Dissertation

Assessment of Randomization Procedures in the Presence of Selection and Chronological Bias

David Schindler

Submitted in fulfillment of the
requirements for the degree of
"Doktor der Naturwissenschaften"
at the faculty of statistics
of the TU Dortmund University
in August, 2016

Referees:
Prof. Dr. Joachim Kunert
Prof. Dr. Christine H. Müller

Date of oral examination: December 14th, 2016

Acknowledgement

This thesis is embedded in the IDeAl project, which has received funding from the European Union's 7th Framework Programme for research, technological development and demonstration under Grant Agreement no. 602552.

Contents

Chapter 1

Introduction

A randomized controlled clinical trial is considered to be the gold standard for clinical trials. If it is possible, the randomized clinical trial should be masked or blinded for both physicians as well as patients before and after the treatment and is therefore called double-blind. In this thesis a new framework for the assessment of randomization procedures with respect to several objectives on a two-armed clinical trial with parallel group design is presented. The objectives of particular interest are the susceptibility of a randomization procedure to both selection and chronological bias. It is assumed that each objective (for example a type of bias) is measured on the basis of a corresponding criterion or even several criteria. The new framework is based on desirability functions, which have been first introduced by Harrington Jr. (1965) with the aim to ensure the quality of produced goods. In general, a desirability function maps the values of a criterion to a dimensionless scale in the interval $[0, 1]$. Afterwards, the desirability scores of several criteria are combinable with the geometric mean. In this way a unified score for several objectives, which are reflected by the corresponding criteria, is derived. On the basis of the desirability scores the objectives are linked and it is possible to assess the objectives of a randomization procedure simultaneously. One major aim of this thesis is to start a scientific discussion about the choice of an appropriate randomization procedure dependent on the clinical trial and its objectives.

The thesis focuses on the application of the new framework to two types of bias, which have already been discussed extensively in the literature: The selection bias (see Proschan, 1994; Tamm et al., 2012; Kennes et al., 2011;

Berger, 2005b) and the chronological bias (see Rosenkranz, 2011; Tamm and Hilgers, 2014; Rosenberger and Lachin, 2016). Both the potential and the occurrence of selection bias in clinical trials without concealment after the treatment (which can be impossible or even neglected) are investigated. The potential of selection bias is measured on the basis of the expected proportion of correct guesses. A correct guess is completed if the next to be assigned treatment (before its execution) is predicted correctly with the knowledge of past treatment allocations. In the ICH E9 (1998) guideline it is recommended to measure the influence of bias on the test decision. Therefore, the influence of selection bias on the test decision of Student's t-test, which is conducted to detect a difference in the effects of the given treatments, in dependency of a given strength of selection bias η is determined (see Proschan, 1994; Kennes, 2013). In contrast to Langer (2014) a more general approach for deriving the influence of bias on the test decision of Student's t-test is examined.

Beside the susceptibility of a randomization procedure to selection bias its susceptibility to chronological bias is investigated. Chronological bias is assumed to be caused by long recruitment times of patients or a learning curve/effect of a surgeon (see Hopper et al., 2007). The only possibility to correct the analysis of a randomized clinical trial for chronological bias is to consider the observed bias in the analysis. However, the true form of the bias is often unknown and thus randomization can be used to limit the influence of chronological bias on the distorted treatment effect estimation. Chronological bias is treated as an unobserved covariate in a linear model which consists of a general mean and the treatment effect. The influence of chronological bias on the test decision of Student's t-test, which is performed to detect a difference in the effects of the given treatments, is computed. The resulting criteria for assessing chronological bias are the distorted type-I-error probability and power conducting Student's t-test in the analysis, which is assumed to be conducted unadjustedly for a predefined time trend.

Two approaches for the assessment of randomization procedures are contrasted: The sequence-based approach and the average-based approach. The first approach measures the investigated criteria conditioned on the individual randomization sequences of a randomization procedure, whereas the second approach assesses the average values of the criteria of a given randomization procedure. The above mentioned criteria use different scales, hence adjustable desirability functions are used to map the values of the

criteria or the criteria‘s mean values to the interval $[0, 1]$. The transformed values are summarized with the (weighted) geometric mean to a unified score. In this way the performance of a randomization procedure with respect to multiple criteria can be investigated. Both the sequence-based approach and the average based approach are used for the assessment of randomization procedures in the presence of selection as well as chronological bias. Finally, the derived results from both approaches are compared.

The structure of the thesis is as follows: In Chapter 2 the work on randomization in clinical trials in the literature is reviewed. First, the notation and terminology is introduced. Therefore, various randomization procedures are presented; these are: Permuted Block Randomization, Randomized Permuted Block Randomization, Random Allocation Rule, Truncated Binomial Design, Complete Randomization, (see for these five randomization procedures Rosenberger and Lachin, 2016, Chapter 3), Efron's Biased Coin Design (see Efron, 1971), and Big Stick Design (see Soares and Wu, 1983). The last three mentioned randomization procedures belong to the class of the generalized biased coin design presented in Chen (1999). Afterwards, the properties of several randomization procedures are discussed and additional terms are defined. In the third section of Chapter 2, the homogeneous population model and the requirements for the evaluation of a randomized clinical trial with Student's t-test are presented. The chapter closes with the definition of the population models affected by selection and chronological bias.

For the investigation of chronological and selection bias on the basis of the type-I-error probability, as it is recommended in the ICH E9 (1998) guideline, theoretical derivations are necessary. Chapter 3 presents the foundation for the assessment of both chronological and selection bias on the basis of the distorted rejection probability under Student's t-test conditioned on the randomization sequence. It is assumed that Student's t-test is conducted to detect a difference in the effects of the given treatments. This treatment comparison is conducted unadjustedly for the present bias. The exact distribution of the test statistic of Student's t-test conditioned on a randomization sequence is derived. Therefore, it is assumed that the treatment effects are interfered with a bias vector, which is not considered in the analysis. The derived distribution of the test statistic conditioned on a randomization sequence is a doubly-noncentral t-distribution with noncentrality parameters δ and λ. Particularly, the effect of chronological bias on the noncentrality

parameters is discussed and some dependencies are shown. Furthermore, the influence of δ and λ on the density function of the doubly-noncentral t-distribution is examined and visualized exemplarily for a linear time trend. The first section of Chapter 3 closes with an example, which shows the impact of an unadjusted linear time trend on the distorted test decision of Student's t-test conditioned on the randomization sequence. In the second section of Chapter 3 the derived results for modelling chronological bias are transferred to the situation of selection bias. In the last section of Chapter 3 a more general bias model is considered under the Random Allocation Rule. In contrast to Rosenkranz (2011) it is derived that the pooled variance of the two treatment groups is an unbiased and not an asymptotically unbiased estimator for the variance of the difference between the two treatment groups under the assumption that the clinical trial is affected by a fixed bias. Furthermore, it is shown that if the bias in a clinical trial suffices the central limit theorem for samples from a finite population according to Hajek (1960), the test statistic of Student's t-test is asymptotically normally distributed. Finally, for a linear time trend and a step time trend some asymptotical results concerning the rejection probability under the Random Allocation Rule using Student's t-test are derived.

Chapter 4 introduces the class of desirability functions according to Derringer and Suich (1980). To give an example, a desirability function is applied on a given criterion of randomization procedure. Afterwards, the combination of several desirability scores is discussed. Chapter 4 closes with the presentation of the properties of desirability scores. A sensitivity analysis of the new framework is carried out in Chapter 5. Initially the sample size $N = 4$ is considered to show how the new framework works. The randomization procedure Efron's Biased Coin Design is assessed. The sensitivity analysis is performed for both the sequence-based approach and the average-based approach. An extensive sensitivity analysis is conducted for the sample size $N = 12$. The sample sizes $N = 50$ and $N = 200$ are only discussed briefly because many results are transferable from the situation with $N = 12$. In Chapter 6 the new framework is used to assess several settings of randomization procedures in the presence of both selection and chronological bias. In repetitive manner the sequence-based as well as the average-based approach are investigated for the sample sizes $N \in \{12, 50, 200\}$. For the sample size $N = 12$ all possible randomization sequences of the randomization procedures are assessed and

for the larger sample sizes $N = 50$ and $N = 200$ simulations are used to reflect the behavior of the corresponding randomization procedures. The thesis closes with a discussion of the results in Chapter 7.

Chapter 2

Literature review

In the beginning of this chapter the notation and terminology is introduced. Unless explicitly stated otherwise, throughout this thesis a two-armed trial with parallel group design and a continuous endpoint is considered. Furthermore, the sample size N is assumed to be even. In the second section of the chapter various randomization procedures and their properties are presented. In the third section the homogeneous population model is introduced and the evaluation of a randomized clinical trial is discussed. Finally, the assumed models for selection bias as well as chronological bias are described and the impact on the homogeneous population model is shown.

2.1 Notation and terminology

Let $\Gamma_N := \{-1, 1\}^N$ be the space of all possible randomization sequences in a two-armed clinical trial with parallel group design and a total number of N patients. Let $\boldsymbol{T} = (T_1, \ldots, T_N)^T$ be a random vector which takes values in Γ_N. A randomization sequence $\boldsymbol{t} = (t_1, \ldots, t_N)^T \in \Gamma_N$ is a realization of $\boldsymbol{T}$ and its according probability of occurrence is denoted as $p_{\boldsymbol{t}} \in [0, 1]$. Furthermore, T_n defines the nth element of $\boldsymbol{T}$ and t_n the nth element of $\boldsymbol{t}$ with $n = 1, \ldots, N$. When the nth patient is entering the trial, he or she is assigned in the following manner:

$$T_n = \begin{cases} 1, & \text{thus treatment } E \text{ is assigned to the } n\text{th patient} \\ -1, & \text{thus treatment } C \text{ is assigned to the } n\text{th patient} \end{cases}, \qquad (2.1)$$

where treatment E defines the experimental treatment and treatment C the control treatment. Let $N_E(n, \boldsymbol{T}) = 1/2 \sum_{i=1}^{n} (T_i + 1)$ be a random variable denoting the number of patients assigned to treatment E after $n \leq N$ allocations and $N_C(n, \boldsymbol{T}) = n - N_E(n, \boldsymbol{T})$ be a random variable denoting the number of patients assigned to treatment C after $n \leq N$ allocations. If $\boldsymbol{t}$ is a realization of $\boldsymbol{T}$, the term $N_j(n, \boldsymbol{t})$ with $j \in \{E, C\}$ denotes the number of patients assigned to treatment E or alternatively treatment C after $n \leq N$ allocations. In the strict sense, $N_j(n, \boldsymbol{t})$ with $j \in \{E, C\}$ depends only on the first $n < N$ elements of $\boldsymbol{t}$. To be concise $N_j(n, \boldsymbol{t})$ with $j \in \{E, C\}$ is used instead of $N_j(n, (t_1, t_2, \ldots, t_n)^T)$. This terminology holds for all further defined terms and expressions. Below, random variables for the allocation ratios, the imbalance, the number of returns to origin, and the number of deterministic allocations are defined.

Definition 2.1:

a) The allocation ratios for treatment E and treatment C after $n \leq N$ assigned patients with $\boldsymbol{t} \in \Gamma_N$ are defined as:

$$AR_E(n, \boldsymbol{t}) = \frac{N_E(n, \boldsymbol{t})}{n} \quad \textit{and} \quad AR_C(n, \boldsymbol{t}) = \frac{N_C(n, \boldsymbol{t})}{n}\,. \tag{2.2}$$

b) The imbalance in the number of patients assigned to the treatments E and C after $n \leq N$ assigned patients with $\boldsymbol{t} \in \Gamma_N$ is defined as:

$$D(n, \boldsymbol{t}) = N_E(n, \boldsymbol{t}) - N_C(n, \boldsymbol{t}) = \sum_{i=1}^{n} t_i\,. \tag{2.3}$$

c) The number of returns to origin (nro) after $n \leq N$ assigned patients with $\boldsymbol{t} \in \Gamma_N$ is defined as:

$$nro(n, \boldsymbol{t}) = \sum_{i=1}^{n} \mathbb{1}_{\{D(i,\boldsymbol{t})=0\}}\,, \tag{2.4}$$

where $\mathbb{1}_{\{D(i,\boldsymbol{t})=0\}}$ is one if $D(i, \boldsymbol{t}) = 0$ and zero otherwise.

d) The number of deterministic allocations (nda) after $n \leq N$ assigned patients with $\boldsymbol{t} \in \Gamma_N$ is defined as:

$$nda(n, \boldsymbol{t}) = \sum_{i=1}^{n} \mathbb{1}_{\{P(T_i=1|T_1=t_1, T_2=t_2, \ldots, T_{i-1}=t_{i-1}) \in \{0,1\}\}}\,, \tag{2.5}$$

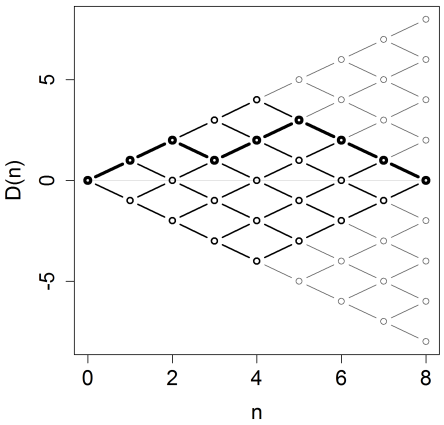
where $\mathbb{1}_{\{P(T_i=1|T_1=t_1,T_2=t_2,\ldots,T_{i-1}=t_{i-1})\in\{0,1\}\}}$ *is one if* $P(T_i = 1|T_1 = t_1, T_2 = t_2, \ldots, T_{i-1} = t_{i-1}) \in \{0, 1\}$ *and zero otherwise.*

2.2 Overview of various randomization procedures

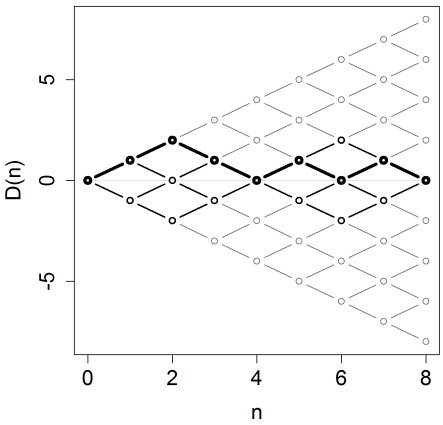
In this section the following randomization procedures are introduced: Permuted Block Randomization, Randomized Permuted Block Randomization, Random Allocation Rule, Truncated Binomial Design, Complete Randomization (see for these five randomization procedures Rosenberger and Lachin, 2016, Chapter 3), Efron's Biased Coin Design (Efron, 1971), and Big Stick Design (Soares and Wu, 1983). Furthermore, the most specific properties of the randomization procedures are discussed and at the end of the section additional terms are defined.

2.2.1 Random Allocation Rule

If the the total sample number N is known and it is desired that the treatment groups are equally sized at the end of the clinical study, it is imaginable to assign randomly both treatments E and C $N/2$ times – randomly means that all possible randomization sequences with balanced groups at the end of the the trial are equiprobable. In the literature this kind of randomization is called Random Allocation Rule (RAR) (see Rosenberger and Lachin, 2016, pp. 38-39). The Random Allocation Rule can be described by a simple urn model. An urn is filled with $N/2$ black and $N/2$ white balls. For the first selected patient one ball is drawn randomly and not replaced. If the drawn ball is black, the patient will be assigned to treatment group E, if not, the patient will be part of treatment group C. For the second patient one ball is drawn from the left balls in the urn and so on. All possible allocation paths of the Random Allocation Rule are presented on the left in Figure 2.1 with $N = 8$. In the figure the imbalances $D(n) := D(n, \boldsymbol{t})$ (see Equation (2.3)) are plotted for all randomization sequences against their corresponding patient number n. Afterwards, the points $(n, D(n))$ of each randomization sequence are connected. The diagram visualizes all possible randomization sequences. Furthermore, one particular randomization sequence is highlighted in heavy bold.



(a) Possible paths of RAR with $N = 8$ in light bold. A particular path is highlighted in heavy bold.



(b) Possible paths of PBR(4) with $N = 8$ in light bold. A particular path is highlighted in heavy bold.

Figure (2.1): The subfigures show all possible allocation paths of RAR and PBR(4) with $N = 8$.

2.2.2 (Randomized) Permuted Block Randomization

In comparison to the Random Allocation Rule, the (Randomized) Permuted Block Randomization has an additional (random) blocking factor. Patients with the same blocking factor form a unit during the randomization process. Let $k \in \mathbb{N}$ be even and $N/k = K$ be a positive integer. Thus, the corresponding Permuted Block Randomization consists of K blocks with block length k. Permuted Block Randomization with block length k is denoted by PBR(k). In each block of PBR(k) half of the patients are assigned to the treatment groups E and C, respectively. All randomization paths are equiprobable. Thus, Permuted Block Randomization is the same as conducting K times the Random Allocation Rule. Figure 2.1 shows alle possible allocations paths of Permuted Block Randomization with block length four and $N = 8$.

Because of its simple feasibility and its good balancing behavior Permuted Block Randomization is probably the most commonly used randomization procedure in randomized clinical trials. Permuted Block Randomization is already reviewed extensively in the literature (see for example Zelen, 1974) and even the covariance matrix of the patients' allocations of this randomization procedure is derived (see Rosenberger and Lachin, 2016). In

contrast to the Random Allocation Rule the total sample size N can be unknown at the beginning of the clinical trial and it is even possible that the last block is unfilled. A block is termed as unfilled when only $i < k$ patients are assigned to one block with length k.

The ICH E9 (1998) guideline recommends using two or more different block lengths selected randomly for each block in order to prevent selection bias. In Shao and Rosenberger (2016) Randomized Permuted Block Randomization (RPBR) is reviewed. In a similar manner to the PBR(k) one block length k, where k is an even integer, is defined. Afterwards, blocks of the lengths $2, 4, 6, \ldots, k$ are selected equiprobablely until all N patients are included into the clinical trial. The used randomization procedure is denoted by RPBR(k). In comparison to the Random Allocation Rule the property of equally sized treatment groups at the end of the clinical trial is not ensured for Randomized Permuted Block Randomization. The maximal possible absolute value of the final imbalance in the number of patients assigned to the treatments E and C is $k/2$.

2.2.3 Truncated Binomial Design

Blackwell and Hodges Jr. (1957) presented the Truncated Binomial Design (TBD) as an alternative to the Random Allocation Rule to fullfil the aim to assign exactly half of the patients to each treatment group. The conditional allocation probability $P(T_{n+1} = 1 | T_1 = t_1, \ldots, T_n = t_n)$ of the Truncated Binomial Design is given by:

$$P(T_{n+1} = 1 | T_1 = t_1, \ldots, T_n = t_n) = \begin{cases} 1, & \text{if } 1/2 \sum_{i=1}^{n} (t_i - 1) = -N/2 \\ 0, & \text{if } 1/2 \sum_{i=1}^{n} (t_i + 1) = N/2 \\ 0.5, & \text{else} \end{cases} . \tag{2.6}$$

Consequently, a fair coin is tossed for both treatment groups as long as each treatment is assigned less than $N/2$ times. Afterwards, the randomization list would be filled with the opposite treatment until N patients are included. Blackwell and Hodges Jr. (1957) have proven that this randomization procedure is the least susceptible one to their model for selection bias. The model that they used is introduced in Section 2.3.2. The possible paths of

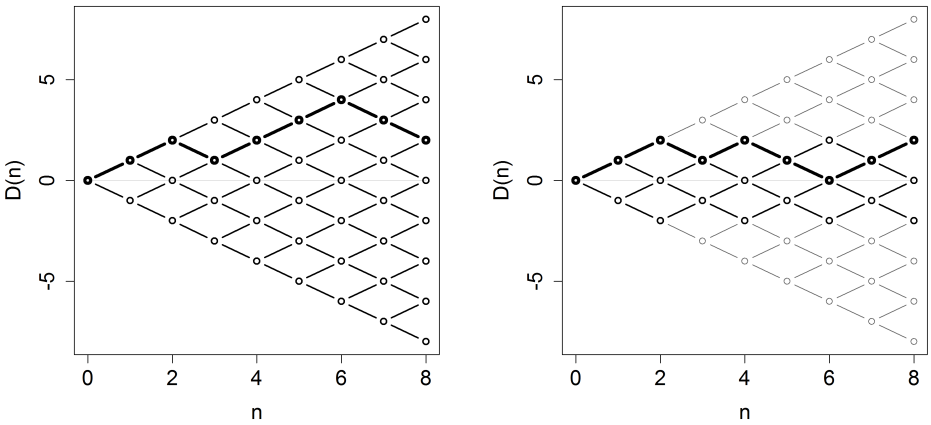
the Truncated Binomial Design are the same as for the Random Allocation Rule. The different probabilities of the individual randomization sequences are the only difference between these two randomization procedures. For the Truncated Binomial Design not all possible randomization sequences are equiprobable, which is one property of the Random Allocation Rule.

2.2.4 Generalized biased coin design

In this section the generalized biased coin design defined by Chen (1999) is presented. The generalized biased coin design is described as follows: A fair coin for the allocation of the $(n+1)$th patient is tossed if the sum of the first $n < N$ allocations is zero, otherwise a biased coin is tossed if the difference in the number of patients assigned to each of the treatment groups does not exceed a prespecified fixed tolerated imbalance boundary. In the literature this boundary is called the maximum tolerated imbalance (MTI). If the absolute value of the difference in the number of patients assigned to each of the treatment groups exceeds the MTI, a deterministic allocation for the $(n+1)$th patient is completed, so that the absolute value of the imbalance in patient numbers in the two treatment groups is reduced. Let $p \in [0.5, 1]$ and MTI $\in \mathbb{N}$, hence the conditional allocation probability $P(T_{n+1} = 1 | T_1 = t_1, \ldots, T_n = t_n)$ of the generalized biased coin design introduced by Chen (1999) is given by:

$$P(T_{n+1} = 1 | T_1 = t_1, \ldots, T_n = t_n) = \begin{cases} 1, & \text{if } \sum_{i=1}^{n} t_i = -\text{MTI} \\ p, & \text{if } 0 > \sum_{i=1}^{n} t_i > -\text{MTI} \\ 0.5, & \text{if } \sum_{i=1}^{n} t_i = 0 \\ 1-p, & \text{if } 0 < \sum_{i=1}^{n} t_i < \text{MTI} \\ 0, & \text{if } \sum_{i=1}^{n} t_i = \text{MTI} \end{cases} . \tag{2.7}$$

The generalized biased coin design represents three randomization procedures, which are of particular interest in this thesis. First, let MTI $= \infty$ and $p = 0.5$, thus Equation (2.7) describes a fair coin toss for each patient. The corresponding randomization procedure is called Complete Randomization (CR) – every patient allocation is done completely random. One property of Complete Randomization is, that the random variables T_i and T_j with



(a) Possible paths of CR and EBC(p) with $N = 8$ in light bold. A particular path is highlighted in heavy bold.

(b) Possible paths of BSD(2) with $N = 8$ in light bold. A particular path is highlighted in heavy bold.

Figure (2.2): The subfigures show all possible allocation paths of CR, EBC and BSD(2) with $N = 8$.

$i \neq j$ and $i, j \in \{1, 2, \ldots, N\}$ are independent. Consequently, the covariance matrix of Complete Randomization is the identity matrix (see Rosenberger and Lachin, 2016, pp. 38-39). One disadvantage of using Complete Randomization is, that there is a non-negligible probability of realizing a high absolute value of the final imbalance concerning patient numbers in the clinical trial. In particular, this is a problem for small clinical trials. In order to prevent this problem Efron (1971) presented Efron's Biased Coin Design (EBC(p)), which is dependent on the parameter $p \in [0.5, 1]$ according to Equation (2.7). Efron suggested in his work to set the parameter p to $2/3$ while keeping MTI $= \infty$ as in the setting of Complete Randomization. Thus, the probability of realizing undesired high absolute values of of the final imbalance concerning patient numbers in the clinical trial is minimized, but not ruled out. Twelve years later, Soares and Wu (1983) introduced Big Stick Design (BSD). This special design only deals with the MTI and p is set to 0.5 as in the setting for Complete Randomization. A Big Stick Design with a fixed MTI is denoted by BSD(MTI). Figure 2.2 shows the possible paths of Complete Randomization, Efron's Biased Coin Design, and BSD(2) with $N = 8$. The only difference between Complete Randomization and Efron's Biased Coin Design is that the probabilities of occurrence of the

individual allocation paths are different. These probabilities are not visible in the printed graphs.

2.2.5 Properties of randomization procedures

A randomization procedure (RP) is a probability distribution over the set $\Gamma_N := \{-1, 1\}^N$. Thus, a randomization procedure is uniquely defined by the probability of occurrence of the individual randomization sequences. Below, the set of possible randomization sequences of a given randomization procedure is specified.

Definition 2.2: (Set of possible randomization sequences of an RP)
Let $\boldsymbol{t} \in \Gamma_N$ and $p_{\boldsymbol{t}} := p_{\boldsymbol{t}}^{RP} \in [0, 1]$ be the probability of occurrence of the randomization sequence $\boldsymbol{t}$ dependent on a given randomization procedure RP. Thus, the set of possible randomization sequences of a given randomization procedure $\Gamma_N^{RP} \subseteq \Gamma_N$ is defined as follows:

$$\Gamma_N^{RP} := \{\boldsymbol{t} \in \Gamma_N | p_{\boldsymbol{t}} > 0\}. \tag{2.8}$$

If a randomization sequence is an element of the set Γ_N^{RP}, the randomization sequence $\boldsymbol{t}$ is said to be produced or generated by the randomization procedure RP.

Two randomization procedures are the same if all possible randomization sequences have the same probability of occurrence. Below, the maximal tolerated imbalance, the average number of returns to origin, the average number of deterministic allocations, and the probability for $T_n = 1$ with fixed n for a given randomization procedure are defined.

Definition 2.3:

a) The maximal tolerated imbalance (MTI) of a given randomization procedure RP is defined as:

$$MTI_{RP} := \max_{\substack{n=1,2,\ldots,N \\ \boldsymbol{t} \in \Gamma_N^{RP}}} |D(n, \boldsymbol{t})|. \tag{2.9}$$

b) The average number of returns to origin (avnro) of a given randomization procedure RP is defined as:

$$avnro_{RP} := \sum_{\boldsymbol{t} \in \Gamma_N^{RP}} p_{\boldsymbol{t}}\, nro(N, \boldsymbol{t})\,. \tag{2.10}$$

c) The average number of deterministic allocations (avnda) of a given randomization procedure RP is defined as:

$$avnda_{RP} := \sum_{\boldsymbol{t} \in \Gamma_N^{RP}} p_{\boldsymbol{t}}\, nda(N, \boldsymbol{t})\,. \tag{2.11}$$

d) The probability for $T_n = 1$ with fixed $n \in \{1, \ldots, N\}$ of a given randomization procedure RP is defined as:

$$P_{RP}(T_n = 1) = \sum_{\boldsymbol{t} \in \Gamma_N^{RP}} \mathbb{1}_{\{t_n = 1\}}\, p_{\boldsymbol{t}}\,, \tag{2.12}$$

where $\mathbb{1}_{\{t_n=1\}}$ is one if $t_n = 1$ and zero otherwise.

If $P_{\mathrm{RP}}(T_n = 1) = \mathrm{AR}_E$ holds for all $n = 1, 2, \ldots N$ with $\mathrm{AR}_E \in [0, 1]$, the randomization procedure preserves a fixed target allocation ratio of the experimental treatment AR_E. Unless explicitly stated otherwise, throughout this thesis the target allocation ratios of the treatments E and C are assumed to be 0.5. That means no treatment arm is preferred during the whole clinical trial and the two treatments are equiprobable for every enrolled patient independent of his or her enrollment number n. Complete Randomization serves as an example that not every final allocation ratio $\mathrm{AR}_E(N, \boldsymbol{t})$ for each randomization sequence of a given randomization procedure has to attain its target allocation ratio. For this randomization procedure the final allocation ratio is maintained, although there exists one randomization sequence where each patient is assigned to the treatment group E or alternatively to the treatment group C. The final allocation ratio 0.5 is maintained only on average.

2.3 Evaluation and assumptions of an (un)biased clinical trial

At the beginning of this section the homogeneous population model and its evaluation according to Lachin (1988) are introduced. In the second and third subsection the assumed models and the measurements of both selection and chronological bias are presented. Furthermore, the impact of selection as well as chronological bias on the evaluation of a randomized clinical trial is shown. Particularly, the influences of selection as well as chronological bias on the homogeneous population model are discussed.

2.3.1 Homogeneous population model and its evaluation

The population model presented by Lachin (1988) assumes that the sample of patients included into a clinical trial is representative of a reference population. Furthermore, the target values, which are measured on the patients after the treatments, are independent and identically distributed (i.i.d.) according to some probability function $H_{\psi_j}(y)$. The parameter vector $\boldsymbol{\psi}_j$ of $H_{\psi_j}(y)$ is dependent on the given treatment $j \in \{E, C\}$. The term $Y_{n,j}$ with $j \in \{E, C\}$ defines a random variable for the nth patient's response. In the homogeneous population model for a two-armed clinical trial with parallel group design discussed by Lachin (1988) there is one population assumed, from which all included patients are taken. After the assignment to treatment E or alternatively treatment C according to the corresponding element of the randomization list, the nth patient's response is distributed according to some probability function $H_{\psi_j}(y)$ with $j \in \{E, C\}$ (written $Y_{n,j} \sim H_{\psi_j}$). The assumed homogeneous population model is shown in Figure 2.3, where $N_j := N_j(N, \boldsymbol{T})$ with $j \in \{E, C\}$ is the final number of patients assigned to the treatment group E or alternativeley treatment group C. Due to the fact that the nth patient is allocated randomly to one of the two treatment groups either the corresponding response $y_{n,E}$ or the corresponding response $y_{n,C}$ is observable.

In Lachin et al. (1988) it is pointed out that under the assumption of a homogeneous population model the randomization process is ignored in the analysis. If the distribution $H_{\psi_j}(y)$ with $j \in \{E, C\}$ is specified, optimal estimators for the parameters $\boldsymbol{\psi}_C$ and $\boldsymbol{\psi}_E$ can be constructed and an optimal

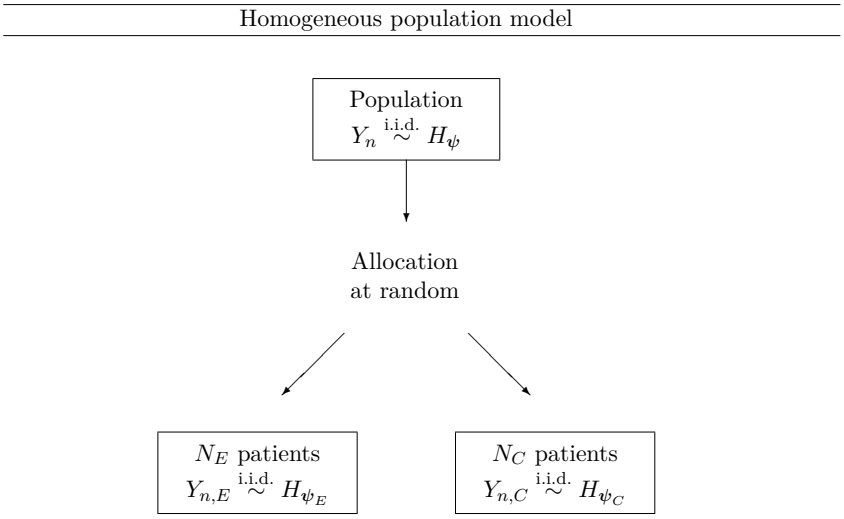


Figure (2.3): The homogeneous population model according to Lachin (1988).

test for the hypothesis $H_0 : \boldsymbol{\psi}_E = \boldsymbol{\psi}_C$ against $H_1 : \boldsymbol{\psi}_E \neq \boldsymbol{\psi}_C$ can be derived. Throughout this thesis $H_{\boldsymbol{\psi}_j}(y)$ describes the distribution function of a normal distribution with parameter vector $\boldsymbol{\psi}_j = (\mu_j, \sigma^2)^T$ and $j \in \{E, C\}$. The variance $\sigma^2 \in \mathbb{R}_+$ is assumed to be unknown and the same in both treatment groups. Thus, the investigated hypotheses are:

$$\begin{aligned} & H_0 : \boldsymbol{\psi}_E = \boldsymbol{\psi}_C && \text{vs.} && H_1 : \boldsymbol{\psi}_E \neq \boldsymbol{\psi}_C \\ \Leftrightarrow \quad & H_0 : \mu_E = \mu_C && \text{vs.} && H_1 : \mu_E \neq \mu_C \,. \end{aligned} \tag{2.13}$$

For the test problem in Equation (2.13) Student's t-test is the uniformly most powerful unbiased test. However, when Student's t-test is used the randomization process itself is not taken into account in the analysis. In order to include the randomization process in the analysis, the randomization model introduced by Lachin (1988) should be used.

For later derivations a regression model in the situation of a homogeneous population model is used. Therefore it is assumed that the observed patient responses $y_1, \ldots, y_N$ are realizations of stochastically independent random variables $Y_1, \ldots, Y_N$ dependent on $\boldsymbol{T}$, which satisfy the following model:

$$Y_n = a(T_n)^T \boldsymbol{\theta} + \epsilon_n \,, \tag{2.14}$$

where ϵ_n with $n = 1, 2, \ldots, N$ is an independent and identically normally distributed random variable with expectation zero and variance $\sigma^2 \in \mathbb{R}_+$. The vector $\boldsymbol{\theta} \in \mathbb{R}^p$ is an unknown parameter vector and $a : T_n \to \mathbb{R}^p$ a known regression function. With $\boldsymbol{Y} = (Y_1, \ldots, Y_N)^T$, $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_N)^T$, and $\boldsymbol{A} = (a(T_1)^T, a(T_2)^T, \ldots, a(T_N)^T)^T$ the Model (2.14) can be written as follows:

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{\theta} + \boldsymbol{\epsilon}\,, \tag{2.15}$$

where $\boldsymbol{\epsilon} \in \mathbb{R}^N$ is a vector of independent and identically normally distributed random variables with expectation zero and variance $\sigma^2 \in \mathbb{R}^+$. The matrix $\boldsymbol{A} \in \mathbb{R}^{N \times p}$ is called the design matrix. Transferring Model (2.15) to the homogeneous population model presented in Figure 2.3 the parameter vector $\boldsymbol{\theta} = (\theta_0, \theta_1)^T$ consists of two elements. The parameter θ_0 defines the intercept and the parameter θ_1 distinguishes between the two treatments E and C. Let $a(T_n) = (1, T_n)$ in the Model (2.14), thus it follows:

$$\boldsymbol{Y} = \begin{pmatrix} 1 & T_1 \\ 1 & T_2 \\ \vdots & \vdots \\ 1 & T_N \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{pmatrix} = \underbrace{(\mathbf{1}_N, \boldsymbol{T})}_{\boldsymbol{A}} \boldsymbol{\theta} + \boldsymbol{\epsilon}\,, \tag{2.16}$$

where the vector $\mathbf{1}_N \in \mathbb{R}^N$ contains the number one N times and $\boldsymbol{T}$ is a random variable taking values in $\{-1, 1\}^N$ (see Section 2.1).

2.3.2 Definition and model of selection bias

According to Berger (2005a) selection bias can be divided in four different types. First-order selection bias might occur either when patients are allowed to choose their own treatment group or the investigator is able to choose the treatment group for each patient after learning about his characteristics. If the clinical trial is randomized, but the randomization list is known to the investigator, he or she could influence the enrollment or the exclusion of suitable patients due to soft inclusion or exclusion criteria. This type of selection bias is called second-order selection bias. Consider a clinical trial which is randomized but only future allocations are concealed, thus the investigator might be able to predict future allocations based on past assignments

due to restrictions of the randomization method. Bias which arises in this manner is denoted as third-order selection bias. If the investigator is blinded to future allocations and past allocations are masked, then the clinical trial is only susceptible to fourth-order selection bias. This bias occurs despite masking and blinding and is also called residual selection bias (see Berger, 2005a).

Throughout this thesis clinical trials which are susceptible to third-order selection bias are of primary interest. This quantity of clinical trials includes at least all those clinical trials where the treatment concealment after its execution is not possible. This is the case in all surgical trials and in all clinical trials where the compared treatments have different obvious side-effects. In Joppi et al. (2012) 66 clinical trials for orphan drugs are reported, of which 22 were planned open label. One third of those presented clinical trials would be highly susceptible to third-order selection bias. In this thesis a two-armed clinical trial is considered because this class of clinical trials is the most susceptible one to third-order selection bias. All randomization sequences exist of only two different elements. Thus, the investigator assumes that both treatments E and C occur equiprobablely and in this way he or she can guess correctly for each patient in 50% of the cases the corresponding treatment allocation, before the patient is included. If there are some known restrictions of the randomization method to the investigator, he or she can increase this value enormously. Below, the correct guesses dependent on a guessing strategy are defined.

Correct guesses

If the investigator will not have any premonition about the randomization process, he or she will not be able to select the included patients in the clinical trial consciously. In the setting of third-order selection bias it is assumed that the investigator knows past treatment allocations. Furthermore, due to the power aspect of the treatment comparison the investigator assumes equally sized treatment groups at the end of the trial. Under these circumstances it is in accordance with Blackwell and Hodges Jr. (1957) opportune for the investigator to guess the nth allocation of a given randomization sequence

according to the following guess function:

$$G_{\mathrm{CS}}(n, \boldsymbol{t}) = \begin{cases} 1, & \text{if } N_E(n-1, \boldsymbol{t}) < N_C(n-1, \boldsymbol{t}) \\ 2\,\mathbb{B}(0.5) - 1, & \text{if } N_E(n-1, \boldsymbol{t}) = N_C(n-1, \boldsymbol{t}) \\ -1, & \text{if } N_E(n-1, \boldsymbol{t}) > N_C(n-1, \boldsymbol{t}) \end{cases}, \tag{2.17}$$

where $\mathbb{B}(p)$ defines a random variable from the Bernoulli distribution with success probability $p = 0.5$. For $n = 1$ the terms $N_E(0, \boldsymbol{t})$ and $N_C(0, \boldsymbol{t})$ are assumed to be zero. The guessing strategy presented in Equation (2.17) is called convergence strategy (CS) because the investigator assumes that the sum of the elements of the randomization sequence convergences to zero. Below, the expected number of correct guesses dependent on a guess function G of a given randomization procedure RP is defined.

Definition 2.4: (Expected number of correct guesses of an RP)
Let $G : \{1, 2, \ldots, N\} \times \Gamma_N \to \{-1, 1\}$ be a guess function and $p_{\boldsymbol{t}}$ be the probability of occurrence of the randomization sequence $\boldsymbol{t} \in \Gamma_N$ dependent on a given randomization procedure RP. Then, the expected number of correct guesses dependent on G of a given randomization procedure RP is defined as:

$$\begin{aligned} \overline{CG}_{RP}(G) &= \sum_{\boldsymbol{t} \in \Gamma_N^{RP}} \sum_{n=1}^{N} E\left(\mathbb{1}_{\{t_n = G(n, \boldsymbol{t})\}}\right) p_{\boldsymbol{t}} \\ &= \sum_{\boldsymbol{t} \in \Gamma_N^{RP}} CG(\boldsymbol{t})\, p_{\boldsymbol{t}}\,, \end{aligned} \tag{2.18}$$

where $CG(\boldsymbol{t})$ defines the expectation of correct guesses of a given randomization sequence. The bar in the term $\overline{CG}_{RP}(G)$ indicates that a weighted average value for the correct guesses of all randomization sequences is calculated.

Due to the fact that the probabilities of occurrence of the randomization sequences are strictly dependent on a given randomization procedure, the expected number of correct guesses differs dependent on the randomization procedure. In the later analysis, the expected proportion of correct guesses is often used instead of the expected number of correct guesses. Therefore, two more terms are defined as

$$\mathrm{prop}\overline{\mathrm{CG}}_{\mathrm{RP}}(G) = \frac{\overline{\mathrm{CG}}_{\mathrm{RP}}(G)}{N} \tag{2.19}$$

$$\text{and propCG}(\boldsymbol{t}) = \frac{\text{CG}(\boldsymbol{t})}{N}\,. \tag{2.20}$$

One advantage of the expected proportion of correct guesses is that if N increases, the term will not increase The expected proportion of correct guesses takes values in the interval $[0, 1]$ - in the one extreme case the investigator guesses correctly for all elements of all randomization sequences of a randomization procedure and in the other extreme case the investigator guesses wrongly for all elements of all randomization sequences of a randomization procedure.

The expectation of the correct guesses of a given randomization sequence from Definition 2.4 under the guess function G_{CS} from Equation (2.17) can be written as follows:

$$\text{CG}(\boldsymbol{t}) = \sum_{n=1}^{N} E\left(\mathbb{1}_{\{t_n = G_{\text{CS}}(n,\boldsymbol{t})\}}\right) \text{ with}$$

$$E\left(\mathbb{1}_{\{t_n = G_{\text{CS}}(n,\boldsymbol{t})\}}\right) = \begin{cases} 1, & \text{if } D(n-1,\boldsymbol{t}) \neq 0 \,\wedge\, |D(n-1,\boldsymbol{t})| > |D(n,\boldsymbol{t})| \\ 0.5, & \text{if } D(n-1,\boldsymbol{t}) = 0 \\ 0, & \text{otherwise} \end{cases}.$$

The convergence strategy will always decide for the correct treatment, whenever the randomization sequence reduces the difference between the numbers of patients assigned to each of the treatment groups. If the numbers of assigned patients are equal, there is a 50% probability of guessing correctly. Consequently, 0.75 is the maximal possible value. This value is reached for any randomization procedure with the property $D(n, \boldsymbol{t}) = 0$ for all possible randomization sequences with n even. In the literature such randomization sequences are called alternating sequences. The randomization procedure PBR(2), which is introduced in Section 2.2.2, consists of only such alternating randomization sequences. Thus, PBR(2) is the most susceptible randomization procedure to the presented convergence strategy. In the appendix the following relationship between the expected number of correct guesses with the guessing function G_{CS} from Equation (2.17) and the number of returns to origin from Equation (2.4) of a given randomization procedure

RP is derived:

$$\begin{aligned}\overline{\mathrm{CG}}_{\mathrm{RP}}(G_{\mathrm{CS}}) &= \sum_{\boldsymbol{t}\in\Gamma_N^{\mathrm{RP}}} \sum_{n=1}^{N} E\left(\mathbb{1}_{\{t_n=G_{\mathrm{CS}}(n,\boldsymbol{t})\}}\right) p_{\boldsymbol{t}} \\ &= \sum_{\boldsymbol{t}\in\Gamma_N^{\mathrm{RP}}} \left(\frac{\mathrm{nro}(N,\boldsymbol{t})}{2} + \frac{N}{2} - \mathbb{1}_{\{|D(N,\boldsymbol{t})|>0\}} \frac{|D(N,\boldsymbol{t})|-1}{2}\right) p_{\boldsymbol{t}}\,.\end{aligned} \tag{2.21}$$

Under the assumption of $D(N,\boldsymbol{t}) = 0$ for all randomization sequences of a given randomization procedure, the expected number of correct guesses is strictly dependent on the number of returns to origin. Consequently, minimizing the expected number of correct guesses from Equation (2.21) is equivalent to minimizing the number of returns to origin of a randomization procedure. The derived relationship helps to attain a deeper understanding of the convergence strategy. To protect a clinical trial against selection bias achieved by the convergence strategy, a randomization procedure should not increase the number of returns to origin and should achieve a high value of the absolute value of the final imbalance. However, this is a conflict with the power aspect of clinical trial. The power for the treatment comparison is maximized if the clinical trial is balanced at the end.

Other guess functions than G_{CS} defined in Equation (2.17) are possible. The divergence strategy (DS) (see Blackwell and Hodges Jr., 1957) forms a contrast to the convergence strategy. In this scenario the investigator assumes $N_E(N,\boldsymbol{t}) > N_C(N,\boldsymbol{t})$ or $N_E(N,\boldsymbol{t}) < N_C(N,\boldsymbol{t})$. Thus, the investigator supposes that one treatment is favored by randomization at the end. This strategy is important in response adaptive clinical trials (see Hu and Rosenberger, 2006). With the divergence strategy the investigator would maximize the expected number of correct guesses in such clinical trials. One last strategy for the investigator is to guess deterministic allocations due to the restriction of the randomization method. Randomization methods which have restrictions due to its maximal tolerated imbalance are susceptible to this guessing strategy. Finally, other guessing strategies for the investigator are imaginable and even a combination of already presented strategies are possible to maximize the expected number of correct guesses.

Modelling selection bias

In accordance with Proschan (1994) the investigator can use a so called biasing policy when selecting patients. Dependent on a guess function G the investigator can select the nth patient in the following manner:

$$Y_n \sim \begin{cases} \mathcal{N}(\mu+\eta,\sigma^2), & \text{if } E(G(n,\boldsymbol{t})) = \quad 1 \\ \mathcal{N}(\mu-\eta,\sigma^2), & \text{if } E(G(n,\boldsymbol{t})) = -1 \\ \mathcal{N}(\mu,\sigma^2), & \text{else} \end{cases}, \tag{2.22}$$

where $G : \{1,2,\ldots,N\} \times \Gamma_N \to \{-1,1\}$ defines a guess function and $\eta \in \mathbb{R}$ is the strength of selection bias. The parameter η can be interpreted as a shift in the expectation $\mu \in \mathbb{R}$ of the nth patient. In the presented Model (2.22) it is assumed that the variance σ^2 is not affected by η.

Influence of selection bias on the population model

Below, the influence of the selection bias model according to Proschan (1994) on the homogeneous population model presented in Section 2.3.1 is discussed. The population model affected by selection bias is illustrated in Figure 2.4. It is assumed that the underlying population can be divided into three subgroups: The good responders, the normal responders, and the poor responders. The subgroup, the nth patient belongs to, is always chosen freely by the investigator. Furthermore, this additional attribute of a patient's response is independent of the given treatment. In what follows, the guess function according to the convergence strategy defined in Equation (2.17) and the biasing policy presented in Equation (2.22) is considered. Thus, the investigator enrolls a poorly responding patient into the clinical trial if the occurred imbalance in the clinical trial is positive. If the occurred imbalance is negative, it is assumed that the investigator enrolls a well responding patient and alternatively if the occurred imbalance is zero, a normally responding patient is included.

In this thesis unobserved selection bias is investigated. That means the population model presented in Section 2.3.1 is assumed in the analysis, but in fact the situation of a population model affected by selection bias is present. Thus, in the evaluation of this thesis an analysis unadjusted for selection bias for the test problem described in Equation (2.13) is undertaken. In

the literature there are already several tests for selection bias available (see Berger, 2005b; Kennes, 2013). However, always incorporating selection bias in the analysis would lead to a loss of power in situations when the effects of the treatments are different and the trial is not affected by selection bias.

Revisiting the regression model introduced in Equation (2.16), the situation of a population model affected by selection bias can be expressed by the function $a(T_n) = (1, T_n, b_n)$ with $b_n = \text{sgn}(D(n-1, \boldsymbol{T}))$. The sign function $\text{sgn}(D(n-1, \boldsymbol{T}))$ is one if the imbalance after the inclusion of $n-1$ patients is positive, minus one if the imbalance after the inclusion of $n-1$ patients is negative, and zero otherwise. Consequently, the following model is derived:

$$\boldsymbol{Y} = \boldsymbol{A}_{\text{SB}}\boldsymbol{\theta} + \boldsymbol{\epsilon} = \begin{pmatrix} 1 & T_1 & b_1 \\ 1 & T_2 & b_2 \\ \vdots & \vdots & \vdots \\ 1 & T_N & b_N \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \\ \eta \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{pmatrix} . \tag{2.23}$$

For later derivations this model is often written as follows:

$$\boldsymbol{Y} = \begin{pmatrix} 1 & T_1 \\ 1 & T_2 \\ \vdots & \vdots \\ 1 & T_N \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{pmatrix} \eta + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{pmatrix} = \boldsymbol{A}\boldsymbol{\theta} + \boldsymbol{B}_{\text{SB}}\,\eta + \boldsymbol{\epsilon}\,, \tag{2.24}$$

where the matrix $\boldsymbol{A} \in \mathbb{R}^{N\times 2}$ is the design matrix of the homogeneous population model (see Model (2.16)). The random vector $\boldsymbol{B}_{\text{SB}} \in \mathbb{R}^N$ defines a general bias vector for modelling selection bias and takes values in $\{-1, 0, 1\}^N$. The strength of selection bias is expressed by the parameter $\eta \in \mathbb{R}$.

2.3.3 Definition and model of chronological bias

The term chronological bias was first defined in Matts and McHugh (1978). A chronologic bias may arise due to systematic temporal changes in for example patient characteristics or changes of personnel. A long run of one treatment may be particularly problematic. In Tamm and Hilgers (2014) it is mentioned that the treatment effect can be biased if patients with a worse prognosis are included at the beginning of a clinical trial and patients with a better prognosis are included at the end of a clinical trial. In the literature

several randomization procedures like the Maximal Procedure (see Berger et al., 2003) and Hadamard Randomization (see Bailey and Nelson, 2003) are discussed for controlling the imbalances in patient numbers between the two treatment groups. The ICH E9 (1998) guideline explicitly recommends the usage of randomization in blocks like Permuted Block Randomization (see. Section 2.2.2) in situations when chronological bias is expected.

As pointed out in the introduction it is assumed that chronological bias appears in form of time trends. Below, two possible functions for modelling a time trend $\tau_\vartheta(n)$ dependent on the patient number n and the strength ϑ of the time trend are defined. The presented time trends are reviewed in Rosenkranz (2011) and Tamm and Hilgers (2014).

Definition 2.5: (Time trends)
Let $n \leq N$ be the number of the nth patient and $\vartheta \in \mathbb{R}$. Thus, the

a) Linear time trend is defined as:

$$\tau_{lin,\vartheta}(n) = \frac{n-1}{N}\,\vartheta\,. \tag{2.25}$$

b) Step trend is defined as:

$$\tau_{step,\vartheta}(n) = \mathbb{1}_{\{n \geq s\}}\,\vartheta\,, \tag{2.26}$$

where $\mathbb{1}_{\{n \geq s\}}$ is one if $n \geq s$ with $s \in \{1, \ldots, N\}$ and zero otherwise.

For practical applications a step trend represents the adaptation or relaxation of inclusion criteria in a study. Further applications of a step trend are a change of the attending physician or seasonal changes. The linear time trend represents a linear increase in the patients' characteristics.

Influence of chronological bias on the population model

In this section the influence of chronological bias on the homogeneous population model introduced in Section 2.3.1 is presented. The population model affected by chronological bias due to a time trend is shown in Figure 2.5. It is assumed that the expectation of the nth patient's response is shifted by a time trend $\tau_\vartheta(n)$. Examples of possible time trend functions $\tau_\vartheta(n)$ are given in Definition (2.5). The time trend is a fixed effect, which occurs additive to

the the treatment effect. The variance σ^2 is assumed to remain unaffected by the time trend function $\tau_\vartheta(n)$.

In this thesis unobserved time trends are investigated. That means the population model presented in Section 2.3.1 is assumed in the analysis, but in fact the population model affected by chronological bias in form of a time trend is present. Thus, in the later evaluation of this thesis the test problem described in Equation (2.13) is investigated under the assumption that the test decision is not adjusted for the additional time trend function $\tau_\vartheta(n)$. Due to the fact that the time trend is assumed to be unobserved, relatively small values for the time trend factor ϑ are considered.

Revisiting the regression model introduced in Equation (2.16), the situation of a population model affected by chronological bias in form of a linear time trend can be expressed by the function $a(T_n) = (1, T_n, \tau_{\text{lin},\vartheta}(n))$. Thus, for the design matrix it follows:

$$\boldsymbol{Y} = \boldsymbol{A}_{\text{CB}}\boldsymbol{\theta} + \boldsymbol{\epsilon} = \begin{pmatrix} 1 & T_1 & 0 \\ 1 & T_2 & {}^1\!/_N \\ \vdots & \vdots & \vdots \\ 1 & T_N & {}^{(N-1)}\!/_N \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vartheta \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{pmatrix}. \tag{2.27}$$

For later derivations this model is often transformed in the following manner:

$$\boldsymbol{Y} = \begin{pmatrix} 1 & T_1 \\ 1 & T_2 \\ \vdots & \vdots \\ 1 & T_N \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix} + \begin{pmatrix} 0 \\ {}^1\!/_N \\ \vdots \\ {}^{(N-1)}\!/_N \end{pmatrix} \vartheta + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{pmatrix} = \boldsymbol{A}\boldsymbol{\theta} + \boldsymbol{B}_{\text{CB}}\,\vartheta + \boldsymbol{\epsilon}\,, \tag{2.28}$$

where $\boldsymbol{B}_{\text{CB}} = (0, {}^1\!/_N, \ldots, {}^{(N-1)}\!/_N)^T$ defines a general bias vector for chronological bias dependent on the expected time trend function $\tau_{\text{lin},\vartheta}(n)$. The matrix $\boldsymbol{A} \in \mathbb{R}^{N\times 2}$ is again the design matrix of the homogeneous population model (see Model (2.16)). In case of a step time trend introduced in Equation (2.26) the vector $\boldsymbol{B}_{\text{CB}} \in \mathbb{R}^N$ consists of $s-1$ zeros followed by $N-s+1$ ones.

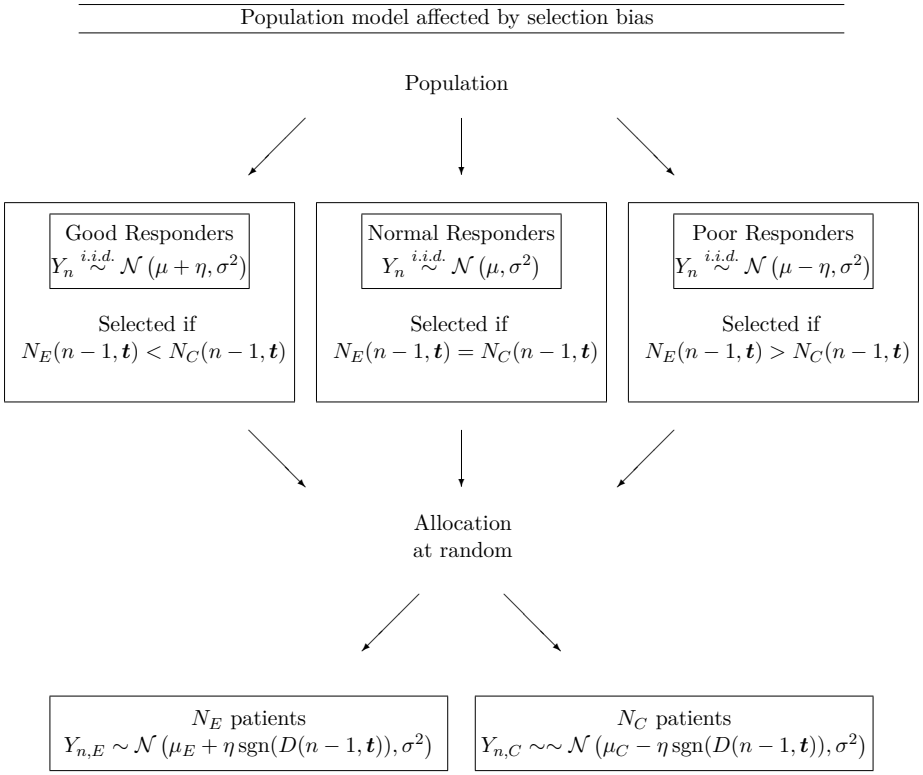


Figure (2.4): The population model affected by selection bias introduced by Proschan (1994).

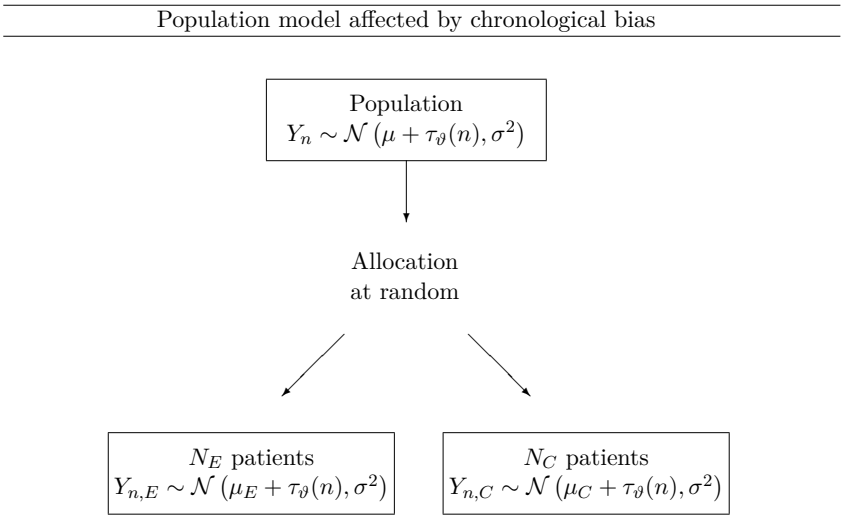


Figure (2.5): The population model affected by chronological bias in form of a a time trend $\tau_\vartheta(n)$.

Chapter 3

Measurement of selection and chronological bias

The derivations in this chapter are based on the homogeneous population model presented in Section 2.3.1. If not stated otherwise, a homogeneous population is assumed and the regression model defined in Equation (2.16) is investigated under misspecification for both a time trend parameter ϑ (see Model (2.27)) and a selection bias effect η (see Model (2.23)). In accordance with the ICH E9 (1998) guideline both an external time trend $\tau_\vartheta(n)$ and a possible selection effect η is understood as bias in a clinical trial. The ICH E9 guideline recommends that "the interpretation of statistical measures of uncertainty of [...] treatment comparisons should involve consideration of [...] bias to the p-value [...]." Thus, the influence of chronological and selection bias on the p-value of the test with the following hypotheses is investigated:

$$H_0 : \theta_1 = 0 \quad \text{vs.} \quad H_1 : \theta_1 \neq 0\,, \tag{3.1}$$

with θ_1 defined in Model (2.16). These hypotheses are equivalent to:

$$H_0 : \mu_E = \mu_C \quad \text{vs.} \quad H_1 : \mu_E \neq \mu_C\,, \tag{3.2}$$

where μ_E is the expectation of the effect associated with treatment E and μ_C the expectation of the effect associated with treatment C. In what follows, the errors in the models presented in Chapter 2 are assumed to

be independent and identically normally distributed. The null hypothesis presented in Equation (3.2) is tested with Student's t-test. At the beginning of this chapter the influence of a model affected by chronological bias presented in Equation (2.28) on the distorted test decision of Student's t-test is derived. Therefore the rejection probability under Student's t-test is investigated conditioned on the randomization sequence $\boldsymbol{t}$. In the situation of chronological bias it is assumed that the analysis is not adjusted for the additional term $\boldsymbol{B}_{\mathrm{CB}}\,\vartheta$. Afterwards, the results are transferred to the model affected by selection bias presented in Equation (2.24). In this particular case, the analysis is not adjusted for the additional term $\boldsymbol{B}_{\mathrm{SB}}\,\eta$. At the end of the chapter a more general bias model is investigated under the Random Allocation Rule. For this randomization procedure the properties of the central limit theorem for samples from a finite population, which was first introduced by Hajek (1960), are used to show that the nominal significance level $\alpha \in (0, 1)$ of Student's t-test is controlled asymptotically in case of a linear as well as a step time trend. Finally, asymptotical quantiles of the rejection probability under Student's t-test conditioned on the randomization sequence $\boldsymbol{t}$ are compared to exact/simulated quantiles under the Random Allocation Rule in case of a linear and a step time trend.

3.1 Influence of chronological bias on the test decision

In this section, the parameter θ_1 is tested to be zero, assuming Model (2.16). Thus, the influence of an unadjusted linear time trend on the test decision $H_0 : \theta_1 = 0$ against $H_1 : \theta_1 \neq 0$ is investigated. At the end of this section the derived results for a linear time trend are transferred to the situation in which the clinical trial is affected by a step time trend.

Revisiting Model (2.28), the design matrix has the following structure: $\boldsymbol{A} = (\mathbb{1}_N, \boldsymbol{T}) \in \mathbb{R}^{N\times 2}$, where $\mathbb{1}_N \in \mathbb{R}^N$ defines a vector containing the number one N times. The random vector $\boldsymbol{T} = (T_1, T_2, \ldots, T_N)^T$ takes values in $\{-1, 1\}^N$ and represents a two level factor for the two treatments E and C in a clinical trial. For the following calculations the case of $T_i = T_j$ for all $i \neq j$ with $i, j \in \{1, 2, \ldots, N\}$ is explicitly excluded. Furthermore, $N \geq 3$ and N being even are assumed. Consequently, the rank of the matrix

$\boldsymbol{A} \in \mathbb{R}^{N\times 2}$ is considered to be two (written: $rk(\boldsymbol{A}) = 2$) and it follows that the inverse $(\boldsymbol{A}^T\boldsymbol{A})^{-1}$ exists. In the analysis the model $\boldsymbol{Y} = \boldsymbol{A\theta} + \boldsymbol{\epsilon}$ (see Equation (2.16)) is wrongly assumed to be the correct one, although the model $\boldsymbol{Y} = \boldsymbol{A\theta} + \boldsymbol{B}_{\mathrm{CB}}\,\vartheta + \boldsymbol{\epsilon}$ (see Equation (2.28)) holds. The parameter θ_1 is tested to be zero. Hence, special emphasis is placed on the test statistic W with:

$$W := \frac{\hat{\theta}_1}{\sqrt{\widehat{Var}(\hat{\theta}_1)}}\,, \tag{3.3}$$

where $\hat{\theta}_1$ is the corresponding least square estimator for θ_1 and $\widehat{\mathrm{Var}}(\hat{\theta}_1)$ the corresponding estimator for the variance of $\hat{\theta}_1$. It is well known from the literature (see Fahrmeir et al., 2007, p. 120) that if the true parameter θ_1 is zero and the model $\boldsymbol{Y} = \boldsymbol{A\theta} + \boldsymbol{\epsilon}$ is correctly assumed with an independent and identically normally distributed random vector $\boldsymbol{\epsilon}$, that the test statistic W is t-distributed with $N - 2$ degrees of freedom.

Below, the influence of an unadjusted analysis for the bias vector $\boldsymbol{B}_{\mathrm{CB}}\,\vartheta$ on the distribution of the test statistic W is investigated, when the analysis is not adjusted for it. According to Fahrmeir et al. (2007, p. 92) the parameter θ_1 in the model $\boldsymbol{Y} = \boldsymbol{A\theta} + \boldsymbol{\epsilon}$ (see Equation (2.16)) can be estimated by the corresponding least-square estimator:

$$\hat{\theta}_1 = \boldsymbol{k}^T(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{Y}\,, \tag{3.4}$$

with $\boldsymbol{k} = (0,1)^T = \boldsymbol{e}_2 \in \mathbb{R}^2$, where $\boldsymbol{e}_i$ defines the unit vector with the number one on the position i. Before the distribution of $\hat{\theta}_1$ is derived, two more lemmas are given.

Lemma 3.1: (see Searle, 1971, pp. 180–181)
A function of parameters $\boldsymbol{k}^T\boldsymbol{\theta}$ in a linear model with $\boldsymbol{Y} = \boldsymbol{A\theta}+\boldsymbol{\epsilon}$ is estimable, if there is an $\boldsymbol{l} \in \mathbb{R}^N$ with $\boldsymbol{l}^T\boldsymbol{A} = \boldsymbol{k}^T$.

Lemma 3.2:
Let $\boldsymbol{A} \in \mathbb{R}^{N\times p}$ and $(\boldsymbol{A}^T\boldsymbol{A})^{-1}$ exists (which is equivalent to $rk(\boldsymbol{A}) = p$). Then, in the model $\boldsymbol{Y} = \boldsymbol{A\theta} + \boldsymbol{\epsilon}$ all combinations of $\boldsymbol{k}^T\boldsymbol{\theta}$ with $\boldsymbol{k} \in \mathbb{R}^p$ are estimable.

In a first step the distribution of the nominator of the test statistic W defined in Equation (3.3) is derived. The estimator $\hat{\theta}_1$ from Equation (3.4) can be transformed as follows:

$$\begin{aligned}\hat{\theta}_1 &= \boldsymbol{k}^T(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{Y} \\ &= \boldsymbol{l}^T\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{Y} \\ &= \boldsymbol{l}^T\omega(\boldsymbol{A})\boldsymbol{Y}\,. \end{aligned} \tag{3.5}$$

The vector $\boldsymbol{l} \in \mathbb{R}^N$ is chosen according to Lemma 3.1 and the vector is not explicitly specified. The matrix $\omega(\boldsymbol{A}) = \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ is the projection matrix on the column span of $\boldsymbol{A} \in \mathbb{R}^{N\times 2}$. The matrix $\omega(A)$ has the following properties (see Fahrmeir et al., 2007, p. 93):

1. $\omega(\boldsymbol{A})\,\omega(\boldsymbol{A}) = \omega(\boldsymbol{A})$
2. $\omega(\boldsymbol{A})^T = \omega(\boldsymbol{A})$
3. $\omega(\boldsymbol{A})\,\boldsymbol{A} = \boldsymbol{A}$

Analogously to the projection matrix $\omega(\boldsymbol{A})$, the corresponding orthogonal matrix $\omega^{\perp}(\boldsymbol{A}) = \boldsymbol{I}_{N\times N} - \omega(\boldsymbol{A})$, with identity matrix $\boldsymbol{I}_{N\times N} \in \mathbb{R}^{N\times N}$, is defined. For $\omega^{\perp}(\boldsymbol{A})$ the following properties hold:

4. $\omega^{\perp}(\boldsymbol{A})\,\boldsymbol{A} = \boldsymbol{0}$
5. $\omega(\boldsymbol{A})\,\omega^{\perp}(\boldsymbol{A}) = \boldsymbol{0}$.

The distribution of $\hat{\theta}_1 = \boldsymbol{l}^T\omega(\boldsymbol{A})\boldsymbol{Y}$ follows from a lemma found in Fahrmeir et al. (2007, p. 464).

Lemma 3.3:
Let $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\nu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\nu} \in \mathbb{R}^N$, $\boldsymbol{d} \in \mathbb{R}^q$, and $\boldsymbol{D} \in \mathbb{R}^{q\times N}$. The covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{N\times N}$ is assumed to be positive definite. Then, for $\boldsymbol{X} = \boldsymbol{d} + \boldsymbol{D}\boldsymbol{Y}$ it follows:

$$\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{d} + \boldsymbol{D}\boldsymbol{\nu}, \boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D}^T)\,.$$

Particularly, $\boldsymbol{Y} \in \mathbb{R}^N$ has the dimension N and $\boldsymbol{X} \in \mathbb{R}^q$ has the dimension q.

In what follows, the random vector $\boldsymbol{\epsilon}$ in Model (2.28) is assumed to be multivariate normally distributed with covariance matrix $\sigma^2\,\boldsymbol{I}_{N\times N} \in \mathbb{R}^{N\times N}$ and $\sigma \in \mathbb{R}_+$. For $\boldsymbol{Y} \in \mathbb{R}^N$ defined in Equation (2.28) it holds $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{A\theta} + \boldsymbol{B}_{\mathrm{CB}}\,\vartheta, \sigma^2\,\boldsymbol{I}_{N\times N})$. Thus, applying Lemma 3.3 with $\boldsymbol{D} = \boldsymbol{l}^T\omega(\boldsymbol{A}) \in \mathbb{R}^{1\times N}$ and $\boldsymbol{d} = \boldsymbol{0} \in \mathbb{R}^N$ the distribution of $\hat{\theta}_1$ can be transformed as follows:

$$\begin{aligned}
\boldsymbol{l}^T\omega(\boldsymbol{A})\boldsymbol{Y} &\sim \mathcal{N}\left(\boldsymbol{l}^T\omega(\boldsymbol{A})\,(\boldsymbol{A\theta} + \boldsymbol{B}_{\mathrm{CB}}\,\vartheta), \boldsymbol{l}^T\omega(\boldsymbol{A})\,\sigma^2\,\boldsymbol{I}_{N\times N}\,(\boldsymbol{l}^T\omega(\boldsymbol{A}))^T\right) \\
&= \mathcal{N}\left(\boldsymbol{l}^T\omega(\boldsymbol{A})\boldsymbol{A\theta} + \boldsymbol{l}^T\omega(\boldsymbol{A})\boldsymbol{B}_{\mathrm{CB}}\,\vartheta, \sigma^2\,\boldsymbol{l}^T\omega(\boldsymbol{A})\omega(\boldsymbol{A})\boldsymbol{l}\right) \\
&= \mathcal{N}\left(\boldsymbol{l}^T\boldsymbol{A\theta} + \tilde{\delta}, \sigma^2\,\boldsymbol{l}^T\omega(\boldsymbol{A})\boldsymbol{l}\right) \\
&= \mathcal{N}(\boldsymbol{k}^T\boldsymbol{\theta} + \tilde{\delta}, \sigma^2\,\boldsymbol{l}^T\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{l}) \\
&= \mathcal{N}\left(\theta_1 + \tilde{\delta}, \sigma^2\,\boldsymbol{k}^T(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{k}\right), \qquad (3.6)
\end{aligned}$$

with $\tilde{\delta} = \boldsymbol{l}^T\omega(\boldsymbol{A})\boldsymbol{B}_{\mathrm{CB}}\,\vartheta$. Consequently, for $\hat{\theta}_1 = \boldsymbol{l}^T\omega(\boldsymbol{A})\boldsymbol{Y}$ according to Equation (3.5) it follows:

$$\frac{\hat{\theta}_1}{\sqrt{\sigma^2\,\boldsymbol{k}^T(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{k}}} \sim \mathcal{N}(\delta, 1)\,, \qquad (3.7)$$

with $\delta = {}^{(\theta_1+\tilde{\delta})}\!/\!_{(\sqrt{\sigma^2\,\boldsymbol{k}^T(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{k}})} = {}^{(\theta_1+\boldsymbol{l}^T\omega(\boldsymbol{A})\boldsymbol{B}_{\mathrm{CB}}\,\vartheta)}\!/\!_{(\sqrt{\sigma^2\,\boldsymbol{k}^T(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{k}})}$ and $\boldsymbol{l}^T\boldsymbol{A} = \boldsymbol{k}^T = (0, 1)$. Now, the parameter δ in the case that $\boldsymbol{B}_{\mathrm{CB}} \in \mathbb{R}^N$ is a (random) vector is derived. Model (2.28) is transformed for the calculation of $\omega(\boldsymbol{A})$, in the following manner:

$$\begin{aligned}
\boldsymbol{Y} = \boldsymbol{A\theta} + \boldsymbol{B}_{\mathrm{CB}}\,\vartheta + \boldsymbol{\epsilon} \quad \Leftrightarrow \quad \tilde{\boldsymbol{Y}} &= \tilde{\boldsymbol{A}}\boldsymbol{\theta} + \tilde{\boldsymbol{B}}_{\mathrm{CB}}\,\vartheta + \boldsymbol{\epsilon} \qquad (3.8) \\
\text{with } \tilde{\boldsymbol{A}} &= \begin{pmatrix} \mathbb{1}_{N_E} & \mathbb{1}_{N_E} \\ \mathbb{1}_{N_C} & -\mathbb{1}_{N_C} \end{pmatrix},
\end{aligned}$$

with $N_E := {}^1\!/\!_2 \sum_{i=1}^{N}(T_i + 1)$ and $N_C := N - N_E$. Thus, $\tilde{\boldsymbol{Y}} \in \mathbb{R}^N$ is an ordered vector of $\boldsymbol{Y} \in \mathbb{R}^N$ such that at the beginning of the vector all entries Y_i with $T_i = 1$ are appearing with its corresponding bias element $b_i \in \boldsymbol{B}_{\mathrm{CB}}$ for $i \in \{1, \ldots, N\}$. For $\tilde{\boldsymbol{A}}$ the vector $\boldsymbol{l} = {}^1\!/\!_2\,(\boldsymbol{e}_1 - \boldsymbol{e}_{N_E+1}) \in \mathbb{R}^N$ with the property $\boldsymbol{l}^T\tilde{\boldsymbol{A}} = (0, 1) = \boldsymbol{k}^T$ is assumed to be fixed. For $\omega(\tilde{\boldsymbol{A}})$ it follows:

$$\omega(\tilde{\boldsymbol{A}}) = \tilde{\boldsymbol{A}}(\tilde{\boldsymbol{A}}^T\tilde{\boldsymbol{A}})^{-1}\tilde{\boldsymbol{A}}^T$$

$$
\begin{aligned}
&= \tilde{\boldsymbol{A}} \begin{pmatrix} N & \underbrace{N_E - N_C}_{:=d} \\ N_E - N_C & N \end{pmatrix}^{-1} \tilde{\boldsymbol{A}}^T \\
&= \frac{1}{N^2 - d^2} \tilde{\boldsymbol{A}} \begin{pmatrix} N & -d \\ -d & N \end{pmatrix} \tilde{\boldsymbol{A}}^T \\
&= \frac{1}{N^2 - d^2} \tilde{\boldsymbol{A}} \begin{pmatrix} (N-d)\mathbb{1}_{N_E}^T & (N+d)\mathbb{1}_{N_C}^T \\ (N-d)\mathbb{1}_{N_E}^T & (-N-d)\mathbb{1}_{N_C}^T \end{pmatrix} \\
&= \frac{1}{N^2 - d^2} \begin{pmatrix} \mathbb{1}_{N_E \times N_E}\, 2\,(N-d) & \mathbf{0} \\ \mathbf{0} & \mathbb{1}_{N_C \times N_C}\, 2\,(N+d) \end{pmatrix} \\
&= \begin{pmatrix} \mathbb{1}_{N_E \times N_E}\, \frac{2}{(N+d)} & \mathbf{0} \\ \mathbf{0} & \mathbb{1}_{N_C \times N_C}\, \frac{2}{(N-d)} \end{pmatrix} \\
&\overset{(1)}{=} \begin{pmatrix} \mathbb{1}_{N_E \times N_E}\, \frac{1}{N_E} & \mathbf{0} \\ \mathbf{0} & \mathbb{1}_{N_C \times N_C}\, \frac{1}{N_C} \end{pmatrix}. \qquad (3.9)
\end{aligned}
$$

The matrix $\omega(\tilde{\boldsymbol{A}})$ depends on $\boldsymbol{T}$ only in the number of times assigning the treatment E or C, respectively. Below, the parameter δ from Equation (3.7) with $\boldsymbol{l} = 1/2\,(\boldsymbol{e}_1 - \boldsymbol{e}_{N_E+1}) \in \mathbb{R}^N$ and $\tilde{\boldsymbol{B}}_{\mathrm{CB}} = (\tilde{b}_1, \tilde{b}, \ldots, \tilde{b}_N)^T \in \mathbb{R}^N$ is derived:

$$
\begin{aligned}
\delta &= \frac{\theta_1 + \boldsymbol{l}^T \omega(\tilde{\boldsymbol{A}}) \tilde{\boldsymbol{B}}_{\mathrm{CB}}\, \vartheta}{\sqrt{\sigma^2\, \boldsymbol{l}^T \omega(\tilde{\boldsymbol{A}}) \boldsymbol{l}}} \\
&= \frac{1}{\sqrt{\frac{\sigma^2}{2}\left(\frac{1}{N_E}\mathbb{1}_{N_E}^T, -\frac{1}{N_C}\mathbb{1}_{N_C}^T\right)\boldsymbol{l}}} \left(\theta_1 + \left(\frac{1}{N_E}\mathbb{1}_{N_E}^T, -\frac{1}{N_C}\mathbb{1}_{N_C}^T\right) \tilde{\boldsymbol{B}}_{\mathrm{CB}} \frac{\vartheta}{2}\right) \\
&= \frac{\vartheta}{2\sqrt{\frac{\sigma^2}{4}\left(\frac{1}{N_E} + \frac{1}{N_C}\right)}} \Bigg(\frac{2\theta_1}{\vartheta} + \underbrace{\frac{1}{N_E} \sum_{i=1}^{N_E} \tilde{b}_i}_{:=\tilde{\bar{B}}_E} - \underbrace{\frac{1}{N_C} \sum_{i=N_E+1}^{N} \tilde{b}_i}_{:=\tilde{\bar{B}}_C} \Bigg) \\
&= \frac{\vartheta}{\sqrt{\sigma^2\left(\frac{N_E+N_C}{N_E N_C}\right)}} \left(\frac{2\theta_1}{\vartheta} + \tilde{\bar{B}}_E - \tilde{\bar{B}}_C \right) \\
&= \frac{\vartheta}{\sigma} \sqrt{\frac{N_E N_C}{N_E + N_C}} \left(\frac{2\theta_1}{\vartheta} + \bar{B}_E - \bar{B}_C \right). \qquad (3.10)
\end{aligned}
$$

The term $\bar{B}_j$ with $j \in \{E, C\}$ is the mean value of the elements of the bias vector assigned to the corresponding treatment group j dependent

(1) $N + d = N_E + N_C + N_E - N_C = 2N_E$ and $N - d = N_E + N_C - (N_E - N_C) = 2N_C$.

on the randomization sequence. This mean value is independent of the transformation used in Equation (3.8). Under the assumption $H_0 : \theta_1 = 0$ it follows that the noncentrality parameter δ is zero, if and only if $\bar{B}_E = \bar{B}_C$ holds - otherwise $\delta \neq 0$. In summary, δ is shifted dependent on the parameter θ_1 as well as the allocation of the corresponding elements of the bias vector to the two treatment groups.

In Fahrmeir et al. (2007, p. 102) an unbiased estimator for the denominator $\widehat{\mathrm{Var}}(\hat{\theta}_1)$ according to Equation (3.3) is given:

$$\begin{aligned}
\widehat{\mathrm{Var}}(\hat{\theta}_1) &= \hat{\sigma}^2 \, \boldsymbol{k}^T (\boldsymbol{A}^T \boldsymbol{A})^{-1} \boldsymbol{k} \\
&= \frac{1}{N-r} \, \boldsymbol{Y}^T (\boldsymbol{I}_{N\times N} - \omega(\boldsymbol{A})) \boldsymbol{Y} \boldsymbol{k}^T (\boldsymbol{A}^T \boldsymbol{A})^{-1} \boldsymbol{k} \\
&= \frac{1}{N-r} \, \sigma^2 \, \boldsymbol{k}^T (\boldsymbol{A}^T \boldsymbol{A})^{-1} \, \boldsymbol{k} \boldsymbol{Y}^T \, \frac{1}{\sigma^2} \, \omega^{\perp}(\boldsymbol{A}) \boldsymbol{Y} \, , \qquad (3.11)
\end{aligned}$$

with $\boldsymbol{k} = \boldsymbol{e}_2 = (0,1)^T$. The term r defines the rank of the matrix $\boldsymbol{A} \in \mathbb{R}^{N\times 2}$, which is two under the assumption of Model (2.16). For the derivation of the distribution of $\widehat{\mathrm{Var}}(\hat{\theta}_1)$ the following lemma from Koch (2004, p. 146) is applied.

Lemma 3.4:
Let $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\nu}, \sigma^2 \boldsymbol{I}_{N\times N})$, thus if $\boldsymbol{D}\,\sigma^2$ is an idempotent matrix then $\boldsymbol{X}^T \boldsymbol{D} \boldsymbol{X} \sim \chi^2(h, \boldsymbol{\nu}^T \boldsymbol{D} \boldsymbol{\nu})$ with $h = rk(\boldsymbol{D})$, where $\chi^2(h,\lambda)$ defines the χ^2-distribution with h degrees of freedom and noncentrality parameter λ.

Below, the distribution of the last term $\boldsymbol{Y}^T \, {}^1\!/\!_{\sigma^2} \, \omega^{\perp}(\boldsymbol{A}) \boldsymbol{Y}$ defined in Equation (3.11) is derived. For this, Lemma 3.4 with $\boldsymbol{D} = {}^1\!/\!_{\sigma^2} \, \omega^{\perp}(\boldsymbol{A})$ is applied. The matrix $\boldsymbol{D}\,\sigma^2 = \omega^{\perp}(\boldsymbol{A})$ is idempotent. Furthermore, $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{\theta} + \boldsymbol{B}_{\mathrm{CB}}\,\vartheta, \sigma^2 \boldsymbol{I}_{N\times N})$ with $\boldsymbol{\nu} = \boldsymbol{A}\boldsymbol{\theta} + \boldsymbol{B}_{\mathrm{CB}}\,\vartheta$ is considered. Thus, it follows:

$$\begin{aligned}
\boldsymbol{Y}^T \, \frac{1}{\sigma^2} \, \omega^{\perp}(\boldsymbol{A}) \boldsymbol{Y} &\sim \chi^2 \left(h, (\boldsymbol{A}\boldsymbol{\theta} + \boldsymbol{B}_{\mathrm{CB}}\,\vartheta)^T \, \frac{1}{\sigma^2} \, \omega^{\perp}(\boldsymbol{A}) \, (\boldsymbol{A}\boldsymbol{\theta} + \boldsymbol{B}_{\mathrm{CB}}\,\vartheta) \right) \\
&= \chi^2 \Big(h, \frac{1}{\sigma^2} (\boldsymbol{\theta}^T \underbrace{\boldsymbol{A}^T \omega^{\perp}(\boldsymbol{A})}_{=(\omega^{\perp}(\boldsymbol{A})\boldsymbol{A})^T = \boldsymbol{0}} + \vartheta \, \boldsymbol{B}_{\mathrm{CB}}^T \, \omega^{\perp}(\boldsymbol{A})) \, (\boldsymbol{A}\boldsymbol{\theta} + \boldsymbol{B}_{\mathrm{CB}}\,\vartheta) \Big) \\
&= \chi^2 \left(h, \frac{1}{\sigma^2} \left(\vartheta \, \boldsymbol{B}_{\mathrm{CB}}^T \, \omega^{\perp}(\boldsymbol{A}) \boldsymbol{B}_{\mathrm{CB}} \, \vartheta \right) \right) \\
&= \chi^2 \left(h, \frac{\vartheta^2}{\sigma^2} \left(\boldsymbol{B}_{\mathrm{CB}}^T \, \omega^{\perp}(\boldsymbol{A}) \boldsymbol{B}_{\mathrm{CB}} \right) \right) \, , \qquad (3.12)
\end{aligned}$$

with $h = rk(\omega^{\perp}(\boldsymbol{A})) = N - rk(\omega(\boldsymbol{A})) = N - 2$. Finally, the noncentrality parameter $\lambda = {}^{\vartheta^2}/_{\sigma^2}\,(\boldsymbol{B}_{\mathrm{CB}}^T\,\omega^{\perp}(\boldsymbol{A})$ is simplified. The situation with $\boldsymbol{A} = (\mathbb{1}_N, \boldsymbol{T}) \in \mathbb{R}^{N\times 2}$ and $\boldsymbol{B}_{\mathrm{CB}} = (b_1, b_2, \ldots, b_N)^T \in \mathbb{R}^N$ is assumed. For the following calculations the transformed Model (3.8) and the equality $\omega^{\perp}(\tilde{\boldsymbol{A}}) = \boldsymbol{I}_{N\times N} - \omega(\tilde{\boldsymbol{A}})$ are used. The matrix $\omega(\tilde{\boldsymbol{A}})$ is derived in Equation (3.9). Now, $\lambda = {}^{\vartheta^2}/_{\sigma^2}\,(\tilde{\boldsymbol{B}}_{\mathrm{CB}}^T\,\omega^{\perp}(\tilde{\boldsymbol{A}})\tilde{\boldsymbol{B}}_{\mathrm{CB}})$ is transformed:

$$
\begin{aligned}
\lambda &= \frac{\vartheta^2}{\sigma^2}\left(\tilde{\boldsymbol{B}}_{\mathrm{CB}}^T\boldsymbol{I}_{N\times N} - \tilde{\boldsymbol{B}}_{\mathrm{CB}}^T\begin{pmatrix}\mathbb{1}_{N_E\times N_E}\,\frac{1}{N_E} & \boldsymbol{0}\\ \boldsymbol{0} & \mathbb{1}_{N_C\times N_C}\,\frac{1}{N_C}\end{pmatrix}\right)\tilde{\boldsymbol{B}}_{\mathrm{CB}}\\
&= \frac{\vartheta^2}{\sigma^2}\left(\tilde{\boldsymbol{B}}_{\mathrm{CB}}^T - \left(\mathbb{1}_{N_E}^T\frac{1}{N_E}\sum_{i=1}^{N_E}\tilde{b}_i, \mathbb{1}_{N_C}^T\frac{1}{N_C}\sum_{i=N_E+1}^{N}\tilde{b}_i\right)\right)\tilde{\boldsymbol{B}}_{\mathrm{CB}}\\
&= \frac{\vartheta}{\sigma^2}\left(\tilde{\boldsymbol{B}}_{\mathrm{CB}}^T\,\tilde{\boldsymbol{B}}_{\mathrm{CB}} - \left(\mathbb{1}_{N_E}^T\tilde{\bar{B}}_E, \mathbb{1}_{N_C}^T\tilde{\bar{B}}_C\right)\tilde{\boldsymbol{B}}_{\mathrm{CB}}\right)\\
&= \frac{\vartheta^2}{\sigma^2}\left(\sum_{i=1}^{N}\tilde{b}_i^2 - \tilde{\bar{B}}_E\sum_{i=1}^{N_E}\tilde{b}_i - \tilde{\bar{B}}_C\sum_{i=N_E+1}^{N}\tilde{b}_i\right)\\
&= \frac{\vartheta^2}{\sigma^2}\left(\sum_{i=1}^{N_E}\tilde{b}_i^2 - \tilde{\bar{B}}_E\sum_{i=1}^{N_E}\tilde{b}_i + \sum_{i=N_E+1}^{N}\tilde{b}_i^2 - \tilde{\bar{B}}_C\sum_{i=N_E+1}^{N}\tilde{b}_i\right)\\
&= \frac{\vartheta^2}{\sigma^2}\left(\sum_{i=1}^{N_E}\left(\tilde{b}_i^2 - 2\,\tilde{\bar{B}}_E\tilde{b}_i + \tilde{\bar{B}}_E^2\right) + \sum_{i=N_E+1}^{N}\left(\tilde{b}_i^2 - 2\,\tilde{\bar{B}}_C\tilde{b}_i + \tilde{\bar{B}}_C^2\right)\right)\\
&= \frac{\vartheta^2}{\sigma^2}\left(\sum_{i=1}^{N_E}\left(\tilde{b}_i - \tilde{\bar{B}}_E\right)^2 + \sum_{i=N_E+1}^{N}\left(\tilde{b}_i - \tilde{\bar{B}}_C\right)^2\right)\\
&= \frac{\vartheta^2}{\sigma^2}\left((N_E-1)\tilde{S}_{E,\boldsymbol{b}}^2 + (N_C-1)\tilde{S}_{C,\boldsymbol{b}}^2\right)\\
&= \frac{\vartheta^2}{\sigma^2}\left((N_E-1)\,S_{E,\boldsymbol{b}}^2 + (N_C-1)\,S_{C,\boldsymbol{b}}^2\right), \qquad (3.13)
\end{aligned}
$$

where $S_{E,\boldsymbol{b}}^2 = {}^1/_{(N_E-1)}\sum_{i=1}^{N}\mathbb{1}_{\{T_i=1\}}\left(b_i - {}^1/_{N_E}\sum_{j=1}^{N}b_j\,\mathbb{1}_{\{T_j=1\}}\right)^2$ is the variance of the elements of the bias vector allocated to treatment group E and $S_{C,\boldsymbol{b}}^2 = {}^1/_{(N_C-1)}\sum_{i=1}^{N}\mathbb{1}_{\{T_i=-1\}}\left(b_i - {}^1/_{N_C}\sum_{j=1}^{N}b_j\,\mathbb{1}_{\{T_j=-1\}}\right)^2$ is the variance of the elements of the bias vector allocated to treatment group C. Although the noncentrality parameter λ is independent of the transformation of the design matrix $\boldsymbol{A} \in \mathbb{R}^{N\times 2}$ used in Equation (3.8), the noncentrality parameter λ is strongly dependent on $\boldsymbol{T}$. The noncentrality parameter λ is zero, if and only if the variance of the elements of the additional bias term $\boldsymbol{B}_{\mathrm{CB}} \in \mathbb{R}^N$ in both treatment groups E and C is zero. In particular, the noncentrality

parameter λ is zero if all the patients' responses of any given group are shifted by a constant – otherwise it is not. Finally, the results are summarized and the distribution of the test statistic W is derived:

$$\begin{aligned} W &:= \frac{\hat{\theta}_1}{\sqrt{\widehat{\mathrm{Var}}(\hat{\theta}_1)}} \\ &= \frac{\frac{\hat{\theta}_1}{\sqrt{\sigma^2\, \boldsymbol{k}^T (\boldsymbol{A}^T \boldsymbol{A})^{-1} \boldsymbol{k}}}}{\sqrt{\boldsymbol{Y}^T \frac{1}{\sigma^2}\, \omega^{\perp}(\boldsymbol{A}) \boldsymbol{Y} \frac{1}{N-2}}} \\ &= \frac{\frac{1}{2\,\sigma} \sqrt{\frac{N_E\, N_C}{N_E + N_C}}\, \hat{\theta}_1}{\sqrt{\boldsymbol{Y}^T \frac{1}{\sigma^2}\, \omega^{\perp}(\boldsymbol{A}) \boldsymbol{Y} \frac{1}{N-2}}} \sim t_{N-r,\delta,\lambda}\,, \end{aligned} \tag{3.14}$$

with $\sigma^{-1} \sqrt{(N_E - N_C)/(N_E N_C)}\, \hat{\theta}_1 \sim \mathcal{N}(\delta, 1)$ and $\sqrt{\boldsymbol{Y}^T \sigma^{-2}\, \omega^{\perp}(\boldsymbol{A}) \boldsymbol{Y}\, {}^1/_{(N-2)}} \sim \chi^2_{N-2,\lambda}$. In Equation (3.10) $\delta = {}^{\vartheta}/_{\sigma} \sqrt{(N_E N_C)/(N_E + N_C)} \left({}^{(2\,\theta_1)}/_{\vartheta} + \bar{B}_E - \bar{B}_C \right)$ and in Equation (3.13) $\lambda = {}^{\vartheta^2}/_{\sigma^2} \left((N_E - 1)\, S^2_{E,\boldsymbol{b}} + (N_C - 1)\, S^2_{C,\boldsymbol{b}} \right)$ with $N_E := {}^1/_2 \sum_{i=1}^{N} (T_i + 1)$ and $N_C := N - N_E$ are derived. After Searle (1971, p. 99) the nominator and denominator are independently distributed. Thus, W follows a doubly noncentral t-distribution $t_{N-2,\delta,\lambda}$ with $N - 2$ degrees of freedom and noncentrality parameters δ and λ. Both noncentrality parameters are dependent on $\boldsymbol{T}$. The properties of the doubly noncentral t-distribution are discussed in more detail in Section 3.1.2. Below, the correlation between the sum of the investigated noncentrality parameters δ and λ is more closely investigated. Under the assumption of the model $\boldsymbol{Y} = \boldsymbol{A\theta} + \boldsymbol{\epsilon}$ defined in Equation (2.16) and $H_0 : \theta_1 = 0$ the test statistic W is t-distributed. The test statistic W is noncentral t-distributed if $H_1 : \theta_1 \neq 0$ and Model (2.16) holds. Particularly, in this situation the noncentrality parameter λ is zero and only the noncentrality parameter δ is shifted.

3.1.1 Properties of noncentrality parameters

Proposition 3.5:
Let $\boldsymbol{Y} = \boldsymbol{A\theta} + \boldsymbol{B}_{CB}\, \vartheta + \boldsymbol{\epsilon}$ be the model defined in Equation (2.28) and $\theta_1 = 0$. Then, for the noncentrality parameters $\delta = {}^{\vartheta}/_{\sigma} \sqrt{(N_E N_C)/(N_E + N_C)} \left(\bar{B}_E - \bar{B}_C \right)$ and $\lambda = {}^{\vartheta^2}/_{\sigma^2} \left((N_E - 1)\, S^2_{E,\boldsymbol{b}} + (N_C - 1)\, S^2_{C,\boldsymbol{b}} \right)$ with $N_E = \sum_{i=1}^{N} {}^{(T_i+1)}/_2$ and

$N_C = N - N_E$ the following equality holds:

$$\delta^2 + \lambda = \frac{\vartheta^2}{\sigma^2}\,(N-1)\,S^2_{\boldsymbol{B}_{CB}}\,, \tag{3.15}$$

with $S^2_{\boldsymbol{B}_{CB}} = {}^1\!/_{(N-1)} \sum_{i=1}^{N} (b_i - \bar{B}_{CB})^2$ *and* $\bar{B}_{CB} = {}^1\!/_N \sum_{i=1}^{N} b_i$.

Proof

$$\begin{aligned}
\delta^2 + \lambda &= \frac{\vartheta^2}{\sigma^2}\,\frac{N_E\,N_C}{N_E + N_C}\left(\bar{B}_E - \bar{B}_C\right)^2 \\
&\quad + \frac{\vartheta^2}{\sigma^2}\left((N_E-1)\,S^2_{E,\boldsymbol{b}} + (N_C-1)\,S^2_{C,\boldsymbol{b}}\right) \\
&= \frac{\vartheta^2}{\sigma^2}\left(\frac{N_E\,N_C}{N}\bar{B}_E^2 - 2\,\frac{N_E\,N_C}{N}\,\bar{B}_E\,\bar{B}_C + \frac{N_E\,N_C}{N}\bar{B}_C^2\right. \\
&\quad \left. + \sum_{i=1}^{N_E}\left(\tilde{b}_i - \bar{B}_E\right)^2 + \sum_{i=N_E+1}^{N}\left(\tilde{b}_i - \bar{B}_C\right)^2\right) \\
&= \frac{\vartheta^2}{\sigma^2}\left(\frac{N_C}{N\,N_E}\left(\sum_{i=1}^{N_E}\tilde{b}_i\right)^2 - \frac{2}{N}\sum_{i=1}^{N_E}\tilde{b}_i\sum_{i=N_E+1}^{N}\tilde{b}_i + \frac{N_E}{N\,N_C}\left(\sum_{i=N_E+1}^{N}\tilde{b}_i\right)^2\right. \\
&\quad + \sum_{i=1}^{N_E}\tilde{b}_i^2 - 2\frac{1}{N_E}\sum_{i=1}^{N_E}\tilde{b}_i\sum_{i=1}^{N_E}\tilde{b}_i + \frac{1}{N_E}\left(\sum_{i=1}^{N_E}\tilde{b}_i\right)^2 \\
&\quad \left. + \sum_{i=N_E+1}^{N}\tilde{b}_i^2 - 2\frac{1}{N_C}\sum_{i=N_E+1}^{N}\tilde{b}_i\sum_{i=N_E+1}^{N}\tilde{b}_i + \frac{1}{N_C}\left(\sum_{i=N_E+1}^{N}\tilde{b}_i\right)^2\right) \\
&= \frac{\vartheta^2}{\sigma^2}\left(\frac{N_C}{N\,N_E}\left(\sum_{i=1}^{N_E}\tilde{b}_i\right)^2 - \frac{2}{N}\sum_{i=1}^{N_E}\tilde{b}_i\sum_{i=N_E+1}^{N}\tilde{b}_i + \frac{N_E}{N\,N_C}\left(\sum_{i=N_E+1}^{N}\tilde{b}_i\right)^2\right. \\
&\quad \left. + \sum_{i=1}^{N_E}\tilde{b}_i^2 - \frac{1}{N_E}\left(\sum_{i=1}^{N_E}\tilde{b}_i\right)^2 + \sum_{i=N_E+1}^{N}\tilde{b}_i^2 - \frac{1}{N_C}\left(\sum_{i=N_E+1}^{N}\tilde{b}_i\right)^2\right).
\end{aligned}$$

The ordered elements $\tilde{b}_i$ with $i \in \{1, 2, \ldots, N\}$ of the bias vector $\boldsymbol{B}_{\mathrm{CB}} \in \mathbb{R}^N$ are dependent on $\boldsymbol{T}$ (see Equation (3.8)). With some further transformations the proposition follows:

$$\delta^2 + \lambda = \frac{\vartheta^2}{\sigma^2}\left(\sum_{i=1}^{N}\tilde{b}_i^2 - \frac{1}{N}\left(-\frac{N_C}{N_E}\left(\sum_{i=1}^{N_E}\tilde{b}_i\right)^2 + 2\sum_{i=1}^{N_E}\tilde{b}_i\sum_{i=N_E+1}^{N}\tilde{b}_i\right.\right.$$

$$
\begin{aligned}
&- \frac{N_E}{N_C}\left(\sum_{i=N_E+1}^{N}\tilde{b}_i\right)^2 + \frac{N}{N_E}\left(\sum_{i=1}^{N_E}\tilde{b}_i\right)^2 + \frac{N}{N_C}\left(\sum_{i=N_E+1}^{N}\tilde{b}_i\right)^2\Bigg)\Bigg) \\
&= \frac{\vartheta^2}{\sigma^2}\left(\sum_{i=1}^{N}\tilde{b}_i^2 - \frac{1}{N}\left(\left(\sum_{i=1}^{N_E}\tilde{b}_i\right)^2\left(\frac{N_E+N_C}{N_E} - \frac{N_C}{N_E}\right) + 2\sum_{i=1}^{N_E}\tilde{b}_i\sum_{i=N_E+1}^{N}\tilde{b}_i\right.\right. \\
&+ \left.\left.\left(\sum_{i=N_E+1}^{N}\tilde{b}_i\right)^2\left(\frac{N_E+N_C}{N_C} - \frac{N_E}{N_C}\right)\right)\right) \\
&= \frac{\vartheta^2}{\sigma^2}\left(\sum_{i=1}^{N}\tilde{b}_i^2 - \frac{1}{N}\left(\left(\sum_{i=1}^{N_E}\tilde{b}_i\right)^2 + 2\sum_{i=1}^{N_E}\tilde{b}_i\sum_{i=N_E+1}^{N}\tilde{b}_i + \left(\sum_{i=N_E+1}^{N}\tilde{b}_i\right)^2\right)\right) \\
&= \frac{\vartheta^2}{\sigma^2}\left(\sum_{i=1}^{N}\tilde{b}_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N_E}\tilde{b}_i + \sum_{i=N_E+1}^{N}\tilde{b}_i\right)^2\right) \\
&= \frac{\vartheta^2}{\sigma^2}\left(\sum_{i=1}^{N}\tilde{b}_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N}\tilde{b}_i\right)^2\right) \\
&= \frac{\vartheta^2}{\sigma^2}\left(\sum_{i=1}^{N}b_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N}b_i\right)^2\right) \\
&= \frac{\vartheta^2}{\sigma^2}\left(\sum_{i=1}^{N}b_i^2 - \frac{2}{N}\left(\sum_{i=1}^{N}b_i\right)^2 + \frac{1}{N}\left(\sum_{i=1}^{N}b_i\right)^2\right) \\
&= \frac{\vartheta^2}{\sigma^2}\left(\sum_{i=1}^{N}b_i^2 - 2\,\bar{B}_{\mathrm{CB}}\sum_{i=1}^{N}b_i + N\,\bar{B}_{\mathrm{CB}}^2\right) \\
&= \frac{\vartheta^2}{\sigma^2}\sum_{i=1}^{N}\left(b_i - \bar{B}_{\mathrm{CB}}\right)^2 \\
&= \frac{\vartheta^2}{\sigma^2}\,(N-1)\,S^2_{\boldsymbol{B}_{\mathrm{CB}}}\,.
\end{aligned}
$$

□

The result derived in Proposition 3.5 can be applied directly to the distribution of the test statistic W from Equation (3.14). Under the assumptions of $H_0 : \theta_1 = 0$ and a fixed bias vector $\boldsymbol{B}_{\mathrm{CB}} \in \mathbb{R}^N$, it follows that the sum $\delta^2 + \lambda$ is independent of $\boldsymbol{T}$. However, δ and λ themselves are strongly dependent on $\boldsymbol{T}$.

Below, the noncentrality parameter δ derived in Equation (3.10) is transformed. Therefore, the true difference between the effects of the two treat-

ments is considered to be an additional element of the bias vector $\boldsymbol{B}_{\text{CB}} \in \mathbb{R}^N$:

$$\begin{aligned}
\delta_\star &= \frac{\vartheta}{\sigma}\sqrt{\frac{N_E\,N_C}{N_E+N_C}}\left(\frac{2\,\theta_1}{\vartheta} + \bar{B}_E - \bar{B}_C\right) \\
&= \frac{\vartheta}{\sigma}\sqrt{\frac{N_E\,N_C}{N_E+N_C}}\left(\left(\bar{B}_E + \frac{\theta_1}{\vartheta}\right) - \left(\bar{B}_C - \frac{\theta_1}{\vartheta}\right)\right) \\
&= \frac{\vartheta}{\sigma}\sqrt{\frac{N_E\,N_C}{N_E+N_C}}\left(\bar{B}^\star_E - \bar{B}^\star_C\right), \qquad (3.16)
\end{aligned}$$

with $\bar{B}^\star_E = {}^1\!/_{N_E}\sum_{i=1}^{N_E}\left(\tilde{b}_i + {}^{\theta_1}\!/_{\vartheta}\right)$ and $\bar{B}^\star_C = {}^1\!/_{N_C}\sum_{i=N_E+1}^{N}\left(\tilde{b}_i - {}^{\theta_1}\!/_{\vartheta}\right)$, where $\tilde{b}_i$ are the elements of the ordered vector $\tilde{\boldsymbol{B}}_{\text{CB}}$ dependent on $\boldsymbol{T}$ (see Equation (3.8)). Analogously to Proposition 3.5 the following equality holds for $\delta_\star = {}^{\vartheta}\!/_{\sigma\sqrt{{}^{(N_E\,N_C)}\!/_{(N_E+N_C)}}}\cdot\left(\bar{B}^\star_E - \bar{B}^\star_C\right)$:

$$\delta^2_\star + \lambda = \frac{\vartheta^2}{\sigma^2}\,(N-1)\,S^2_{\boldsymbol{B}^\star_{\text{CB}}}, \qquad (3.17)$$

with $\boldsymbol{B}^\star_{\text{CB}} = \boldsymbol{B}_{\text{CB}} + \boldsymbol{T}\,{}^{\theta_1}\!/_{\vartheta}$. Particularly, the noncentrality parameter $\lambda = {}^{\vartheta^2}\!/_{\sigma^2}\left((N_E-1)\,S^2_{E,\boldsymbol{b}} + (N_C-1)\,S^2_{C,\boldsymbol{b}}\right)$ is independent of the constant shift θ_1 between the two treatment groups, but dependent on $\boldsymbol{T}$ in the number of patients allocated to the two treatment groups treatment groups E and C. In the next section, the boundary ${}^{\vartheta^2}\!/_{\sigma^2}\,(N-1)\,S^2_{\boldsymbol{B}_{\text{CB}}}$ for $\delta^2+\lambda$ in situations of a linear time trend and a step trend is derived. For these calculations the parameter θ_1 is assumed to be zero.

Boundary for $\delta^2 + \lambda$ in case of a linear time trend

Under the assumption of Model (2.28) with $\theta_1 = 0$ the bias vector has the following form $\boldsymbol{B}_{\text{CB}} = {}^1\!/_N\,(0,1,2,\ldots,N-1)^T$. The term ${}^{\vartheta^2}\!/_{\sigma^2}\,(N-1)\,S^2_{\boldsymbol{B}_{\text{CB}}}$ defined in Proposition 3.5 is calculated as follows:

$$\begin{aligned}
\frac{\vartheta^2}{\sigma^2}\,(N-1)\,S^2_{\boldsymbol{B}_{\text{CB}}} &= \frac{\vartheta^2}{\sigma^2}\sum_{i=1}^{N}\left(b_i - \bar{B}_{\text{CB}}\right)^2 \\
&= \frac{\vartheta^2}{\sigma^2}\left(\sum_{i=1}^{N} b_i^2 - 2\sum_{i=1}^{N} b_i\,\bar{B}_{\text{CB}} + N\,\bar{B}^2_{\text{CB}}\right) \\
&= \frac{\vartheta^2}{\sigma^2}\left(\sum_{i=1}^{N} b_i^2 - \frac{2}{N}\left(\sum_{i=1}^{N} b_i\right)^2 + \frac{1}{N}\left(\sum_{i=1}^{N} b_i\right)^2\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\vartheta^2}{\sigma^2}\left(\sum_{i=1}^{N} b_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N} b_i\right)^2\right) \\
&= \frac{\vartheta^2}{N^2\,\sigma^2}\sum_{i=1}^{N-1} i^2 - \frac{1}{N^3}\left(\sum_{i=1}^{N-1} i\right)^2 \\
&\overset{(2)}{=} \frac{\vartheta^2}{6\,N^2\,\sigma^2}(N-1)\,N\,(2\,N-1) - \frac{1}{N^3}\left(\frac{(N-1)\,N}{2}\right)^2 \\
&= \frac{\vartheta^2}{6\,N\,\sigma^2}(N-1)\,(2\,N-1) - \frac{1}{4\,N}(N-1)^2 \\
&= \frac{\vartheta^2}{N\,\sigma^2}\left(\frac{1}{6}(2\,N^2-3\,N+1) - \frac{1}{4}(N^2-2\,N+1)\right) \\
&= \frac{\vartheta^2}{12\,N\,\sigma^2}\left(4\,N^2-6\,N+2-3\,N^2+6\,N-3\right) \\
&= \frac{\vartheta^2}{12\,N\,\sigma^2}(N^2-1)\,. \qquad (3.18)
\end{aligned}
$$

Assuming a linear time trend and $N_E \geq 2$ and $N_C \geq 2$, the term $\max(\delta^2)$ is unequal to ${}^{\vartheta^2}/_{\sigma^2}\,(N-1)\,S^2_{\boldsymbol{B}_{\mathrm{CB}}}$. Particularly, the noncentrality parameter λ is always unequal to zero. The bias vector $\boldsymbol{B}_{\mathrm{CB}} \in \mathbb{R}^N$ in case of a linear time trend is not a constant shift in the expectations of the patients' responses in both treatment groups.

Boundary for $\delta^2 + \lambda$ in case of a step time trend

In what follows, Model (2.28) with $\theta_1 = 0$ is assumed. Furthermore, the bias vector has the following form $\boldsymbol{B}_{\mathrm{CB}} = (\mathbf{0}^T_{N-s}, \mathbf{1}^T_s)^T\,\vartheta$ with $s \in \mathbb{N}$ and $0 < s < N$. The parameter $\vartheta \in \mathbb{R}$ defines the strength and $s \in \mathbb{N}$ the length of the step time trend. Now, the term ${}^{\vartheta^2}/_{\sigma^2}\,(N-1)\,S^2_{\boldsymbol{B}_{\mathrm{CB}}}$ according to Proposition 3.5 is exactly calculated:

$$
\begin{aligned}
\frac{\vartheta^2}{\sigma^2}(N-1)\,S^2_{\boldsymbol{B}_{\mathrm{CB}}} &= \frac{\vartheta^2}{\sigma^2}\sum_{i=1}^{N}\left(b_i - \bar{B}_{\mathrm{CB}}\right)^2 \\
&= \frac{\vartheta^2}{\sigma^2}\left(\sum_{i=1}^{N} b_i^2 - 2\sum_{i=1}^{N} b_i\,\bar{B}_{\mathrm{CB}} + N\,\bar{B}^2_{\mathrm{CB}}\right) \\
&= \frac{\vartheta^2}{\sigma^2}\left(\sum_{i=1}^{N} b_i^2 - \frac{2}{N}\left(\sum_{i=1}^{N} b_i\right)^2 + \frac{1}{N}\left(\sum_{i=1}^{N} b_i\right)^2\right)
\end{aligned}
$$

(2) $\sum_{i=1}^{N} i = \frac{N\,(N+1)}{2}$ and $\sum_{i=1}^{N} i^2 = \frac{1}{6}\,N\,(N+1)\,(2\,N+1)$

$$
\begin{aligned}
&= \frac{\vartheta^2}{\sigma^2}\left(\sum_{i=1}^{N} b_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N} b_i\right)^2\right) \\
&= \frac{\vartheta^2}{\sigma^2}\left(s - \frac{s^2}{N}\right) \\
&= \frac{\vartheta^2}{\sigma^2}\, s\left(\frac{N-s}{N}\right). \qquad (3.19)
\end{aligned}
$$

Below, the maximum of $t(s) = s - {}^{s^2}/_{N}$ is derived. Therefore the first derivative is set equal to zero (necessary condition for an extremum):

$$
\begin{aligned}
\frac{\partial t(s)}{\partial s} &= 1 - \frac{2\,s}{N} = 0 \\
\Leftrightarrow s &= \frac{N}{2}.
\end{aligned}
$$

Due to the fact that the second derivative of $t(s)$ is always smaller than zero, the function $t(s)$ has a maximum at the position $s = {}^{N}/_{2}$. Thus, the maximal extend of $\delta^2 + \lambda$ in case of a step trend is:

$$
\max_{s\in\{1,2,\ldots,N\}}\left(\frac{\vartheta^2}{\sigma^2}\,(N-1)\,S^2_{\boldsymbol{B}_{\mathrm{CB}}}\right) = \frac{\vartheta^2}{\sigma^2}\left(\frac{N}{2} - \frac{N^2}{4\,N}\right) = \frac{\vartheta^2\,N}{\sigma^2\,4}. \qquad (3.20)
$$

It is obvious, that as $s \to 0$ or as $s \to N$ the term $\delta^2 + \lambda$ approaches zero (see Equation (3.19)). Provided that the analysis is unadjusted for the step time trend, a step trend with moderate strength ϑ at the beginning or the end of a clinical trial will not have any or just less influence on the analysis of the clinical trial.

3.1.2 The doubly noncentral t-distribution

The doubly noncentral t-distribution is an extension of the noncentral t-distribution. The noncentral t-distribution depends only on the noncentrality parameter δ. Hence, the second noncentrality parameter λ is by default zero for the noncentral t-distribution. According to Kocherlakota and Kocherlakota (1991) the distribution function of the doubly noncentral t-distribution with h degrees of freedom and noncentrality parameters δ and λ can be ex-

pressed as follows:

$$F_{t_{h,\delta,\lambda}}(x) = \sum_{k=0}^{\infty} f_{\mathrm{Poi}(\lambda/2)}(k) \cdot F_{t_{h+2k,\delta}}\left(x\sqrt{\frac{h+2\,k}{h}}\right) . \tag{3.21}$$

The term $f_{\mathrm{Poi}(\lambda/2)}(k)$ defines the density function of a Poisson distribution on the position k with parameter $\lambda/2$ and $F_{t_{h+2k,\delta}}\left(x\sqrt{(h+2\,k)/h}\right)$ defines the distribution function of a noncentral t-distribution on the position $\left(x\sqrt{(h+2\,k)/h}\right)$ with $h + 2k$ degrees of freedom and noncentrality parameter δ. Analogously to the characterization of the distribution function of the doubly noncentral t-distribution, the corresponding density function can be expressed as follows:

$$f_{t_{h,\delta,\lambda}}(x) = \sum_{k=0}^{\infty} f_{\mathrm{Poi}(\lambda/2)}(k) \cdot f_{t_{h+2k,\delta}}\left(x\sqrt{\frac{h+2\,k}{h}}\right) . \tag{3.22}$$

Both the density and the distribution function of the doubly noncentral t-distribution at the position x depend on an infinite sum, which starts at zero. For all presented results in this thesis, this infinite sum is evaluated at the positions $0, 1, \ldots, q$, where q is the 99% quantile of the poisson distribution function with parameter $\lambda/2$. A fixed value of the sum $\delta^2 + \lambda$ (see Proposition 3.5) is of particular interest in the later analysis. Figure 3.1 shows several density functions for a doubly noncentral t-distribution with 10 degrees of freedom and $\delta^2 + \lambda = 12$. In general, the figure shows that the greater the noncentrality parameter δ is, the greater the shift of the density function to the right becomes. Comparing the density function of the t-distribution ($\delta = \lambda = 0$) to the density function of the doubly noncentral t-distribution with $\delta = 0$ it becomes apparent that the density function of the doubly noncentral t-distribution is narrower, whereas the tails of the density function of the t-distribution are heavier. This behavior is derivable from the density function presented in Equation (3.22). The smaller the noncentrality parameter λ is, the greater the values of the density function of the Poisson distribution for small values of the parameter k become. The values of the Poisson density function can be interpreted as weights for the density function of the noncentral t-distribution. The smaller the noncentrality parameter λ is, the more weight is set on greater values of the noncentral t-distribution. The term $f_{t_{h+2k,\delta}}\left(x\sqrt{(h+2\,k)/h}\right)$ approaches zero for growing k with $x \neq 0$ and moderate values of h and δ.

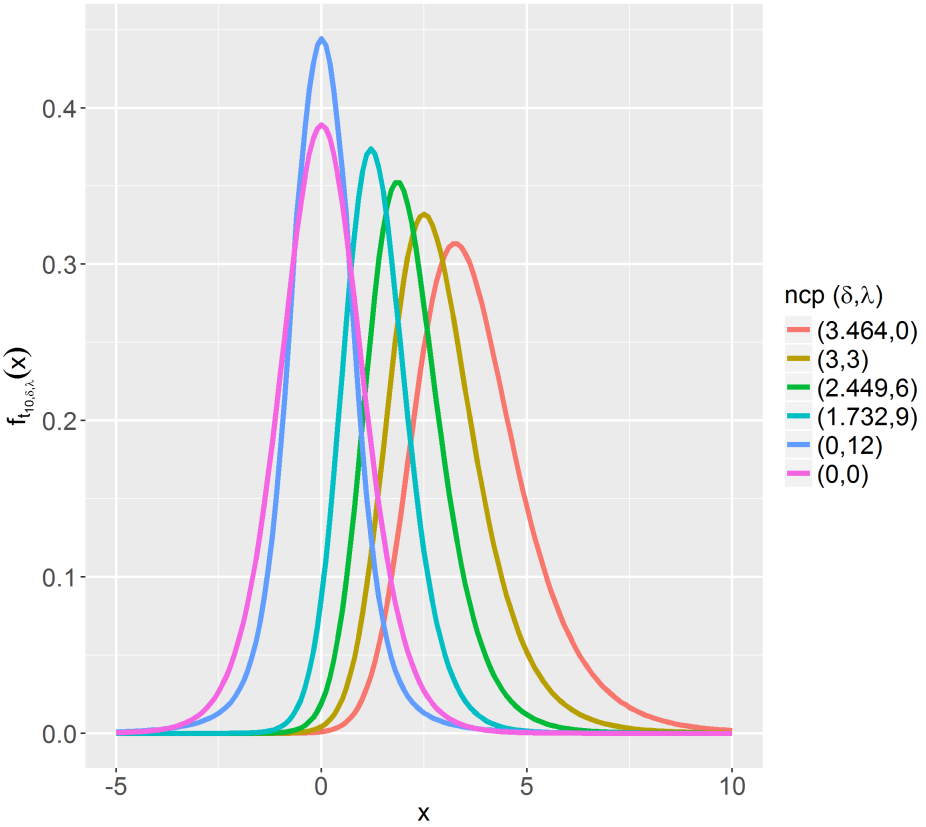


Figure (3.1): Several settings of the density function of the doubly noncentral t-distribution with 10 degrees of freedom and $\delta^2 + \lambda = 12$, where δ and λ are the noncentrality parameters.

Implementation of the doubly-noncentral t-distribution in R

The doubly noncentral t-distribution does not belong to the standard implementation of statistical software. The software R (R Core Team, 2016), which is used in this thesis, might have some inaccuracies with the calculation of the density function or distribution function of the noncentral t-distribution. Particularly for extreme noncentrality parameters (called ncp in the help of the "The Student t-Distribution" in the R help) some inaccuracies may occur:

> »Supplying ncp = 0 uses the algorithm for the non-central distribution, which is not the same algorithm used if ncp is omitted. This is to give consistent behavior in extreme cases with values of ncp very near zero. The code for non-zero ncp is principally

> intended to be used for moderate values of ncp: it will not be highly accurate, especially in the tails, for large values.«
>
> (R help from `dt` (distribution function of the t-distribution))

Despite warnings from `R` on using their own implemented function, it is assumed that this problem does not affect any results in this thesis. All functions using the distribution function or quantiles of the noncentral t-distribution with "extreme" noncentrality parameters were cross-checked with results from the literature (tables found in Kocherlakota and Kocherlakota (1991)).

3.1.3 Conditional rejection probabilities

The type-I-error probability $\alpha \in (0, 1)$ or size of a test (for example of Student's t-test) is defined as the probability of a false rejection of the null hypothesis $H_0 : \mu_C = \mu_E$ (see Equation (3.2)). The opposite of the type-I-error probability is the type-II-error probability $\beta \in (0, 1)$. The power $1 - \beta$ of a test (for example of Student's t-test) is the probability to reject the null hypothesis and accept the alternative $H_1 : \mu_C \neq \mu_E$, when the alternative is true. In general, the power and the type-I-error probability form a contrast. That means it is not possible to guarantee a high power and a low type-I-error probability. In clinical trials the planned power is often 80% and the size α is set to 5% beforehand. Thus, a false rejection of the null hypothesis is allowed in 5% of the cases and in 80% of the cases the alternative should be correctly accepted, if it is true. Section 3.1 shows that under the assumption of a homogeneous population model $\boldsymbol{Y} = \boldsymbol{A\theta} + \boldsymbol{\epsilon}$ (see Model (2.16)) and $2\,\theta_1 = \mu_E - \mu_C = 0$ the test statistic W defined in Equation (3.3) is t-distributed with $N - 2$ degrees of freedom. Thus, a two-sided α-level test is derived by rejecting the null hypothesis if the test statistic W is greater than the $(1 - \alpha/2)$-quantile of the t-distribution with $N - 2$ degrees of freedom or the test statistic W is smaller than the $\alpha/2$-quantile of the t-distribution with $N - 2$ degrees of freedom. In the literature this test is called two-sided two-sample Student's t-test (see Student, 1908). In this section the population model affected by chronological bias $\boldsymbol{Y} = \boldsymbol{A\theta} + \boldsymbol{B}_{\mathrm{CB}}\,\vartheta + \boldsymbol{\epsilon}$ according to Equation (2.28) is considered. The situation of conducting Student's t-test without adjusting for the additional bias vector $\boldsymbol{B}_{\mathrm{CB}} \in \mathbb{R}^N$ and its strength ϑ is investigated. On the one hand the

planned size of the test α_0 is distorted due to an unadjusted analysis for the bias term $\boldsymbol{B}_{\mathrm{CB}}\,\vartheta$ and on the other hand the planned power $1-\beta_0$ is distorted in the same manner. In Section 3.1 it was shown that the distortion of the rejection probability under Student's t-test is strongly dependent on the randomization sequence. Under the assumption of the population model affected by chronological bias (see Equation (2.28)), the type-I-error probability α_{CB} of Student's t-test is determined as follows:

$$\alpha_{\mathrm{CB}} = 1 - F_{t_{N-2,\delta,\lambda}}\left(t_{1-\alpha_0/2,N-2}\right) + F_{t_{N-2,\delta,\lambda}}\left(t_{\alpha_0/2,N-2}\right)\,, \tag{3.23}$$

where the noncentrality parameters $\delta = {}^{\vartheta}\!/_{\sigma}\sqrt{{}^{(N_E\,N_C)}\!/_{(N_E+N_C)}}\left(\bar{B}_E - \bar{B}_C\right)$ (see Equation (3.10)) and $\lambda = {}^{\vartheta^2}\!/_{\sigma^2}\left((N_E-1)\,S^2_{E,\boldsymbol{b}} + (N_C-1)\,S^2_{C,\boldsymbol{b}}\right)$ (see Equation (3.13)) of the distribution function of the doubly noncentral t-distribution are dependent on the bias vector $\boldsymbol{B}_{\mathrm{CB}} \in \mathbb{R}^N$, its strength ϑ, and $\boldsymbol{T}$. The parameter θ_1 in Model (2.28) is assumed to be zero, otherwise α_{CB} will not be a type-I-error. Table 3.1 shows the noncentrality parameters and the corresponding type-I-error probabilities conditioned on the possible randomization sequences under the Random Allocation Rule for $N = 4$ (see Section 2.2.1). The nominal significance level α_0 of the conducted test is set to 5%. For the corresponding quantiles of the t-distribution it follows $t_{0.025,2} = -t_{0.975,2} = 4.303$. In the presented setting it is assumed that the clinical trial is affected by a comparatively strong linear time trend with $\vartheta = 4$ (see Equation (2.25)). The term $\boldsymbol{B}^T_{\mathrm{CB}}\,\vartheta$ is independent of the randomization sequence and corresponds to the vector $(0,1,2,3)^T$. The boundary for $\delta^2+\lambda$, which is derived in Equation (3.18), is five for $\sigma^2 = 1$.

Table 3.1 shows the realizations of the type-I-error probabilities and the noncentrality parameters δ and λ conditioned on the randomization sequences under the Random Allocation Rule for $N = 4$ (see Section 2.2.1). The expected or average values of the investigated parameters can be calculated by weighting the presented realizations with their probability of appearance. Under the Random Allocation Rule all possible randomization sequences are equiprobable. Thus, the mean values of the realizations of the type-I-error probabilities and the noncentrality parameters δ and λ conditioned on the randomization sequences correspond to the expected values under the Random Allocation Rule. Only the average value of the type-I-error probability in the situation of chronological bias has been assessed extensively

i	$\boldsymbol{t}_i^T$	$\boldsymbol{B}_{\text{CB}}^T\vartheta$	δ_i	λ_i	$\delta_i^2+\lambda_i$	$F_{t_{2,\delta_i,\lambda_i}}(-4.303)$	1-$F_{t_{2,\delta_i,\lambda_i}}(4.303)$	$\alpha_{\text{CB}}(\boldsymbol{t}_i)$
0	$\{(T_1,T_2,T_3,T_4): \sum_{i=1}^4 T_i=0\}$	(0,0,0,0)	0	0	0	0.025	0.025	0.050
1	(1,1,-1,-1)	(0,1,2,3)	-2	1	5	0.146	0.000	0.146
2	(1,-1,1,-1)	(0,1,2,3)	-1	4	5	0.016	0.001	0.016
3	(-1,1,1,-1)	(0,1,2,3)	0	5	5	0.002	0.002	0.005
4	(1,-1,-1,1)	(0,1,2,3)	0	5	5	0.002	0.002	0.005
5	(-1,1,-1,1)	(0,1,2,3)	1	4	5	0.001	0.016	0.016
6	(-1,-1,1,1)	(0,1,2,3)	2	1	5	0.000	0.146	0.146
	average values:		0	3.333	5	0.028	0.028	0.056

Table (3.1): Realizations of α_{CB}, δ, and λ conditioned on the randomization sequences under RAR with $N=4$, $\alpha_0=0.05$, $\vartheta=4$, and $\sigma^2=1$.

in the literature (see Rosenkranz, 2011; Rosenberger and Lachin, 2016). Usually, this average value is time-intensively simulated. Tamm and Hilgers (2014) conducted a worst case sequence analysis in situations of a population model affected by chronological bias. A worst case randomization sequence of a randomization procedure is defined as a randomization sequence, which attains the greatest value of the type-I-error probabilities. Under the Random Allocation Rule with $N=4$ these are the randomization sequences $\boldsymbol{t}_1$ and $\boldsymbol{t}_6$ (see Table 3.1) with the value $\alpha_{\text{CB}}=0.146$. Table 3.1 shows that the average value of the type-I-error probabilities under the Random Allocation Rule in case of a strong linear time trend is 5.6%. This value does not differ as much from the planned type-I-error probability $\alpha_0=5\%$. Although any type-I-error probability of any randomization sequence corresponds with the nominal significance level of 5%, the nominal significance level is preserved almost in average. This property of the Random Allocation Rule is more closely investigated in Section 3.3.

Below, the setting when the alternative hypothesis $H_1: \mu_E \neq \mu_C$ is true is investigated. Analogously to the type-I-error probabilities, the power of Student's t-test conditioned on the randomization sequences of a particular randomization procedure is calculated. When investigating the setting under the alternative hypothesis, the difference between the effects of the expectations of the two treatment groups has to be quantified. It is assumed that the effect of the experimental treatment is superior to the effect of the control treatment (i.e. $\mu_E > \mu_C$) and should be correctly detected with a probability of 80%. For further calculations the effect size $\Delta_0 = |\mu_E - \mu_C|/\sigma = |2\,\theta_1|/\sigma$, which attains a power of 80%, has to be derived. For the setting

$N = 4$, $\sigma^2 = 1$, $N_E = N_C = 2$, $\alpha_0 = 0.05$, and $1 - \beta_0 = 80\%$ an effect size $\Delta_0 = 5.653$ is computed with the statistical software `R` (R Core Team, 2016). Under the assumption of the population model affected by chronological bias defined in Equation (2.28), the power $1 - \beta_{\text{CB}}$ of Student's t-test can be derived as follows:

$$1 - \beta_{\text{CB}} = 1 - F_{t_{N-2,\delta,\lambda}}\left(t_{1-\alpha_0/2,N-2}\right) + F_{t_{N-2,\delta,\lambda}}\left(t_{\alpha_0/2,N-2}\right), \tag{3.24}$$

with $\delta = {}^{\vartheta}/_{\sigma\sqrt{{}^{(N_E\,N_C)}/_{(N_E+N_C)}}}\left(2\,{}^{\theta_1}/_{\vartheta} + \bar{B}_E - \bar{B}_C\right)$ (see Equation (3.10)) and $\lambda = {}^{\vartheta^2}/_{\sigma^2}\left((N_E - 1)\,S^2_{E,\boldsymbol{b}} + (N_C - 1)\,S^2_{C,\boldsymbol{b}}\right)$ (see Equation (3.13)). The non-centrality parameters δ and λ of the distribution function of the doubly non-central t-distribution are strongly dependent on the bias vector $\boldsymbol{B}_{\text{CB}} \in \mathbb{R}^N$, its strength ϑ, the parameter θ_1, and $\boldsymbol{T}$. Furthermore, the term $S^2_{\boldsymbol{B}_{\text{CB}^\star}}$ according to Equation (3.17) with $\boldsymbol{B}^\star_{\text{CB}}\,\vartheta = \boldsymbol{B}_{\text{CB}}\,\vartheta + \boldsymbol{T}\,\theta_1$ is not constant. Table 3.2 shows the realizations of the power, δ, and λ conditioned on the possible randomization sequences under the Random Allocation Rule with $N = 4$ for $\theta_1 = {}^{\Delta_0}/_2 = 2.83$ and $\lambda = 4$. Again, it is assumed that the clinical trial is affected by a comparatively strong linear time trend with $\vartheta = 4$ (see Equation (2.25)), which is not taken into account in the analysis. Table 3.2 shows that the values for the power $1 - \beta_{\text{CB}}$, δ, and λ are strongly dependent on the randomization sequence. A worst case randomization sequence is defined as a sequence with the lowest value of the power $1 - \beta_{\text{CB}}$. Under the Random Allocation Rule with $N = 4$ this is the randomization sequence $\boldsymbol{t}_2$ with $\beta_{\text{CB}} = 0.231$. Furthermore, neither the power of any randomization sequence $\boldsymbol{t}_i$ with $i \in \{1, \ldots, 6\}$ corresponds with the planned power $1 - \beta_0 = 0.8$ nor the planned power is preserved in average. When the clinical

i	$\boldsymbol{t}_i^T$	$(\boldsymbol{B}^\star_{\text{CB}})^T\,\vartheta$	δ_i	λ_i	$\delta_i^2 + \lambda_i$	$F_{t_{2,\delta_i,\lambda_i}}(-4.303)$	$1\text{-}F_{t_{2,\delta_i,\lambda_i}}(4.303)$	$1 - \beta_{\text{CB}}(\boldsymbol{t}_i)$
0	$\{(T_1,T_2,T_3,T_4):$ $\sum_{i=1}^4 T_i = 0\}$	$\{(b_1,b_2,b_3,b_4):$ $b_i = T_i\theta_1 \vert \boldsymbol{T}\}$	5.653	0	31.753	0	0.8	0.8
1	(1,1,-1,-1)	(2.83, 3.83, -0.83, 0.17)	3.653	1	14.348	0	0.368	0.368
2	(1,-1,1,-1)	(2.83, -1.83, 4.83, 0.17)	4.653	4	25.655	0	0.231	0.231
3	(-1,1,1,-1)	(-2.83, 3.83, 4.83, 0.17)	5.653	5	36.962	0	0.274	0.274
4	(1-,1,-1,1)	(2.83, -1.83, -0.83, 5.83)	5.653	5	36.962	0	0.274	0.274
5	(-1,1,-1,1)	(-2.83, 3.83, -0.83, 5.83)	6.653	4	48.269	0	0.476	0.476
6	(-1,-1,1,1)	(-2.83, -1.83, 4.83, 5.83)	7.653	1	59.576	0	0.867	0.867
	average values:		5.653	3.333	36.962	0	0.415	0.415

Table (3.2): Realizations of the power $1 - \beta_{\text{CB}}$, δ, and λ conditioned on the randomization sequences under RAR with $N = 4$, $1 - \beta_0 = 0.8$, $\vartheta = 4$, $\sigma^2 = 1$, and $\theta_1 = 2.83$.

trial is affected by such a strong time trend, which is assumed not to be taken into account in the analysis, the average power under the Random Allocation Rule decreases to 0.415. In other words, the test statistic W of Student's t-test is distorted so much by the misspecification of the time trend, that it is nearly impossible to detect an effect of the size $\vartheta_1 = 2.83$ in the difference between the effects of the two treatment groups. As already pointed out the realizations of the noncentrality parameter λ are independent of a difference between the effects of the two treatment groups. Thus, the values of λ_i with $i \in \{1, \ldots, 6\}$ are the same in Table 3.2 and Table 3.1. This property is of particular interest in Section 3.3.

3.2 Influence of selection bias on the test decision

In this section the results from Section 3.1 and particularly from the Subsection 3.1.3 are transferred to the model $\boldsymbol{Y} = \boldsymbol{A\theta} + \boldsymbol{B}_{\mathrm{SB}}\,\eta + \boldsymbol{\epsilon}$ introduced in Equation (2.24). Analogously to α_{CB} the type-I-error probability α_{SB} is defined as follows:

$$\alpha_{\mathrm{SB}} = 1 - F_{t_{N-2,\delta,\lambda}}\left(t_{1-\alpha_0/2,N-2}\right) + F_{t_{N-2,\delta,\lambda}}\left(t_{\alpha_0/2,N-2}\right), \tag{3.25}$$

where the noncentrality parameters $\delta = \eta/\sigma\sqrt{(N_E\,N_C)/(N_E+N_C)}\left(\bar{B}_E - \bar{B}_C\right)$ and $\lambda = \eta^2/\sigma^2\left((N_E-1)\,S^2_{E,\boldsymbol{b}} + (N_C-1)\,S^2_{C,\boldsymbol{b}}\right)$ are defined in the Equations (3.10) and (3.13), respectively. The noncentrality parameters of the distribution function of the doubly-noncetral t-distribution are dependent on the bias vector $\boldsymbol{B}_{\mathrm{SB}} \in \mathbb{R}^N$, its strength η, and $\boldsymbol{T}$. The vector $\boldsymbol{B}_{\mathrm{SB}} \in \mathbb{R}^N$ (see Equation (2.24)) is in contrast to $\boldsymbol{B}_{\mathrm{CB}} \in \mathbb{R}^N$ strongly dependent on $\boldsymbol{T}$. Thus, the equality $\delta^2 + \lambda = \eta^2/\sigma^2\,(N-1)\,S^2_{\boldsymbol{B}_{\mathrm{SB}}}$ (see Proposition 3.5) holds, but $S^2_{\boldsymbol{B}_{\mathrm{SB}}}$ is not constant as it is in the presence of a known time trend (see Table 3.1). Table 3.3 shows the realizations of the noncentrality parameters and of the corresponding type-I-error probabilities conditioned on the randomization sequences under the Random Allocation Rule with $N = 4$ (see Section 2.2.1). The nominal significance level α_0 of the test is set to 5% and for the strength of selection bias the effect $\eta = 1$ is assumed. The randomization sequences $\boldsymbol{t}_2$ and $\boldsymbol{t}_5$ attain with 0.091 a maximal elevation of the type-I-error probability. Thus, these randomization sequences are the worst case sequences. The average value of the type-I-error probabilities in

i	$\boldsymbol{t}_i^T$	$\boldsymbol{B}_{\text{SB}}^T\,\eta$	δ_i	λ_i	$\delta_i^2+\lambda_i$	$F_{t_{2,\delta_i,\lambda_i}}(-4.303)$	1-$F_{t_{2,\delta_i,\lambda_i}}(4.303)$	$\alpha_{\text{SB}}(\boldsymbol{t}_i)$
1	(1,1,-1,-1)	(0,-1,-1,-1)	0.5	0.5	0.75	0.008	0.041	0.049
2	(1,-1,1,-1)	(0,-1,0,-1)	1.0	0.0	1.00	0.004	0.091	0.095
3	(-1,1,1,-1)	(0,1,0,-1)	1.0	1.0	2.00	0.002	0.059	0.061
4	(1,-1,-1,1)	(0,-1,0,1)	1.0	1.0	2.00	0.002	0.059	0.061
5	(-1,1,-1,1)	(0,1,0,1)	1.0	0.0	1.00	0.004	0.091	0.095
6	(-1,-1,1,1)	(0,1,1,1)	0.5	0.5	0.75	0.008	0.041	0.049
	average values:		0.833	0.5	1.25	0.005	0.064	0.068

Table (3.3): Realizations of α_{SB}, δ, and λ conditioned on the randomization sequences under RAR with $N = 4$, $\alpha_0 = 0.05$, $\eta = 1$, and $\sigma^2 = 1$.

the presence of selection bias with the strength $\eta = 1$ is 0.068. Compared to $\alpha_0 = 0.05$, the average type-I-error probability under the Random Allocation Rule is increased significantly. Langer (2014) extensively reviewed the behavior of the Random Allocation Rule in the situation of selection bias with $N = 6$. Particularly, the behavoir of the individual randomization sequences dependent on the strength of selection bias η is investigated.

Analogously to the power in a model affected by chronological bias (see Section 3.1.3), for the calculation of the power in a model affected by selection bias it follows:

$$1-\beta_{\text{SB}} = 1 - F_{t_{N-2,\delta,\lambda}}\left(t_{1-\frac{\alpha_0}{2},N-2}\right) + F_{t_{N-2,\delta,\lambda}}\left(t_{\frac{\alpha_0}{2},N-2}\right), \tag{3.26}$$

with $\delta = {}^{\vartheta}/_{\sigma}\sqrt{{}^{(N_E\,N_C)}/_{(N_E+N_C)}}\left(2\,\theta_1/\vartheta + \bar{B}_E - \bar{B}_C\right)$ (see Equation (3.10)) and $\lambda = {}^{\vartheta^2}/_{\sigma^2}\left((N_E-1)\,S^2_{E,\boldsymbol{b}} + (N_C-1)\,S^2_{C,\boldsymbol{b}}\right)$ (see Equation (3.13)). The noncentrality parameters δ and λ of the distribution function of the doubly noncentral t-distribution are dependent on the bias vector $\boldsymbol{B}_{\text{SB}} \in \mathbb{R}^N$, its strength η, the parameter θ_1, and $\boldsymbol{T}$. Analogously to Table 3.2, the effect size $\Delta_0 = 5.653$ is assumed in the setting for generating Table 3.4. This effect size attains with $N = 4$ and $N_E = N_C = 2$ a power of 80% if and only if the trial will not be affected by selection bias. For further calculations it is assumed that $\mu_E > \mu_C$ holds. Surprisingly, Table 3.4 shows that although the investigator tries to select well responding patients using the convergence strategy to the experimental group, the power will not exceed the prespecified power $1 - \beta_0 = 80\%$ for all randomization sequences. The randomization sequences $\boldsymbol{t}_3$ and $\boldsymbol{t}_4$ attain only a power of less than 78%. Thus, even here worst case randomization sequences are found. The decreased power of these two randomization sequences can be explained partly by the fact that

i	$\boldsymbol{t}_i^T$	$(\boldsymbol{B}_{\text{SB}}^{\star})^T\eta$	δ_i	λ_i	$\delta_i^2+\lambda_i$	$F_{t_{2,\delta_i,\lambda_i}}(-4.303)$	1-$F_{t_{2,\delta_i,\lambda_i}}(4.303)$	$1-\beta_{\text{SB}}(\boldsymbol{t}_i)$
1	(1,1,-1,-1)	(2.83, 1.83, -3.83, -3.83)	6.153	0.5	38.365	0	0.786	0.786
2	(1,-1,1,-1)	(2.83, -3.83, 2.83, -3.83)	6.653	0.0	44.269	0	0.890	0.890
3	(-1,1,1,-1)	(-2.83, 3.83, 2.83, -3.83)	6.653	1.0	45.269	0	0.779	0.779
4	(1,-1,-1,1)	(2.83, -3.83, -2.83, 3.83)	6.653	1.0	45.269	0	0.779	0.779
5	(-1,1,-1,1)	(-2.83, 3.83, -2.83, 3.83)	6.653	0.0	44.269	0	0.890	0.890
6	(-1,-1,1,1)	(-2.83, -1.83, 3.83, 3.83)	6.153	0.5	38.365	0	0.786	0.786
	average values:		6.487	0.500	42.634	0	0.819	0.819

Table (3.4): Realizations of the power $1-\beta_{\text{SB}}$, δ, and λ conditioned on the randomization sequences under RAR with $N=4$, $1-\beta_0=0.8$, $\eta=1$, $\sigma^2=1$, and $\theta_1=2.83$.

the investigator guesses incorrectly once. He or she enrolls a well (badly) responding patient when a badly (well) responding patient would be the right choice from his or her point of view. Mathematically Table 3.4 shows that the randomization sequences $\boldsymbol{t}_1$ and $\boldsymbol{t}_6$ have the lowest noncentrality parameter δ in comparison to the other randomization sequences. Additionally, the noncentrality parameter λ for these randomization sequences is unequal to zero. To attain a great value of the power, a great value of δ and a low value of λ are required. The average value of the power under the Random Allocation Rule in the presence of selection bias with strength $\eta=1$ is 0.819. Compared to the planned power $1-\beta_0=0.8$, the average value of 0.819 can be interpreted as an increase in power.

3.3 Student's t-test under Random Allocation Rule

In this section the standard Model (2.16), which depends on the parameters θ_0 and θ_1, is investigated. The null hypothesis $H_0:\theta_1=0$ is tested against the alternative hypothesis $H_1:\theta_1\neq 0$ with Student's t-test. Furthermore, it is assumed that in Model (2.16) a bias vector $\boldsymbol{b}=(b_1,b_2,\ldots,b_N)^T\in\mathbb{R}^N$ as additional fixed effect is present. The model is not adjusted for this additional effect, so that the fixed effect $\boldsymbol{b}\in\mathbb{R}^N$ interferes the estimation of θ_1. Particularly, the test decision $H_0:\theta_1=0$ against $H_1=\theta_1\neq 0$ is distorted. As randomization procedure the Random Allocation Rule (see Section 2.2.1) is considered, so that half of the patients are allocated to each treatment group. According to Lemma A.3 in the appendix the covariance of two random allocations T_i and T_j with $i\neq j$ and $i,j\in\{1,2,\ldots,N\}$ under

the Random Allocation Rule is given by:

$$\mathrm{Cov}(T_i, T_j) = E(T_i\, T_j) = \frac{-1}{N-1}\,. \tag{3.27}$$

In this section the regression models introduced for chronological and selection bias in Chapter 2 are investigated more generally. Analogously to the regression model for chronological bias defined in Equation (2.27) the following model is assumed for further investigations:

$$\boldsymbol{Y} = \begin{pmatrix} 1 & T_1 \\ 1 & T_2 \\ \vdots & \vdots \\ 1 & T_N \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{pmatrix} + \sigma \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{pmatrix} = \boldsymbol{A\theta} + \boldsymbol{b} + \sigma\, \boldsymbol{\epsilon}\,. \tag{3.28}$$

Formally, the presented model describes a fixed effects model. Alternatively, the ith patient's response can be expressed as $Y_i = \theta_0 + T_i\,\theta_1 + b_i + \sigma\,\epsilon_i$. The error terms ϵ_i with $i \in \{1, 2, \ldots, N\}$ are assumed to be normally distributed with expectation zero and variance one. Furthermore, the error terms ϵ_i with $i \in \{1, 2, \ldots, N\}$ of the vector $\boldsymbol{\epsilon}$ are assumed to be independent and identically distributed. Particularly, the random vector $\boldsymbol{\epsilon}$ is independent of the randomization sequence $\boldsymbol{T}$. The (known and fixed) bias vector $\boldsymbol{b} \in \mathbb{R}^N$ is a fixed effect and is independent of $\boldsymbol{\epsilon}$ and $\boldsymbol{T}$.

The test statistic W from Equation (3.14) can be transformed in the following manner:

$$\begin{aligned} W &:= \frac{\hat{\theta}_1}{\sqrt{\widehat{\mathrm{Var}}(\hat{\theta}_1)}} \\ &= \sqrt{\frac{N_E\, N_C}{N_E + N_C}}\, \frac{\bar{Y}_E - \bar{Y}_C}{S_p} \\ &\text{with } S_p^2 = \frac{(N_E - 1)\, S_E^2 + (N_C - 1)\, S_C^2}{N - 2}\,, \end{aligned} \tag{3.29}$$

where S_p^2 defines the pooled variance of the investigated treatment groups. The term $S_E^2 = {}^1\!/_{(N_E-1)} \sum_{i=1}^N \mathbb{1}_{\{T_i=1\}} \left(Y_i - {}^1\!/_{N_E} \sum_{j=1}^N Y_j\, \mathbb{1}_{\{T_j=1\}}\right)^2$ is the variance of the patients' responses in treatment group E and $S_C^2 = {}^1\!/_{(N_C-1)}$ $\sum_{i=1}^N \mathbb{1}_{\{T_i=-1\}} \left(Y_i - {}^1\!/_{N_C} \sum_{j=1}^N Y_j\, \mathbb{1}_{\{T_j=-1\}}\right)^2$ is the variance of the patients'

responses in treatment group C. Below, a lemma is given that the variance within a treatment group is independent of the parameters θ_0 and θ_1.

Lemma 3.6: (Variance within a treatment group independent of θ_0 and θ_1) *Under the assumption of Model* (3.28) *and* $Y_i = \theta_0 + T_i\,\theta_1 + b_i + \sigma\,\epsilon_i$ *the sampling variances* $S_E^2 = {}^1\!/_{(N_E-1)} \sum_{i=1}^{N} \mathbb{1}_{\{T_i=1\}} \left(Y_i - {}^1\!/_{N_E} \sum_{j=1}^{N} Y_j\,\mathbb{1}_{\{T_j=1\}}\right)^2$ *and* $S_C^2 = {}^1\!/_{(N_C-1)} \sum_{i=1}^{N} \mathbb{1}_{\{T_i=-1\}} \left(Y_i - {}^1\!/_{N_C} \sum_{j=1}^{N} Y_j\,\mathbb{1}_{\{T_j=-1\}}\right)^2$ *are independent of the effects* θ_0 *and* θ_1.

Proof

The proof is executed for the sampling variance of the experimental group and can be easily transferred on the sampling variance of the control group. Model (3.28) is assumed for $Y_i = \theta_0 + T_i\,\theta_1 + b_i + \sigma\,\epsilon_i$. Thus, for S_E^2 it follows:

$$
\begin{aligned}
S_E^2 &= \frac{1}{N_E-1} \sum_{i=1}^{N} \mathbb{1}_{\{T_i=1\}} \left(Y_i - \frac{1}{N_E} \sum_{j=1}^{N} Y_j\,\mathbb{1}_{\{T_j=1\}}\right)^2 \\
&= \frac{1}{N_E-1} \sum_{i=1}^{N} \mathbb{1}_{\{T_i=1\}} \Big(\theta_0 + T_i\,\theta_1 + b_i + \sigma\,\epsilon_i \\
&\quad - \frac{1}{N_E} \sum_{j=1}^{N} \Big((\theta_0 + T_j\,\theta_1 + b_j + \sigma\,\epsilon_j)\ \mathbb{1}_{\{T_j=1\}}\Big)\Big)^2 \\
&= \frac{1}{N_E-1} \sum_{i=1}^{N} \mathbb{1}_{\{T_i=1\}} \Big(\theta_0 + \theta_1 + b_i + \sigma\,\epsilon_i \\
&\quad - \frac{1}{N_E} \sum_{j=1}^{N} \Big((\theta_0 + \theta_1 + b_j + \sigma\,\epsilon_j)\ \mathbb{1}_{\{T_j=1\}}\Big)\Big)^2 \\
&= \frac{1}{N_E-1} \sum_{i=1}^{N} \mathbb{1}_{\{T_i=1\}} \Big(b_i + \sigma\,\epsilon_i - \frac{1}{N_E} \sum_{j=1}^{N} \Big((b_j + \sigma\,\epsilon_j)\ \mathbb{1}_{\{T_j=1\}}\Big)\Big)^2 \\
&= \frac{1}{N_E-1} \sum_{i=1}^{N} \mathbb{1}_{\{T_i=1\}} \Big(\tilde{Y}_i - \frac{1}{N_E} \sum_{j=1}^{N} \Big(\tilde{Y}_j\,\mathbb{1}_{\{T_j=1\}}\Big)\Big)^2 .
\end{aligned}
$$

Hence, the variance of the experimental group depends only on $\tilde{Y}_i = b_i + \sigma\,\epsilon_i$ and is independent of θ_0 and θ_1.

□

A consequence of Lemma 3.6 is that the pooled variance S_p^2 is independent of the parameters θ_0 and θ_1. This independence is given by the following lemma.

Lemma 3.7: (S_p^2 independent of θ_0 and θ_1)
Under the assumption of Model (3.28) *and* $Y_i = \theta_0 + T_i\,\theta_1 + b_i + \sigma\,\epsilon_i$ *the pooled sampling variance* $S_p^2 = {((N_E-1)\,S_E^2+(N_C-1)\,S_C^2)}/{(N-2)}$ *from Equation* (3.29) *is independent of the parameters* θ_0 *and* θ_1.

Proof
From Lemma 3.6 it follows that S_E^2 and S_C^2 are independent of the parameters θ_0 and θ_1. Consequently, S_p^2 is independent of the parameters θ_0 and θ_1.

□

Finally, under the Random Allocation Rule and its property $N_E = N_C = {N}/{2}$ the test statistic W can be expressed as:

$$\begin{aligned} W &:= \frac{\bar{Y}_E - \bar{Y}_C}{\sqrt{\frac{N_E+N_C}{N_E\,N_C}}\,S_p} = \frac{\bar{Y}_E - \bar{Y}_C}{\sqrt{\frac{N/2+N/2}{N/2\,N/2}}\,S_p} = \frac{\bar{Y}_E - \bar{Y}_C}{\sqrt{\frac{4}{N}}\,S_p} \\ &= \frac{\bar{Y}_E - \bar{Y}_C}{\sqrt{\frac{2}{N}\,(S_E^2 + S_C^2)}} \\ &\text{with } S_p^2 = \frac{(N/2-1)\,S_E^2 + (N/2-1)\,S_C^2}{N-2} = \frac{1}{2}\,(S_E^2 + S_C^2)\,. \end{aligned} \tag{3.30}$$

Rosenkranz (2011) and Tamm and Hilgers (2014) investigated the difference between $\mathrm{Var}(\bar{Y}_E - \bar{Y}_C)$ and S_p^2 dependent on the randomization procedure when the model $Y_i = \theta_0 + T_i\,\theta_1 + \sigma\,\epsilon_i$ is assumed in the analysis, although the true model is $Y_i = \theta_0 + T_i\,\theta_1 + b_i + \sigma\,\epsilon_i$. Rosenkranz (2011) shows that S_p^2 is an asymptotically unbiased estimator of $\mathrm{Var}(\bar{Y}_E - \bar{Y}_C)$ under the Random Allocation Rule. However, the following result shows that S_p^2 is even an unbiased estimator of $\mathrm{Var}(\bar{Y}_E - \bar{Y}_C)$ under the Random Allocation Rule

Theorem 3.8:
Under the Random Allocation Rule and Model (3.28) *the squared denominator* ${2}/{N}\,(S_E^2 + S_C^2)$ *of* W *defined in Equation* (3.30) *is an unbiased estimator for* $Var(\bar{Y}_E - \bar{Y}_C)$.

For the proof of Theorem 3.8 the following two lemmas are necessary. First the variance of the squared denominator $^2/_N\,(S_E^2 + S_C^2)$ of W defined in Equation (3.30) is derived and afterwards the expectation of $\mathrm{Var}(\bar{Y}_E - \bar{Y}_C)$ is determined.

Lemma 3.9:
Under the assumption of Model (3.28) *and conducting the Random Allocation Rule, the variance of the numerator* $\bar{Y}_E - \bar{Y}_C$ *of* W *defined in Equation* (3.30) *is given by:*

$$Var(\bar{Y}_E - \bar{Y}_C) = \frac{4\,\sigma^2}{N} + \frac{4}{N\,(N-1)} \sum_{i=1}^{N} \left(b_i - \bar{b}\right)^2 . \tag{3.31}$$

Proof
Under the Random Allocation Rule the random variables $\bar{Y}_E$ and $\bar{Y}_C$ can be written as follows:

- $\bar{Y}_E = \dfrac{1}{N_E} \displaystyle\sum_{i=1}^{N} \frac{T_i + 1}{2}\, Y_i = \frac{2}{N} \sum_{i=1}^{N} \frac{T_i + 1}{2}\, Y_i$
- $\bar{Y}_C = \dfrac{1}{N_C} \displaystyle\sum_{i=1}^{N} \frac{1 - T_i}{2}\, Y_i = \frac{2}{N} \sum_{i=1}^{N} \frac{1 - T_i}{2}\, Y_i\,.$

Thus, the difference J between the effects of the responses in the two treatment groups under the assumption of Model (3.28) can be expressed as follows:

$$\begin{aligned} J = \bar{Y}_E - \bar{Y}_C &= \frac{2}{N} \sum_{i=1}^{N} T_i\, Y_i = \frac{2}{N} \sum_{i=1}^{N} T_i(\theta_0 + T_i\,\theta_1 + b_i + \sigma\,\epsilon_i) \\ &= \frac{2}{N} \left(\underbrace{\sum_{i=1}^{N} T_i\,\theta_0}_{=0} + \underbrace{\sum_{i=1}^{N} T_i^2\,\theta_1}_{=N\,\theta_1} + \sum_{i=1}^{N} T_i b_i + \sum_{i=1}^{N} T_i\,\sigma\,\epsilon_i \right) \\ &= 2\,\theta_1 + \frac{2}{N} \sum_{i=1}^{N} T_i\, b_i + \frac{2}{N} \sum_{i=1}^{N} T_i\,\sigma\,\epsilon_i\,. \end{aligned}$$

Applying the law of total variance according to Weiss (2005, pp. 385-386), the variance of J is given by:

$$\mathrm{Var}(J) = E\left(\mathrm{Var}(J|\boldsymbol{T})\right) + \mathrm{Var}\left(E(J|\boldsymbol{T})\right) .$$

In order to dissolve the expected conditional variance, the conditional variance $\mathrm{Var}(J|\boldsymbol{T} = \boldsymbol{t})$ is calculated as follows:

$$\begin{aligned}\mathrm{Var}(J|\boldsymbol{T} = \boldsymbol{t}) &= \mathrm{Var}\left(2\,\theta_1 + \frac{2}{N}\sum_{i=1}^{N} t_i\, b_i + \frac{2}{N}\sum_{i=1}^{N} t_i\,\sigma\,\epsilon_i\right) \\ &= \mathrm{Var}\left(\frac{2}{N}\sum_{i=1}^{N} t_i\,\sigma\,\epsilon_i\right) = \frac{4\,\sigma^2}{N^2}\sum_{i=1}^{N}\underbrace{t_i^2}_{=1}\underbrace{\mathrm{Var}(\epsilon_i)}_{=1} \\ &= \frac{4\,\sigma^2}{N} . \end{aligned} \tag{3.32}$$

The conditional variance $\mathrm{Var}(J|\boldsymbol{T} = \boldsymbol{t})$ is constant. Consequently, it is $E\left(\mathrm{Var}(J|\boldsymbol{T})\right) = {}^{(4\,\sigma^2)}/_{N}$. The conditional expectation $E(J|\boldsymbol{T} = \boldsymbol{t})$ can be transformed as follows:

$$\begin{aligned}E(J|\boldsymbol{T} = \boldsymbol{t}) &= E\left(2\,\theta_1 + \frac{2}{N}\sum_{i=1}^{N} t_i\, b_i + \frac{2}{N}\sum_{i=1}^{N} t_i\,\sigma\,\epsilon_i\right) \\ &= 2\,\theta_1 + \frac{2}{N}\sum_{i=1}^{N} t_i\, b_i + \frac{2}{N}\sum_{i=1}^{N} t_i\,\sigma\,\underbrace{E(\epsilon_i)}_{=0} \\ &= 2\,\theta_1 + \frac{2}{N}\sum_{i=1}^{N} t_i\, b_i =: d + b(\boldsymbol{t}) . \end{aligned} \tag{3.33}$$

The term $\mathrm{Var}(E(J|\boldsymbol{T})) = \mathrm{Var}(b(\boldsymbol{T})) = E((b(\boldsymbol{T}))^2) - (E(b(\boldsymbol{T}))^2 = E((b(\boldsymbol{T}))^2)$ can be expressed as:

$$\begin{aligned}\mathrm{Var}(b(\boldsymbol{T})) &= \mathrm{Var}\left(2\,\theta_1 + \frac{2}{N}\sum_{i=1}^{N} T_i\, b_i\right) = \frac{4}{N^2}\mathrm{Var}\left(\sum_{i=1}^{N} T_i\, b_i\right) \\ &= \frac{4}{N^2}E\left(\left(\sum_{i=1}^{N} T_i\, b_i\right)^2\right) = \frac{4}{N^2}E\left(\left(\sum_{i=1}^{N} T_i\, b_i\right)\left(\sum_{j=1}^{N} T_j\, b_j\right)\right) \\ &= \frac{4}{N^2}E\left(\sum_{i,j\in\{1,\ldots,N\}} T_i\, T_j\, b_i\, b_j\right)\end{aligned}$$

$$= \frac{4}{N^2} E\left(\sum_{i=1}^{N} T_i^2 \, b_i^2 + \sum_{\substack{i,j \in \{1,2,\ldots,N\} \\ i \neq j}} T_i \, T_j \, b_i \, b_j \right)$$

$$= \frac{4}{N^2} \left(\sum_{i=1}^{N} \underbrace{E\left(T_i^2\right)}_{=1} \, b_i^2 + \sum_{\substack{i,j \in \{1,2,\ldots,N\} \\ i \neq j}} E\left(T_i \, T_j\right) \, b_i \, b_j \right)$$

$$= \frac{4}{N^2} \left(\sum_{i=1}^{N} b_i^2 + \sum_{\substack{i,j \in \{1,2,\ldots,N\} \\ i \neq j}} E\left(T_i \, T_j\right) \, b_i \, b_j \right).$$

In Lemma A.3 in the appendix for the Random Allocation Rule $E\left(T_i \, T_j\right) = {}^{-1}/_{(N-1)}$ is derived. Furthermore, in Lemma A.2 in the appendix the following equality is shown:

$$\sum_{i=1}^{N} b_i^2 - \frac{1}{N-1} \sum_{\substack{i,j \in \{1,2,\ldots,N\} \\ i \neq j}} b_i \, b_j = \frac{N}{N-1} \sum_{i=1}^{N} (b_i - \bar{b})^2 \, .$$

Thus, it follows:

$$\begin{aligned} \mathrm{Var}(b(\boldsymbol{T})) &= \frac{4}{N^2} \left(\sum_{i=1}^{N} b_i^2 - \sum_{\substack{i,j \in \{1,2,\ldots,N\} \\ i \neq j}} \frac{1}{N-1} \, b_i \, b_j \right) \\ &= \frac{4}{N\,(N-1)} \sum_{i=1}^{N} \left(b_i - \bar{b} \right)^2 . \end{aligned} \tag{3.34}$$

Taking Equations (3.32) and (3.34) into account, the variance of J can be computed as follows:

$$\begin{aligned} \mathrm{Var}(J) &= E(\mathrm{Var}(J|\boldsymbol{T})) + \mathrm{Var}(E(J|\boldsymbol{T})) \\ &= \frac{4\,\sigma^2}{N} + \frac{4}{N\,(N-1)} \sum_{i=1}^{N} \left(b_i - \bar{b} \right)^2 . \end{aligned}$$

The derived result is the same as Result 2 found by Rosenkranz (2011).

□

Lemma 3.10:

Under the Random Allocation Rule and the assumption of Model (3.28)

the expectation of the squared denominator $^2/_N\,(S_E^2 + S_C^2)$ *of* W *defined in Equation* (3.30) *is given by:*

$$E\left(\frac{2}{N}\,(S_E^2 + S_C^2)\right) = \frac{4\,\sigma^2}{N} + \frac{4}{N\,(N-1)}\sum_{i=1}^{N}(b_i - \bar{b})^2\,. \tag{3.35}$$

The term $S_E^2 = {}^1/_{(N_E-1)}\sum_{i=1}^{N}\mathbb{1}_{\{T_i=1\}}\left(Y_i - {}^1/_{N_E}\sum_{j=1}^{N}Y_j\,\mathbb{1}_{\{T_j=1\}}\right)^2$ *is the variance of the patients' responses in treatment group* E *and* $S_C^2 = {}^1/_{(N_C-1)}$ $\sum_{i=1}^{N}\mathbb{1}_{\{T_i=-1\}}\left(Y_i - {}^1/_{N_C}\sum_{i=j}^{N}Y_j\,\mathbb{1}_{\{T_j=-1\}}\right)^2$ *is the variance of the patients' responses in treatment group* C*.*

Proof

In a first step the property $N_E = N_C = {}^N/_2$ is used to simplify the term $S_E^2 + S_C^2$:

$$\begin{aligned}
S_E^2 + S_C^2 &= \sum_{i=1}^{N}\left(\frac{1}{{}^N/_2 - 1}\,\frac{1+T_i}{2}\,\left(Y_i - \bar{Y}_E\right)^2 + \frac{1}{{}^N/_2 - 1}\,\frac{1-T_i}{2}\,\left(Y_i - \bar{Y}_C\right)^2\right)\\
&= \frac{2}{N-2}\sum_{i=1}^{N}\left(\frac{1+T_i}{2}\,\left(Y_i^2 - 2\,Y_i\,\bar{Y}_E + \bar{Y}_E^2\right) + \frac{1-T_i}{2}\,\left(Y_i^2 - 2\,Y_i\,\bar{Y}_C + \bar{Y}_C^2\right)\right)\\
&= \frac{2}{N-2}\sum_{i=1}^{N}\left(\frac{1+T_i}{2}\,Y_i^2 + \frac{1-T_i}{2}\,Y_i^2 - 2\,\frac{1+T_i}{2}\,Y_i\,\frac{2}{N}\,\left[\sum_{j=1}^{N}Y_j\,\frac{1+T_j}{2}\right]\right.\\
&\quad + \frac{1+T_i}{2}\,\frac{4}{N^2}\left[\sum_{k=1}^{N}Y_k\,\frac{1+T_k}{2}\right]\left[\sum_{j=1}^{N}Y_j\,\frac{1+T_j}{2}\right] - 2\,\frac{1-T_i}{2}\,Y_i\,\frac{2}{N}\,\left[\sum_{j=1}^{N}Y_j\,\frac{1-T_j}{2}\right]\\
&\quad \left. + \frac{1-T_i}{2}\,\frac{4}{N^2}\left[\sum_{k=1}^{N}Y_k\,\frac{1-T_k}{2}\right]\left[\sum_{j=1}^{N}Y_j\,\frac{1-T_j}{2}\right]\right)\\
&= \frac{2}{N-2}\left(\sum_{i=1}^{N}Y_i^2 - \frac{1}{N}\sum_{i,j\in\{1,2,\ldots,N\}}(1+T_i)\,(1+T_j)\,Y_i\,Y_j\right.\\
&\quad + \frac{1}{2\,N}\sum_{i,j\in\{1,2,\ldots,N\}}(1+T_i)\,(1+T_j)\,Y_i\,Y_j - \frac{1}{N}\sum_{i,j\in\{1,2,\ldots,N\}}(1-T_i)\,(1-T_j)\cdot\\
&\quad \left. Y_i\,Y_j + \frac{1}{2\,N}\sum_{i,j\in\{1,2,\ldots,N\}}(1-T_i)\,(1-T_j)\,Y_i\,Y_j\right)\\
&= \frac{2}{N-2}\left(\sum_{i=1}^{N}Y_i^2 - \frac{1}{N}\sum_{i,j\in\{1,2,\ldots,N\}}\left((1+T_i)\,(1+T_j) + (1-T_i)\,(1-T_j)\right)\cdot\right.
\end{aligned}$$

$$Y_i\,Y_j + \frac{1}{2\,N} \sum_{i,j\in\{1,2,\ldots,N\}} \left((1+T_i)\,(1+T_j) + (1-T_i)\,(1-T_j)\right)\,Y_i\,Y_j\Bigg)$$

$$= \frac{2}{N-2}\left(\sum_{i=1}^{N} Y_i^2 - \frac{1}{2\,N} \sum_{i,j\in\{1,2,\ldots,N\}} Y_i\,Y_j \cdot \right.$$

$$\left.\left((1+T_i)\,(1+T_j) + (1-T_i)\,(1-T_j)\right)\right).$$

According to Lemma 3.6 both variances S_E^2 and S_C^2 are independent of the parameters θ_1 and θ_0 in Model (3.28). Consequently, the parameters θ_1 and θ_0 do not influence the expectation $E(S_E^2 + S_C^2)$. Thus, in the following calculations the easier model $\tilde{Y}_i = b_i + \sigma\,\epsilon_i$ instead of $Y_i = \theta_0 + T_i\,\theta_1 + b_i + \sigma\,\epsilon_i$ is considered. Particularly, $\tilde{Y}_i$ is in comparison to Y_i independent of T_i. For the expectation $E(S_E^2 + S_C^2)$ it follows:

$$\begin{aligned}
\frac{N-2}{2}\,E(S_E^2 + S_C^2) &= \sum_{i=1}^{N} E(\tilde{Y}_i^2) - \frac{1}{2\,N} \sum_{i,j\in\{1,2,\ldots,N\}} E(\tilde{Y}_i\,\tilde{Y}_j) \cdot \\
&\qquad E\left((1+T_i)\,(1+T_j) + (1-T_i)\,(1-T_j)\right) \\
&= \sum_{i=1}^{N} E(\tilde{Y}_i^2) - \frac{1}{2\,N} \sum_{i,j\in\{1,2,\ldots,N\}} E(\tilde{Y}_i\,\tilde{Y}_j) \cdot \\
&\quad E\left(1 + T_j + T_i + T_i\,T_j + 1 - T_i - T_j + T_i\,T_j\right) \\
&= \sum_{i=1}^{N} E(\tilde{Y}_i^2) - \frac{1}{2\,N} \sum_{i,j\in\{1,2,\ldots,N\}} E(\tilde{Y}_i\,\tilde{Y}_j)\,2\,E\left(T_i\,T_j + 1\right) \\
&= \sum_{i=1}^{N} E(\tilde{Y}_i^2) - \frac{1}{N} \sum_{\substack{i,j\in\{1,2,\ldots,N\} \\ i\neq j}} E(\tilde{Y}_i\,\tilde{Y}_j)\,E\left(T_i\,T_j + 1\right) \\
&\quad - \frac{1}{N}\sum_{i=1}^{N} E(\tilde{Y}_i^2)\,\underbrace{E(T_i^2 + 1)}_{=2} \\
&= \left(1 - \frac{2}{N}\right)\sum_{i=1}^{N} E(\tilde{Y}_i^2) - \frac{1}{N} \sum_{\substack{i,j\in\{1,2,\ldots,N\} \\ i\neq j}} E(\tilde{Y}_i\,\tilde{Y}_j)\,E\left(T_i\,T_j + 1\right) \\
&= \frac{N-2}{N}\sum_{i=1}^{N} E(\tilde{Y}_i^2) - \frac{1}{N} \sum_{\substack{i,j\in\{1,2,\ldots,N\} \\ i\neq j}} E(\tilde{Y}_i\,\tilde{Y}_j)\,E\left(T_i\,T_j + 1\right)\,. \quad (3.36)
\end{aligned}$$

For $\tilde{Y}_i = b_i + \sigma\,\epsilon_i$ the individual terms can be expressed as:

- $\sum_{i=1}^{N} E(\tilde{Y}_i^2) = \sum_{i=1}^{N} E(b_i^2 + 2\,b_i\,\sigma\,\epsilon_i + \sigma^2\,\epsilon_i^2) = \sum_{i=1}^{N} b_i^2 + N\,\sigma^2$
- $E(\tilde{Y}_i\,\tilde{Y}_j) = E\left((b_i + \sigma\,\epsilon_i)\,(b_j + \sigma\,\epsilon_j)\right) = b_i\,b_j \quad (i \neq j)$
- $E(T_i\,T_j + 1) = E(T_i\,T_j) + 1 = \dfrac{-1}{N-1} + 1 = \dfrac{N-2}{N-1}\,.$

For the last transformation the covariance $\mathrm{Cov}(T_i, T_j) = {}^{-1}/_{(N-1)}$ for the Random Allocation Rule derived in Lemma A.3 in the appendix is used. Taking the results presented above into account, it follows:

$$\begin{aligned}
\frac{N-2}{2}\,E(S_E^2 + S_C^2) &= \frac{N-2}{N}\left(N\,\sigma^2 + \sum_{i=1}^{N} b_i^2\right) - \frac{1}{N}\sum_{\substack{i,j\in N\\ i\neq j}} b_i\,b_j\,\frac{N-2}{N-1} \\
&= (N-2)\,\sigma^2 + \frac{N-2}{N}\sum_{i=1}^{N} b_i^2 - \frac{N-2}{N\,(N-1)}\sum_{\substack{i,j\in\{1,2,\ldots,N\}\\ i\neq j}} b_i\,b_j \\
&= (N-2)\,\sigma^2 + \frac{N-2}{N}\left(\sum_{i=1}^{N} b_i^2 - \frac{1}{N-1}\sum_{\substack{i,j\in\{1,2,\ldots,N\}\\ i\neq j}} b_i\,b_j\right).
\end{aligned}$$

The following equality is shown in Lemma A.2 in the appendix:

$$\sum_{i=1}^{N} b_i^2 - \frac{1}{N-1}\sum_{\substack{i,j\in\{1,2,\ldots,N\}\\ i\neq j}} b_i\,b_j = \frac{N}{N-1}\sum_{i=1}^{N}(b_i - \bar{b})^2\,.$$

Using this equality the lemma follows:

$$\begin{aligned}
& \frac{N-2}{2}\,E(S_E^2 + S_C^2) = (N-2)\,\sigma^2 + \frac{N-2}{N-1}\sum_{i=1}^{N}(b_i - \bar{b})^2 \\
\Leftrightarrow \quad & E(S_E^2 + S_C^2) = 2\,\sigma^2 + \frac{2}{N-1}\sum_{i=1}^{N}(b_i - \bar{b})^2 \\
\Leftrightarrow \quad & E\left(\frac{2}{N}\,(S_E^2 + S_C^2)\right) = \frac{4\,\sigma^2}{N} + \frac{4}{N\,(N-1)}\sum_{i=1}^{N}(b_i - \bar{b})^2\,.
\end{aligned}$$

This result is different to the one Rosenkranz (2011) presents in the appendix of his paper. Even Result 3 in Rosenkranz (2011) is affected by this mistake.

For the sake of completeness, Result 3 from Rosenkranz (2011) can be corrected with $T_i, T_j \in \{-1, 1\}$ as follows:

$$\begin{aligned} E(\bar{V}) &:= E\left(\frac{2}{N}\left(S_E^2 + S_C^2\right)\right) \\ &= \frac{4\,\sigma^2}{N} + \frac{4}{N^2}\sum_{i=1}^{N} b_i^2 - \frac{4}{N^2\,(N-2)} \sum_{\substack{i,j \in \{1,2,\ldots,N\} \\ i \neq j}} b_i\,b_j\,E(T_i\,T_j + 1)\,. \end{aligned}$$

Particularly, the first term $(4\,\sigma^2)/N$ differs from the result derived by Rosenkranz (2011). The expectation $E(T_i\,T_j + 1)$ has to be calculated dependent on the randomization procedure.

□

The Theorem 3.8 follows directly from Lemma 3.9 and Lemma 3.10. In contrast to Rosenkranz (2011) it is obvious that the denominator from Equation (3.30) is an unbiased and not an asymptotically unbiased estimator for the variance of the nominator.

Consequently, using Student's t-test under the Random Allocation Rule the squared denominator of the test statistic W (see Equation (3.30)) under Model (3.28) is an unbiased estimator for the variance of the nominator. For other randomization procedures like the Truncated Binomial Design this property does not hold (see Rosenkranz (2011)). Tamm and Hilgers (2014) extensively reviewed the Permuted Block Randomization. An explicit formula for the difference between the terms $\mathrm{Var}(\bar{Y}_E - \bar{Y}_C)$ and $E(2/N\,(S_E^2 + S_C^2))$ dependent on the block constellation is derived. Below, Model (3.28) is considered and the dependence of the test statistic W according to Equation (3.30) on the randomization sequence $\boldsymbol{T}$ is of particular interest:

$$\begin{aligned} W(\boldsymbol{Y}, \boldsymbol{T}) &:= \frac{\bar{Y}_E(\boldsymbol{T}) - \bar{Y}_C(\boldsymbol{T})}{\sqrt{\frac{2}{N}\left(S_E^2(\boldsymbol{Y}, \boldsymbol{T}) + S_C^2(\boldsymbol{Y}, \boldsymbol{T})\right)}} \\ &= \frac{\frac{2}{N}\left(\sum_{i=1}^{N}(T_i\,\theta_0 + T_i^2\,\theta_1 + T_i\,b_i + T_i\,\sigma\,\epsilon_i)\right)}{\sqrt{\frac{2}{N}\left(S_E^2(\boldsymbol{Y}, \boldsymbol{T}) + S_C^2(\boldsymbol{Y}, \boldsymbol{T})\right)}} \\ &= \frac{\frac{2}{N}\left(\sum_{i=1}^{N}(\theta_1 + T_i\,\sigma\,\epsilon_i)\right) + b(\boldsymbol{T})}{\sqrt{\frac{2}{N}\left(S_E^2(\boldsymbol{Y}, \boldsymbol{T}) + S_C^2(\boldsymbol{Y}, \boldsymbol{T})\right)}} \end{aligned}$$

$$= \frac{\frac{2}{N}\left(\sum_{i=1}^{N}(\theta_1 + \sigma\,\epsilon_i)\right) + b(\boldsymbol{T})}{\sqrt{\frac{2}{N}\left(S_E^2(\boldsymbol{Y},\boldsymbol{T}) + S_C^2(\boldsymbol{Y},\boldsymbol{T})\right)}} \tag{3.37}$$

with $b(\boldsymbol{T}) = {}^{2}/_{N} \sum_{i=1}^{N} T_i\, b_i$ according to Equation (3.33). The last transformation is important, because the product $T_i\,\epsilon_i$ has the same distribution as ϵ_i. A proof is given in Lemma A.4 in the appendix. For the distribution of the expression ${}^{2}/_{N}\left(\sum_{i=1}^{N}(\theta_1 + \sigma\,\epsilon_i)\right)$ it follows:

$$\frac{2}{N}\left(\sum_{i=1}^{N}(\theta_1 + \sigma\,\epsilon_i)\right) = 2\,\theta_1 + \frac{2}{N\,\sigma}\sum_{i=1}^{N}\epsilon_i \sim \mathcal{N}\left(2\,\theta_1, \frac{4\,\sigma^2}{N}\right). \tag{3.38}$$

Next, it is shown that the term $b(\boldsymbol{T})$ is under the Random Allocation Rule asymptotically normally distributed. For that, the properties of the Simple Random Sampling are used.

Definition 3.11: (Simple Random Sampling (SRS))
Let N be the total sample number from which a subsample of size $k < N$ is randomly selected. For the Simple Random Sampling (see Hajek, 1981, pp. 49-52) the probability of inclusion of the ith patient with $i \in \{1, 2, \ldots, N\}$ in the subsample is given by:

$$\pi_i = \frac{k}{N} \quad \textit{for } 1 \leq i \leq N\,.$$

For the probability of inclusion of two patients i and j with $i, j \in \{1, 2, \ldots, N\}$ and $i \neq j$ in the subsample it follows:

$$\pi_{ij} = \frac{k\,(k-1)}{N\,(N-1)} \quad \textit{for } 1 \leq i \neq j \leq N\,.$$

For $k = {}^{N}/_{2}$ it follows that the Simple Random Sampling leads to $\binom{N}{N/2}$ possible samples. All of these samples are equiprobable. These possible samples correspond to the reference set of the Random Allocation Rule defined in Section 2.2.1. Thus, the following corollary can be established.

Corollary 3.12: (Equality between RAR and SRS for $k = {}^{N}/_{2}$)
Let N be the total sample number from which a subsample of size $k = {}^{N}/_{2}$ is selected. Hence, the Simple Random Sampling (SRS) with $k = {}^{N}/_{2}$ defined

in Definition 3.11 has the same properties as the Random Allocation Rule defined in Section 2.2.1.

A consequence of Corollary 3.12 is that the random allocation of the bias vector $\boldsymbol{b} \in \mathbb{R}^N$ (see Model (3.28)) to the two treatments can be understood as sampling from a finite population. The known and fixed bias vector $\boldsymbol{b} \in \mathbb{R}^N$ in Model (3.28) is randomly distributed to the two treatment groups dependent on $\boldsymbol{T}$. In this section the random allocation of the bias vector to the two treatment groups is investigated under the Random Allocation Rule. For this random allocation the elements of $\boldsymbol{b} \in \mathbb{R}^N$ are distributed equally to the two treatment groups. In the literature this approach is also termed as completely randomized design or (randomized) complete block design (with one block) (see Hinkelmann and Kempthorne, 2008, Chapter 6 and Chapter 9). Below, the central limit theorem for samples from a finite population, which was first introduced by Hajek (1960), is presented.

Theorem 3.13: (Central limit theorem (see Lehmann, 1999, p. 116))
Considering Model (3.28) *and let* $\boldsymbol{T}$ *be a random variable for a randomization sequence generated by the Random Allocation Rule. Furthermore, let* $\boldsymbol{b} = (b_1, b_2, \ldots, b_N)^T$ *be fixed. Thus for the term* $\bar{B}_E = {}^2\!/_N \sum_{i=1}^{N} \frac{T_i+1}{2} b_i$ *the central limit theorem holds:*

$$\frac{\bar{B}_E - E(\bar{B}_E)}{\sqrt{Var(\bar{B}_E)}} = \frac{\bar{B}_E - \bar{b}}{\sqrt{Var(\bar{B}_E)}} \underset{asymp.}{\sim} \mathcal{N}(0,1)\,, \tag{3.39}$$

provided that

(1) both the total sample number N *and the number of patients assigned to treatment* E *tend to infinity*

and the following condition is fulfilled:

(2) $\left(\sum_{i=1}^{N} \frac{T_i+1}{2}\right)/_N$ *is bounded away from* 0 *and* 1 *as* $N \to \infty$ *and*

$$\frac{\max\limits_{i\in\{1,2,\ldots,N\}} \left(b_i - \bar{b}\right)^2}{\sum_{i=1}^{N} \left(b_i - \bar{b}\right)^2} \to 0 \quad (N \to \infty) \tag{3.40}$$

For the Random Allocation Rule the number of patients assigned to treatment E is ${}^N\!/_2$. It follows that the condition 1 from Theorem 3.13 is fulfilled as

N approaches infinity. Furthermore, the term $\left(\sum_{i=1}^{N} \frac{T_i+1}{2}\right)/N$ is 0.5 under the usage of the Random Allocation Rule. Consequently, for any $\boldsymbol{b} = (b_1, b_2, \ldots, b_N)^T$, which satisfies Equation (3.40), the central limit theorem for samples from a finite population holds.

Proposition 3.14:
Assuming a linear time trend defined in Model (2.28) *with* $\boldsymbol{b} = \vartheta/N\,(0, 1, 2, \ldots, N-1)^T$*, the condition 2 in Theorem 3.13 is fulfilled.*

Proof
Under the assumption of a linear time trend, the following equality for the nominator in Equation (3.40) holds:

$$\begin{aligned}
\max_{i\in\{1,2,\ldots,N\}} \left(b_i - \bar{b}\right)^2 &= \max_{i\in\{1,2,\ldots,N\}} \left(b_i - \frac{\vartheta}{N^2}\,\frac{(N-1)\,N}{2}\right)^2 \\
&= \max_{i\in\{1,2,\ldots,N\}} \left(b_i - \frac{(N-1)\,\vartheta}{2\,N}\right)^2 \\
&= \left(0 - \frac{(N-1)\,\vartheta}{2\,N}\right)^2 \\
&= \frac{(N-1)^2}{N^2}\,\frac{\vartheta^2}{4} \to \frac{\vartheta^2}{4} \quad (N \to \infty)\,.
\end{aligned}$$

Using the results from Equation (3.18), the denominator of Equation (3.40) can be written as:

$$\sum_{i=1}^{N} \left(b_i - \bar{b}\right)^2 = \frac{\vartheta^2}{12\,N}\,(N^2 - 1) \to \infty \quad (N \to \infty)\,.$$

Combining the derived results for the nominator and denominator, it follows that the condition 2 of Theorem 3.13 is fulfilled.

□

Proposition 3.15:
Assuming a step time trend defined in Model (2.28) *with a step of the strength* ϑ *in the middle of a clinical trial and* $\boldsymbol{b} = \vartheta\,(\mathbf{0}_{N/2}^T, \mathbf{1}_{N/2}^T)^T$*, the condition 2 in Theorem 3.13 is fulfilled.*

Proof

Under the assumption of a step time trend, the nominator of Equation (3.40) can be expressed as:

$$\begin{aligned}\max_{i\in\{1,2,\ldots,N\}} \left(b_i - \bar{b}\right)^2 &= \max_{i\in\{1,2,\ldots,N\}} \left(b_i - \frac{\vartheta}{2}\right)^2 \\ &= \left(0 - \frac{\vartheta}{2}\right)^2 \to \frac{\vartheta^2}{4} \quad (N \to \infty)\,.\end{aligned}$$

Using the results from Equation (3.20), the denominator of Equation (3.40) can be written as:

$$\sum_{i=1}^{N} \left(b_i - \bar{b}\right)^2 = \frac{\vartheta^2\, N}{4} \to \infty \quad (N \to \infty)\,.$$

Bringing the derived results for the nominator and denominator together, it follows that the condition 2 of Theorem 3.13 is fulfilled.

□

The term $b(\boldsymbol{T})$ defined in Equation (3.37) can be transformed as:

$$b(\boldsymbol{T}) = {}^{2}\!/_{N} \sum_{i=1}^{N} T_i\, b_i = \bar{B}_E - \bar{B}_C\,.$$

For a fixed vector $\boldsymbol{b} = (b_1, b_2, \ldots, b_N)^T$ with $\bar{b} = \frac{1}{N}\sum_{i=1}^{N} b_i$ the following equality holds:

$$\begin{aligned}\bar{b} &= \frac{1}{2}\left(\bar{B}_E + \bar{B}_C\right) \\ \Leftrightarrow\ \bar{B}_C &= 2\,\bar{b} - \bar{B}_E\,.\end{aligned}$$

Using this property it follows:

$$b(\boldsymbol{T}) = \bar{B}_E - \bar{B}_C = \bar{B}_E - 2\,\bar{b} + \bar{B}_E = 2\,\bar{B}_E - 2\,\bar{b}\,.$$

For a linear time trend or a step time trend (see Definition 2.5) the term $\bar{b}$ is constant and Theorem 3.13 holds. Thus, for both of these time trends the term $b(\boldsymbol{T})$ is asymptotically normally distributed with expectation $\mu = E(2\,\bar{B}_E - 2\,\bar{b}) = 0$ and variance $\sigma^2 = \mathrm{Var}(2\,\bar{B}_E - 2\,\bar{b}) = 4\,\mathrm{Var}(\bar{B}_E)$.

Considering the test statistic $W(\boldsymbol{Y}, \boldsymbol{T})$ from Equation (3.37), it follows that the nominator is the sum of a normally distributed random variable and an asymptotically normally distributed random variable. Both random variables are independent. Using characteristic functions and applying continuous mapping theorem (see Billingsley, 1999, p. 21), it can be shown that the sum of a normally distributed random variable and an independent asymptotically normally distributed random variable is asymptotically normally distributed. Thus, the nominator of the test statistic $W(\boldsymbol{Y}, \boldsymbol{T})$ from Equation (3.37) is asymptotically normally distributed. According to Theorem 3.8 the denominator of $W(\boldsymbol{Y}, \boldsymbol{T})$ is an unbiased estimator for the standard deviation of the nominator. Furthermore, the denominator is a consistent estimator for the standard deviation σ (see Lehmann and Casella, 1998, p. 55). Finally, the nominator of the test statistic is asymptotically normally distributed with expectation $2\,\theta_1 = \mu_E - \mu_C$ and variance σ^2. The denominator converges against the constant σ for large N. Applying the Slutsky-Theorem (see Chow and Teicher, 2008, p. 272), it follows that the test statistic $W(\boldsymbol{Y}, \boldsymbol{T})$ is asymptotically normally distributed with variance $\sigma^2 = 1$ and expectation $2\,\theta_1 = \mu_E - \mu_C$. Assuming $2\,\theta_1 = \mu_E - \mu_C = 0$ (the null hypothesis of Student's t-test), it follows that the test statistic $W(\boldsymbol{Y}, \boldsymbol{T})$ is asymptotically normally distributed. Thus, the following lemma can be established.

Lemma 3.16: (Asymptotical control of the α-level under RAR)
Under the Random Allocation Rule and the assumption of a fixed bias vector $\boldsymbol{b} \in \mathbb{R}^N$ which fulfills the central limit theorem (see Theorem 3.13), the test statisic $W(\boldsymbol{Y}, \boldsymbol{T})$ from Equation (3.37) *is asymptotically normally distributed as N approaches infinity. Consequently, under the assumption of $\vartheta_1 = 0$ according to Model* (3.28) *the α-level with $\alpha \in (0, 1)$ of Student's two-tailed t-test (unadjusted for $\boldsymbol{b} \in \mathbb{R}^N$) for testing $H_0 : \theta_1 = 0$ vs. $H_1 : \theta_1 \neq 0$ is asymptotically preserved.*

Proof
Follows directly from the rejection probability of $W(\boldsymbol{Y}, \boldsymbol{T})$ according to Equation (3.37) with $\vartheta_1 = 0$ using Student's two-tailed t-test with level $\alpha \in (0, 1)$:

$$P(|W(\boldsymbol{Y}, \boldsymbol{T})| > t_{N-2,1-\alpha/2}) \; \to \alpha \; (N \to \infty)\,.$$

Two properties are used as N approaches infinity. First, the $(1-\alpha/2)$-quantile of the t-distribution with $N-2$ degrees of freedom converges against the $z_{1-\alpha/2}$ quantile, where $z_{1-\alpha/2}$ describes the $(1-\alpha/2)$-quantile of the standard normal distribution. Second, as N approaches infinity the central limit theorem defined in Theorem 3.13 holds and $W(\boldsymbol{Y},\boldsymbol{T}) \underset{\text{asymp.}}{\sim} \mathcal{N}(0,1)$. For that, the property that the denominator of $W(\boldsymbol{Y},\boldsymbol{T})$ in Equation (3.37) is a consistent estimator (see Lehmann and Casella, 1998, p. 55) for the standard deviation σ is used.

□

In Table 3.1 one example for $N=4$ and $\boldsymbol{b}=(0,1,2,3)^T$ is provided. It shows that the expected type-I-error probability of Student's t-test conditioned on the possible randomization sequences under the Random Allocation Rule is not controlled. The asymptotical considerations of $W(\boldsymbol{Y},\boldsymbol{T})$ do not hold for such a strong bias vector $\boldsymbol{b}\in\mathbb{R}^N$ and such a small sample size. Particularly, the term $\big(\max_{i\in\{1,2,\ldots,N\}}(b_i-\bar{b})^2\big)/\big(\sum_{i=1}^N(b_i-\bar{b})^2\big) = 1.5/5 = 0.3$ from Equation (3.40) is clearly greater than zero for the considered situation. Thus, the central limit theorem does not hold for $N=4$ and $\boldsymbol{b}=(0,1,2,3)^T$. Below, the asymptotical and exact power of Student's t-test is computed in situations when the patients' responses are affected by a linear time or a step trend.

Transferring the results to the situation of a linear time trend

In this subsection the theoretical considerations from above are transferred to the situation of a linear time trend. Table 3.5 shows the effect size $\Delta_0 := |\mu_E-\mu_C|/\sigma$ for a clinical trial which leads to a power of 80% using Student's t-test when the groups are equally sized at the end of the trial dependent on N. If not explicitly stated otherwise, the planned variance $\sigma\in\mathbb{R}_+$ is assumed to be one. If the clinical trial is not affected by chronological bias,

	Sample size N				
	4	12	50	200	1000
Δ_0	5.6535	1.7955	0.8087	0.3981	0.1774

Table (3.5): Effect size Δ_0 for achieving a power of 80% using Student's t-test, when the groups are equally sized at the end of a trial dependent on N.

the variance of $\bar{Y}_E - \bar{Y}_C$ is $\sigma_0^2 = {}^{(4\,\sigma^2)}/_N$ (see Equation (3.32)). In Table 3.6 the ratio of the caused variance by the linear time trend $\tilde{\sigma}^2$ and σ_0^2 dependent on the strength ϑ and the sample size N is depicted. The caused variance by the linear time trend $\tilde{\sigma}^2$ (see Lemma 3.9) and the ratio ${}^{\tilde{\sigma}^2}/_{\sigma_0^2}$ are computed dependent on the bias vector $\boldsymbol{b} = {}^{\vartheta}/_N\,(0, 1, 2, \ldots, N-1)^T$:

$$\begin{aligned}
\operatorname{Var}\left(\bar{Y}_E - \bar{Y}_C\right) &= \frac{4\,\sigma^2}{N} + \frac{4}{N\,(N-1)} \underbrace{\sum_{i=1}^{N} (b_i - \bar{b})^2}_{\text{see Eq. (3.18)}} \\
&= \sigma_0^2 + \frac{4}{N\,(N-1)}\,\frac{\vartheta^2}{12\,N}(N^2-1) \\
&= \sigma_0^2 + \frac{(N+1)\,\vartheta^2}{3\,N^2} = \tilde{\sigma}^2 \qquad (3.41) \\
\Rightarrow \quad \frac{\tilde{\sigma}^2}{\sigma_0^2} &= 1 + \frac{(N+1)\,\vartheta^2}{3\,N^2}\,\frac{N}{4\,\sigma^2} = 1 + \frac{(N+1)\,\vartheta^2}{12\,N\,\sigma^2}\,. \qquad (3.42)
\end{aligned}$$

The derived result shows that $\tilde{\sigma}^2$ is in comparison to σ_0^2 inflated dependent on ϑ and N. Thus, the linear time trend leads to a higher variability in the data for the treatment comparison. The ratio ${}^{\tilde{\sigma}^2}/_{\sigma_0^2}$ for a linear time trend for different shapes of ϑ, different sample sizes N, and fixed $\sigma^2 = 1$ is shown in Table 3.6. It is visible that the greater the parameter ϑ of the linear time trend is, the greater the ratio ${}^{\tilde{\sigma}^2}/_{\sigma_0^2}$ becomes. Furthermore, as N approaches infinity the ratio ${}^{\tilde{\sigma}^2}/_{\sigma_0^2}$ converges against the constant $1 + {}^{\vartheta}/_{(12\,\sigma^2)}$, which is dependent on ϑ and $\sigma \in \mathbb{R}^+$ (see Equation (3.42)).

	Sample size N				
ϑ	4	12	50	200	10000
0.50	1.0260	1.0226	1.0212	1.0209	1.0209
1.00	1.1042	1.0903	1.0850	1.0837	1.0834
2.00	1.4167	1.3611	1.3400	1.3350	1.3337

Table (3.6): The ratio ${}^{\tilde{\sigma}^2}/_{\sigma_0^2}$ according to Equation (3.42) for a linear time trend dependent on different shapes of ϑ, different sample sizes N, and fixed $\sigma^2 = 1$.

In what follows, the influence of a linear time trend under the Random Allocation Rule on the power of Student's t-test is investigated. Analogously to the presented average value of the power in Table 3.2, an asymptotical value of the power under the Random Allocation Rule using Student's t-test

ϑ	$N = 4$		$N = 12$		$N = 50$		$N = 200$		$N = 10.000$
	exact	asymp.	exact	asymp.	sim.	asymp.	sim.	asymp.	asymp.
0.50	0.7919	0.7919	0.7931	0.7913	0.7920	0.7917	0.7918	0.7918	0.7919
1.00	0.7680	0.7683	0.7655	0.7655	0.7671	0.7672	0.7679	0.7677	0.7679
2.00	0.6782	0.6837	0.6696	0.6717	0.6766	0.6774	0.6787	0.6790	0.6795

Table (3.7): Comparison (asymp. vs. exact/sim.) of the achieved power of the Random Allocation Rule using Student's t-test in case of a linear time trend (assuming $W(\boldsymbol{Y}, \boldsymbol{T}) \sim t_{N-2}$). The planned power $1 - \beta$ is 80%.

is computed. Afterwards, the exact/simulated average value of the power is compared to the corresponding asymptotical value of the power. For five different sample sizes the effect sizes, which lead to a power of 80% when the groups are equally sized at the end of the clinical trial, are shown in Table 3.5. The presented effect sizes Δ_0 are calculated under σ^2 (no additional time trend in the data). The derived values for the effect size Δ_0 lead to noncentrality parameters $\delta_0 = \Delta_0 \sqrt{N/(4\,\sigma^2)} = \Delta_0/\sigma_0$ according to Equation (3.10) with $N_E = N_C = N/2$ and $\Delta_0 = 2\,\theta_1$. However, the noncentrality parameter $\tilde{\delta} = \Delta_0/\tilde{\sigma}$ with $\tilde{\sigma} \in \mathbb{R}_+$ from Equation (3.41) is underlying in the data in the situation of a time trend.

As a first approach it is in concordance with Proposition 3.14 assumed that as N approaches infinity, the test statistic $W(\boldsymbol{Y}, \boldsymbol{T})$ from Equation (3.37) is normally distributed with variance one. Table A.1 in the appendix shows the achieved average power (asymptotical and simulated/exact) for the Random Allocation Rule using Student's t-test dependent on the shape of ϑ and the sample size N. For the exact values, the whole reference set of the Random Allocation Rule is computed and the achieved power of the individual randomization sequences is summarized (approach analogous to Table 3.2). For larger sample sizes 100 000 randomization sequences from the Random Allocation Rule are generated with the `randomizeR` package (Schindler et al., 2015) and the average power of Student's t-test is computed. Table A.1 in the appendix shows that the asymptotical consideration does not apply to small N including $N = 50$. Even for $N = 200$ the asymptotical value for the power of Student's t-test is inconsistent with the simulated value. Table 3.7 shows the distribution of $W(\boldsymbol{Y}, \boldsymbol{T})$ approximated by the t-distribution, whose density function has heavier tails in comparison to the density function of the normal distribution. It is assumed that on the

one hand the term $b(\boldsymbol{T})$ from Equation (3.37) is asymptotically normally distributed and on the other hand the distribution of the test statistic itself is t-distributed with $N-2$ degrees of freedom. In contrast to the situation before the distribution of the test statistic $W(\boldsymbol{Y}, \boldsymbol{T})$ did not converged against a normal distribution, yet. From Table 3.7 it can be concluded that this asymptotical consideration is the better type of asymptotical consideration and even holds for a small sample size like $N = 4$. Due to the fact that δ_0 is greater than $\tilde{\delta}$, the power is in comparison to the planned power of 80% in all situations deflated. The greater the amount of ϑ of the linear time trend is, the greater the deviation of the planned power $1 - \beta = 0.8$ becomes.

Next, the rejection probability of the test statistic W conditioned on the distribution of $b(\boldsymbol{T})$ dependent on the strength ϑ of the bias vector $\boldsymbol{b} = {}^{\vartheta}/_{N}\,(0, 1, 2, \ldots, N-1)^T$ is investigated. Proposition 3.14 proves that the central limit theorem holds for the considered linear time trend. Thus, corresponding asymptotical quantiles of the distribution of $b(\boldsymbol{T})$ according to Equation (3.37) can be derived. The variance of $b(\boldsymbol{T})$ in case of a linear time trend with $\boldsymbol{b} = {}^{\vartheta}/_{N}\,(0, 1, 2, \ldots, N-1)^T$ is given by:

$$\operatorname{Var}(b(\boldsymbol{T})) \underset{\text{Eq. (3.34)}}{=} \frac{4}{N(N-1)} \underbrace{\sum_{i=1}^{N} (b_i - \bar{b})^2}_{\text{see Eq. (3.18)}} = \frac{(N+1)\,\vartheta^2}{3\,N^2} = \sigma^2_{\text{linT}}\,. \tag{3.43}$$

Under the assumption $b(\boldsymbol{T}) \underset{\text{asymp.}}{\sim} \mathcal{N}(0, \sigma^2_{\text{linT}})$, the asymptotically greatest absolute deviations of $b(\boldsymbol{T})$ from zero can be derived. Knowing worst case deviations of $b(\boldsymbol{T})$ from zero the corresponding noncentrality parameter δ under the null hypothesis and alternative hypothesis can be computed. For a given δ the corresponding value of λ can be calculated according to Proposition 3.5. The vector $\boldsymbol{b} = {}^{\vartheta}/_{N}\,(0, 1, 2, \ldots, N-1)^T$ and the parameter ϑ are assumed to be known. The boundary for $\delta^2 + \lambda = {}^{(\vartheta^2\,(N^2-1))}/_{(12\,N\,\sigma^2)}$ in case of a linear time trend is given by Equation (3.18). Table 3.8 shows corresponding asymptotical quantiles for $|b(\boldsymbol{T})|$ and the resulting parameters δ and λ for $N = 50$ and $\vartheta = 1$. The situation of a two-tailed Student's t-test is considered. Thus, under the null hypothesis ($H_0 : \theta_1 = 0$) deviations of δ from zero in both directions lead to the same distorted type-I-error probability. Consequently, the distributions of $|\delta|$ and $|b(\boldsymbol{T})|$ are of interest. Table A.2 in the appendix shows the asymptotical quantiles of the noncen-

	$\tilde{q}_{0.01}$	$\tilde{q}_{0.05}$	$\tilde{q}_{0.25}$	$\tilde{q}_{0.5}$	$\tilde{q}_{0.75}$	$\tilde{q}_{0.95}$	$\tilde{q}_{0.99}$
$\lvert b(\boldsymbol{T})\rvert$	0.0010	0.0052	0.0263	0.0556	0.0949	0.1616	0.2124
δ_{H_0}	0.0037	0.0183	0.0929	0.1966	0.3354	0.5714	0.7510
λ	4.1650	4.1647	4.1564	4.1263	4.0525	3.8385	3.6010

Table (3.8): Asymptotical quantiles for $|b(\boldsymbol{T})|$, δ, and λ. A linear time trend is assumed for $N = 50$, $\vartheta = 1$, and $\sigma^2 = 1$ under the null hypothesis ($\theta_1 = 0$).

trality parameters δ and λ for $N = 50$ under the alternative hypothesis (planned power 80%). This table is derived analogously to Table 3.8.

The asymptotical quantiles of the noncentrality parameters presented in Tables 3.8 and A.2 in the appendix are used to derive the corresponding rejection probabilities under Student's t-test conditioned on the distribution of $b(\boldsymbol{T})$. Table 3.9 shows for $N \in \{12, 50, 200\}$ the quantiles of the asymptotical and the exact/simulated rejection probability. For the exact values, the quantiles of the whole reference set of the Random Allocation Rule are computed. For larger sample sizes of $N \geq 50$, 100 000 randomization sequences from the Random Allocation Rule are generated with the `randomizeR` package (Schindler et al., 2015) and the corresponding quantiles are calculated. It is assumed that the clinical trial is planned with size $\alpha = 0.05$ and power $1-\beta = 0.8$. For the calculations of the asymptotical quantiles of the distorted type-I-error probability α_{CB} and power $1 - \beta_{\text{CB}}$ the Equations (3.23) and (3.24) are used. To given an example, for the derived asymptotical value of the 0.95-quantile for α_{CB} and $N = 50$ in Table 3.9 the noncentrality parameters $\delta = 0.5714$ and $\lambda = 3.8385$ from Table 3.8 are used. The greater N is, the better the accordance between the simulated and the asymptotical quantiles becomes. In general, the distributions of both the α_{CB} values and the $1 - \beta_{\text{CB}}$ values converge against stable distributions. The differences of the values of the corresponding quantiles for $N = 50$ and $N = 200$ are only marginal. Table A.3 in the appendix shows the quantiles of the rejection probability under Student's t-test conditioned on the distribution of $b(\boldsymbol{T})$ in case of linear time trend with strength $\vartheta = 2$. The asymptotical quantiles of the rejection probabilities are in accordance with the exact/simulated quantiles even for $N = 12$. Table 3.9 and Table A.3 in the appendix show that the greater the value of the parameter ϑ is, the greater the range of the distributions of both the α_{CB} values and the $1 - \beta_{\text{CB}}$ values becomes.

Quantile	$N = 12$				$N = 50$				$N = 200$			
	α_{CB}		$1 - \beta_{\text{CB}}$		α_{CB}		$1 - \beta_{\text{CB}}$		α_{CB}		$1 - \beta_{\text{CB}}$	
	exact	asymp.	exact	asymp.	sim.	asymp.	sim.	asymp.	sim.	asymp.	sim.	asymp.
$\tilde{q}_{0.01}$	0.0416	0.0415	0.5742	0.5482	0.0414	0.0413	0.5419	0.5374	0.0413	0.0413	0.5385	0.5362
$\tilde{q}_{0.05}$	0.0416	0.0415	0.6319	0.6190	0.0414	0.0414	0.6158	0.6134	0.0414	0.0414	0.6123	0.6130
$\tilde{q}_{0.25}$	0.0425	0.0423	0.7171	0.7131	0.0422	0.0422	0.7128	0.7140	0.0422	0.0422	0.7147	0.7144
$\tilde{q}_{0.5}$	0.0450	0.0452	0.7707	0.7718	0.0451	0.0452	0.7747	0.7757	0.0452	0.0452	0.7769	0.7766
$\tilde{q}_{0.75}$	0.0521	0.0525	0.8199	0.8243	0.0528	0.0526	0.8286	0.8294	0.0527	0.0527	0.8305	0.8303
$\tilde{q}_{0.95}$	0.0740	0.0750	0.8829	0.8874	0.0747	0.0749	0.8905	0.8911	0.0749	0.0748	0.8917	0.8917
$\tilde{q}_{0.99}$	0.0934	0.1018	0.9167	0.9223	0.0986	0.1006	0.9224	0.9239	0.0998	0.1002	0.9238	0.9241

Table (3.9): Quantiles of the rejection probability under Student's t-test conditioned on the distribution of $b(\boldsymbol{T})$ using the Random Allocation Rule in case of a linear time trend with strength $\vartheta = 1$, $\sigma^2 = 1$, and planned $\alpha = 0.05$ and $1 - \beta = 0.8$.

Transferring the results to the situation of a step time trend

In comparison to the linear time trend the step time trend defined in Equation (2.26) depends on two parameters: The strength ϑ and the time point of the step $s \in \mathbb{N}$. In Equation (3.19) the following equality is derived for $\boldsymbol{b} = \vartheta\,(\mathbf{0}_s^T, \mathbf{1}_{N-s}^T)^T$:

$$\sum_{i=1}^{N}(b_i - \bar{b})^2 = \vartheta^2\, s\left(\frac{N-s}{N}\right).$$

This term is maximized for $s = {}^{N}\!/_{2}$ with $\max\limits_{s\in\{1,2,\ldots,N\}}\left(\sum_{i=1}^{N}(b_i - \bar{b})^2\right) = {}^{(\vartheta^2\,N)}\!/_{4}$ (see Equation (3.20)). The following calculations are based on this worst case scenario of a step time trend. The caused variance by a step time trend with $s = {}^{N}\!/_{2}$ and $\boldsymbol{b} = \vartheta\,(\mathbf{0}_{N/2}^T, \mathbf{1}_{N/2}^T)^T$ for the variance of $\bar{Y}_E - \bar{Y}_C$ can be

	Sample size N				
ϑ	4	12	50	200	1000
0.50	1.0833	1.0682	1.0638	1.0628	1.0625
1.00	1.3333	1.2727	1.2551	1.2513	1.2500
2.00	2.3333	2.0909	2.0204	2.0050	2.0001

Table (3.10): The ratio $\tilde{\sigma}_0^2/\sigma_0^2$ according to Equation (3.45) for a step time trend dependent on different shapes of ϑ, different sample sizes N, and fixed $\sigma^2 = 1$.

derived analogously to Equation (3.41):

$$\begin{aligned}
\operatorname{Var}\left(\bar{Y}_E - \bar{Y}_C\right) &= \frac{4\,\sigma^2}{N} + \frac{4}{N\,(N-1)} \underbrace{\sum_{i=1}^{N} (b_i - \bar{b})^2}_{\text{see Eq. (3.20)}} \\
&= \sigma_0^2 + \frac{4}{N\,(N-1)}\,\frac{\vartheta^2\,N}{4} \\
&= \sigma_0^2 + \frac{\vartheta^2}{N-1} = \tilde{\sigma}^2 \qquad (3.44) \\
\Rightarrow \quad \frac{\tilde{\sigma}^2}{\sigma_0^2} &= 1 + \frac{\vartheta^2\,N}{(N-1)\,4\,\sigma^2}\,. \qquad (3.45)
\end{aligned}$$

As in the situation of a linear time trend, the caused variance by a step time trend $\tilde{\sigma}^2$ is inflated. An unadjusted step time trend in the data leads to a higher variability for the treatment comparison. Table 3.10 illustrates the ratio $\tilde{\sigma}_0^2/\sigma_0^2$ for a step time trend dependent on different shapes of ϑ, different sample sizes N, and fixed $\sigma^2 = 1$. As N approaches infinity, the ratio $\tilde{\sigma}_0^2/\sigma_0^2$ converges against the constant $1 + \vartheta^2/(4\,\sigma^2)$. The greater the amount of the step ϑ is, the greater the caused variance by the step time trend becomes. For $N = 200$ and $\vartheta = 2$ the caused variance $\tilde{\sigma}^2$ by a step time trend is already doubled in comparison to σ_0^2.

Table 3.11 shows the achieved average power of the Random Allocation Rule (asymptotical and simulated/exact) using Student's t-test dependent on the shape of ϑ and the sample size N. The table is derived analogously to Table 3.7 and can be interpreted in the same manner. The planned power $1 - \beta$ of the trial is 80%. Thus, a step time trend decreases the power of Student's t-test dependent on ϑ and N. For the calculations of the exact/simulated values the `randomizeR` package (Schindler et al., 2015)

ϑ	$N = 4$		$N = 12$		$N = 50$		$N = 200$		10.000
	exact	asymp.	exact	asymp.	sim.	asymp.	sim.	asymp.	asymp.
0.50	0.7743	0.7745	0.7739	0.7738	0.7753	0.7753	0.7756	0.7757	0.7758
1.00	0.7011	0.7047	0.6995	0.7006	0.7055	0.7057	0.7066	0.7070	0.7075
2.00	0.4728	0.5128	0.4782	0.4933	0.5019	0.5045	0.5072	0.5074	0.5084

Table (3.11): Comparison (asymp. vs. exact/sim.) of the achieved power of the Random Allocation Rule using Student's t-test dependent on the shape of ϑ and the sample size N in case of a step time trend (assuming $W(\boldsymbol{Y}, \boldsymbol{T}) \sim t_{N-2}$).

is used. In case of a simulated value 100 000 randomization sequences are generated at random from the Random Allocation Rule. The greater N and the lower ϑ are, the better is the accordance between the asymptotical and exact/simulated value.

Finally, the variance of $b(\boldsymbol{T})$ in case of a step time trend $\text{Var}(b(\boldsymbol{T})) = {}^{\vartheta^2}/_{(N-1)}$ is used to derive asymptotical quantiles for the rejection probability under Student's t-test conditioned on the distribution of $b(\boldsymbol{T})$ (approach analogous to Table 3.9). Table 3.12 shows the resulting quantiles (asymp. and exact/sim.) of the rejection probability for the Random Allocation Rule using Student's t-test for a step time trend with strength $\vartheta = 1$. Initially, the corresponding asymptotical quantiles for the noncentrality parameter δ are computed. Proposition 3.15 proves that the central limit theorem holds for the considered step time trend. Afterwards the boundary for $\delta^2 + \lambda = {}^{(\vartheta^2\, N)}/_{(\sigma^2\, 4)}$ derived in Equation (3.20) is used to calculate the corresponding value for λ. For the presented exact values, the quantiles of the whole reference set of the Random Allocation Rule are computed. For larger sample sizes the `randomizeR` package (Schindler et al., 2015) is used to generate 100 000 randomization sequences from the Random Allocation Rule and to compute the corresponding quantiles of the rejection probability under Student's t-test. Table 3.12 can be interpreted in the same manner as Table 3.9. The asymptotical quantiles are in good agreement with the corresponding simulated/exact quantiles. The greatest differences between the asymptotical quantiles and their simulated/exact quantiles are observed for $N = 12$. In general, the distribution of both the α_{CB} values and the $1-\beta_{\text{CB}}$ values seem to converge against a stable distribution. The differences

Quantile	$N = 12$				$N = 50$				$N = 200$			
	α_{CB}		$1-\beta_{\text{CB}}$		α_{CB}		$1-\beta_{\text{CB}}$		α_{CB}		$1-\beta_{CB}$	
	exact	asymp.	exact	asymp.	sim.	asymp.	sim.	asymp.	sim.	asymp.	sim.	asymp.
$\tilde{q}_{0.01}$	0.0287	0.0285	0.3655	0.3048	0.0299	0.0284	0.2697	0.2919	0.0284	0.0284	0.3045	0.2906
$\tilde{q}_{0.05}$	0.0287	0.0286	0.5311	0.4216	0.0299	0.0285	0.4672	0.4160	0.0284	0.0285	0.4079	0.4160
$\tilde{q}_{0.25}$	0.0287	0.0304	0.5311	0.5983	0.0299	0.0304	0.5745	0.6051	0.0300	0.0304	0.6280	0.6070
$\tilde{q}_{0.5}$	0.0531	0.0372	0.7105	0.7141	0.0426	0.0375	0.7693	0.7259	0.0348	0.0376	0.7282	0.7284
$\tilde{q}_{0.75}$	0.0531	0.0552	0.8618	0.8144	0.0698	0.0557	0.8451	0.8256	0.0548	0.0558	0.8130	0.8277
$\tilde{q}_{0.95}$	0.1464	0.1169	0.8618	0.9195	0.1144	0.1144	0.9027	0.9232	0.1153	0.1138	0.9270	0.9238
$\tilde{q}_{0.99}$	0.1464	0.1992	0.9544	0.9633	0.1790	0.1866	0.9693	0.9621	0.1793	0.1842	0.9589	0.9620

Table (3.12): Quantiles of the rejection probability under Student's t-test conditioned on the distribution of $b(\boldsymbol{T})$ using the Random Allocation Rule in case of a step time trend with strength $\vartheta = 1$, $\sigma^2 = 1$, and planned $\alpha = 0.05$ and $1 - \beta = 0.8$.

of the values of the corresponding quantiles for $N = 50$ and $N = 200$ are only marginal. Table A.4 in the appendix shows the quantiles of the rejection probability conditioned on the distribution of $b(\boldsymbol{T})$ under the the Random Allocation Rule in case of step time trend with strength $\vartheta = 2$. The asymptotical quantiles of the rejection probabilities conditioned on the distribution of $b(\boldsymbol{T})$ are still in good agreement with the exact/simulated quantiles even for $N = 12$. Table 3.12 and Table A.4 in the appendix show the greater the value of the parameter ϑ is, the greater the range of the distribution of the α_{CB} values and the $1 - \beta_{\mathrm{CB}}$ values becomes.

Chapter 4

Desirability index

In Chapters 2 and 3 different criteria for the assessment of randomization procedures in the presence of chronological and selection bias were introduced. One aim of this thesis is to give a general statement of a randomization procedure with respect to its behavior in the presence of both selection and chronological bias. Up to now, the performance of a randomization procedures is only investigated with respect to one objective like the susceptibility to selection bias or balancing behavior. The objectives are not assessed simultaneously. Neither Berger et al. (2015) nor Zhao et al. (2012) managed to recommend a randomization procedure when both the susceptibility to selection bias and the balancing behavior are important aspects. Antognini et al. (2015) derived an optimal Efron's biased coin design with respect to balancing behavior and susceptibility to selection bias. Nevertheless, no comparison to other randomization procedures is presented. Rosenberger and Lachin (2016, Chapter 8) suggest to plot the mean values of two criteria's distributions against each other. A so called trade-off plot can be used to find an appropriate randomization procedure in the situation of two criteria. However, no solution is found for the situation when three or more criteria are of importance.

The main difficulty is that the criteria use different scales. It remains unclear how to combine the criteria to one optimality score. Furthermore, it would be good if the importance of the criteria should be adjustable dependent on the practical situation. In what follows, desirability functions and a desirability index are introduced. Based on this index the different criteria are combinable and it is possible to take the importance of the criteria into

account. Below, the desirability function according to Derringer and Suich (1980) is introduced. Two approaches of the application of a desirability function on a criterion are discussed. When using the first approach, the mean value of the criterion's distribution is assessed and the second approach applies the desirability function to the individual values of the criterion's distribution. The combination and properties of desirability scores are discussed in the second and third section.

4.1 Desirability functions according to Derringer-Suich

Definition 4.1: (Desirability function)
A function with the property $d : \mathbb{R} \to [0, 1]$ *is called a desirability function. The value* $d(x)$ *is called desirability score or desirability value of* x*. The desirability score of a given* x *should be chosen such that it can be interpreted as recommended in Table 4.1.*

In this thesis the main emphasis is put on the class of the desirability functions introduced by Derringer and Suich (1980). The key property of this class is that all x smaller/greater than a lower/an upper specified border are mapped to zero. The upper specified border is called upper specification limit (USL) and the lower specified border is called lower specification limit (LSL). If the value of x corresponds to its optimal desired value, the desirability function $d(x)$ is one. This optimal desired value is called target value (TV). Below,

$d(x)$	Meaning
1	Ultimate satisfaction - improvement beyond this value has no benefit.
$[0.8, 1)$	Excellent satisfaction.
$[0.6, 0.8)$	Good satisfaction.
$[0.4, 0.6)$	Acceptable, but poor satisfaction.
$[0.3, 0.4)$	Borderline.
$(0, 0.3)$	Unacceptable satisfaction.
0	Completely unacceptable satisfaction.

Table (4.1): Interpretation of desirability scores $d(x)$ according to Harrington Jr. (1965).

the one-sided and two-sided desirability functions according to Derringer and Suich (1980) are defined.

Definition 4.2: (Desirability functions after Derringer and Suich (1980))
Let x, TV, LSL, $USL \in \mathbb{R}$ and $b_r, b_l \in \mathbb{R}^+$. Thus, the

a) Left-One-sided Derringer-Suich desirability function is defined as:

$$d(x) = \begin{cases} 0, & \text{if } x \leq LSL \\ \left[\frac{x-LSL}{TV-LSL}\right]^{b_l}, & \text{if } LSL < x < TV \\ 1, & \text{if } x \geq TV \end{cases}. \tag{4.1}$$

b) Right-One-sided Derringer-Suich desirability function is defined as:

$$d(x) = \begin{cases} 0, & \text{if } x \geq USL \\ \left[\frac{USL-x}{USL-TV}\right]^{b_r}, & \text{if } TV < x < USL \\ 1, & \text{if } x \leq TV \end{cases}. \tag{4.2}$$

c) Two-sided Derringer-Suich desirability function is defined as:

$$d(x) = \begin{cases} \left[\frac{x-LSL}{TV-LSL}\right]^{b_l}, & \text{if } LSL \leq x \leq TV \\ \left[\frac{USL-x}{USL-TV}\right]^{b_r}, & \text{if } TV < x \leq USL \\ 0, & \text{if } x > USL \text{ or } x < LSL \end{cases}. \tag{4.3}$$

The values $b_l \in \mathbb{R}^+$ and $b_r \in \mathbb{R}^+$ are exponential factors for weighting standardized deviations from the target value. Figure 4.1 shows several settings of the two-sided Derringer-Suich desirability functions with different values of b_l and b_r. For $b_l = b_r = 1$ the desirability scores decrease linearly. For $b_l = b_r = 0$ every value of $d(x)$ with $x \in [\text{LSL}, \text{USL}]$ is mapped to one, otherwise to zero.

Application of a desirability function

In this section two approaches of the application of a desirability function on a given criterion are presented. Initially, a suitable desirability function for the investigated criterion has to adjusted. The first approach applies the

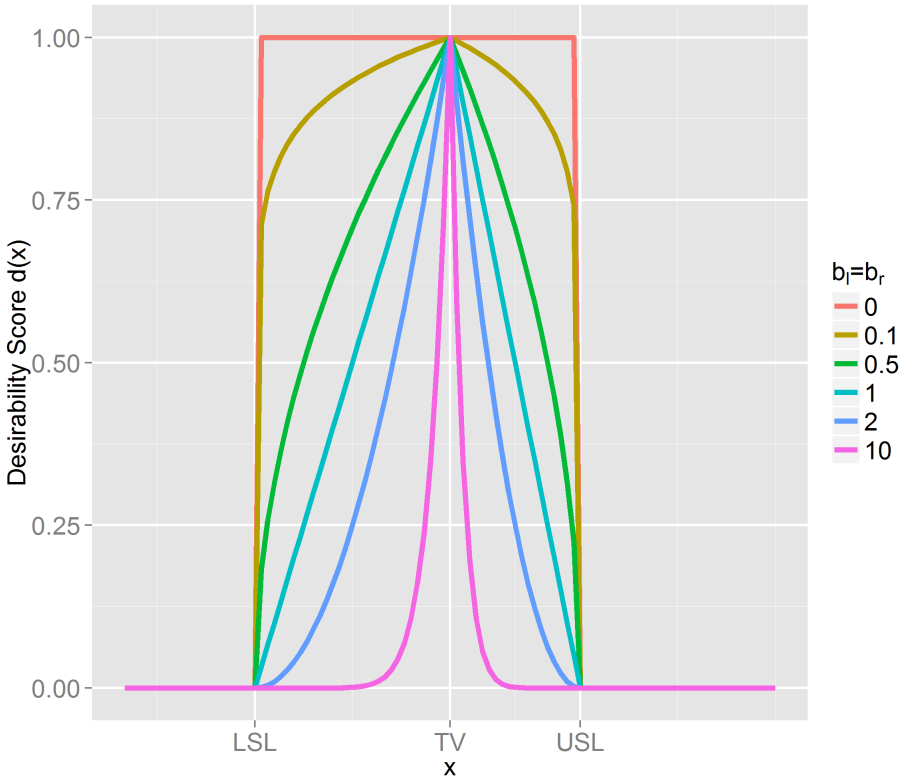


Figure (4.1): Exemplary settings of the two-sided Derringer-Suich desirability function dependent on b_l and b_r.

specified desirability function on the mean value of the criterion's distribution. When the full reference set of a randomization procedure is investigated, the mean value is a (weighted) average – the values of the criterion are weighted with their probability of occurrence. In later sections this approach of applying a desirability function is called the **average-based approach** (see Section 5.2.2). Considering the average-based approach, a high desirability score associates that the mean value of a criterion's distribution is desired. The second approach applies the desirability function on the particular values of the criterion's distribution. Thus, the dependency between the criterion's values and the randomization sequences is taken into account. This advance of applying a desirability function is called the **(randomization) sequence-based approach** (see Section 5.2.1). In the sequence-based approach a high

desirability score leads to the conclusion that the individual values of the criterion's distribution are desired.

In Chapters 2 and 3 several criteria for the measurement of selection and chronological bias on a given randomization procedure are introduced. Table 3.1 in Section 3.1.3 illustrates the realizations of the rejection probabilities $\alpha_{\mathrm{CB}}(\boldsymbol{t}_i)$ with $i \in \{1, 2, \ldots, 6\}$ conditioned on the randomization sequences of the Random Allocation Rule (see Section 2.2.2) under the null hypothesis $(H_0 : \theta_1 = 0)$ in the presence of a linear time trend with strength $\vartheta = 4$. To give an example, a right-sided Derringer-Suich desirability function with $\mathrm{TV} = \alpha_0 = 0.05$, $b_r = 1$, and $\mathrm{USL} = 2\,\alpha_0 = 0.1$ is applied on the mean value $\bar{\alpha}_{\mathrm{CB}} = 0.056$. Consequently, the desirability value 0.88 with respect to the type-I-error probabilities in case of chronological bias for the Random Allocation Rule is derived. Interpreting the desirability score $d(\bar{\alpha}_{\mathrm{CB}})$ according to Table 4.1, it can be concluded that the Random Allocation Rule is with respect to its expected type-I-error probability an excellent randomization procedure. Now, the previously defined right-sided Derringer-Suich desirability function is applied on the type-I-error probabilities conditioned on the randomization sequences of the Random Allocation Rule (presented in Table 3.1). The results are shown in Table 4.2. Four randomization sequences attain a type-I-error probability lower than the target value and are mapped to one and the two worst case randomization sequences with $\alpha_{\mathrm{CB}}(\boldsymbol{t}_i) > 0.1$ are mapped to zero. Hence, the average desirability of the sequence-based approach $\bar{d}(\alpha_{\mathrm{CB}}) = 1/6 \sum_{i=1}^{6} d(\alpha_{\mathrm{CB}}(\boldsymbol{t}_i))$ corresponds to $2/3$. Interpreting the summarized desirability scores of the individual randomization sequences according to Table 4.1, it can be concluded that the Random Allocation Rule is a good randomization procedure in the presence of the assumed linear time trend.

Desirability functions can easily be applied to other criteria, which were presented in the previous chapters. An extensive sensitivity analysis is

i	1	2	3	4	5	6
$\boldsymbol{t}_i^T$	(1,1,-1,-1)	(1,-1,1,-1)	(-1,1,1,-1)	(1,-1,-1,1)	(-1,1,-1,1)	(-1,-1,1,1)
$\alpha_{\mathrm{CB}}(\boldsymbol{t}_i)$	0.146	0.016	0.005	0.005	0.016	0.146
$d(\alpha_{\mathrm{CB}}(\boldsymbol{t}_i))$	0	1	1	1	1	0

Table (4.2): A right-one-sided Derringer-Suich desirability function with $\mathrm{TV} = 0.05$, $\mathrm{USL} = 0.1$, and $b_r = 1$ is used to map the $\alpha_{\mathrm{CB}}(\boldsymbol{t}_i)$ values for RAR to the interval $[0, 1]$.

conducted in Chapter 5. In that chapter both the average-based approach and the sequence-based approach are investigated in more detail to determine an appropriate randomization procedure. The next section discusses the combination of desirability scores.

4.2 Combination of desirability scores

The target value, the specification limits, and the weights of a desirability function are set by a person's point of view and should be chosen with care. When establishing a suitable desirability function the central question is, which values are desired for a given criterion and which values are not. The fitted desirability function should give an informative answer to this question. The class of the Derringer-Suich desirability functions presented in Section 4.1 is only one class of possible desirability functions. Other classes and forms of desirability functions are possible, for example the desirability function according to Harrington Jr. (1965). One key property of the class of the Derringer-Suich desirability functions is that completely undesired values of a criterion are mapped to zero. This property does not hold for any class of desirability functions, see for example the class of desirability functions defined by Harrington Jr. (1965).

A desirability function $d(x)$ maps the value x to a scale in the interval $[0, 1]$. This scale is dimensionless and represents the desirability of x dependent on the target value, the specification limits, and the weights of the corresponding desirability function. A value of zero is an undesired value and a value of one is a completely desired value. In this thesis the situation of having several objectives on one randomization procedure or on one randomization sequence is investigated. These objectives are measured based on corresponding criteria. Let $l \in \mathbb{N}$ be the number of different criteria. Furthermore, for each criterion an individual desirability function is fixed. Hence, l different desirability scores $d_1(x_1), \ldots, d_l(x_l)$ for the corresponding values $x_1, \ldots, x_l$ are derived. The aim is to give a universal statement for the randomization procedure or the randomization sequence and their values $x_1, \ldots, x_l$. Thus, the l desirability scores of a given randomization procedure or a given randomization sequence are combinable with the weighted geometric mean

according to Derringer (1994):

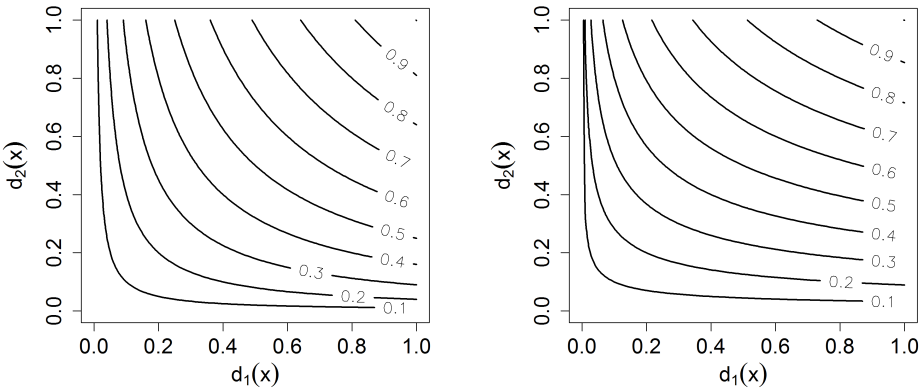
$$\bar{d}_{\text{geo}} := d_1^{\omega_1}(x_1) \cdot d_2^{\omega_2}(x_2) \cdot \ldots \cdot d_l^{\omega_l}(x_l)\,, \tag{4.4}$$

where $\omega_1, \ldots, \omega_l$ with $\sum_{i=1}^{l} \omega_i = 1$ are the weights of the desirability scores $d_1(x_1), \ldots, d_l(x_l)$ of $x_1, \ldots, x_l$. An important property of the geometric mean $\bar{d}_{\text{geo}}$ is that it is zero if and only if there is at least one $i \in \{1, \ldots, l\}$ with $\omega_i > 0$, such that $d_i(x_i) = 0$, i.e. that the corresponding desirability score does not fulfill its criterion. Consequently, a summarized desirability value with the geometric mean greater than zero is only possible if and only if all $d_i(x_i)$ with $w_i > 0$ are greater than zero for $i \in \{1, \ldots, l\}$. A randomization sequence which attains a summarized desirability value with the geometric mean of zero is called **undesired randomization sequence**. In Kim and Lin (2000) it is pointed out that a summarized desirability value with the geometric mean does not allow a clear interpretation, except that greater summarized desirability values with the geometric mean are preferred. However, in the later analysis the summarized desirability scores with the geometric mean are interpreted according to Table 4.1.

4.3 Properties of desirability scores

The combinability of desirability scores is the main reason why the transformation to desirability scores is indispensable. The criteria presented in Chapter 2 and Chapter 3 use different scales. Thus, it is not possible to make a unified assessment of one randomization procedure with respect to several criteria. By the usage of desirability functions it is possible to conduct a multiobjective assessment of randomization procedures with respect to several criteria. On the basis of the desirability scores the criteria are linked and it is possible to investigate the criteria of a randomization procedure simultaneously. It is obvious, that the randomization procedure with the greatest desirability score can be classified as the best one.

Originally, Harrington Jr. (1965) recommended to summarize desirability scores by the geometric mean with equal weight on each desirability score. Later Derringer (1994) pointed out that when the desirability scores are of different importance, a weighted geometric mean could be used. In Figure 4.2 the situation of two desirability scores is assumed, which are summarized



(a) $\omega_1 = \omega_2 = 1/2$.

(b) $\omega_1 = 1/3$ and $\omega_2 = 2/3$.

Figure (4.2): Contour plots when two desirability scores $d_1(x)$ and $d_2(x)$ are summarized with the geometric mean.

with the geometric mean. The contour plot on the left shows the value of the summarized desirability score when the desirability scores of the two criteria are of equal importance ($\omega_1 = \omega_2 = 1/2$). For the generation of the contour plot on the right, it is assumed that the importance of the desirability score of the second criterion is doubled ($\omega_1 = 1/3$, $\omega_2 = 2/3$). Particularly in the edges of the contour plots differences are visible. If both desirability scores are more or less the same, the weight does not influence the summarized desirability score that much.

For the adaptation of desirability functions the following situation should be considered: Let l be large and all desirability functions of the different criteria be chosen in that way, that there is only a small possibility to reach a desirability value greater than zero. Furthermore, several criteria work in opposite direction. Summarizing the desirability scores in this situation with the geometric mean would lead to a $\bar{d}_{\text{geo}}$ value of zero. The transformation of the criteria to desirability scores would not bring any benefit. Thus, the specification limits of the used desirability functions should be chosen as liberal as necessary, but also not more liberal than necessary.

For the weights ω_i with $i \in \{1, 2, \ldots, l\}$ of the geometric mean defined in Equation (4.4) it is assumed that the sum of the weights is one. This assumption is important in the situation when all desirability scores $d_j(x_j) =$

$d_k(x_k)$ for $j, k \in \{1, 2, \ldots, l\}$ are the same. Under the assumption that the sum of the weights ω_i with $i \in \{1, 2, \ldots, l\}$ is one, it follows:

$$\bar{d}_{\text{geo}} = d_1(x_1) = d_1^{\omega_1}(x_1) \cdot d_2^{\omega_2}(x_2) \cdot \ldots \cdot d_l^{\omega_l}(x_l)\,.$$

Without the assumption that the sum of the weights ω_i with $i \in \{1, 2, \ldots, l\}$ is one, the equality $\bar{d}_{\text{geo}} = d_1(x_1)$ would not hold. Although it is not necessary that the sum of the weights is one, all presented investigations in this thesis are based on the assumption that the sum of the weights is one.

Chapter 5

Sensitivity analysis of the desirability index

One aim of this thesis is to make a general statement of a randomization procedure with respect to its behavior in the presence of both selection and chronological bias. In Chapters 2 and 3 different criteria for the assessment of randomization procedures in the presence of chronological and selection bias were introduced. Chapter 4 is the foundation for the assessment of randomization procedures based on a desirability index. Below, the susceptibility of a randomization procedure to selection bias is measured by the expected proportion of correct guesses. The distorted rejection probabilities (type-I-error probability and power in case of a linear time trend) serve as criteria for the assessment of a randomization procedure in the presence of chronological bias.

To give an example, the application of desirability functions is illustrated on Efron's Biased Coin Design with parameter $p = 2/3$ (see Section 2.2.4). For the application of the desirability functions both the sequence-based approach and the average-based approach, which were introduced in Section 4.1, are extensively discussed. First, all randomization sequences for $N = 4$ are considered. Afterwards, the sample sizes $N \in \{12, 50, 200\}$ are investigated. For $N = 4$ and $N = 12$ the full reference set of Efron's Biased Coin Design is assessed, while for $N = 50$ and $N = 200$ a simulation study with 100 000 randomization sequences of EBC(2/3) is conducted. For the presented figures the additional `R` package `ggplot2` (Wickham, 2009) is used.

5.1 Assessment of Efron's Biased Coin Design

i	$\boldsymbol{t}_i^T$	$p_{\boldsymbol{t}}$	propCG($\boldsymbol{t}_i$)	d(propCG($\boldsymbol{t}_i$))	$\alpha_{\text{CB}}(\boldsymbol{t}_i)$	$d(\alpha_{\text{CB}}(\boldsymbol{t}_i))$	$1-\beta_{\text{CB}}(\boldsymbol{t}_i)$	$d(1-\beta_{\text{CB}}(\boldsymbol{t}_i))$	$\bar{d}_{\text{geo}}(\boldsymbol{t}_i)$
1	(-1,-1,-1,-1)	0.019	0.125	1.000	0.000	1.000	0.000	0.000	0.000
2	(1,-1,-1,-1)	0.056	0.500	1.000	0.055	0.892	0.620	0.099	0.445
3	(-1,1,-1,-1)	0.056	0.500	1.000	0.045	1.000	0.637	0.187	0.571
4	(1,1,-1,-1)	0.074	0.625	0.500	0.060	0.804	0.730	0.649	0.639
5	(-1,-1,1,-1)	0.037	0.375	1.000	0.045	1.000	0.679	0.395	0.734
6	(1,-1,1,-1)	0.111	0.750	0.000	0.047	1.000	0.734	0.668	0.000
7	(-1,1,1,-1)	0.111	0.750	0.000	0.043	1.000	0.755	0.776	0.000
8	(1,1,1,-1)	0.037	0.375	1.000	0.055	0.892	0.620	0.099	0.445
9	(-1,-1,-1,1)	0.037	0.375	1.000	0.055	0.892	0.743	0.715	0.861
10	(1,-1,-1,1)	0.111	0.750	0.000	0.043	1.000	0.755	0.776	0.000
11	(-1,1,-1,1)	0.111	0.750	0.000	0.047	1.000	0.792	0.961	0.000
12	(1,1,-1,1)	0.037	0.375	1.000	0.045	1.000	0.637	0.187	0.571
13	(-1,-1,1,1)	0.074	0.625	0.500	0.060	0.804	0.842	1.000	0.738
14	(1,-1,1,1)	0.056	0.500	1.000	0.045	1.000	0.679	0.395	0.734
15	(-1,1,1,1)	0.056	0.500	1.000	0.055	0.892	0.743	0.715	0.861
16	(1,1,1,1)	0.019	0.125	1.000	0.000	1.000	0.000	0.000	0.000
average values:			0.597	0.481	0.047	0.951	0.702	0.605	0.344

Table (5.1): Overview and properties of the 16 randomization sequences generated by EBC(2/3) for the sample size $N = 4$. The desirability functions for the assessment of the criteria are set according to Table 5.2.

One property of Efron's Biased Coin Design is that all imaginable randomization sequences are possible, but in contrast to Complete Randomization they are not equiprobable. Table 5.1 shows all possible randomization sequences under EBC(2/3) for the sample size $N = 4$ with their corresponding probability of appearance $p_{\boldsymbol{t}}$. In the fourth column of the table the expected proportion of correct guesses (see Section 2.3.2) under the usage of the convergence strategy dependent on the corresponding randomization sequence (propCG($\boldsymbol{t}_i$) = $\text{CG}(\boldsymbol{t}_i)/N$) is presented. In the next column, this value is assessed with a right-sided Derringer-Suich desirability function (see Section 4.1) with the following settings: TV = 0.50, USL = 0.75, and $b_r = 1$. The value of the upper specification limit corresponds to the maximal possible value of the expected proportion of correct guesses achieved by the convergence strategy. The target value is set to the prior probability of a correct guess in a two-armed clinical trial. As already mentioned before, this probability is 0.5. The column d(propCG($\boldsymbol{t}_i$)) in Table 5.1 shows the corresponding desirability value of the expected proportion of correct guesses. The values in the columns of the conditional rejection probabilities (type-I-error probability and power in case of a linear time trend) are derived in analogy to Tables 3.1 and 3.2 in Chapter 3. The parameter ϑ of the linear time trend is set to one (see Section 2.5). Thus, the time trend is assumed to be a comparatively small one. For assessing the type-I-error probabilities

j	Criterion$_j$	TV$_j$	USL$_j$/LSL_j	b_{l_j}/b_{r_j}	ω_j
1	propCG	0.50	0.75	1	1/3
2	α_{CB}	0.05	0.10	1	1/3
3	$1-\beta_{\text{CB}}$	0.80	0.60	1	1/3

Table (5.2): Standard setting of the desirability functions of the three investigated criteria.

conditioned on the randomization sequences a right-sided Derringer-Suich desirability function with TV = 0.05, USL = 0.10, and $b_r = 1$ is used. The conditional values for the power in the situation of the assumed linear time trend are mapped by a left sided Derringer-Suich desirability function with TV = 0.80, LSL = 0.60, and $b_l = 1$. The desirability values of the three criteria are summarized with the geometric mean (see Equation (4.4)). All of the three weights are set to one third. Table 5.2 summarizes the used settings of the desirability functions and the used weights ω_j for the three criteria. The presented setting of the desirability functions is used for all examples in this chapter. The row on the bottom in Table 5.1 shows the mean or average values of the corresponding columns under EBC(2/3) for $N = 4$. For that, every value in the corresponding column is weighted with $p_{\boldsymbol{t}}$, which is the probability of appearance of the randomization sequence under EBC(2/3). The average value of the summarized desirability scores is 0.344. As pointed out in Section 4.1 this value represents the average desirability of the randomization sequences under EBC(2/3). The average value of the summarized desirability scores is called $\bar{d}$(RS) (average value of the summarized desirability scores of the individual randomization sequences). In this thesis this approach of summarizing desirability scores is called the **(randomization) sequence-based approach** (see Section 4.1). The name of the approach refers solely to the application of the desirability functions. In this approach the desirability functions are directly applied on the realizations of the criteria's distributions conditioned on the randomization sequences.

The desirability scores of the criteria's average values under EBC(2/3) are provided by Table 5.3. The same desirability functions as established for the sequence-based approach are used to map the corresponding average values of the criteria to the interval $[0, 1]$ (see Table 5.2). Afterwards, the three desirability scores of the criteria's average values are summarized with

prop$\overline{\text{CG}}$	d(prop$\overline{\text{CG}}$)	$\bar{\alpha}_{\text{CB}}$	$d(\bar{\alpha}_{\text{CB}})$	$1-\bar{\beta}_{\text{CB}}$	$d(1-\bar{\beta}_{\text{CB}})$	$\bar{d}_{\text{geo}}$(AV)
0.597	0.611	0.047	1.000	0.702	0.509	0.678

Table (5.3): Desirability of the criteria's average values under EBC(2/3) for $N = 4$. The scores prop$\overline{\text{CG}}$, $\bar{\alpha}_{\text{CB}}$, and $1-\bar{\beta}_{\text{CB}}$ are derived in Table 5.1. The geometric mean of the three desirability values of the criteria is termed as $\bar{d}_{\text{geo}}$(AV).

the geometric mean. The weights of the geometric mean are again all set to one third. The summarized desirability score of the criteria's average values under EBC(2/3) is 0.678. This score is called $\bar{d}_{\text{geo}}$(AV) (summarized desirability score of the criteria's average values). In this thesis this approach of summarizing desirability scores is called the **average-based approach** (see Section 4.1). The summarized desirability score reflects the behavior of the criteria's average values under a given randomization procedure. The name average-based approach refers solely on the application of the desirability functions. In the given example the desirability functions are in the strict sense applied on the expected or mean values of the criteria's distributions. However, for larger sample sizes the mean values of the criteria's distributions are estimated by simulations. Let N be large, so that it is not possible to assess all randomization sequences of a given randomization procedure, thus the (weighted) arithmetic mean is an estimator for the expected value of a criterion's distribution. In this thesis no distinction is made between whether a desirability function is applied on a simulated average value or an expected value of a criterion's distribution. Nevertheless, it is always clearly pointed out whether the exact distribution of a criterion is taken into account or simulations are used.

For EBC(2/3) and $N = 4$ the desirability of the criteria's average values is 0.678. This value is clearly greater than the average desirability of the summarized desirability scores for EBC(2/3), which is 0.344 (see Table 5.1). However, being 0.509, the desirability of the average power in case of the assumed linear time trend value is smaller than the average of the desirability scores derived in the sequence-based approach (see Table 5.1). Considering the other two criteria, the desirability scores of the average values are greater than the corresponding average values derived in the sequence-based approach (see Table 5.1).

Summarizing the results, the desirability scores of both approaches are clearly greater than zero. Considering the three investigated criteria and the sample size $N = 4$, no randomization procedure can be highly recommended. In Table 5.1 it can be seen that six of sixteen possible randomization sequences have a desirability score of zero. Thus, for a good randomization procedure the probability of appearance of these six randomization sequences should be minimized. The sample size $N = 4$ is very small. For this sample size not all investigated trade-offs are solvable. The example is only selected to show how the application of desirability functions on several criteria of a randomization procedure works. The influence of different target values and of the upper/lower specification limits on the desirability scores is investigated in more detail based on the sample size $N = 12$ in the following section.

5.2 Parameters of the desirability functions

For the sample size $N = 12$ there are $2^{12} = 4096$ possible randomization sequences under EBC(2/3). Consequently, it is not possible to discuss or show the behavior of all randomization sequences in detail. As in the previous section it is assumed that the clinical trial is affected by a linear time trend with the strength one ($\vartheta = 1$ according to Equation (2.25)). Furthermore, the expected proportion of correct guesses reflects the susceptibility to the convergence strategy. Table 5.4 shows a summary of all possible randomization sequences under EBC(2/3) for $N = 12$. The values in the table are generated with the `randomizeR` package (Schindler et al., 2015). Let $c_j(\boldsymbol{t}_i) = x_{i,j}$ be the value of the jth criterion with $j \in \{1, 2, 3\}$ of the ith randomization sequence $\boldsymbol{t}_i$, thus the expectation of the jth criterion is computed as follows:

$$E(X_j) = \sum_{\{i:\boldsymbol{t}_i \in \Gamma_N\}} x_{i,j} \cdot p_{\boldsymbol{t}_i} = \bar{x}_j \,, \tag{5.1}$$

where $p_{\boldsymbol{t}_i}$ is the probability of appearance of the ith randomization sequence $\boldsymbol{t}_i$. The mean value of the jth criterion is computed by weighting the realizations of a criterion with their probability of appearance. In situations when the set of possible randomization sequences of a randomization procedure gets too large, simulations are used to estimate the expected value of a criterion. In these situations the randomization sequences are sampled from an algorithm, which represents the corresponding randomization procedure. Afterwards, the

(weighted) arithmetic mean of the criterion's values is used as estimator for the criterion's expected value. When the whole reference set of randomization procedure is assessed, the term $\bar{x}_j$ stands for criterion's expected value. In situations when randomization sequences from a randomization procedure are sampled, the term $\bar{x}_j$ is an estimator for the expected value of the corresponding criterion.

The variance of the jth criterion with $I = |\Gamma_N|$ is computed as follows:

$$s^2(X_j) = \sum_{\{i:\boldsymbol{t}_i \in \Gamma_N\}} p_{\boldsymbol{t}} \left(x_{i,j} - \mu\right)^2 = \sum_{\{i:\boldsymbol{t}_i \in \Gamma_N\}} p_{\boldsymbol{t}} \left(x_{i,j} - \bar{x}_j\right)^2 . \tag{5.2}$$

where the different probabilities of occurrence of the realizations $x_{i,j}$ are taken into account. The value of the maximum (max) and minimum (min) in Table 5.4 is the maximum or the minimum value of the vector $\boldsymbol{x}_j \in \mathbb{R}^I$ of the jth criterion. The value $\tilde{x}_{q,j}$ with $q \in [0, 1]$ is the q-quantile of the vector $\boldsymbol{x}_j \in \mathbb{R}^I$ of the jth criterion. Due to the fact that all imaginable randomization sequences are possible under Efron's Biased Coin, the depicted values of the minimum/maximum in Table 5.4 show the maximal possible range of the criteria. In case of the assumed linear time trend and in comparison to the planned power of 80%, the average power of EBC(2/3) is only slightly deflated. The worst case randomization sequence for the power (twelve times the same treatment in a row) has a power of zero. The average value of the expected proportion of correct guesses is 61.3% with range $[0.042, 0.750]$. In the situation of the assumed linear time trend nearly one quarter of the randomization sequences maintains an inflated type-I-error probability. The nominal significance level α is set to 5%.

Based on the summary of EBC(2/3) presented in Table 5.4 several settings of desirability functions are discussed. For assessing the distributions of the expected proportion of correct guesses and the distorted type-I-error probability in case of the linear time trend several settings of right-sided Derringer-Suich desirability functions are investigated. For the evaluation

j	Criterion	$\bar{x}_j$	$s(X_j)$	$\max(\boldsymbol{x}_j)$	$\min(\boldsymbol{x}_j)$	$\tilde{x}_{0.05,j}$	$\tilde{x}_{0.25,j}$	$\tilde{x}_{0.50,j}$	$\tilde{x}_{0.75,j}$	$\tilde{x}_{0.95,j}$
1	propCG	0.613	0.096	0.750	0.042	0.417	0.542	0.625	0.667	0.750
2	α_{CB}	0.047	0.009	0.119	0.000	0.041	0.042	0.044	0.049	0.066
3	$1 - \beta_{CB}$	0.756	0.072	0.944	0.000	0.626	0.717	0.765	0.806	0.861

Table (5.4): Summary of the distributions of the criteria of the randomization sequences produced by EBC(2/3) for $N = 12$.

of the distribution of the distorted power in case of the linear time trend several left-sided Derringer-Suich desirability functions are adjusted. In what follows, both the sequence-based approach and the average-based approach are investigated and the results are compared.

5.2.1 Sequence-based approach

The one-sided Derringer-Suich desirability function, which is presented in Section 4.1, depends on three parameters: The upper/lower specification limit, the target value, and the corresponding weight b_l or alternatively b_r. In what follows, the target values of the three criteria are fixed and not changed anymore. The target values of the criteria are set as follows: $\text{TV}_{\text{propCG}} = 0.50$, $\text{TV}_{\alpha_{\text{CB}}} = 0.05$, and $\text{TV}_{1-\beta_{\text{CB}}} = 0.80$. Additionally to the settings of the desirability functions, the choice of the weights ω_j with $j \in \{1, 2, 3\}$ of the geometric mean, which is used for summarizing the three desirability scores of the three criteria to a unified score, is investigated. Table 5.2 shows the standard setting of the three desirability functions of the three criteria and their weights ω_j. If it is not explicitly stated otherwise, all criteria are investigated in this section with respect to the values of the standard setting. In this setting all criteria have a weight of $\omega_j = 1/3$. The upper specification limit for the expected proportion of correct guesses is set to 0.75, which is the maximal possible value of this criterion. For both the planned type-I-error probability $\alpha_0 = 0.05$ and the planned type-II-error probability $\beta_0 = 0.2$ maximal a duplication in the presence of the assumed linear time trend is allowed - otherwise the corresponding values are mapped to zero. Thus, the specification limits for the type-I-error probability and power in case of the assumed linear time trend are set to 0.10 and 0.60, respectively. In the standard setting the b_l/b_r values of all desirability functions are set to one (linear loss of deviations from the target value).

Table 5.5 shows the summary of the distribution of the (summarized) desirability scores. The values of the criteria of the individual randomization sequences are assessed with the desirability functions defined in the standard setting. The criterion RS represents the distribution of the summarized desirability scores of the individual randomization sequences ($\bar{d}_{\text{geo}}(\boldsymbol{t}_i)$). The used weights ω_j for the geometric mean are dictated in the standard setting (see Table 5.2). The probability of generating an undesired value for the

j	Criterion$_j$	$\bar{d}(j)$	$P(d(j)=0)$	$\tilde{x}_{0.05,d(j)}$	$\tilde{x}_{0.25,d(j)}$	$\tilde{x}_{0.50,d(j)}$	$\tilde{x}_{0.75,d(j)}$	$\tilde{x}_{0.95,d(j)}$
1	propCG	0.516	0.088	0.000	0.333	0.500	0.833	1.000
2	α_{CB}	0.955	0.003	0.675	1.000	1.000	1.000	1.000
3	$1-\beta_{\text{CB}}$	0.738	0.028	0.129	0.585	0.827	1.000	1.000
4	RS	0.611	0.117	0.000	0.522	0.667	0.794	0.941

Table (5.5): Summary of the desirability scores under the standard setting for EBC(2/3) for $N = 12$.

distribution of the expected proportion of correct guesses is 8.8%. The average value of the desirability of the expected proportion of correct guesses corresponds to 0.516, which is an acceptable value (see Table 4.1). The quantiles of the distribution function of the desirability scores of the four criteria represent the distribution function of the desirability scores. The distribution functions of the type-I-error probability and the power in case of the assumed linear time trend are remarkable. For the desirability distribution of the type-I-error probability the probability of a desirability score of one is 75%. In other words, the probability of generating a randomization sequence with a deflated or preserved type-I-error probability is greater than three quarters (see Table 5.4). The probability of generating a randomization sequence with an inflated or preserved power in case of the investigated linear time trend is greater than 25%. In what follows, the influence of a change in the standard setting with respect to the upper/lower specification limit, the b_l or b_r value, and the weights ω_j with $j \in \{1, 2, 3\}$ on both the average desirability score of the distribution of the corresponding criterion and the average of the summarized desirability scores is investigated. When the standard setting is used, the probability of generating an undesired randomization sequence is 11.7% and the average of the summarized desirability scores corresponds to 0.611. Considering the standard setting of the desirability functions (see Table 5.2), EBC(2/3) for $N = 12$ is a randomization procedure with good properties with respect to selection and chronological bias. Below, the influence of the choice of the upper/lower specification limits on the derived desirability scores is investigated.

Influence of the specification limits on the desirability scores

In this section the specification limits (USL and LSL) of the criteria's desirability functions are altered successively. The criteria, which are not

investigated, are always assessed with the desirability functions defined in the standard setting (see Table 5.2). When assessing the upper/lower specification limit of one criterion in the analysis, there are three values of special interest:

1.) The expectation of the summarized desirability scores:

$$E(d(\text{RS})) = \bar{d}(\text{RS}) = \sum_{\{i:\boldsymbol{t}_i \in \Gamma_N\}} \bar{d}_{\text{geo}}(\boldsymbol{t}_i)\, p_{\boldsymbol{t}_i}\,,$$

where $\bar{d}_{\text{geo}}(\boldsymbol{t}_i)$ is the summarized desirability value of the ith randomization sequence with the geometric mean.

2.) The expected desirability of the investigated criterion:

$$E(d(\text{criterion}_j)) = \bar{d}(\text{criterion}_j) = \sum_{\{i:\boldsymbol{t}_i \in \Gamma_N\}} d(x_{i,j})\, p_{\boldsymbol{t}_i}\,,$$

where $d(x_{i,j})$ is the desirability of the value x_i of the ith randomization and the jth criterion.

3.) The probability of generating an undesired randomization sequence dependent on the criterion:

$$P(d(\text{criterion}_j) = 0) = \sum_{\{i:\boldsymbol{t}_i \in \Gamma_N\}} \mathbb{1}_{\{d(x_{i,j})=0\}}\, p_{\boldsymbol{t}_i}\,,$$

where $\mathbb{1}_{\{d(x_{i,j})=0\}}$ is one if $d(x_{i,j}) = 0$ and zero otherwise.

Figure 5.1 consists of three subfigures – one for each of the investigated criteria: propCG, α_{CB}, and $1 - \beta_{\text{CB}}$. The three computed values of interest are visualized in each subfigure dependent on the altered specification limit of the investigated criterion. The influence of the upper specification limit of the desirability function defined for the type-I-error probability in case of the assumed linear time trend is assessed when the upper specification limit 0.2 converges against the target value 0.05 of the criterion. The upper specification limit for for the expected proportion of correct guesses is varied between 0.75 and 0.5. The influence of the lower specification limit of the desirability function defined for the power in case of the assumed linear time trend is investigated when the lower specification limit lies anywhere

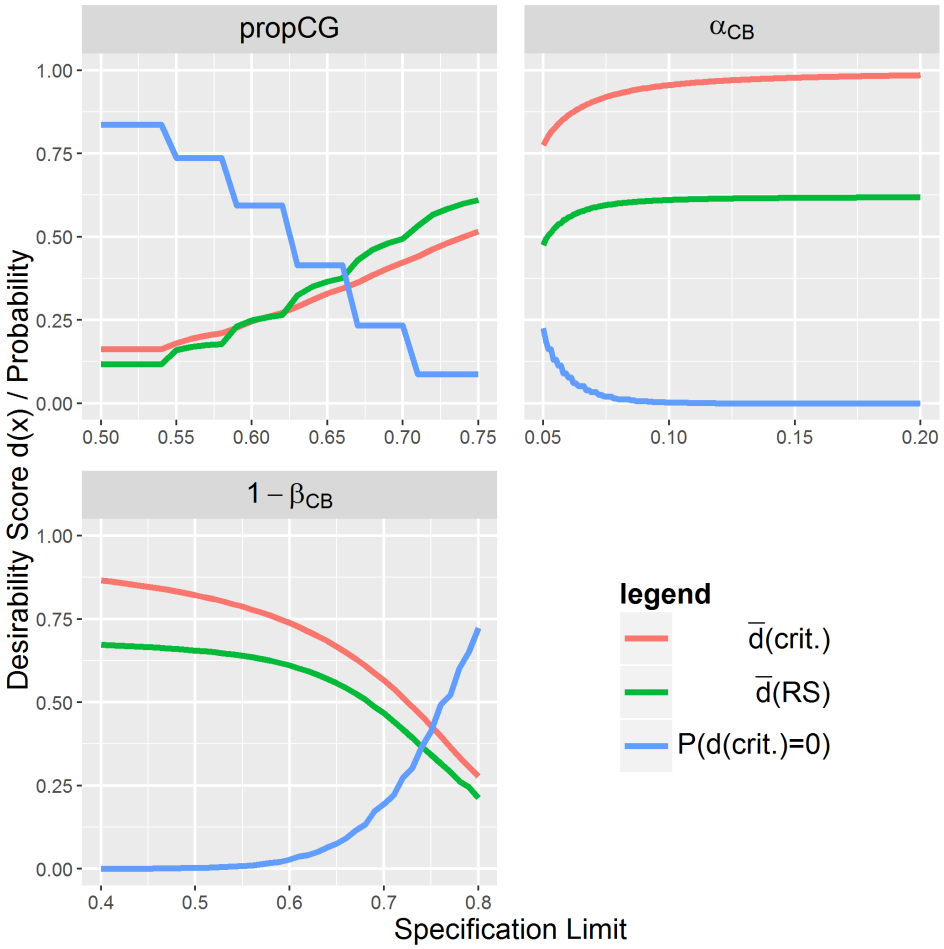


Figure (5.1): Assessment of the three investigated criteria dependent on different specification limits under EBC(2/3).

between 0.4 and the target value 0.8 of the criterion. Figure 5.1 shows that the closer the specification limits are chosen to the corresponding target value of the criterion, the greater the probability of generating an undesired randomization sequence becomes. As mentioned above there is a probability of less than 25% of generating a randomization sequence, which attains an inflated type-I-error probability for the investigated linear time trend. The expected proportion of correct guesses is the most susceptible criterion with respect to the probability of generating an undesired randomization sequence when the specification limit converges against the target value. Consequently, the setting of the upper specification limit of the desirability function for the expected proportion of correct guesses is the most sensitive with respect

to the influence on the average of the summarized desirability scores under EBC(2/3). Furthermore, the average of the summarized desirability scores is more sensitive with respect to the setting of the lower specification limit for the power than for the setting of the upper specification limit for the type-I-error probability in case of the assumed linear time trend. Finally, Figure 5.1 shows that the bigger the difference between the specification limit and the target value of a criterion are, the greater the expected desirability score of the corresponding desirability scores of the criterion becomes.

Summarizing the results shown in Figure 5.1 the specification limits (USL and LSL) of the three criteria have strong influence on the derived desirability scores under EBC(2/3). Particularly, the probability of generating an undesired randomization sequence is determined by the setting of the specification limits. The other parameters, which are investigated in the next paragraph, will not have any influence on the probability of generating an undesired randomization sequence. The settings of the other parameters only influence the average values of the desirability scores of the individual criteria and/or the expected desirability of the summarized desirability scores of the individual randomization sequences.

Influence of the parameter b on the desirability scores

In this section the influence of the parameter b of the three one-sided desirability functions on the desirability scores is investigated. The parameters of the weights and of the specification limits are set in all situations according to the standard setting (see Table 5.2). Figure 5.2 shows the influence of a change of a b value on both the average desirability of a corresponding criterion and the average of the summarized desirability scores under EBC(2/3). Setting the parameter b to one corresponds to a linear decreasing of the desirability function. If the parameter b is selected greater than one, the desirability function is a convex one and the setting of a b value of smaller than one corresponds to a concave decreasing of the desirability function (see Figure 4.1). The subfigures of Figure 5.2 show the smaller the parameter b of a criterion is, the greater the expected desirability of the summarized desirability scores of the individual randomization sequences under EBC(2/3) becomes. Selecting a linear desirability ($b = 1$) seems to be a good solution of weighting standardized deviations from the target value neither too

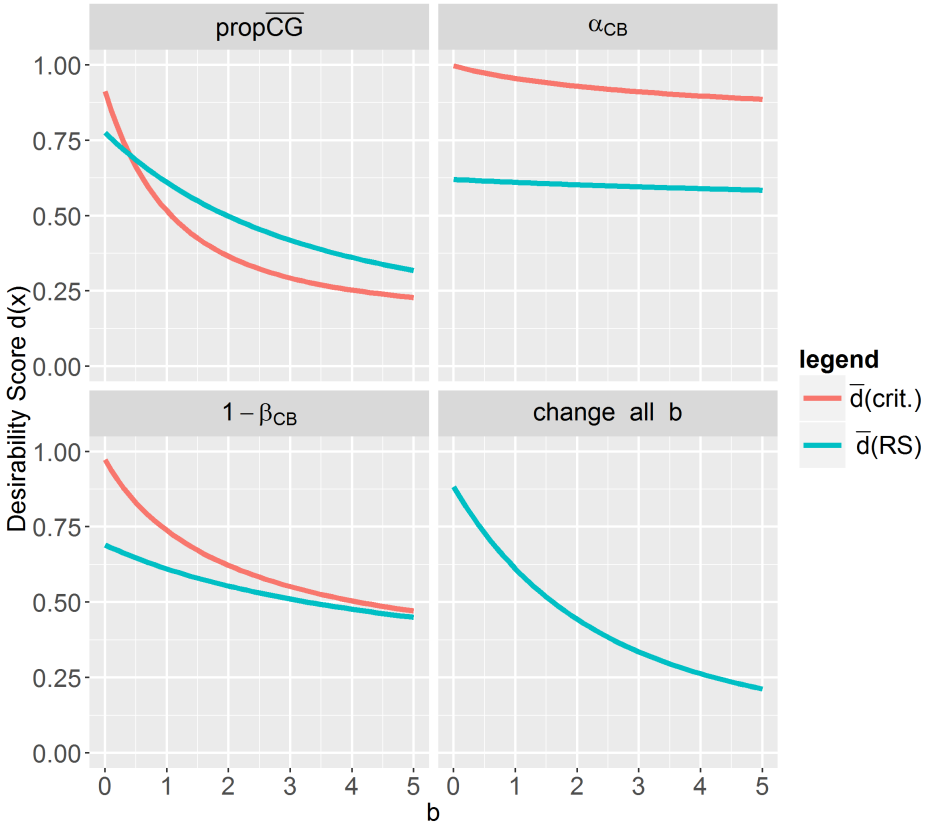


Figure (5.2): Assessment of the three investigated criteria dependent on the choice of the parameter b under EBC(2/3).

hard nor too soft. The desirability of the expected proportion of correct guesses criteria is most sensitive concerning the setting of the parameter b. The subfigure on the bottom right of Figure 5.2 shows the change of the expected desirability of the summarized desirability scores of the individual randomization sequences under EBC(2/3), when the b values of all three investigated criteria are altered simultaneously. Change all b values means that the parameters of all three desirability functions are changed at the same time. Considering this setting, the average of the summarized desirability scores decreases in form of a concave curve as the parameter b grows.

Influence of the weights on the desirability scores

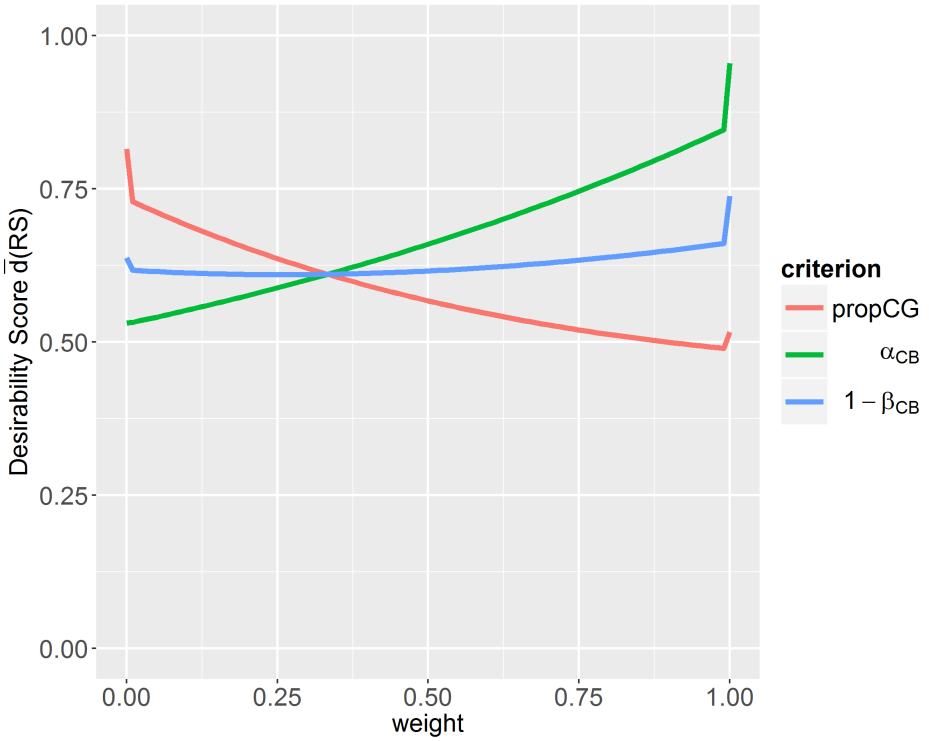


Figure (5.3): Assessment of the average of the summarized desirability scores under EBC(2/3) dependent on the weights ω_j of the three investigated criteria.

The analysis of the weights ω_j with $j \in \{1, 2, 3\}$ of the three criteria is based on the desirability functions defined by the standard setting (see Table 5.2). Solely the summarized desirability scores of the individual randomization sequences and therefore the expected desirability of the summarized desirability scores of the individual randomization sequences depend on the choice of the weights ω_j. The expected desirability values of a criterion ($\bar{d}$(criterion) with criterion $\in \{\text{propCG}, \alpha_{\text{CB}}, 1 - \beta_{\text{CB}}\}$) are determined by the standard setting and do not differ from the statistics shown in Table 5.5.

Figure 5.3 shows three different graphs – one for each investigated criterion. The average of the summarized desirability scores under EBC(2/3) is plotted against a value for the weight ω_j used for the jth criterion. Thus, the x-axis shows a value for the weight for the corresponding criterion. All three weights have to sum up to one. Hence, the values for the weights of the other two

criteria are set both to $(1-\omega_j)/2$. Three settings are of special interest: If a weight corresponds to the value one, the weights of the other two criteria are zero – the expected desirability of the summarized desirability scores of the individual randomization sequences corresponds to the expected desirability of the jth criterion, which is shown in Table 5.5. Next, if a weight corresponds to the value zero, the weight of the other two criteria is 0.5 – in this situation the criterion itself is weighted with zero and thus the desirability of the criterion does not influence the further analysis. If any criterion j has weight $\omega_j > 0$, while its desirability is zero, then the whole desirability index is zero, even if $\omega_j \to 0$. However, when $\omega_j = 0$, then the influence of criterion j vanishes completely. This explains the discontinuities of the graphs in Figure 5.3. Finally, all criteria are equally weighted if ω_j corresponds to $1/3$ (standard setting). This is the point of intersection of the three lines in the Figure 5.3.

Summarizing Figure 5.3 it is obvious, that the desirability of the expected proportion of correct guesses is small in comparison to the other two criteria. The smaller the influence of the expected proportion of correct guesses on the desirability of the individual randomization sequences is, the greater both the individual desirability values of the randomization sequences and the average of the summarized desirability scores under EBC($2/3$) become. Vice versa, the greater the weight of the expected proportion of correct guesses is, the lower the average of the summarized desirability scores under EBC($2/3$) becomes. Setting the weight ω_j of the expected proportion of correct guesses to $1/2$ and the weights of the other two criteria to $1/4$, a summarized desirability score of 0.57 is derived. In this setting selection and chronological bias are equally weighted.

5.2.2 Average-based approach

In this section the behavior of the average values of the three criteria under EBC($2/3$) is investigated. A summary of the criteria's average values is shown in Table 5.4 for $N = 12$. The standard setting presented in Table 5.2 is used for assessing the behavior of the criteria's average values under EBC($2/3$). Table 5.6 shows both the desirability of the criteria's average values under EBC($2/3$) and their summarized value with the geometric mean. The summarized desirability score of the criteria's average values under

the standard setting of EBC(2/3) is greater than 0.75. Hence, for $N = 12$ EBC(2/3) is a randomization procedure with good satisfaction. Table 5.6 shows that the expected type-I-error in case of the assumed linear time trend is deflated under EBC(2/3). Thus, the right-sided Derringer-Suich desirability function with target value 0.05, which is used for mapping the expected type-I-error probability in case of the assumed linear time trend to the interval $[0, 1]$, is always one regardless of the setting of the weight b and the upper specification limit. Consequently, the sensitivity analysis is solely done for the other two criteria. The desirability of the expected type-I-error probability in case of the investigated linear time trend is assumed to be one below.

Figure A.1 in the appendix shows a sensitivity analysis of the desirability of the average of the expected proportion of correct guesses and the average power $1 - \bar{\beta}_{\text{CB}}$. The figure depicts both the desirability score of the criteria's average values and the summarized desirability score of the criteria's average values with the weights ω_j of the standard setting. The closer the upper/lower specification limit is selected to the target value, the smaller the desirability of the average value of the criterion becomes. Furthermore, the smaller the desirability score of the average value of a criterion is, the smaller the summarized desirability score of the criteria's average values becomes. If the upper/lower specification limit is selected lower/higher than the average score of one criterion, the desirability value of both the criterion and the summarized desirability score of the criteria's average values is zero. Figure A.2 in the appendix shows the influence of the parameter b of the one-sided Derringer-Suich desirability function on both the desirability value of the average value of the criterion and the summarized desirability score of the criteria's average values. The closer the value b is selected to zero, the greater the desirability value of the average value of the criterion becomes. The greater the desirability value of the criterion's average value is, the greater the summarized desirability score of the criteria's average values

$\text{prop}\overline{\text{CG}}$	$d(\text{prop}\overline{\text{CG}})$	$\bar{\alpha}_{\text{CB}}$	$d(\bar{\alpha}_{\text{CB}})$	$1 - \bar{\beta}_{\text{CB}}$	$d(1 - \bar{\beta}_{\text{CB}})$	$\bar{d}_{\text{geo}}(\text{AV})$
0.613	0.549	0.047	1.000	0.756	0.782	0.754

Table (5.6): Desirability of the behavior of the criteria's average values using the standard setting from Table 5.2 under EBC(2/3) for $N = 12$. The scores $\text{prop}\overline{\text{CG}}$, $\bar{\alpha}_{\text{CB}}$, and $1 - \bar{\beta}_{\text{CB}}$ are derived in Table 5.4.

becomes. Figure A.3 in the appendix shows the influence of the weights ω_j on the summarized desirability score of the criteria's average values. The figure has to be interpreted analogously to Figure 5.3 presented in the previous section. Again the summarized desirability score of the criteria's average values is mainly influenced by the choice of the weight ω_j for the expected proportion of correct guesses. Setting the weight ω_j of the expected proportion of correct guesses to $1/2$ and the weights of the other two criteria to $1/4$, a summarized desirability score of the average values of the criteria of 0.70 is derived. Below, the desirability scores under EBC($2/3$) are investigated for the sample sizes $N = 50$ and $N = 200$.

5.3 Investigations of other sample sizes

In this section the desirability functions from the standard setting presented in Table 5.2 are investigated for the sample sizes $N = 50$ and $N = 200$ under EBC($2/3$). Again the desirability scores of both approaches, the sequence-based approach and the average-based approach, are of interest. For $N = 12$ Table 5.4 shows a summary of the distribution of the investigated criteria under EBC($2/3$). For $N = 50$ there are already $2^{50} = 1.1259 \cdot 10^{15}$ possible randomization sequences under Efron's Biased Coin Design. Thus, not the full reference set of all possible randomization sequences is assessed. The additional package `randomiezR` (Schindler et al., 2015) in `R` (R Core Team, 2016) is used for generating 100 000 randomization sequences under EBC($2/3$) for the sample sizes $N = 50$ and $N = 200$. The strength of the linear time trend is for all sample sizes set to one ($\vartheta = 1$ according to Equation (2.25)). The assessed criteria are the average of the expected proportion of correct guesses and the distorted rejection probabilities in case of the linear time trend under the null hypothesis and under the alternative.

Tables A.5 and A.6 in the appendix show the results of the simulation study. These tables are generated in analogy to Table 5.4. Tables 5.4, A.5, and A.6 show the greater the sample size N is, the lower the variance in the proportion of correct guesses becomes. The average value of the proportion of correct guesses converges to a constant and does not differ much in the presented tables. The variance of the rejection probabilities of the individual randomization sequences ($\alpha_{\text{CB}}(\boldsymbol{t}_i)$ and $1 - \beta_{\text{CB}}(\boldsymbol{t}_i)$) in case of a linear time trend of strength $\vartheta = 1$ becomes lower as N grows. In other words, the

behavior of the individual randomization sequences concerning the three criteria becomes more homogenous as the sample size N grows. The average values of the rejection probabilities are nearly on the same level for all investigated sample sizes. Particularly, the average values of both the type-I-error probability and power in case of the linear time trend are for all sample sizes deflated. In the next step, the desirability functions from the standard setting presented in Table 5.2 are applied to the simulated randomization sequences for the sample size $N = 50$.

Desirability of EBC(2/3) for $N = 50$

The target values of the desirability functions are set according to the values of the standard setting presented in Table 5.2. The desirability functions are directly applied to the realizations of the criteria conditioned on the randomization sequences from the simulation study. The realizations of the criteria from the simulation study are summarized in Table A.5 in the appendix. Table 5.7 shows a summary of the derived desirability scores by the sequence-based approach for EBC(2/3) and $N = 50$. The average desirability of the desirability scores of the expected proportions of the correct guesses is the lowest. The proportion of undesired randomization sequences is only 0.3%. The average desirability of the summarized desirability scores of the individual randomization sequences in Table 5.7 under EBC(2/3) is 0.728. Thus, EBC(2/3) is a good randomization procedure concerning the three investigated criteria. The desirability of the average values of the criteria under EBC(2/3) is shown in Table 5.8. For that, the desirability functions defined in the standard setting presented in Table 5.2 are used. The summarized desirability score of the criteria's average values under EBC(2/3) for $N = 50$ is 0.761. The derived desirability values of both approaches are

j	Criterion$_j$	$\bar{d}(j)$	$\hat{P}(d(j) = 0)$	$\tilde{x}_{0.05,d(j)}$	$\tilde{x}_{0.25,d(j)}$	$\tilde{x}_{0.50,d(j)}$	$\tilde{x}_{0.75,d(j)}$	$\tilde{x}_{0.95,d(j)}$
1	propCG	0.512	0.000	0.240	0.400	0.480	0.640	0.840
2	α_{CB}	0.990	0.000	0.957	1.000	1.000	1.000	1.000
3	$1 - \beta_{\text{CB}}$	0.825	0.003	0.450	0.728	0.874	1.000	1.000
4	RS	0.728	0.003	0.565	0.661	0.731	0.804	0.896

Table (5.7): Summary of the desirability scores derived by the sequence-based approach with the standard setting under EBC(2/3) for $N = 50$. The results are based on 100 000 simulated randomization sequences.

prop$\overline{\text{CG}}$	d(prop$\overline{\text{CG}}$)	$\bar{\alpha}_{\text{CB}}$	$d(\bar{\alpha}_{\text{CB}})$	$1-\bar{\beta}_{\text{CB}}$	$d(1-\bar{\beta}_{\text{CB}})$	$\bar{d}_{\text{geo}}$(AV)
0.622	0.513	0.043	1.00	0.772	0.861	0.761

Table (5.8): Desirability of the average behavior of EBC(2/3) in case of $N = 50$. The scores prop$\overline{\text{CG}}$, $\bar{\alpha}_{\text{CB}}$, and $1-\bar{\beta}_{\text{CB}}$ are derived in Table A.5 in the appendix.

very close to each other. Thus, both derived desirability scores lead to the conclusion that EBC(2/3) for $N = 50$ is a good randomization procedure with respect to the investigated criteria.

Desirability of EBC(2/3) for $N = 200$

As in the previous section, the desirability functions are set to the values defined in the standard setting presented in Table 5.2. The desirability functions are directly applied on the realizations of the criteria conditioned on the randomization sequences from the simulation study. The realizations of the criteria from the simulation study are summarized in Table A.6 in the appendix. Table 5.9 shows a summary of the derived desirability scores by the sequence-based approach for EBC(2/3) and $N = 200$. In general, Table 5.9 is very similar to Table 5.7. In particular, the proportion of undesired randomization sequences under EBC(2/3) for $N = 200$ is zero. The desirability scores of the average values of the three assessed criteria under EBC(2/3) for $N = 200$ are presented in Table 5.10. Both the average desirability of the summarized desirability scores shown in Table 5.9 and the summarized desirability score of the criteria's average values derived in Table 5.10 are equal to 0.76. Thus, both the sequence-based and the average-based approach lead to the conclusion that EBC(2/3) is a good randomization procedure with respect to chronological as well as selection bias.

j	Criterion	$\bar{d}(j)$	$\hat{P}(d(j)=0)$	$\tilde{x}_{0.05,d(j)}$	$\tilde{x}_{0.25,d(j)}$	$\tilde{x}_{0.50,d(j)}$	$\tilde{x}_{0.75,d(j)}$	$\tilde{x}_{0.95,d(j)}$
1	propCG	0.504	0.000	0.370	0.440	0.500	0.560	0.650
2	α_{CB}	1.000	0.000	1.000	1.000	1.000	1.000	1.000
3	$1-\beta_{\text{CB}}$	0.870	0.000	0.677	0.811	0.883	0.952	1.000
4	RS	0.755	0.000	0.672	0.719	0.754	0.790	0.843

Table (5.9): Summary of the desirability scores derived by the sequence-based approach with the standard setting under EBC(2/3) for $N = 200$. The results are based on 100 000 simulated randomization sequences.

prop$\overline{\text{CG}}$	prop$\overline{\text{CG}}$	$\bar{\alpha}_{\text{CB}}$	$d(\bar{\alpha}_{\text{CB}})$	$1-\bar{\beta}_{\text{CB}}$	$d(1-\bar{\beta}_{\text{CB}})$	$\bar{d}_{\text{geo}}(\text{AV})$
0.624	0.504	0.042	1.00	0.776	0.879	0.762

Table (5.10): Desirability of the average behavior of EBC(2/3) in case of $N = 200$. The scores prop$\overline{\text{CG}}$, $\bar{\alpha}_{\text{CB}}$, and $1-\bar{\beta}_{\text{CB}}$ are derived in Table A.6 in the appendix.

5.4 Sensitivity analysis in a nutshell

The presented sensitivity analysis has shown that the desirability scores of EBC(2/3) are strongly dependent on the specification limits and the used weights of the desirability functions. Furthermore, the influence of the weights for the geometric mean, which is used for summarizing the desirability scores to a unified score, and the sample size N were examined. For EBC(2/3) the specification limits are the most sensitive parameter with respect to the desirability scores derived by both the sequence-based and the average-based approach. The probability of generating an undesired randomization sequence is solely dependent on the setting of the specification limit. The standard setting presented in Table 5.2 is a robust setting for meaningful desirability scores under EBC(2/3) for all sample sizes. The standard setting neither weights standardized deviations from the corresponding target value of a criterion to hard nor to soft. So the standard setting can be used for both the sequence-based and the average-based approach.

In the following chapter several randomization procedures will be compared based on the desirability functions given in the standard setting (see Table 5.2) at the beginning of this chapter. For EBC(2/3) the derived desirability scores of the two approaches in this chapter are very close to each other in all investigated scenarios. In the next chapter the desirability of other randomization procedures for both approaches (the sequence-based and the average-based approach) is investigated.

Chapter 6

Assessment of randomization procedures

As pointed out at the beginning of Chapter 4, in the literature, there is no advice for the choice of a randomization procedure in the presence of both selection and chronological bias given. Even the ICH E9 (1998) is contradictory in itself. In order to prevent chronological bias it is recommended to choose small block lengths under the usage of the Permuted Block Randomization, whereas greater (random) block lengths should be preferred to prevent selection bias. The situation when both selection and chronological bias are present remains unclear. In this chapter, several settings of different randomization procedures are assessed when both selection and chronological bias are present. Selection bias is measured by the expected proportion of correct guesses and chronological bias is assessed by the distorted type-I-error probability and power of Student's t-test in the situation of a linear time trend. On the basis of these three criteria several settings of different randomization procedures are assessed for the sample sizes $N \in \{12, 50, 200\}$. Both the sequence-based approach and the average-based approach are investigated and the results of both approaches are compared. The chapter closes with a general conclusion of the derived results.

In the previous chapter a sensitivity analysis of the desirability index under EBC(2/3) was conducted. In the sensitivity analysis the standard setting from Table 5.2 was investigated. This setting was shown to be a robust setting for meaningful desirability scores under EBC(2/3). In this chapter, randomization procedures are assessed in the presence of selection as well

j	Criterion$_j$	TV$_j$	SL$_j$	b_{l_j}/b_{r_j}	ω_j
1	propCG	0.50	0.75	1	1/2
2	α_{CB}	0.05	0.10	1	1/4
3	$1 - \beta_{\text{CB}}$	0.80	0.60	1	1/4

Table (6.1): Setting of the desirability functions and the weights ω_j for the comparison of the randomization procedures for the investigated samples sizes $N \in \{12, 50, 200\}$.

as chronological bias. In the following assessment it is assumed that both objectives (selection and chronological bias) are of equal importance. Thus, half of the weight ω_j is put on each objective. The investigated selection bias is measured on the basis of the expected proportion of correct guesses. Hence, the corresponding weight ω_j for this criterion is set to 1/2. Chronological bias is assessed by the distorted test decision conducting a two-sided Student's t-test with level $\alpha = 0.05$ without adjusting for a linear time trend in the analysis. Therefore, both the distorted type-I-error probability α_{CB} in the situation of $H_0 : \theta_1 = 0$ (see Model (2.27)) and the distorted power $1 - \beta_{\text{CB}}$ in the situation of $H_1 : \theta_1 \neq 0$ are assessed in the presence of a linear time trend with the strength $\vartheta = 1$ (see Equation (2.25)). The weights for both of these two criteria are set to 1/4, so that the sum of all three weights is one. Table 6.1 summarizes the choice of the parameters for the desirability functions and the weights that are used for the following assessment of the randomization procedures. The upper/lower specification limits and weights b_l/b_r of the desirability functions are set according to the standard setting, because in the sensitivity analysis it was shown that the standard setting is reliable for all investigated sample sizes $N \in \{12, 50, 200\}$. Below, several settings of different randomization procedures for the sample size $N = 12$ are investigated with respect to their susceptibility to both selection and chronological bias.

6.1 Investigation of the sample size $N = 12$

Table 6.2 shows a detailed overview of the assessed randomization procedures and their parameters for $N = 12$. For all investigated randomization procedures up to the Randomized Permuted Block Randomization the full distribution of possible randomization sequences with their corresponding

Design	$\bar{d}_{\text{geo}}$(AV)	$\bar{d}$(RS) (sd)	$P(d(RS)=0)$	$\bar{\alpha}_{\text{CB}}$ (sd)	$1-\bar{\beta}_{\text{CB}}$ (sd)	prop$\overline{\text{CG}}$ (sd)
BSD(2)	0.7244	0.6704	0.0195	0.0462	0.7618	0.6042
		(0.185)		(0.006)	(0.061)	(0.063)
BSD(3)	0.8121	0.7287	0.0291	0.0485	0.7585	0.5648
		(0.211)		(0.009)	(0.073)	(0.071)
BSD(4)	0.8653	0.7159	0.0527	0.0501	0.7476	0.5319
		(0.241)		(0.011)	(0.083)	(0.093)
CR	0.8890	0.6503	0.1331	0.0500	0.7250	0.5000
		(0.302)		(0.011)	(0.106)	(0.127)
EBC(2/3)	0.6970	0.5673	0.1167	0.0473	0.7563	0.6126
		(0.266)		(0.009)	(0.072)	(0.096)
PBR(4)	0.3919	0.3199	0.2963	0.0425	0.7699	0.7083
		(0.222)		(0.001)	(0.029)	(0.034)
PBR(6)	0.4952	0.4338	0.1600	0.0437	0.7692	0.6833
		(0.218)		(0.003)	(0.042)	(0.044)
PBR(12)\ RAR	0.6237	0.5199 (0.211)	0.0942	0.0500 (0.012)	0.7654 (0.079)	0.6430 (0.058)
RPBR(6)*	0.4573	0.3699	0.2666	0.0437	0.7669	0.6928
		(0.247)		(0.003)	(0.042)	(0.048)
RPBR(8)*	0.5268	0.4426	0.1871	0.0446	0.7657	0.6738
		(0.248)		(0.005)	(0.050)	(0.057)
RPBR(12)*	0.5996	0.5092	0.1256	0.0466	0.7636	0.6506
		(0.24)		(0.008)	(0.063)	(0.065)
TBD	0.6654	0.5029	0.1548	0.0594	0.7603	0.6128
		(0.252)		(0.020)	(0.111)	(0.058)

Table (6.2): Summary of the behavior of the investigated randomization procedures for $N = 12$. The whole reference set of the presented randomization procedures is used for the assessment.

probability of appearance is evaluated. For Randomized Permuted Block Randomization 100 000 randomization sequences are generated at random with the `R` package `randomizeR` (Schindler et al., 2015). The last three columns of Table 6.2 show the expected values of the investigated criteria and their standard deviation dependent on the randomization procedure. For the average-based approach these three values are mapped with desirability functions to the interval $[0, 1]$. Afterwards, the three desirability scores of the three expected values of the criteria are summarized with the weighted geometric mean. The desirability functions and weights are set according to Table 6.1. The summarized desirability score of the average-based approach $\bar{d}_{\text{geo}}$(AV) is provided in the second column of Table 6.2. A detailed overview of the desirability values of the expected values of the criteria is shown in Table A.7 in the appendix. The column of the $\bar{d}$(RS) values in Table 6.2

*Values are based on 100 000 simulated randomization sequences.

represents the average desirability of the summarized desirability scores of the individual randomization sequences (desirability of the sequence-based approach). For that, the summarized desirability scores of the individual randomization sequences are weighted with their corresponding probability of appearance. In brackets the standard deviation (see Equation (5.2)) of the desirability values attained by the individual randomization sequences is given. The probability of generating an undesired randomization sequence (i.e. a randomization sequence with a $\bar{d}_{\text{geo}}(\boldsymbol{t}_i)$ value of zero) is shown in the fourth column of Table 6.2.

The results presented in Table 6.2 are used for the assessment of the randomization procedures. First, the expected or average values of the three criteria of the investigated randomization procedures are assessed (average-based approach). For Complete Randomization the expected proportion of correct guesses is 0.5. For all other randomization procedures this average value is inflated. Under the assumption of a linear time trend with strength $\vartheta = 1$, the Random Allocation Rule (PBR(12)), BSD(4), and Complete Randomization preserve the type-I-error probability of Student's t-test (size of the test is 0.05). For all other investigated randomization procedures up to the Truncated Binomial Design the type-I-error probability in case of the assumed linear time trend is deflated. For Complete Randomization the expected power in the presence of the assumed linear time trend is the lowest. PBR(4), which is the most susceptible investigated randomization procedure to the convergence strategy, attains the greatest value of the power in case of the assumed linear time trend. For the Big Stick Design and the (Randomized) Permuted Block Randomization the following rule holds: The greater the maximal tolerated imbalance/(maximal possible) block length is, the lower the average of the expected proportion of correct guesses, the lower the average of the power in case of the assumed linear time trend, and the lower the deflation of the average of the type-I-error probability in case of the assumed linear time trend become. That means the greater the restrictions on the randomization process are, the greater the average of the expected proportion of correct guesses, the lower the average of the type-I-error probability in case of the assumed linear time trend, and the greater the average of the power in case of the assumed linear time trend become. Due to the fact that Complete Randomization maintains the type-I-error probability in average and has in expectation a proportion of 0.5

correct guesses, it is the most desired randomization procedure for $N = 12$. The desirability scores for both the average of the type-I-error probability in case of the assumed linear time trend and the average proportion of correct guesses for Complete Randomization are one (see Table A.7 in the appendix). In general, the desirability scores of the average scores lead to the conclusion that the smaller the restrictions on the randomization procedure are, the greater the desirability of the randomization procedure in the investigated setting becomes. Small (random) block lengths seem to be the greatest restriction on the randomization process and a boundary for the maximal tolerated imbalance seems to be the smallest restriction. Particularly, the class of the Big Stick Design has better properties than the investigated EBC(2/3) and the Truncated Binomial Design.

The results of the assessment of the randomization procedures on the basis of the sequence-based approach differ a bit from the results derived on the basis of the average-based approach. First, the column of the probability of generating an undesired randomization sequence is considered. Although PBR(4) has the greatest average value of the power in case of the assumed linear time trend, more than 30% of its generated randomization sequences are undesired. Table A.7 in the appendix shows that for PBR(4) all of these undesired randomization sequences are caused by the criterion of the expected proportion of correct guesses. The lowest probability of generating an undesired randomization sequence has BSD(2). The greater the (maximal possible) block length of the (Randomized) Permuted Block Randomization is, the lower the probability of generating an undesired randomization sequence becomes. The probability of generating an undesired randomization sequence using Complete Randomization is greater than 13%. Among the assessed randomization procedures in Table 6.2 all investigated Big Stick Designs, EBC(2/3), PBR(12), and RPBR(12) have less probability of generating an undesired randomization sequence than Complete Randomization. However, the average desirability of the summarized desirability scores of the individual randomization sequences under Complete Randomization is the fourth highest of the assessed randomization procedures. Solely the investigated Big Stick Designs with MTIs of two, three, and four have greater average desirability scores of the summarized desirability scores of the individual randomization sequences than Complete Randomization. EBC(2/3) performs better than the investigated settings of the (Randomized) Permuted Block

Randomization and the Truncated Binomial Design. Altogether, in the class of the (Randomized) Permuted Block Randomization the Permuted Block Randomization with only one block (Random Allocation Rule/PBR(12)) seems to handle the investigated criteria the best. The properties of the Random Allocation Rule are better than the properties of the Truncated Binomial Design, too.

Summarizing the results for $N = 12$ the class of the Big Stick Design and Complete Randomization have the best properties. Complete Randomization maintains two of the three target values of the criteria. However, Table 6.2 shows that the variance of the three criteria using Complete Randomization is the greatest of all investigated randomization procedures. Dealing with the parameter of the maximal tolerated imbalance is a good approach to reduce the risk of realizing undesired randomization sequences. The greater the parameter of the MTI is, the lower the expected proportion of correct guesses and the lower the power in case of the investigated linear time trend become. If the expected proportion of correct guesses is the main criterion on the randomization process, greater values of the MTI should be chosen. However, if the power in case of a linear time trend is the main criterion smaller values of the MTI should be chosen. The type-I-error probability in case of the assumed linear time trend is maintained or deflated on average for all investigated Big Stick Designs.

6.2 Investigation of the sample size $N = 50$

The performance of several randomization procedures for $N = 50$ is depicted in Table 6.3. In comparison to the situation for $N = 12$, some additional settings of the randomization procedures are investigated. The performance of the randomization procedures is quite similar to the one for $N = 12$. The presented results in the table are based on 100 000 simulated randomization sequences generated for each randomization procedure with the `randomizeR` package (Schindler et al., 2015). The average values of the criteria of the randomization procedures are assessed with the same desirability functions used for $N = 12$ (see Table 6.1). The summarized desirability of the criteria's average values of Complete Randomization is 0.9438. For Complete Randomization the average values of both the type-I-error probability in case of the assumed linear time trend and the expected proportion of correct guesses

Design	$\bar{d}_{\text{geo}}$(AV)	$\bar{d}$(RS) (sd)	$\hat{P}(d(RS)=0)$	$\bar{\alpha}_{\text{CB}}$ (sd)	$1-\bar{\beta}_{\text{CB}}$ (sd)	prop$\overline{\text{CG}}$ (sd)
BSD(2)	0.6965	0.6875	<0.0001	0.0425	0.7741	0.6200
		(0.075)		(0.001)	(0.031)	(0.026)
BSD(3)	0.7976	0.7828	<0.0001	0.0433	0.7730	0.5790
		(0.084)		(0.003)	(0.041)	(0.032)
BSD(4)	0.8485	0.8146	0.0024	0.0454	0.7701	0.5548
		(0.113)		(0.005)	(0.059)	(0.036)
BSD(5)	0.8739	0.8220	0.0092	0.0468	0.7688	0.5422
		(0.142)		(0.007)	(0.068)	(0.039)
BSD(6)	0.8933	0.8163	0.0226	0.0484	0.7663	0.5312
		(0.178)		(0.009)	(0.077)	(0.043)
CR	0.9438	0.7984	0.0515	0.0499	0.7587	0.4999
		(0.228)		(0.012)	(0.087)	(0.067)
EBC(2/3)	0.6899	0.6611	0.0034	0.044	0.7721	0.6217
		(0.125)		(0.005)	(0.048)	(0.045)
PBR(4)	0.4182	0.4060	0.0050	0.042	0.7746	0.7032
		(0.090)		(0.001)	(0.024)	(0.019)
PBR(6)	0.5191	0.5084	0.0004	0.0422	0.7745	0.6779
		(0.088)		(0.001)	(0.027)	(0.023)
PBR(8)	0.5767	0.5660	0.0001	0.0424	0.7741	0.6609
		(0.088)		(0.001)	(0.030)	(0.026)
PBR(50)\ RAR	0.7903	0.7058	0.0408	0.0499	0.7672	0.5792
		(0.182)		(0.012)	(0.084)	(0.037)
RPBR(6)	0.4315	0.4197	0.0046	0.0418	0.7749	0.7002
		(0.094)		(0.001)	(0.019)	(0.020)
RPBR(8)	0.4937	0.4826	0.0011	0.0419	0.7749	0.6849
		(0.094)		(0.001)	(0.022)	(0.024)
RPBR(10)	0.5378	0.5266	0.0005	0.0421	0.7744	0.6726
		(0.094)		(0.001)	(0.025)	(0.026)
RPBR(12)	0.5719	0.5602	0.0001	0.0423	0.7743	0.6624
		(0.095)		(0.001)	(0.028)	(0.028)
TBD	0.7437	0.5740	0.2216	0.0667	0.7529	0.5563
		(0.327)		(0.033)	(0.139)	(0.036)

Table (6.3): Summary of the behavior of the investigated randomization procedures for $N = 50$. For each presented randomization procedure 100 000 randomization sequences are generated for the assessment.

correspond nearly to their according target values of the used desirability functions. The greater the parameter of the MTI of the investigated Big Stick Design is, the greater the summarized desirability value of the criteria's average values becomes. The summarized desirability of the criteria's average values for the Truncated Binomial Design is greater than for EBC(2/3). The Random Allocation Rule has a summarized desirability value of the criteria's average values of greater than 0.75. All other summarized desirability values of the average values of the criteria under the (Randomized) Permuted Block Randomization are lower than 0.6. The greater the (possible) block length of the (Randomized) Permuted Block Randomization is, the greater

the summarized desirability values of the criteria's average values become. A detailed overview of the desirability values of the three criteria of the investigated randomization procedures is given in Table A.8 in the appendix. For all randomization procedures up to the Truncated Binomial Design the desirability score of the average of the type-I-error probability in case of the assumed linear time trend is one. All average values for the power in case of the assumed linear time trend are nearly on the same level in Table 6.3. The lower specification limit of the desirability function for the power in case of the assumed linear time trend is set to 0.6 and the target value is set to 0.8, thus the desirability function is not sensitive in detecting differences in values around 0.77 ± 0.01. Finally, the summarized desirability value of the criteria's average values is mainly influenced by the desirability of the average value of the expected proportion of correct guesses. Complete Randomization maintains the target value 0.5, which is used for assessing the expected proportion of correct guesses. Hence, for Complete Randomization the desirability of this criterion is one. In general, the desirability scores for the average value of both the type-I-error probability and the power in case of the assumed linear time trend of the randomization procedures do not distinguish a lot. In this way, Complete Randomization – considering the average-based approach – is the most recommendable randomization procedure.

The probability of generating an undesired randomization sequence sheds light on the distribution of the criteria. For Complete Randomization the probability of generating an undesired randomization sequence is greater than 5%. Only for the Truncated Binomial Design this value is greater with 22.16%. For all other investigated randomization procedures the probability of generating an undesired randomization sequence is less than 4.1%. BSD(5) has the greatest average desirability score of the summarized desirability scores of the individual randomization sequences. The attained value of 0.822 is almost on the same level than for the Big Stick Design with MTIs of three, four, and six and for Complete Randomization. The best randomization procedure of the class of the (Randomized) Permuted Block Randomization is again the Random Allocation Rule, which has nearly the same average desirability score of the summarized desirability scores of the individual randomization sequences than EBC(2/3). The Truncated Binomial Design

attains an average desirability score of the summarized desirability scores of the individual randomization sequences of 0.574 and performs adequately.

Finally, it should be pointed out that the derived probability of generating an undesired randomization sequence under the Random Allocation Rule presented in Table 6.3 coincides with the asymptotical results derived in Section 3.3. Table 3.9 shows that 99% of the generated randomization sequences under the Random Allocation Rule attain asymptotically a type-I-error probability in case of the assumed linear time trend smaller than 10.06%. Thus, defining a Derringer-Suich desirability function with an upper specification limit of 10% for the type-I-error probability in case of the assumed linear time trend would lead to nearly 1% randomization sequences with a desirability score of zero. The simulated value of generating an undesired randomization sequence under the Random Allocation Rule corresponds to 0.92% (see Table A.8 in the appendix). Thus the asymptotical results under the Random Allocation Rule coincide with the simulated results already for $N = 50$. Furthermore, Table 3.9 shows that 95% randomization sequences under the Random Allocation Rule attain asymptotically a power of more than 61.34% in case of a linear time trend with strength $\vartheta = 1$. This value is close to the used specification limit 0.6 for the desirability function of the power in case of the linear time trend. In the simulation study, which is presented in Table A.8 in the appendix, the Random Allocation Rule has 3.65% undesired randomization sequences concerning the power in case of the assumed linear time trend. The asymptotical results coincide with the simulated results.

6.3 Investigation of the sample size $N = 200$

The performance of several randomization procedures for $N = 200$ is shown in Table 6.4. As for $N = 50$ a simulation study of 100 000 randomization sequences for each randomization procedure is conducted. Considering the average-based approach, again Complete Randomization is the best randomization procedure. In general, the performance of all investigated randomization procedures with respect to the criteria α_{CB} and $1 - \beta_{\text{CB}}$ is nearly the same as for the sample size $N = 50$. Table A.9 in the appendix shows that

Design	$\bar{d}_{\text{geo}}$(AV)	$\bar{d}$(RS) (sd)	$\hat{P}(d(RS)=0)$	$\bar{\alpha}_{\text{CB}}$ (sd)	$1-\bar{\beta}_{\text{CB}}$ (sd)	prop$\overline{\text{CG}}$ (sd)
BSD(2)	0.6884	0.6869 (0.038)	<0.0001	0.0416 (<0.001)	0.7763 (0.015)	0.6238 (0.013)
BSD(3)	0.7936	0.7913 (0.041)	<0.0001	0.0417 (<0.001)	0.7761 (0.018)	0.5822 (0.015)
BSD(4)	0.8423	0.8364 (0.048)	<0.0001	0.0422 (0.001)	0.7755 (0.028)	0.5607 (0.017)
BSD(5)	0.8692	0.8601 (0.053)	<0.0001	0.0427 (0.002)	0.7749 (0.034)	0.5480 (0.018)
BSD(6)	0.8871	0.8721 (0.063)	<0.0001	0.0435 (0.003)	0.7739 (0.043)	0.5391 (0.019)
CR	0.9531	0.8367 (0.221)	0.0464	0.0501 (0.012)	0.7653 (0.086)	0.5000 (0.034)
EBC(2/3)	0.6871	0.6815 (0.062)	<0.0001	0.0420 (0.001)	0.7758 (0.024)	0.6241 (0.022)
PBR(4)	0.3956	0.3935 (0.040)	<0.0001	0.0414 (<0.001)	0.7765 (0.002)	0.7084 (0.008)
PBR(6)	0.5053	0.5033 (0.042)	<0.0001	0.0416 (<0.001)	0.7763 (0.013)	0.6820 (0.011)
PBR(8)	0.5614	0.5598 (0.042)	<0.0001	0.0414 (<0.001)	0.7765 (0.004)	0.6661 (0.012)
PBR(200)\ RAR	0.8729	0.7875 (0.197)	0.0426	0.0500 (0.012)	0.7676 (0.085)	0.5419 (0.021)
RPBR(6)	0.4240	0.4216 (0.044)	<0.0001	0.0414 (<0.001)	0.7764 (0.009)	0.7021 (0.010)
RPBR(8)	0.4855	0.4833 (0.045)	<0.0001	0.0415 (<0.001)	0.7764 (0.011)	0.6872 (0.011)
RPBR(12)	0.5626	0.5603 (0.046)	<0.0001	0.0415 (<0.001)	0.7763 (0.013)	0.6657 (0.013)
TBD	0.7606	0.5982 (0.372)	0.2576	0.0717 (0.042)	0.7504 (0.153)	0.5282 (0.020)

Table (6.4): Summary of the behavior of the investigated randomization procedures for $N = 200$. For each presented randomization procedure 100 000 randomization sequences are generated for the assessment.

the main difference between the randomization procedures is caused by different average values of the expected proportion of correct guesses. With the exception of the Truncated Binomial Design all presented randomization procedures maintain or reduce the average type-I-error probability in the presence of the investigated linear time trend. The Truncated Binomial Design attains with an average power of about 75% the lowest power in the presence of the assumed linear time trend of all assessed randomization procedures. All other presented randomization procedures have an average power between 76.5% and 77.7%. The summarized desirability of the criteria's average values under the Random Allocation Rule is 0.8729. This value is close to the corresponding value for BSD(6). All other investigated settings of the

(Randomized) Permuted Block Randomization have a summarized desirability of the criteria's average values which is lower than 0.57. The summarized desirability of the criteria's average values under the Truncated Binomial Design is greater with 0.76 than the corresponding value for EBC(2/3) with 0.69.

Altogether, for $N = 200$ there are only three investigated randomization procedures left, which have a measureable proportion of undesired randomization sequences. The probability of generating an undesired randomization sequence using the Truncated Binomial Design is the greatest with more than 25%. For Complete Randomization and the Random Allocation Rule the probability of generating an undesired randomization sequence is less than 5%. Thus, for these randomization procedures the probability of generating an undesired randomization sequence is nearly the same as for $N = 50$. The greatest average desirability score of the summarized desirability scores of the individual randomization sequences of all investigated randomization procedures has BSD(6) with 0.8721. Excluding the Random Allocation Rule, all investigated settings of the (Randomized) Permuted Block Randomization have an average desirability score of the summarized desirability scores of the individual randomization sequences less than 0.57. Considering the sequence-based approach, it turns out that EBC(2/3) is better than the Truncated Binomial Design. The average desirability score of the summarized desirability scores of the individual randomization sequences of the Truncated Binomial Design is greater than the one for EBC(2/3). The average desirability score of the summarized desirability scores of the individual randomization sequences value for Complete Randomization is 0.8367. This score is at the same level as the corresponding value for BSD(4).

For the Random Allocation Rule the proportions of undesired randomization sequences in the simulation study with respect to both the type-I-error probability and the power in case of the assumed linear time trend are depicted in Table A.9 in the appendix. These simulated values coincide with the derived asymptotical 99% quantile or alternatively 5% quantile in Table 3.9. As already pointed out in Section 3.3 the asymptotical distributions of both the type-I-error probability and the power in case of a linear time trend under the Random Allocation Rule converge against constant distributions. Consequently, under the Random Allocation Rule the proportion of randomization sequences, which attain greater/lower values for both the type-I-error proba-

bility and the power in case of a linear time trend than a given specification limit, is constant. Finally, it seems that both rejection probabilities (α_{CB} and $1 - \beta_{\mathrm{CB}}$) under Complete Randomization converge against a constant value as it is the case for the Random Allocation Rule (see Tables A.8 and A.9). For all other investigated randomization procedures in Table 6.4 this property does not hold.

6.4 Summary and conclusion

Aggregating the results for all sample sizes the class of the Big Stick Design has the best performance under the used settings of the desirability functions. In general, the class of the Big Stick Design with moderate values of the MTI has a low probability of generating undesired randomization sequences and attains high average desirability values of the summarized desirability scores of the individual randomization sequences. The greater the MTI is, the lower the expected proportion of correct guesses and the lower the average value of the power in the presence of the investigated linear time trend become. Due to the fact that the weight for the criterion of the expected proportion of correct guesses is with $1/2$ the greatest, the summarized desirability of the criteria's average values is very sensitive for the criterion of the expected proportion of correct guesses. Complete Randomization corresponds to the randomization procedure described by BSD(N). Under Complete Randomization the target values of the expected proportion of correct guesses and of the type-I-error probability in case of the investigated linear time trend are maintained. However, the variances of these criteria (shown in brackets in Tables 6.2, 6.3, and 6.4), which are not assessed, are the greatest ones of the investigated randomization procedures. Furthermore, the probability of generating undesired randomization sequences is greater for Complete Randomization than for any Big Stick Design with a moderate value of the MTI. Although Complete Randomization has the greatest summarized desirability values of the criteria's average values for all investigated sample sizes, the class of the Big Stick Design with moderate values of the MTI is more recommendable. This is reflected by the fact that on the one hand the class of the Big Stick Design with moderate values of the MTI has a lower or even zero probability of generating undesired randomization sequences and on the other hand the Big Stick Design attains partly greater average

desirability values of the summarized desirability scores of the individual randomization sequences than Complete Randomization.

For the class of the widespread used (Randomized) Permuted Block Randomization it holds that the greater the (possible) block length is, the greater the derived desirability scores for both the average-based and the sequence-based approach becomes. Thus, the Random Allocation Rule (PBR(N)) performs the best of this class. Due to the fact that the susceptibility of the Random Allocation to the convergence strategy decreases as N increases, the desirability of the Random Allocation Rule increases in N - the distributions of both rejection probabilites (α_{CB} and $1 - \beta_{\text{CB}}$) converge against constant distributions (see Section 3.3). For $N = 12$ the Random Allocation Rule is less desired than EBC(2/3) and for $N = 200$ the Random Allocation Rule performs better than EBC(2/3). However, under the Random Allocation Rule there is a non-negligible probability of generating an undesired randomization sequence. If N increases, the Random Allocation Rule has in comparison to the Truncated Binomial Design a worse performance of the expected proportion of correct guesses and a better performance with respect to criteria concerning the rejection probability in case of the assumed linear time trend. Overall, the Truncated Binomial Design performs somewhere between EBC(2/3) and the Random Allocation Rule. Nevertheless, the Truncated Binomial Design should be used carefully in situations of a linear time trend, due to its clearly elevated type-I-error probability in case of the investigated linear time trend.

Chapter 7

Discussion

In the beginning of this thesis a review on the work on randomization procedures in the literature is given. Inter alia the notation and terminology is introduced and the later investigated randomization procedures are presented. Furthermore, the evaluation and assumptions of an (un)biased randomized clinical trial are described. For that, the population models affected by selection as well as chronological bias are introduced. One criterion of measuring selection and chronological bias is the influence of the according bias on the distorted test decision. This type of assessment is recommended in the ICH E9 (1998) guideline.

In Chapter 3 a two-armed clinical trial is investigated which is planned to be evaluated with the two sample Student's t-test to detect a difference between the effects of the given treatments. For the assessment of chronological bias the influence of a time trend on the test statistic of Student's t-test is derived. The analysis is assumed to be conducted unadjustedly for this time trend. It is shown that in the presence of a time trend and a possible difference between the effects of the treatments the test statistic of Student's t-test is doubly-noncentral t-distributed. Based on this result the rejection probability of Student's t-test in the presence of a (linear) time trend and a difference between the effects of the two treatments is computed. It is assumed that all randomization sequences, which are generated by a given randomization procedure, are affected in the same manner by a (linear) time trend. Furthermore, both scenarios with and without a difference between the effects of the two treatments are investigated. It is shown that the rejection probability of the null hypothesis of Student's t-test in the situation of a time

trend is strongly dependent on the realized randomization sequence. Thus, an expected rejection probability for a fixed randomization procedure is derived by weighting the rejection probabilities of the individual randomization sequences with their probability of appearance. Additional to the rejection probability of a randomization procedure, the standard deviation and even a density function of the rejection probabilities is derived. Finally, the results for chronological bias are transferred on the model assumed for selection bias. At the end of Chapter 3, the central limit theorem for samples from a finite population according to Hajek (1960) is applied on occurring chronological bias under the Random Allocation Rule. The asymptotical behavior of the rejection probability of Student's t-test under the Random Allocation Rule is derived for a linear time trend and a step trend.

Based on the theoretical results derived in Chapter 3 the aim of this thesis is to assess several settings of restricted randomization procedures with respect to their susceptibility to selection as well as chronological bias. Thus, in Chapter 4 a desirability index is introduced, which is applicable on all imaginable criteria measured on randomization sequences or randomization procedures. A generalization of the presented framework is depicted in Figure 7.1. At the top of the figure the clinical study with its corresponding objectives on the randomization process is depicted. Afterwards, the criteria, which are used for measuring the objectives, and suitable desirability functions have to be selected. Finally, a unified assessment score can be used for the comparison of different randomization procedures. In this thesis the investigated objectives on the randomization process are the susceptibility to selection as well as chronological bias. The measured criteria are the expected proportion of correct guesses, and the biased rejection probabilities under both the null hypothesis and the alternative hypothesis in case of a small (linear) time trend. The desirability functions are selected from the class of the desirability functions introduced by Derringer and Suich (1980). Furthermore, two approaches of applying desirability functions on the criteria are contrasted - the sequence-based approach and the average-based approach. In the sequence-based approach the criteria are considered to be problems handled conditioned on the individual randomization sequences and in the average-based approach the average behavior of a criterion conditioned on a randomization procedure is assessed. In both approaches the derived values for the criteria are evaluated with desirability functions. The derived

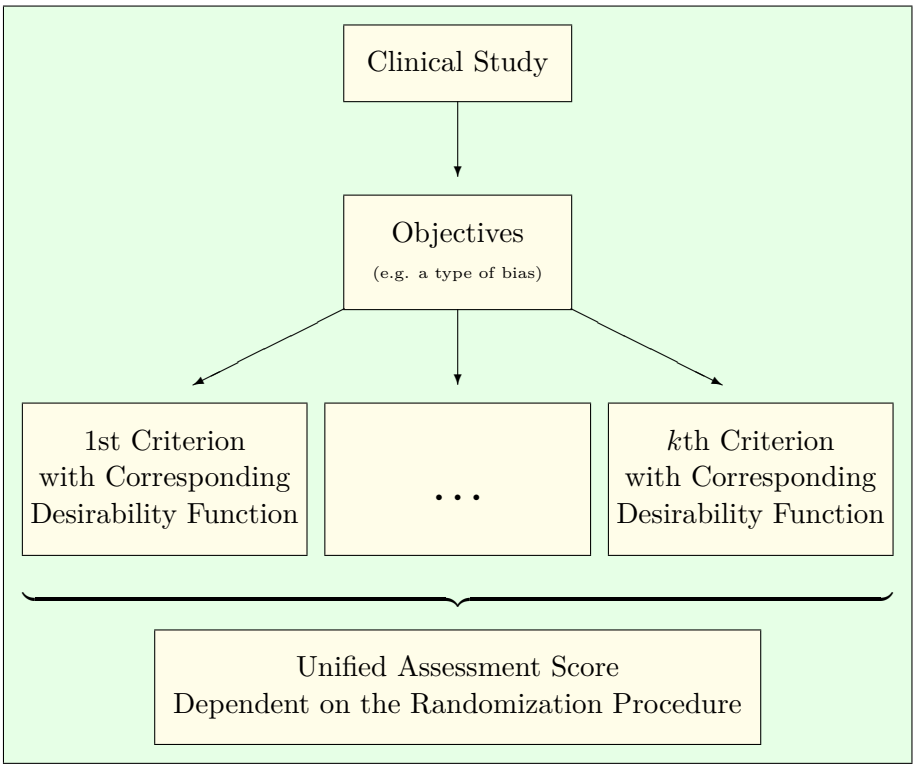


Figure (7.1): Generalization of the presented framework for the assessment of randomization procedures dependent on the clinical study and its objectives on the randomization process.

desirability scores for the criteria are summarized with the geometric mean, which serves as linked optimization criterion. In this way a unified assessment score is derived. This assessment score is used for the selection of an appropriate randomization procedure dependent on the objectives of a clinical study. In Chapter 5 a sensitivity analysis of both approaches is conducted for EBC(2/3). Therefore, the sample sizes $N \in \{4, 12, 50, 200\}$ are investigated. The assessment of several settings of different randomization procedures with respect to their susceptibility to selection as well as chronological bias is carried out in Chapter 6. Both the average-based approach and the sequence-based approach are used to assess randomization procedure. Therefore, the derived settings of the desirability functions in the sensitivity analysis are taken into account. Due to the low probability of generating undesired randomization sequences (at least one desirability score of one of the criteria conditioned on a randomization sequence corresponds to zero) the Big Stick

Design is a highly recommendable randomization procedure. Furthermore, the Big Stick Design with its parameter MTI is easily adjustable for the different sample sizes. The greater the sample size becomes, the greater the parameter of the MTI should be selected. When summarizing the derived desirability scores of the criteria with the geometric mean to one unified score, half of the weight is put on the expected proportion of correct guesses and the other half is split equally between the type-I-error probability and power in the presence of the assumed linear time trend.

Under Complete Randomization the allocations of the patients to the treatment groups are independent. There is no increased probability for the investigator to guess future allocations with the knowledge of past patient allocations. The expectation of the proportion of correct guesses is 0.5. Consequently, Complete Randomization has the best properties with respect to the expected proportion of correct guesses. One disadvantage of using Complete Randomization is the probability of the realization of randomization sequences which lead to high imbalances within the numbers of patients assigned to the two treatment groups at the end of a clinical trial. This objective on the randomization process is not reflected enough by the derived desirability scores. Solely, the derived value of the power in the presence of a (linear) time trend takes this objective partially into account. Considering the average-based approach Complete Randomization is the best randomization procedure. However, the sequence-based approach reveals that there is a non-negligible probability of generating an undesired randomization sequence using Complete Randomization independent of the investigated sample sizes.

The investigated settings of the (Randomized) Permuted Block Randomization are the most susceptible to the convergence strategy. The key property of the Permuted Block Randomization, which is the realization of always equally sized treatment groups at the end of a clinical trial, is not reflected enough by the desirability scores. The class of the investigated Randomized Permuted Block Randomization introduced in Rosenberger and Lachin (2016, Chapter 3) does not have the property of equally sized treatment groups at the end of a clinical trial. In general, this class has a lower expected proportion of correct guesses in comparison to the Permuted Block Randomization, but does not perform better in case of the investigated linear time trend. In summary, the performance of the class of the (Randomized) Permuted Block Randomization is the worst of the investigated randomization procedures.

Solely, the Random Allocation Rule seems to have adequate properties with respect to its susceptibility to the investigated selection and chronological bias. The Truncated Binomial Design, which is the only randomization procedure with an inflated type-I-error probability in case of the assumed linear time trend, and EBC(2/3) perform worse than the class of the Big Stick Designs and better than the investigated settings of the class of the (Randomized) Permuted Block Randomization, except the Random Allocation Rule.

It is hard to say which approach of applying desirability functions is optimal. Both the sequence-based and the average-based approach have their own advantages and disadvantages. All in all, it is permittable to use both approaches for the assessment of randomization procedures. It should be pointed out that the results and consequences of both approaches are different and should be interpreted differently. In any case, the probability of generating an undesired randomization sequence should be an important criterion for the assessment of randomization procedures. If for example one half of the randomization sequences of a randomization procedure is undesired, the randomization procedure cannot be a desired one even if the target value is maintained. Last but not least an undesired randomization sequence is defined by self-adjustable desirability functions. Thus, the core for the validity of both approaches are the defined desirability functions for the criteria. The used desirability functions according to Derringer and Suich (1980) form only one suitable class of desirability functions. Other forms of desirability functions like the class of desirability functions introduced in Harrington Jr. (1965) are imaginable. Nevertheless, the class of the Derringer-Suich desirability functions provide a huge flexibility to define a suitable desirability function for any presented criterion on a randomization sequence or a randomization procedure.

One criticism on the investigated criterion of the expected proportion of correct guesses could be, that this criterion has no direct influence on the the rejection probability of the test used in the analysis. Investigations with the `randomizeR` package (Schindler et al., 2015) have shown that the criterion of the expected proportion of the correct guesses is strongly correlated with other criteria of measuring selection bias. Including the distorted type-I-error probability researched by Tamm et al. (2012) in case of selection bias with a strength η dependent on the planned effect size

Δ_0 instead of the expected proportion of correct guesses would lead to nearly the same recommendations as derived in Chapter 6. Furthermore, the correct guesses are strongly dependent on the guessing strategy. For Complete Randomization and the Big Stick Design the used convergence strategy is even the wrong guessing strategy. However, for these randomization procedures in this thesis as also in the literature, it is assumed that the investigator is wrongly assuming a balanced clinical trial. Investigating the Big Stick Design with the correct guessing strategy (guess only deterministic allocations) leads to nearly the same recommendations as presented in Chapter 6. For Complete Randomization even no best guessing strategy is possible - it would be the random guess for every patient. The conclusion would be that Complete Randomization would be even more desired than it is already in the extensive analysis of this thesis. The presented recommendation of an appropriate randomization procedure in this thesis in the presence of chronological bias is strongly dependent on the assumed (small) linear time trend in the analysis. The presented approach for measuring the influence of an unadjusted time trend on the test decision of Student's t-test is transferable to any imaginable time trend or even any unobserved covariate in the presented linear model. It is obvious, that if a stronger/smaller linear time trend is assumed to affect the data, the upper/lower specification limits of the used desirability functions of the according criteria should be adjusted. Further investigations concerning a step time trend in the middle of a clinical trial instead of the linear time trend would lead to nearly the same recommendations of the randomization procedures as presented in this thesis. Only the scale and range of the desirability scores would be changed. In Chapter 3 the asymptotical behavior of Student's t-test under the Random Allocation Rule is derived. For these calculations no time costly simulations are necessary. Thus it is imaginable to use the behavior of the Random Allocation Rule as benchmark for other randomization procedures. In this way useful boundaries for the desirability functions can be established.

Last but not least, the ICH E9 (1998) guideline suggests to incorporate the randomization process itself in the analysis. For the Permuted Block Randomization it is imaginable that the analysis is performed adjustedly for the used block lengths. Tamm and Hilgers (2014) investigated such a scenario under the presence of chronological bias and showed that incorporating the used block length in the analysis leads to a preserved type-I-error probability.

Unfortunately, no statement with respect to the impact on the power is given. However, the aspect of incorporating a randomization procedure itself in the analysis was not of major interest in this thesis.

Finally, the flexibility of the presented framework, which is shown in Figure 7.1, for the assessment of randomization procedures with the help of desirability functions should be pointed out again. It is imaginable that selection and chronological bias are assessed with other criteria as presented in this thesis. Even the inclusion of further criteria is imaginable. As pointed out above one could favor the Permuted Block Randomization due to its good balancing behavior and punish the high possible imbalances of Complete Randomization. The loss (see Atkinson, 2001) in the estimation of the treatment difference would be an excellent criterion for the assessment of the balancing behavior. In the `R` package `randomizeR` (Schindler et al., 2015) all presented randomization procedures in this thesis are implemented and the printed results in the Chapters 5 and 6 in this thesis can be repeated with other/more criteria. Additionally, the desirability scores can be visualized in a so called radar plot.

For all practical purposes it is recommended that before starting a clinical trial the statistician and the physician determine together both the objectives and its corresponding criteria for the assessment of randomization procedures. Afterwards, they adjust on the one hand the desirability functions for the selected criteria and on the other hand the weights for the summary of the single desirability values with the geometric mean. In this way, several randomization procedures are compared on the basis of the unified assessment criterion. At the end, the randomization procedure with greatest unified assessment score for the planned clinical trial is selected. The process of selecting an appropriate randomization procedure should be part of the trial and analysis plan of any planned clinical trial. On the basis of the presented framework in this thesis a scientific choice of an appropriate randomization procedure dependent on a clinical trial and its objectives is possible.

Appendix A

Additional derivations, graphics, and tables

A.1 Derivations

Proposition A.1:
Under the assumption that the coin is tossed fairly in case of balance between the two treatment groups, the relationship between the average proportion of correct guesses under the convergence strategy introduced in Equation (2.17) *and the number of returns to origin defined in Equation* (2.4) *is given by:*

$$\begin{aligned}\overline{CG}_{RP}(G_{CS}) &= \sum_{\boldsymbol{t}\in\Gamma_N^{RP}} \sum_{n=1}^{N} E\left(\mathbb{1}_{\{t_n = G_{CS}(n,\boldsymbol{t})\}}\right) p_{\boldsymbol{t}} \\ &= \sum_{\boldsymbol{t}\in\Gamma_N^{RP}} \left(\frac{nro(N,\boldsymbol{t})}{2} + \frac{N}{2} - \mathbb{1}_{\{|D(N,\boldsymbol{t})|>0\}} \frac{|D(N,\boldsymbol{t})|-1}{2}\right) p_{\boldsymbol{t}},\end{aligned}$$

where $D(N,\boldsymbol{t})$ according to Equation 2.3 defines the imbalance of assigned patients to the two treatment groups.

Proof
Basis:
Without loss of generality the proposition is shown for any arbitrary fixed $\boldsymbol{t} \in \Gamma_N$. Afterwards, the Proposition (A.1) follows directly. Below, the

following equality is proven with mathematical induction:

$$\sum_{n=1}^{N} E\left(\mathbb{1}_{\{t_n=G_{\mathrm{CS}}(n,\boldsymbol{t})\}}\right) = \frac{\mathrm{nro}(N,\boldsymbol{t})}{2} + \frac{N}{2} - \mathbb{1}_{\{|D(N,\boldsymbol{t})|>0\}}\frac{|D(N,\boldsymbol{t})|-1}{2},$$

Let $N = 1$, thus it follows:

$$\begin{aligned}\sum_{n=1}^{1} E\left(\mathbb{1}_{\{t_n=G_{\mathrm{CS}}(n,\boldsymbol{t})\}}\right) &= E\left(\mathbb{1}_{\{t_1=G_{\mathrm{CS}}(1,\boldsymbol{t})\}}\right) = \frac{1}{2} \\ &= \underbrace{\frac{\mathrm{nro}(1,\boldsymbol{t})}{2}}_{0} + \frac{1}{2} - \underbrace{\mathbb{1}_{\{|D(1,\boldsymbol{t})|>0\}}\frac{|D(1,\boldsymbol{t})|-1}{2}}_{0}.\end{aligned}$$

Inductive step:
Let $N \to N+1$ and $\boldsymbol{t} \in \{-1,1\}^K$ with $K \geq N+1$, thus it follows:

$$\begin{aligned}&\sum_{n=1}^{N+1} E\left(\mathbb{1}_{\{t_n=G_{\mathrm{CS}}(n,\boldsymbol{t})\}}\right) \\ &= \sum_{n=1}^{N} E\left(\mathbb{1}_{\{t_n=G_{\mathrm{CS}}(n,\boldsymbol{t})\}}\right) + E\left(\mathbb{1}_{\{t_{N+1}=G_{\mathrm{CS}}(N+1,\boldsymbol{t})\}}\right) \\ &\overset{\text{Basis}}{=} \frac{\mathrm{nro}(N,\boldsymbol{t})}{2} + \frac{N}{2} - \mathbb{1}_{\{|D(N,\boldsymbol{t})|>0\}}\frac{|D(N)|-1}{2} + E\left(\mathbb{1}_{\{t_{N+1}=G_{\mathrm{CS}}(N+1,\boldsymbol{t})\}}\right) \\ &= \frac{\mathrm{nro}(N+1,\boldsymbol{t})}{2} + \frac{N+1}{2} - \mathbb{1}_{\{|D(N+1,\boldsymbol{t})|>0\}}\frac{|D(N+1,\boldsymbol{t})|-1}{2}\end{aligned}$$

Hence, for the mathematical induction it is necessary to show:

$$\begin{aligned}&E\left(\mathbb{1}_{\{t_{N+1}=G_{\mathrm{CS}}(N+1,\boldsymbol{t})\}}\right) \\ &= \frac{\mathrm{nro}(N+1,\boldsymbol{t})}{2} + \frac{N+1}{2} - \mathbb{1}_{\{|D(N+1,\boldsymbol{t})|>0\}}\frac{|D(N+1,\boldsymbol{t})|-1}{2} \\ &- \left(\frac{\mathrm{nro}(N,\boldsymbol{t})}{2} + \frac{N}{2} - \mathbb{1}_{\{|D(N,\boldsymbol{t})|>0\}}\frac{|D(N,\boldsymbol{t})|-1}{2}\right) \\ &= \frac{\mathrm{nro}(N+1,\boldsymbol{t})}{2} + \frac{1}{2} - \mathbb{1}_{\{|D(N+1,\boldsymbol{t})|>0\}}\frac{|D(N+1,\boldsymbol{t})|-1}{2} \\ &- \left(\frac{\mathrm{nro}(N,\boldsymbol{t})}{2} - \mathbb{1}_{\{|D(N,\boldsymbol{t})|>0\}}\frac{|D(N,\boldsymbol{t})|-1}{2}\right) \qquad \text{(A.1)}\end{aligned}$$

Now, the following three cases must be differentiated:

$$E\left(\mathbb{1}_{\{t_{N+1}=G_{\text{CS}}(N+1,\boldsymbol{t})}\right) = \begin{cases} 1 & \Leftrightarrow t_{N+1} = G_{\text{CS}}(N+1,\boldsymbol{t}) \quad (1) \\ 0.5 & \Leftrightarrow G_{\text{CS}}(N+1,\boldsymbol{t}) = 2\,\mathbb{B}(0.5) - 1 \quad (2) \\ 0 & \Leftrightarrow t_{N+1} \neq G_{\text{CS}}(N+1,\boldsymbol{t}) \quad (3) \end{cases}$$

$$\Rightarrow \begin{cases} N_E(N,\boldsymbol{t}) \neq N_C(N,\boldsymbol{t}) & \Leftrightarrow |D(N,\boldsymbol{t})| \geq 1 \quad (1) \\ N_E(N,\boldsymbol{t}) = N_C(N,\boldsymbol{t}) & \Leftrightarrow |D(N,\boldsymbol{t})| = 0 \quad (2) \\ N_E(N,\boldsymbol{t}) \neq N_C(N,\boldsymbol{t}) & \Leftrightarrow |D(N,\boldsymbol{t})| \geq 1 \quad (3) \end{cases}$$

Considering the case (1) which means that the $(N+1)$th patient is assigned to the treatment group with fewer included patients before his enrollment. Two further differentiations must be done for $|D(N+1,\boldsymbol{t})|$ to show the equality in Equation (A.1) assuming case (1):

$$\begin{cases} |D(N+1,\boldsymbol{t})| & > 0 \Leftrightarrow |D(N,\boldsymbol{t})| > 1 \quad (1.\text{a}) \\ |D(N+1,\boldsymbol{t})| & = 0 \Leftrightarrow |D(N,\boldsymbol{t})| = 1 \quad (1.\text{b}) \end{cases}$$

For the first case $(1.a)$ it follows:

$$\begin{aligned} & E\left(\mathbb{1}_{\{t_{N+1}=G_{\text{CS}}(N+1,\boldsymbol{t})\}}\right) = 1 \\ & = \underbrace{\frac{\text{nro}(N+1,\boldsymbol{t})}{2}}_{\overset{(1.\text{a})}{=}\text{nro}(N,\boldsymbol{t})\,0.5} + \frac{1}{2} - \underbrace{\mathbb{1}_{\{|D(N+1,\boldsymbol{t})|>0\}}}_{\overset{(1.\text{a})}{=}1} \underbrace{\frac{|D(N+1,\boldsymbol{t})|-1}{2}}_{\overset{(1.\text{a})}{=}(|D(N,\boldsymbol{t})|-1)\,0.5-0.5} \\ & - \left(\frac{\text{nro}(N,\boldsymbol{t})}{2} - \mathbb{1}_{\{|D(N,\boldsymbol{t})|>0\}}\frac{|D(N,\boldsymbol{t})|-1}{2}\right) \\ & = \frac{1}{2} + \frac{1}{2} = 1 \end{aligned}$$

For the second case $(1.b)$ it follows:

$$\begin{aligned} & E\left(\mathbb{1}_{\{t_{N+1,\boldsymbol{t}}=G_{\text{CS}}(N+1,\boldsymbol{t})\}}\right) = 1 \\ & = \underbrace{\frac{\text{nro}(N+1,\boldsymbol{t})}{2}}_{\overset{(1.\text{b})}{=}\text{nro}(N,\boldsymbol{t})\,0.5+0.5} + \frac{1}{2} - \underbrace{\mathbb{1}_{\{|D(N+1,\boldsymbol{t})|>0\}}}_{\overset{(1.\text{b})}{=}0} \frac{|D(N+1,\boldsymbol{t})|-1}{2} \end{aligned}$$

$$
\begin{aligned}
&- \left(\frac{\mathrm{nro}(N, \boldsymbol{t})}{2} - \underbrace{\mathbb{1}_{\{|D(N,\boldsymbol{t})|>0\}}}_{\overset{(1.b)}{=}1} \underbrace{\frac{|D(N, \boldsymbol{t})| - 1}{2}}_{\overset{(1.b)}{=}0} \right) \\
&= \frac{1}{2} + \frac{1}{2} = 1
\end{aligned}
$$

Considering the case (2) one further differentiation is not necessary and it holds $|D(N, \boldsymbol{t})| = 0$ and $|D(N+1, \boldsymbol{t})| = 1$. The equality in Equation (A.1) can be shown analogously to the case (1). Finally, the equality in Equation (A.1) for the last case (3) must be shown. For this, it holds $|D(N+1, \boldsymbol{t})| = |D(N, \boldsymbol{t})| + 1$ and the equality in Equation (A.1) follows directly. Thus, Equation (A.1) holds for all $\boldsymbol{t} \in \Gamma_N$ and the Proposition (A.1) is shown.

□

Lemma A.2:
Let $\boldsymbol{b} \in \mathbb{R}^N$, thus the following equality holds:

$$
\sum_{i=1}^{N} b_i^2 - \frac{1}{N-1} \sum_{\substack{i,j \in \{1,2,\ldots,N\} \\ i \neq j}} b_i \, b_j = \frac{N}{N-1} \sum_{i=1}^{N} (b_i - \bar{b})^2 \, .
$$

Proof

$$
\begin{aligned}
\sum_{i=1}^{N} b_i^2 - \sum_{\substack{i,j \in \{1,2,\ldots,N\} \\ i \neq j}} \frac{1}{N-1} \, b_i \, b_j &= \frac{1}{N-1} \left((N-1) \sum_{i=1}^{N} b_i^2 - \sum_{\substack{i,j \in \{1,2,\ldots,N\} \\ i \neq j}} b_i \, b_j \right) \\
&= \frac{1}{N-1} \left(N \sum_{i=1}^{N} b_i^2 - \sum_{\substack{i,j \in \{1,2,\ldots,N\} \\ i \neq j}} b_i \, b_j - \sum_{i=1}^{N} b_i^2 \right) \\
&= \frac{1}{N-1} \left(N \sum_{i}^{N} b_i^2 - \sum_{i,j \in \{1,2,\ldots,N\}} b_i \, b_j \right) \\
&= \frac{1}{N-1} \left(N \sum_{i=1}^{N} b_i^2 - 2 \sum_{i,j \in \{1,2,\ldots,N\}} b_i \, b_j + \sum_{i,j \in \{1,2,\ldots,N\}} b_i \, b_j \right)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{N-1}\left(N\sum_{i=1}^{N} b_i^2 - 2\sum_{i=1}^{N} b_i \frac{N}{N}\sum_{j=1}^{N} b_j + \frac{N^2}{N^2}\left(\sum_{i=1}^{N} b_i\right)^2\right) \\
&= \frac{N}{N-1}\left(\sum_{i=1}^{N} b_i^2 - 2\sum_{i=1}^{N} b_i\,\bar{b} + N\,\bar{b}^2\right) \\
&= \frac{N}{N-1}\sum_{i=1}^{N}\left(b_i^2 - 2\,b_i\,\bar{b} + \bar{b}^2\right) \\
&= \frac{N}{N-1}\sum_{i=1}^{N}\left(b_i - \bar{b}\right)^2 .
\end{aligned}
$$

□

Lemma A.3: (Covariance under RAR)
Assuming $\Gamma_N := \{-1, 1\}^N$*, the covariance of two random allocations* T_i *and* T_j *with* $i \neq j$ *and* $i, j \in \{1, 2, \ldots, N\}$ *under the Random Allocation Rule, which is introduced in Section 2.2.2, is given by:*

$$Cov(T_i, T_j) = E(T_i\,T_j) = \frac{-1}{N-1}\,. \tag{A.2}$$

Proof
Under the Random Allocation Rule the expectation of any T_i with $i \in \{1, 2, \ldots, N\}$ is given by $E(T_i) = 0$. The expectation of $T_i\,T_j$ with $i, j \in \{1, 2, \ldots, N\}$ and $i \neq j$ can be expressed as follows:

$$
\begin{aligned}
E(T_i\,T_j) &= P(T_i = 1|T_j = 1)\,P(T_j = 1) + P(T_i = -1|T_j = -1)\,P(T_j = -1) \\
&\quad - P(T_i = 1|T_j = -1)\,P(T_j = -1) - P(T_i = -1|T_j = 1)\,P(T_j = 1) \\
&= 2\,P(T_i = 1|T_j = 1)\,P(T_j = 1) - 2\,P(T_i = 1|T_j = -1)\,P(T_j = -1)\,.
\end{aligned}
$$

For the Random Allocation Rule $P(T_i = 1|T_j = 1) = P(T_i = -1|T_j = -1)$ holds. Below, for the Random Allocation Rule the following calculations are carried out:

- $P(T_i = 1|T_j = 1)\,P(T_j = 1) = \dfrac{N/2 - 1}{N-1} \cdot \dfrac{1}{2} = \dfrac{N-2}{4\,(N-1)}$
- $P(T_i = 1|T_j = -1)\,P(T_j = -1) = \dfrac{N/2}{N-1} \cdot \dfrac{1}{2} = \dfrac{N}{4\,(N-1)}\,.$

Thus, $E(T_i\,T_j)$ can be written as:

$$\begin{aligned} E(T_i\,T_j) &= 2\left(\frac{N-2}{4\,(N-1)}\right) - 2\left(\frac{N}{4\,(N-1)}\right) \\ &= 2\left(\frac{-2}{4\,(N-1)}\right) \\ &= \frac{-1}{N-1}\,. \end{aligned}$$

For the covariance $\mathrm{Cov}(T_i, T_j) = E(T_i\,T_j) - E(T_i)E(T_j)$ with $E(T_i) = E(T_j) = 0$ Equation (A.2) follows.

□

Lemma A.4: (Distribution of $\epsilon\,T$)
Let $\Gamma := \{-1, 1\}$ *and* T *be a random variable which takes values in* Γ *with* $P(T = 1) = P(T = -1) = 0.5$. *The random variable* ϵ *is assumed to be standard normally distributed and is independent of* T. *Then, for* $P(T\,\epsilon \leq a)$ *with* $a \in \mathbb{R}$ *it follows:*

$$P(T\,\epsilon \leq a) = P(\epsilon \leq a)\,.$$

Proof

$$\begin{aligned} P(T\,\epsilon \leq a) &= P(T = 1)\,P(\epsilon \leq a) + P(T = -1)\,P(-\epsilon \leq a) \\ &= \frac{1}{2}\,(P(\epsilon \leq a) + 1 - P(\epsilon > -a)) \\ &= \frac{1}{2}\,(P(\epsilon \leq a) + 1 - 1 + P(\epsilon \leq a)) \\ &= P(\epsilon \leq a)\,. \end{aligned}$$

□

A.2 Graphics

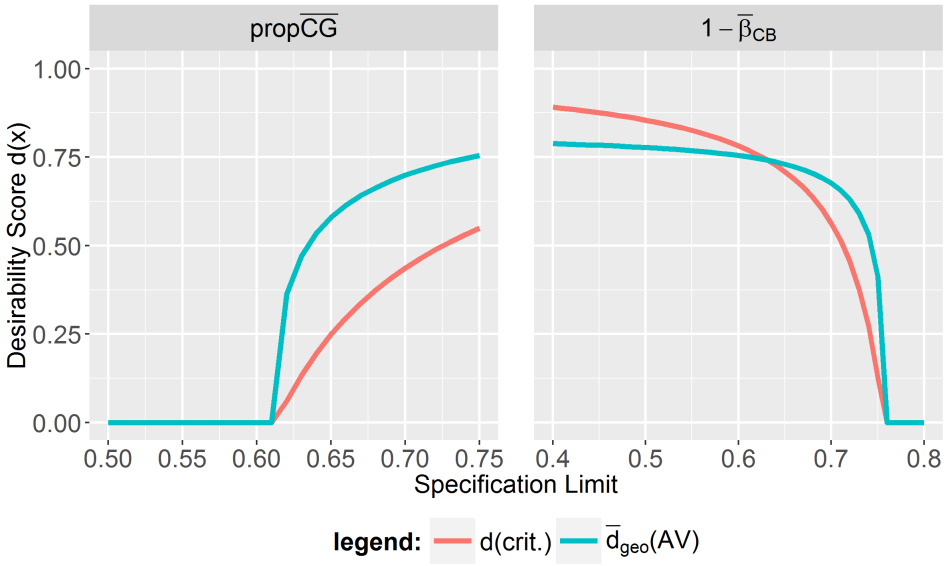


Figure (A.1): Assessment of the desirability scores of the criteria's average values under EBC(2/3) for $N = 12$ dependent on different specification limits.

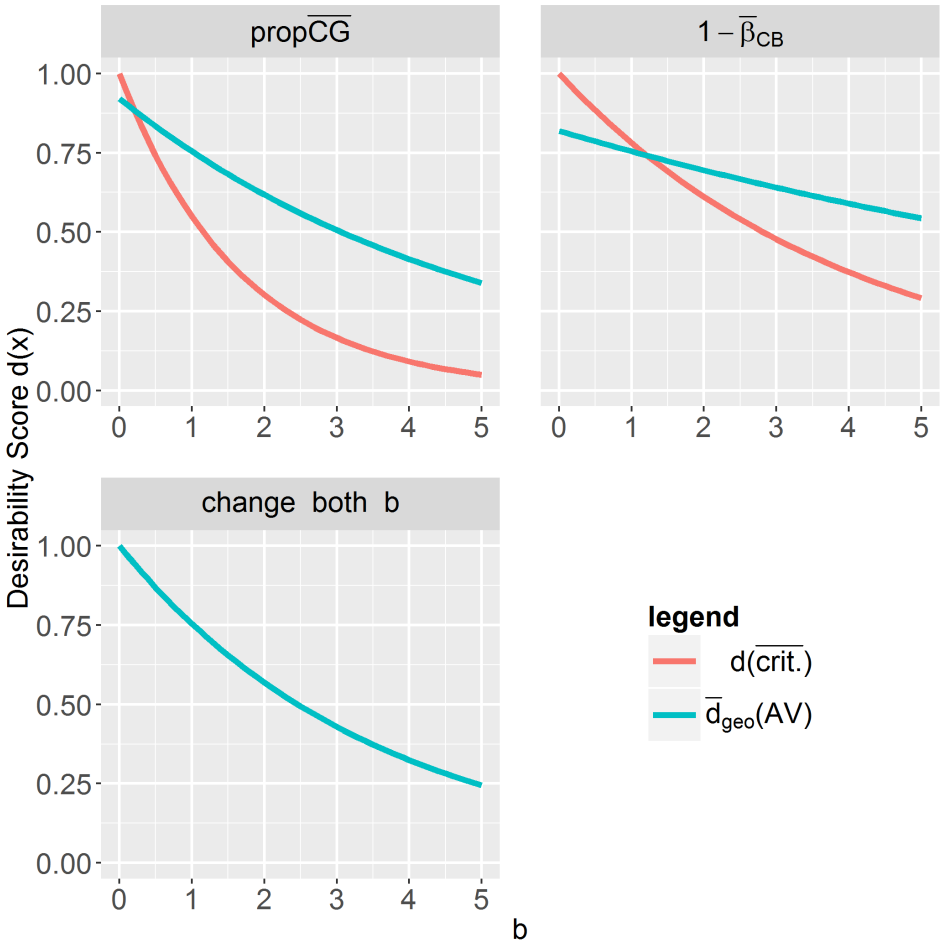


Figure (A.2): Assessment of desirability scores of the criteria's average values under EBC(2/3) for $N = 12$ dependent on the choice of the parameter b.

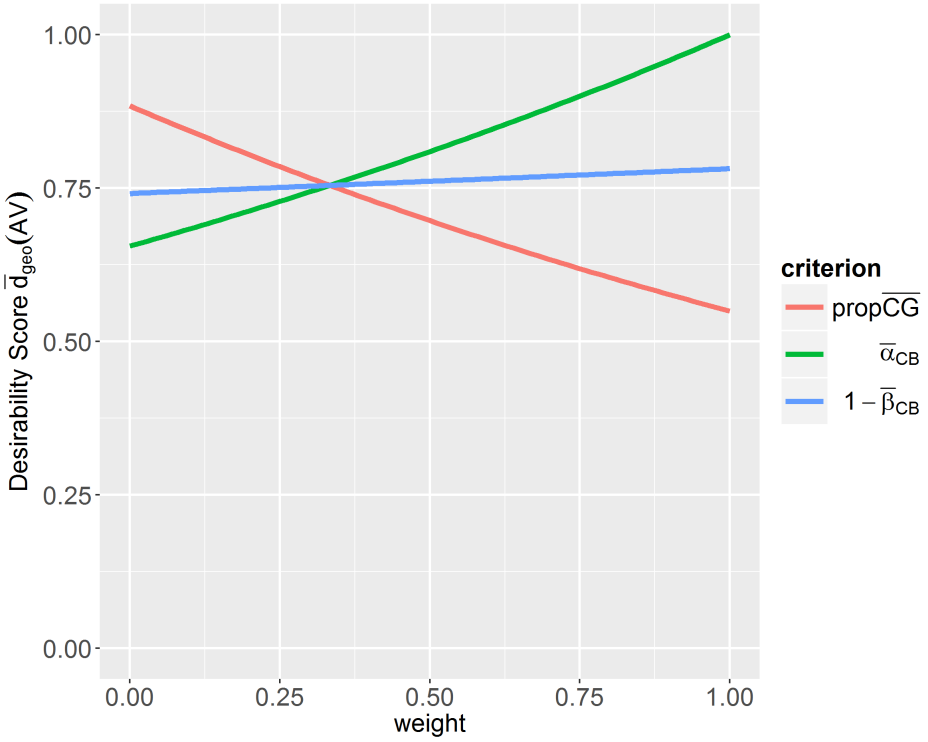


Figure (A.3): Assessment of the summarized desirability score of the average values under EBC(2/3) for $N = 12$ dependent on the weights ω_j of the three criteria.

A.3 Tables

ϑ	$N = 4$		$N = 12$		$N = 50$		$N = 200$		$N = 10.000$
	exact	asymp.	exact	asymp.	sim.	asymp.	sim.	asymp.	asymp.
0.50	0.7919	0.9999	0.7931	0.8677	0.7920	0.8077	0.7918	0.7957	0.7920
1.00	0.7680	0.9997	0.7655	0.8458	0.7671	0.7838	0.7679	0.7717	0.7680
2.00	0.6782	0.9974	0.6696	0.7598	0.6766	0.6950	0.6787	0.6832	0.6796

Table (A.1): Comparison (asymp. vs. exact/sim.) of the achieved power under the Random Allocation Rule using Student's t-test in case of a linear time trend (assuming $W(\boldsymbol{Y}, \boldsymbol{T}) \sim \mathcal{N}(0, 1)$).

	$\tilde{q}_{0.01}$	$\tilde{q}_{0.05}$	$\tilde{q}_{0.25}$	$\tilde{q}_{0.5}$	$\tilde{q}_{0.75}$	$\tilde{q}_{0.95}$	$\tilde{q}_{0.99}$
$b(\boldsymbol{T})$	-0.1918	-0.1356	-0.0556	0.0000	0.0556	0.1356	0.1918
δ_{H_1}	2.1810	2.3797	2.6626	2.8592	3.0559	3.3388	3.5375
λ	3.7050	3.9350	4.1263	4.1650	4.1263	3.9350	3.7050

Table (A.2): Asymptotical quantiles using the Random Allocation Rule for $b(\boldsymbol{T})$, δ, and λ. A linear time trend is assumed for $N = 50$, $\vartheta = 1$, and $\sigma^2 = 1$ under the alternative hypothesis with $\Delta_0 = 0.8087$ and $1 - \beta = 0.8$.

Quantile	$N = 12$				$N = 50$				$N = 200$			
	α_{CB}		$1 - \beta_{\text{CB}}$		α_{CB}		$1 - \beta_{\text{CB}}$		α_{CB}		$1 - \beta_{\text{CB}}$	
	exact	asymp.	exact	asymp.	sim.	asymp.	sim.	asymp.	sim.	asymp.	sim.	asymp.
$\tilde{q}_{0.01}$	0.0240	0.0238	0.2888	0.2776	0.0236	0.0236	0.2199	0.2218	0.0236	0.0236	0.2161	0.2132
$\tilde{q}_{0.05}$	0.0240	0.0239	0.3851	0.3779	0.0237	0.0237	0.3435	0.3464	0.0237	0.0237	0.3404	0.3418
$\tilde{q}_{0.25}$	0.0262	0.0260	0.5590	0.5524	0.0259	0.0259	0.5536	0.5568	0.0260	0.0260	0.5586	0.5586
$\tilde{q}_{0.5}$	0.0331	0.0340	0.6813	0.6811	0.0348	0.0343	0.6977	0.6997	0.0345	0.0344	0.7036	0.7036
$\tilde{q}_{0.75}$	0.0539	0.0556	0.7933	0.7987	0.0566	0.0564	0.8206	0.8184	0.0567	0.0566	0.8228	0.8221
$\tilde{q}_{0.95}$	0.1288	0.1339	0.9175	0.9215	0.1315	0.1303	0.9294	0.9298	0.1286	0.1295	0.9310	0.9312
$\tilde{q}_{0.99}$	0.2074	0.2426	0.9647	0.9689	0.2229	0.2238	0.9695	0.9698	0.2209	0.2201	0.9694	0.9700

Table (A.3): Quantiles of the rejection probability under Student's t-test conditioned on the distribution of $b(\boldsymbol{T})$ using the Random Allocation Rule in case of a linear time trend with strength $\vartheta = 2$, $\sigma^2 = 1$, $\alpha_0 = 0.05$, and $1 - \beta_0 = 0.8$.

Quantile	$N = 12$				$N = 50$				$N = 200$			
	α		$1 - \beta$		α		$1 - \beta$		α		$1 - \beta$	
	exact	asymp.	exact	asymp.	sim.	asymp.	sim.	asymp.	sim.	asymp.	sim.	asymp.
$\tilde{q}_{0.01}$	0.0055	0.0055	0.0413	0.0146	0.0074	0.0055	0.0103	0.0118	0.0056	0.0056	0.0142	0.0115
$\tilde{q}_{0.05}$	0.0055	0.0055	0.1554	0.0571	0.0074	0.0056	0.0892	0.0533	0.0056	0.0056	0.0492	0.0532
$\tilde{q}_{0.25}$	0.0055	0.0076	0.1554	0.2399	0.0074	0.0079	0.2075	0.2541	0.0075	0.0080	0.2951	0.2582
$\tilde{q}_{0.5}$	0.0446	0.0165	0.4548	0.4676	0.0262	0.0178	0.6076	0.5017	0.0140	0.0182	0.5087	0.5094
$\tilde{q}_{0.75}$	0.0446	0.0492	0.8304	0.7193	0.0845	0.0517	0.7969	0.7510	0.0500	0.0523	0.7206	0.7573
$\tilde{q}_{0.95}$	0.3321	0.2289	0.8304	0.9495	0.2121	0.2121	0.9189	0.9521	0.2137	0.2093	0.9570	0.9528
$\tilde{q}_{0.99}$	0.3321	0.5166	0.9872	0.9937	0.4157	0.4388	0.9948	0.9912	0.4117	0.4262	0.9889	0.9909

Table (A.4): Quantiles of the rejection probability under Student's t-test conditioned on the distribution of $b(\boldsymbol{T})$ using the Random Allocation Rule in case of a step time trend with strength $\vartheta = 2$, $\sigma^2 = 1$, $\alpha_0 = 0.05$, and $1 - \beta_0 = 0.8$.

j	Criterion$_j$	$\bar{\boldsymbol{x}}_j$	$s(\boldsymbol{x}_j)$	$\max(\boldsymbol{x}_j)$	$\min(\boldsymbol{x}_j)$	$\tilde{x}_{0.05,j}$	$\tilde{x}_{0.25,j}$	$\tilde{x}_{0.50,j}$	$\tilde{x}_{0.75,j}$	$\tilde{x}_{0.95,j}$
1	propCG	0.622	0.045	0.750	0.340	0.540	0.590	0.630	0.650	0.690
2	α_{CB}	0.044	0.005	0.179	0.041	0.041	0.042	0.042	0.044	0.052
3	$1 - \beta_{\text{CB}}$	0.772	0.048	0.969	0.396	0.690	0.746	0.775	0.803	0.844

Table (A.5): Summary of the properties of the randomization sequences generated by EBC(2/3) for $N = 50$. The results are based on 100 000 simulated randomization sequences.

j	Criterion$_j$	$\bar{\boldsymbol{x}}_j$	$s(\boldsymbol{x}_j)$	$\max(\boldsymbol{x}_j)$	$\min(\boldsymbol{x}_j)$	$\tilde{x}_{0.05,j}$	$\tilde{x}_{0.25,j}$	$\tilde{x}_{0.50,j}$	$\tilde{x}_{0.75,j}$	$\tilde{x}_{0.95,j}$
1	propCG	0.625	0.021	0.685	0.550	0.588	0.610	0.625	0.640	0.660
2	α_{CB}	0.042	0.001	0.053	0.041	0.041	0.041	0.042	0.042	0.044
3	$1-\beta_{\text{CB}}$	0.775	0.025	0.863	0.671	0.731	0.761	0.777	0.792	0.814

Table (A.6): Summary of the properties of the randomization sequences generated by EBC(2/3) for $N = 200$. The results are based on 100 000 simulated randomization sequences.

Design	$d(\bar{\alpha}_{\text{CB}})$	$\bar{d}(\alpha_{\text{CB}})$ (sd)	$P(d(\alpha_{\text{CB}}) = 0)$	$d(1-\bar{\beta}_{\text{CB}})$	$\bar{d}(1-\beta_{\text{CB}})$ (sd)	$P(d(1-\beta_{\text{CB}}) = 0)$	$d(\text{prop}\overline{\text{CG}})$	$\bar{d}$(propCG) (sd)	$P(d(\text{propCG}) = 0)$
BSD(2)	1.0000	0.9763 (0.069)	0.0000	0.8092	0.7609 (0.246)	0.0039	0.5833	0.5807 (0.245)	0.0156
BSD(3)	1.0000	0.9470 (0.126)	0.0000	0.7926	0.7287 (0.284)	0.0134	0.7407	0.7209 (0.259)	0.0156
BSD(4)	0.9983	0.9202 (0.17)	0.0034	0.7379	0.6756 (0.320)	0.0354	0.8724	0.7799 (0.268)	0.0156
CR	0.9998	0.9195 (0.177)	0.0059	0.6248	0.6150 (0.355)	0.1145	1.0000	0.7941 (0.271)	0.0156
EBC(2/3)	1.0000	0.9554 (0.130)	0.0028	0.7817	0.7380 (0.279)	0.0275	0.5495	0.5161 (0.314)	0.0878
PBR(4)	1.0000	1.0000 (<0.001)	0.0000	0.8496	0.8394 (0.128)	0.0000	0.1667	0.1667 (0.136)	0.2963
PBR(6)	1.0000	0.9970 (0.018)	0.0000	0.8459	0.8193 (0.172)	0.0000	0.2667	0.2667 (0.177)	0.1600
PBR(12)\ RAR	0.9992	0.9191 (0.179)	0.0087	0.8272	0.7467 (0.290)	0.0206	0.4278	0.4278 (0.231)	0.0693
RPBR(12)*	1.0000	0.9655 (0.112)	0.0021	0.8181	0.7702 (0.249)	0.0107	0.3974	0.3963 (0.255)	0.1138
RPBR(6)*	1.0000	0.9951 (0.028)	0.0000	0.8345	0.8125 (0.182)	0.0002	0.2289	0.2289 (0.194)	0.2664
RPBR(8)*	1.0000	0.9885 (0.051)	0.0000	0.8285	0.7971 (0.208)	0.0027	0.3049	0.3047 (0.229)	0.1844
TBD	0.8119	0.7744 (0.316)	0.0703	0.8014	0.6799 (0.368)	0.0884	0.5488	0.5488 (0.234)	0.0312

Table (A.7): Summary of the desirability scores for all investigated randomization procedures for $N = 12$. The whole reference set of the presented randomization procedures is used for the assessment.

*Values are based on 100 000 simulated randomization sequences.

Design	$d(\bar{\alpha}_{\text{CB}})$	$\bar{d}(\alpha_{\text{CB}})$ (sd)	$\hat{P}(d(\alpha_{\text{CB}}) = 0)$	$d(1-\bar{\beta}_{\text{CB}})$	$\bar{d}(1-\beta_{\text{CB}})$ (sd)	$\hat{P}(d(1-\beta_{\text{CB}}) = 0)$	$d(\text{prop}\overline{\text{CG}})$	$\bar{d}(\text{propCG})$ (sd)	$\hat{P}(d(\text{propCG}) = 0)$
BSD(2)	1.0000	1.000 (<0.001)	<0.0001	0.8706	0.8542 (0.133)	<0.0001	0.5199	0.5199 (0.106)	<0.0001
BSD(3)	1.0000	0.9986 (0.01)	<0.0001	0.8649	0.8349 (0.167)	<0.0001	0.6840	0.6840 (0.126)	<0.0001
BSD(4)	1.0000	0.9833 (0.057)	<0.0001	0.8505	0.7955 (0.229)	0.0024	0.7806	0.7784 (0.141)	<0.0001
BSD(5)	1.0000	0.9670 (0.091)	<0.0001	0.8438	0.7746 (0.260)	0.0092	0.8313	0.8242 (0.145)	<0.0001
BSD(6)	1.0000	0.9457 (0.132)	0.0012	0.8317	0.7533 (0.287)	0.0219	0.8750	0.8554 (0.147)	<0.0001
CR	1.0000	0.9203 (0.183)	0.0092	0.7936	0.7218 (0.314)	0.0471	1.0000	0.8927 (0.149)	<0.0001
EBC(2/3)	1.0000	0.9898 (0.057)	0.0004	0.8605	0.8251 (0.191)	0.0032	0.5131	0.5123 (0.176)	<0.0001
PBR(4)	1.0000	1.0000 (<0.001)	<0.0001	0.8732	0.8625 (0.105)	<0.0001	0.1871	0.1871 (0.075)	0.0050
PBR(6)	1.0000	1.0000 (<0.001)	<0.0001	0.8724	0.8592 (0.118)	<0.0001	0.2885	0.2885 (0.093)	0.0004
PBR(8)	1.0000	1.0000 (<0.001)	<0.0001	0.8704	0.8555 (0.127)	<0.0001	0.3565	0.3565 (0.104)	0.0001
PBR(50)\ RAR	1.0000	0.9198 (0.183)	0.0092	0.8359	0.7504 (0.300)	0.0365	0.6831	0.6831 (0.148)	<0.0001
RPBR(6)	1.0000	1.0000 (<0.001)	<0.0001	0.8745	0.8688 (0.085)	<0.0001	0.1991	0.1991 (0.081)	0.0046
RPBR(8)	1.0000	1.0000 (0.001)	<0.0001	0.8747	0.8666 (0.099)	<0.0001	0.2606	0.2606 (0.095)	0.0011
RPBR(12)	1.0000	0.9998 (0.004)	<0.0001	0.8717	0.8591 (0.121)	<0.0001	0.3503	0.3503 (0.112)	0.0001
TBD	0.6661	0.7096 (0.370)	0.1320	0.7647	0.6614 (0.394)	0.1559	0.7749	0.7749 (0.144)	<0.0001

Table (A.8): Summary of the desirability scores for all investigated randomization procedures for $N = 50$. For each presented randomization procedure 100 000 randomization are generated for the assessment.

Design	$d(\bar{\alpha}_{\text{CB}})$	$\bar{d}(\alpha_{\text{CB}})$ (sd)	$\hat{P}(d(\alpha_{\text{CB}}) = 0)$	$d(1-\bar{\beta}_{\text{CB}})$	$\bar{d}(1-\beta_{\text{CB}})$ (sd)	$\hat{P}(d(1-\beta_{\text{CB}}) = 0)$	$d(\text{prop}\overline{\text{CG}})$	$\bar{d}(\text{propCG})$ (sd)	$\hat{P}(d(\text{propCG}) = 0)$
BSD(2)	1.0000	1.0000 (<0.001)	<0.0001	0.8814	0.8808 (0.075)	<0.0001	0.5048	0.5048 (0.051)	<0.0001
BSD(3)	1.0000	1.0000 (<0.001)	<0.0001	0.8807	0.8777 (0.088)	<0.0001	0.6711	0.6711 (0.061)	<0.0001
BSD(4)	1.0000	1.0000 (<0.001)	<0.0001	0.8775	0.8639 (0.121)	<0.0001	0.7574	0.7574 (0.068)	<0.0001
BSD(5)	1.0000	0.9999 (0.002)	<0.0001	0.8746	0.8526 (0.143)	<0.0001	0.8079	0.8078 (0.072)	<0.0001
BSD(6)	1.0000	0.9976 (0.015)	<0.0001	0.8697	0.835 (0.174)	<0.0001	0.8438	0.8437 (0.075)	<0.0001
CR	0.9986	0.9180 (0.187)	0.0100	0.8266	0.7436 (0.306)	0.0416	0.9999	0.9448 (0.078)	<0.0001
EBC(2/3)	1.0000	0.9999 (0.004)	<0.0001	0.8791	0.8702 (0.105)	<0.0001	0.5035	0.5035 (0.087)	<0.0001
PBR(4)	1.0000	1.0000 (<0.001)	<0.0001	0.8827	0.8827 (0.010)	<0.0001	0.1666	0.1666 (0.033)	<0.0001
PBR(6)	1.0000	1.0000 (<0.001)	<0.0001	0.8813	0.8811 (0.067)	<0.0001	0.2720	0.2720 (0.044)	<0.0001
PBR(8)	1.0000	1.0000 (<0.001)	<0.0001	0.8826	0.8826 (0.018)	<0.0001	0.3355	0.3355 (0.049)	<0.0001
PBR(200)\ RAR	1.0000	0.9194 (0.184)	0.0091	0.8378	0.7514 (0.302)	0.0382	0.8325	0.8325 (0.083)	<0.0001
RPBR(6)	1.0000	1.0000 (<0.001)	<0.0001	0.8819	0.8819 (0.046)	<0.0001	0.1914	0.1914 (0.039)	<0.0001
RPBR(8)	1.0000	1.0000 (<0.001)	<0.0001	0.8819	0.8817 (0.054)	<0.0001	0.2510	0.2510 (0.045)	<0.0001
RPBR(12)	1.0000	1.0000 (<0.001)	<0.0001	0.8816	0.8807 (0.064)	<0.0001	0.3371	0.3371 (0.053)	<0.0001
TBD	0.5656	0.6770 (0.390)	0.1662	0.7518	0.6580 (0.402)	0.1736	0.8872	0.8872 (0.079)	<0.0001

Table (A.9): Summary of the desirability scores for all investigated randomization procedures for $N = 200$. For each presented randomization procedure 100 000 randomization are generated for the assessment.

Appendix B

Acronyms and notations

B.1 Acronyms

General

Acronym	Meaning
AR	Allocation ratio
C	(Treatment of) the control group
CB	Chronological bias
CG	Correct guesses
CS	Convergence strategy
E	(Treatment of) the experimental group
linT	Linear time trend
ncp	Noncentrality parameter
prob	Probability
prop	Proportion
RP	Randomization procedure
SB	Selection bias
TT	Time trend

Randomization procedures

Acronym	Meaning
BSD(MTI)	Big Stick Design with parameter MTI
CR	Complete Randomization
EBC(p)	Efron's Biased Coin Design with parameter p
PBR(k)	Permuted Block Randomization with block length k
RAR	Random Allocation Rule
RPBR(k)	Randomized Permuted Block Randomization with maximal block length k
TBD	Truncated Binomial Design

B.2 Notations

General

Notation	Meaning
$\mathbf{1}_{n(\times n)} \in \mathbb{R}^{n(\times n)}$	Vector/matrix containing the number one $n(\times n)$ times
α, α_0	(Planned) type-I-error probability
$\mathbb{B}(p)$	Bernoulli distribution with success probability p
β, β_0	(Planned) power
$c_j(\boldsymbol{t})$	Value of the jth criterion of a randomization sequence $\boldsymbol{t}$
$\chi^2(h, \lambda)$	χ^2 distribution with h degrees of freedom and noncentrality parameter λ
Δ, Δ_0	(Planned) effect size of a clinical trial
δ, λ	Noncentrality parameters of the doubly noncentral t-distribution
$\boldsymbol{e}_i \in \mathbb{R}^N$	Unit vector with a one at position i
$F(x)$	Distribution function
$f(x)$	Density function
$H_{\boldsymbol{\psi}}(x)$	Probability mass function of a distribution H dependent on the parameter vector $\boldsymbol{\psi}$
μ_C, μ_E	Expectation of the patients assigned to C/E
N	Total number of patients included into the trial

Notation	Meaning
$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	Normal distribution with expectation vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$
$\text{Poi}(\lambda)$	Poisson distribution with parameter λ
$\text{sgn}(x)$	Sign function for $x \in \mathbb{R}$
$t_{\alpha,n,\delta,\lambda}$	α-quantil of the doubly noncentral t-distribution with n degrees of freedom and noncentrality parameters δ and λ
$t_{n,\delta,\lambda}(x)$	Density of the doubly noncentral t-distribution with n degrees of freedom and noncentrality parameters δ and λ at the position $x \in \mathbb{R}$
$\tau_\vartheta(n)$	Time trend function dependent on the patient number n with the strength ϑ
W	Test statistic of Student's t-test

Randomization

Notation	Meaning
$\text{AR}_C(n, \boldsymbol{T})$, $\text{AR}_E(n, \boldsymbol{T})$	Allocation ratio of treatment group C/E after n assigned patients dependent on $\boldsymbol{T}$
avnda_{RP}	Average number of deterministic allocations of a given randomization procedure
$\text{avnCG}_{\text{RP}}(G)$, $\text{avpCG}_{\text{RP}}(G)$	Average number/proportion of correct guesses of a given randomization procedure dependent on the guessing function G
avnro_{RP}	Average number of returns to origin of a given randomization procedure
$D(n, \boldsymbol{T})$	Imbalance between C and E after n assigned patients dependent on $\boldsymbol{T}$
η	Strength of selection bias
$G(n, \boldsymbol{t})$	Guessing function for the nth patient dependent on $\boldsymbol{t} \in \Gamma^N$
Γ_N, Γ_N^{RP}	Space of all possible randomization sequences (of a given randomization procedure) in a two-armed clinical trial with N included patients

Notation	Meaning
MTI, $\mathrm{MTI}_{\mathrm{RP}}$	Maximal tolerated imbalance (of a given randomization procedure)
$N_C(n, \boldsymbol{T})$, $N_E(n, \boldsymbol{T})$	Number of patients assigned to C/E after n allocations dependent on $\boldsymbol{T}$
N_C, N_E	Final number of patients assigned to C/E
$\mathrm{nda}(n, \boldsymbol{T})$	Number of deterministic allocations after n assigned patients dependent on $\boldsymbol{T}$
$\mathrm{nro}(n, \boldsymbol{T})$	Number of returns to origin after n assigned patients dependent on $\boldsymbol{T}$
$p_{\boldsymbol{t}}^{\mathrm{RP}}$, $p_{\boldsymbol{t}}$	Appearance probability of $\boldsymbol{t}$ dependent on the randomization procedure RP
$\boldsymbol{T}$	Random vector for a randomization sequence, which takes values in Γ_N
$\boldsymbol{t} \in \Gamma_N$	Realized randomization sequence

Linear model

Notation	Meaning
$\boldsymbol{A}^{-1}$, $\boldsymbol{A}^{-}$	(g-)Inverse of the matrix $\boldsymbol{A}$
$\boldsymbol{A}$	Design matrix
$a(T_n)$	Regression function for T_n
$\boldsymbol{B}_{\mathrm{SB}}$, $\boldsymbol{B}_{\mathrm{CB}}$	Bias vector for modelling selection/chronological bias
$\boldsymbol{\epsilon}$	Random vector for error terms, which takes values in $\mathbb{R}^N$
$\boldsymbol{I}_{n\times n} \in \mathbb{R}^{N\times N}$	Identity matrix with dimension $n \times n$
$r = rk(\boldsymbol{A})$	Rank of the matrix $\boldsymbol{A}$
σ^2	Variance
S_C^2, S_E^2	Sampling variance of the treatment group C/E
$\boldsymbol{\theta} \in \mathbb{R}^p$	Vector of regression coefficients
$\omega(\boldsymbol{A}) = \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$	Projection matrix on the column span of $\boldsymbol{A}$
$\boldsymbol{Y}$	Random vector for the patients' responses, which takes values in $\mathbb{R}^N$
$\boldsymbol{y} \in \mathbb{R}^N$	Vector of the observed patients' responses

Desirability terms

Notation	Meaning
b_l, b_r	Left/right sided weight for the desirability function
$\bar{d}_{\text{geo}}(\text{AV})$	Summarized desirability score of the average-based approach
$d(\text{RS}), \bar{d}(\text{RS})$	(Average) desirability of the randomization sequence-based approach
$d(x)$	Desirability function at the position $x \in \mathbb{R}$
LSL	Lower specification limit
TV	Target value
USL	Upper specification limit

Bibliography

Antognini, A. B., W. F. Rosenberger, W. Y., and Z. M. (2015). Exact optimium coin bias in Efron's randomization procedure. *Statistics in Medicine 34*, 3760–3768.

Atkinson, A. C. (2001). The comparison of designs for sequential clinical trials with covariate information. *Journal of the Royal Statistical Society 165*, 349–373.

Bailey, R. A. and P. Nelson (2003). Hadamard randomization: A valid restriction of random permuted blocks. *Biometrical Journal 45*, 554–560.

Berger, V., A. Ivanova, and D. Knoll (2003). Minimizing predictability while retaining balance through the use of less restrictive randomization procedures. *Statistics in Medicine 22*, 3017–3028.

Berger, V. W. (2005a). Quantifying the magnitude of baseline covariate imbalances resulting from selection bias in randomized clinical trials. *Biometrical Journal 47*, 119–127.

Berger, V. W. (2005b). *Selection bias and covariate imbalances in randomized clinical trials*. Wiley.

Berger, V. W., B. K., and A. R. (2015). Comparing MTI randomization procedures to the blocked randomization. *Statistics in Medicine 35*, 685–694.

Billingsley, P. (1999). *Convergence of Probability Measures*, Volume 2. JOHN WILEY & SONS INC.

Blackwell, D. and J. L. Hodges Jr. (1957). Design for the control of selection bias. *Annals of Mathematical Statistics 25*, 449–460.

Chen, Y.-P. (1999). Biased coin design with imbalance tolerance. *Communications in Statistics. Stochastic Models 15*, 953–975.

Chow, Y. S. and H. Teicher (2008). *Probability theory* (3. Edition). Springer.

Derringer, G. (1994). A balancing act: Optimizing a product's property. *Quality Progress 27*, 51–58.

Derringer, G. and R. Suich (1980). Simultaneous optimization of several response variables. *Journal of Quality Technology 12*, 214–219.

Efron, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika 58*, 403–417.

Fahrmeir, L., T. Kneib, and S. Lang (2007). *Regression.* Springer.

Hajek, J. (1960). Limiting distributions in simple random sampling from a finite population. *Publications of Mathematical Institute of Hungarian Academy of Sciences 5*, 361–374.

Hajek, J. (1981). *Sampling from a finite population.* Marcel Dekker.

Harrington Jr., E. C. (1965). The desirability function. *Industrial Quality Control 21*, 494–498.

Hinkelmann, K. and O. Kempthorne (2008). *Design and analysis of experiments* (2. Edition), Volume 1. Wiley series in probability and statistics.

Hopper, A. N., M. H. Jamison, and W. G. Lewis (2007). Learning curves in surgical practice. *Postgraduate Medical Journal 83*, 777–779.

Hu, F. and W. Rosenberger (2006). *The theory of response-adaptive randomization in clinical trials.* Wiley series in probability and statistics.

ICH E9 (1998). Statistical principles for clinical trials. *Current Step 4 version dated 5 Februrary 1998. Last access in August 2016. Available from:* `http://www.ich.org`.

Joppi, R., V. Bertele, and S. Garattini (2012). Orphan drugs, orphan diseases. The first decade of orphan drug legislation in the EU. *European Journal of Clinical Pharmacology 69*, 1009–1024.

Kennes, L. N. (2013). *The effect of and adjustment for selection bias in randomized concontrol clinical trials.* Doctoral dissertation, RWTH Aachen University.

Kennes, L. N., E. Cramer, R.-D. Hilgers, and N. Heussen (2011). The impact of selection bias on test decision in randomized clinical trials. *Statistics in Medicine 30*, 2573–2581.

Kim, K.-J. and D. K. J. Lin (2000). Simultaneous optimization of mechanical properties of steel by maximizing exponential desirability functions. *Applied Statistics 49*, 311–326.

Koch, K.-R. (2004). *Parameterschätzung und Hypothesentests in linearen Modellen.* Dümmler. URL: `http://www.igg.uni-bonn.de/tg/fileadmin/publication/media/buch97_format_neu.pdf`.

Kocherlakota, K. and S. Kocherlakota (1991). On the doubly noncentral t distribution. *Communications in Statistics 20*, 23–31.

Lachin, J. M. (1988). Statistical properties of randomization in clinical trials. *Controlled Clinical Trials 9*, 289–311.

Lachin, J. M., J. P. Matts, and L. Wei (1988). Randomization in clinical trials: Conclusions and recommendations. *Controlled Clinical Trials 9*, 365–374.

Langer, S. (2014). *The modified distribution of the t-test statistic under the influence of selection bias based on random allocation rule.* Master thesis, RWTH Aachen University.

Lehmann, E. L. (1999). *Elements of large-sample theory.* Springer.

Lehmann, E. L. and G. Casella (1998). *Theory of Point Estimation* (2. Edition). Springer.

Matts, J. P. and R. B. McHugh (1978). Analysis of accrual randomized clinical trials with balanced groups in strata. *Journal of Chronic Diseases 31*, 735–740.

Proschan, M. (1994). Influence of selection bias on type 1 error rate under random permuted block designs. *Statistica Sinica 4*, 219–231.

R Core Team (2016). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Rosenberger, W. F. and J. M. Lachin (2016). *Randomization in clinical trials - Theory and practice.* (2. Edition). Wiley series in probability and statistics.

Rosenkranz, G. K. (2011). The impact of randomization on the analysis of clinical trials. *Statistics in Medicine 30*, 3475–3487.

Schindler, D., D. Uschner, R.-D. Hilgers, and N. Heussen (2015). *randomizeR: Randomization for clinical trials.* R package version 1.2.

Searle, S. R. (1971). *Linear models.* John Wiley and Sons.

Shao, H. and W. F. Rosenberger (2016). *Properties of the random block design for clinical trials,* in *mODa 11 - Advances in model-oriented design and analysis,* pub. by Kunert J., C. H. Müller, and A. C. Atkinson, pp. 225–233. Springer.

Soares, J. F. and C. Wu (1983). Some restricted randomization rules in sequential designs. *Communications in Statiscs - Theory and Methods 12*, 2017–2034.

Student (1908). The probable error of a mean. *Biometrika 6*, 1–25.

Tamm, M., E. Cramer, L. N. Kennes, and N. Heussen (2012). Influence of selection bias on the test decision - A simulation study. *Methods of Information in Medicine 51*, 138–143.

Tamm, M. and R.-D. Hilgers (2014). Chronological bias in randomized clinical trials under different types of unobserved time trends. *Methods of Information in Medicine 53*, 501–510.

Weiss, N. A. (2005). *A course in probability.* Addison-Wesley.

Wickham, H. (2009). *ggplot2: elegant graphics for data analysis.* Springer.

Zelen, M. (1974). The randomization and stratification of patients to clinical trials. *Journal of Chronic Diseases 27*, 365–375.

Zhao, W., Y. Weng, Q. Wu, and Y. Palesch (2012). Quantitative comparison of randomization designs in sequential clinical trials based on treatment balance and allocation randomness. *Pharmaceutical Statistics 11*, 39–48.