1. Introduction

1.1. Randomized Controlled Trials

In the last decades, *randomized clinical trials* have become the gold standard approach for evaluating the efficacy of new experimental treatments in drug development. In the classical one-to-one design, one drug for a given disease and population is investigated and patients receiving this drug are compared to patients in the control arm, which can be either the current standard of care or placebo.

The course of a clinical trial is usually divided into four phases. The aim of phase I is to determine the possible toxic effects and establish the tolerated dose, which will be used for further examination. Once the safety and tolerability of the new drug has been assessed in phase I, the trial progresses to phase II, where the therapeutic effects of the treatment are analyzed. Here, the objective is to prove that the new intervention provides sufficient efficacy to justify further investigation. If this holds, the new drug is compared to the current standard of care for the given disease in a phase III study. This comparative study aims to assess the efficacy and toxicity of the compared treatments in a setting with larger sample sizes. Results from phase III study pose the basis for regulatory agencies to decide whether the new treatment will be approved for the general population. Phase IV study is conducted after the drug has been approved for the market in order to monitor and document rare side effects and other problems that might occur in the population after a longer period of time.

In order to guarantee the comparability of the two treatment groups in a phase III study, patients are allocated randomly, i.e. *randomized*, to the groups, to ensure that the assigned treatment is independent of the baseline characteristics. These classical randomized controlled clinical trials (RCTs), have been considered the most reliable form of scientific evidence for evaluating drug efficacy, since the randomization reduces spurious causality and bias, which could occur if the physicians or patients were involved in the decision, or could arise from other factors that affect the outcome. If proper randomization is employed, the groups are alike on average, which enables the researchers to make causal inference about the new treatment.

In the traditional setting, the differences in outcomes between these two groups are only assessed in the final analysis at the end of the trial in order to determine whether the new treatment is effective or not. It might, however, be more ethical and economical to monitor the trial data periodically and implement the possibility to stop the trial earlier. Clinical trials that allow for early stopping after an interim analysis are referred to as *group sequential trials*. An interim analysis may be performed at multiple time points. The trial should not be carried on, if, for example, one treatment is clearly

superior, so that new patients can benefit from this therapy more quickly. It can also be observed at the interim analysis that the new drug does not offer any benefit, or even poses safety concerns. In this case, the trial should be stopped early for futility to prevent the administration of this treatment to more patients [3]. The design parameters, however, such as the decision rules used in the interim analysis, have to be specified in advance, during the planning of the clinical trial. Because of this so called *pre-specified adaptivity* of group sequential designs, the data from the interim analysis cannot be used to modify the course of the trial, which reduces the flexibility of the trial design [4].

In recent years, there have been efforts to address these drawbacks by introducing *adaptive trial designs*, where modifications are allowed to be made after interim analyses, in order to optimize the further progress of the trial. Besides early stopping of the treatment arms, these adaptations may also include sample size reassessment or reallocation based on the observed data. Such designs also enable investigating multiple drugs and allow the conduct of a phase II and phase III study within one trial, in a so-called *seamless* manner [3].

1.2. Platform Trials

Platform trials are adaptive multi-arm multi-stage clinical trials that aim at evaluating the efficacy of multiple treatments for a single disease simultaneously [5]. The number of experimental treatment arms is not known in advance and the treatments are also allowed to enter and leave the progressing trial. This setting enables faster evaluation of drugs that are being developed during the ongoing trial, as they can be directly included to the shared platform. Moreover, interim analyses can be conducted before the final analysis and modifications of the trial design can be made based on the observed information.

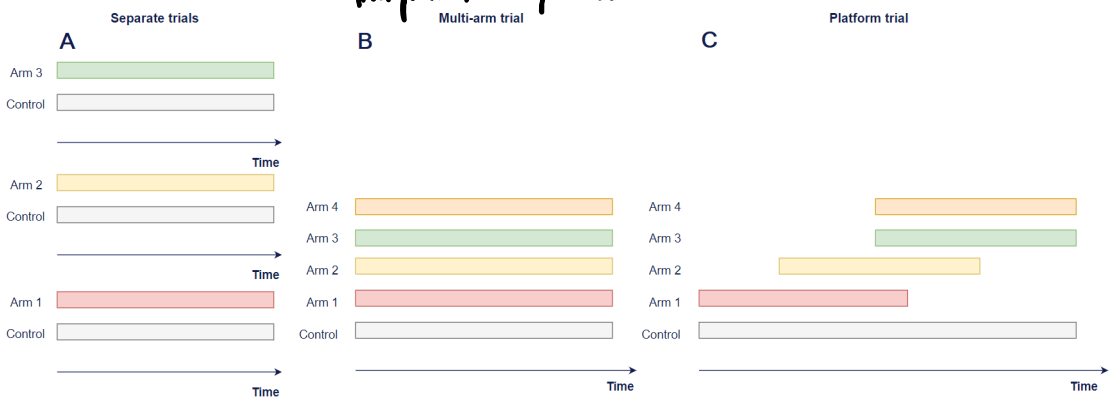

Figure 1.1.: Testing multiple experimental treatments in different clinical trial frameworks. **A)** Separate randomized clinical trials, each with its own control group. **B)** A multi-arm trial, where all interventions enter the trial at the same time. **C)** A platform trial, permitting the arms to enter and leave at different time points.

(A) In the recent decades, adaptive designs have been proposed that allow design changes to be made during the course of the trial based on interim results. Examples are sample size reassessment, subgroup selection, and endpoint selection. Methods have been proposed to enable these adaptations while controlling the T1E and resulting in unbiased estimators.

Testing multiple treatments in the classical trial framework would require setting up separate RCTs for all investigated interventions, each one with its own control group, as illustrated in Figure 1.1-A. In platform trials, the control group is shared across all treatment arms and thus, due to this sharing of resources, less patients are required to be randomized into this arm as compared to separate RCTs. This poses an advantage primarily from the ethical and patient perspective. The control may also change in the course of the trial, if an effective treatment arm becomes the new standard of care. Fig. 1.1-C illustrates a platform trial with 4 experimental treatment arms and a shared control arm.

Due to their higher flexibility, platform trials bring many operational and statistical challenges. In particular, their planing is much more complex as compared to the classical RCTs, and requires extensive computer simulations, since the progress of the trial (i.e. how many experimental arms will be evaluated and the times of their adding and dropping) is not known upfront. Concerns about multiplicity arise from the fact that multiple experimental treatments are investigated within the trial, which may additionally be tested with regard to multiple endpoints or in multiple subgroups [6]. An optimal decision on the adaptation rules, such as the timing of the interim analyses, stopping rules, and randomization is not trivial, as their choice should not inflate the type I error rate. Challenging is also the definition of estimands, the targets of estimation, as platform trials might have multiple objectives and different estimands might be needed for different treatments. Moreover, estimands may need to be modified in the course of the trial, for example, if the control arm changes [7]. Another statistical issue that has recently been controversially discussed is the use of the shared control group in the trial analysis [8].

1.3. Concurrent and Non-concurrent Controls

For treatment arms that enter the ongoing trial later on, the control group is divided into two separate groups; the *concurrent controls* (CC), which includes patients that were randomized to the control arm at the same time as the given treatment arm was active in the trial; and the *non-concurrent controls* (NCC), which denotes patients randomized to the control group before the evaluated treatment arm entered the platform. Fig. 1.2 shows a platform trial with 2 experimental arms, where the 2nd arm joined the trial later. NCC data for this arm are highlighted.

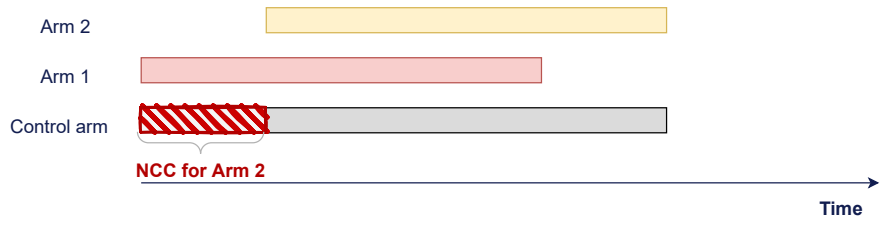

Figure 1.2.: Non-concurrent control data for arm 2.

The question is if and how to incorporate the NCC data to the analysis of the given treatment arm and there is still no consensus regarding this issue. Including NCC data to the analysis may, on one hand, improve the statistical power of the treatment-control comparison, as the underlying sample size increases. However, bias might be introduced to the analysis, if time trends are present in the trial. Such time trends might be caused for instance by changes in the standard of care, patient population or by seasonal effects. Therefore, appropriate methodology that takes into account possible temporal changes is required in order to analyze such trials properly.

Non-concurrent controls share some characteristics with *historical controls*, as they both refer to data collected prior to the data on the treatment under study, and thus both might introduce calendar time bias if used in the analysis [9]. The underlying difference between historical and non-concurrent controls is that non-concurrent controls comprise patients that have been part of the same randomized trial as the investigated treatment (i.e. part of the common trial framework), hence the same inclusion and exclusion criteria applied and the assessment of the endpoint was performed analogously in both groups. However, methodology proposed for historical controls, which is a broader class of methods than has been discussed specifically for NCC in platform trials, can be also applied to non-concurrent controls [10].

There are two naive approaches for utilizing historical or non-concurrent controls - the separate and pooled analysis. In the separate approach, the experimental treatment is only compared to the concurrent controls and the historical information is completely discarded. Pooled analysis naively pools CC and NCC data without further adjustments.

Frequentist methods discussed in the context of historical controls include for instance the "test-then-pool" approach, where the distribution of historical and CC data is first tested for equality with a frequentist test and subsequently either separate or pooled analysis is performed based on the rejection or acceptance of the null hypothesis [11]. Another method where historical data is borrowed dynamically is the so-called dynamic pooling, where a weight parameter is assigned to the historical data, which controls the proportion of the historical information that will be used in the analysis [12]. Moreover, propensity score methods can be used to balance the differences between historical and concurrent controls [13]. The analysis can also be directly adjusted for the baseline covariates using a frequentist regression model.

As far as Bayesian approaches are concerned, power prior, commensurate power prior and meta-analytic-predictive (MAP) prior approaches for incorporating historical data have been proposed in the literature. Power prior methods down-weight the historical information by introducing a power parameter to account for the differences between historical and concurrent control data. A modification to the power prior approach is the commensurate prior approach that uses conditional prior distributions for the concurrent controls, which adjusts the weight parameter through a measure of commensurability. **This was taken directly from your review, I'll adapt it once I understand better what it means ;).** In the meta-analytic-predictive (MAP) prior approaches, the historical information is used to obtain the MAP prior distribution to estimate the response of the concurrent controls.

1.4. Thesis Contribution

This thesis focuses on using non-concurrent controls for individual treatment-control comparisons in platform trials with continuous endpoints in the presence of time trends. In Chapter 2, we first review the current methods presented in the literature so far and discuss their assumptions and limitations. Second, we introduce novel frequentist model-based methods for incorporating NCC data into the analysis. This comprises extending the current frequentist methods to trials with a flexible number of treatment arms, as well as proposing more advanced modelling approaches for adjusting for the time trend, such as mixed models or polynomial splines. In Chapter 3, we evaluate the performance of the proposed methods in a simulation study and assess the conditions under which these approaches lead to valid statistical inference. The considered methods and simulations are implemented in an R package presented in Chapter 4. The package, called `NCC`, enables users to simulate platform trials with an arbitrary number of experimental treatment arms and a common control group and assess the efficacy of individual treatments, while utilizing NCC data. We finish the master thesis by summarizing the work and discussing further research.