta}_{B_{(m)}}, \mathbf{z}) = \Phi\left\{ \frac{\log\{\hat{h}_A(t|\mathbf{z}_m)\} - \log\{\hat{h}_B(t|\mathbf{z}_m)\}}{SE[\log\{\hat{h}_A(t|\mathbf{z}_m)\} - \log\{\hat{h}_B(t|\mathbf{z}_m)\}] + \widehat{HR(\mathbf{z}_m)}} \right\}. \quad (3.17)$$

This design scales the covariate adjusted treatment difference by the hazard ratio between the two treatment arms. This means that, if the hazard of an event for a particular treatment group is greater than that for the other, there would be less chance of allocating the next patient to that particular treatment arm. Let $\beta_{-k}$ be the model parameters ignoring the treatment effect. The probit link function in (3.17) has mean zero and the standard deviation as $h(t|\mathbf{z}_{m+1}) = \exp(\beta^T_{-k}\mathbf{z}_{m+1})$.

Let $\mathcal{X}_m$, $\mathcal{Y}_m$,$\mathcal{Z}_m$ and $\mathbf{z}_{m+1}$ denote, respectively, the past allocation history, responses history, the prognostic factors for the first $m$ patients and the covariates of the incoming patient. Based on the probit link, the conditional probability that the $(m+1)^{th}$ patient with covariate vector $\mathbf{z}_{m+1}$, will be assigned to treatment $A$ is given by

$$P(X_{(m+1)} = 1|\mathcal{X}_m, \mathcal{Y}_m, \mathcal{Z}_m, \mathbf{z}_{m+1}) = \Phi\left\{ \frac{\log\{\hat{h}_A(t|\mathbf{z}_m)\} - \log\{\hat{h}_B(t|\mathbf{z}_m)\}}{SE[\log\{\hat{h}_A(t|\mathbf{z}_m)\} - \log\{\hat{h}_B(t|\mathbf{z}_m)\}] + \widehat{HR(\mathbf{z}_m)}} \right\}, \quad (3.18)$$

where, conditional on $\chi_m$,

$$[\log\{\hat{h}_A(t|\mathbf{z}_m)\} - \log\{\hat{h}_B(t|\mathbf{z}_m)\}] \xrightarrow{d} N([\log\{h_A(t|\mathbf{z}_m)\} - \log\{h_B(t|\mathbf{z}_m)\}], \sigma^2).$$

as $m \to \infty$. The variance $\sigma^2$ is calculated from inverting the observed information matrix given by (3.14). Let $\zeta(m) = \mathrm{P}\ (X_{m+1} = 1)$. Then the sequence $\{\zeta(m) : m \geq (2m_0 + 1)\}$ converges to $\Phi\left[\frac{\log\{h_A(t|\mathbf{z}_m)\} - \log\{h_B(t|\mathbf{z}_m)\}}{HR(\mathbf{z}_m)}\right]$. The limiting proportion of the allocation of patients on treatment $A$ is also $\Phi\left[\frac{\log\{h_A(t|\mathbf{z}_m)\} - \log\{h_B(t|\mathbf{z}_m)\}}{HR(\mathbf{z}_m)}\right]$.

Most phase III clinical trials of two treatments employ an equal allocation scheme. Such schemes are often unattractive to clinicians and volunteers, as they assign almost half of the patients to the less effective treatment even if a treatment effect exists. For many years, adaptive designs have been proposed as a compromise. Apart from using a link function or a formal optimization procedure, CARA designs for such scenario can

also be developed by the method of treatment effect mappings similar to Rosenberger and Sheshaiyer (1997). The score function $u_k(0)$ can be used to map the past histories to the present allocation pattern. When a patient is ready for randomization, a function of the current value of the score function from the fitted Cox regression model is used to bias a coin, which is then used for randomization. As mentioned earlier, $u_A(0)$ measures the difference between the sum over the survival times of the observed number of events from treatment $A$ and the expected number of events under the null hypothesis. In general, the score function can be used to develop a mapping onto [0,1] that exceeds 0.5 if treatment $A$ has been performing better thus far, and is less than 0.5 if treatment $B$ has been performing better. Let $N_A(m)$ and $N_B(m)$ be the numbers of patients allocated to treatments $A$ and $B$, respectively, up to stage $m$. Define

$$Q_k = \frac{u_k(0)}{\{max(N_A(m), N_B(m)\}\frac{h_A(t|\mathbf{z}_m)}{h_B(t|\mathbf{z}_m)}\sum_{i=1}^{d_m}(\frac{1}{N-i}) + h(t|\mathbf{z}_{m+1})}$$

as the treatment effect mapping factor . Then the allocation function to treatment A can be suggested with the following mapping :

$$P[X_{(m+1)} = 1|\mathcal{X}_m, \mathcal{Y}_m, \mathcal{Z}_m, z_{m+1}] = \frac{1}{2}(1 - Q_A). \tag{3.19}$$

It can be seen from (3.19) above that if the null hypothesis of the equality in the covariate adjusted treatment effects is true, the treatment effect mapping factor $Q_k$ is zero and therefore the probability of allocating the next patient to one of the two treatment arms would be 0.5. If the null hypothesis is false, the allocation proportion to treatment $A$ drifts away from 0.5 according to the magnitude of $Q_k$. This idea is similar to that of Rosenberger (1993) in dealing with immediate continuous outcomes using a nonparametric rank test. Most clinical trials in the United States use a fixed sample design and interim monitoring. The design in (3.19) assumes a fixed sample size and also skews the allocatio proportions according to the covariate-adjusted relative efficacy of the treatments, making it fully randomized.

The allocation proportions obtained by using the formal optimization procedure can be targeted using the covariate-adjusted doubly-adaptive biased coin design (CADBCD) or the covariate-adjusted efficient-randomized adaptive design (CAERADE). The CADBCD is a randomization procedure which is used to target the allocation proportions

$\{\pi_{Ai}^S(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B, \mathbf{z})\}_{i=1}^2$ and applies to the cases where the desired allocation proportions are unknown, but estimated sequentially. When the $(m+1)^{th}$ patient enters the clinical trial with covariate vector $\mathbf{z}_{m+1}$, let $\hat{\pi}_m = \pi_A^S(\hat{\boldsymbol{\beta}}_{A,m}, \hat{\boldsymbol{\beta}}_{B,m}, \mathbf{z}_{m+1})$ represent the estimate of $\pi_A^S(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B, \mathbf{z})$ based on the responses observed from the $m$ patients, adjusted for the covariate $\mathbf{z}_{m+1}$ of the incoming patient. Let $\hat{\rho}_{Am} = \{\sum_{i=1}^m \pi_A^S(\hat{\boldsymbol{\beta}}_{A,m}, \hat{\boldsymbol{\beta}}_{B,m}, \mathbf{z}_i)/m\}$ be an estimate of the average target allocations for treatment $A$ based on the data for the first $m$ patients. Using the CADBCD procedure, the $(m+1)^{th}$ patient can be assigned to treatment $A$ with probability $j_{m+1}[\{N_A(m)/m\}, \hat{\pi}_m, \hat{\rho}_{Am}]$, where $\{N_A(m)/m\}$ is the observed proportion of patients who have been assigned to treatment $A$ after $m$ allocations. Therefore, the mathematical form of the allocation rule for the $(m+1)^{th}$ patient entering the clinical trial with covariate vector $\mathbf{z}_{m+1}$,to be assigned to treatment $A$ is

$$j_{m+1}\left\{\frac{N_A(m)}{m}, \hat{\pi}_m, \hat{\rho}_{Am}\right\} = \begin{cases} \frac{\hat{\pi}_m\{\hat{\rho}_{Am}/\frac{N_A(m)}{m}\}^\alpha}{\hat{\pi}_m\{\hat{\rho}_{Am}/\frac{N_A(m)}{m}\}^\alpha + (1-\hat{\pi}_m)\{\{1-\hat{\rho}_{Am})/(1-\frac{N_A(m)}{m})\}^\alpha}, & \text{if } 0 < \frac{N_A(m)}{m} < 1, \\ 1 - \frac{N_A(m)}{m}, & \text{if } \frac{N_A(m)}{m} = 0 \text{ or } \frac{N_A(m)}{m} = 1. \end{cases} \tag{3.20}$$

Here $\alpha$ is a non-negative parameter controlling the degree of randomness of the CADBCD procedure. A value of $\alpha = 0$ corresponds to the procedure being most random and a value of $\alpha = \infty$ corresponds to it being most deterministic. The allocation function $j_{m+1}[\{N_A(m)/m\}, \hat{\pi}_m, \hat{\rho}_m]$ is strictly decreasing in $\{N_A(m)/m\}$ and strictly increasing in $(\hat{\pi}_m, \hat{\rho}_m)$ on [0,1]×[0,1].

On the other hand the allocation function for the CAERADE is discrete. When efficiency is the sole criterion for developing a CARA design, the CAERADE performs the best compared to CADBCD. Let $\hat{\pi}_m = \pi_A^S(\hat{\boldsymbol{\beta}}_{A,m}, \hat{\boldsymbol{\beta}}_{B,m}, \mathbf{z}_{m+1})$ denote an estimate of the target allocation proportion $\pi_A(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B, \mathbf{z})$ based on the data from $m$ patients, adjusted for the covariate $\mathbf{z}_{m+1}$. Then, according to the CAERADE allocation rule, the probability of the $(m+1)^{th}$ patient with covariate vector $\mathbf{z}_{m+1}$ being assigned to

treatment $A$ is given by

$$j_{m+1}\left\{\frac{N_A(m)}{m}, \hat{\pi}_m, \hat{\rho}_{Am}\right\} = \begin{cases} \alpha'\hat{\pi}_m & \text{if } \frac{N_A(m)}{m} > \hat{\rho}_{Am}, \\ \hat{\pi}_m & \text{if } \frac{N_A(m)}{m} = \hat{\rho}_{Am}, \\ 1-\alpha'(1-\hat{\pi}_m) & \text{if } \frac{N_A(m)}{m} < \hat{\rho}_{Am}, \end{cases} \tag{3.21}$$

where $0 \leq \alpha' < 1$ is a constant that reflects the degree of randomization. Hu,Zhang and He (2009), following Burman (1996), recommended the value of $\alpha'$ between 0.4 and 0.7. This gives a family of CARA designs that are fully randomized and also asymptotically efficient. The CAERADE with covariates ignored can be viewed as a generalization of the Efron's biased coin design for any desired allocation function, which may depend on the unknown parameters. If the response distribution belong to the exponential family, the CAERADE for any $\alpha \in$ [0,1) is fully efficient. However, this can also be used to target the optimal allocation proportions when the survival responses follow a semi-parametric survival model. Though there is hadrly a reference for the theoretical asymptotic properties of the ERADE or the CAERADE for response distribution which do not follow a parametric model, the performance of these randomization procedures can be viewed using simulations.

Atkinson and Biswas (2005) proposed a class of CARA designs for which the randomization probabilities are sequentially determined by maximizing a utility function that combines inferential and ethical criteria. The authors concentrated on normally distributed responses. However, their approach can be applied in the context of a survival responses of a patient following a semi-parametric model between the treatment arms. If $m$ patients have been enrolled in the trial and $N_A(m)$ patients have been assigned to treatment $A$ and $N_B(m) = m$ - $N_A(m)$ have been allocated to treatment $B$, a randomization probability for the $(m+1)^{th}$ patient is determined by optimizing a utility function that combines the inferential and ethical criteria of a CARA design. The inferential criteria can be addressed by taking a convex criterion based on the observed information matrix $J(\boldsymbol{\beta}) = diag\{J_A(\boldsymbol{\beta}_A), J_B(\boldsymbol{\beta}_B)\}$, where

$J_k(\boldsymbol{\beta}_k) = \sum_{i=1}^{n_k} \delta_{ik} V(t_i, \boldsymbol{\beta}_k)$ for $k = A, B$. Since the main interest lies in maximizing the information about the model parameters, let $\nu_k = \log\{1 + d(k, \boldsymbol{\beta}_k, \mathbf{z}_{m+1})\} + \Sigma_{k=A}^{B} J_k(\boldsymbol{\beta}_k)$ be some measure of information from applying treatment $k$ to the $(m+1)^{th}$ patient, where $d(k, \boldsymbol{\beta}_k, \mathbf{z}_{m+1})$ is the directional derivative of the $D$-optimal criterion given by $d(k, \boldsymbol{\beta}_k, \mathbf{z}_{m+1}) = \epsilon_k(\mathbf{z}; \boldsymbol{\beta}_k) \mathbf{z}_{m+1}^T J_k^{-1}(\boldsymbol{\beta}_k) \mathbf{z}_{m+1}$. The $(m+1)^{th}$ patient is assigned to the treatment for which $d(k, \boldsymbol{\beta}_k, \mathbf{z}_{m+1})$ is the maximum. For an ethical criterion, $p_A(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B, \mathbf{z}_{m+1}) = h_B(t|\mathbf{z}_{m+1}) / \{h_A(t|\mathbf{z}_{m+1}) + h_B(t|\mathbf{z}_{m+1})\}$ can be considered as a mapping of the treatment effect, adjusted for the covariates of the new patient. The treatment allocation probabilities, $\pi_A^S$ and $\pi_B^S = 1 - \pi_A^S$, for the $(m+1)^{th}$ patient are determined by maximizing the utility function $U = \sum_{k=A}^{B} \pi_k^S \nu_k - \eta \sum_{k=A}^{B} \pi_k^S \log(\pi_k^S / p_k)$. Using Lagrange multiplier optimization, the optimal solution is obtained as

$$\pi_A^S = \frac{p_A\{1 + d(A, \boldsymbol{\beta}_A, \mathbf{z}_{m+1})\}^{1/\eta}}{p_A\{1 + d(A, \boldsymbol{\beta}_A, \mathbf{z}_{m+1})\}^{1/\eta} + p_B\{1 + d(B, \boldsymbol{\beta}_B, \mathbf{z}_{m+1})\}^{1/\eta}}. \tag{3.22}$$

The optimal allocation probabilities in (3.22) depend on the model parameters, which must be replaced by their partial likelihood estimates. When $\eta \to 0$, the allocation function maximizes the efficiency of the study design, whereas, when $\eta \to \infty$, it satisfies the ethical standpoint of the trial.

## 3.5 Simulation Results

In order to compare the performances of the different derived CARA randomization procedures, a two-arm survival trial with 800 patients has been considered. A patient's arrival time here is simulated from a uniform (0,365) distribution, whereas the response time of a patient is added to the recruitment time of the patient and patients whose outcomes have not been observed by the pre-specified study time, are said to be generalized type I right censored. The recruitment period has been considered to be 365 days, and the overall trial duration is taken to be 581.66 days.

Following Rosenberger, Vidyashankar and Agarwal (2001), a covariate structure of three independent covariates has been generated. These are gender (Bernoulli, p = 0.5), age {Uniform(30,75)} and cholesterol level {Normal (200,400)}. The censoring

time of the patients is simulated from a uniform (0,581.66) distribution. The survival time of a patient with covariate vector $\mathbf{z} = (1, z_1, z_2, z_3)^T$ in treatment group $k$ is simulated from the Weibull distribution with scale parameter $\mu_k(\mathbf{z})$ =exp($\boldsymbol{\beta}_k^T\mathbf{z}$) and shape parameter $\gamma_k = 2.07527$. Three choices of the treatment effects vector have been considered in this case, which are neutral effect of either treatment, positive effect of treatment $A$ and negative effect of treatment $A$. The effects of the corresponding covariates for the simulation model $\mu_k(\mathbf{z})$ =exp($\boldsymbol{\beta}_k^T\mathbf{z}$) are the same as summarized in Table 2.1. The procedure used here is a fully sequential one that recalculates the randomization probabilities over 5000 simulation runs. The competing randomization procedures and the corresponding design numbers are listed in Table 3.1.

| **Design** | **Competing Randomization Procedures** |
|---|---|
| I | Completely randomized design |
| II | Efron's biased coin design with p = 2/3 |
| III | Pocock and Simon design with p = 3/4 |
| IV | Taves minimization procedure with p = 1 |
| V | CARA CADBCD with(3.15) as the target |
| VI | CARA CADBCD with(3.16) as the target |
| VII | CARA CAERADE with(3.15) as the target |
| VIII | CARA CAERADE with(3.16) as the target |
| IX | CARA design based on the probit link function |
| X | CARA design based on (3.19) |
| XI | CARA design based on (3.22) with $\eta = 1$ |
| XII | CARA design based on (3.22) with $\eta = \infty$ |

Table 3.1: List of the competing designs

In survival trials, the delay time for a patient is the patient's survival or censoring time. To facilitate CARA designs with delayed responses, it is required that, at the $i^{th}$ patient$'$s randomization time, only data from those patients who have responded before the $i^{th}$ patient's arrival are used in computing the randomization probability for the $i^{th}$

patient. For the implementation of the derived CARA designs, initially $2m_0$ patients have been equally allocated to the two treatment arms using Efron's biased coin design. Here, $m_0$ is a positive number and $2m_0$ is chosen to be 160. After the model parameters are estimated from the initial stage of the design, new patients arrive sequentially into the trial and the Cox proportional hazards model (3.5) is fitted to the available responses and covariate information of the patients. Next, an allocation proportion is calculated which is a functions of the partial likelihood estimates obtained from the fitted Cox model. A randomization rule is then used to allocate the new patients to a treatment arm. This randomization rule can be based on any one of the derived allocation functions which target a specific allocation proportion. A pseudo-random number generator is then used to draw a uniform random number between 0 and 1. If the derived probability from the randomization rule is greater than or equal to this random number, the patient is assigned to treatment $A$ or else the patient is assigned to treatment $B$. The procedure described is repeated for the subjects entering the trial in the future.

For the CADBCD designs in Table 3.1 above, following Hu and Zhang (2004), Zhang and Hu (2009), and Rosenberger and Hu (2004), the trade-off parameter for randomness is taken to be $\alpha = 2$. Based on Burman's (1996) studies on the expected p-value deficiency to evaluate the performance of a particular design, Hu, Zhang and He (2009) recomended that for the appropriate implementation of the ERADE designs, it is reasonable to choose $\alpha'$ between 0.4 to 0.7. The parameter $\alpha'$ controls the degree of randomness of the design. When $\alpha'$ is smaller, the CAERADE is more deterministic and has a smaller variability. Here, for the appropriate implementation of the CAERADE designs, $\alpha'$ is chosen to be 0.55.

Among the balanced randomization procedures, apart from the completely randomized design, Efron's biased coin design is a restricted randomization procedure, whereas Pocock and Simon's (1975) design and Taves' (1974) design are covariate adaptive randomization procedures. These covariate-adaptive randomization procedures are is called minimization procedures, as they minimize treatment imbalances marginally on important covariates. For their implementation, all covariates must be categorical. The

covariate age has been dichotomized according to age $< 53$ and age$\geq 53$, whereas the covariate cholesterol level has been dichotomized according to cholestrol level$<200$ and cholestrol level $\geq 200$. According to Pocock and Simon (1975), for an incoming patient, one computes the treatment imbalance at each observed levels of the patients covariates, and with probability 3/4 the patient is assigned to the treatment arm that reduces the overall covariate imbalance. If the imbalance is zero, the patient is randomized to any competing treatment arms with equal probability. Taves (1974), on the other hand, proposed a deterministic method to allocate treatments designed to minimize the overall covariate imbalance marginally on important covariates.

The performances of the competing randomization procedures in Table 3.1 for semiparametric survival models, can be analysed from their operating characteristics in Table 3.2 and Table 3.3. All simulations were carried out using the R statistical software. For each experimental design, the significance level of the Wald test for testing the treatment difference has been set to 0.05.

| **Design** | $\frac{N_A}{n}$ **(SE)** | $N_{AM}$-$N_{BM}$ | $N_{AF}$-$N_{BF}$ | **Event** | **Type I Error** |
|---|---|---|---|---|---|
| I | 0.50 (0.018) | 200-200 | 200-200 | 650 | 0.05 |
| II | 0.50 (0.001) | 200-200 | 200-200 | 650 | 0.04 |
| III | 0.50 (0.015) | 200-200 | 200-200 | 649 | 0.05 |
| IV | 0.50 (0.004) | 200-200 | 200-200 | 650 | 0.05 |
| V | 0.50 (0.014) | 200-200 | 200-200 | 649 | 0.05 |
| VI | 0.51 (0.020) | 202-198 | 204-196 | 650 | 0.05 |
| VII | 0.51 (0.015) | 203-197 | 203-197 | 650 | 0.04 |
| VIII | 0.51 (0.019) | 202-198 | 204-196 | 650 | 0.04 |
| IX | 0.50 (0.075) | 200-200 | 200-200 | 650 | 0.07 |
| X | 0.49 (0.022) | 196-204 | 196-204 | 650 | 0.05 |
| XI | 0.50 (0.026) | 200-200 | 200-200 | 650 | 0.04 |
| XII | 0.50 (0.026) | 200-200 | 200-200 | 650 | 0.04 |

Table 3.2: Simulation Results of the competing designs for neutral model

Table 3.2 presents the operating characteristics of the various competing designs when there is no difference in the effects of the two treatment arms. It compares the derived semi-parametric CARA designs with the traditional balanced randomization procedures, under the null model. It can be seen that, almost all designs result in an equal allocation of patients to the two treatment arms. Efron's biased coin design achieves the lowest variability for the allocation proportions. This is because the asymptotic distribution of the allocation proportion for this design converges to a single point. However, all of the derived allocation proportions for the CARA randomization procedures, apart from designs IV and IX, are similarly variable compared to the completely randomized design. It has been shown by Cox (1972) that the asymptotic behaviour of the partial likelihood estimator is similar to the maximum likelihood estimator. Therefore, the Wald test can be used to compare the covariate-adjusted treatment effects from the fitted Cox regression model after the patients arrive sequentially. With 800 patients, the type I error rates obtained from the Wald test for almost all of the competing randomization procedures are close to the nominal significance level of 0.05, whereas that of design IX is inflated. The diference in the standard errors for of the type I error rates range from 0.001 to 0.004. The variabilities of the type I error rates for the balanced randomization designs are the lowest and that for design IX is the highest . This phenomenon was further explored by simulation for a range of sample sizes and the simulated type I error rates revealed similar results. The average numbers of events are almost the same for all the designs and their standard errors were no more than 11.

Table 3.3 presents the operating characteristics of the various competing randomization procedures when the covariate adjusted effects of the two treatment arms differ. From the negative model, it can be seen that all of the derived CARA designs result in skewed allocation towards the better treatment arm, but the degree of skewness is different. In this case, it is treatment $B$ which is superior. Design IX is most skewed towards the better treatment arm and thus result in the least number of events on average. On the other hand, designs VII and X least skewed compared to the other

| **Model** | **Design** | $\frac{N_A}{n}$ **(SE)** | $N_{AM}$-$N_{BM}$ | $N_{AF}$-$N_{BF}$ | **Event** | **Power** |
|---|---|---|---|---|---|---|
| Negative | I | 0.50 (0.018) | 200-200 | 200-200 | 334 | 0.99 |
| | II | 0.50 (0.001) | 200-200 | 200-200 | 334 | 0.99 |
| | III | 0.50 (0.016) | 200-200 | 200-200 | 334 | 0.99 |
| | IV | 0.50 (0.005) | 200-200 | 200-200 | 334 | 0.99 |
| | V | 0.45 (0.015) | 180-220 | 181-219 | 332 | 0.99 |
| | VI | 0.39 (0.020) | 157-243 | 158-242 | 332 | 0.98 |
| | VII | 0.48 (0.016) | 192-208 | 197-203 | 332 | 0.99 |
| | VIII | 0.38 (0.019) | 150-250 | 152-248 | 332 | 0.98 |
| | IX | 0.30 (0.070) | 120-280 | 120-280 | 322 | 0.90 |
| | X | 0.47 (0.080) | 188-212 | 190-210 | 332 | 0.87 |
| | XI | 0.43 (0.026) | 172-228 | 172-228 | 329 | 0.97 |
| | XII | 0.40 (0.028) | 160-240 | 161-239 | 328 | 0.96 |
| Positive | I | 0.50 (0.018) | 200-200 | 200-200 | 726 | 0.99 |
| | II | 0.50 (0.001) | 200-200 | 200-200 | 726 | 0.99 |
| | III | 0.50 (0.016) | 200-200 | 200-200 | 725 | 0.99 |
| | IV | 0.50 (0.005) | 200-200 | 200-200 | 725 | 0.99 |
| | V | 0.56 (0.014) | 223-177 | 226-174 | 723 | 0.99 |
| | VI | 0.62 (0.020) | 248-152 | 249-151 | 720 | 0.98 |
| | VII | 0.52 (0.015) | 208-192 | 209-191 | 723 | 0.99 |
| | VIII | 0.62 (0.019) | 248-150 | 248-154 | 723 | 0.98 |
| | IX | 0.71 (0.074) | 299-101 | 272-128 | 710 | 0.89 |
| | X | 0.52 (0.083) | 122-118 | 122-118 | 725 | 0.84 |
| | XI | 0.57 (0.025) | 227-173 | 227-173 | 722 | 0.98 |
| | XII | 0.60 (0.025) | 228-172 | 228-172 | 721 | 0.97 |

Table 3.3: Performance of the competing designs when Treatment effects differ

competing CARA randomization procedures. All of the competing designs, apart from IX and X, have similar powers for the Wald test, whereas skewing the patient allocation towards the better treatment arm results in fewer of events on average during the trial. The standard errors of the latter are no more than 14. Therefore, using the CARA

designs instead of the balanced randomization procedures can result in more patients being treated with the better-performing treatment arm. Apart from designs IX and X, this ethical benefit is achieved without compromising much on the power of the test for the covariate-adjusted treatment differences.

Similar results can be seen for the positive model in Table 3.3. All of the derived CARA randomization procedures significantly skew the patient allocation towards the better treatment arm, with the design based on the probit link function achieving the highest skewness. Apart from designs IX and X, they have similar powers and the skewness of the CARA designs towards the better treatment arm results in fewer events compared to the balanced randomization procedures. This time, the average number of events have standard errors no more than 10. Therefore, the optimal CARA designs can be considered as suitable alternatives to the traditional balanced randomization procedures.

The observed allocation proportions in Figure 3.1 depict the performances of the derived CARA procedures compared to the traditional balanced designs. The distributions of the observed allocation proportions for the various designs appear to be very close to a symmetric distribution, but with different means and with different variability. It can be seen that the CARA designs allocate more patients in the trial to the better performing treatment arm.

The impact of sample size on the type I error rates and powers for different designs has been explored. Figure 3.2 shows the type I error rates and powers for sample sizes between 200 to 800.

It can be seen from Figure 3.2 that the type I error rates are in the range 0.05 to 0.07 and that the power are very similar for a sample size of around 800 patients. This means that, for the statistical power of the Wald test to be greater than 0.9, most of the competing designs are equally efficient. However, if the the power for treatment difference is relaxed slightly by the clinicians to greater than 0.8, which is often the case in late phase clinical trials due to patient recruitment problems, then it can be seen that most of the proposed CARA designs are more powerful than the traditional

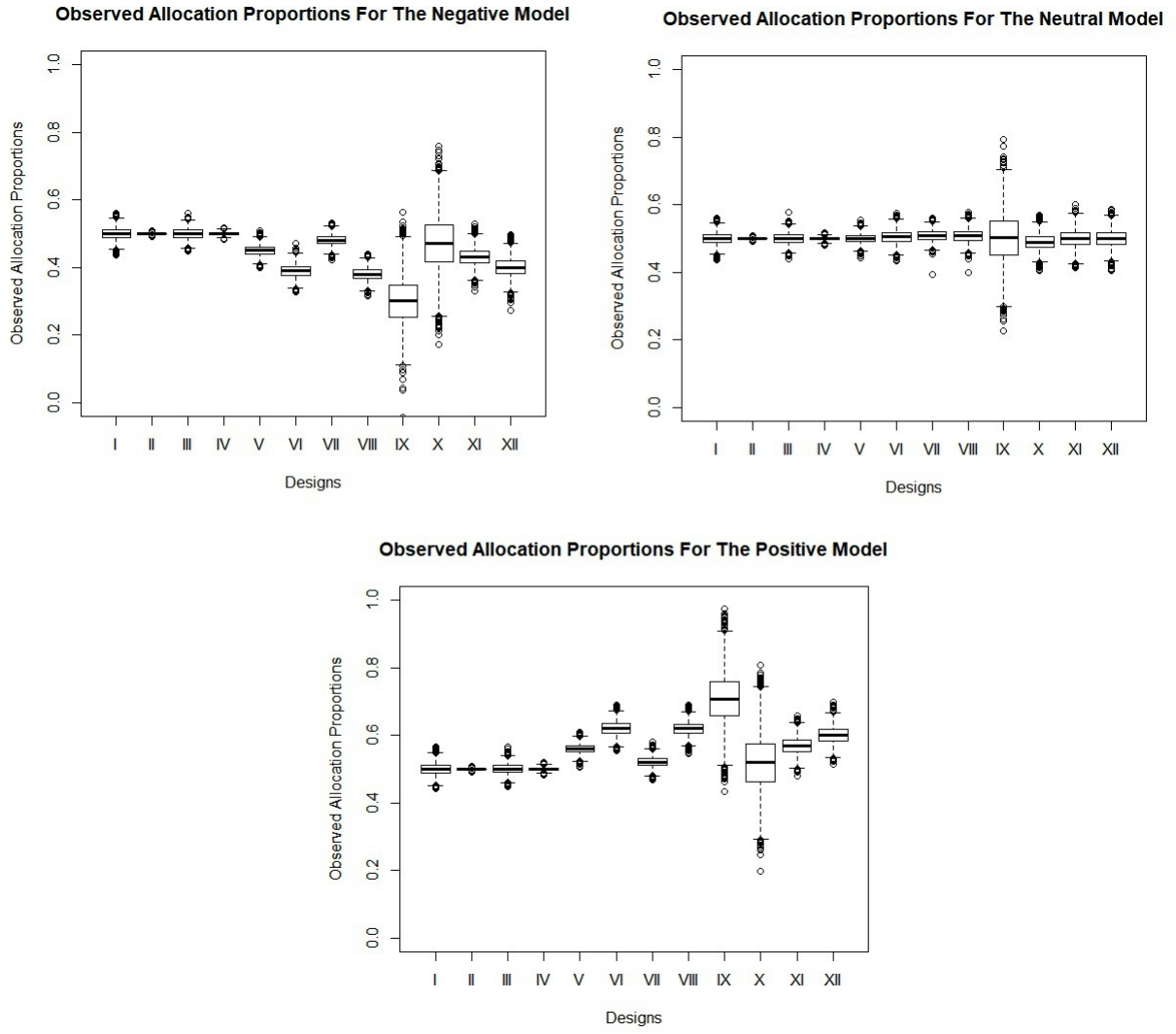


Figure 3.1: Effect Size for Semi-parametric models skews CARA allocation proportions

balanced designs. It must be noted that the simulations here considered a phase 3 trial with $n = 800$ patients instead of $n = 400$ patients as in chapter 3. This is because such a sample size is large enough to enable asymptotic properties of CARA procedures to hold with power less than 1. This is reflected in Figure 3.2 above. More patients are needed to achieve equivalent level of power in this scenario because chapter 3 assumed a parametric model for the survival responses and therefore needed less number of patients to achieve the desired power than the designs based on a semi-parametric model which is more dependent on the observed data. For the Wald test, the significance level is set to 0.05.

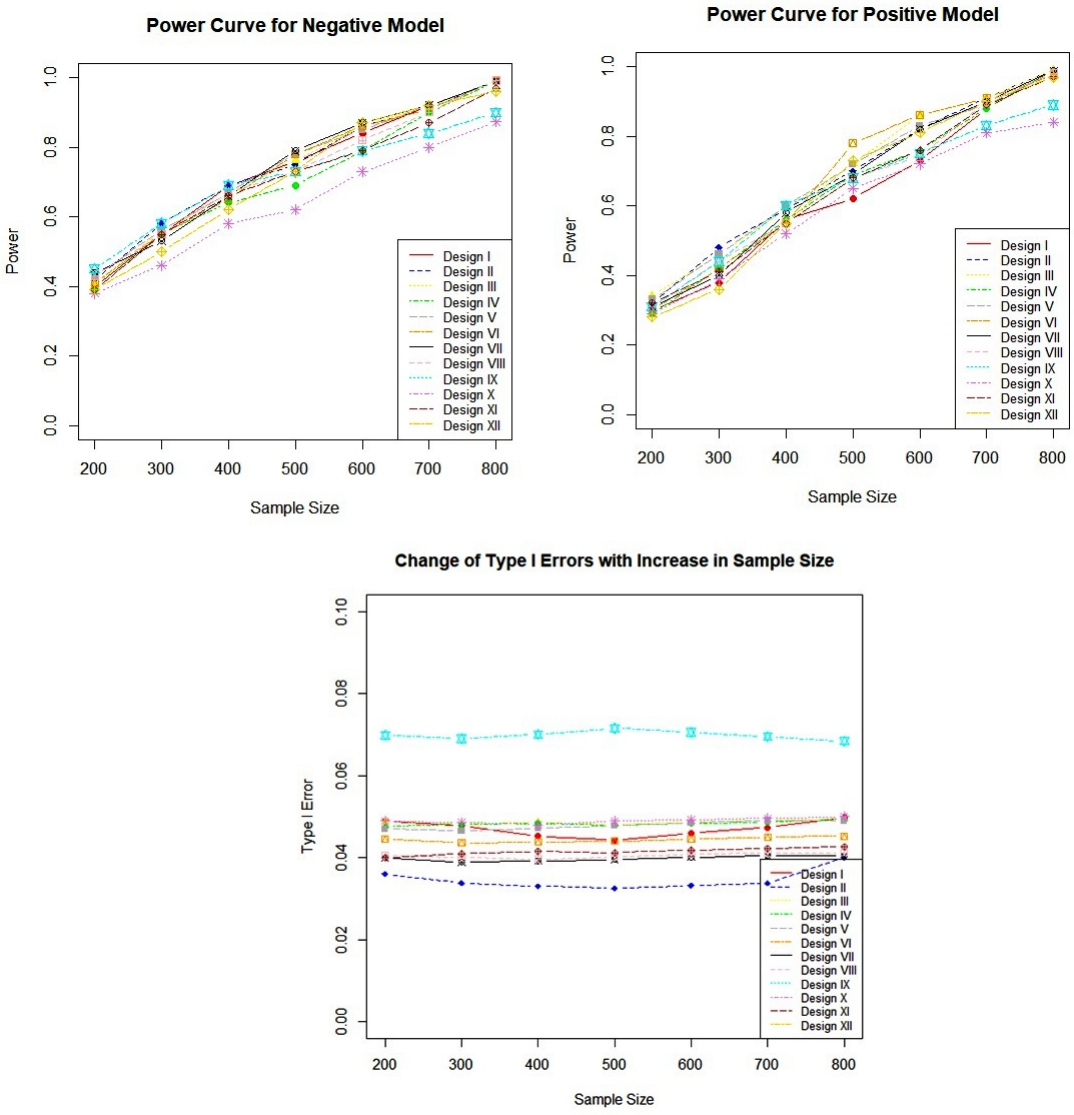


Figure 3.2: Type I error rates and Powers for the semi-parametric Competing Designs

## 3.6 Re-designing a Real-Life Clinical Trial

In order to apply the derived methodologies to re-design a clinical trial, a randomized trial in advanced colorectal cancer has been considered. The study was previously explored by Sverdlov, Rosenberger and Ryeznik (2013), where they re-designed the trial to establish the usefulness of CARA designs for exponentially distributed survival responses. In this chapter, the trial is re-designed using the derived semi-parametric CARA designs for survival responses. For a period of 21 months, 572 eligible patients

were randomized in a 1:1 ratio cetuximab plus best supportive care (treatment $A$) and best supportive care alone (treatment $B$). The primary endpoint of this study was overall survival, defined as the time to death from any cause. The model-adjusted median overall survival was 6.1 months for treatment $A$ versus 4.6 months for treatment $B$. However, the degree and the direction of the effect of cetuximab differed between the levels of $K-ras$ mutation status. Patients with wild-type K-ras tumours benefited from cetuximab (median overall survival 9.5 versus. 4.8 months; hazard ratio for death 0.55), whereas patients with a colorectal tumour bearing mutated K-ras did not benefit from cetuximab (median overall survival 4.6 versus. 4.5 months; hazard ratio for death 0.98).

A simulation study with 5,000 replications was conducted to compare the CARA designs with the traditional balanced randomization procedures. The CARA designs considered here are the ones derived by using an optimality criterion. For appropriate implementation of the derived CARA designs, 110 patients were initially equally randomized to the two treatment arms using Efron's biased coin design before the adaptive randomization process started. Following Karapetis *etal.* (2008), the parameters were chosen as follows: $R = 21$ months; trial duration $D = 27$ months; : $z = 0$ (wild-type $K - ras$ tumour) with probability 0.59 and $z = 1$ (mutated $K - ras$ tumour) with probability 0.41. Patient survival times were simulated from an exponential distribution with mean $\exp(\beta_{k0} + \beta_{k1}z)$, $k = A, B$ and $z = 0, 1$. The model parameters were set to $\beta_{A0} = 2.62$, $\beta_{A1}$ = -0.68, $\beta_{B0} = 1.87$, and $\beta_{B1} = 0.02$.

| **Designs** | **I** | **II** | **VI** | **VIII** | **XI** | **XII** |
|---|---|---|---|---|---|---|
| $N_A$/n (SE) | 0.50(0.02) | 0.50(0.002) | 0.65(0.02) | 0.65(0.02) | 0.62(0.02) | 0.68(0.03) |
| $N_{A-}$—$N_{B-}$ | 171—171 | 146—146 | 248—102 | 242—104 | 233—110 | 257—102 |
| $N_{A+}$—$N_{B+}$ | 115—115 | 140—140 | 124—98 | 130—96 | 122—107 | 132—81 |
| Events | 377 | 377 | 357 | 356 | 361 | 355 |
| Power | 0.99 | 0.99 | 0.96 | 0.97 | 0.95 | 0.94 |

Table 3.4: Re-design of a survival trial in colorectal cancer

Table 3.4 presents the simulation summary for six different designs in Table 3.1. It can be seen that using the CARA designs instead of the traditional balanced designs results in more patients being treated with the better treatment during the trial. This is achieved without compromising much on the power of the Wald test for the of treatment-biomarker interaction in the Cox regression model, which is fitted after patients arrive squentially in the trial. It can also be seen that the CARA designs account the treatment-covariate interaction, whereas the balanced designs do not. The CARA designs result in greater skewing to treatment $A$ in the wild-type $K-ras$ subgroup than in the mutated $K-ras$ subgroup. On the other hand, the balanced randomization procedures balances the treatment allocations in both the $K-ras$ subgroups. Therefore, the CARA designs on average result in many fewer events during the trial compared to the balanced randomization procedures, the standard errors being no more than 3. Since there is a significant interaction between the treatment effects and the $K-ras$ biomarker subgroup effects, the Wald test is testing the hypothesis of equality in the treatment effects within each biomarker stratum. Unlike the exponential regression model, the Cox regression model treats the baseline hazard as a nuisance parameter. So it does not provide an estimate of the intercept. Simulations were also run for 572 patients assuming no differences between the covariate adjusted treatment effects for $A$ and $B$. The simulated type I error rates for the designs were between 0.04 and 0.05.

## 3.7 Discussion

A new CARA randomization procedure has been introduced in this chapter for two-arm right censored survival trials. This is done to enhance the applicability of CARA designs for treatment comparisons in real life clinical trials. The procedures in this chapter obviate any parametric assumptions about the underlying distribution of the survival responses of the patients. However, in order to preserve the robustness of parameter estimation, a lighter assumption has been made about the hazard rate at any given time-point for an individual in one treatment group being proportional to the hazard rate at the same time-point for an individual in the other treatment group.

The designs developed here are based on both formal optimization procedures as well as non-optimal procedures. The approaches to optimality includes a covariate adjusted optimal target approach (Zhang and Hu, 2009) and a weighted optimality approach (Atkinson and Biswas, 2005). The non-optimal procedures such as the probit link design and the score design are developed with the intention of addressing the ethical issue of treating more patients with the better treatment during the trial. The operating characteristics of the proposed CARA designs have been compared with the balanced randomization designs through extensive simulation for a two-arm survival trial with three predictive covariates, right-censored data and staggered entry. It has been found that many of the proposed CARA randomization designs generate skewed allocations according to covariate-specific treatment differences and result in fewer events in the trial, while having similar powers for the Wald test for treatment differences compared to the balanced randomization designs. An application of the proposed methodology has been illustrated by redesigning a two-arm survival trial from the literature.

When comparing CARA designs based on covariate adjusted optimal targets, two distinct approaches have been used: the covariate-adjusted doubly adaptive biased coin design (Hu and Zhang, 2004) and the covariate-adjusted efficient randomized adaptive design (CAERADE) which is motivated from Hu, Zhang and He (2009). The work of Hu, Zhang and He (2009) can be regarded as a generalization of Efron's biased coin design for any desired allocation proportion, which may depend on the unknown parameters. Through simulations it is seen that the CAERADE has got lower variability as compared to the corresponding covariate-adjusted doubly adaptive biased coin design (CADBCD). Hu and Rosenberger (2003) established that optimality, variability and power can be considered as the essential components for selecting a suitable adaptive randomization procedure. In the situation where efficiency is critically important, the CAERADE should be the best choice among all of the CARA randomization procedures. However, sometimes the CAERADE does not converge to the target allocation proportion as fast as the CADBCD does, although its finite-sample variances are always small. This is mainly because the allocation probabilities in the CAERADE are not stable. The allocation function being discrete, they always jump from one value to another. A continuous allocation function like the DBCD can make the allocation

probabilities stable and speed up the convergence of the sample allocation proportions.

When comparing each class of CARA designs, it has been seen that the designs in each class have their merits. CARA designs with targets have established asymptotic properties. By appropriately choosing the trade-off parameter $\eta$, the weighted optimality CARA designs also give skewed allocations towards the better treatment arm and balance the objectives of statistical efficiency and ethics. However, selecting the value of the parameter $\eta$ depends on the experimental criteria. The non-optimal designs also work well compared to the balanced randomization procedures. If the sole criterion is to treat more patients with the better performing treatment during the trial, the design based on the probit link function works the best and also achieves the least number of events on average. When there is no difference between the effects of the two treatments, the design based on the score function achieves the nominal type I error rate. On the other hand, when one of the treatment arms performs better than the other, this design happens to be the least powerful. The merits of the developed methodologies have also been elucidated by applying them to redesign an existing clinical trial.

The methodology described here is free from any distributional assumption on the survival responses, but relies on a lighter assumption of the hazard rate at any given time-point for an individual in one treatment group being proportional to the hazard rate at the same time-point for an individual in the other treatment group. Conceptually, instead of considering the proportionality of the hazard rates of individuals from the two treatment groups, one can also use the proportionality of the odds of survival of patients from the two treatment groups as the basis for the development of CARA randomization procedures. Another important area of potential application of CARA randomization procedures is time-to-event trials with more than two treatment arms. All of the CARA designs for survival trials developed so far in the literature concentrated on trials with events due to a single cause. However, in many real-life trials, patients experience events due to multiple causes. Therefore, the information related to the cause of the events is very important when developing a CARA design, which would make such designs more applicable in real-life survival trials.

It should be acknowledged that the proposed methodology is only suitable for sur-

vival trials with long recruitment periods. The amount of concomitant information in the model impacts sequential estimation at the design stage. One should consider implementing the proposed CARA designs only with a limited number of the most predictive baseline covariates.

# Chapter 4

# Covariate-Adjusted Response-Adaptive Designs for Competing Risk Survival Models

## 4.1 Introduction

A branch of healthcare science that determines the safety and efficacy of treatment regimens is called clinical research. It involves any test article from its inception in the laboratory to its introduction to the consumer market and beyond. Clinical trials are complex experiments which are designed through clinical research to answer specific questions about biomedical or behavioral interventions, including new treatments and known interventions that warrant further study and comparison. A purpose of phase III clinical trials is to obtain information on the performance of the competing treatments. It involves large-scale trials in which a new treatment is compared with one or more standard treatments in terms of its safety and efficacy .

Patients usually arrive sequentially in a clinical trial and are assigned to one of the competing treatment arms. Information about the performance of the competing treatments accrues sequentially as results become available on an increasing number of patients. There has been extensive research in establishing the superiority of a treat-

ment early in the trial. Adaptive designs play an important role in effectively utilizing the sequential information and serving the purpose. Adaptive allocation procedures are sequential designs in which the method of allocation of treatments to patients are modified based on the results obtained in the previous stage until a particular treatment is declared to be a clear winner over the others.

The past several years have witnessed significant contributions on the development adaptive designs for the effective comparison of treatments. These involve the development of balanced allocation schemes such as restricted randomization or covariate-adaptive randomization designs. Restricted randomization balances the patient allocation based on only the history of the previous allocations, whereas covariate-adaptive randomization considers the patient heterogeniety as well while balancing the distributions of the covariates across treatment arms. However, since clinical trials involve human patients, there is an ethical concern to treat as many patients as possible in the trial with the best treatment without compromising much on the power to test for the treatment differences. Covariate-adjusted response-adaptive (CARA) designs serves this purpose by allocating more subjects to the better-performing treatment arm thus far in the trial, and, at the same time, estimate with high efficiency any difference in the covariate-adjusted treatment effects while maintaining the randomness in treatment assignment. The patient allocation to a treatment arm for CARA designs is based on the history of previous patient treatment assignments, the response history, the covariate profiles of the previous patients and also the covariate information of the incoming patient. Such designs are especially used when a non-linear, heteroscedastic model determines the relationship between the covariate profiles of the patients and their responses to a treatment. It is also useful when the effect of a treatment significantly varies among the levels of the covariates of the patients and when multiple experimental objectives such as the ones mentioned in Section 3.4 are being pursued.

There has been very little work on developing CARA designs for survival trials. Those that have accounted for survival trials explored right-censored time-to-event responses with a single cause of failure. However, real-life clinical trials often face complex situations where an individual may experience events from different causes. Competing

risks arise in studies where the failure of an individual may be classified into one of the various mutually exclusive causes of failure such as death from different causes. This makes the usual survival analysis techniques inappropriate, as specialized methods are needed to account for multiple causes of failures. Competing risks are events whose occurrence either precludes the occurrence of another event or fundamentally alters the probability of the occurrence of this other event. For example, while measuring the time to relapse of chronic myeloid leukemia due to the increase of BCR-ABL1 oncopotein in human bone marrow, patients may die due to cardiovascular disease. This is because the interventions known to treat chronic myeloid leukemia are also known for QT prolongation and ventricular tachycardia. Therefore, monitoring such information could be vital, not only for informing patients of the risks that they face in certain situations, but also for making decisions about which treatment to assign a patient, how best to allocate health resources and for understanding the longer-term outcomes of chronic conditions. Handling such information during the design phase of a clinical trial along with that about the patient heterogeneiety would make the design more applicable in real-life clinical trials. In this chapter, CARA designs have been developed for survival trials with competing risk scenarios without relying on any parametric distributional assumptions about the survival responses of the patients. The various CARA designs for survival trials with competing risks are developed and the performance of such designs has been validated using simulation. An existing clinical trial has also been re-designed using the derived methodologies.

An outline of the chapter is as follows. The background material relating to regression models for survival trials with competing risks is explained in Section 4.2. Section 4.3 proposes the various semi-parametric CARA randomization procedures for a survival trial with competing risks. The validation of the findings of Section 4.3 is detailed in Section 4.4 using extensive simulation studies. The results obtained from applying the proposed CARA designs to re-design a real-life clinical trial are presented in Section 4.5. The conclusions of the findings with a discussion and an outline of some future research in this direction are discussed in Section 4.6.

## 4.2 Regression Models for Competing Risk Problems

Often in real-life clinical trials, failure for a person may be due to several distinct causes. It may be desirable to distinguish different kinds of events that may lead to failure and treat the patients differently during the design phase of a clinical trial . For example, to evaluate the efficacy of heart transplants, one would certainly want to treat deaths due to heart failure differently from deaths due to other causes, such as accidents and cancer. These different causes of failure are considered competing events, which introduce competing risks. Thus, problems arising in a clinical trial with multiple causes of failure are commonly referred to as competing risk problems. Treating failures due to other causes as censored observations would not account for the fact that these patients have already faced an event which is of a different cause. This would severely bias the calculation of the risk set of the partial likelihood estimation of the model parameters from the experiment. Therefore, it would result in biased estimates of the model parameters. Figure 4.1 below simulates a cohort of competing events, where the event of interest can be considered as cancer but its observations are made impossible for patients 6 and 8 by a preceding competing event.

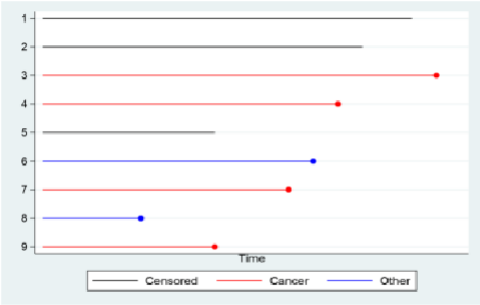


Figure 4.1: Scenario for competing risk

An approach to handling competing risk problems assumes the existence of $c$ failure times, one for each possible type of failure. In a clinical study, one then observes the minimum of the latent failure times ($T$) and the corresponding cause of failure ($\mathcal{C}$). This

means that, if the times to the events from $c$ different causes are $T_j$ for $j = 1, ....c$, then in a clinical trial one observes $T = \min(T_j)$. The problem with this approach is that neither the joint distribution of the failure times nor the correspondng marginal distributions are identifiable from the observed data without additional assumptions, such as the independence of the different latent failure times. Another approach focuses on using the proportional hazards model to identify significant prognostic or risk factors when competing risks are present. This approach considers the joint distribution of failure time $T$ and the cause of failure $\mathcal{C}$ for the covariate vector $\mathbf{z}$. The $j^{th}$ cause-specific hazard at time $t$ is defined by

$$h_j(t|\mathbf{z}) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t, \mathcal{C} = j | T \geq t; \mathbf{z})}{\Delta t}. \tag{4.1}$$

This is the instantaneous failure rate of cause $j$ at time $t$ given $\mathbf{z}$ in the presence of all other causes of failure. The overall hazard $h(t|\mathbf{z})$ from any cause is the sum of the cause-specific hazards over each failure type.

The cumulative incidence function (CIF) can also be used to handle competing risk problems. Using the CIFs instead of the cause-specific hazards helps a clinician to have a direct interpretation in terms of survival probabilities for the particular failure type. This gives the marginal failure probabilities for a particular cause, which are intuitively appealing and more easily explained to non-statisticians. The risk factors which have a significant effect on the cause-specific hazards may not have an effect on the CIFs for that cause. This is because as shown in (4.2) below, the CIF for each cause is a function of the overall survivor function which considers a sum of the hazards for all the different causes. The CIF is a function of the cause-specific hazards and is given by

$$F_j(t|\mathbf{z}) = P(T \leq t, \mathcal{C} = j|\mathbf{z}) = \int_0^t \left[ h_j(u|\mathbf{z}) \exp\left\{ -\int_0^t \sum_{i=1}^{c} h_i(w|\mathbf{z}) dw \right\} \right] du. \tag{4.2}$$

It represents the probability that an event of type $j$ has occurred by time $t$ for patients with covariate $\mathbf{z}$. In the presence of competing risks, the complement of the Kaplan-Meier estimator overestimates the cumulative incidence function of cause $j$. That is, if

$\hat{S}_j(t|\mathbf{z})$ is the estimator of the survivor function for events of cause $j$ given covariates $\mathbf{z}$, then $1 - \hat{S}_j(t|\mathbf{z}) \geq \hat{F}_j(t|\mathbf{z})$. It is possible that the sum of the complements of the Kaplan-Meier estimators at time $t$ over all possible causes of failure is greater than 1, which is impossible, as this sum is the cumulative probability for failure from any cause.

Strategies for regression modelling in survival trials with competing risks can be formed by modelling the cause-specific hazards or the cumulative incidence functions. When the aim is to assess the effect of the risk factors on the risk of failure due to a certain cause, the Cox proportional hazards model can be used for each cause-specific hazard to analyse these seperately by treating individuals failing from other causes as censored observations. However, when the failure times are dependent and when interest lies in assessing the effect of the risk factors on the incidence of a given cause, the cumulative incidence functions can be modelled using the Fine and Gray (Fine and Gray, 1999) approach.

The classical regression analysis of competing risks establishes a Cox proportional hazards model (Prentice *et al.*, 1978) for each cause-specific hazard given by

$$h_{jk}(t|\mathbf{z}) = h_{0jk}(t)e^{\boldsymbol{\beta}_{jk}^T\mathbf{z}} \quad \text{for} \quad j = 1, 2, ...., c, \tag{4.3}$$

where $\mathbf{z}$ is a $p \times 1$ vector of covariates and $\boldsymbol{\beta}_{jk}$ is a $p \times 1$ vector of regression coefficients for cause $j$ and treatment arm $k = A, B$. The effect of the covariates is assumed to act multiplicatively on an unknown baseline hazard function $h_{0j}(t)$, which is considered to be a nuisance parameter in the model. Estimation of the regression parameters $\boldsymbol{\beta}_{jk}$ is based on the partial likelihood approach. If $C_{ik}$ denotes the censoring time for the $i^{th}$ patient on treatment arm $k$, then the partial likelihood function is defined only for the event times, yielding

$$L_k(\boldsymbol{\beta}_{1k}, \boldsymbol{\beta}_{2k}, ....., \boldsymbol{\beta}_{ck}) = \prod_{i=1}^{n_k}\prod_{j=1}^{c}\left(\frac{e^{\boldsymbol{\beta}_{jk}^T\mathbf{z_i}}}{\sum_{l\in R_i} e^{\boldsymbol{\beta}_{jk}^T\mathbf{z_l}}}\right)^{\delta_{ijk}} \tag{4.4}$$

for $k = A, B$, where the risk set is $R_i = \{l : t_l \geq t_i\}$ and $\delta_{ijk} = I(T_{ijk} < C_{ik}, \mathcal{C} = j)$. The risk set can be diminished by the occurrence of an event from any cause. Maximizing each factor of the partial likelihood function in (4.4) provides an estimator $\hat{\boldsymbol{\beta}}_{jk}$ which is consistent and asymptotically normal under regularity conditions. Given $\hat{\boldsymbol{\beta}}_{jk}$, the generalized Nelson-Aalen estimates (Njamen-Njomen and Ngatchou-Wandji, 2014) for the cause-specific baseline cumulative hazards are given by

$$\hat{H}_{0jk}(t) = \sum_{i:t_i \leq t} \left( \frac{\delta_{ijk}}{\sum_{l \in R_i} e^{\hat{\boldsymbol{\beta}}_{jk}^T \mathbf{z_l}}} \right), \quad \text{for} \quad j = 1, 2, ...., c, \quad \text{for} \quad k = A, B.$$

.

A drawback of modelling the cause-specific hazards is that, in order to estimate the overall survival

$$S(t|\mathbf{z}) = \exp\left\{ -\int_0^t \sum_{i=1}^{c} h_i(w|\mathbf{z}) dw \right\},$$

models need to be fitted for all types of events. The Fine and Gray model on the other hand relates the cumulative incidence functions more directly to the covariates and enables a clinician to assess the effects of the risk factors on the probability of an event of a particular type. The method makes use of the sub-distribution hazard, which is a function of the cumulative incidence for the corresponding cause of failure. The sub-distribution hazard is defined as

$$h_j^{sub}(t|\mathbf{z}) = \lim_{\Delta t \to 0} \frac{P\{t \leq T < t + \Delta t, \mathcal{C} = j | T \geq t \cup (T < t, \mathcal{C} \neq j); \mathbf{z}\}}{\Delta t}. \tag{4.5}$$

This would be the hazard obtained from the cumulative incidence function. Unlike the Cox proportional hazards model for the cause-specific hazards, the Fine and Gray model does not treat individuals failing from other causes as censored observations. The risk set at time $t$ for the latter model includes not only patients who have not yet experienced the event of interest, but also those who have failed from other causes before $t$. Fine and Gray (1999) pointed out that patients who have failed from causes other than the event of interest remain at risk for the cause of interest. They established a semi-parametric proportional hazards model for the sub-distribution hazard of the event of interest given by

$$h_{jk}^{sub}(t|\mathbf{z}) = h_{0jk}^{sub}(t)e^{\boldsymbol{\beta}_{jk}^T\mathbf{z}} \tag{4.6}$$

for $j = 1, 2, ...., c$, where $h_{0jk}^{sub}(t)$ is the baseline sub-hazard and is treated as a nuisance parameter in the model. This does not address the probability of failure from any cause. However, it directly relates to the cumulative incidence function through

$$h_{jk}^{sub}(t|\mathbf{z}) = -\frac{d}{dt}\log\{1 - F_{jk}(t|\mathbf{z})\} \tag{4.7}$$

for $j = 1, 2, ...., c$. Parameter estimation for the model depends on the right-censoring mechanism. It has been assumed throughout that the patients experience generalized type I right censoring, and therefore the censoring time is known, even for those who fail for other causes before the administrative censoring time. The partial likelihood approach is used to estimate the model parameters. For survival trials with right-censored observations, Fine and Gray (1999) developed a weighted score function to deal with dependent censoring. For a sample size of $n$ patients, if $\delta_{ijk}$ is the censoring indicator as before for the $i^{th}$ patient on treatment $k$ from cause $j$, the partial likelihood is defined by

$$L_k(\boldsymbol{\beta}_{1k}, \boldsymbol{\beta}_{2k}, ....., \boldsymbol{\beta}_{ck}) = \prod_{i=1}^{n_k}\prod_{j=1}^{c}\left(\frac{e^{\boldsymbol{\beta}_{jk}^T\mathbf{z_i}}}{\sum_{l\in R_i^{sub}} w_{il}e^{\boldsymbol{\beta}_{jk}^T\mathbf{z_l}}}\right)^{\delta_{ijk}}, \tag{4.8}$$

where the risk set for cause $j$ at time $t_i$ is $R_i^{sub} = \{l : \{t_l \geq t_i\} \cup \{t_l < t_i, \mathcal{C} \neq j\}\}$. Fine and Gray (1999) justified that the subjects experiencing a competing event before $t_i$ remain at risk from the main cause of failure. The weights for the Fine and Gray model can be calculated as

$$w_{il} = \begin{cases} 1 & \text{if } t_l \geq t_i \\ \frac{\hat{G}(t_i)}{\hat{G}(\min(t_l, t_i))} & \text{if } t_l < t_i,\ \mathcal{C} \neq j, \end{cases}$$

where $\hat{G}$ is an estimate of the survivor function for the censoring distribution, that is, the cumulative probability of still being followed up at time $t_i$. It can be estimated by the usual product limit method by treating the censored observations as event times.

Taking the logarithm of the partial likelihood in (4.8), the Breslow log-likelihood function for cause $j$ is given by

$$l(\boldsymbol{\beta}_{jk}) = \sum_{i=1}^{n_k} \delta_{ijk}\{\boldsymbol{\beta}_{jk}^T \mathbf{z_i} - \log(\sum_{l \in R_i^{sub}} w_{il} e^{\boldsymbol{\beta}_{jk}^T \mathbf{z_l}})\}.$$

Therefore, the partial score function for cause $j$ and treatment $k = A, B$ is given by;

$$\nabla l(\boldsymbol{\beta}_{jk}) = \sum_{i=1}^{n_k} \delta_{ijk} \left( \mathbf{z_i} - \frac{\sum_{l \in R_i^{sub}} \mathbf{z_l} w_{il} e^{\boldsymbol{\beta}_{jk}^T \mathbf{z_l}}}{\sum_{l \in R_i^{sub}} w_{il} e^{\boldsymbol{\beta}_{jk}^T \mathbf{z_l}}} \right).$$

The Hessian matrix of the partial log likelihood for patients with the $j^{th}$ cause of failure from treatment $k = A, B$ is

$$\nabla^2 l(\boldsymbol{\beta}_{jk}) = -\sum_{i=1}^{n_k} \delta_{ijk} \left\{ \frac{\sum_{l \in R_i^{sub}} \mathbf{z}_l \mathbf{z}_l^T w_{il} e^{\boldsymbol{\beta}_{jk}^T \mathbf{z}_l}}{\sum_{l \in R_i^{sub}} w_{il} e^{\boldsymbol{\beta}_{jk}^T \mathbf{z}_l}} - \frac{(\sum_{l \in R_i^{sub}} \mathbf{z}_l w_{il} e^{\boldsymbol{\beta}_{jk}^T \mathbf{z}_l})(\sum_{l \in R_i^{sub}} \mathbf{z}_l w_{il} e^{\boldsymbol{\beta}_{jk}^T \mathbf{z}_l})^T}{(\sum_{l \in R_i^{sub}} w_{il} e^{\boldsymbol{\beta}_{jk}^T \mathbf{z}_l})^2} \right\}.$$

Let,

$$\bar{\mathbf{z}}(t_i, \boldsymbol{\beta}_{jk}^T) = \frac{\sum_{l \in R_i^{sub}} \mathbf{z}_l w_{il} e^{\boldsymbol{\beta}_{jk}^T \mathbf{z}_l}}{\sum_{l \in R_i^{sub}} w_{il} e^{\boldsymbol{\beta}_{jk}^T \mathbf{z}_l}} = \sum_{l \in R_i^{sub}} \mathbf{z}_l w_l,$$

where;

$$w_l = \frac{w_{il} e^{\boldsymbol{\beta}_{jk}^T \mathbf{z}_l}}{\sum_{l \in R_i^{sub}} w_{il} e^{\boldsymbol{\beta}_{jk}^T \mathbf{z}_l}}$$

is the weight that is proportional to the hazard of the patient experiencing the event for a specific cause $j$ due to treatment $k$. Therefore, $\bar{\mathbf{z}}(t_i, \boldsymbol{\beta}_{jk})$ can be interpreted as the weighted average of the covariate vectors among those individuals still at risk from cause $j$ at time $t_i$ from treatment $k$ with weights $w_l$. The quantity,

$$V^{sub}(t_i, \boldsymbol{\beta}_{jk}) = \sum_{l \in R_i^{sub}} \{\mathbf{z}_l - \bar{\mathbf{z}}(t_i, \boldsymbol{\beta}_{jk})\}\{\mathbf{z}_l - \bar{\mathbf{z}}(t_i, \boldsymbol{\beta}_{jk})\}^T w_l$$

can be interpreted as the weighted covariance matrix of the covariates among those individuals still at risk from cause $j$ at time point $t_i$. If $\{V_{ss}^{sub}(t_i, \boldsymbol{\beta}_{jk})\}_{s=1}^p$ are the diagonal entries of the matrix $V^{sub}(t_i, \boldsymbol{\beta}_{jk})$, then $V_{ss}^{sub}(t_i, \boldsymbol{\beta}_{jk}) \geq 0$. Therefore the log

partial likelihood function has a unique maximizer, which can be obtained by equating the score function to zero and solving for $\boldsymbol{\beta}_{jk}$. So the weighted covariance matrix of the covariates among those individuals at risk from cause $j$ and treatment $k$ at time $t_i$ is positive definite. Hence, the observed information matrix for $\boldsymbol{\beta}_{jk}$, given by $J_{k(\text{sub})}(\boldsymbol{\beta}_{jk}) = \sum_{i=1}^{n_k} \delta_{ijk} V^{sub}(t_i, \boldsymbol{\beta}_{jk})$, is also positive definite. When there are $p$ covariates, the $p$-dimensional vector $\hat{\boldsymbol{\beta}}_{jk}$ of the estimated regression coefficients converges to a multivariate normal distribution with mean $\boldsymbol{\beta}_{jk}$ and covariance matrix $\mathrm{J}^{-1}_{k(\text{sub})}(\boldsymbol{\beta}_{jk})$. Since $\mathrm{J}_{k(\text{sub})}(\boldsymbol{\beta}_{jk})$ is positive definite, its unique inverse exists and is also positive definite. This is the basis of the Wald test, which is used for treatment comparisons in the presence of the concomitant information of the patients.

## 4.3 Proposed CARA Designs Based on Competing Risk Models

Information on patients accrues sequentially in a clinical trial as they are assigned to a treatment arm based on the design plan which the trial follows. Since clinical trials handle human patients, there is always an ethical concern of treating more patients with the better performing treatment arm during the trial phase. However, while comparing different competing treatments, this skewness of patient allocation should not affect the power of the Wald test for a treatment difference. Therefore, a design is needed which balances these two requirements. The CARA designs serves this purpose quite well. Now, often in survival trials patients do not fail due to a single cause. During the design phase of the trial, ignoring the information about the causes of failures of patients would severely bias the results of the experiment as well as the analysis. It is therefore appropriate to develop a CARA design which accounts for such information.

The designs introduced in this section are semi-parametric in nature. This means that they are free from any distributional assumption on the survival responses, but a lighter assumption about the proportionality of the sub-distribution hazards or the proportionality of the cumulative incidence functions between the two treatment arms has been made. Proportionality here refers to the ratio of the sub-distribution hazard

functions for patients or their cumulative incidence functions for a specific cause between the two treatment arms at a given point of time in the trial not being dependent on time. The censoring scheme assumed throughout this chapter is generalized type I right censoring , where the recruitment period is of length $R > 0$ and $D$ is the overall duration of the clinical trial. At time $D$, the subjects who have not experienced the event or have not already been right censored are considered to be generalized right censored of type I.

At the beginning of the trial, patients are equally allocated between the treatment arms using some kind of restricted randomization scheme. This is performed to collect initial data to estimate the unknown model parameters. After the initial allocation, one computes the partial likelihood estimates of the model parameters. From this stage, onwards, when a new patient enters the trial with his/her covariate information, he/she is randomly allocated to the better treatment arm found thus far in the trial in regards to the events due to the primary cause of interest. This decision of allocating the incoming patient to a particular treatment arm is based on the history of treatment assignments, responses, and covariate vectors of the previous patients, and also the covariate information of the incoming patient. Appropriate allocation functions are thus derived here which skews the patient allocation towards the better treatment arm for the main cause without compromising much on the power of the Wald test for a treatment difference. If the recruitment phase is sufficiently long and the number of accumulating responses during this phase is substantial, then it is possible to facilitate CARA randomization.

### 4.3.1 Semi-Parametric CARA Designs with Target for Competing Risks Models

One approach to derive CARA designs is to establish a suitable target allocation proportion and then use a randomization procedure to target the derived alloccation proportions. Survival trials are considered where patients experience the event due to multiple causes. Therefore, the causes of failure needs to be considered when deriving

the suitable target allocation proportions. These allocation proportions are functions of the Fine and Gray model paramters, which are estimated sequentially using the partial likelihood method with the arrival of every new patient in the clinical trial. Thus, it is useful to construct a CARA design based on the sub-distribution hazard function.

Let $\epsilon_{jk}(\mathbf{z};\boldsymbol{\beta}_{jk})$ be the probability of an event from the $j^{th}$ cause before censoring for a patient on treatment $k$ and with covariate vector $\mathbf{z}$. Then,if $C_i$ denotes the censoring time for the $i^{th}$ patient, we have $\epsilon_{jk}(\mathbf{z};\boldsymbol{\beta}_{jk}) = \mathrm{P}(T_{ik} \leq C_i \mid \mathbf{z};\ \boldsymbol{\beta}_{jk}\ ,\ \mathcal{C} = j)$.To meet most of the multiple experimental objectives in a clinical trial, one can minimize the overall sub-distribution hazard for a patient failing due to a specific cause and with a given set of covariates subject to the constraint of keeping the asymptotic variance of the difference between the estimated sub-distribution hazard functions for the two treatment groups achieved constant. This is by minimizing

$$n_{Aj}h_{jA}^{sub}(t|\mathbf{z}) + n_{Bj}h_{jB}^{sub}(t|\mathbf{z}),$$

$$\text{subject to}: \mathbf{z}^T\{\mathbf{z}^T J_{k(\text{sub})}^{-1}(\boldsymbol{\beta}_{jk})\mathbf{z}\}\mathbf{z}e^{2\boldsymbol{\beta}_{jk}^T\mathbf{z}} + \mathbf{z}^T\{\mathbf{z}^T J_{k(\text{sub})}^{-1}(\boldsymbol{\beta}_{jk})\mathbf{z}\}\mathbf{z}e^{2\boldsymbol{\beta}_{jk}^T\mathbf{z}} = \mathrm{k} > 0.$$

If $Va_{jk}\{\hat{h}_{jk}(t|\mathbf{z})\} = \mathbf{z}^T\{\mathbf{z}^T\bar{V}^{-1sub}(t,\boldsymbol{\beta}_{jk})\mathbf{z}\}\mathbf{z}e^{2\boldsymbol{\beta}_{jk}^T\mathbf{z}}$ denotes the variability adjustment factor for treatment $k$ and cause $j$, and $q_{jk}(\mathbf{z};\boldsymbol{\beta}_{jk}) = \epsilon_{jk}(\mathbf{z};\boldsymbol{\beta}_{jk})h_{jk}^{sub}(t|\mathbf{z})$, the optimal allocation proportion for treatment $A$ is given by:

$$\pi_{A1j}^{G}(\boldsymbol{\beta}_{Aj},\boldsymbol{\beta}_{Bj},\mathbf{z}) = \frac{\sqrt{q_{jB}(\mathbf{z};\boldsymbol{\beta}_{jB})Va_{jA}\{\hat{h}_{jA}(t|\mathbf{z})\}}}{\sqrt{q_{jB}(\mathbf{z};\boldsymbol{\beta}_{jB})Va_{jA}\{\hat{h}_{jA}(t|\mathbf{z})\}} + \sqrt{q_{jA}(\mathbf{z};\boldsymbol{\beta}_{jA})Va_{jB}\{\hat{h}_{jB}(t|\mathbf{z})\}}}. \tag{4.9}$$

For having a design with high power, the Neyman allocation function minimizes

$$n_{Aj} + n_{Bj},$$

$$\textit{subject to}: \mathbf{z}^T\{\mathbf{z}^T J_{k(\text{sub})}^{-1}(\boldsymbol{\beta}_{jk})\mathbf{z}\}\mathbf{z}_e^{2\boldsymbol{\beta}_{jk}^T\mathbf{z}} + \mathbf{z}^T\{\mathbf{z}^T J_{k(\text{sub})}^{-1}(\boldsymbol{\beta}_{jk})\mathbf{z}\}\mathbf{z}e^{2\boldsymbol{\beta}_{jk}^T\mathbf{z}} = \mathrm{k} > 0.$$

The optimal allocation proportion for treatment $A$ is given by:

$$\pi_{A2j}^{G}(\boldsymbol{\beta}_{Aj},\boldsymbol{\beta}_{Bj},\mathbf{z}) = \frac{\sqrt{\epsilon_{jB}(\mathbf{z};\boldsymbol{\beta}_{jB})Va_{jA}\{\hat{h}_{jA}(t|\mathbf{z})\}}}{\sqrt{\epsilon_{jB}(\mathbf{z};\boldsymbol{\beta}_{jB})Va_{jA}\{\hat{h}_{jA}(t|\mathbf{z})\}} + \sqrt{\epsilon_{jA}(\mathbf{z};\boldsymbol{\beta}_{jA})Va_{jB}\{\hat{h}_{jB}(t|\mathbf{z})\}}}. \tag{4.10}$$

In order to have a direct interpreation in terms of the probability of an event, one can also minimize the overall cumulative incidence functions for each cause subject to the asymptotic variance of the difference between the estimated cumulative incidence rates for the two treatments being constant. It is shown in Appendix F by an application of the delta method that

$$Var\{\hat{F}_{jk}(t|\mathbf{z})\} = (t\mathbf{z}^T)Var\{\hat{h}^{sub}_{jk}(t|\mathbf{z})\}(t\mathbf{z})(\exp[-2\int_0^t\{h^{sub}_{jk}(u|\mathbf{z})\}du]),$$

for $k = A, B$. Therefore minimize

$$n_{Aj}F_{jA}(t|\mathbf{z}) + n_{Bj}F_{jB}(t|\mathbf{z}),$$

$$subject\ to: Var\{\hat{F}_{jA}(t|\mathbf{z})\} + Var\{\hat{F}_{jB}(t|\mathbf{z})\} = \mathrm{k} > 0.$$

Let

$$Vr_{jk}(\boldsymbol{\beta}_{jk}, \mathbf{z}) = (t\mathbf{z}^T)Va_{jk}\{\hat{h}_{jk}(t|\mathbf{z})\}(t\mathbf{z})(\exp[-2\int_0^t\{h^{sub}_{jk}(u|\mathbf{z})\}du]).$$

If $q'_{jk}(\mathbf{z}; \boldsymbol{\beta}_{jk}) = \epsilon_{jk}(\mathbf{z}; \boldsymbol{\beta}_{jk})F_{jk}(t|\mathbf{z})$. The optimal allocation proportion for treatment $A$ for cause $j$ is given by

$$\pi^G_{A3j}(\boldsymbol{\beta}_{Aj}, \boldsymbol{\beta}_{Bj}, \mathbf{z}) = \frac{\sqrt{q'_{jB}(\mathbf{z}; \boldsymbol{\beta}_{jB})Vr_{jA}(\boldsymbol{\beta}_{jA}, \mathbf{z})}}{\sqrt{q'_{jB}(\mathbf{z}; \boldsymbol{\beta}_{jB})Vr_{jA}(\beta_{jk}, \mathbf{z})} + \sqrt{q'_{jA}(\boldsymbol{\beta}_{jA}, \mathbf{z})Vr_{jB}(\mathbf{z}; \boldsymbol{\beta}_{jk})}}. \quad (4.11)$$

The derivation of (4.11) is detailed in Appendix G.

One can also minimize the overall sample size for events from a specific cause subject to the asymptotic variance of the difference between the estimated cumulative incidence rate for the two treatments being constant. This yields the Neyman allocation given by

$$\pi^G_{A4j}(\boldsymbol{\beta}_{Aj}, \boldsymbol{\beta}_{Bj}, \mathbf{z}) = \frac{\sqrt{\epsilon_{jB}(\mathbf{z}; \boldsymbol{\beta}_{jB})Vr_{jA}(\mathbf{z}; \boldsymbol{\beta}_{jA})}}{\sqrt{\epsilon_{jB}(\mathbf{z}; \boldsymbol{\beta}_{jB})Vr_{jA}(\boldsymbol{\beta}_{jA}, \mathbf{z})} + \sqrt{\epsilon_{jA}(\mathbf{z}; \boldsymbol{\beta}_{jA})Vr_{jA}(\boldsymbol{\beta}_{jA}, \mathbf{z})}}. \quad (4.12)$$

### 4.3.2 Rules to Target the Derived Allocation Proportions

The dervied allocation proportions are functions of the competing risk model parameters which are estimated sequentially using the partial likelihood method. As soon as a new patient enters the trial, their covariate information is recorded. This is used along

with the treatment allocation history, response history and the covariate profile history to allocate the incoming patient to the better performing treatment arm for a given cause of failure, using a randomization procedure with low variability . An appropriate randomization procedure needs to be used which approaches the derived allocation proportions in the long run.

The covariate-adjusted doubly-adaptive biased coin design (CADBCD) serves this purpose well. This is a continuous allocation function that targets the derived allocation proportions. After $m$ patients are randomized, on the arrival of the $(m+1)^{th}$ patient with covariate vector $z_{m+1}$, let $\hat{\pi}_{mj} = \pi_{Aj}(\hat{\boldsymbol{\beta}}_{Aj,m}, \hat{\boldsymbol{\beta}}_{Bj,m}, \mathbf{z}_{m+1})$ represent the estimate of $\pi_{Aj}(\boldsymbol{\beta}_{Aj}, \boldsymbol{\beta}_{Bj}, \mathbf{z})$ based on the responses observed from the $m$ patients, adjusted for the covariate $\mathbf{z}_{m+1}$ of the incoming patient. Let $\{\hat{\rho}_{Amj} = \sum_{i=1}^{m} \pi_{Aj}(\hat{\boldsymbol{\beta}}_{Aj,m}, \hat{\boldsymbol{\beta}}_{Bj,m}, \mathbf{z}_i)/m\}$ be an estimate of the average target allocations for treatment A based on the data for the first $m$ patients. Using the CADBCD procedure, the $(m+1)^{th}$ patient can be assigned to treatment $A$ with probability $j_{m+1}[\{N_{Aj}(m)/m\}, \hat{\pi}_{mj}, \hat{\rho}_{Amj}]$, where $\{N_{Aj}(m)/m\}$ is the proportion of patients with events from cause $j$ who have been assigned to treatment $A$ after $m$ allocations. Therefore, the mathematical form of the allocation rule for the $(m+1)^{th}$ patient entering the clinical trial with covariate vector $\mathbf{z}_{m+1}$ is

$$j_{m+1}\left\{\frac{N_{Aj}(m)}{m}, \hat{\pi}_{mj}, \hat{\rho}_{Amj}\right\} = \begin{cases} \frac{\hat{\pi}_{mj}\{\hat{\rho}_{Amj}/\frac{N_{Aj}(m)}{m}\}^{\alpha}}{\hat{\pi}_{mj}\{\hat{\rho}_{Amj}/\frac{N_{Aj}(m)}{m}\}^{\alpha}+(1-\hat{\pi}_{mj})\{(1-\hat{\rho}_{Amj})/(1-\frac{N_{Aj}(m)}{m})\}^{\alpha}}, & \text{if } 0 < \frac{N_{Aj}(m)}{m} < 1, \\ 1 - \frac{N_{Aj}(m)}{m}, & \text{if } \frac{N_{Aj}(m)}{m} = 0 \text{ or } \frac{N_{Aj}(m)}{m} = 1. \end{cases} \tag{4.13}$$

If the clinician decides to have an ethical design with minimum variability, then the covariate-adjusted efficient randomized adaptive design (CAERADE) may be suitted better than the CADBCD. This is a discrete allocation function and is given by

$$j_{m+1}\left\{\frac{N_{Aj}(m)}{m}, \hat{\pi}_{mj}, \hat{\rho}_{Amj}\right\} = \begin{cases} \alpha'\hat{\pi}_{mj} \text{ if } \frac{N_{Aj}(m)}{m} > \hat{\rho}_{Amj}, \\ \hat{\pi}_{mj} \text{ if } \frac{N_{Aj}(m)}{m} = \hat{\rho}_{Amj}, \\ 1 - \alpha'(1-\hat{\pi}_{mj}) \text{ if } \frac{N_{Aj}(m)}{m} < \hat{\rho}_{Amj}, \end{cases} \tag{4.14}$$

In the above allocation rules, for CAERADE, $0 \leq \alpha' < 1$ is a constant that reflects the degree of randomization, whereas for CADBCD the degree of randomization is determined by $0 \leq \alpha < \infty$. Both the CADBCD and the CAERADE helps the observed allocation proportion approach the target allocation proportion for a given cause, in the long run. However, the CAERADE is first-order efficient and therefore attains the Cramer-Rao lower bound. Efron's biased coin design with $\alpha' = 2/3$ is a special case of it. For the CADBCD, $\alpha = 0$ corresponds to the most random case of the allocation rule, whereas $\alpha = \infty$ is the most deterministic scenario. On the other hand, for the CAERADE, it is considered that $\alpha' = 0$ is the most random scenario and $\alpha' = 1$ is the most deterministic rule. Note that the theory of the CADBCD and the CAERADE are based on parametric assumptions on the responses. There is currently no work on semi-parametric case. These can however be checked using simulations.

## 4.4 Simulation Results

A comparison of the different derived CARA randomization procedures is made by considering a two-arm survival trial with 800 patients who are failing due to two different causes. A patient's arrival time here is simulated from a uniform (0,365) distribution. The response time of a patient is added to the recruitment time of the patient and patients whose outcomes have not been observed by the pre-specified study time are said to be generalized type I right censored. The recruitment period here has been considered to be 365 days, and the overall trial duration is taken to be 581.66 days. A covariate structure of two independent covariates has been generated, which are gender (Bernoulli, p = 0.381) and age (Uniform(40,80)). The censoring time of the patients is simulated from a uniform (0,581.66) distribution. The survival time of a patient with covariate vector $\mathbf{z} = (1, z_1, z_2)^T$ in treatment group $k$ is simulated from the Weibull distribution with scale parameter $\mu_k(\mathbf{z}) = \exp(\boldsymbol{\beta}_k^T \mathbf{z})$ and shape parameter $\gamma_k = 2.48$ for patients failing from cause 1 and $\gamma_k = 0.53$ for patients failing from cause 2. Three choices of the treatment effects vector have been considered in this case, which are neutral effect of either treatment, positive effect of treatment $A$ and negative effect of

treatment $A$. The effects of the corresponding covariates for the simulation model $\mu_k(\mathbf{z})$ =exp($\boldsymbol{\beta}_k^T\mathbf{z}$), are summarized in Table 4.1 below :-

| **Model** | **Treatment** | **Covariate Effects** | | | **Covariate Effects** | | |
|---|---|---|---|---|---|---|---|
| | | **Cause 1** | | | **Cause 2** | | |
| | | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\beta_0$ | $\beta_1$ | $\beta_2$ |
| Neutral | A | 3.803 | 0.810 | 0.06 | 2.535 | 0.001 | 0.06 |
| | B | 3.803 | 0.810 | 0.06 | 2.535 | 0.001 | 0.06 |
| Negative | A | 0.429 | 0.810 | 0.06 | - 1.024 | 0.001 | 0.06 |
| | B | 3.803 | 0.810 | 0.06 | 2.535 | 0.001 | 0.06 |
| Positive | A | 7.176 | 0.810 | 0.06 | 6.094 | 0.001 | 0.06 |
| | B | 3.803 | 0.810 | 0.06 | 2.535 | 0.001 | 0.06 |

Table 4.1: Values of model parameters

In Table 4.1 the neutral treatment effect refers to the hypothetical experimental scenario where treatments $A$ and $B$ are equally effective. In the case of comparing a new treatment with a control, this scenario refers to the situation where the new treatment is as good as the existing control. The positive treatment effect refers to the hypothetical experimental scenario where treatment $A$ is more effective than treatment $B$, or the new treatment performs better than the control.The negative treatment effect refers to the hypothetical experimental scenario where treatment $B$ is more effective than treatment $A$, or in the case of comparing a new treatment with a control,this means that the new treatment is not as effective as the control. The procedure used here is a fully sequential one that recalculates the treatment effects using the Fine and Gray regression model after the arrival of every new patient and there are 5000 simulation runs. The competing randomization procedures and the corresponding designs number is listed in Table 4.2.

| Design | Competing Randomization Procedures |
|---|---|
| I | Completely randomized design |
| II | Efron's biased coin design with p = 2/3 |
| III | CARA DBCD with (4.9) as the target |
| IV | CARA DBCD with (4.10) as the target |
| V | CARA DBCD with (4.11) as the target |
| VI | CARA DBCD with (4.12) as the target |
| VII | CARA ERADE with (4.9) as the target |
| VIII | CARA ERADE with (4.10) as the target |
| IX | CARA ERADE with (4.11) as the target |
| X | CARA ERADE with (4.12) as the target |

Table 4.2: List of the Competing Designs

For the implementation of the designs in Table 4.2, for the $i^{th}$ patient′s randomization, only data from those patients who have responded before the $i^{th}$ patient's arrival are used in computing the randomization probability for the $i^{th}$ patient. Initially $2m_0$ patients have been equally allocated to the two treatment arms using Efron's biased coin design. Here, $m_0$ is a positive number and $2m_0$ is chosen to be 220, which is sufficiently large for the restricted randomization procedure to accurately estimate the model parameters. After the model parameters are estimated from the initial stage of the design using a Fine and Gray model, a randomization probability is calculated after each new patient who arrives sequentially into the trial. This randomization probability can be based on any one of the derived allocation functions. For the implementation of the CARA designs, at each stage of the trial, the number of events from each causes from the previous stages is recorded. The incoming patient is then assigned to a treatment arm according to the derived allocation proportion of the primary cause of interest in the clinical trial. A pseudo-random number generator is then used to draw a uniform random number between 0 and 1. If the derived randomization probability is greater than or equal to this random number, the patient is assigned to treatment $A$ or else the patient is assigned to treatment $B$. The procedure described is repeated for the subjects entering the trial in the future. For the CADBCD designs in Table 4.2, it must

be noted that following Zhang and Hu (2009) and Rosenberger and Hu (2004), the trade-off parameter for randomness is taken to be $\alpha = 2$, whereas, for the appropriate implementation of the CAERADE designs, $\alpha'$ is chosen to be 0.55. This number is a balance between 0.4 and 0.7 which have been set as the boundaries by Burman(1996).

The completely randomized design and Efron's biased coin design have been considered as the traditional balanced randomization procedures to be compared against the derived CARA designs. The performances of the competing randomization procedures in Table 4.2 for patients experiencing the event from cause 1, when the treatment effects between the arms do not differ, can be analysed from their operating characteristics in Table 4.3. For each experimental design, the significance level of the Wald test for testing the treatment difference has been set to 0.05.

| **Model** | **Design** | $\frac{N_A}{n}$ **(SE)** | **Event** | **Type I Error** |
|---|---|---|---|---|
| Neutral | I | 0.500 ( 0.033) | 0.60 | 0.05 |
| | II | 0.50 ( 0.020) | 0.60 | 0.05 |
| | III | 0.50 (0.030) | 0.60 | 0.04 |
| | IV | 0.50 (0.040) | 0.60 | 0.04 |
| | V | 0.50 (0.030) | 0.60 | 0.04 |
| | VI | 0.50 (0.038) | 0.60 | 0.04 |
| | VII | 0.50 ( 0.037) | 0.60 | 0.04 |
| | VIII | 0.50 (0.022) | 0.60 | 0.04 |
| | IX | 0.50 (0.037) | 0.60 | 0.05 |
| | X | 0.50 (0.022) | 0.60 | 0.04 |

Table 4.3: Performance of the Competing Designs for Cause 1 when Treatment Effects are Similar

It can be seen from Table 4.3 that, when there is no differences between the treatment effects, all of the procedures allocate equal numbers of patients to the treatment arms. The proportion of patients experiencing events from the main cause of interest is the same for all the designs. Since the trial consists of patients experiencing events

due to multiple causes, and the main interest is in the primary cause, therefore the proportion of patients experiencing events due to cause 1 has been considered here An important observation here is that the simulated type I error rate is slightly conservative for the derived CARA designs, apart from design IX. The standard errors for the type I error rates of the Wald test for treatment comparisons for the proportion of events are about 0.025. They are all around 0.003. The implementations of the integral function for the designs V, VI, IX and X are taken as summation of the sub-hazard fucntion upto time $t$. The performances of the competing designs for patients experiencing the event from cause 1 when the treatment effects between the arms differ significantly, can be analysed from their operating characteristics in Table 4.4.

On an average, the CARA designs with delayed responses result in a slight reduction in the proportion of events from the main cause of interest when there is any difference between the treatment effects, the standard errors being 0.025. This is because, unlike the balanced randomization procedures, the CARA designs result in more patients being allocated to the better treatment arm. This ethical gain is achieved by the derived CARA designs without compromising much on the power of the Wald test for treatment comparisons. The variability of the power is the lowest for Efron's biased coin design and CARA design X. Most of the powers for the CARA designs have the standard errors of 0.004, but those for designs I, VII, VIII and IX are 0.002. It can also be seen that, when targeting the Neyman allocation proportions, the CADBCD procedures are more variable than the corresponding CAERADE ones but this is not the case while targeting the other allocation proportions. This inconsistency in the performance of the randomization procedures may arise because the CADBCD and the CAERADE procedures are theoretically well defined for parametric resonses. Designs VIII and X are the most powerful CARA designs considered

The boxplots given in Figure 4.2 depicts the performances of the competing designs in the individual trials.

| **Models** | **Design** | $\frac{N_A}{n}$ **(SE)** | **Event** | **Power** |
|---|---|---|---|---|
| Positive | I | 0.50 (0.031) | 0.65 | 0.98 |
| | II | 0.50 (0.019) | 0.65 | 0.99 |
| | III | 0.65 (0.031) | 0.63 | 0.95 |
| | IV | 0.69 (0.039) | 0.63 | 0.92 |
| | V | 0.61 (0.031) | 0.63 | 0.96 |
| | VI | 0.62 (0.037) | 0.63 | 0.91 |
| | VII | 0.64 (0.035) | 0.63 | 0.93 |
| | VIII | 0.68 (0.022) | 0.63 | 0.96 |
| | IX | 0.62 ( 0.032) | 0.63 | 0.95 |
| | X | 0.63 (0.022) | 0.63 | 0.97 |
| Negative | I | 0.50 (0.032) | 0.62 | 0.98 |
| | II | 0.50 ( 0.020) | 0.62 | 0.99 |
| | III | 0.34 (0.031) | 0.59 | 0.93 |
| | IV | 0.31 (0.041) | 0.59 | 0.90 |
| | V | 0.39 (0.031) | 0.59 | 0.94 |
| | VI | 0.39 (0.039) | 0.59 | 0.90 |
| | VII | 0.35 (0.037) | 0.59 | 0.91 |
| | VIII | 0.32 (0.022) | 0.59 | 0.97 |
| | IX | 0.40 (0.034) | 0.59 | 0.92 |
| | X | 0.37 (0.022) | 0.59 | 0.97 |

Table 4.4: Performance of the Competing Designs from Cause 1 when Treatment Effects Differ Significantly

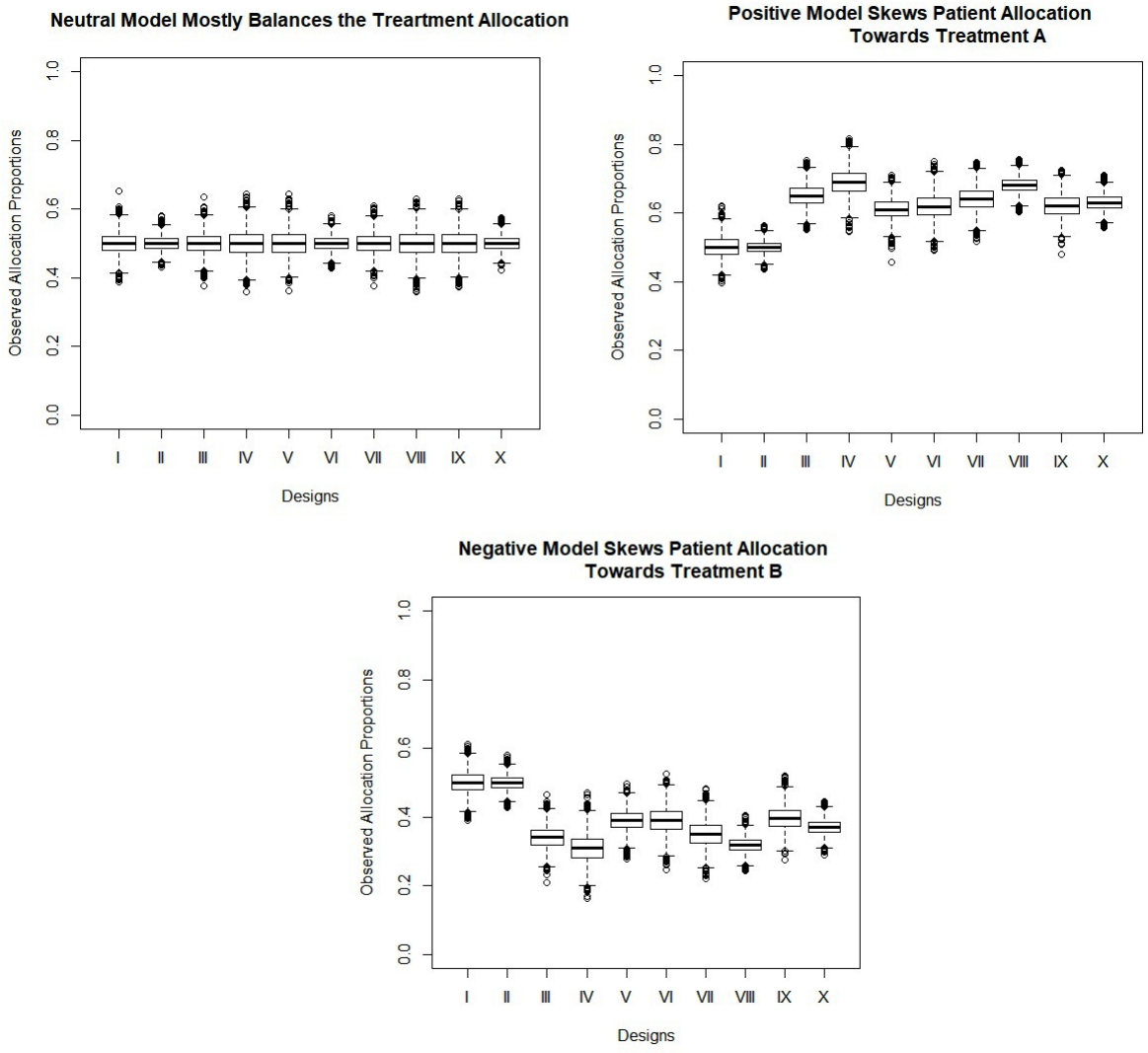


Figure 4.2: Performance of the Competing Designs based on the primary cause in the Individual Trials

The distribution of the observed allocation proportions appear to be very close to a symmetric distribution, but with different means and with different variability. When there is no difference between the treatment effects, all of the competing designs allocate equal numbers of patients on an average to the two treatment arms. On the other hand when there is a difference in the treatment effects the derived CARA designs allocate more patients to the better performing treatment in relation to the main cause of interest. As expected, Efron's biased coin design, which is a special case of the CAERADE designs, is the most efficient. This is because the asymptotic variability

of this design converges to zero. After taking care of the other risks due to competing events, in the presence of a difference between the treatment effects, the CARA designs allocate more patients to the better treatment arm for the main cause of interest. This ethical objective of a clinical trial is achieved without compromising much on the power of the Wald test for treatment differences.

## 4.5 Re-designing a Clinical Trial Using the Derived Procedures

The performance of the derived methodologies can be assessed after applying them to an existing clinical trial. Follicular cell lymphoma data from Pintilie (2007) can be considered which consists of 541 patients with early disease stage follicular cell lymphoma and treated in a 1:1 ratio with radiation alone (chemo = 0) or a combination treatment of radiation and chemotherapy (chemo = 1). Disease relapse or no response and death in remission are the two competing risks that are considered in this trial. The concomitant information of the patients that was recorded is their age (age: mean = 57 and sd = 14) and haemoglobin levels (mean = 138 and sd = 15) recorded. The median follow-up time was recorded to be 5.5 years. Out of the 541 patients, 272 experienced events due to the disease, with 76 competing risk events (death without relapse) being recorded and the rest were censored individuals.

A simulation study with 5,000 replications was conducted to compare four randomization designs: complete randomization, Efron's biased coin design,the CADBCD with target (4.9), and the CAERADE with target (4.9), have been considered. Based on data from Pintilie (2007), the parameters were chosen as follows : $R = 365$ days, $D = 2007.5$ days, age is uniformly distributed between 31.752 and 81.248 and the haemoglobin levels is uniformly distributed between 112.02 and 163.98. For each of these designs, the aim is to assess their performance in terms of treatment allocation and efficiency. The survival time of a patient with covariate vector $\mathbf{z} = (1, z_1, z_2)^T$ in treatment group $k$ is simulated from the exponential distribution with mean $\mu_k(\mathbf{z}) = \exp(\boldsymbol{\beta}_k^T \mathbf{z})$, for $k = A, B$ . Based on the reported overall median survival times, the covariate-adjusted treatment

effects have been set to 2.79 for the main cause of interest and to 1.72 for death without relapse the covariate adjusted haemoglobin effect has been set to 1.16 and the covariate adjusted effect of age is 4.12 for the main cause of interest, and these are 0.0127 and 0.09, respectively, for death without relapse. The intercept term from the exponential model is estimated to be 3.29 for the disease and 8.62 for the competing risk event. For appropriate implementation of the derived CARA designs 140 patients have been initially equally randomized to the two treatment arms using Efron's biased coin design before the adaptive randomization process started. The existing design has been compared with Efron's biased coin design and the CADBCD and CAERADE with target (4.9). The results are given in Table 4.5.

| **Designs** | **I** | **II** | **III** | **VII** |
|---|---|---|---|---|
| $N_A$/n (SE) | 0.50(0.035) | 0.50(0.018) | 0.66(0.032) | 0.67(0.035) |
| Event Proportion | 0.503 | 0.501 | 0.458 | 0.469 |
| Power | 0.95 | 0.96 | 0.91 | 0.89 |

Table 4.5: Comparison of the original study design with the derived CARA designs

It can be seen from Table 4.5 above that using the CARA designs instead of the completely randomized design or the Efron's biased coin design, results in more patients being allocated on an average to the better performing treatment. The balanced designs equalize the patient allocations between the two treatment arms. As a result, the CARA design has, on average lesser proportion of events for relapse after accounting for the other caues of events occuring in the trial, as compared to the balanced randomization procedures. It can be seen that the average power using the CARA designs are not significantly reduced as compared to the balanced randomization procedure. After performing a simulation study with 541 patients assuming that the treatment effects to be identical, the simulated type I error rates of the four procedures were between 0.04 and 0.05. Note that the standard errors for the proportions of events are no more than 0.06.

## 4.6 Discussion

This chapter has proposed new CARA randomization procedures for two-arm survival trials with competing risks. When patients are failing due to competing causes, an appropriate design needs to consider the causes of failure while allocating a patient to a particular treatment arm. Therefore, an attempt has been made here to develop methods to allocate patients to a particular treatment arm based on the main cause of failure after taking care of the other causes which exists in the trial.

CARA designs have been developed here based on on two distinct approaches to optimality: the CADBCD and the CAERADE, which targets various derived allocation proportions. These allocation proportions are derived by formal optimization techniques to achieve the ethical criteria of a clinical trial, but not at the cost of compromising on the efficiency. The target allocation proportions are derived for a model without covariates and their covariate-adjusted versions are used for the sequential allocation of patients.

The operating characteristics of the proposed CARA designs have been compared with two balanced randomization designs through simulation for a two-arm survival trial with two predictive covariates, right-censored delayed responses and staggered entry. It has been found that the proposed CARA designs generate skewed allocations according to covariate-specific treatment differences and result in fewer events in the trial for the main cause of failure, while having similar statistical properties to the balanced randomization designs. The proposed CARA designs achieve modest reductions in the proportion of events due to the main cause of interest in the study.

When comparing CARA designs based on the CADBCD and the CAERADE procedures, it has been seen that, when targeting some of the derived allocation proportions, the CADBCD gives the more efficient designs than the CAERADE. This is because the theoretical properties of the DBCD and the ERADE procedures were derived based on parametric assumptions on the responses. There is hardly any reference to their theoretical properties for the semi-parametric case. Simulation results here shows that when dealing with more complicated semi-parametric models such as the competing risk

Fine and Gray models, the usual properties of these designs do not hold. Therefore, an extended scope of this research work might be to develop the theoretical properties of these designs when the parameter estimates are not based on maximum likelihood procedures.

Another important area of potential application of CARA randomization procedures is competing risk survival trials with more than two treatment arms. Noted that the target allocation proportions derived here are based on optimal allocation for a model without covariates and then using its covariate-adjusted version for the sequential allocation of patients. Therefore, the information on the covariate history of the patients already enrolled in the trial is not taken into account when deriving this allocation proportion. It is only accounted for while calculating the Fine and Gray regression coefficient estimates, which the estimated covariate-adjusted allocation proportions are a function of. A useful alternative which can be explored in the future might be to derive an allocation proportion using some information about the distribution of the covariate profile of the previous patients and taking an integral function with respect to the covariate history of the product of the joint density function of the covariate history and the estimated covariate-adjusted version of the target allocation proportion.

It should be noted that our proposed designs are only appropriate for survival trials with long recruitment periods where the majority of patients contribute towards outcome data during the recruitment phase. The number of covariates in the model impacts sequential estimation at the design stage. Implementation of CARA designs should be considered only with a limited number of predictive baseline covariates.

# Chapter 5

# Discussion

## 5.1 Summary of the Main Conclusions

The thesis presented herewith reflects an attempt made to develop covariate-adjusted response-adaptive designs for phase III clinical trials for treatment comparison, with a view towards enhancing its scope of applications beyond what has been already considered. The clinical trials considered here are the ones which records time to event outcomes of patients who are administered to a particular treatment arm. This often occurs in oncology trials where the primary endpoints considered are often duration of remission, progression free survival, overall survival or event free survival of patients following a treatment or complete remission (CR). In such cases reducing the number of events during the trial phase becomes very important. This makes the clinical trial more ethically attractive.

Addressing the ethical criteria of a clinical trial has become imperative in every clinical trial experiments nowadays. The World Health Organization (WHO) has appealed to the pharmaceutical industries to adopt measures based on the ethical criteria as appropriate as possible, and monitor and enforce their standards. They mentioned that ethical criteria for drug promotion should lay the foundation for proper behaviour concerning the promotion of medicinal drugs, consistent with the search for truthfulness and righteousness. While the focus of ethical clinical trial conduct has been on proto-

col review in advance of the research, there has been a huge emphasis on monitoring subject welfare during the conduct of research.

Clinical research involving humans should be scientifically sound and conducted in accordance with basic ethical principles, which have their origin in the Declaration of Helsinki. The Good Clinical Practice (GCP) standard is followed by the FDA and the clinical domain maintains an international ethical and scientific quality for standard of designing, conducting, recording, and reporting reporting trials that involve the participation of human subjects.This is the reason why the primary principle of the WHO about GCP is to maintain ethical conduct during the process of a clinical trial. They emphasized three basic ethical principles with equal importance, namely respect for people, beneficence, and justice, that permeate all other GCP principles.

The clinical trials, specially conducted in the oncology theraputic area, which records time to event of patients thus become very attractive if it reduces the number of events during the process of clinical trial. This may be achieved by skewing the patient allocation towards the better performing treatment arm, during the course of the trial, by making use of the sequential arrival of patients. However, since clinical trials deal with human patients, they differ based on their personal characteristics. Such heterogeneity needs to be accounted for while skewing the patient allocation towards the better performing treatment arm.

The idea behind the traditional balanced randomization procedures has been to compare treatments with highest statistical efficiency. However while tending to achieve the balance, such designs allocates almost half the patients in the worse treatment arm, thus making the clinical trial ethically not very attractive. Also when the responses of the patients to a treatment arm follow a non-linear and heteroscedastic model, balanced allocation may not always give optimum results. The covariate-adjusted response-adaptive designs developed throughout this thesis balances these competing goals of addressing the ethical objective of the trial by skewing the patient allocation towards the better treatment arm, without compromising much on the statistical efficiency for treatment comparison. Such designs are specifically useful when the responses of the patients follow a heteroskedastic non-linear model and when the degree and the direction

of the treatment effect differs across patient subgroups. While allocating an incoming patient in the clinical trials to a particular treatment arm the covariate-adjusted response-adaptive designs considers the history of previous treatment assignments, response history of patients during the trial phase, the history of covariate information of patients already admitted in the trial, and also the covariate information of this incoming patient in the trial. Throughout the thesis the arrival of the patients has been considered to be fully sequential. For the implementation of the derived covariate-adjusted response-adaptive designs, initially certain number of patients are equally allocated to the two competing treatment arms using a balanced randomization procedure. After the superiority of a treatment arm is established in this interim stage of the trial, patients who arrive sequentially in the trial are allocated to the better treatement arm according to the derived covariate-adjusted response-adaptive randomization procedure. This process continues until a treatment is declared to be a clear winner. Throughout this thesis the overall sample size in the clinical trial is considered to be fixed, while performing the adaptation process.

Very limited research work exists in the literature about developing covariate-adjusted response-adaptive designs for surival trials. Sverdlov, Rosenberger and Ryznik (2013) developed such designs considering the survival responses of the patients to be exponentially distributed. The work in this thesis extends the applicability of such designs beyond the boundary of exponential survival responses, making it more appealing in real-life clinical trials. After pointing out the limitations of the applicability of the designs based on exponential survival responses due to its constant hazard property, covariate-adjusted response-adaptive designs have been developed for Weibull surival responses. It is known that Weibull distribution is categorized with an extra shape parameter which determines the shape of the hazard function. Exponential distriution is a special case of the Weibull model when the shape parameter is unity. Thus considering the development of covariate-adjusted response-adaptive designs for Weibull responses, enhances the scope of application of such designs in real-life clinical trials when the hazard of the patients to an event is non-constant.

The covariate-adjusted response-adaptive designs for Weibull distributed survival

responses are based on two distinct approaches to optimality: the covariate-adjusted doubly adaptive biased coin design (Zhang and Hu 2009) and the covariate-adjusted efficient randomized adaptive design (CAERADE) and also a non-optimality based approach : the glink function. The design based on the glink function is derived based on the cumulative distribution function of a Gumbel model whose location parameter is calculated as the reciprocal of the scale paramenter of the Weibull accelerated life model calculated from the covariate information of the incoming patient and its scale parameter is calculated by the inverse of the shape parameter calculated from the Weibull accelerated life model based on the information based on the previous patients. This bridges the past allocation, response histories and the present allocation pattern after allowing for the incorporation of prognostic factors. The Gumbel model being asymmetric and light-tailed, provides more weight to the available data, and tends to allocate more patients to the better treatment. Moreover, when the response of the patients follow a Weibull distribution, the design based on Gumbel model is more appropriate as compared to other continuous models because the theoretical errors in the Weibull Aceelerated Life regression model follow a Gumbel distribution. The arbitrariness of choosing a value for the tuning parameter T present in the design based on the probit link function by Bandyopadhyay and Biswas (2001) is not present in the design based on the glink function. Scaling the the estimated covariate adjusted treatment difference by its standard error plus the estimated hazard ratio makes this design more consistently applicable.

After comparing the operating characteristics of the derived designs along with the response-adaptive and the traditional balanced randomization designs it is found that all the proposed covariate-adjusted response-adaptive designs generate skewed allocations towards the better treatment, according to covariate-specific treatment effects and thus result in fewer events in the trial, without compromising much on the statistical efficiency as compared to the balanced randomization designs. The degree of skewness also varies according to the background model the design is based on. It has been established that the ethical gain of allocating more patients to the better treatment arm persists for Weibull survival responses based designs, without heavily compromising on the statistical power of the Wald test for the difference of covariate adjusted treatment

effects. It has been established that when there is a significant treatment difference, among all the derived allocation proportions, it is the Neyman allocation proportion that assigns more patients to the worse treatment and therefore is not ethically quite attactive as compared to the other competing designs. The Taylor series expansion of the random non-centrality parameter of the asymptotic Wald test for the difference of covariate adjusted treatment effects estalishes the inverse proportionality between its power and the variance of the random observed allocation proportions. It is this relation which has been heavily used for comparing the derived designs.

It has been seen that when the survival responses of the patients follow a Weibull distribution, the CAERADE being the asymptotically most efficient, increases the power or treatment comparison as compared to the corresponding covariate-adjusted Doubly Adaptive Biased Coin Design (CADBCD). However, sometimes the CAERADE does not converge to the target allocation proportion as fast as the CADBCD does. This is mainly because the allocation probabilities in the CAERADE are not stable. The allocation function being discrete, they always jump from one value to another. A continuous allocation function like the CADBCD can make the allocation probabilities stable and speed up the convergence of the sample allocation proportions. Therefore if the requirement of the clinician is to solely have an efficient CARA design, the CAERADE should be the best choice among all the covariate-adjusted response-adaptive randomization procedures. However in a real-life clinical trial, ethical considerations would appear to be quite unavoidable. Thus the CADBCD is the best achievable balance between ethics and efficiency when the survival responses conform to a Weibull model.

Often in real-life clinical trials the response of patients fail to conform to a parametric model. However, parametric assumption enhances the robustness and efficiency of the parameter estimates. Therefore to strike a balance between robustness and practicality, attempt has also been made to develop covariate-adjusted response-adaptive designs for survival responses following a semi-parametric model. Here designs are developed obviating any parametric assumptions about survival responses of the patients but only considering that the hazard of patients at any given time-point is proportional

to each other. The underlying model used to develop such kind of covariate-adjusted response-adaptive designs is the Cox proportional hazard model. Here the derived designs are a function of the Cox regression coefficients which are used to obtain the partial likelihood estimate of the treatment effect in order to allocate more patients to the better performing treatment arm during this course of the trial. As in the case of the derived covariate-adjusted response-adaptive designs for Weibull survival responses, here the designs are derived optimally by fixing the asymptotic variance of the covariate adjusted treatment difference to a constant value. This helps the statistical efficiency of the design not to get compromised much while achieving the ethical objective of skewing the patient allocation to the better performing treatment arm. Various optimal allocation proportions are derived minimizing the overall hazard during the trial and also by minimizing the overall trial size which gives the Neyman allocation. As with the designs for Weibull distributed survival responses, the CADBCD and the CAERADE are also used here to target the derived optimal allocation proportions.

Among the other optimality approach, the weighted optimality approach (Atkinson and Biswas, 2005) has also been considered. By appropriately choosing the tradeoff parameter $\eta$, the weighted optimality CARA designs also gives skewed allocations towards the better treatment arm and balances the objectives of statistical efficiency and ethics. However judiciously selecting the parameter depends on the experimental crteria of the clinician and the obejective of the clinical trial.

Apart from the optimal designs, the cumulative distribution function of the normal model with mean zero and standard deviation determined by the covariates of the incoming patient, has also been used to derive the covariate-adjusted response-adaptive designs for survival responses following a semi-parametric model. The design is developed making the cumulative distriution function of such a normal model depend on the partial likelihood estimate of the covariate adjusted treatment difference obtained from the sequentially fitted Cox regression model. The probit link function being an increasing function of the covariate adjusted treatment difference make this allocation procedure favour the treatment doing better at the particular stage of the clinical trial. The design scales the covariate adjusted treatment differences according to the hazard

ratio between the two treatments and the standard error of the treatment difference for the $mth$ patient. This means that if the hazard of an event for a particular treatment group is greater than the other, there would be less chance of allocating the next patient to that particular treatment arm. The probit link function here thus bridges the past history to the present allocation pattern for patients whose responses belong to the Lehmann family.

Apart from using a link function or a formal optimization procedure, CARA designs can also be developed by the method of treatment effect mapping similar to Rosenberger and Sheshaiyer (1997). The score function of the $s^{th}$ covariate in the Cox regression model has been used to develop a mapping onto [0,1] that exceeds 0.5 if treatment $A$ has been doing better thus far, and is less than 0.5 if treatment $B$ has been doing better. The idea is similar to that of Rosenberger (1993) in dealing with immediate continuous outcomes using a nonparametric rank test.

When comparing CARA designs based on covariate adjusted optimal targets, the results obtained were very similar to the ones obtained for Weibull distributed survival responses. However the score based design tends to have the lowest power for treatment comparison as compared to other designs and therefore is not recommended. The design based on the probit link is ethically most attractive as it allocates more of the patients to the better treatment arm. However, when there is no treatment difference it has a over-inflated type I error rate. Therefore it is also not recommended. The covariate-adjusted response-adaptive designs based on optimality approach thus stands out and is a suitable alternative to the traditional balanced randomization procedures for semi-parametric survival responses of the patients.

Often in medical research response to a treatment can be classified in terms of failure from disease processes and/or non-disease-related causes. In such cases, time to the event of interest cannot be observed because of a preceding event i.e. a competing event occuring before. An example can be of an event of interest being a specific cause of death where death from any other cause can be called a competing event. Such scenarios in survival analysis is termed as competing risk. When patients are failing due to competing causes, an appropriate design needs to consider the causes of failure

while allocating a patient to a particular treatment arm. The underlying model used to develop such kind of covariate-adjusted response-adaptive designs is the Fine and Gray sub-distribution hazard model. Here the derived designs are a function of the Fine and Gray regression coefficients which are used to obtain the partial likelihood estimate of the treatment effect in order to allocate more patients to the better performing treatment during this course of the trial. New covariate-adjusted response-adaptive randomization designs have been proposed for two-arm survival trials with competing risks. CARA designs for competing risk scenario have been developed here based on on two distinct approaches to optimality: the CADBCD and the CAERADE, which targets various derived allocation proportions. Operating characteristics of the proposed CARA designs have been compared with two balanced randomization designs through simulation, for a two-arm survival trial with two predictive covariates, right-censored delayed responses, and staggered entry. It has been seen that the derived CARA designs are a suitable alternative to the traditional balanced randomization designs. However it has been seen that while targeting some of the derived allocation proportions, the CADBCD gives the most efficient designs. This is because the theoretical properties of the CADBCD and the CAERADE procedures were derived based on parametric assumption of the responses. There is hardly any reference about the theoretical properties for semi-parametric alternative of these procedures. Simulation results reveals that while moving to more complicated semi-parametric models than Cox regression model, such as the competing risk models, the usual properties of these designs does not hold. Therefore an extended scope of this research work might be to develop these designs when the parameter estimates are not based on maximum likelihood procedures.

It should be noted that our proposed designs are only appropriate for survival trials with long recruitment periods where majority patients contribute towards outcome data during the recruitment phase. The number of covariates in the model impacts sequential estimation at the design stage. Implementation of CARA designs should be considered only with a limited number of the predictive baseline covariates.

The idea behind developing covariate-adjusted response-adaptive randomization designs in this thesis has been to make such designs more fruitfully applicable in survival

trials in industrial enterprise. Such trials are primarily conducted in oncology related clinical trials. This studied inadequacy may be attributed to a desire to enable even a non-expert in the field of adaptive designs, to have a fairly adequate understanding about the development of this area of research.

## 5.2 Critical Evaluation

In recent years, the use of adaptive design methods in pharmaceutical research and development has become popular due to its flexibility and efficiency for identifying potential signals of clinical benefit of the treatment under investigation. The flexibility and efficiency, however, increase the risk of operational biases, resulting in decrease in the accuracy and reliability for assessing the treatment effect of the treatment under investigation. This is because fully sequential strategies require outcomes from all previous allocations prior to the next allocation. This can prolong an experiment unduly. Thus there has been a lot of research going on for establishing a suitable delayed response model. Hardwick,Oehmke and Stout (2006) proposed a delayed rate response bandit model where they showed that except when the delay rate is several orders of magnitude different than the patient arrival rate, the delayed response bandit is nearly as efficient as the immediate response bandit. However such model does not take care of individual heterogeneity of patients.

The biggest challenge in applying such designs in real-life clinical trials is that it faces a lot of logiistical issues. Specially in survival trials, one needs to wait until a long period of time until they obseve an event or a censoring. Thus in fully sequential trial we are not always in a position to observe the previous response when the next patient arrives. This may have a disadvantage of dampening the convergence of the randomization procedure to its target. A possible way out of this problem is to obtain a surrogate endpoint in survival trial and build the design on the surrogate endpoint instead of the real clinical enpoint.

In clinical trials, a surrogate endpoint is a measure of effect of a specific treatment that may correlate with a real clinical endpoint but does not necessarily have

a guaranteed relationship. Surrogate endpoints are used when the primary endpoint is undesired, or when the number of events in the trial is very small, thus making it impractical to conduct a clinical trial to gather a statistically significant number of endpoints. The FDA and other regulatory agencies will often accept evidence from clinical trials that show a direct clinical benefit to surrogate endpoints. Surrogate endpoints can be obtained from different modalities, such as, behavioural or cognitive scores, or biochemical biomarkers. A correlation does not make a surrogate. It is a common misconception that if an outcome is correlated with the true clinical outcome, it can be used as a valid surrogate end point. However, proper justification for such replacement requires that the effect of the intervention on the surrogate end point predicts the effect on the clinical outcome. Progression Free Survival (PFS) is a prominent example of a surrogate endpoint in oncology contexts. However there are examples of cancer drugs (eg: Avastin) approved on the basis of progression-free survival, failed to show subsequent improvements in overall survival in subsequent studies. There have also been a number of instances when studies using surrogate endpoints have been used to show benefit from a particular treatment, but later, a repeat study looking at endpoints has not shown a benefit, or has even shown harm.

As pointed out by Rosenberger and Lachin (2002), a common argument against practically implementing response-adaptive designs or even covariate-adjusted response-adaptive designs is that the ethical advantage gained using such experimental designs are on an average over several trials. It does not guarantee such success rate in every single experiment.

In order to make the derived covariate-adjusted response-adaptive designs more applicable for implementation in real-life clinical trials, building these designs or more complicated survival models makes the interim phase of the clinical trial significantly long before the actual adaptation may start. These are the major issues for which the Food and Drug Administration (FDA) has shown extremely reserved approach in implementing adaptive randomization procedure in real-life clinical trials.

## 5.3 Directions for Future Research

Despite various reservations from the FDA about implementing adaptive randomization procedures, active research on developing covariate-adjusted response-adaptive designs has continued to be a hot topic of discussion in variours symposiums and conferences across the globe. The interest of researchers on developing such designs have significantly grown in order to give it more applied look so that it eventually gets applied in real-life clinical trials.

A significant step forward in developing covariate-adjusted response-adaptive designs for real-life clinical trials would be on building a theoretical model for estimation of the size of the interim stage given the theoretical model the design would be based on when the adaptation would start. Till now very limited discussion exists in the literature about this point and the ones which do, relies on simulation procedure to determine the size of the interim stage for equal allocation of patients. Building a theoretical model for estimating the size of the interim stage for equal allocation of patients to the competing treatment arms would enhance the stance of the statisticians to have a more robust justification to the clinicians about the waiting period before the model based adaptation process would start.

The exponential distribution is a special case of the Weibull model, but the Weibull model belongs to a wider class of two-parameter location-scale survival distributions, that encompasses distributions such as log-normal and log-logistic. When considering the development of CARA designs for survival responses belonging to the parametric family of distributions, the ambit of applications of the CARA designs can be enhanced further if the derived CARA designs based on Weibull model can be developed to include all models in the two-parameter location-scale class of distributions.

If the shape parameter of the Weibull distribution is known beforehand, the relationship between the exponential and the Weibull distribution can be used to simplify the results in Chapter 2. The Weibull distributed random variable can be raised to the power of the value of this known shape parameter and as a new patient sequentially arrives in a clinical trial, the CARA designs based on the exponential distribution can

be applied on the transformed response after considering all the information which a CARA design conditions on while calculating the randomization probability.

Often in clinical trials hazards of an event are non-proportional. However often the odds of surviving beyond a given time-point is proportional between treatment groups. Conceptually, instead of considering the proportionality of the hazard rate of individuals from the two treatment groups, one can also use the proportionality of the odds of survival of patients from the two treatment groups as the basis for development of CARA randomization procedures.

The essence of using CARA designs in real-life clinical trials becomes slightly more prominent when more than two competing treatment arms are considered. Instead of considering the overall sample size to be fixed, one can develop a suitable stopping rule which would help dropping a treatment arm in the trial before all patients are randomized. This would make the clinical research more cost effective.

Often in real-life clinical trials one observe patients dropping out due to efficacy related causes of a treatment under consideration. Such cases are known as informative censoring. In this case, the survivor function of the event time is a power function of that of the censoring time. Very little work has been done till date on developing covariate-adjusted response-adaptive designs for survival responses considering informative censoring. A possible and fruitful area of exploration might be to develop Covariate-adjusted response-adaptive designs for all the types of different survival models considered in this thesis but considering informative censoring instead of generalized type I right censoring.

In all the CARA designs developed till date using a formal optimization method, the derived target allocation proportions are based on optimal allocation for a model without covariates. The covariate-adjusted version of these derived target allocation proportion are then used for the sequential allocation of patients. Therefore the information of the covariate history of the patients already enrolled in the trial is not taken into account while deriving this allocation proportion. An useful alternative which can be explored in the future might be to derive an allocation proportion using

some information about the distribution of the covariate profile of the previous patients and marginalizing the distribution of this covariate profile over the covariate adjusted version of the target allocation proportion. This would make the designs more ethical as well as it would fit the best within the actual definition of covariate-adjusted response-adaptive designs.

Finally, more research has to be put into the development of a robust randomization procedure to target the derived allocation proportions when the distribution of the survival responses move away from conforming to a parametric model. It has been seen through simulation that while obviating any parametric assumption about the distribution of the survival responses, when the underlying model becomes more complex than the usual Cox proportional hazard model, the usual properties of the CADBCD and the CAERADE does not hold true. As pointed out before, in real-life clinical trials survival responses rarely follow a parametric form of distribution. Moreover non-proportionality of hazard is an active area of discussion in various industrial research. Therefore the adequacy of a robust randomization procedure to target the formally optimized target allocation proportion seem to be an important ingredient of research for developing covariate-adjusted response-adaptive designs. In practice, applied researchers do not only want to be protected from adhering to a model based on a parametric assumption on survival responses , but they also want to obtain a robust randomization procedure which would be effficient in all practical scenario. Therefore finding such semi-parametric or non-parametric randomization procedures which would target the allocation proportions derived by formal optimization methods might be a big step forward in the development of covariate-adjusted response-adaptive designs for its future application in real-life clinical trials.

# Appendix A

# Deriving the variance of the Weibull distribution parameter estimates

Consider two random samples ($t_{ik}$,$\delta_{ik}$), (where $k = A$ and $B$) from Weibull distributions with parameters $\{\mu_A(\mathbf{z}), \gamma_A\}$ and $\{\mu_B(\mathbf{z}), \gamma_B\}$. Therefore two such distributions corresponding to treatments $A$ and $B$ can be compared using the Wald test. To test the hypotheses,

$$H_0 : \log\{\mu_A(\mathbf{z})\} = \log\{\mu_B(\mathbf{z})\}$$

$$H_A : \log\{\mu_A(\mathbf{z})\} \neq \log\{\mu_B(\mathbf{z})\},$$

the Wald test statistic

$$T_n = \frac{\log\{\hat{\mu}_A(\mathbf{z})\} - \log\{\hat{\mu}_B(\mathbf{z})\}}{\sqrt{var[\log\{\hat{\mu}_A(\mathbf{z})\}] + var[\log\{\hat{\mu}_B(\mathbf{z})\}]}}$$

can be used, where $T_n \xrightarrow{d} \mathrm{N}(0,1)$, and $\hat{\mu}_A(\mathbf{z})$ and $\hat{\mu}_B(\mathbf{z})$ are the maximum likelihood estimators of $\mu_A(\mathbf{z})$ and $\mu_B(\mathbf{z})$ respectively. Let $\hat{\gamma}_k$ be the maximum likelihood estimator of $\gamma_k$ . The MLE of $\log\{\mu_k(\mathbf{z})\}$ and $(1/\gamma_k)$ can be obtained by numerically solving the equations (2.12) and (2.13). The Fisher information matrix is obtained by finding the Hessian matrix from equations (2.12) and (2.13) and taking the expectations after changing the sign of the entries in the Hessian matrix. The Fisher information matrix

here is

$$I[\log\{\mu_k(\mathbf{z})\},(1/\gamma_k)] = \begin{bmatrix} n_k\epsilon_k(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k)\gamma_k^2 & n_kE(\varsigma_{ik}e^{\varsigma_{ik}})\gamma_k^2 \\ n_kE(\varsigma_{ik}e^{\varsigma_{ik}})\gamma_k^2 & n_k\epsilon_k(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k)\gamma_k^2 + n_kE(\varsigma_{ik}^2e^{\varsigma_{ik}})\gamma_k^2 \end{bmatrix},$$

where, $\varsigma_{ik} = (\gamma_k)[y_{ik} - \log\{\mu_k(\mathbf{z})\}]$ , $\epsilon_k(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k)$ is the probability of an event before censoring, and $n_k$ is the number of patients in treatment $k$ ($k$= $A, B$). The determinant of the matrix $I[\log\{\mu_k(\mathbf{z})\},(1/\gamma_k)]$ is given by

$$det\ I[\log\{\mu_k(\mathbf{z})\},(1/\gamma_k)] = n_k^2\gamma_k^4\{\epsilon_k^2(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k) + \epsilon_k(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k)E(\varsigma_{ik}^2e^{\varsigma_{ik}}) - E(\varsigma_{ik}e^{\varsigma_{ik}})^2\}.$$

Hence the variance-covariance matrix of $\log\{\mu_k(\mathbf{z})\}$ and $1/\gamma_k$ can be obtained from $M = I^{-1}[\log\{\mu_k(\mathbf{z})\}, 1/\gamma_k]$ which is given by

$$M = \begin{bmatrix} \frac{1}{n_k\gamma_k^2}\frac{\epsilon_k(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k)+E(\varsigma_{ik}^2e^{\varsigma_{ik}})}{\epsilon_k^2(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k)+\epsilon_k(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k)E(\varsigma_{ik}^2e^{\varsigma_{ik}})-E(\varsigma_{ik}e^{\varsigma_{ik}})^2} & -\frac{1}{n_k\gamma_k^2}\frac{E(\varsigma_{ik}e^{\varsigma_{ik}})}{\epsilon_k^2(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k)+\epsilon_k(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k)E(\varsigma_{ik}^2e^{\varsigma_{ik}})-E(\varsigma_{ik}e^{\varsigma_{ik}})^2} \\ -\frac{1}{n_k\gamma_k^2}\frac{E(\varsigma_{ik}e^{\varsigma_{ik}})}{\epsilon_k^2(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k)+\epsilon_k(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k)E(\varsigma_{ik}^2e^{\varsigma_{ik}})-E(\varsigma_{ik}e^{\varsigma_{ik}})^2} & \frac{1}{n_k\gamma_k^2}\frac{\epsilon_k(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k)}{\epsilon_k^2(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k)+\epsilon_k(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k)E(\varsigma_{ik}^2e^{\varsigma_{ik}})-E(\varsigma_{ik}e^{\varsigma_{ik}})^2} \end{bmatrix}.$$

Now, equation (2.16) gives the mathematical form of $G_k$. Therefore, $M$ can be written as

$$M = \begin{bmatrix} \frac{G_k}{n_k\gamma_k^2} & -\frac{1}{n_k\gamma_k^2}\frac{E(\varsigma_{ik}e^{\varsigma_{ik}})}{\epsilon_k^2(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k)+\epsilon_k(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k)E(\varsigma_{ik}^2e^{\varsigma_{ik}})-E\{\varsigma_{ik}e^{\varsigma_{ik}}\}^2} \\ -\frac{1}{n_k\gamma_k^2}\frac{E(\varsigma_{ik}e^{\varsigma_{ik}})}{\epsilon_k^2(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k)+\epsilon_k(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k)E(\varsigma_{ik}^2e^{\varsigma_{ik}})-E(\varsigma_{ik}e^{\varsigma_{ik}})^2} & \frac{1}{n_k\gamma_k^2}\frac{\epsilon_k(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k)}{\epsilon_k^2(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k)+\epsilon_k(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k)E(\varsigma_{ik}^2e^{\varsigma_{ik}})-E(\varsigma_{ik}e^{\varsigma_{ik}})^2} \end{bmatrix}.$$

Therefore,

$$\hat{\sigma}^2 = \widehat{var}[\log\{\hat{\mu}_k(\mathbf{z})\}] = \frac{G_k}{n_k\hat{\gamma}_k^2},$$

and,

$$\begin{aligned}\widehat{var}((1/\hat{\gamma}_k)) &= \frac{1}{n_k\hat{\gamma}_k^2}\frac{\hat{\epsilon}_k(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k)}{\hat{\epsilon}_k^2(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k) + \hat{\epsilon}_k(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k)E(\varsigma_{ik}^2e^{\varsigma_{ik}}) - E(\varsigma_{ik}e^{\varsigma_{ik}})^2} \\ &= \frac{\hat{\sigma}^2\hat{\epsilon}_k(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k)}{\hat{\epsilon}_k(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k) + E(\varsigma_{ik}^2e^{\varsigma_{ik}})}.\end{aligned}$$

# Appendix B

# Derving the Analytical Form for $\epsilon_k(\mathbf{z}; \boldsymbol{\beta}_k, \gamma_k)$

It is known that $\epsilon_k(\mathbf{z}; \boldsymbol{\beta}_k, \gamma_k)$ = P ( $T_{ik} \leq C_{ik}$ $|\mathbf{z}; \boldsymbol{\beta}_k,\ \gamma_k)$ and $\mathcal{T}_{ik} = \min(T_{ik}, C_{ik})$. The survival outcomes are assumed to conform to a Weibull distribution with scale parameter $\mu_k(\mathbf{z})$ and the shape parameter $\gamma_k$ and the right censored times $C_{ik}$ are assumed to follow uniform distribution with parameters 0 and $D$. $D$ is considered to be the trial duration. Let A be a set such that A = $(t_{ik} : T_{ik} \leq C_{ik})$. Therefore, $\epsilon_k(\mathbf{z}; \boldsymbol{\beta}_k, \gamma_k)$ = P { $(T_{ik}, C_{ik})$,$\in$ A } or,

$$\epsilon_k(\mathbf{z}; \boldsymbol{\beta}_k, \gamma_k) = \int_0^D \int_0^{c_{ik}} f(t_{ik}, C_{ik}) dt_{ik} dC_{ik},$$

substituting the joint density function of the event times and the censored times we get,

$$\epsilon_k(\mathbf{z}; \boldsymbol{\beta}_k, \gamma_k) = \int_0^D \frac{1}{D} \int_0^{c_{ik}} \frac{\gamma_k}{\mu_k(\mathbf{z})} \{t_{ik}/\mu_k(\mathbf{z})\}^{(\gamma_k - 1)} e^{-\{t_{ik}/\mu_k(\mathbf{z})\}^{(\gamma_k - 1)}} dt_{ik} dC_{ik}.$$

(Assuming that the survival times and the right censored times are independent); Performing a change of coordinates in the set A,

$$\epsilon_k(\mathbf{z}; \boldsymbol{\beta}_k, \gamma_k) = \int_0^D \frac{1}{D} \int_{t_{ik}}^D \frac{\gamma_k}{\mu_k(z)} \{t_{ik}/\mu_k(\mathbf{z})\}^{(\gamma_k - 1)} e^{-\{t_{ik}/\mu_k(\mathbf{z})\}^{(\gamma_k - 1)}} dC_{ik} dt_{ik}.$$

Integrating with respect to $C_{ik}$ we get,

$$\epsilon_k(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k) = \frac{1}{D}\int_0^D (D-t_{ik})\frac{\gamma_k}{\mu_k(\mathbf{z})}\{t_{ik}/\mu_k(\mathbf{z})\}^{(\gamma_k-1)}e^{-\{t_{ik}/\mu_k(\mathbf{z})\}^{(\gamma_k-1)}}dt_{ik}.$$

Opening up the brackets,

$$\epsilon_k(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k) = \frac{1}{D}\int_0^D D\frac{\gamma_k}{\mu_k(\mathbf{z})}\{t_{ik}/\mu_k(\mathbf{z})\}^{(\gamma_k-1)}e^{-\{t_{ik}/\mu_k(\mathbf{z})\}^{(\gamma_k-1)}}dt_{ik}$$

$$-\frac{1}{D}\int_0^D t_{ik}\frac{\gamma_k}{\mu_k(\mathbf{z})}\{t_{ik}/\mu_k(\mathbf{z})\}^{(\gamma_k-1)}e^{-\{t_{ik}/\mu_k(\mathbf{z})\}^{(\gamma_k-1)}}dt_{ik}.$$

Cancelling out the contants in the numerator and the denominator,

$$\epsilon_k(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k) = \int_0^D \frac{\gamma_k}{\mu_k(\mathbf{z})}\{t_{ik}/\mu_k(\mathbf{z})\}^{(\gamma_k-1)}e^{-\{t_{ik}/\mu_k(\mathbf{z})\}^{(\gamma_k-1)}}dt_{ik}$$

$$-\frac{1}{D}\int_0^D t_{ik}\frac{\gamma_k}{\mu_k(\mathbf{z})}\{t_{ik}/\mu_k(\mathbf{z})\}^{(\gamma_k-1)}e^{-\{t_{ik}/\mu_k(\mathbf{z})\}^{(\gamma_k-1)}}dt_{ik}.$$

Therefore,

$$\epsilon_k(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k) = \int_0^D f(t_{ik})dt_{ik} - \frac{1}{D}\int_0^D t_{ik}f(t_{ik})dt_{ik}.$$

Therefore,

$$\epsilon_k(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k) = P(t_{ik}\leq D) - \frac{1}{D}E\{t_{ik}I_{(t_{ik}\leq D)}\}.$$

This gives,

$$\epsilon_k(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k) = F(D) - \frac{1}{D}E\{t_{ik}I_{(t_{ik}\leq D)}\},$$

where F($D$) is the distribution function of the Weibull model at $D$ and $I_{(t_{ik}\leq D)}$ is the indicator function such that;

$$I_{\{t_{ik}\leq D\}} = \begin{cases} 1; \text{if } t_{ik}\leq D, \\ 0; \text{if } t_{ik} > D. \end{cases}$$

Now $E\{t_{ik}I_{(t_{ik}\leq D)}\}$ can be approximated by the average of all the observed event time for a particular treatment arm. Let's call this $\bar{x}_{ik}$. Therefore, the estimate of $\epsilon_k(\mathbf{z};\beta_k,\gamma_k)$ can be written as;

$$\hat{\epsilon}_k(\mathbf{z};\boldsymbol{\beta}_k,\gamma_k) = F(D) - \frac{1}{D}\bar{x}_{ik}.$$

# Appendix C

# Deriving the Asymptotic Variance of the Hazard Ratio

From the theory it is known that $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \rightarrow N_p\{\mathbf{0}, J^{-1}(\beta)\}$.

According to the Cox proportional hazard model,

$$\log\{h(t|\mathbf{z})\} = \log\{h(t|\mathbf{z}=0)\} + \boldsymbol{\beta}^T\mathbf{z}.$$

Here, $h(t|\mathbf{z}=0)$ is the baseline hazard function and is treated as a nuisance parameter in the model.

Therefore, to fit the Cox proportional hazard model to a set of survival data, the fitted model can be written as

$$\log\{\hat{h}(t|\mathbf{z})\} = \hat{\boldsymbol{\beta}}^T\mathbf{z}.$$

this is because the baseline hazard function is treated as a nuisance parameter and cannot be estimated in the Cox proportional hazard model.

Therefore,

$$Var[\log\{\hat{h}(t|\mathbf{z})\}] = Var(\hat{\boldsymbol{\beta}}^T\mathbf{z})$$

. Which means,

$$Var[\log\{\hat{h}(t|\mathbf{z})\}] = \mathbf{z}^T J^{-1}(\hat{\boldsymbol{\beta}})\mathbf{z}.$$

According to the multivariate version of the delta method, if $(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \xrightarrow{d} N\{\mathbf{0}, J^{-1}(\boldsymbol{\beta})\}$, then for $f$ being a one-one continuous function of the estimator $\hat{\boldsymbol{\beta}}$,

$$\{f(\hat{\boldsymbol{\beta}}) - f(\boldsymbol{\beta})\} \xrightarrow{d} N_p[\mathbf{0}, \nabla\{f(\boldsymbol{\beta})\}^T J^{-1}(\beta)\nabla f(\boldsymbol{\beta})].$$

Now, $\log\{\hat{h}(t|\mathbf{z})\}$ being the partial likelihood estimator of $\log\{h(t|\mathbf{z})\}$,

$$[\log\{\hat{h}(t|\mathbf{z})\}] \xrightarrow{d} N(\log\{h(t|\mathbf{z})\}, Var[\log\{\hat{h}(t|\mathbf{z})\}]).$$

Therefore,

$$Var(\exp\{\log[\hat{h}(t|\mathbf{z})]\}) = \nabla[\exp(\boldsymbol{\beta}^T\mathbf{z})]^T Var[\log\{\hat{h}(t|\mathbf{z})\}]\nabla[\exp(\boldsymbol{\beta}^T\mathbf{z})].$$

Which means,

$$Var\{\hat{h}(t|\mathbf{z})\} = \mathbf{z}^T Var[\log\{\hat{h}(t|\mathbf{z})\}]\mathbf{z}\exp(2\boldsymbol{\beta}^T\mathbf{z}).$$

For each treatment arm $k$,

$$Var\{\hat{h}_k(t|\mathbf{z})\} = \mathbf{z}^T\{\mathbf{z}^T J_k^{-1}(\boldsymbol{\beta}_k)\mathbf{z}\}\mathbf{z}\exp(2\boldsymbol{\beta}_k^T\mathbf{z}).$$

Now, $J_k(\boldsymbol{\beta}_k) = \sum_{i=1}^{n_k}\delta_{ik}V(t_{(i)}, \boldsymbol{\beta}_k)$. The law of large numbers can be used to estimate $J_k(\boldsymbol{\beta}_k)$ in order to get the variance estimate of the hazard ratio. Therefore,

$$E\{J_k(\boldsymbol{\beta}_k)\} = E\left\{\sum_{i=1}^{n_k}\delta_{ik}V(t_{(i)}, \boldsymbol{\beta}_k)\right\}.$$

This can also be written as

$$E\{J_k(\boldsymbol{\beta}_k)\} = \sum_{i=1}^{n_k} E\{\delta_{ik} V(t_{(i)}, \boldsymbol{\beta}_k)\}.$$

This can also be written as

$$E\{J_k(\boldsymbol{\beta}_k)\} = \sum_{i=1}^{n_k} E\{\delta_{ik}] V(t_{(i)}, \boldsymbol{\beta}_k).$$

Which deduces to

$$E\{J_k(\boldsymbol{\beta}_k)\} = \sum_{i=1}^{n_k} \epsilon_k(\mathbf{z}; \boldsymbol{\beta}_k) V(t_{(i)}, \boldsymbol{\beta}_k).$$

This can be written as

$$E\{J_k(\boldsymbol{\beta}_k)\} = n_k \epsilon_k(\mathbf{z}; \boldsymbol{\beta}_k) \bar{V}(t, \boldsymbol{\beta}_k).$$

Thus,

$$Var\{\hat{h}_k(t|\mathbf{z})\} = \mathbf{z}^T \left\{ \mathbf{z}^T \frac{\bar{V}^{-1}(t, \boldsymbol{\beta}_k)}{\epsilon_k(\mathbf{z}; \boldsymbol{\beta}_k) n_k} \mathbf{z} \right\} \mathbf{z} \exp(2\boldsymbol{\beta}_k^T \mathbf{z}),$$

where $k = A, B$.

# Appendix D

# Deriving the Optimal Allocation Proportion for minimizing the Total Cumulative Hazard Function

The total cumulative hazard at time $D > 0$ can be minimized subject to the asymptotic variance for the covariate adjusted treatment difference remaining fixed to a constant. This yields the optimal allocation proportion of (2.18) as follows :

$$\text{min: } n_A \left\{ \frac{D^{\gamma_A}}{\mu_A^{\gamma_A}(\mathbf{z})} \right\} + n_B \left\{ \frac{D^{\gamma_B}}{\mu_B^{\gamma_B}(\mathbf{z})} \right\}$$

$$\text{subject to : } \left\{ \frac{G_A}{n_A \gamma_A^2} \right\} + \left\{ \frac{G_B}{n_B \gamma_B^2} \right\} = \text{k} > 0$$

Re-arranging the constraint we get

$$\text{k} - \left\{ \frac{G_B}{n_B \gamma_B^2} \right\} = \left\{ \frac{G_A}{n_A \gamma_A^2} \right\},$$

Re-arranging the constraint further we get

$$\frac{\text{k} n_B \gamma_B^2 - G_B}{n_B \gamma_B^2} = \left\{ \frac{G_A}{n_A \gamma_A^2} \right\},$$

Solving for $n_A$ we get;

$$\frac{G_A n_B \gamma_B^2}{\mathrm{k} n_B \gamma_B^2 \gamma_A^2 - G_B \gamma_A^2} = n_A.$$

The ethical objective here is to minimize,

$$n_A \left\{ \frac{D^{\gamma_A}}{\mu_A^{\gamma_A}(\mathbf{z})} \right\} + n_B \left\{ \frac{D^{\gamma_B}}{\mu_B^{\gamma_B}(\mathbf{z})} \right\}.$$

However

$$\frac{G_A n_B \gamma_B^2}{\mathrm{k} n_B \gamma_B^2 \gamma_A^2 - G_B \gamma_A^2} \left\{ \frac{D^{\gamma_A}}{\mu_A^{\gamma_A}(\mathbf{z})} \right\} + n_B \left\{ \frac{D^{\gamma_B}}{\mu_B^{\gamma_B}(\mathbf{z})} \right\}$$

can be minimized in order to achieve the ethical objective.

To achieve the minimum value of the objective function

$$\frac{\partial}{\partial n_B} \left[ \frac{G_A n_B \gamma_B^2}{\mathrm{k} n_B \gamma_B^2 \gamma_A^2 - G_B \gamma_A^2} \left\{ \frac{D^{\gamma_A}}{\mu_A^{\gamma_A}(\mathbf{z})} \right\} + n_B \left\{ \frac{D^{\gamma_B}}{\mu_B^{\gamma_B}(\mathbf{z})} \right\} \right] = 0.$$

Differentiating with respect to $n_B$ we get

$$\left\{ \frac{D^{\gamma_B}}{\mu_B^{\gamma_B}(\mathbf{z})} \right\} + \left\{ \frac{D^{\gamma_A}}{\mu_A^{\gamma_A}(\mathbf{z})} \right\} \frac{-G_A G_B \gamma_A^2 \gamma_B^2}{(\mathrm{k} n_B \gamma_B^2 \gamma_A^2 - G_B \gamma_A^2)^2} = 0.$$

This can be further re-arranged to obtain

$$\left\{ \frac{D^{\gamma_A}}{\mu_A^{\gamma_A}(\mathbf{z})} \right\} \frac{G_A G_B \gamma_A^2 \gamma_B^2}{(\mathrm{k} n_B \gamma_B^2 \gamma_A^2 - G_B \gamma_A^2)^2} = \left\{ \frac{D^{\gamma_B}}{\mu_B^{\gamma_B}(\mathbf{z})} \right\}.$$

This can be further re-arranged to obtain

$$\left\{ \frac{D^{\gamma_A}}{\mu_A^{\gamma_A}(\mathbf{z})} \right\} G_A G_B \gamma_A^2 \gamma_B^2 = (\mathrm{k} n_B \gamma_B^2 \gamma_A^2 - G_B \gamma_A^2)^2 \left\{ \frac{D^{\gamma_B}}{\mu_B^{\gamma_B}(\mathbf{z})} \right\}.$$

Taking the positive square-root on both sides we get

$$\sqrt{\left\{\frac{D^{\gamma_A}}{\mu_A^{\gamma_A}(\mathbf{z})}\right\}G_A G_B \gamma_A^2 \gamma_B^2} = (\mathrm{k} n_B \gamma_B^2 \gamma_A^2 - G_B \gamma_A^2)\sqrt{\left\{\frac{D^{\gamma_B}}{\mu_B^{\gamma_B}(\mathbf{z})}\right\}}.$$

Substituting for k we get

$$\sqrt{\left\{\frac{D^{\gamma_A}}{\mu_A^{\gamma_A}(\mathbf{z})}\right\}G_A G_B \gamma_A^2 \gamma_B^2} = \left\{\frac{G_A n_B \gamma_B^2}{n_A} - \frac{G_B n_B \gamma_A^2}{n_B} - G_B \gamma_A^2\right\}\sqrt{\left\{\frac{D^{\gamma_B}}{\mu_B^{\gamma_B}(\mathbf{z})}\right\}}.$$

Simplifying we get

$$\sqrt{\left\{\frac{D^{\gamma_A}}{\mu_A^{\gamma_A}(\mathbf{z})}\right\}G_A G_B \gamma_A^2 \gamma_B^2} = \frac{n_B}{n_A} G_A \gamma_B^2 \sqrt{\left\{\frac{D^{\gamma_B}}{\mu_B^{\gamma_B}(\mathbf{z})}\right\}}.$$

Solving for $\frac{n_B}{n_A}$ we get

$$\frac{n_B}{n_A} = \frac{\sqrt{\left\{\frac{D^{\gamma_A}}{\mu_A^{\gamma_A}(\mathbf{z})}\right\}G_A G_B \gamma_A^2 \gamma_B^2}}{\sqrt{\left\{\frac{D^{\gamma_B}}{\mu_B^{\gamma_B}(\mathbf{z})}\right\}G_A \gamma_B^2}}.$$

Now,

$$\pi_{A_0}^W(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B, \gamma_k, \mathbf{z}) = \frac{n_A}{n_A + n_B}.$$

Dividing the numerator and the denominator in the right hand side of the equation by $n_B$ we get

$$\pi_{A_0}^W(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B, \gamma_k, \mathbf{z}) = \frac{n_A/n_B}{1 + n_A/n_B}.$$

Substituting for $\frac{n_A}{n_B}$ we get

$$\pi^W_{A_0}(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B, \gamma_k, \mathbf{z}) = \frac{\sqrt{\left\{\frac{D^{\gamma_B}}{\mu_B^{\gamma_B}(\mathbf{z})}\right\}\frac{G_A}{\gamma_A^2}}}{\sqrt{\left\{\frac{D^{\gamma_B}}{\mu_B^{\gamma_B}(\mathbf{z})}\right\}\frac{G_A}{\gamma_A^2}} + \sqrt{\left\{\frac{D^{\gamma_A}}{\mu_A^{\gamma_A}(\mathbf{z})}\right\}\frac{G_B}{\gamma_B^2}}}.$$

Re-arranging this we obtain

$$\pi^W_{A_0}(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B, \gamma_k, \mathbf{z}) = \frac{\gamma_B\sqrt{\{1/(\mu_B(\mathbf{z})\}^{\gamma_B}D^{\gamma_B}G_A}}{\gamma_B\sqrt{\{1/(\mu_B(\mathbf{z})\}^{\gamma_B}D^{\gamma_B}G_A} + \gamma_A\sqrt{\{1/(\mu_A(\mathbf{z})\}^{\gamma_A}D^{\gamma_A}G_B}}.$$

# Appendix E

# Deriving the Optimal Allocation Proportion for minimizing the Overall trial size for Semi-parametric Survival Models

The overall trial size can be minimized subject to the asymptotic variance for the difference in the covariate adjusted hazard ratios for the two treatment arms with the baseline hazard to remain fixed to a constant. The variance of the covariate adjusted hazard ratio for a treatment arm $k$ with respect to the baseline hazard is;

$$Var\{\hat{h}_k(t|\mathbf{z})\} = \mathbf{z}^T \left\{ \mathbf{z}^T \frac{\bar{V}^{-1}(t, \boldsymbol{\beta}_k)}{\epsilon_k(\mathbf{z}; \boldsymbol{\beta}_k) n_k} \mathbf{z} \right\} \mathbf{z} \exp(2\boldsymbol{\beta}_k^T \mathbf{z}),$$

where $k = A, B$.

Therefore the objective is to

$$\text{min: } n_A + n_B$$

$$\text{subject to : } Var\{\hat{h}_A(t|\mathbf{z})\} + Var\{\hat{h}_B(t|\mathbf{z})\} = \text{k} > 0.$$

Re-arranging the constraint we get,

$$\mathrm{k} - \mathbf{z}^T \left\{ \mathbf{z}^T \frac{\bar{V}^{-1}(t, \boldsymbol{\beta}_B)}{\epsilon_B(\mathbf{z}; \boldsymbol{\beta}_B) n_B} \mathbf{z} \right\} \mathbf{z} \exp(2\boldsymbol{\beta}_B^T \mathbf{z}) =$$

$$\mathbf{z}^T \left\{ \mathbf{z}^T \frac{\bar{V}^{-1}(t, \boldsymbol{\beta}_A)}{\epsilon_A(\mathbf{z}; \boldsymbol{\beta}_A) n_A} \mathbf{z} \right\} \mathbf{z} \exp(2\boldsymbol{\beta}_A^T \mathbf{z}).$$

This can also be written as

$$\frac{\mathrm{k}\epsilon_B(\mathbf{z}; \boldsymbol{\beta}_B) n_B - \mathbf{z}^T \{\mathbf{z}^T \bar{V}^{-1}(t, \boldsymbol{\beta}_B)\mathbf{z}\} \mathbf{z} \exp(2\boldsymbol{\beta}_B^T \mathbf{z})}{\epsilon_B(\mathbf{z}; \boldsymbol{\beta}_B) n_B} =$$

$$\mathbf{z}^T \left\{ \mathbf{z}^T \frac{\bar{V}^{-1}(t, \boldsymbol{\beta}_A)}{\epsilon_A(\mathbf{z}; \boldsymbol{\beta}_A) n_A} \mathbf{z} \right\} \mathbf{z} \exp(2\boldsymbol{\beta}_A^T \mathbf{z}).$$

Solving for $n_A$ we get

$$n_A = \frac{\mathbf{z}^T \{\mathbf{z}^T \bar{V}^{-1}(t, \boldsymbol{\beta}_A)\mathbf{z}\} \mathbf{z} \exp(2\boldsymbol{\beta}_A^T \mathbf{z}) \epsilon_B(\mathbf{z}; \boldsymbol{\beta}_B) n_B}{\mathrm{k}\epsilon_A(\mathbf{z}; \boldsymbol{\beta}_A)\epsilon_B(\boldsymbol{\beta}_B, \mathbf{z}) n_B - \epsilon_A(\boldsymbol{\beta}_A, \mathbf{z}) \mathbf{z}^T \{\mathbf{z}^T \bar{V}^{-1}(t, \boldsymbol{\beta}_B)\mathbf{z}\} \mathbf{z} \exp(2\boldsymbol{\beta}_B^T \mathbf{z})}.$$

The objective here is to minimize $n_A + n_B$. Therefore,

$$\min : \frac{\mathbf{z}^T \{\mathbf{z}^T \bar{V}^{-1}(t, \boldsymbol{\beta}_A)\mathbf{z}\} \mathbf{z} \exp(2\boldsymbol{\beta}_A^T \mathbf{z}) \epsilon_B(\mathbf{z}; \boldsymbol{\beta}_B) n_B}{\mathrm{k}\epsilon_A(\mathbf{z}; \boldsymbol{\beta}_A)\epsilon_B(\mathbf{z}; \boldsymbol{\beta}_B) n_B - \epsilon_A(\mathbf{z}; \boldsymbol{\beta}_A) \mathbf{z}^T \{\mathbf{z}^T \bar{V}^{-1}(t, \boldsymbol{\beta}_B)\mathbf{z}\} \mathbf{z} \exp(2\boldsymbol{\beta}_B^T \mathbf{z})} + n_B$$

is the objective.

Let $Va_k\{\hat{h}_k(t|\mathbf{z})\} = \mathbf{z}^T \{\mathbf{z}^T \bar{V}^{-1}(t, \boldsymbol{\beta}_k)\mathbf{z}\} \mathbf{z} \exp(2\boldsymbol{\beta}_k^T \mathbf{z})$ denote the variability adjustment factor for treatment $k$.

To achieve the minimum value of the objective function,

$$\frac{\partial}{\partial n_B} \left[ \frac{Va_A\{\hat{h}_A(t|\mathbf{z})\} \epsilon_B(\mathbf{z}; \boldsymbol{\beta}_B) n_B}{\mathrm{k}\epsilon_A(\mathbf{z}; \boldsymbol{\beta}_A)\epsilon_B(\mathbf{z}; \boldsymbol{\beta}_B) n_B - \epsilon_A(\mathbf{z}; \boldsymbol{\beta}_A) Va_B\{\hat{h}_B(t|\mathbf{z}_m)\}} + n_B \right] = 0.$$

Differentiating both sides with respect to $n_B$ and re-arranging we get,

$$\frac{\{Va_A\{\hat{h}_A(t|\mathbf{z})\}\epsilon_B(\mathbf{z};\boldsymbol{\beta}_B)\}\{\mathrm{k}\epsilon_A(\mathbf{z};\boldsymbol{\beta}_A)\epsilon_B(\mathbf{z};\boldsymbol{\beta}_B)n_B - \epsilon_A(\mathbf{z};\boldsymbol{\beta}_A)Va_B\{\hat{h}_B(t|\mathbf{z})\}\}}{\{\mathrm{k}\epsilon_A(\mathbf{z};\boldsymbol{\beta}_A)\epsilon_B(\mathbf{z};\boldsymbol{\beta}_B)n_B - \epsilon_A(\mathbf{z};\boldsymbol{\beta}_A)Va_B\{\hat{h}_B(t|\mathbf{z})\}\}^2} -$$
$$\frac{\{Va_A\{\hat{h}_A(t|\mathbf{z})\}\epsilon_B(\mathbf{z};\boldsymbol{\beta}_B)n_B\}\{\mathrm{k}\epsilon_A(\mathbf{z};\boldsymbol{\beta}_A)\epsilon_B(\mathbf{z};\boldsymbol{\beta}_B)\}}{\{\mathrm{k}\epsilon_A(\mathbf{z};\boldsymbol{\beta}_A)\epsilon_B(\mathbf{z};\boldsymbol{\beta}_B)n_B - \epsilon_A(\mathbf{z};\boldsymbol{\beta}_A)Va_B\{\hat{h}_B(t|\mathbf{z})\}\}^2} = -1.$$

Simplifying the expression we get,

$$\epsilon_A(\mathbf{z};\boldsymbol{\beta}_A)\epsilon_B(\mathbf{z};\boldsymbol{\beta}_B)Va_A\{\hat{h}_A(t|\mathbf{z})\}Va_B\{\hat{h}_B(t|\mathbf{z})\} =$$

$$\{\mathrm{k}\epsilon_A(\mathbf{z};\boldsymbol{\beta}_A)\epsilon_B(\mathbf{z};\boldsymbol{\beta}_B)n_B - \epsilon_A(\mathbf{z};\boldsymbol{\beta}_A)Va_B\{\hat{h}_B(t|\mathbf{z})\}\}^2.$$

Taking the positive square-root on both sides we get,

$$\sqrt{\epsilon_A(\mathbf{z};\boldsymbol{\beta}_A)\epsilon_B(\mathbf{z};\boldsymbol{\beta}_B)Va_A\{\hat{h}_A(t|\mathbf{z})\}Va_B\{\hat{h}_B(t|\mathbf{z})\}} =$$

$$\mathrm{k}\epsilon_A(\mathbf{z};\boldsymbol{\beta}_A)\epsilon_B(\mathbf{z};\boldsymbol{\beta}_B)n_B - \epsilon_A(\mathbf{z};\boldsymbol{\beta}_A)Va_B\{\hat{h}_B(t|\mathbf{z})\}.$$

Substituting for k we get,

$$\sqrt{\epsilon_A(\mathbf{z};\boldsymbol{\beta}_A)\epsilon_B(\mathbf{z};\boldsymbol{\beta}_B)Va_A\{\hat{h}_A(t|\mathbf{z})\}Va_B\{\hat{h}_B(t|\mathbf{z})\}} =$$

$$Va_B\{\hat{h}_B(t|\mathbf{z})\}\epsilon_A(\mathbf{z};\boldsymbol{\beta}_A) + Va_A\{\hat{h}_A(t|\mathbf{z})\}\epsilon_B(\mathbf{z};\boldsymbol{\beta}_B)\frac{n_B}{n_A} - \epsilon_A(\mathbf{z};\boldsymbol{\beta}_A)Va_B\{\hat{h}_B(t|\mathbf{z})\}.$$

Solving for $\frac{n_B}{n_A}$ we get,

$$\frac{n_B}{n_A} = \frac{\sqrt{\epsilon_A(\mathbf{z};\boldsymbol{\beta}_A)\epsilon_B(\mathbf{z};\boldsymbol{\beta}_B)Va_A\{\hat{h}_A(t|\mathbf{z})\}Va_B\{\hat{h}_B(t|\mathbf{z})\}}}{Va_A\{\hat{h}_A(t|\mathbf{z})\}\epsilon_B(\mathbf{z};\boldsymbol{\beta}_B)}.$$

Which can also be written as

$$\frac{n_B}{n_A} = \frac{\sqrt{\epsilon_A(\mathbf{z};\boldsymbol{\beta}_A)Va_B\{\hat{h}_B(t|\mathbf{z})\}}}{\sqrt{Va_A\{\hat{h}_A(t|\mathbf{z})\}\epsilon_B(\mathbf{z};\boldsymbol{\beta}_B)}}.$$

Now,

$$\pi^S_{A_2}(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B, \mathbf{z}) = \frac{n_A}{n_A + n_B}.$$

Dividing the numerator and the denominator in the right hand side of the equation by $n_B$ we get,

$$\pi^S_{A_2}(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B, \mathbf{z}) = \frac{n_A/n_B}{1 + n_A/n_B}.$$

Substituting for $\frac{n_A}{n_B}$ we get,

$$\pi^S_{A_2}(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B, \mathbf{z}) = \frac{\sqrt{Va_A\{\hat{h}_A(t|\mathbf{z})\}\epsilon_B(\mathbf{z};\boldsymbol{\beta}_B)}/\sqrt{\epsilon_A(\mathbf{z};\boldsymbol{\beta}_A)Va_B\{\hat{h}_B(t|\mathbf{z})\}}}{1 + \sqrt{Va_A\{\hat{h}_A(t|\mathbf{z})\}\epsilon_B(\mathbf{z};\boldsymbol{\beta}_A)}/\sqrt{\epsilon_A(\mathbf{z};\boldsymbol{\beta}_A)Va_B\{\hat{h}_B(t|\mathbf{z})\}}}.$$

This can be written as

$$\pi^S_{A_2}(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B, \mathbf{z}) = \frac{\sqrt{Va_A\{\hat{h}_A(t|\mathbf{z})\}\epsilon_B(\mathbf{z};\boldsymbol{\beta}_B)}}{\sqrt{\epsilon_A(\mathbf{z};\boldsymbol{\beta}_A)Va_B\{\hat{h}_B(t|\mathbf{z})\}} + \sqrt{Va_A\{\hat{h}_A(t|\mathbf{z})\}\epsilon_B(\mathbf{z};\boldsymbol{\beta}_B)}}.$$

This can be further written as

$$\pi^S_{A_2}(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B, \mathbf{z}) =$$

$$\frac{\sqrt{\mathbf{z}^T\{\mathbf{z}^T\bar{V}^{-1}(t,\boldsymbol{\beta}_A)\mathbf{z}\}\mathbf{z}e^{2\boldsymbol{\beta}_A^T\mathbf{z}}\epsilon_B(\mathbf{z};\boldsymbol{\beta}_B)}}{\sqrt{\epsilon_A(\mathbf{z};\boldsymbol{\beta}_A)\mathbf{z}^T\{\mathbf{z}\bar{V}^{-1}(t,\beta_B)\mathbf{z}\}\mathbf{z}e^{2\boldsymbol{\beta}^T\mathbf{z}}} + \sqrt{\mathbf{z}^T\{\mathbf{z}^T\bar{V}^{-1}(t,\boldsymbol{\beta}_A)\mathbf{z}\}\mathbf{z}e^{2\boldsymbol{\beta}_A^T\mathbf{z}}\epsilon_B(\mathbf{z};\boldsymbol{\beta}_B)}}.$$

# Appendix F

# Deriving the Asymptotic Variance of the Cause Specific Cumulative Incidence Function

From (4.7) the relationship between the subdistribution hazard $h_{jk}^{sub}(t|\mathbf{z})$ for cause $j$ at treatment $k$ and the corresponding cumulative incidence function $F_{jk}(t|\mathbf{z})$) is given by

$$h_{jk}^{sub}(t|\mathbf{z}) = -\frac{d}{dt}\log\{1 - F_{jk}(t|\mathbf{z})\}, \quad \text{for} \quad j = 1, ...., c.$$

Therefore,

$$\log\{1 - F_{jk}(t|\mathbf{z})\} = -\int_0^t \{h_{jk}^{sub}(u|\mathbf{z})\}du, \quad \text{for} \quad j = 1, ...., c.$$

This can also be writted as

$$F_{jk}(t|\mathbf{z}) = (1 - \exp[-\int_0^t \{h_{jk}^{sub}(u|\mathbf{z})\}du]), \quad \text{for} \quad j = 1, ...., c.$$

Now, $\hat{h}_{jk}^{sub}(t|\mathbf{z}_i)$ being the partial likelihood estimator of $h_{jk}^{sub}(t|\mathbf{z}_i)$,

$$\hat{h}_{jk}^{sub}(t|\mathbf{z}) \xrightarrow{d} N_p[h_{jk}^{sub}(t|\mathbf{z}), Var\{\hat{h}_{jk}^{sub}(t|\mathbf{z})\}].$$

Also,

$$Var\{\hat{h}_{jk}^{sub}(t|\mathbf{z})\} = \mathbf{z}^T \left\{ \mathbf{z}^T \frac{\bar{V}^{-1}(t, \boldsymbol{\beta}_{jk})}{\epsilon_k(\mathbf{z}; \boldsymbol{\beta}_{jk}) n_k} \mathbf{z} \right\} \mathbf{z} \exp(2\boldsymbol{\beta}_{jk}^T \mathbf{z}),$$

where $k = A, B$.

Therefore,

$$\hat{F}_{jk}(t|\mathbf{z}) \xrightarrow{d} N_p[F_{jk}(t|\mathbf{z}), Var\{\hat{F}_{jk}(t|\mathbf{z})\}],$$

where

$$Var\{\hat{F}_{jk}(t|\mathbf{z})\} = \nabla(1-\exp[-\int_0^t \{h_{jk}^{sub}(u|\mathbf{z})\}du])^T Var\{\hat{h}_{jk}^{sub}(t|\mathbf{z})\} \nabla(1-\exp[-\int_0^t \{h_{jk}^{sub}(u|\mathbf{z})\}du]).$$

This can be written as

$$Var\{\hat{F}_{jk}(t|\mathbf{z})\} = \left\{ \int_0^t \nabla h_{jk}^{sub}(u|\mathbf{z})du \right\}^T Var\{\hat{h}_{jk}^{sub}(t|\mathbf{z})\} \left\{ \int_0^t \nabla h_{jk}^{sub}(u|\mathbf{z})du \right\}$$
$$\left( \exp\left[ -2 \int_0^t \{h_{jk}^{sub}(u|\mathbf{z})\}du \right] \right),$$

which mean

$$Var\{\hat{F}_{jk}(t|\mathbf{z})\} = (t\mathbf{z}^T) Var\{\hat{h}_{jk}^{sub}(t|\mathbf{z})\} (t\mathbf{z}) (\exp[-2 \int_0^t \{h_{jk}^{sub}(u|\mathbf{z})\}du]),$$

for $k = A, B$.

# Appendix G

# Deriving the Optimal Allocation Proportion for minimizing the Overall Cumulative Incidence Function for a given Cause

The overall cumulative incidence function for cause $j$ can be minimized subject to the asymptotic variance for the difference in the covariate adjusted cumulative incidence function for the two treatment arms to remain fixed to a constant. This means,

$$min : n_{Aj}F_{jA}(t|\mathbf{z}) + n_{Bj}F_{jB}(t|\mathbf{z}),$$

$$subject\ to : Var\{\hat{F}_{jA}(t|\mathbf{z})\} + Var\{\hat{F}_{jB}(t|\mathbf{z})\} = \mathrm{k} > 0$$

Re-arranging the constraint we get,

$$\mathrm{k} - \frac{(t\mathbf{z}^T)Va_{jB}\{\hat{h}_{jB}(t|\mathbf{z})\}(t\mathbf{z})(\exp[-2\int_0^t\{h_{jB}^{sub}(u|\mathbf{z})\}du])}{\epsilon_{jB}(\mathbf{z};\boldsymbol{\beta}_{jB})n_{Bj}} =$$

$$\frac{(t\mathbf{z}^T)Va_{jA}\{\hat{h}_{jA}(t|\mathbf{z})\}(t\mathbf{z})(\exp[-2\int_0^t\{h_{jA}^{sub}(u|\mathbf{z})\}du])}{\epsilon_{jA}(\mathbf{z};\boldsymbol{\beta}_{jA})n_{Aj}}.$$

This can also be written as

$$\frac{\mathrm{k}\epsilon_{jB}(\mathbf{z};\boldsymbol{\beta}_{jB})n_{Bj} - (t\mathbf{z}^T)Va_{jB}\{\hat{h}_{jB}(t|\mathbf{z})\}(t\mathbf{z})(\exp[-2\int_0^t\{h_{jB}^{sub}(u|\mathbf{z})\}du])}{\boldsymbol{\epsilon}_{jB}(\mathbf{z};\boldsymbol{\beta}_{jB})n_{Bj}} =$$

$$\frac{(t\mathbf{z}^T)Va_{jA}\{\hat{h}_{jA}(t|\mathbf{z})\}(t\mathbf{z})(\exp[-2\int_0^t\{h_{jA}^{sub}(u|\mathbf{z})\}du])}{\epsilon_{jA}(\mathbf{z};\boldsymbol{\beta}_{jA})n_{Aj}}.$$

Solving for $n_{Aj}$ we get

$$n_{Aj} = \frac{(t\mathbf{z}^T)Va_{jA}\{\hat{h}_{jA}(t|\mathbf{z})\}(t\mathbf{z})(\exp[-2\int_0^t\{h_{jA}^{sub}(u|\mathbf{z})\}du])\epsilon_{jB}(\mathbf{z};\boldsymbol{\beta}_{jB})n_{Bj}}{\mathrm{k}\epsilon_{jA}(\mathbf{z};\boldsymbol{\beta}_{jA})\epsilon_{jB}(\mathbf{z};\boldsymbol{\beta}_{jB})n_{Bj} - \epsilon_{jA}(\mathbf{z};\boldsymbol{\beta}_{jA})(t\mathbf{z}^T)Va_{jB}\{\hat{h}_{jB}(t|\mathbf{z})\}(t\mathbf{z})(e^{[-2\int_0^t\{h_{jB}^{sub}(u|\mathbf{z})\}du]})}.$$

The objective here is to minimize $n_{Aj}F_{jA}(t|\mathbf{z}) + n_{Bj}F_{jB}(t|\mathbf{z})$. Let

$$Vr_{jA}(\boldsymbol{\beta}_{jA},\mathbf{z}) = (t\mathbf{z}^T)Va_{jA}\{\hat{h}_{jA}(t|\mathbf{z})\}(t\mathbf{z})(\exp[-2\int_0^t\{h_{jA}^{sub}(u|\mathbf{z})\}du])$$

and

$$Vr_{jB}(\boldsymbol{\beta}_{jB},\mathbf{z}) = (t\mathbf{z}^T)Va_{jB}\{\hat{h}_{jB}(t|\mathbf{z})\}(t\mathbf{z})(e^{[-2\int_0^t\{h_{jB}^{sub}(u|\mathbf{z})\}du]}).$$

Therefore,

$$\min : \frac{Vr_{jA}(\boldsymbol{\beta}_{jA},\mathbf{z})\epsilon_{jB}(\mathbf{z};\boldsymbol{\beta}_{jB})n_{Bj}}{\mathrm{k}\epsilon_{jA}(\mathbf{z};\boldsymbol{\beta}_{jA})\epsilon_{jB}(\mathbf{z};\boldsymbol{\beta}_{jB})n_{Bj} - \epsilon_{jA}(\mathbf{z};\boldsymbol{\beta}_{jA})Vr_{jB}(\boldsymbol{\beta}_{jB},\mathbf{z})}F_{jA}(t|\mathbf{z}) + n_{Bj}F_{jB}(t|\mathbf{z})$$

is the objective.

To achieve the minimum value of the objective function,

$$\frac{\partial}{\partial n_{Bj}}\left[\frac{Vr_{jA}(\boldsymbol{\beta}_{jA},\mathbf{z}_i)\epsilon_{jB}(\mathbf{z};\boldsymbol{\beta}_{jB})n_{Bj}}{\mathrm{k}\epsilon_{jA}(\mathbf{z};\boldsymbol{\beta}_{jA})\epsilon_{jB}(\mathbf{z};\boldsymbol{\beta}_{jB})n_{Bj} - \epsilon_{jA}(\mathbf{z};\boldsymbol{\beta}_{jA})Vr_{jB}(\boldsymbol{\beta}_{jB},\mathbf{z}_i)}F_{jA}(t|\mathbf{z}_i) + n_{Bj}F_{jB}(t|\mathbf{z}_i)\right] = 0.$$

Differentiating both sides with respect to $n_{Bj}$ and re-arranging we get,

$$\frac{\{Vr_{jA}(\boldsymbol{\beta}_{jA},\mathbf{z})\epsilon_{jB}(\mathbf{z};\boldsymbol{\beta}_{jB})F_{jA}(t|\mathbf{z})\}\{\mathrm{k}\epsilon_{jA}(\mathbf{z};\boldsymbol{\beta}_{jA})\epsilon_{jB}(\mathbf{z};\boldsymbol{\beta}_{jB})n_{Bj} - \epsilon_{jA}(\mathbf{z};\boldsymbol{\beta}_{jA})Vr_{jB}(\boldsymbol{beta}_{jB},\mathbf{z})\}}{\{\mathrm{k}\epsilon_{jA}(\mathbf{z};\boldsymbol{\beta}_{jA})\epsilon_{jB}(\mathbf{z};\boldsymbol{\beta}_{jB})n_{Bj} - \epsilon_{jA}(\mathbf{z};\boldsymbol{\beta}_{jA})Vr_{jB}(\boldsymbol{\beta}_{jB},\mathbf{z})\}^2} -$$

$$\frac{\{Vr_{jA}(\boldsymbol{\beta}_{jA},\mathbf{z})\epsilon_{jB}(\mathbf{z};\boldsymbol{\beta}_{jB})F_{jA}(t|\mathbf{z})n_{Bj}\}\{\mathrm{k}\epsilon_{jA}(\mathbf{z};\boldsymbol{\beta}_{jA})\epsilon_{jB}(\mathbf{z};\boldsymbol{\beta}_{jB})\}}{\{\mathrm{k}\epsilon_{jA}(\mathbf{z};\boldsymbol{\beta}_{jA})\epsilon_{jB}(\mathbf{z};\boldsymbol{\beta}_{jB})n_{Bj} - \epsilon_{jA}(\mathbf{z};\boldsymbol{\beta}_{jA})Vr_{jB}(\boldsymbol{\beta}_{jB},\mathbf{z})\}^2} = -F_{jB}(t|\mathbf{z}).$$

Simplifying the expression we get,

$$\epsilon_{jA}(\mathbf{z};\boldsymbol{\beta}_{jA})\epsilon_{jB}(\mathbf{z};\boldsymbol{\beta}_{jB})F_{jA}(t|\mathbf{z})Vr_{jA}(\boldsymbol{\beta}_{jA},\mathbf{z})Vr_{jB}(\boldsymbol{\beta}_{jB},\mathbf{z}) =$$

$$F_{jB}(t|\mathbf{z})\{\mathrm{k}\epsilon_{jA}(\mathbf{z};\boldsymbol{\beta}_{jA})\epsilon_{jB}(\mathbf{z};\boldsymbol{\beta}_{jB})n_{Bj} - \epsilon_{jA}(\mathbf{z};\boldsymbol{\beta}_{jA})Vr_{jB}(\mathbf{z};\boldsymbol{\beta}_{jB})\}^2.$$

Taking the positive square-root on both sides we get,

$$\sqrt{\epsilon_{jA}(\mathbf{z};\boldsymbol{\beta}_{jA})\epsilon_{jB}(\mathbf{z};\boldsymbol{\beta}_{jB})F_{jA}(t|\mathbf{z})Vr_{jA}(\boldsymbol{\beta}_{jA},\mathbf{z})Vr_{jB}(\boldsymbol{\beta}_{jB},\mathbf{z})} =$$

$$\sqrt{F_{jB}(t|\mathbf{z})}\{\mathrm{k}\epsilon_{jA}(\mathbf{z};\boldsymbol{\beta}_{jA})\epsilon_{jB}(\mathbf{z};\boldsymbol{\beta}_{jB})n_{Bj} - \epsilon_{jA}(\mathbf{z};\boldsymbol{\beta}_{jA})Vr_{jB}(\boldsymbol{\beta}_{jB},\mathbf{z})\}.$$

Substituting for k we get,

$$\sqrt{\epsilon_{jA}(\mathbf{z};\boldsymbol{\beta}_{jA})\epsilon_{jB}(\mathbf{z};\boldsymbol{\beta}_{jB})F_{jA}(t|\mathbf{z})Vr_{jA}(\boldsymbol{\beta}_{jA},\mathbf{z})Vr_{jB}(\boldsymbol{\beta}_{jB},\mathbf{z})} =$$

$$\left[\sqrt{F_{jB}(t|\mathbf{z})}Vr_{jB}(\boldsymbol{\beta}_{jB},\mathbf{z})\epsilon_{jA}(\mathbf{z};\boldsymbol{\beta}_{jA}) + \right.$$

$$\left.\sqrt{F_{jB}(t|\mathbf{z})}Vr_{jA}(\boldsymbol{\beta}_{jA},\mathbf{z})\epsilon_{jB}(\mathbf{z};\boldsymbol{\beta}_{jB})\frac{n_{Bj}}{n_{Aj}} - \sqrt{F_{jB}(t|\mathbf{z})}\epsilon_{jA}(\mathbf{z};\boldsymbol{\beta}_{jA})Vr_{jB}(\boldsymbol{\beta}_{jB},\mathbf{z})\right].$$

Solving for $\frac{n_{Bj}}{n_{Aj}}$ we get,

$$\frac{n_{Bj}}{n_{Aj}} = \frac{\sqrt{\epsilon_{jA}(\mathbf{z};\boldsymbol{\beta}_{jA})\epsilon_{jB}(\mathbf{z};\boldsymbol{\beta}_{jB})F_{jA}(t|\mathbf{z})Vr_{jA}(\boldsymbol{\beta}_{jA},\mathbf{z})Vr_{jB}(\boldsymbol{\beta}_{jB},\mathbf{z})}}{\sqrt{F_{jB}(t|\mathbf{z})}Vr_{jA}(\boldsymbol{\beta}_{jA},\mathbf{z})\epsilon_{jB}(\mathbf{z};\boldsymbol{\beta}_{jB})}.$$

Which can also be written as

$$\frac{n_{Bj}}{n_{Aj}} = \frac{\sqrt{\epsilon_{jA}(\mathbf{z};\boldsymbol{\beta}_{jA})F_{jA}(t|\mathbf{z})Vr_{jB}(\boldsymbol{\beta}_{jB},\mathbf{z})}}{\sqrt{F_{jB}(t|\mathbf{z})Vr_{jA}(\boldsymbol{\beta}_{jA},\mathbf{z})\epsilon_{jB}(\mathbf{z};\boldsymbol{\beta}_{jB})}}..$$

Now,

$$\pi^G_{A_{3j}}(\boldsymbol{\beta}_{Aj},\boldsymbol{\beta}_{Bj},\mathbf{z}) = \frac{n_{Aj}}{n_{Aj}+n_{Bj}}.$$

Dividing the numerator and the denominator in the right hand side of the equation by $n_B$ we get,

$$\pi^G_{A_{3j}}(\boldsymbol{\beta}_{Aj}, \boldsymbol{\beta}_{Bj}, \mathbf{z}) = \frac{n_{Aj}/n_{Bj}}{1 + n_{Aj}/n_{Bj}}.$$

Substituting for $\frac{n_{Aj}}{n_{Bj}}$ we get,

$$\pi^G_{A_{3j}}(\boldsymbol{\beta}_{Aj}, \boldsymbol{\beta}_{Bj}, \mathbf{z}) = \frac{\sqrt{F_{jB}(t|\mathbf{z})Vr_{jA}(\boldsymbol{\beta}_{jA}, \mathbf{z})\epsilon_{jB}(\mathbf{z}; \boldsymbol{\beta}_{jB})}/\sqrt{\epsilon_{jA}(\mathbf{z}; \boldsymbol{\beta}_{jA})F_{jA}(t|\mathbf{z})Vr_{jB}(\boldsymbol{\beta}_{jB}, \mathbf{z})}}{1 + \sqrt{F_{jB}(t|\mathbf{z})Vr_{jA}(\boldsymbol{\beta}_{jA}, \mathbf{z})\epsilon_{jB}(\mathbf{z}; \boldsymbol{\beta}_{jB})}/\sqrt{\epsilon_{jA}(\mathbf{z}; \boldsymbol{\beta}_{jA})F_{jA}(t|\mathbf{z})Vr_{jB}(\mathbf{z}; \boldsymbol{\beta}_{jB})}}.$$

This can be written as

$$\pi^G_{A_{3j}}(\boldsymbol{\beta}_{Aj}, \boldsymbol{\beta}_{Bj}, \mathbf{z}) = \frac{\sqrt{F_{jB}(t|\mathbf{z})Vr_{jA}(\boldsymbol{\beta}_{jA}, \mathbf{z})\epsilon_{jB}(\mathbf{z}; \boldsymbol{\beta}_{jB})}}{\sqrt{F_{jB}(t|\mathbf{z})Vr_{jA}(\boldsymbol{\beta}_{jA}, \mathbf{z})\epsilon_{jB}(\boldsymbol{\beta}_{jB}, \mathbf{z})} + \sqrt{\epsilon_{jA}(\mathbf{z}; \boldsymbol{\beta}_{jA})F_{jA}(t|\mathbf{z})Vr_{jB}(\mathbf{z}; \boldsymbol{\beta}_{jB})}}.$$

# Bibliography

Anderson, P.K., and Gill, R.D. (1982), "Cox's Regression Model For Counting Process: A Large Sample Study.", *The Annals of Statistics*, **10**, 1100–1120.

Atkinson, A.C. (1982), "Optimum Biased Coin Designs for Sequential Clinical Trials With Prognostic Factors.", *Biometrika*, **69**, 61–67.

Atkinson, A.C., and Biswas, A. (2005), "Bayesian Adaptive Biased- Coin Designs for Clinical Trials With Normal Responses.", *Biometrics*, **61**, 118–125.

Bandyopadhyay, U., and Biswas, A.(2001), "Adaptive designs for normal responses with prognostic factors.", *Biometrika* **88**, 409–419.

Bandyopadhyay, U., Biswas, A., and Bhattacharya, R. (2010), "A Covariate-Adjusted Adaptive Design for Two-Stage Clinical Trials With Survival Data.", *Statistica Neerlandica*, **64**, 202–226.

Biswas, A., Bhattacharya, R., and Park, E. (2016), "On a class of optimal covariate-adjusted response adaptive designs for survival outcomes.", *Statistical Methods in Medical Research*, **25(6)**, 2444–2456.

Burman, C-F. (1996). "On sequential treatment allocation in clinical trials.", PhD Thesis. Chalmers University of Technology, Göteborg.

Bolland, K., Sooriyarachchi, M.R., and Whitehead, J. (1998), "Sample Size Review in a Head Injury 1270 Trial with Ordered Categorical Responses.", *Statistics in Medicine*, **17(24)**, 2835–2847.

Cox, D. R. (1972). "Regression models and life-tables (with discussion).", *Journal of the Royal Statistical Society*, **B 34**, 187–220.

Cox, D. R. (1975). "Partial likelihood.", *Biometrika*, **62**, 269–76.

Crowley, J. (1974). "A note on some recent likelihoods leading to the log rank test.", *Biometrika*, **61**, 533–538.

Dragalin, V., (2006), " Adaptive designs: terminology and classification.", *Drug Information Journal*, **40(6)**, 425–435.

Efron, B. (1971). "Forcing a sequential experiment to be balanced.", *Biometrika*, **58**, 403–417.

Eisele, J. R., and Woodroofe, M. B. (1995). "Central Limit Theorems for Doubly Adaptive Biased Coin Designs.", *The Annals of Statistics*, **23**, 234–254.

Fine, J. P., and Gray, R.J. (1999). "A Proportional Hazards Model for the Subdistribution of a Competing Risk.", *Journal of the American Statistical Association*, **94**, 496–509.

Fidalgo, L.J., and Lopez, R.M.J. (2014). "Optimal experimental designs for partial likelihood information.", *Computational Statistics and Data Analysis*, **71**, 859–867.

Food and Drug Administration (FDA). (2004). "Innovation or stagnation: Challenge and opportunity on the critical path to the new medical products."

Food and Drug Administration (FDA). (2006). "Investigational New Drug Applications Prepared and Submitted by Sponsor-Investigators ."

Food and Drug Administration (FDA). (2006b). "Innovation or stagnation: Critical path opportunities report."

Food and Drug Administration (FDA). (2010). "Guidance for industry. Adaptive design clinical trials for drugs and biologics (draft document)."

Gupta, R.D., Gupta, R.C., and Sankaran, P.G. (2004)." Some characterization results based on the (reversed) hazard rate function.",*Comm. Statist.TheoryMethods*, **33 (12)**, 3009–3031.

Hardwick,J., Oehmke,R., and Stout. Q.F. (2006). "New Adaptive Designs for Delayed Response Models." , *Journal of Sequential Planning and Inference*, **136**, 1940–1955.

Hu, F., and Rosenberger, W.F. (2003), "Optimality, Variability, Power: EvaluatingResponse-Adaptive Randomization Procedures for Treatment Comparisons. ", *Journal of the American Statistical Association*,**98**, 671–678.

Hu, F., Zhang, L.X, and He, X. (2009), "Efficient Randomized-Adaptive Designs.", *The Annals of Statistics*, **37**, 2543–2560.

Hu, F., and Zhang, L.X. (2004), "Asymptotic Properties of Doubly Adaptive Biased Coin Designs for Multitreatment Clinical Trials.", *The Annals of Statistics*, **32**, 268–301.

Hu, F., Rosenberger, W. F. and Zhang, L.X. (2006). "Asymptotically best response-adaptive randomization procedures.", *Journal of Statistical Planning and Inference*, **136**, 1911–1922.

Hu, F. and Rosenberger, W.F. (2006)." The Theory of Response-Adaptive Randomiza-

tioin Clinical Trials.", *Wiley*. New York.

Hu, J., Zhu, H., and Hu, F. (2015), " A Unified Family of Covariate-Adjusted Response-Adaptive Designs Based on Efficiency and Ethics, Journal of the American Statistical Association.", **110(509)**, 357–367.

International Conference On Harmonisation Off Technical Requirements For Registration Of Pharmaceuticals For Human Use(ICH). (1996) . "Guideline For Good Clinical Practice E6(R1),"

International Conference On Harmonisation Off Technical Requirements For Registration Of Pharmaceuticals For Human Use(ICH). (1998) . "Guideline For Good Clinical Practice E9,"

Ivanova, A. (2003). "A play-the-winner type urn model with reduced variability.", *Metrika*, **58**, 1–13.

Kalbfleisch, J.D. and Prentice, R.L. (1980)." The statistical analysis of failure time data.", *John Wiley and Sons*, New York.

Kalish, A. L. and Harrington, P. D. (1988)."Efficiency of Balanced Treatment Allocation for Survival Analysis.", *Biometrics*, **44**, 815–821.

Karapetis, C.S., Khambata-Ford, S., Jonker, D.J., O' Callaghan, C.J., Tu, D., Tebbutt, N.C., Simes, R.J., Chalchal, H., Shapiro, J.D., Robitaille, S., Price, T.J., Shepherd, L., Au, H-J, Langer, C., Moore, M.J., and Zalchberg, J.R. (2008), "K-ras Mutations and Benefit From Cetuximab in Advanced Colorectal Cancer.", *The New England Journal of Medicine*, **359**, 1757–1765. [39,50].

Konstantinou, M., Biedermann, S., and Kimber, A.C. (2015)."Optimal Designs For Two-Parameter Nonlinear Models With Application To Survival Models.", *Statistica*

*Sinica*, **24**,415–428.

Koziol, J.A. and Green, S.B. (1976). "A Cramer-von Mises statistic for randomly censored data.", *Biometrika*, **63**, 465–474.

Latta, R.B. (1981), "A Monte Carlo Simulation of Some Two-Sample Rank TestsWith Censored Data.", *Journal of the American Statistical Association*, **76**, 713–719.

Lawless, J. (2003),"Statistical Models and Methods for Lifetime Data.", *Wiley*, New York.

McCullagh, P. and Nelder, J. A. (1989), "Generalized linear models.", *Chapman and Hall London* New York.

Melfi, V.F., Page, C., and Geraldes, M. (2001), "An adaptive randomized design with application to estimation.", *Canadian Journal of Statistics*, **20**, 328–342.

Njamen-Njomen, D. and Ngatchou-Wandji, J. (2014), "Nelson-Aalen and Kaplan-Meier Estimators in Competing Risks.", *Applied Mathematics*, **5**, 765–776.

Pintille, P. (2007), "Analysing and interpreting competing risk data.", *Statistics in Medicine*, **26(6)**, 1360–1367.

Pocock, S.J., and Simon, R. (1975), "Sequential Treatment Assignment With Balancing for Prognostic Factors in the Controlled Clinical Trial.", *Biometrics*, **31**, 103–115.

Porta, P., Guadalupe, G., Calle,M.L., and Nuria, M. (2007), "Competing Risk Methods.", PhD Thesis. UPC, Campus Nord, Barcelona.

Prentice,R. L., Kalbfleisch, J. D. , Peterson,A. V., Jr., Flournoy, N., Farewell V. T., and Breslow, N. E. (1978), "The Analysis of Failure Times in the Presence of Competing

Risks.", *Biometrics*, **34**, 541–554.

Rosenberger,W.F., and Seshaiyer, P. (1997), "Adaptive Survival Trials.", *Journal of Biopharmaceutical Statistics*, **7**, 617–624.

Rosenberger, W . F. and Lachin,J . M. (2002). "Randomization in Clinical Trials Theory and Practice." ,*Wiley*, New York.

Rosenberger, W.F., and Sverdlov, O. (2008), "Handling Covariates in the Design of Clinical Trials.", *Statistical Science*, **23**, 404–419.

Rosenberger, W.F., Vidyashankar, A.N., and Agarwal, D.K. (2001), "Covariate-Adjusted Response-Adaptive Designs for Binary Response.", *Journal of Biopharmaceutical Statistics*, **11**, 227–236.

Sverdlov, O., Rosenberger, W.F., and Ryeznik, Y. (2013), " Utility of Covariate-Adjusted Response-Adaptive Randomization in Survival Trials.", *Statistics in Biopharmaceutical Research*,**5:1**,38–53.

Tsiatis, A.A. (1981), "A Large Sample Study of Cox's Regression Model.", *The Annals of Statistics*, **9**, 93–108.

Zhang, L-X, and Hu, F. (2009), "A New Family of Covariate- Adjusted Response-Adaptive Designs and Their Properties.", *Applied Mathematics A Journal of Chinese Universities*, **24**, 1–13.

Zhang, L. and Rosenberger,W. E (2006) "Response-adaptive randomization in clinical trials with continuous outcomes.", *Biometrics*, **62**, 562–569.

Zhang, L.-X., Hu, F., Cheung, S.H., and Chan,W.S. (2007), "Asymptotic Properties of Covariate-Adjusted Response-Adaptive Designs.", *The Annals of Statistics*, **35**,

1166–1182.

Zhang, L., and Rosenberger, W.F. (2007), "Response-Adaptive Randomization for Survival Trials: The Parametric Approach", *Applied Statistics*, **56**, 153–165.