Using Estimands for Reporting Clinical Trials

Statistical Methods in Clinical Trials

Master's Thesis

Kristoffer Segerstrøm Mørk

4 June 2018

Aalborg University
Department of Mathematical Sciences

Department of Mathematical Sciences
Aalborg University
http://www.math.aau.dk

AALBORG UNIVERSITY
STUDENT REPORT

Title:
Using Estimands for Reporting Clinical Trials

Theme:
Statistical Methods in Clinical Trials

Project Period:
September 2017 - June 2018

Participants:
Kristoffer Segerstrøm Mørk

Supervisor:
Poul Svante Eriksen

Number of Copies: 1

Page Numbers: 104

Date of Completion:
4 June 2018

Abstract:

The purpose of this thesis is to investigate the use of estimands in clinical trials. In order to do this, different statistical methods are applied. The statistical methods are among other things based on the theory of missing data including multiple imputation, the theory of mixed models for repeated measurements and the theory of logistic regressions. The thesis introduces the different applied statistical theories in order to form the basis for the investigation of the use of estimands in clinical trials. It is worth noticing that this thesis mainly focuses on the efficacy of a new treatment rather than the safety.

The clinical trial NN1998-2076 conducted by Novo Nordisk is used as an example to explore and illustrate the use of estimands in clinical trials. The trial compared the inhaled insulin called AERx against Insulin Aspart. NN1998-2076 was conducted before the regulatory authorities started to focus on missing data and estimands.

In this thesis, four different estimands are formulated for NN1998-2076. The four estimands are based on different strategies for addressing intercurrent events. Based on the results from the different estimands for NN1998-2076, the overall conclusion is that AERx cannot be shown to be non-inferior to Insulin Aspart.

Contents

Preface

Aalborg University, 4 June 2018

Since August 2016, I have been working as a scientific student assistant in one of the biostatistics departments of Novo Nordisk in Aalborg. From September 2017 until June 2018, I have been writing this thesis as a part of my Master's program in Mathematics at the Department of Mathematical Sciences at Aalborg University. The thesis concerns the use of estimands in clinical trials, which is a topic proposed by Claus Dethlefsen, Principal Statistician at Novo Nordisk. The thesis involves different statistical topics like multiple imputation, mixed models for repeated measurements and logistic regressions. The thesis primarily focuses on the frequentist approach.

Novo Nordisk has granted access to some datasets, which have been used as examples to investigate the use of estimands in this thesis. I would like to thank my supervisor Poul Svante Eriksen for his valuable guidance during the project period. Furthermore, I would like to thank Claus Dethlefsen for the topic proposal and for his helpful inputs to this thesis. A special thanks to Novo Nordisk for granting access to the data.

The thesis will refer to sources by [*Author(s)*, *Publication year*]. Eventually the publication year is followed by a page number or a section number. Equations are referred to as (*Number of Equation*). Chapters, sections, definitions etc. are referred to by mentioning the type followed by the number, e.g. Definition 2.1.1.

Kristoffer Segerstrøm Mørk
<kmark13@student.aau.dk>

Danish Abstract

Formålet med dette speciale er at undersøge brugen af estimander i kliniske studier. Til dette anvendes forskellige statistiske metoder som blandt andet baserer sig på teorien vedrørende manglende data herunder to forskellige metoder til at drage likelihood baseret inferens samt multipel imputation. Derudover er metoderne også baseret på teorien vedrørende mixed models herunder mixed models for gentagne målinger samt teorien vedrørende logistiske regressioner. Specialet introducerer de forskellige anvendte statistiske teorier med henblik på at danne grundlaget for undersøgelsen af brugen af estimander i kliniske studier. Det er værd at bemærke, at specialet primært fokuserer på effekten af et nyt lægemiddel fremfor sikkerheden. Effekt og sikkerhed er de to overordnede fokusområder indenfor kliniske studier.

Et studie kaldet NN1998-2076 bruges som et eksempel til at undersøge og illustrere brugen af estimander i kliniske studier. NN1998-2076 er et klinisk studie udført af Novo Nordisk, som startede tilbage i 2006. NN1998-2076 var en del af et klinisk projekt, som havde til formål at undersøge både effekten samt sikkerheden af et inhalerbart insulin kaldet AERx. I det kliniske studie NN1998-2076 blev AERx sammenlignet med Insulin Aspart, hvilket var et lægemiddel, som allerede var på markedet da studiet blev udført. Insulin Aspart skulle modsat AERx indsprøjtes i kroppen. NN1998-2076 var et fase 3a studie, hvor både effekten og sikkerheden af AERx blev sammenlignet med Insulin Aspart. Et af de primære endepunkter til at undersøge effekten var glykosyleret hæmoglobin (%), som noteres HbA1c. Testpersonerne, som blev inddraget i studiet, blev randomiseret 1:1 til enten at modtage AERx eller Insulin Aspart. I alt blev 597 testpersoner randomiseret, hvoraf 586 startede på deres planlagte behandling. Af de 586 testpersoner modtog 292 AERx og 294 Insulin Aspart. Studiet var planlagt til at skulle løbe over 24 måneder, hvor hver af testpersonerne i alt skulle møde op til 20 planlagte besøg på deres behandlingscenter. NN1998-2076 blev stoppet før tid, da Novo Nordisk besluttede sig for at stoppe udviklingen af AERx. En af de primære årsager til dette var, at det syntes at være usandsynligt, at den inhalerbare insulin skulle opnå kliniske fordele i forhold til insulin, som skulle indsprøjtes i kroppen. Det er vigtigt at påpege, at NN1998-2076 blev udført før myndighederne begyndte at fokusere på manglende data og estimander. Derfor var der ikke inddraget nogle estimander i den kliniske rapport for NN1998-2076.

Teorien vedrørende estimander omhandler blandt andet fem forskellige strategier til at håndtere sammenfaldende hændelser, som kan have indflydelse på det målte endepunkt. En sammenfaldende hændelse kan eksempelvis være hovedpine, kvalme eller et slagtilfælde. De fem strategier er fornyligt blevet introduceret og de udgør en vigtig del af teorien vedrørende estimander. Derfor fokuserer dette speciale i særdeleshed på disse stragetier og beskriver nogle af deres fordele og ulemper. Fire af de fem strategier er anvendt i forbindelse med undersøgelsen af brugen af estimander i NN1998-2076. Den sidste strategi er ikke mulig at undersøge i NN1998-2076, da denne kræver et bestemt studie design. I undersøgelsen af brugen af estimander fokuseres der på HbA1c, og der bliver blandt andet testet for om AERx er ikke-dårligere end Insulin Aspart.

Fire forskellige estimander formuleres for NN1998-2076 med henblik på at illustrere samt undersøge brugen af estimander. Baseret på resultaterne fra de forskellige estimander

for NN1998-2076 konkluderes det, at AERx ikke kan påvises at være ikke-dårligere end Insulin Aspart.

1. Introduction

The purpose of this thesis is to investigate the use of estimands in clinical trials. An estimand can be seen as a description of what has to be estimated in order to answer a scientific question of interest in a clinical trial. A formal definition of an estimand is given in Chapter 2. The theory of estimands in this thesis is primarily based on [ICH, 2017].

The thesis mainly focuses on the effect of a treatment also known as *efficacy* in the pharmaceutical industry. Efficacy is one of the two main focus areas within clinical trials. The other main focus area is *safety*, which among other things concerns adverse events like *headache*, *nausea*, *stroke* or even *death*. In this thesis, the actual treatment difference between treatment A and treatment B is the average difference between the effect of treatment A and treatment B measured on the same subject at the same time. But the problem is that it is only possible to measure the effect of one treatment on a subject at a time. One way to address this problem is to allocate different subjects to different treatment conditions randomly, which is known as *randomisation*. Randomisation is a keystone in a clinical trial together with *blinding*. Blinding ensures that each subject's assigned treatment is hidden to some or even all parties in a clinical trial until a specific point in time, cf. [ICH, 2000].

In this thesis, different statistical methods are used in order to estimate the target of an estimand. The statistical methods are based on the theory in Chapter 3, Chapter 4 and Chapter 5. The theory in Chapter 3 deals with missing data and it is one of the main parts in this thesis because handling of missing data plays a central role in the statistical analysis of a clinical trial, cf. [National Research Council, 2010]. In this thesis, the frequentist approach is in scope unless otherwise stated. The following section presents a clinical trial conducted by Novo Nordisk. In Chapter 6, the clinical trial is used as an example to explore and illustrate the practical use of estimands and the related statistical analyses. Hence, this thesis should not be seen as an analysis of the clinical trial.

1.1 Clinical Trial Investigating Inhaled Mealtime Insulin

In this thesis, the use of estimands is investigated in the clinical trial called *NN1998-2076*, which was a clinical trial conducted by Novo Nordisk and started back in 2006. The clinical trial was a part of a project in which the efficacy and safety of the inhaled insulin called AERx were investigated. NN1998-2076 was a phase 3a trial. Phase 3a is the last phase before a regulatory submission of a new drug where both efficacy and safety are investigated in a relatively large number of subjects. The trial consisted of two parallel treatment arms, which are referred to as *the AERx arm* and *the Insulin Aspart arm* in this thesis. Subjects in both treatment arms received Insulin Detemir as a basal insulin. Moreover, the subjects in the AERx arm should inhale AERx just before the three daily main meals, and the subjects in the Insulin Aspart arm should inject Insulin Aspart just before the three daily main meals. Only subjects with type 1 diabetes were included in the trial. Other inclusion and exclusion criteria were used as well but further details are not described in this thesis. The included subjects were randomised 1:1 across the two treatment arms. The trial was planned to last for 24 months with a total of 20 planned visits at the 82 treatment centers. The treatment centers were located in both Canada

and the United States of America. The trial was unblinded, i.e. each subject's assigned treatment was not hidden. An example of a device used to inhale AERx is illustrated in Figure 1.1.

Figure 1.1: AERx insulin device.

The purpose of the trial was to investigate both the efficacy and safety of AERx compared to Insulin Aspart. One of the main efficacy endpoints was glycosylated haemoglobin (%) denoted by HbA1c, which was planned to be measured at visit 1, 7, 9, 11, 12, 13, 15, 17, 19 and 20. In this thesis, HbA1c is the endpoint of interest. Visit 1 is the baseline visit, which is the last visit before the subjects start their treatment period. An overview of the planned timing of the visits, where HbA1c was measured, is given in Table 1.1.

Visit	1	7	9	11	12	13	15	17	19	20
Weeks	-2	12	24	26	44	52	64	80	96	104

Table 1.1: The visits, where HbA1c was measured, together with the planned number of weeks after the beginning of the treatment period.

A total of 597 subjects were randomised. Out of the 597 randomised subjects, 586 were treated. 292 of the treated subjects received AERx and 294 of the treated subjects received Insulin Aspart. In January 2008, Novo Nordisk decided to discontinue the development of AERx. In [Novo Nordisk, 2008], Lars Rebien Sørensen, president and CEO of Novo Nordisk back in 2008, said:

"*The AERx® system has been developed for delivering fast-acting insulin in connection with meals, and we have concluded that fast-acting inhaled insulin in the form it is known today is unlikely to offer significant clinical or convenience benefits over injections of modern insulin with pen devices such as Novo Nordisk's FlexPen®*".

As a consequence of the discontinued development of AERx, NN1998-2076 was terminated prematurely. An overview of how long each subject stayed in the trial is given in Figure 1.2.

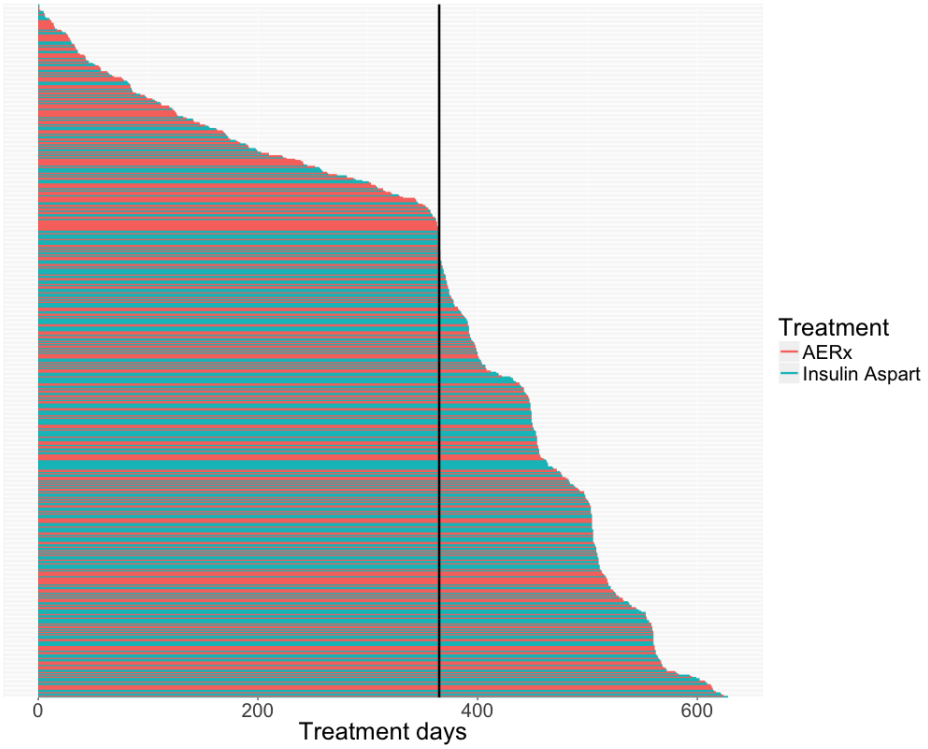

Figure 1.2: Number of treatment days for each subject.

In this thesis, only the treatment period from visit 1 to visit 13 is considered due to the large number of subjects who discontinue the trial after 12 months of treatment. In Figure 1.2 the black vertical line represents 12 months of treatment.

2. Estimands

In this chapter, the theory of estimands is introduced. The chapter is based on [ICH, 2017]. In the following, it is explained how and when an estimand is incorporated in a clinical trial.

The planning phase of a clinical trial starts with a specification of *a trial objective*. Afterwards, the trial objective is translated into a scientific question of interest by formulating *an estimand*, which specifies a target of estimation according to the trial objective. A suitable statistical analysis is then selected in order to estimate the target of the estimand. The statistical analysis is referred to as *the primary analysis* and it defines an estimator called *the primary estimator* which is related to the target of the estimand. The corresponding estimate is referred to as *the primary estimate*. The assumptions, which the primary analysis relies on, have to be clearly stated. Hereafter, different statistical analyses have to be specified in order to investigate how robust the results from the primary analysis are to both deviations from the underlying assumptions and limitations in the data. A statistical analysis of this type is referred to as *a sensitivity analysis*. The corresponding estimator and estimate are referred to as *a sensitivity estimator* and *a sensitivity estimate*, respectively. Remark, it is important that the sensitivity estimators target the same estimand as the primary estimator. The explained flow of the trial planning is illustrated in Figure 2.1, which is based on [ICH, 2017].

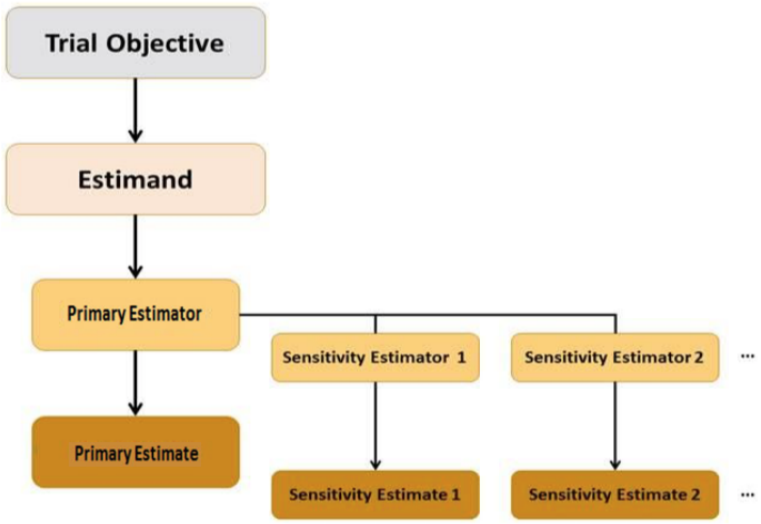

Figure 2.1: Flow of the trial planning.

In a clinical trial there may be multiple trial objectives and therefore multiple estimands. The different estimands may address the same trial objective, but at least one estimand should be defined for each trial objective. The estimand of most interest in a clinical trial is referred to as the primary estimand. The others are either referred to as secondary or supplementary estimands. Remark that the trial objectives, the estimands and the related statistical analyses are prespecified in a clinical trial in order to avoid bias. It is also worth noticing that the choice of trial design depends on the chosen estimands, especially the primary estimand. Some reasons for this are given later in this chapter.

2.1 Definition

In the following, a formal definition of an estimand is introduced. Afterwards, examples are given for each part of the definition.

Definition 2.1.1
An estimand consists of four parts:

A. The population of interest.

B. The endpoint of interest.

C. Addressing of intercurrent events.

D. The population-level summary.

The population of interest is the subjects who are targeted by the scientific question of interest. E.g. the subjects who are eligible to participate in a clinical trial based on some inclusion and exclusion criteria.

The endpoint of interest is the variable which has to be obtained for each subject in the population of interest. An example could be *the change from baseline to week 24 in HbA1c*. In this case, the endpoint of interest is a function of measurements but it could also just consist of measurements itself. Another example may be *the average of the change from baseline in body weight until discontinuation* for which the endpoint of interest incorporate the intercurrent event *discontinuation*. Discontinuation occurs when a subject discontinues the trial prematurely.

Part C. in Definition 2.1.1 specifies how to address different intercurrent events that in some way can be forseen prior to a clinical trial. Usually, the intercurrent events included in an estimand are based on some a priori knowledge. It could e.g. be based on knowledge from a previous trial of the same treatment. An example of part C. may be *use of rescue medication is ignored*, where rescue medication is a drug that the subjects are allowed to receive besides the assigned treatment if necessary. According to [ICH, 2017] there exist different strategies to address intercurrent events. These strategies are presented in the next section.

The last part of Definition 2.1.1 describes how the endpoint of interest is summarized and compared across different treatment conditions. It could e.g. be *the difference in means between the treatment arm and the comparator arm*. In case of a binary endpoint of interest, part D. in Definition 2.1.1 may be *the odds ratio of a successful response between the treatment conditions*.

2.2 Strategies for Addressing Intercurrent Events

The strategies introduced in this section are based on [ICH, 2017] and they specify how to address an intercurrent event. Remark, in a clinical trial there may be several intercurrent events to address when formulating an estimand. It is possible to use a combination of the strategies across different intercurrent events. In the following, only one intercurrent

event is considered in order to keep it simple. But it is worth noticing that all important intercurrent events should be addressed in an estimand. Recall from Chapter 1 that this thesis mainly focuses on efficacy.

2.2.1 Treatment Policy Strategy

In the treatment policy strategy, all occurrences of the intercurrent event are ignored. This means that any observed value should be included in the statistical analysis regardless of the intercurrent event has occurred or not. Therefore, it is important to retain the subjects in the trial even though some of the subjects experience the intercurrent event. For a diabetes trial, an example of an estimand based on the treatment policy strategy could be:

A. All exposed subjects.

B. Change from baseline to week 52 in HbA1c.

C. *Use of rescue medication* is ignored.

D. Difference in means between the treatment conditions.

In the example above, the treatment conditions are compared by estimating the treatment difference. In general, the estimated treatment difference based on the treatment policy strategy includes the effect of the intercurrent event. In the example above, where the treatment policy strategy has been used to address *use of rescue medication*, the estimated treatment difference will be based on both *the effect of the assigned treatment* and eventually *the effect of the rescue medication*. In order to construct a reliable estimator of the actual treatment difference described in Chapter 1, the expected impact of the intercurrent event on the endpoint of interest has to be minor when using the treatment policy strategy. If this is not the case, then this strategy can result in a biased estimator of the actual treatment difference, which can lead to an incorrect conclusion. As an example:

Consider a clinical trial with two parallel treatment arms, one arm with the treatment of interest and the other with a comparator treatment. A greater proportion of the subjects assigned to the treatment of interest end up using rescue medication, while almost none of the subjects in the comparator arm end up using rescue medication. The rescue medication has a major positive impact on the endpoint of interest and hence the results based on the treatment policy strategy reflect that the treatment of interest is better than the comparator treatment. But the truth is that the treatment of interest is worse than the comparator treatment.

In the example above, the estimated treatment difference actually reflects the effect of *the treatment of interest + rescue medication* versus *the comparator treatment*, which easily can lead to an incorrect conclusion if it is not treated with caution. It is worth remarking that the regulatory authorities often prefer this strategy, cf. [ICH, 2017, Section A.3.3.2].

2.2.2 Hypothetical Strategy

The hypothetical strategy examines a scenario in which the intercurrent event would not occur. This implies that all measurements after an occurrence of the intercurrent event should be considered as missing for the given subject. Therefore, it is not necessary to retain the subjects in the trial after they have experienced the intercurrent event. An example of an estimand based on the hypothetical strategy could be:

A. All exposed subjects.

B. Change from baseline to week 52 in HbA1c.

C. If *use of rescue medication* would not occur.

D. Difference in means between the treatment conditions.

In general, the hypothetical strategy gives an estimate of the treatment difference in a scenario where the intercurrent event would not occur. In order to construct a reliable estimator of the actual treatment difference using the hypothetical strategy, it is important that only a minor proportion of the subjects is expected to experience the intercurrent event. If the proportion of subjects, who experience the intercurrent event, is major, then the hypothetical strategy causes a greater proportion of missing data and hence an increased uncertainty about the corresponding results. Notice, the impact of the intercurrent event on the endpoint of interest is not important when using this strategy. The reason is that the hypothetical strategy leaves out the effect of the intercurrent event.

2.2.3 Composite Strategy

The idea of the composite strategy is to incorporate the occurrence of the intercurrent event as a part of the endpoint of interest. There are several ways to incorporate the intercurrent event, but in order to avoid any complex scientific questions, it might be a good idea to specify the endpoint as a binary response variable. For example in a diabetes trial, an estimand based on the composite strategy could be:

A. All exposed subjects.

B. A binary response variable indicating a successful response if the change from baseline to week 52 in HbA1c is below 0, and *use of rescue medication* did not occur.

C. *Use of rescue medication* is captured through the definition of the endpoint of interest.

D. Odds ratio of a successful response between the treatment conditions.

In the example above based on the composite strategy, the treatment conditions are compared by estimating the odds ratio of a successful response. In a clinical trial where both efficacy and safety are of great interest, it seems convenient to use this strategy in the primary estimand since it incorporates the intercurrent event into the endpoint of interest. In trials with a major focus on efficacy, the composite strategy may suit well in a secondary or supplementary estimand to make a summary of the response. Based on [Akacha et al., 2016], a summary of the response could be of interest to both patients, physicians, payers and regulatory authorities.

2.2.4 Principal Stratum Strategy

In the principal stratum strategy, the intercurrent event is incorporated as a part of the population of interest. The idea is to consider *a principal stratum*, which is a subpopulation specified according to the intercurrent event. In a clinical trial with two treatment arms, say A and B, there are four different principal strata with respect to one intercurrent event:

- The subpopulation in which the intercurrent event would not occur regardless of the assigned treatment.
- The subpopulation in which the intercurrent event would occur on treatment A but not on B.
- The subpopulation in which the intercurrent event would occur on treatment B but not on A.
- The subpopulation in which the intercurrent event would occur regardless of the assigned treatment.

Often, the principal stratum of interest is the subpopulation in which the intercurrent event would not occur regardless of the assigned treatment. In practice, it is not possible to identify the subjects in the principal stratum since the subjects can only be assigned to one treatment at a time. Therefore, it is necessary to use a trial design which gives an indication of the subjects in the principal stratum of interest. It could e.g. be a cross-over design or a run-in period. In a cross-over design, each subject receives the different treatments during the treatment period, which makes it possible to determine the subjects in the principal stratum after the treatment period. In a run-in period, the subjects receive the different treatments prior to the treatment period in order to determine the subjects in the principal stratum before they are randomised. It is only necessary to randomise the subjects in the principal stratum when using a run-in period. An example of an estimand under the principal stratum strategy for a diabetes trial could be:

A. All exposed subjects who would not *use rescue medication* in the treatment period regardless of which treatment they are assigned to.

B. Change from baseline to week 52 in HbA1c.

C. *Use of rescue medication* is captured through the definition of the population of interest.

D. Difference in means between the treatment conditions.

The principal stratum strategy estimates the treatment difference among the subjects in the principal stratum of interest. Therefore, this strategy seems to be preferable in clinical trials where the treatment difference of interest is within a subpopulation defined by the intercurrent event. Both the estimated treatment difference based on this strategy and the proportion of subjects, which the principal stratum contains, should be considered before drawing any conclusion. If the principal stratum is the subpopulation in which the intercurrent event would not occur regardless of the assigned treatment, and the stratum consists of a greater proportion of the subjects in the trial, then this strategy may give a reliable estimate of the actual treatment difference in general and not just within the principal stratum. Based on [Akacha et al., 2016], the principal stratum strategy could be of great interest to patients and physicians since it can be used to reflect the expected treatment difference among patients who tolerates the different treatment conditions.

2.2.5 While on Treatment Strategy

The while on treatment strategy estimates the treatment difference prior to the occurrence of the intercurrent event. Therefore, all measurements after the occurrence of the

intercurrent event for a given subject are irrelevant when using this strategy. Like the hypothetical strategy, it is not necessary to retain the subjects in the trial after they have experienced the intercurrent event. An example of an estimand based on the while on treatment strategy for a diabetes trial:

A. All exposed subjects.

B. Average of the change from baseline in HbA1c until *use of rescue medication* or end of treatment period, whichever comes first.

C. *Use of rescue medication* is captured through the definition of the endpoint of interest.

D. Difference in means between the treatment conditions.

In the example above, the endpoint of interest is based on the average of the measurements prior to an eventually occurrence of the intercurrent event. But it could also be based on the last measurement before an eventually occurrence of the intercurrent event. The while on treatment strategy seems to be convenient in situations where the intercurrent event is likely to occur once for each subject. Especially, if the intercurrent event is expected to have a major impact on the subsequent measurements. The intercurrent event could e.g. be discontinuation or even death. An example could be a clinical trial where the treatment of interest is expected to increase the quality of life for subjects with a terminal cancer diagnosis. In such a trial, it is likely that a greater proportion of the subjects die before the treatment period is over. But if a subject dies during such a trial, it does not mean that the treatment did not have any effect. Therefore, it seems more convenient to use the while on treatment strategy in such a trial compared to the other strategies in which the endpoint of interest is measured at a fixed point in time. Based on [Akacha et al., 2016], an estimand using the while on treatment strategy could be of great interest to payers because this strategy can be used to reflect the expected treatment difference as long as the treatment is taken as described in the protocol or label.

3. Missing Data

When analysing datasets that include missing data, it is necessary to make some assumptions about the mechanism that caused the missing data. In this chapter different assumptions about the missing data mechanism are introduced. The chapter is based on [Little and Rubin, 2002]. Let $\mathbf{Y}$ be the stochastic vector of the response variables Y_i, $i = 1, \ldots, n$. A realization of $\mathbf{Y}$ is denoted by $\mathbf{y}$. In the context of a clinical trial, $\mathbf{Y}$ is considered to be a vector of n repeated measurements for a given subject unless otherwise stated. The theory in this chapter can easily be extended to the case with several subjects by assuming that the subjects are independent of each other. Let $\mathbf{M}$ be the stochastic vector of indicators M_i, $i = 1, \ldots, n$, for which a realization is denoted by $\mathbf{m}$, where $m_i = 1$ if y_i is missing or $m_i = 0$ if y_i is observed. The vectors $\mathbf{y}_{obs}$ and $\mathbf{y}_{mis}$ are defined as the observed values and the missing values of $\mathbf{y}$, respectively. The observed and missing values, $\mathbf{y}_{obs}$ and $\mathbf{y}_{mis}$, can be seen as realizations of the stochastic vectors $\mathbf{Y}_{obs}$ and $\mathbf{Y}_{mis}$. Notice, $\mathbf{Y}_{obs}$ and $\mathbf{Y}_{mis}$ are subvectors of $\mathbf{Y}$.

3.1 Types of Missing Data

In this section, different types of missing data are introduced. Moreover, a special pattern of the missing data is defined in the end of this section.

Definition 3.1.1
Missing data are said to be *missing completely at random* (MCAR) if $\mathbf{M}$ and $\mathbf{Y}$ are independent, which is denoted by:

$$\mathbf{M} \perp\!\!\!\perp \mathbf{Y}. \tag{3.1}$$

The interpretation of MCAR is that the reason for the data is missing neither depends on the observed nor unobserved data. An example of MCAR in the setting of clinical trials could be that the endpoint of interest is not measured for all subjects because of economic aspects related only to the trial. In such cases, the missing data are said to be missing by design.

Definition 3.1.2
Missing data are said to be *missing at random* (MAR) if $\mathbf{M}$ and $\mathbf{Y}_{mis}$ are conditional independent given $\mathbf{Y}_{obs}$, which is denoted by:

$$(\mathbf{M} \perp\!\!\!\perp \mathbf{Y}_{mis}) \mid \mathbf{Y}_{obs}. \tag{3.2}$$

The interpretation of MAR is that the reason for the data is missing does not depend on the missing data given the observed data. In a clinical trial, an example of MAR could be that the subjects discontinue the trial when an certain level of the endpoint of interest is observed. It could e.g. be that the subjects are advised to discontinue the trial because their HbA1c is too low. It is worth remarking, that MCAR implies MAR.

Definition 3.1.3
Missing data are said to be *missing not at random* (MNAR) if (3.2) does not hold.

The interpretation of MNAR is that $\mathbf{M}$ depends not only on $\mathbf{Y}_{obs}$ but also $\mathbf{Y}_{mis}$. In the context of a clinical trial, MNAR could be that the subjects with a certain unobserved level of the endpoint of interest are more likely to discontinue the trial. The next definition is related to the pattern of the missing data.

Definition 3.1.4
The pattern of the missing data is said to be *monotone* if $m_i = 1$ implies that $m_j = 1$ for all $j > i$.

In a clinical trial with repeated measurements, monotone missing data could e.g. be caused by discontinuations. In the following, the *full data density* is considered, which is defined as:

$$f(\mathbf{y}, \mathbf{m} \mid X, W, \boldsymbol{\beta}, \boldsymbol{\psi}), \tag{3.3}$$

where X and W are collections of covariates for $\mathbf{Y}$ and $\mathbf{M}$, respectively. Furthermore, $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$ are vectors of unknown parameters associated with $\mathbf{Y}$ and $\mathbf{M}$, respectively. Unless otherwise stated, $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$ are assumed to be *variation independent* in the sense that the parameter space Ω of $(\boldsymbol{\beta}, \boldsymbol{\psi})$ is a product space $\Omega_{\boldsymbol{\beta}} \times \Omega_{\boldsymbol{\psi}}$ where $\boldsymbol{\beta} \in \Omega_{\boldsymbol{\beta}}$ and $\boldsymbol{\psi} \in \Omega_{\boldsymbol{\psi}}$. Recall from Chapter 1, the frequentist approach is in scope in this thesis unless otherwise stated. Therefore, X and W are both assumed to be fixed. The two next sections describe two different factorizations of the full data density in (3.3). The two factorizations specify two different frameworks to handle missing data.

3.2 Selection Model

This section presents a factorization of the full data density in (3.3). The factorization is used to introduce a way to draw inference for $\boldsymbol{\beta}$ based on the likelihood of ($\boldsymbol{\beta}$, $\boldsymbol{\psi}$) given the observed data $(\mathbf{y}_{obs}, \mathbf{m})$, which is given by:

$$L(\boldsymbol{\beta}, \boldsymbol{\psi} \mid \mathbf{y}_{obs}, \mathbf{m}, X, W) \propto f(\mathbf{y}_{obs}, \mathbf{m} \mid X, W, \boldsymbol{\beta}, \boldsymbol{\psi}), \tag{3.4}$$

where $f(\mathbf{y}_{obs}, \mathbf{m} \mid X, W, \boldsymbol{\beta}, \boldsymbol{\psi})$ is the density of the observed data. Remark, the density in (3.4) can be obtained by marginalizing $\mathbf{y}_{mis}$ out of the full data density in (3.3). In the following, the factorization is defined.

Definition 3.2.1
The full data density in (3.3) can be factorized into:

$$f(\mathbf{y}, \mathbf{m} \mid X, W, \boldsymbol{\beta}, \boldsymbol{\psi}) = f(\mathbf{y} \mid X, \boldsymbol{\beta}) f(\mathbf{m} \mid \mathbf{y}, W, \boldsymbol{\psi}), \tag{3.5}$$

which is called *the selection model*.

The name *selection model* is based on *the missing data mechanism* $f(\mathbf{m} \mid \mathbf{y}, W, \boldsymbol{\psi})$, which also can be seen as a selection of the observed values, i.e. if $m_i = 0$ then y_i is observed. The selection model gives a framework where $\mathbf{Y}$ is handled regardless of $\mathbf{M}$ since the first factor in (3.5) is the conditional density of $\mathbf{Y}$ given X and $\boldsymbol{\beta}$. The selection model in (3.5) can be factorized even more by considering the observed and missing part of $\mathbf{y}$:

$$f(\mathbf{y}, \mathbf{m} \mid X, W, \boldsymbol{\beta}, \boldsymbol{\psi}) = f(\mathbf{y}_{obs} \mid X, \boldsymbol{\beta}) f(\mathbf{y}_{mis} \mid \mathbf{y}_{obs}, X, \boldsymbol{\beta}) f(\mathbf{m} \mid \mathbf{y}, W, \boldsymbol{\psi}). \tag{3.6}$$

Based on (3.6), the likelihood in (3.4) can be written as:

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\psi} \mid \mathbf{y}_{obs}, \mathbf{m}, X, W) &\propto f(\mathbf{y}_{obs}, \mathbf{m} \mid X, W, \boldsymbol{\beta}, \boldsymbol{\psi}) \\ &= \int f(\mathbf{y}, \mathbf{m} \mid X, W, \boldsymbol{\beta}, \boldsymbol{\psi}) \, d\mathbf{y}_{mis} \\ &= \int f(\mathbf{y}_{obs} \mid X, \boldsymbol{\beta}) f(\mathbf{y}_{mis} \mid \mathbf{y}_{obs}, X, \boldsymbol{\beta}) f(\mathbf{m} \mid \mathbf{y}, W, \boldsymbol{\psi}) \, d\mathbf{y}_{mis} \\ &= f(\mathbf{y}_{obs} \mid X, \boldsymbol{\beta}) \int f(\mathbf{y}_{mis} \mid \mathbf{y}_{obs}, X, \boldsymbol{\beta}) f(\mathbf{m} \mid \mathbf{y}, W, \boldsymbol{\psi}) \, d\mathbf{y}_{mis}. \end{aligned} \tag{3.7}$$

Let $L_{sel}(\boldsymbol{\beta}, \boldsymbol{\psi} \mid \mathbf{y}_{obs}, \mathbf{m}, X, W)$ denote the likelihood of $(\boldsymbol{\beta}, \boldsymbol{\psi})$ in (3.7). If the missing data are MAR, then the density of the observed data can be written as:

$$\begin{aligned} f(\mathbf{y}_{obs}, \mathbf{m} \mid X, W, \boldsymbol{\beta}, \boldsymbol{\psi}) &= f(\mathbf{y}_{obs} \mid X, \boldsymbol{\beta}) \int f(\mathbf{y}_{mis} \mid \mathbf{y}_{obs}, X, \boldsymbol{\beta}) f(\mathbf{m} \mid \mathbf{y}_{obs}, W, \boldsymbol{\psi}) \, d\mathbf{y}_{mis} \\ &= f(\mathbf{y}_{obs} \mid X, \boldsymbol{\beta}) f(\mathbf{m} \mid \mathbf{y}_{obs}, W, \boldsymbol{\psi}) \int f(\mathbf{y}_{mis} \mid \mathbf{y}_{obs}, X, \boldsymbol{\beta}) \, d\mathbf{y}_{mis} \\ &= f(\mathbf{y}_{obs} \mid X, \boldsymbol{\beta}) f(\mathbf{m} \mid \mathbf{y}_{obs}, W, \boldsymbol{\psi}), \end{aligned} \tag{3.8}$$

which is based on Definition 3.1.2 and (3.6). Let $L_{ign}(\boldsymbol{\beta} \mid \mathbf{y}_{obs}, X) \propto f(\mathbf{y}_{obs} \mid X, \boldsymbol{\beta})$ be the likelihood ignoring the missing data mechanism. If $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$ are variation independent and the missing data are MAR, then:

$$L_{sel}(\boldsymbol{\beta}, \boldsymbol{\psi} \mid \mathbf{y}_{obs}, \mathbf{m}, X, W) \propto L_{ign}(\boldsymbol{\beta} \mid \mathbf{y}_{obs}, X), \tag{3.9}$$

when $\boldsymbol{\psi}$ is fixed. Hence, the inference for $\boldsymbol{\beta}$ drawn from $L_{sel}(\boldsymbol{\beta}, \boldsymbol{\psi} \mid \mathbf{y}_{obs}, \mathbf{m}, X, W)$ will be equivalent to the inference drawn from $L_{ign}(\boldsymbol{\beta} \mid \mathbf{y}_{obs}, X)$ when $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$ are variation independent and the missing data are MAR. The proportionality in (3.9) follows from (3.7) and (3.8). This leads to the following definition.

Definition 3.2.2
The missing data mechanism $f(\mathbf{m} \mid \mathbf{y}, W, \boldsymbol{\psi})$ is ignorable for likelihood-based inference if the two following statements hold:

- The missing data are MAR.
- The parameters $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$ are variation independent.

Remark, the statements in Definition 3.2.2 have to be considered carefully in practice since they cannot be tested.

To conclude, the inference for $\boldsymbol{\beta}$ can be drawn from $L_{ign}(\boldsymbol{\beta} \mid \mathbf{y}_{obs}, X)$ if the two statements in Defintion 3.2.2 are assumed to hold. However, if it does not seem plausible

to assume the two statements in Definition 3.2.2, the inference for $\boldsymbol{\beta}$ should be drawn from $L_{sel}(\boldsymbol{\beta}, \boldsymbol{\psi} \mid \mathbf{y}_{obs}, \mathbf{m}, X, W)$ instead of $L_{ign}(\boldsymbol{\beta} \mid \mathbf{y}_{obs}, X)$. Remark, this requires a specification of the missing data mechanism $f(\mathbf{m} \mid \mathbf{y}, W, \boldsymbol{\psi})$ and an evaluation of the integral in (3.7) which in some situations cannot be done analytically. Therefore, the selection model approach does not seem appropriate in cases where the statements in Definition 3.2.2 are not plausible to hold. In the next section, an alternative to the selection model is introduced.

3.3 Pattern-Mixture Model

This section presents another factorization of the full data density in (3.3). The factorization leads to another way to draw inference for $\boldsymbol{\beta}$ based on the likelihood in (3.4) compared to the one in Section 3.2.

Definition 3.3.1
The full data density in (3.3) can be factorized into:

$$f(\mathbf{y}, \mathbf{m} \mid X, W, \boldsymbol{\beta}, \boldsymbol{\psi}) = f(\mathbf{y} \mid \mathbf{m}, X, \boldsymbol{\beta}) f(\mathbf{m} \mid W, \boldsymbol{\psi}), \tag{3.10}$$

which is called *the pattern-mixture model*.

This factorization is called the pattern-mixture model because $f(\mathbf{y} \mid \mathbf{m}, X, \boldsymbol{\beta})$ allows different models for the observed and missing values, i.e. a mixture of different models. Therefore, the pattern-mixture model seems to be more convenient compared to the selection model in cases where it is reasonable to think that the missing data follow another distribution compared to the observed data. Moreover, the pattern-mixture model allows different types of missing data to follow different distributions. An example could be that the missing data are either caused by an certain type of adverse event or lack of efficacy. In this case, it would be reasonable to think that the two different types of missing data do not follow the same distribution.

The pattern-mixture model can, like the selection model, be factorized even more by considering the observed and missing part of $\mathbf{y}$:

$$f(\mathbf{y}, \mathbf{m} \mid X, W, \boldsymbol{\beta}, \boldsymbol{\psi}) = f(\mathbf{y}_{obs} \mid \mathbf{m}, X, \boldsymbol{\beta}) f(\mathbf{y}_{mis} \mid \mathbf{y}_{obs}, \mathbf{m}, X, \boldsymbol{\beta}) f(\mathbf{m} \mid W, \boldsymbol{\psi}). \tag{3.11}$$

Another representation of the likelihood in (3.4) can be obtained by using the pattern-mixture model approach:

$$\begin{aligned}
L(\boldsymbol{\beta}, \boldsymbol{\psi} \mid \mathbf{y}_{obs}, \mathbf{m}, X, W) &\propto f(\mathbf{y}_{obs}, \mathbf{m} \mid X, W, \boldsymbol{\beta}, \boldsymbol{\psi}) \\
&= \int f(\mathbf{y}, \mathbf{m} \mid X, W, \boldsymbol{\beta}, \boldsymbol{\psi}) \, d\mathbf{y}_{mis} \\
&= \int f(\mathbf{y}_{obs} \mid \mathbf{m}, X, \boldsymbol{\beta}) f(\mathbf{y}_{mis} \mid \mathbf{y}_{obs}, \mathbf{m}, X, \boldsymbol{\beta}) f(\mathbf{m} \mid W, \boldsymbol{\psi}) \, d\mathbf{y}_{mis} \\
&= f(\mathbf{y}_{obs} \mid \mathbf{m}, X, \boldsymbol{\beta}) f(\mathbf{m} \mid W, \boldsymbol{\psi}) \int f(\mathbf{y}_{mis} \mid \mathbf{y}_{obs}, \mathbf{m}, X, \boldsymbol{\beta}) \, d\mathbf{y}_{mis} \\
&= f(\mathbf{y}_{obs} \mid \mathbf{m}, X, \boldsymbol{\beta}) f(\mathbf{m} \mid W, \boldsymbol{\psi}),
\end{aligned} \tag{3.12}$$

where (3.11) has been applied. Let $L_{pm}(\boldsymbol{\beta}, \boldsymbol{\psi} \mid \mathbf{y}_{obs}, \mathbf{m}, X, W)$ denote the likelihood in (3.12). Remark, the inference for $\boldsymbol{\beta}$ drawn from $L_{pm}(\boldsymbol{\beta}, \boldsymbol{\psi} \mid \mathbf{y}_{obs}, \mathbf{m}, X, W)$ is equivalent

to the inference drawn from $L_{ign}(\boldsymbol{\beta} \mid \mathbf{y}_{obs}, X)$ if $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$ are variation independent and the missing data are MCAR. However, MCAR is often a too strict assumption in practice. Thus, the argument to draw inference from $L_{ign}(\boldsymbol{\beta} \mid \mathbf{y}_{obs}, X)$ lies in the selection model approach together with the MAR assumption. It is also worth noticing that the likelihood based on the pattern-mixture model in (3.12) does not involve an integral like the one based on the selection model in (3.7). Hence, the pattern-mixture model approach seems to be more convenient than the selection model approach in situations where an assumption about MAR is not plausible.

3.4 Multiple Imputation

In this section, the theory of multiple imputation is introduced. The section is based on [Carpenter and Kenward, 2013]. Multiple imputation (MI) is an alternative to both the selection model approach and the pattern-mixture model approach to draw inference for $\boldsymbol{\beta}$. The MI procedure is defined as follows:

1. The missing values $\mathbf{y}_{mis}$ are imputed m times in order to get m complete datasets.
2. Each of the m complete datasets is analysed with the same model to get m estimates of both $\boldsymbol{\beta}$ and the associated covariance matrix Σ.
3. The results from the m analyses are combined.

The MI estimate of $\boldsymbol{\beta}$ is:

$$\widehat{\boldsymbol{\beta}}_{MI} = \frac{1}{m}\sum_{i=1}^{m}\widehat{\boldsymbol{\beta}}_i, \tag{3.13}$$

where $\widehat{\boldsymbol{\beta}}_i$ is the estimate of $\boldsymbol{\beta}$ from the analysis of the i'th complete dataset. The average within-imputation covariance matrix is given by:

$$W = \frac{1}{m}\sum_{i=1}^{m}\widehat{\Sigma}_i, \tag{3.14}$$

where $\widehat{\Sigma}_i$ is the estimate of Σ from the analysis of the i'th complete dataset. The between-imputation variability matrix is given by:

$$B = \frac{1}{m-1}\sum_{i=1}^{m}(\widehat{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}}_{MI})(\widehat{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}}_{MI})^{\top}. \tag{3.15}$$

An estimate of the covariance matrix of $\widehat{\boldsymbol{\beta}}_{MI}$ is:

$$\widehat{\Sigma}_{MI} = W + \frac{m+1}{m}B. \tag{3.16}$$

Notice, the estimates in (3.13) and (3.16) are refered to as *Rubin's rules*. Step 3. in the MI procedure is done by using Rubin's rules.

3.4.1 Justification of Multiple Imputation

The justification of multiple imputation is based on a Bayesian approach. For simplicity, X and W are ignored in the notation in this section. Consider the full data posterior distribution of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\psi})$ given $(\mathbf{y}, \mathbf{m})$:

$$f(\boldsymbol{\theta} \mid \mathbf{y}, \mathbf{m}) = \frac{f(\boldsymbol{\theta})f(\mathbf{y}, \mathbf{m} \mid \boldsymbol{\theta})}{f(\mathbf{y}, \mathbf{m})} \propto f(\boldsymbol{\theta})L(\boldsymbol{\theta} \mid \mathbf{y}, \mathbf{m}), \tag{3.17}$$

where $f(\boldsymbol{\theta})$ is the prior distribution of $\boldsymbol{\theta}$. The posterior distribution of $\boldsymbol{\theta}$ given $(\mathbf{y}_{obs}, \mathbf{m})$ can be related to (3.17) in the following way:

$$\begin{aligned} f(\boldsymbol{\theta} \mid \mathbf{y}_{obs}, \mathbf{m}) &= \int f(\boldsymbol{\theta}, \mathbf{y}_{mis} \mid \mathbf{y}_{obs}, \mathbf{m})\, d\mathbf{y}_{mis} \\ &= \int f(\boldsymbol{\theta} \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}, \mathbf{m}) f(\mathbf{y}_{mis} \mid \mathbf{y}_{obs}, \mathbf{m})\, d\mathbf{y}_{mis}. \end{aligned} \tag{3.18}$$

The posterior mean of $\boldsymbol{\theta}$ given $(\mathbf{y}_{obs}, \mathbf{m})$ can be expressed as:

$$\mathrm{E}[\boldsymbol{\theta} \mid \mathbf{y}_{obs}, \mathbf{m}] = \mathrm{E}\left[\mathrm{E}[\boldsymbol{\theta} \mid \mathbf{y}_{obs}, \mathbf{m}, \mathbf{Y}_{mis}] \mid \mathbf{y}_{obs}, \mathbf{m}\right], \tag{3.19}$$

by using the law of total expectation, $\mathrm{E}[X] = \mathrm{E}[\mathrm{E}[X \mid Y]]$. Similarly, the posterior variance of $\boldsymbol{\theta}$ given $(\mathbf{y}_{obs}, \mathbf{m})$ can be expressed as:

$$\begin{aligned} \mathrm{Var}[\boldsymbol{\theta} \mid \mathbf{y}_{obs}, \mathbf{m}] &= \mathrm{E}[\mathrm{Var}[\boldsymbol{\theta} \mid \mathbf{y}_{obs}, \mathbf{m}, \mathbf{Y}_{mis}] \mid \mathbf{y}_{obs}, \mathbf{m}] \\ &\quad + \mathrm{Var}[\mathrm{E}[\boldsymbol{\theta} \mid \mathbf{y}_{obs}, \mathbf{m}, \mathbf{Y}_{mis}] \mid \mathbf{y}_{obs}, \mathbf{m}], \end{aligned} \tag{3.20}$$

by applying the law of total variance, $\mathrm{Var}[X] = \mathrm{E}[\mathrm{Var}[X \mid Y]] + \mathrm{Var}[\mathrm{E}[X \mid Y]]$. Let $\mathbf{y}_{mis}^{(i)}$ be the i'th imputation of $\mathbf{y}_{mis}$ drawn from the posterior predictive distribution $f(\mathbf{y}_{mis} \mid \mathbf{y}_{obs}, \mathbf{m})$ for $i = 1, \ldots, m$. Then an estimate of the posterior mean in (3.19) can be based on an empirical approximation:

$$\widehat{\boldsymbol{\theta}}_{MI} = \frac{1}{m} \sum_{i=1}^{m} \widehat{\boldsymbol{\theta}}_i,$$

where $\widehat{\boldsymbol{\theta}}_i = \mathrm{E}[\boldsymbol{\theta} \mid \mathbf{y}_{obs}, \mathbf{m}, \mathbf{y}_{mis}^{(i)}]$ is an estimate of the posterior mean of $\boldsymbol{\theta}$ based on $\mathbf{y}_{mis}^{(i)}$ and the full data posterior distribution in (3.17). An estimate of the posterior variance in (3.20) can, like the posterior mean, be based on an empirical approximation:

$$\widehat{\Sigma}_{MI} = \frac{1}{m} \sum_{i=1}^{m} \widehat{\Sigma}_i + \frac{1}{m-1} \sum_{i=1}^{m} \left(\widehat{\boldsymbol{\theta}}_i - \widehat{\boldsymbol{\theta}}_{MI}\right) \left(\widehat{\boldsymbol{\theta}}_i - \widehat{\boldsymbol{\theta}}_{MI}\right)^{\top}, \tag{3.21}$$

where $\widehat{\Sigma}_i = \mathrm{Var}[\boldsymbol{\theta} \mid \mathbf{y}_{obs}, \mathbf{m}, \mathbf{y}_{mis}^{(i)}]$ is an estimate of the posterior variance of $\boldsymbol{\theta}$ based on $\mathbf{y}_{mis}^{(i)}$ and the full data posterior distribution in (3.17). In [Carpenter and Kenward, 2013], the last term in (3.21) is multiplied by $(m+1)/m$ to account for the increased uncertainty when the number of imputations is small. This leads to the estimate in (3.16). Remark, the posterior distribution in (3.18) can be approximated by:

$$\frac{1}{m} \sum_{i=1}^{m} f(\boldsymbol{\theta} \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}^{(i)}, \mathbf{m}), \tag{3.22}$$

since the posterior distribution in (3.18) can be written as $\mathrm{E}[f(\boldsymbol{\theta} \mid \mathbf{y}_{obs}, \mathbf{Y}_{mis}, \mathbf{m}) \mid \mathbf{y}_{obs}, \mathbf{m}]$. It is furthermore worth remarking that this justification of MI can easily be extended to a frequentist approach by considering the prior distribution of $\boldsymbol{\theta}$ to be non-informative. If the prior distribution of $\boldsymbol{\theta}$ is non-informative, then the posterior distribution in (3.17) is proportional to the likelihood $L(\boldsymbol{\theta} \mid \mathbf{y}, \mathbf{m})$.

Recall that $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\psi})$. Often, it is only the inference for $\boldsymbol{\beta}$ which is of interest in practice. Since $\boldsymbol{\beta}$ is a subvector of $\boldsymbol{\theta}$, the MI procedure for $\boldsymbol{\beta}$ is also justified. If $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$ are *a priori independent*, i.e. $f(\boldsymbol{\beta}, \boldsymbol{\psi}) = f(\boldsymbol{\beta}) f(\boldsymbol{\psi})$, and the missing data are MAR, then

the MI procedure for $\boldsymbol{\beta}$ can be based on $f(\boldsymbol{\beta} \mid \mathbf{y}_{obs})$ instead of $f(\boldsymbol{\beta} \mid \mathbf{y}_{obs}, \mathbf{m})$. This follows from:

$$\begin{aligned}
f(\boldsymbol{\beta} \mid \mathbf{y}_{obs}, \mathbf{m}) &= \int f(\boldsymbol{\beta}, \boldsymbol{\psi} \mid \mathbf{y}_{obs}, \mathbf{m})\, d\boldsymbol{\psi} \\
&\propto \int f(\boldsymbol{\beta}, \boldsymbol{\psi}) f(\mathbf{y}_{obs}, \mathbf{m} \mid \boldsymbol{\beta}, \boldsymbol{\psi})\, d\boldsymbol{\psi} \\
&= \int f(\boldsymbol{\beta}, \boldsymbol{\psi}) \int f(\mathbf{y}_{obs}, \mathbf{y}_{mis}, \mathbf{m} \mid \boldsymbol{\beta}, \boldsymbol{\psi})\, d\mathbf{y}_{mis}\, d\boldsymbol{\psi} \\
&= \int f(\boldsymbol{\beta}) f(\boldsymbol{\psi}) \int f(\mathbf{y}_{obs}, \mathbf{y}_{mis} \mid \boldsymbol{\beta}) f(\mathbf{m} \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}, \boldsymbol{\psi})\, d\mathbf{y}_{mis}\, d\boldsymbol{\psi} \\
&= f(\boldsymbol{\beta}) \int f(\boldsymbol{\psi}) \int f(\mathbf{y}_{obs}, \mathbf{y}_{mis} \mid \boldsymbol{\beta}) f(\mathbf{m} \mid \mathbf{y}_{obs}, \boldsymbol{\psi})\, d\mathbf{y}_{mis}\, d\boldsymbol{\psi} \\
&= f(\boldsymbol{\beta}) \int f(\boldsymbol{\psi}) f(\mathbf{y}_{obs} \mid \boldsymbol{\beta}) f(\mathbf{m} \mid \mathbf{y}_{obs}, \boldsymbol{\psi})\, d\boldsymbol{\psi} \\
&= f(\boldsymbol{\beta}) f(\mathbf{y}_{obs} \mid \boldsymbol{\beta}) \int f(\boldsymbol{\psi}) f(\mathbf{m} \mid \mathbf{y}_{obs}, \boldsymbol{\psi})\, d\boldsymbol{\psi} \\
&\propto f(\boldsymbol{\beta} \mid \mathbf{y}_{obs}) \int f(\boldsymbol{\psi}) f(\mathbf{m} \mid \mathbf{y}_{obs}, \boldsymbol{\psi})\, d\boldsymbol{\psi}, \qquad (3.23)
\end{aligned}$$

where Definition 3.1.2, Definition 3.2.1 and the a priori independence of $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$ have been applied. The integral in the last equation in (3.23) does not depend on $\boldsymbol{\beta}$. Hence, the inference for $\boldsymbol{\beta}$ can be based on $f(\boldsymbol{\beta} \mid \mathbf{y}_{obs})$ instead of $f(\boldsymbol{\beta} \mid \mathbf{y}_{obs}, \mathbf{m})$ if $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$ are variation independent and the missing data are MAR. Remark, if $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$ are a priori independent, then they are also variation independent.

In practice, it can be difficult to draw imputations from the posterior predictive distribution $f(\mathbf{y}_{mis} \mid \mathbf{y}_{obs}, \mathbf{m})$ since it involves a marginalization over $\boldsymbol{\beta}$:

$$f(\mathbf{y}_{mis} \mid \mathbf{y}_{obs}, \mathbf{m}) = \int f(\mathbf{y}_{mis} \mid \mathbf{y}_{obs}, \mathbf{m}, \boldsymbol{\beta}) f(\boldsymbol{\beta} \mid \mathbf{y}_{obs}, \mathbf{m})\, d\boldsymbol{\beta}. \qquad (3.24)$$

An alternative is to approximate draws from $f(\mathbf{y}_{mis} \mid \mathbf{y}_{obs}, \mathbf{m})$. Based on (3.24), this can be done by drawing $\widetilde{\boldsymbol{\beta}}$ from $f(\boldsymbol{\beta} \mid \mathbf{y}_{obs}, \mathbf{m})$ and afterwards draw $\mathbf{y}_{mis}^{(i)}$ from $f\left(\mathbf{y}_{mis} \mid \mathbf{y}_{obs}, \mathbf{m}, \widetilde{\boldsymbol{\beta}}\right)$. In the next subsection, an algorithm based on (3.24) is introduced.

3.4.2 MI for Monotone MAR Mechanisms

This subsection specifies an algorithm to do multiple imputation of missing data that are both monotone missing and MAR. The algorithm is based on [Rubin, 2004].

Consider $\mathbf{Y}_j$ to be a stochastic vector consisting of n normal distributed variables $Y_{i,j}$ where $i = 1, \ldots, n$ represents the subject and $j = 1, \ldots, l$ represents the measurement. Let X_j be a $n \times p_j$ matrix consisting of the p_j covariates for $\mathbf{y}_j$. Eventually, X_j can consist of some of the previous measurements $\mathbf{y}_{j-1}, \ldots, \mathbf{y}_1$. Moreover, let $\boldsymbol{\beta}_j$ be the vector of the p_j parameters related to a linear regression of $\mathbf{y}_j$ on X_j. The observed and missing values of $\mathbf{y}_j$ are denoted by $\mathbf{y}_j^{obs}$ and $\mathbf{y}_j^{mis}$, respectively. Let n_j be the number of observed values in $\mathbf{y}_j$. The submatrices of X_j related to $\mathbf{y}_j^{obs}$ and $\mathbf{y}_j^{mis}$ are referred to as X_j^{obs} and X_j^{mis}, respectively. Remark in this context, a monotone pattern of the missing data means that if the k'th measurement $y_{i,k}$ for subject i is missing, then $y_{i,j}$ for $j > k$ is also missing.

Assume $\mathrm{Cov}[\mathbf{Y}_j, \mathbf{Y}_k]$ is a diagonal matrix and $\mathrm{Var}[\mathbf{Y}_j] = \sigma_j^2 I_n$ for all $j = 1, \ldots, l$ and $k \neq j$. This means that the subjects are assumed to be independent. Furthermore, assume

$n_1 = n$, $n_j > p_j$ and all the covariates besides the previous measurements in X_j are fully observed for $j = 1, \ldots, l$. This implies that X_j^{obs} are fully observed for $j = 1, \ldots, l$ when the pattern of the missing data is monotone. Notice, X_j^{mis} for $j = 3, \ldots, l$ can have missing values if some of the previous measurements in $\mathbf{y}_{j-1}, \ldots, \mathbf{y}_2$ are missing.

If the missing data are MAR and the missing data pattern is monotone, then the imputation step in the MI procedure can be done by using the following algorithm m times to create m complete datasets.

Algorithm 3.4.1
Iterate the following 4 steps for $j = 2, \ldots, l$:

1. Fit a linear regression of $\mathbf{y}_j^{obs}$ on X_j^{obs} to obtain the ordinary least squares estimates of $\boldsymbol{\beta}_j$ and σ_j^2 denoted by $\widehat{\boldsymbol{\beta}}_j$ and $\widehat{\sigma}_j^2$, respectively.
2. Draw z from $\chi^2_{n_j - p_j}$ to obtain the estimate:

$$\widetilde{\sigma}_j^2 = \frac{\widehat{\sigma}_j^2 (n_j - p_j)}{z}, \tag{3.25}$$

   where p_j is the number of parameters in $\boldsymbol{\beta}_j$.
3. Draw $\widetilde{\boldsymbol{\beta}}_j$ from $N_{p_j}\left(\widehat{\boldsymbol{\beta}}_j, \widetilde{\sigma}_j^2 V_j\right)$, where $V_j = (X_j^{obs\top} X_j^{obs})^{-1}$.
4. Impute the missing values $\mathbf{y}_j^{mis}$ by:

$$\mathbf{y}_j^{mis} = X_j^{mis} \widetilde{\boldsymbol{\beta}}_j + \boldsymbol{\varepsilon}_j, \tag{3.26}$$

   where $\boldsymbol{\varepsilon}_j$ is drawn from $N_{n-n_j}(\mathbf{0}, \widetilde{\sigma}_j^2 I_{n-n_j})$. □

Remark, if X_j^{mis} consists of some previous measurements that are missing, then they are replaced in step 4. of Algorithm 3.4.1 with the imputed values from the previous iterations. The justification of Algorithm 3.4.1 is based on (3.23), (3.24) and the fact that:

$$f(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_l) = f(\mathbf{y}_l \mid \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{l-1}) \cdots f(\mathbf{y}_2 \mid \mathbf{y}_1) f(\mathbf{y}_1). \tag{3.27}$$

Notice, the second and third step in Algorithm 3.4.1 draws $\widetilde{\sigma}_j^2$ and $\widetilde{\boldsymbol{\beta}}_j$ from the posterior distributions, $f\left(\sigma_j^2 \,\middle|\, \mathbf{y}_j^{obs}, X_j^{obs}\right)$ and $f\left(\boldsymbol{\beta}_j \,\middle|\, \mathbf{y}_j^{obs}, X_j^{obs}, \widetilde{\sigma}_j^2\right)$, with non-informative priors. The last step draws $\mathbf{y}_j^{mis}$ from $f\left(\mathbf{y}_j^{mis} \,\middle|\, \mathbf{y}_j^{obs}, X_j^{mis}, \widetilde{\boldsymbol{\beta}}_j, \widetilde{\sigma}_j^2\right)$. If the missing data pattern is close to be monotone, i.e. only a few measurements cause the pattern to be non-monotone, then it is still possible to use Algorithm 3.4.1 to do multiple imputations.

3.4.3 Inference Using MI

This subsection is based on [Barnard and Rubin, 1999] and it introduces a method to draw inference when using multiple imputation. The method considers a single parameter β_i which is the ith entry of the vector $\boldsymbol{\beta}$ from the previous sections. Let ν_{com} denote the

complete-data degrees of freedom and

$$\begin{aligned} r &= \frac{m+1}{m} B_{i,i} \widehat{\Sigma}_{i,i}^{-1}, \\ \nu_1 &= \frac{(m-1)}{r^2}, \\ \nu_2 &= \frac{\nu_{com}+1}{\nu_{com}+3}(1-r)\nu_{com}, \\ \nu_{MI} &= \left(\frac{1}{\nu_1} + \frac{1}{\nu_2}\right)^{-1}, \end{aligned}$$

where m is the number of imputations, $B_{i,i}$ is the ith diagonal entry of (3.15) and $\widehat{\Sigma}_{i,i}^{-1}$ is the ith diagonal entry of (3.16). Remark, $B_{i,i}$ and $\widehat{\Sigma}_{i,i}^{-1}$ are both scalars since β_i is one dimensional. Let $\widehat{\beta}_i$ be the ith entry of (3.13). According to [Barnard and Rubin, 1999], the inference for β_i can be based on the t-statistic:

$$t(\beta_0) = \widehat{\Sigma}_{i,i}^{-1/2} \left(\widehat{\beta}_i - \beta_0\right)$$

with a $t_{\nu_{MI}}$ reference distribution, where $t_{\nu_{MI}}$ is a t-distribution with ν_{MI} degrees of freedom. The p-value for the two-tailed test of $H_0 \colon \beta_i = \beta_0$ is then:

$$p = 2P\left(T \geq |t(\beta_0)|\right),$$

where $T \sim t_{\nu_{MI}}$. The 95% confidence interval for β_i based on multiple imputation is:

$$\left[\widehat{\beta}_i + t_{\nu_{MI}}(0.025)\ \widehat{\Sigma}_{i,i}^{1/2}\ ;\ \widehat{\beta}_i + t_{\nu_{MI}}(0.975)\ \widehat{\Sigma}_{i,i}^{1/2}\right],$$

where $t_{\nu_{MI}}(\cdot)$ is the quantile function of the t-distribution with ν_{MI} degrees of freedom.

4. Mixed Models

In this chapter, the theory of mixed models is introduced. The chapter is based on [Davis, 2002] and [Diggle et al., 1994]. In the following section, the linear mixed model is defined.

4.1 Definition

Consider $\mathbf{Y}$ to be a stochastic vector of n normal distributed variables. Let X be a $n \times p$ matrix consisting of p covariates for $\mathbf{Y}$, and $\boldsymbol{\beta}$ be a vector of p unknown parameters. A realization of $\mathbf{Y}$ is denoted by $\mathbf{y}$. Moreover, the k dimensional zero vector is denoted by $\mathbf{0}_k$.

Definition 4.1.1
Let Z be a $n \times q$ matrix. *The linear mixed model* for $\mathbf{Y}$ is given by:

$$\mathbf{Y} = X\boldsymbol{\beta} + Z\mathbf{U} + \boldsymbol{\varepsilon}, \tag{4.1}$$

where $\mathbf{U} \sim N_q(\mathbf{0}_q, \Psi)$ and $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}_n, \Sigma)$ are uncorrelated.

Notice, from Definition 4.1.1 it follows that $\mathbf{Y} \sim N_n(X\boldsymbol{\beta}, Z\Psi Z^\top + \Sigma)$. If $\mathbf{U} = \mathbf{u}$ is known, then $\mathbf{Y} \mid \mathbf{U} = \mathbf{u} \sim N_n(X\boldsymbol{\beta} + Z\mathbf{u}, \Sigma)$. In the context of a linear mixed model, the variables related to $\boldsymbol{\beta}$ and the variables related to $\mathbf{U}$ are often referred to as *the fixed effects* and *the random effects*, respectively. In practice, a specific structure of the covariance matrix Σ is assumed. The structure could e.g. be independent or even unstructured.

4.2 Parameter Estimation

Estimation of the parameters in a linear mixed model can be done by using the restricted maximum likelihood (REML) approach. Let $V(\boldsymbol{\alpha}) = \text{Var}[\mathbf{Y}] = Z\Psi Z^\top + \Sigma$ and $\boldsymbol{\alpha}$ be the vector of parameters for the covariance matrix of $\mathbf{Y}$. If $\boldsymbol{\alpha}$ is known, then the REML estimate of $\boldsymbol{\beta}$ is:

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \left(X^\top V(\boldsymbol{\alpha})^{-1} X\right)^{-1} X^\top V(\boldsymbol{\alpha})^{-1}\mathbf{y}, \tag{4.2}$$

which is equivalent to the MLE of $\boldsymbol{\beta}$. If $\boldsymbol{\alpha}$ is unknown, then it has to be estimated in order to get the REML estimate of $\boldsymbol{\beta}$. The REML estimate of $\boldsymbol{\alpha}$ is:

$$\widehat{\boldsymbol{\alpha}} = \arg\max_{\boldsymbol{\alpha}} -\frac{1}{2}\left(\log|V(\boldsymbol{\alpha})| + \log|X^\top V(\boldsymbol{\alpha})^{-1}X| + RSS(\boldsymbol{\alpha})\right), \tag{4.3}$$

where $RSS(\boldsymbol{\alpha}) = (\mathbf{y} - X\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}))^\top V(\boldsymbol{\alpha})^{-1}(\mathbf{y} - X\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}))$. In case $\boldsymbol{\alpha}$ is unknown, the REML estimate of $\boldsymbol{\beta}$ is (4.2) evaluated at $\widehat{\boldsymbol{\alpha}}$. The covariance matrix of $\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ is:

$$\begin{aligned}\text{Var}\left[\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})\right] &= \left(X^\top V(\boldsymbol{\alpha})^{-1}X\right)^{-1} X^\top V(\boldsymbol{\alpha})^{-1}\text{Var}[\mathbf{Y}]V(\boldsymbol{\alpha})^{-1}X\left(X^\top V(\boldsymbol{\alpha})^{-1}X\right)^{-1} \\ &= \left(X^\top V(\boldsymbol{\alpha})^{-1}X\right)^{-1},\end{aligned} \tag{4.4}$$

where $\mathrm{Var}[\mathbf{Y}] = V(\boldsymbol{\alpha})$ have been applied. An estimate of (4.4) is:

$$\Sigma_{\widehat{\beta}(\widehat{\alpha})} = \left(X^\top V\left(\widehat{\boldsymbol{\alpha}}\right)^{-1} X\right)^{-1}, \tag{4.5}$$

where $\boldsymbol{\alpha}$ in (4.4) has been replaced by the REML estimate $\widehat{\boldsymbol{\alpha}}$.

4.2.1 Background of REML

Let A be a $n \times (n-p)$ matrix with columns spanning $M^\perp$, where $M = span(X)$ and $M^\perp$ denotes the orthogonal complement of M. Then it follows that $A^\top X = \mathbf{0}_{(n-p)\times p}$, where $\mathbf{0}_{(n-p)\times p}$ is a $(n-p) \times p$ matrix with 0 in all entries. This implies:

$$\begin{aligned}\widetilde{\mathbf{Y}} &= A^\top Y \\ &= A^\top X\boldsymbol{\beta} + A^\top Z\mathbf{U} + A^\top \boldsymbol{\varepsilon} \\ &= A^\top Z\mathbf{U} + A^\top \boldsymbol{\varepsilon},\end{aligned} \tag{4.6}$$

where (4.1) has been applied. Moreover, $\widetilde{\mathbf{Y}}$ is normal distributed with $\mathrm{E}[\widetilde{\mathbf{Y}}] = \mathbf{0}_{(n-p)}$ and $\mathrm{Var}[\widetilde{\mathbf{Y}}] = A^\top V(\boldsymbol{\alpha})A$ which follows from Definition 4.1.1 and (4.6). The REML log-likelihood is given by:

$$\begin{aligned}\ell_{REML}(\boldsymbol{\alpha} \mid \widetilde{\mathbf{y}}) &= -\frac{1}{2}\left(\log|A^\top V(\boldsymbol{\alpha})A| + \widetilde{\mathbf{y}}^\top (A^\top V(\boldsymbol{\alpha})A)^{-1}\widetilde{\mathbf{y}}\right) \\ &= -\frac{1}{2}\left(\log|A^\top V(\boldsymbol{\alpha})A| + \mathbf{y}^\top A\left(A^\top V(\boldsymbol{\alpha})A\right)^{-1} A^\top \mathbf{y}\right) \\ &= -\frac{1}{2}\left(\log|A^\top V(\boldsymbol{\alpha})A| + \left(\mathbf{y} - X\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})\right)^\top V(\boldsymbol{\alpha})^{-1}\left(\mathbf{y} - X\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})\right)\right) \\ &= -\frac{1}{2}\left(\log|V(\boldsymbol{\alpha})| + \log|X^\top V(\boldsymbol{\alpha})^{-1}X| + \log|A^\top A| - \log|X^\top X| + RSS(\boldsymbol{\alpha})\right) \\ &= -\frac{1}{2}\left(\log|V(\boldsymbol{\alpha})| + \log|X^\top V(\boldsymbol{\alpha})^{-1}X| + RSS(\boldsymbol{\alpha})\right) \\ &\quad -\frac{1}{2}\left(\log|A^\top A| - \log|X^\top X|\right),\end{aligned} \tag{4.7}$$

where Proposition A.1.2 in Appendix A has been applied. The REML estimate of $\boldsymbol{\alpha}$ in (4.3) can be obtained from (4.7):

$$\begin{aligned}\widehat{\boldsymbol{\alpha}} &= \arg\max_{\boldsymbol{\alpha}} \ell_{REML}(\boldsymbol{\alpha} \mid \widetilde{\mathbf{y}}) \\ &= \arg\max_{\boldsymbol{\alpha}} -\frac{1}{2}\left(\log|V(\boldsymbol{\alpha})| + \log|X^\top V(\boldsymbol{\alpha})^{-1}X| + RSS(\boldsymbol{\alpha})\right),\end{aligned}$$

where it has been applied that both $\log|A^\top A|$ and $\log|X^\top X|$ are fixed according to $\boldsymbol{\alpha}$. It is worth remarking that the REML estimate of $\boldsymbol{\alpha}$ in (4.3) does not depend on A nor $\boldsymbol{\beta}$.

4.3 Mixed Models for Repeated Measurements

In this section, the theory of mixed models is represented in the context of repeated measurements. Let $\mathbf{Y}_i = (Y_{i,1}, \ldots, Y_{i,t_i})^\top$ be the stochastic vector of t_i repeated measurements from subject i for $i = 1, \ldots, n$. A realization of $\mathbf{Y}_i$ is denoted by $\mathbf{y}_i$. Let X_i be a $t_i \times p$ matrix consisting of p covariates for $\mathbf{Y}_i$, and $\boldsymbol{\beta}$ be a vector of p unknown parameters. The next definition is an extension of the linear mixed model in Definition 4.1.1.

Definition 4.3.1
Let Z_i be a $t_i \times q$ matrix for $i = 1, \ldots, n$. *The linear mixed model for repeated measurements* $\mathbf{Y}_i$ is given by:

$$\mathbf{Y}_i = X_i\boldsymbol{\beta} + Z_i\mathbf{U}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \ldots, n, \tag{4.8}$$

where $\mathbf{U}_i \sim N_q(\mathbf{0}_q, \Psi)$ and $\boldsymbol{\varepsilon}_i \sim N_{t_i}(\mathbf{0}_{t_i}, \Sigma_i)$. Furthermore, $\mathbf{U}_i \perp\!\!\!\perp \mathbf{U}_j$ and $\boldsymbol{\varepsilon}_i \perp\!\!\!\perp \boldsymbol{\varepsilon}_j$, $j \neq i$, and $\mathbf{U}_i \perp\!\!\!\perp \boldsymbol{\varepsilon}_j$ for $i, j = 1, \ldots, n$.

From Definition 4.3.1, it follows that $\mathbf{Y}_i \sim N_{t_i}(X_i\boldsymbol{\beta}, V_i(\boldsymbol{\alpha}))$ and $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ are independent, where $V_i(\boldsymbol{\alpha}) = Z_i\Psi Z_i^\top + \Sigma_i$ for $i = 1, \ldots, n$ and $\boldsymbol{\alpha}$ is the vector of parameters related to the covariance matrix of $\mathbf{Y}_i$. Remark, $\Sigma_1, \ldots, \Sigma_n$ in Definition 4.3.1 are assumed to be submatrices of Σ_k where $k = \max(t_1, \ldots, t_n)$. This means that the parameters in $\boldsymbol{\alpha}$ are assumed to be the same across the different subjects. In practice, a specific structure of $\Sigma_1, \ldots, \Sigma_n$ is assumed. The type of model in Definition 4.3.1 is referred to as a *MMRM*. Definition 4.3.1 is very general since the number of measurements t_i can differ for different subjects i. If t_i differs for different subjects i, then the data are said to be unbalanced. However, if $t_1 = t_2 = \ldots = t_n$, then the data are said to be balanced and it is assumed that $\Sigma_1 = \Sigma_2 = \ldots = \Sigma_n$. The REML estimate of $\boldsymbol{\beta}$ for a MMRM is:

$$\widehat{\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\alpha}}\right) = \left(\sum_{i=1}^{n} X_i^\top V_i\left(\widehat{\boldsymbol{\alpha}}\right)^{-1} X_i\right)^{-1} \left(\sum_{i=1}^{n} X_i^\top V_i\left(\widehat{\boldsymbol{\alpha}}\right)^{-1} \mathbf{y}_i\right), \tag{4.9}$$

where $\widehat{\boldsymbol{\alpha}}$ is the REML estimate of $\boldsymbol{\alpha}$ given by:

$$\widehat{\boldsymbol{\alpha}} = \arg\max_{\boldsymbol{\alpha}} -\frac{1}{2}\sum_{i=1}^{n}\Big(\log|V_i(\boldsymbol{\alpha})| + \log|X_i^\top V_i(\boldsymbol{\alpha})^{-1}X_i| + RSS_i(\boldsymbol{\alpha})\Big), \tag{4.10}$$

where $RSS_i(\boldsymbol{\alpha}) = (\mathbf{y}_i - X_i\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}))^\top V_i(\boldsymbol{\alpha})^{-1}(\mathbf{y}_i - X_i\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}))$ for $i = 1, \ldots, n$. The estimates in (4.9) and (4.10) are based on the fact that:

$$f(\mathbf{y}_1, \ldots, \mathbf{y}_n) = \prod_{i=1}^{n} f(\mathbf{y}_i),$$

when $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ are independent. Let $C = \left(\sum_{i=1}^{n} X_i^\top V_i(\boldsymbol{\alpha})^{-1}X_i\right)$. The covariance matrix of $\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ is:

$$\begin{aligned}
\mathrm{Var}\left[\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})\right] &= C^{-1}\mathrm{Var}\left[\sum_{i=1}^{n} X_i^\top V_i(\boldsymbol{\alpha})^{-1}\mathbf{Y}_i\right] C^{-1} \\
&= C^{-1}\left(\sum_{i=1}^{n} X_i^\top V_i(\boldsymbol{\alpha})^{-1}\mathrm{Var}[\mathbf{Y}_i]V_i(\boldsymbol{\alpha})^{-1}X_i\right) C^{-1} \\
&= \left(\sum_{i=1}^{n} X_i^\top V_i(\boldsymbol{\alpha})^{-1}X_i\right)^{-1},
\end{aligned} \tag{4.11}$$

which is based on that $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ are independent and $\text{Var}[\mathbf{Y}_i] = V_i(\boldsymbol{\alpha})$. The covariance matrix in (4.11) can be estimated by:

$$\Sigma_{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}})} = \left(\sum_{i=1}^{n} X_i^\top V_i \left(\widehat{\boldsymbol{\alpha}} \right)^{-1} X_i \right)^{-1}, \tag{4.12}$$

where $\boldsymbol{\alpha}$ in (4.11) has been replaced by the REML estimate $\widehat{\boldsymbol{\alpha}}$.

4.4 Inference for Fixed Effects

This section introduces a method to draw inference for $\mathbf{w}^\top \boldsymbol{\beta}$, where $\mathbf{w}$ is a p dimensional vector. The inference for $\mathbf{w}^\top \boldsymbol{\beta}$ can be based on the t-statistic:

$$t(\beta_0) = \left(\mathbf{w}^\top \Sigma_{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}})} \mathbf{w} \right)^{-1/2} \left(\mathbf{w}^\top \widehat{\boldsymbol{\beta}} \left(\widehat{\boldsymbol{\alpha}} \right) - \beta_0 \right)$$

with a t_{ν_S} reference distribution, where t_{ν_S} is a t-distribution with ν_S degrees of freedom. The p-value of the two-tailed test of $H_0 \colon \mathbf{w}^\top \boldsymbol{\beta} = \beta_0$ is:

$$p = 2P \left(T \geq |t(\beta_0)| \right),$$

where $T \sim t_{\nu_S}$. The 95% confidence interval for $\mathbf{w}^\top \boldsymbol{\beta}$ is:

$$\left[\mathbf{w}^\top \widehat{\boldsymbol{\beta}} \left(\widehat{\boldsymbol{\alpha}} \right) + t_{\nu_S}(0.025) \left(\mathbf{w}^\top \Sigma_{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}})} \mathbf{w} \right)^{1/2} ; \mathbf{w}^\top \widehat{\boldsymbol{\beta}} \left(\widehat{\boldsymbol{\alpha}} \right) + t_{\nu_S}(0.975) \left(\mathbf{w}^\top \Sigma_{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}})} \mathbf{w} \right)^{1/2} \right],$$

where $t_{\nu_S}(\cdot)$ is the quantile function of the t-distribution with ν_S degrees of freedom. According to [Brown and Prescott, 1999], the degrees of freedom ν_S can be calculated from Satterthwaite's approximation:

$$\nu_S = \frac{2 \left(\mathbf{w}^\top \Sigma_{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}})} \mathbf{w} \right)^2}{\text{Var} \left[\mathbf{w}^\top \Sigma_{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}})} \mathbf{w} \right]}$$

where $\text{Var} \left[\mathbf{w}^\top \Sigma_{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}})} \mathbf{w} \right]$ usually has to be approximated.

5. Logistic Regression

This chapter introduces logistic regressions and it is based on [Harrell, 2001]. In the following section, the logistic regression is defined.

5.1 Definition

Let $\mathbf{Y}$ be a stochastic vector consisting of n independent Bernoulli distributed variables Y_i, $i = 1, \ldots, n$. Realizations of $\mathbf{Y}$ and Y_i are denoted by $\mathbf{y}$ and y_i, respectively. Furthermore, let X be a $n \times p$ matrix consisting of p covariates for $\mathbf{Y}$, and $\boldsymbol{\beta}$ be a vector of p unknown parameters. The ith row of X is denoted by $\mathbf{x}_i^\top$ and it corresponds to the covariates for Y_i, $i = 1, \ldots, n$.

Definition 5.1.1
Let $P(Y_i = k \mid \mathbf{x}_i)$ be the probability of $Y_i = k$ given $\mathbf{x}_i$, where $k \in \{0, 1\}$. *The logistic regression* is given by:

$$\text{logit}\left(P(Y_i = 1 \mid \mathbf{x}_i)\right) = \mathbf{x}_i^\top \boldsymbol{\beta}, \quad i = 1, \ldots, n, \tag{5.1}$$

where $\text{logit}(p) = \log\left(\dfrac{p}{1-p}\right)$ for $p \in]0; 1[$.

From Definition 5.1.1, it follows that the probability of $P(Y_i = 1 \mid \mathbf{x}_i)$ can be written as:

$$\begin{aligned} P(Y_i = 1 \mid \mathbf{x}_i) &= \text{expit}\left(\text{logit}\left(P(Y_i = 1 \mid \mathbf{x}_i)\right)\right) \\ &= \text{expit}\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right) \\ &= \frac{\exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right)}{1 + \exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right)}, \end{aligned} \tag{5.2}$$

where $\text{expit}(x) = \dfrac{\exp(x)}{1 + \exp(x)}$ has been applied. Remark, the left hand side of (5.1) can be written as:

$$\begin{aligned} \text{logit}\left(P(Y_i = 1 \mid \mathbf{x}_i)\right) &= \log\left(\frac{P(Y_i = 1 \mid \mathbf{x}_i)}{1 - P(Y_i = 1 \mid \mathbf{x}_i)}\right) \\ &= \log\left(\frac{P(Y_i = 1 \mid \mathbf{x}_i)}{P(Y_i = 0 \mid \mathbf{x}_i)}\right) \\ &= \log\left(Odds(Y_i = 1 \mid \mathbf{x}_i)\right), \end{aligned} \tag{5.3}$$

where $Odds(Y_i = 1 \mid \mathbf{x}_i)$ is the odds of $Y_i = 1$ given $\mathbf{x}_i$. From (5.1) and (5.3), it follows that a logistic regression models the log odds of $Y_i = 1$ as a linear function of the covariates in $\mathbf{x}_i$. In practice, if a logistic regression includes both an intercept and a categorical covariate, then one of the levels of the categorical covariate is incorporated in the intercept

as a reference level. Furthermore, the parameter corresponding to a certain level of the categorical covariate is the log odds ratio between the certain level and the reference level of the categorical covariate. This follows from (5.1) and (5.3):

$$\begin{aligned}\log\left(\frac{Odds(Y_i = 1 \mid X_i = b)}{Odds(Y_i = 1 \mid X_i = a)}\right) &= \log\left(Odds(Y_i = 1 \mid X_i = b)\right) - \log\left(Odds(Y_i = 1 \mid X_i = a)\right) \\ &= (\alpha + \beta_b) - (\alpha) \\ &= \beta_b, \end{aligned} \tag{5.4}$$

where the considered logistic regression only includes an intercept α and a categorical covariate X_i with to levels $\{a, b\}$ for which β_b denotes the parameter corresponding to $X_i = b$. Remark, if the logistic regression includes other covariates, the result in (5.4) is still true. In the next section, a method to estimate the parameters in (5.1) is introduced.

5.2 Parameter Estimation

The vector of parameters $\boldsymbol{\beta}$ in a logistic regression is estimated by using the maximum likelihood approach. In order to derive the log-likelihood of $\boldsymbol{\beta}$, it is necessary to consider the density of Y_i given $\mathbf{x}_i$ and $\boldsymbol{\beta}$, which is:

$$\begin{aligned} f(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}) &= P(Y_i = 1 \mid \mathbf{x}_i)^{y_i} \left(1 - P(Y_i = 1 \mid \mathbf{x}_i)\right)^{1-y_i} \\ &= \left(\frac{\exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right)}{1 + \exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right)}\right)^{y_i} \left(1 - \frac{\exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right)}{1 + \exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right)}\right)^{1-y_i} \\ &= \left(\frac{\exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right)}{1 + \exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right)}\right)^{y_i} \left(\frac{1}{1 + \exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right)}\right)^{1-y_i} \\ &= \exp\left(y_i \log\left(\frac{\exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right)}{1 + \exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right)}\right) - (1 - y_i) \log\left(1 + \exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right)\right)\right) \\ &= \exp\left(y_i \, \mathbf{x}_i^\top \boldsymbol{\beta} - \log\left(1 + \exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right)\right)\right), \end{aligned} \tag{5.5}$$

where (5.2) has been applied. Since $Y_1, \ldots, Y_n$ are assumed to be independent, the density of $\mathbf{Y}$ given X and $\boldsymbol{\beta}$ can be written as the product of (5.5) for $i = 1, \ldots, n$:

$$\begin{aligned} f(\mathbf{y} \mid X, \boldsymbol{\beta}) &= \prod_{i=1}^{n} f(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}) \\ &= \exp\left(\sum_{i=1}^{n} y_i \, \mathbf{x}_i^\top \boldsymbol{\beta} - \log\left(1 + \exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right)\right)\right) \\ &= \exp\left(\mathbf{y}^\top X \boldsymbol{\beta} - \sum_{i=1}^{n} \log\left(1 + \exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right)\right)\right). \end{aligned} \tag{5.6}$$

Based on (5.6), the log-likelihood of $\boldsymbol{\beta}$ is given by:

$$\ell(\boldsymbol{\beta} \mid \mathbf{y}, X) = \mathbf{y}^\top X \boldsymbol{\beta} - \sum_{i=1}^{n} \log\left(1 + \exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right)\right). \tag{5.7}$$

The MLE of $\boldsymbol{\beta}$ is obtained by maximizing (5.7):

$$\widehat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \mathbf{y}^\top X\boldsymbol{\beta} - \sum_{i=1}^{n} \log\left(1 + \exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right)\right). \tag{5.8}$$

Often, it is not possible to get an explicit formulation of the MLE of $\boldsymbol{\beta}$. In such cases, (5.7) can be maximized by applying an iterative procedure like the Newton-Raphson method. Let $\widehat{p}_i$ denote the estimated probability of $Y_i = 1$ given $\mathbf{x}_i$, which equals:

$$\widehat{p}_i = \frac{\exp\left(\mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}\right)}{1 + \exp\left(\mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}\right)}, \tag{5.9}$$

where (5.2) has been applied. Based on (5.9), an estimate of the variance of Y_i is:

$$\begin{aligned}
\widehat{\sigma}^2_{Y_i} &= \widehat{p}_i\left(1 - \widehat{p}_i\right) \\
&= \frac{\exp\left(\mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}\right)}{1 + \exp\left(\mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}\right)} \left(1 - \frac{\exp\left(\mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}\right)}{1 + \exp\left(\mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}\right)}\right) \\
&= \frac{\exp\left(\mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}\right)}{1 + \exp\left(\mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}\right)} \left(\frac{1}{1 + \exp\left(\mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}\right)}\right) \\
&= \frac{\exp\left(\mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}\right)}{\left(1 + \exp\left(\mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}\right)\right)^2}.
\end{aligned} \tag{5.10}$$

The covariance matrix of $\widehat{\boldsymbol{\beta}}$ is estimated by:

$$\Sigma_{\widehat{\boldsymbol{\beta}}} = \left(X^\top \widehat{\Sigma}_{\mathbf{Y}} X\right)^{-1}, \tag{5.11}$$

where $\widehat{\Sigma}_{\mathbf{Y}}$ is the estimated covariance matrix of $\mathbf{Y}$, which is a $n \times n$ diagonal matrix with $\widehat{\sigma}^2_{Y_i}$ in the ith diagonal entry for $i = 1, \ldots, n$.

5.3 Inference

This section introduces a method to draw inference in the logistic regression setup. Let $\mathbf{w}$ be a p dimensional vector and consider the hypothesis of $\mathbf{w}^\top \boldsymbol{\beta} = \beta_0$. The inference for $\mathbf{w}^\top \boldsymbol{\beta}$ can be based on the z-statistic:

$$z(\beta_0) = \left(\mathbf{w}^\top \Sigma_{\widehat{\boldsymbol{\beta}}} \mathbf{w}\right)^{-1/2} \left(\mathbf{w}^\top \widehat{\boldsymbol{\beta}} - \beta_0\right)$$

with a $N(0,1)$ reference distribution. The p-value of the two-tailed test of $H_0 \colon \mathbf{w}^\top \boldsymbol{\beta} = \beta_0$ is:

$$p = 2P\left(Z \geq |z(\beta_0)|\right),$$

where $Z \sim N(0,1)$. The 95% confidence interval for $\mathbf{w}^\top \boldsymbol{\beta}$ is:

$$\left[\mathbf{w}^\top \widehat{\boldsymbol{\beta}} + q(0.025)\ \left(\mathbf{w}^\top \Sigma_{\widehat{\boldsymbol{\beta}}} \mathbf{w}\right)^{1/2}\ ;\ \mathbf{w}^\top \widehat{\boldsymbol{\beta}} + q(0.975)\ \left(\mathbf{w}^\top \Sigma_{\widehat{\boldsymbol{\beta}}} \mathbf{w}\right)^{1/2}\right],$$

where $q(\cdot)$ is the quantile function of the standard normal distribution.

6. Use of Estimands in NN1998-2076

The clinical trial NN1998-2076 was conducted before the regulatory authorities started to focus on missing data and estimands. Hence, the clinical trial report (CTR) did not include any estimands. In this chapter, four estimands are formulated for NN1998-2076 in order to explore and investigate the practical use of estimands. The estimands only target a single intercurrent event in order to show how the different strategies for addressing intercurrent events from Section 2.2 answers different scientific questions. The intercurrent event of interest is *use of escape therapy*. In the AERx arm, *use of escape therapy* is classified as use of Insulin Aspart. In the Insulin Aspart arm, *use of escape therapy* means that subjects take a higher amount of Insulin Aspart than actually planned according to the protocol. Occurrences of all other intercurrent events are ignored, which corresponds to use the treatment policy strategy for these events, cf. Section 2.2.1. An estimand is formulated for each of the different strategies for addressing intercurrent events except the principal stratum strategy. The principal stratum strategy requires a trial design which gives an indication of the subjects in the principal stratum of interest, cf. Section 2.2.4. In NN1998-2076, it is not possible to determine the subjects in a principal stratum and therefore an estimand based on the principal stratum strategy is not analysed in this thesis. The four estimands are named after the strategy used to address *use of escape therapy*. In general, it is not recommended to name an estimand after the used strategy since an estimand should address all important intercurrent events and the different events do not have to be addressed with the same strategy, cf. Section 2.2.

Recall from Chapter 2 that an estimand and the related statistical analyses have to be prespecified in a clinical trial. Hence, it is not possible to remove or add any covariates in the statistical models afterwards even though it is significant to do so. In this thesis, the estimands and the related statistical analyses are handled as if they were prespecified even though they are not.

Appendix C.1 consists of the `R`-code used to set up the datasets which are analysed in this chapter. The significance level used in this thesis is $\alpha = 0.05$. The following sections describe and analyse the four different estimands.

6.1 Treatment Policy Estimand

In this section, an estimand based on the treatment policy strategy is analysed. The estimand is named *the treatment policy estimand* and it is given by:

A. All exposed subjects with an observed baseline value of HbA1c.

B. Change from baseline to visit 13 in HbA1c.

C. *Use of escape therapy* is ignored. All other intercurrent events are also ignored.

D. Difference in means between the AERx arm and the Insulin Aspart arm.

This estimand tries to estimate the treatment difference between *AERx + escape therapy* and *Insulin Aspart + escape therapy*. The goal of the analysis related to this estimand is to show that *AERx + escape therapy* is non-inferior to *Insulin Aspart + escape therapy*.

A study of this type is called *a non-inferiority study*. Remark, the treatment difference is the effect related to AERx minus the effect related to Insulin Aspart. Furthermore, the target of both treatments is to lower HbA1c. Hence, a negative treatment difference reflects that AERx is better than Insulin Aspart. Therefore, the test for non-inferiority is a left-tailed test in this case. The null and alternative hypotheses are given by:

$$H_0 \colon \beta_T \geq 0.4$$
$$H_A \colon \beta_T < 0.4,$$

where β_T is the treatment difference and 0.4 is a non-inferiority margin commonly accepted by the FDA when the endpoint is HbA1c, cf. [FDA, 2008]. The alternative hypothesis is equivalent to say that *AERx + escape therapy* is non-inferior to *Insulin Aspart + escape therapy*. The p-value related to the test for non-inferiority is given by:

$$p = P(X \leq x),$$

where x is the observed test statistic and X is a stochastic variable following the reference distribution. Since the test for non-inferiority is a one-tailed test, the half of the significance level is used as a threshold, which is 0.025, cf. [ICH, 1998]. If the p-value is below 0.025, then the null-hypothesis can be rejected and non-inferiority can be claimed. In the following, the primary analysis related to the treatment policy estimand is specified followed by a specification of the different sensitivity analyses.

Primary Analysis for The Treatment Policy Estimand

The primary analysis is a MMRM based on the selection model approach, cf. Section 3.2 and Section 4.3. In this analysis, the statements in Definition 3.2.2 are assumed. The analysis is referred to as $MMRM_S$.

The response variable of the MMRM is the change from baseline in HbA1c. The fixed effects are treatment, visit, the interaction between treatment and visit, a covariate adjustment for the HbA1c baseline value and the interaction between the HbA1c baseline value and visit. The random effect is subject. The parameters are estimated by applying the theory in Section 4.3. In `R`, the function `lmer` is used to fit an MMRM. The function is from the `R`-package of the same name.

The parameter of interest is the treatment difference at visit 13, which is denoted by $\beta_{T_{13}}$. The inference for $\beta_{T_{13}}$ is drawn by using the theory from Section 4.4.

Sensitivity Analysis I for The Treatment Policy Estimand

This sensitivity analysis is a MMRM based on the pattern-mixture model approach, cf. Section 3.3 and Section 4.3. The treatment difference is assumed to be different for subjects who complete visit 13 and subjects who do not complete visit 13. The purpose of this sensitivity analysis is to investigate how robust the results from the primary analysis are to deviations from the MAR assumption. This sensitivity analysis is referred to as $MMRM_{PM}$.

A binary variable indicating whether or not subjects complete visit 13 is included in the MMRM and it is referred to as *the completer indicator*. The response variable of the MMRM is the change from baseline in HbA1c. The fixed effects are treatment, visit, the

interaction between treatment and visit, a covariate adjustment for the HbA1c baseline value, the interaction between the HbA1c baseline value and visit, the completer indicator and the interaction between the completer indicator and treatment. The random effect is subject.

The subjects, who complete visit 13, are referred to as *completers* and the subjects, who do not complete visit 13, are referred to as *non-completers*. Let π_C and π_N denote the proportions of completers and non-completers, respectively. Furthermore, let $\beta_{T_{13}^C}$ and $\beta_{T_{13}^N}$ be the treatment differences at visit 13 for completers and non-completers, respectively. The parameter of interest is the average of the treatment differences at visit 13, which is given by:

$$\begin{aligned}\beta_{T_{13}} &= \pi_C\, \beta_{T_{13}^C} + \pi_N\, \beta_{T_{13}^N} \\ &= (1-\pi_N)\beta_{T_{13}^C} + \pi_N\, \beta_{T_{13}^N} \\ &= \beta_{T_{13}^C} + \pi_N \left(\beta_{T_{13}^N} - \beta_{T_{13}^C}\right) \\ &= \beta_{T_{13}^C} + \pi_N\, \beta_{T_{13}^D}, \end{aligned} \tag{6.1}$$

where $\beta_{T_{13}^D} = \beta_{T_{13}^N} - \beta_{T_{13}^C}$. The parameters in the MMRM are estimated by applying the theory from Section 4.3. The proportion of non-completers π_N is estimated by using the MLE of a binomial distribution.

The proportion π_N in (6.1) is considered to be stochastic. Therefore, the inference for $\beta_{T_{13}}$ is drawn by using the Delta Method in Appendix B, where $\beta_{T_{13}^C}$, $\beta_{T_{13}^D}$ and π_N are assumed to follow a normal distribution. The Jacobian matrix of (6.1) is:

$$J_{\beta_{T_{13}}} = \begin{bmatrix} 1 & \pi_N & \beta_{T_{13}^D} \end{bmatrix},$$

where $\beta_{T_{13}}$ is considered as a function of $\left(\beta_{T_{13}^C},\, \beta_{T_{13}^D},\, \pi_N\right)^\top$.

Sensitivity Analysis II for The Treatment Policy Estimand

The second sensitivity analysis is an ANCOVA combined with MI. The analysis is based on the MI procedure described in Section 3.4, where the number of imputations is set to $m = 1000$. The missing data in the AERx arm are imputed according to the copy reference approach which assumes that the subjects in the AERx arm, who discontinue the trial, gradually respond like the subjects in the Insulin Aspart arm, after they discontinue the trial. The missing data in the Insulin Aspart arm are imputed under a assumption of MAR. This sensitivity analysis does also investigate how robust the results from the primary analysis are to deviations from the MAR assumption. The second sensitivity analysis is referred to as MI^{CR}_{ANCOVA}.

Algorithm 3.4.1 is used to impute the missing data in both the AERx arm and the Insulin Aspart arm, but step 1-3 in Algorithm 3.4.1 are only based on the data from the Insulin Aspart arm in order to do a copy reference imputation in the AERx arm. The response variables of the imputation models in Algorithm 3.4.1 are the change from baseline in HbA1c at visit 7, 9, 11, 12, and 13, respectively. All imputation models include the HbA1c baseline value as a covariate. The imputation models for visit 9, 11, 12 and 13 do also include the change from baseline in HbA1c from the previous visit(s) as covariate(s).

The 1000 complete datasets are analysed using an ANCOVA in which the response variable is the change from baseline in HbA1c at visit 13. The covariates in the ANCOVA are treatment and the HbA1c baseline value. The results from the 1000 ANCOVAs are combined using Rubin's rules, cf. Section 3.4.

The parameter of interest is the treatment difference at visit 13 denoted by $\beta_{T_{13}}$. The inference for $\beta_{T_{13}}$ is drawn based on the theory in Section 3.4.3.

The `R`-package `mice` has implemented the MI procedure. The package consists of a function called `mice` in which Algorithm 3.4.1 is implemented. The function `mice` cannot impute according to the copy reference approach. Therefore, a supplementary `R`-function to `mice` has been developed in this thesis and it can be found in Appendix C.2.

Sensitivity Analysis III for The Treatment Policy Estimand

The third sensitivity analysis is also an ANCOVA combined with MI. The only difference between this sensitivity analysis and MI^{CR}_{ANCOVA} is that the missing data in the AERx arm are imputed according to the jump to reference approach instead of the copy reference approach. The jump to reference approach assumes that the subjects in the AERx arm respond like the subjects in the Insulin Aspart arm, after they discontinue the trial. The third sensitivity analysis also investigates how robust the results from the primary analysis are to deviations from the MAR assumption. This sensitivity analysis is referred to as MI^{J2R}_{ANCOVA}.

Algorithm 3.4.1 can also be used to impute according to the jump to reference approach. The algorithm is used to impute the missing data in the Insulin Aspart arm. The missing data in the AERx arm are then imputed based on the mean change from baseline in HbA1c in the Insulin Aspart arm plus some noise. For example, if a subject in the AERx arm is missing the measurement at visit 11, then the measurement is set to be the mean change from baseline in HbA1c at visit 11 in the Insulin Aspart arm plus a random variable ε, where $\varepsilon \sim N(0, \widetilde{\sigma}^2_{11})$ and $\widetilde{\sigma}^2_{11}$ is the drawn estimate of the variance at visit 11 from Algorithm 3.4.1 applied on the Insulin Aspart arm.

The jump to reference approach is not implemented in the `R`-function `mice`. Therefore, a supplementary `R`-function to `mice` has been developed in this thesis in order to do imputations based on the jump to reference approach. The supplementary function can be found in Appendix C.3.

Sensitivity Analysis IV for The Treatment Policy Estimand

Like the two previous sensitivity analyses, the fourth sensitivity analysis consists of an ANCOVA combined with MI. The only difference between this sensitivity analysis and MI^{CR}_{ANCOVA} is how the missing data are imputed. In this sensitivity analysis, the missing data in the AERx arm are imputed separately from the the missing data in the Insulin Aspart arm in order to allow the subjects in one arm to act differently than the subjects in the other arm. The imputed data at visit 13 in the AERx arm are then adjusted according to a value $\delta \in \mathbb{R}$. The value δ^*, which gives the opposite conclusion compared to the primary analysis and for which $|\delta| \geq |\delta^*|$ holds for all $\delta \in \mathbb{R}$, is called *the tipping point*. The goal of this analysis is to find δ^*. An analysis of this type is referred to as *a tipping point analysis*.

Algorithm 3.4.1 is used separately on the AERx arm and the Insulin Aspart arm. Afterwards the imputed values at visit 13 for the AERx arm are adjusted according to a value δ in order to find the tipping point δ^*. The 1000 complete datasets are analysed the same way as in MI^{CR}_{ANCOVA}, and the parameter of interest is also $\beta_{T_{13}}$ for which the inference is drawn based on the theory from Section 3.4.3. This procedure is repeated until the tipping point δ^* is located.

Results Based on The Treatment Policy Estimand

The proportion of missing data at each visit and the missing data pattern related to the treatment policy estimand are illustrated in the following figure.

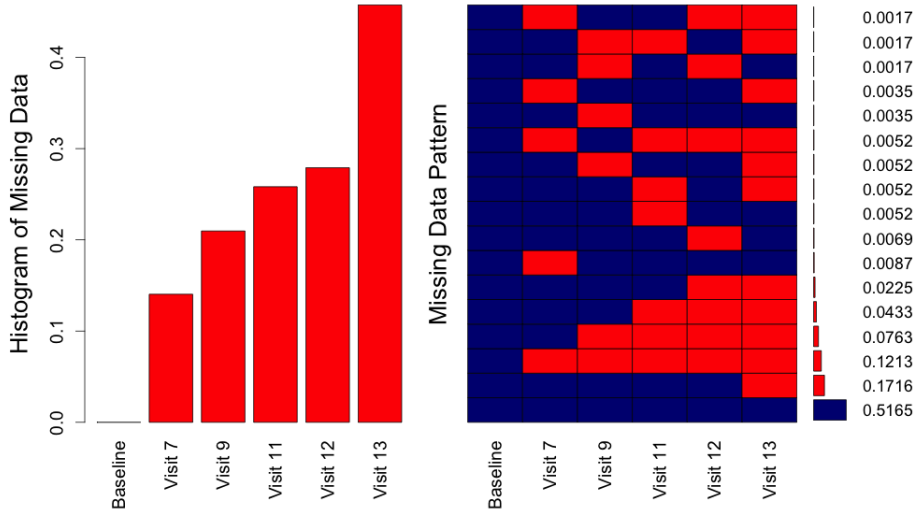

Figure 6.1: The proportions of missing data at each visit and the missing data pattern related to the treatment policy estimand.

From Figure 6.1, it follows that the missing data pattern is almost monotone. Hence, the use of Algorithm 3.4.1 seems to be justified. The R-code in Appendix C.5 is used to analyse the treatment policy estimand and Figure 6.2 gives an overview of the results.

Test for Non-Inferiority

	Estimate	95%-CI	p-value
Primary Analysis:			
$MMRM_S$	0.228	[0.075 ; 0.381]	0.014
Sensitivity Analyses:			
$MMRM_{PM}$	0.234	[0.078 ; 0.39]	0.0187
MI^{CR}_{ANCOVA}	0.182	[0.033 ; 0.332]	0.0022
MI^{J2R}_{ANCOVA}	0.086	[-0.046 ; 0.217]	0
Tipping Point Analysis:			
delta = 0	0.232	[0.069 ; 0.396]	0.0226
delta = 0.007	0.236	[0.072 ; 0.4]	0.025
delta = 0.014	0.24	[0.076 ; 0.404]	0.0276

$H_0 : \beta_{T_{13}} \geq 0.4$
$H_A : \beta_{T_{13}} < 0.4$

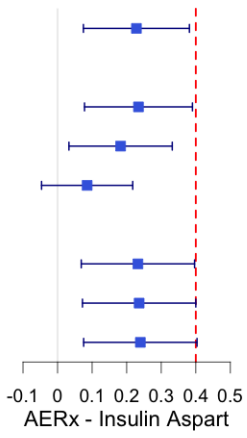

Figure 6.2: Results from the analyses related to the treatment policy estimand.

Based on the results from the primary analysis, it can be concluded that *AERx + escape therapy* is non-inferior to *Insulin Aspart + escape therapy* when *use of escape therapy* is ignored. Each of the three sensitivity analyses $MMRM_{PM}$, MI^{CR}_{ANCOVA} and MI^{J2R}_{ANCOVA} supports the conclusion drawn from the primary analysis. Therefore, the results from the primary analysis seems to be quite robust to deviations from the MAR assumption. But in this case it does not make good sense to use MI^{CR}_{ANCOVA} and MI^{J2R}_{ANCOVA} since both analyses "reward" AERx when subjects discontinue. This is caused by that the observed effect of Insulin Aspart is better than the observed effect of AERx, which can be seen in Figure 6.3 generated by the `R`-code in Appendix C.4.

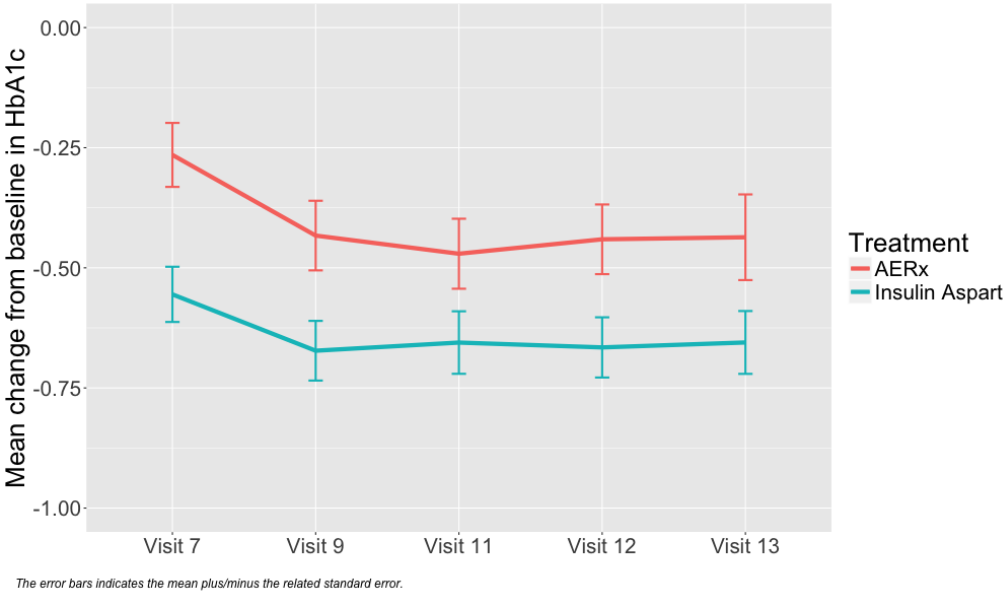

Figure 6.3: Mean of the observed change from baseline in HbA1c at each visit for the two treatment arms.

Even though MI^{CR}_{ANCOVA} and MI^{J2R}_{ANCOVA} do not make good sense, they cannot be excluded since they are considered to be prespecified. The tipping point gives an indication of how much the missing values of the change from baseline in HbA1c at visit 13 should have differed from the imputed values to give the opposite conclusion compared to the primary analysis. The result of the tipping point analysis is that the tipping point is approximately 0.007. A change of 0.007 may seem plausible and hence the results from the primary analysis may not be that robust after all. But in general the tipping point should be evaluated together with a physician.

The proportion of subjects who end up using escape therapy is 58.54% in the AERx arm and 0.69% in the Insulin Aspart arm. Therefore, the estimated treatment difference based on this estimand actually reflects the effect of *AERx + escape therapy* versus *Insulin Aspart* more than *AERx + escape therapy* versus *Insulin Aspart + escape therapy*, which was prespecified. If the tipping point of 0.007 is considered to be plausible, the conclusion based on this estimand actually is that *AERx + escape therapy* cannot be shown to be non-inferior to *Insulin Aspart* when *use of escape therapy* is ignored. In case the tipping point is not plausible, the conclusion is that *AERx + escape therapy* is non-inferior to *Insulin Aspart* when *use of escape therapy* is ignored.

6.2 Hypothetical Estimand

An estimand based on the hypothetical strategy is analysed in this section. The estimand is referred to as *the hypothetical estimand* and it is given as follows:

A. All exposed subjects with an observed baseline value of HbA1c.

B. Change from baseline to visit 13 in HbA1c.

C. If *use of escape therapy for at least 14 consecutive days* would not occur. All other intercurrent events are ignored.

D. Difference in means between the AERx arm and the Insulin Aspart arm.

This estimand targets the treatment difference between *AERx* and *Insulin Aspart* in a scenario where *use of escape therapy for at least 14 consecutive days* would not occur. Like the treatment policy estimand, the goal of the analysis related to the hypothetical estimand is also to show non-inferiority. The primary analysis and sensitivity analysis are the same as in Section 6.1.

Results Based on The Hypothetical Estimand

As a consequence of the hypothetical strategy, all HbA1c measurements after the occurrence of *use of escape therapy for at least 14 consecutive days* for a given subject have to be considered as missing, cf. Section 2.2.2. Therefore, the datasets considered under this estimand is not the same as in Section 6.1. The proportion of missing data at each visit and the missing data pattern related to the hypothetical estimand can be seen in Figure 6.4.

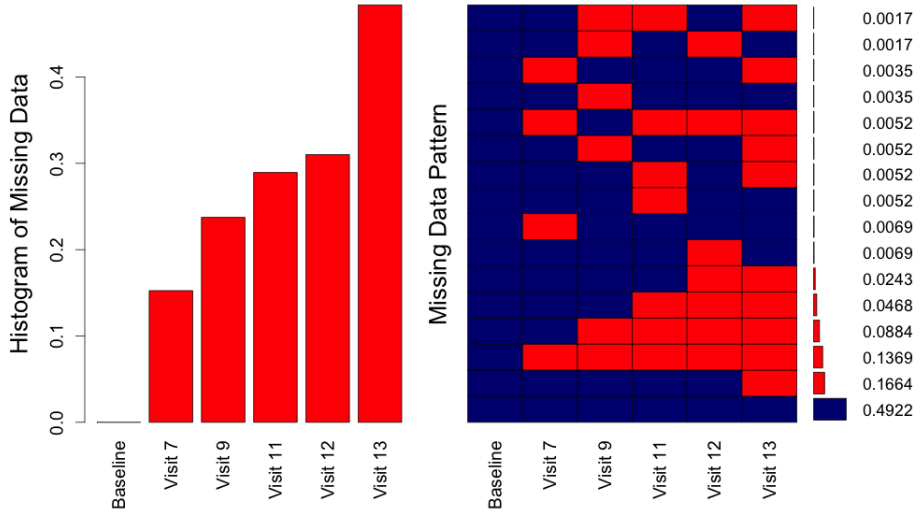

Figure 6.4: The proportions of missing data at each visit and the missing data pattern related to the hypothetical estimand.

The missing data pattern in Figure 6.4 is almost the same as the missing data pattern in Figure 6.1 and it is also close to be monotone. Thus, the use of Algorithm 3.4.1 seems to be justified. The proportions of missing data in Figure 6.4 are slightly increased compared to the proportions in Figure 6.1. The R-code used to analyse the hypothetical estimand

can be found in Appendix C.6. In Figure 6.5 the results from the different analyses related to the hypothetical estimand are presented.

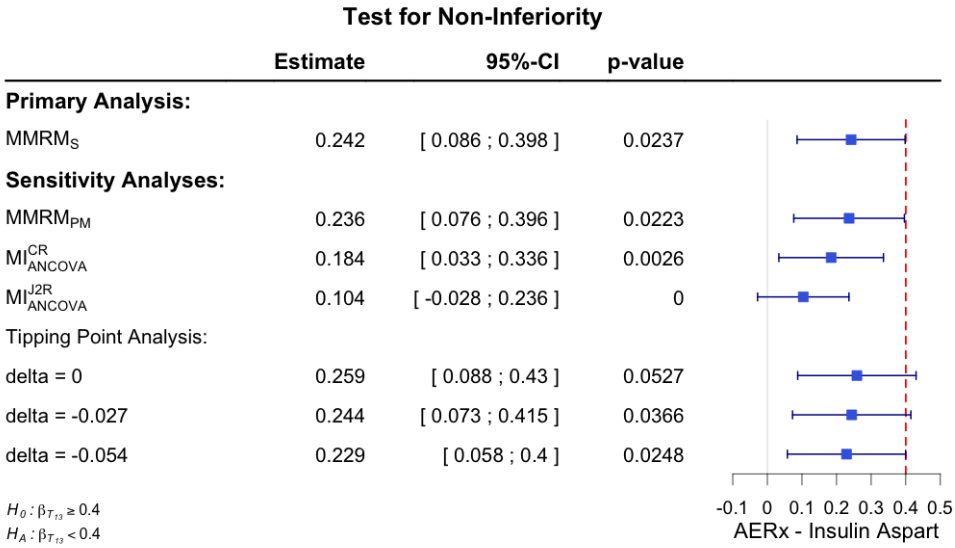

Figure 6.5: Results from the analyses related to the hypothetical estimand.

In the primary analysis, the null hypothesis can be rejected. Hence, the conclusion based on the primary analysis is that non-inferiority can be claimed in the scenario where *use of escape therapy for at least 14 consecutive days* does not occur. The sensitivity analyses MI^{CR}_{ANCOVA} and MI^{J2R}_{ANCOVA} related to the hypothetical estimand do also "reward" AERx when subjects discontinue, as in Section 6.1. Therefore, these sensitivity analyses do not make good sense in this case as well. Remark, the missing data are imputed under a assumption of MAR in the tipping point analysis when $\delta = 0$. A proper tipping point analysis cannot be done because the primary analysis $MMRM_S$ and the tipping point analysis with $\delta = 0$ results in different conclusions. The tipping point analysis with $\delta = 0$ shows that the results from the primary analysis are not very robust at all since both analyses are based on a assumption of MAR. Hence, the conclusion based on the hypothetical estimand is that non-inferiority cannot be shown in the scenario where *use of escape therapy for at least 14 consecutive days* does not occur.

6.3 Composite Estimand

In this section, an estimand called *the composite estimand* is analysed. The estimand is based on the composite strategy and it is given by:

A. All exposed subjects with an observed baseline value of HbA1c.

B. A binary response variable indicating a successful response if the HbA1c at visit 13 is below 7.0%, and *use of escape therapy for at least 14 consecutive days* did not occur.

C. *Use of escape therapy for at least 14 consecutive days* is captured through the definition of the endpoint of interest. All other intercurrent events are ignored.

D. Odds ratio of a successful response between the AERx arm and the Insulin Aspart arm.

In [EMA, 2018], it is suggested to use 7.0% as a threshold for HbA1c. The composite estimand targets the odds ratio of a successful response between the AERx arm and the Insulin Aspart arm, which is given by the odds of a successful response in the AERx arm divided by the corresponding odds in the Insulin Aspart arm. Hence, an odds ratio greater than 1 reflects that the odds is greatest in the AERx arm. The goal of the analysis related to the composite estimand is to show that the odds of a successful response in the AERx arm is greater than the corresponding odds in the Insulin Aspart arm. The goal is tested through a right-tailed test where the null and alternative hypotheses are given as follows:

$$H_0 \colon \log(OR_S) \leq 0$$
$$H_A \colon \log(OR_S) > 0,$$

where OR_S is the odds ratio of a successful response between the AERx arm and the Insulin Aspart arm. The test is done according to the log odds ratio because the log odds ratio is approximately normal distributed, cf. Chapter 5. The p-value related to the test is given by:

$$p = P(X \geq x),$$

where x is the observed test statistic and X is a stochastic variable following the reference distribution. Like the test for the non-inferiority, the half of the significance level is used as a threshold for the test since the test is one-tailed.

It is worth noticing that a logistic regression based solely on the selection model approach or the pattern-mixture model approach is not very suitable for this estimand. The reason is that the subjects with missing HbA1c values at visit 13 will not be included in the analysis and hence not contribute to the results. Therefore, all analyses related to the composite estimand are based on a logistic regression combined with MI. The primary analysis and the sensitivity analyses are introduced in the following.

Primary Analysis for The Composite Estimand

The primary analysis is a logistic regression combined with MI under a assumption of MAR, cf. Section 3.4 and Chapter 5. The number of imputations is set to $m = 1000$. This analysis is referred to as MI^{MAR}_{LogReg}.

The missing data are imputed by applying Algorithm 3.4.1 separately on the two arms. The response variables of the imputation models are HbA1c at visit 7, 9, 11, 12 and 13, respectively. The HbA1c baseline value is included as a covariate in every imputation model. The imputation models for visit 9, 11, 12 and 13 do also include the HbA1c value from the previous visit(s) as covariate(s).

A logistic regression is used to analyse the 1000 complete datasets. The response variable of the logistic regression is a binary variable indicating *Success* if the HbA1c is below 7.0% at visit 13 and *use of escape therapy for at least 14 consecutive days* did not occur. Otherwise, the binary variable indicates *Failure*. The covariates in the logistic regression are treatment and the HbA1c baseline value. The theory from Section 5.2 is used to estimate the parameters in the logistic regression. The 1000 results are combined by applying Rubin's rules, cf. Section 3.4.

The parameter of interest is the odds ratio of *Success* between the AERx arm and the Insulin Aspart arm, which is denoted by OR_S. Remark, that the log odds ratio of

Success between the two treatment conditions is the parameter related to the treatment covariate in the logistic regression, cf. Chapter 5. Hence, an estimate of the odds ratio and a 95% confidence interval can be constructed by applying $\exp(\cdot)$ on the estimate and the boundaries of the 95% confidence interval related to $\log(OR_S)$. The inference for $\log(OR_S)$ is drawn based on the theory in Section 3.4.3.

Sensitivity Analysis I for The Composite Estimand

The first sensitivity analysis is a combination of a logistic regression and MI using the copy reference approach for the AERx arm. The number of imputations is also set to $m = 1000$. The purpose of this sensitivity analysis is to investigate how robust the results from the primary analysis are to deviations from the MAR assumption. This analysis is named MI^{CR}_{LogReg}.

The missing data are imputed by applying Algorithm 3.4.1 the same way as in the second sensitivity analysis in Section 6.1, but the imputation models are modified to be based on HbA1c instead of the change from baseline in HbA1c.

The logistic regression used to analyse the complete datasets is equivalent to the logistic regression used in the primary analysis in this section. The parameter of interest is OR_S. The inference for $\log(OR_S)$ and hence OR_S is drawn the same way as in the primary analysis related to the composite estimand.

Sensitivity Analysis II for The Composite Estimand

This analysis is almost equivalent to the previous sensitivity analysis. The only difference is that the missing data in the AERx arm are imputed according to the jump to reference approach instead of the copy reference approach. The jump to reference approach in this analysis is equivalent to the approach used in the third sensitivity analysis related to the treatment policy estimand. This analysis is called MI^{J2R}_{LogReg} and it investigates how robust the results from the primary analysis are to deviations from the MAR assumption.

Sensitivity Analysis III for The Composite Estimand

This sensitivity analysis is a tipping point analysis and it is a modified version of the tipping point analysis in Section 6.1. The imputation models are modified to be based on HbA1c instead of the change from baseline in HbA1c. Moreover, the 1000 complete datasets are analysed using a logistic regression similar to the one in the primary analysis in this section. The parameter of interest is OR_S and the inference is drawn as in the primary analysis related to the composite estimand.

Results Based on The Composite Estimand

Under this estimand, the proportion of missing data at each visit and the missing data pattern are the same as in Section 6.1. The `R`-code in Appendix C.7 is used to analyse the composite estimand. In Figure 6.6, the results are presented.

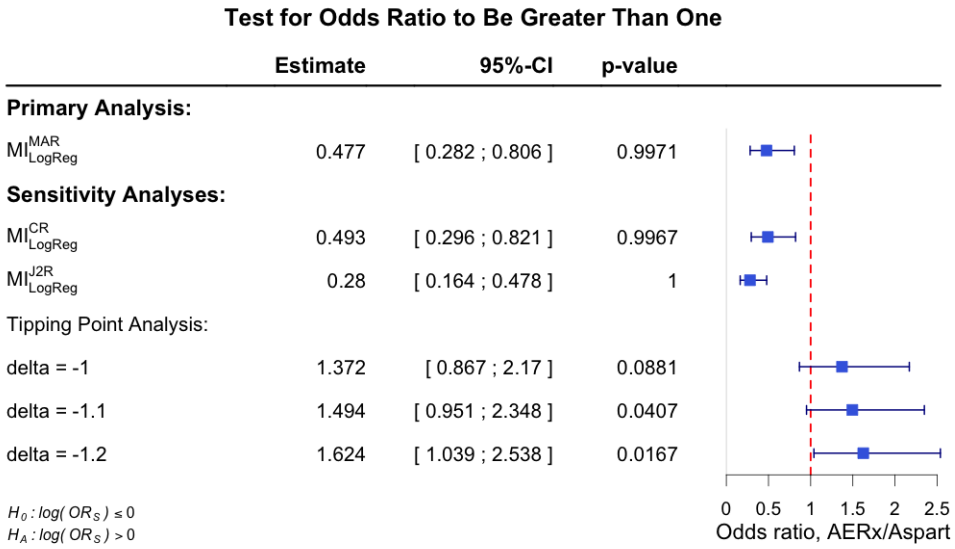

Figure 6.6: Results from the analyses related to the composite estimand.

In the primary analysis MI^{MAR}_{LogReg}, the null-hypothesis cannot be rejected. Thus, the conclusion based on the primary analysis is that there is not enough evidence to claim that the odds of success in the AERx arm is greater than the corresponding odds in the Insulin Aspart arm. This conclusion is supported by Figure 6.7 in which the means of the observed HbA1c values for each treatment arm are plotted.

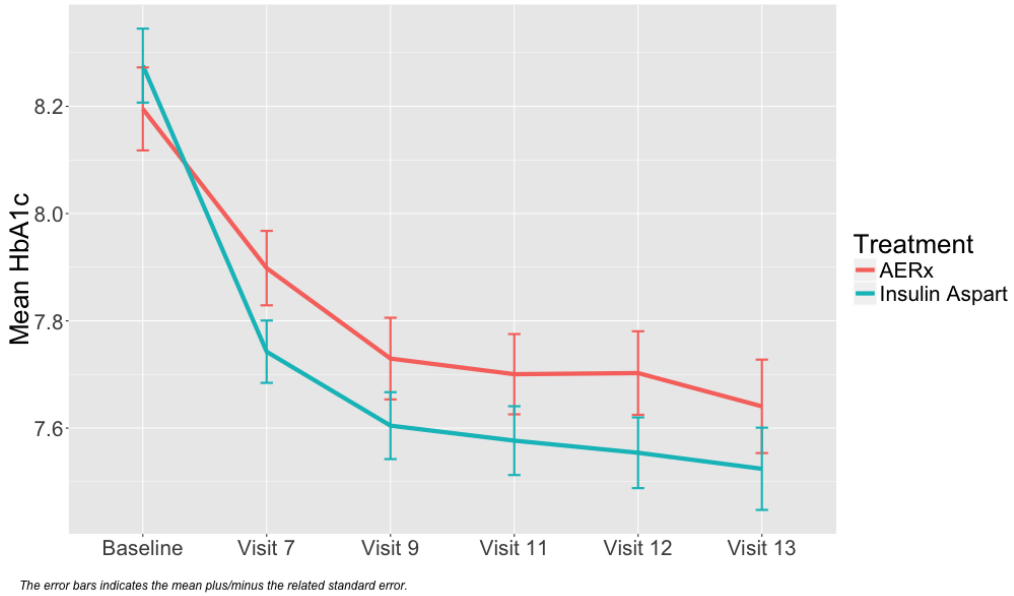

Figure 6.7: Mean of the observed HbA1c values at each visit for the two treatment arms.

Figure 6.7 is generated by the R-code in Appendix C.4. In the sensitivity analyses MI^{CR}_{LogReg} and MI^{J2R}_{LogReg}, the null-hypothesis cannot be rejected. The result of the tipping point analysis is that $\delta^* \in]-1.2\,;-1.1[$. The sensitivity analyses support the conclusion drawn from the primary analysis. Hence, the conclusion based on the composite estimand is that the odds of a successful response in the AERx arm cannot be rejected to be lower

than the corresponding odds in the Insulin Aspart arm.

6.4 While on Treatment Estimand

In this section, an estimand based on the while on treatment strategy is analysed. The estimand is called *the while on treatment estimand* and it is given by:

A. All exposed subjects with an observed baseline value of HbA1c.

B. Change from baseline in HbA1c at visit 13 or until the start of a period with *use of escape therapy for at least 14 consecutive days*, whichever comes first.

C. *Use of escape therapy for at least 14 consecutive days* is captured through the definition of the endpoint of interest. All other intercurrent events are ignored.

D. Difference in means between the AERx arm and the Insulin Aspart arm.

The while on treatment estimand targets the treatment difference between *AERx* and *Insulin Aspart* before an eventually occurrence of the intercurrent event *use of escape therapy for at least 14 consecutive days*. The specified treatment difference is denoted by β_T. The goal of the analysis related to the while on treatment estimand is to show non-inferiority, which is specified in Section 6.1.

Notice that measurements after an occurrence of the intercurrent event for a given subject should be excluded from the dataset considered under this estimand regardless of whether the measurements are observed or missing. A MMRM based solely on the selection model approach or the pattern-mixture model approach is not very suitable for the while on treatment estimand. The reason is that it is not possible to differ between the excluded data and the missing data when using a MMRM.

The primary analysis and the related sensitivity analyses are modified versions of the analyses in Section 6.3. One modification is that the measurements after an occurrence of the intercurrent event for a given subject are excluded from the dataset. Furthermore, the imputation models are modified to be based on the change from baseline in HbA1c instead of HbA1c. Another modification is that the model used to analyse is an ANCOVA instead of a logistic regression. The ANCOVA is specified as in Section 6.1. If the intercurrent event has occurred for a given subject, then the change from baseline in HbA1c at the latest visit before the occurrence is carried forward to visit 13 in order to estimate the specified treatment difference β_T with the ANCOVA.

Results Based on The While on Treatment Estimand

The structure of the dataset related to this estimand is a bit different compared to the datasets considered in the previous sections because some of the data are excluded as a consequence of the while on treatment strategy. The `R`-code in Appendix C.8 analyse the while on treatment estimand and the results are presented in Figure 6.8.

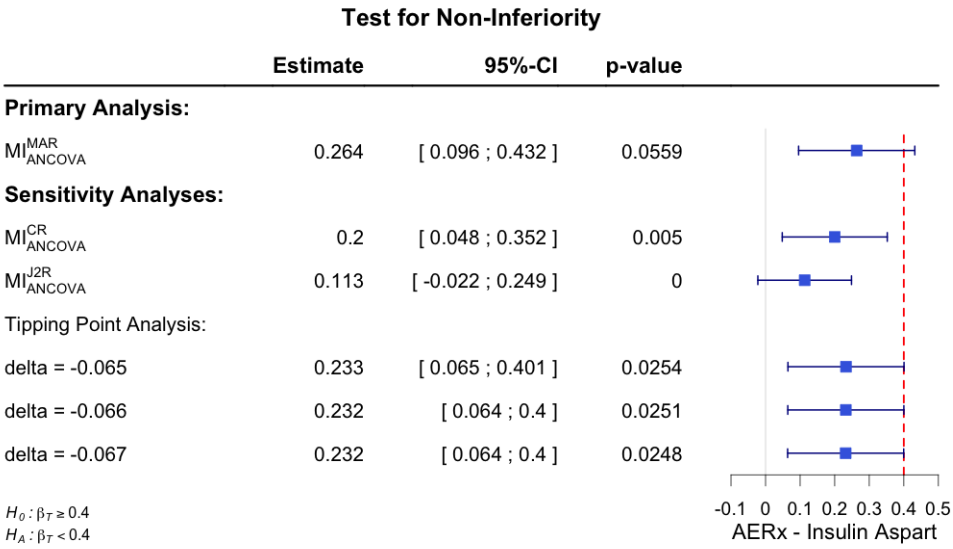

Figure 6.8: Results from the analyses related to the while on treatment estimand.

Based on the results from the primary analysis, the null hypothesis cannot be rejected and it is not possible to claim non-inferiority. As in Section 6.1 and Section 6.2, MI^{CR}_{ANCOVA} and MI^{J2R}_{ANCOVA} do not make sense since the analyses "reward" AERx when subjects discontinue before an eventually occurrence of the intercurrent event. The result of the tipping point analysis is that $\delta^* \approx -0.066$. The conclusion based on the while on treatment estimand is that non-inferiority cannot be shown when considering the treatment difference until an eventually occurrence of *use of escape therapy for at least 14 consecutive days*.

6.5 Results from the CTR

In this section, some of the results from the CTR are presented. An ANCOVA was used to estimate the treatment difference between *AERx* and *Insulin Aspart* at visit 13 under the intention-to-treat (ITT) principle. The ITT principle ignores all occurrence of any intercurrent event, i.e. it corresponds to use the treatment policy strategy to address all intercurrent events. The response variable of the ANCOVA was the change from baseline in HbA1c at visit 13. The covariates in the ANCOVA was treatment, the HbA1c baseline value and center. For subjects who had a missing HbA1c measurement at visit 13, the last observation carried forward (LOCF) method was used. In Figure 6.9, the results from the ANCOVA in the CTR are presented.

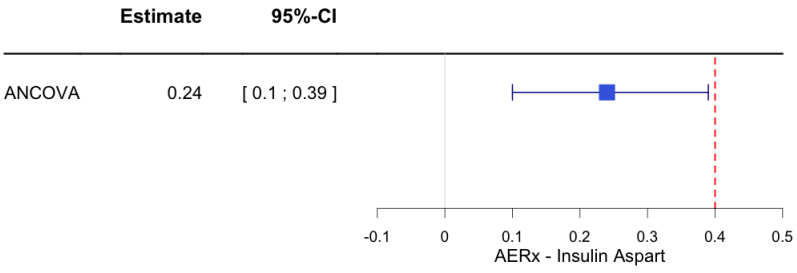

Figure 6.9: Results from the CTR.

Remark, a non-inferiority test as specified in Section 6.1 corresponds to test that the upper bound of the 95% confidence interval of the treatment difference is below 0.4 since the significance level of the one-tailed test is 0.025. Hence, it is possible to claim non-inferiority based on the results from the ANCOVA because the upper bound of the 95% confidence interval is below 0.4.

The ANCOVA from the CTR actually estimates the treatment difference between *AERx + escape therapy* and *Insulin Aspart + escape therapy* before subjects discontinue since it is based on the ITT principle and uses LOCF when a subject discontinues. Therefore, the ANCOVA actually corresponds to an estimand based on both the treatment policy strategy and the while on treatment strategy. The treatment policy strategy is used to address all intercurrent events except *discontinuation* for which the while on treatment strategy is used. The estimand could have been formulated as:

A. All exposed subjects with an observed baseline value of HbA1c.

B. Change from baseline in HbA1c at visit 13 or until *discontinuation*, whichever comes first.

C. *Discontinuation* is captured through the definition of the endpoint of interest. All other intercurrent events than *discontinuation* are ignored.

D. Difference in means between the AERx arm and the Insulin Aspart arm.

Discussion

In Chapter 6, it was illustrated how estimands could have been a part of the clinical trial NN1998-2076 in the light of the recent development of the estimand framework. Recall from Chapter 2 that a clinical trial often consists of several estimands, where one of the estimands is chosen to be the primary estimand and the others to be either secondary or supplementary estimands. None of the four estimands in Chapter 6 were chosen to be the primary estimand because the trial objective regarding HbA1c in the clinical trial did not focus on missing data and hence none of the estimands seemed to be more convenient than the others. In general, even though one estimand is chosen to be the primary estimand, the conclusions drawn from the secondary and supplementary estimands should not be ignored since it is recommended to draw an overall conclusion based on the results from all estimands. The primary estimand should be of greatest interest when drawing the overall conclusion and the secondary and supplementary estimands should be incorporated in order to get a "totality of evidence".

Based on the results from the estimands in Chapter 6, the overall conclusion is that AERx cannot be shown to be non-inferior to Insulin Aspart since none of the estimands gives a convincing result stating that AERx is non-inferior to Insulin Aspart. But it is worth noticing that even though AERx cannot be shown to be non-inferior to Insulin Aspart, it may be more convenient to inhale insulin instead of injecting it.

As described in Section 2.2, the treatment policy strategy is often preferred by the regulatory authorities. But this strategy does not always end up estimating the prespecified treatment difference as seen in Section 6.1, where the estimated treatment difference was expected to reflect the difference between *AERx + escape therapy* and *Insulin Aspart + escape therapy*. The estimated treatment difference did actually reflect the difference between *AERx + escape therapy* and *Insulin Aspart* since *use of escape therapy* did almost not occur in the Insulin Aspart arm. Therefore, it is recommended to be aware of what kind of difference an estimand based on the treatment policy strategy actually reflects before drawing any conclusions.

Compared to the treatment policy strategy, the hypothetical strategy does not cause similar issues since it estimates the treatment difference in a scenario where the intercurrent event does not occur. If the intercurrent event addressed with the hypothetical strategy however occurs for a major proportion of the subjects, then a greater proportion of the data should be considered as missing, cf. Section 2.2. This will cause an increased uncertainty about the related results and it is therefore recommended to reflect this in the conclusion.

In Chapter 6, the composite strategy was used to compare the odds of a successful response between the two treatment arms. In a clinical trial investigating efficacy, the composite strategy may not serve well in a primary estimand since it incorporates the intercurrent event in the endpoint of interest as described in Section 2.2. But an estimand based on the composite strategy may serve well as a secondary or supplementary estimand in order to give a summary of the response.

The principal stratum strategy could be of great interest to patients and physicians as described in Section 2.2. In order to use this strategy, it is necessary to get an indication of

the subjects in the principal stratum of interest through the trial design. In NN1998-2076 this was not possible and hence an estimand based on the principal stratum strategy was not formulated in Chapter 6. Even though the trial design gives an indication of subjects in the principal stratum of interest, the uncertainty about the related results might be increased compared to e.g. the treatment policy strategy and the hypothetical strategy since there is some uncertainty about the subjects assigned to the principal stratum.

An estimand based on the while on treatment strategy may not be of great interest in some situations because it estimates the treatment difference until an eventually occurrence of the intercurrent event, i.e. until a stochastic point in time. Remark, the estimated treatment difference based on the treatment policy strategy, the hypothetical strategy and the principal stratum strategy is at a fixed point in time, which is easier to interpret than a treatment difference at a stochastic point in time. But in some situations it may be more suitable to use the while on treatment strategy than the other strategies. An example could e.g. be a clinical trial where the intercurrent event *death* is likely to occur as described in Section 2.2. Payers may be quite interested in the treatment difference until the subjects eventually discontinue their treatment. Hence, the while on treatment strategy could be of great interest to payers, cf. Section 2.2. Especially, if the strategy is used to address *discontinuation*. The reason is that the payers only have to pay for the treatment as long as the subjects are on their treatment, i.e. until they discontinue.

The results and hence the conclusions in Chapter 6 might be quite uncertain because the proportion of missing data is quite large, see Figure 6.1 and Figure 6.4. In order to lower the proportion of missing data, it is important to ensure a high retention among the subjects in a trial as long as necessary according to the estimands, cf. Chapter 2. But at the same time it is also important to be aware of that the methods used to retain the subjects in the trial do not deviate too much from clinical practice in order to avoid bias, cf. [ICH, 2017, Section A.4].

It is worth remarking that the different analyses based on MI in Chapter 6 could have been influenced by the choice of seed. Another seed could have resulted in different conclusions. Therefore, it is recommended to investigate the results for different seeds. If the conclusion changes for different seeds, then the number of imputations should be increased in order to stabilize the results and hence eliminate the influence of seed, cf. [Lu, 2017].

In Chapter 6, the copy reference approach and the jump to reference approach were not very suitable since both approaches "rewarded" AERx when subjects discontinued the trial. In a trial where the treatment of interest is expected to be better than the alternative(s) it is compared against, both the copy reference and the jump to reference approach may serve well in a sensitivity analysis since both approaches in such cases are expected to "punish" the treatment of interest when subjects discontinue the trial. An alternative that is expected to be worse than the treatment of interest could e.g. be placebo. A trial in which the goal is to show that the treatment of interest is better than the alternative(s) is called *a superiority trial*. Hence, the copy reference and the jump to reference approach may be more suitable in a superiority trial than a non-inferiority trial.

A tipping point analysis cannot always be done properly as seen in Section 6.2, where the primary analysis and the tipping point analysis with $\delta = 0$ resulted in different conclusions. But even though this is the case, the tipping point analysis should not be excluded

since it shows that the results from the primary analysis are not very robust. Hence, a tipping point analysis serves well as a sensitivity analysis even though it cannot be done properly. Before drawing any conclusion based on a tipping point analysis it is recommended to be aware of whether or not the analysis is done properly because failing to do so can result in wrong conclusions. Remark, a tipping point analysis can always be done properly if the primary analysis is based on MI with the imputation models and analysis model being equivalent to the ones used in the tipping point analysis, i.e. the primary analysis is equivalent to the tipping point analysis for $\delta = 0$. This is the case in Section 6.3 and Section 6.4.

The tipping point analyses in Chapter 6 investigated how much the missing data in the AERx arm should have differed from the imputed values in order to give the opposite conclusion compared to the primary analysis. But it could also make sense to perform a tipping point analysis in the Insulin Aspart arm in order to investigate the assumptions about the missing data in this arm. Hence, a tipping point analysis in the comparator arm could also be of interest in some situations.

Even though it is generally not recommended to use LOCF in the pharmaceutical industry, cf. [National Research Council, 2010], the use of LOCF can be justified if the estimand incorporates the while on treatment strategy as in Section 6.5. But for all other strategies in Section 2.2 than the while on treatment strategy, it is not recommend to use LOCF since the method assumes that missing values for a given subject equal the last observed value for the same subject. This assumption is rarely plausible and hence LOCF often results in a biased estimate of the treatment difference, cf. [EMA, 2010, Section 6.3.1].

A. Results for REML

In this Appendix, some of the identities used to derive the REML log-likelihood in (4.7) are proved. The appendix is based on [Olofsson and Andersson, 2012] and [LaMotte, 2007].

Proposition A.1.1
Let $\mathbf{Y}$ be a stochastic vector and $\widetilde{\mathbf{Y}} = D\mathbf{Y}$, where D is an invertible $n \times n$ matrix. Then the densities of $\mathbf{Y}$ and $\widetilde{\mathbf{Y}}$ are related in the following way:

$$f_{\widetilde{\mathbf{Y}}}(\widetilde{\mathbf{y}}) = |D|^{-1} f_{\mathbf{Y}}\left(D^{-1}\widetilde{\mathbf{y}}\right) = |D|^{-1} f_{\mathbf{Y}}(\mathbf{y}).$$

Proposition A.1.2
Let $\mathbf{Y} \sim N_n(X\boldsymbol{\beta}, Z\Psi Z^\top + \Sigma)$ and A be a $n \times (n-p)$ matrix with columns spanning $M^\perp$, where $M = span(X)$ and p is the dimension of the vector of parameters $\boldsymbol{\beta}$. Furthermore, let:

$$D = \begin{bmatrix} A^\top \\ X^\top V^{-1} \end{bmatrix},$$

where $V = \mathrm{Var}[\mathbf{Y}]$. If D is invertible, then the following identities hold:

- $\log\left|A^\top V A\right| = \log|V| + \log\left|X^\top V^{-1} X\right| + \log\left|A^\top A\right| - \log\left|X^\top X\right|$.
- $\mathbf{y}^\top A\left(A^\top V A\right)^{-1} A^\top \mathbf{y} = \left(\mathbf{y} - X\widehat{\boldsymbol{\beta}}\right)^\top V^{-1}\left(\mathbf{y} - X\widehat{\boldsymbol{\beta}}\right)$.

Proof
Let $\widetilde{\mathbf{Y}} = D\mathbf{Y}$ and $\mathbf{0}_{k\times l}$ be a $k \times l$ matrix with 0 in all entries. Remark, $A^\top X = \mathbf{0}_{(n-p)\times p}$ and $A^\top X\boldsymbol{\beta} = \mathbf{0}_{(n-p)}$, where $\mathbf{0}_{(n-p)}$ denotes the $(n-p)$ dimensional zero vector. The mean of $\widetilde{\mathbf{Y}}$ is:

$$\mathrm{E}\left[\widetilde{\mathbf{Y}}\right] = \begin{bmatrix} \mathrm{E}\left[A^\top \mathbf{Y}\right] \\ \mathrm{E}\left[X^\top V^{-1}\mathbf{Y}\right] \end{bmatrix} = \begin{bmatrix} A^\top X\boldsymbol{\beta} \\ X^\top V^{-1} X\boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{0}_{(n-p)} \\ X^\top V^{-1} X\boldsymbol{\beta} \end{bmatrix}, \tag{A.1}$$

where $A^\top X\boldsymbol{\beta} = \mathbf{0}_{(n-p)}$ has been applied. The variance of $\widetilde{\mathbf{Y}}$ is:

$$\begin{aligned} \mathrm{Var}\left[\widetilde{\mathbf{Y}}\right] &= \mathrm{Var}[D\mathbf{Y}] \\ &= DVD^\top \\ &= \begin{bmatrix} A^\top V A & A^\top V V^{-1} X \\ X^\top V^{-1} V A & X^\top V^{-1} V V^{-1} X \end{bmatrix} \\ &= \begin{bmatrix} A^\top V A & \mathbf{0}_{(n-p)\times p} \\ \mathbf{0}_{p\times(n-p)} & X^\top V^{-1} X \end{bmatrix}, \end{aligned} \tag{A.2}$$

where $A^\top X = \mathbf{0}_{(n-p)\times p}$ has been applied. From (A.2) it follows that:

$$\log\left|DVD^\top\right| = \log\left|A^\top VA\right| + \log\left|X^\top V^{-1}X\right|. \tag{A.3}$$

Remark, $|DVD^\top| = |V||DD^\top|$ since $|B_1B_2| = |B_1||B_2|$ for $k\times k$ matrices B_1 and B_2. The determinant of $DD^\top$ is:

$$\begin{aligned}
\left|DD^\top\right| &= \left|\begin{bmatrix} A^\top \\ X^\top V^{-1}\end{bmatrix}\begin{bmatrix} A & V^{-1}X\end{bmatrix}\right| \\
&= \begin{vmatrix} A^\top A & A^\top V^{-1}X \\ X^\top V^{-1}A & X^\top V^{-1}V^{-1}X\end{vmatrix} \\
&= |A^\top A||X^\top V^{-2}X - X^\top V^{-1}A(A^\top A)^{-1}A^\top V^{-1}X| \\
&= |A^\top A||X^\top V^{-1}(I_n - A(A^\top A)^{-1}A^\top)V^{-1}X|,
\end{aligned} \tag{A.4}$$

which follows from $\begin{vmatrix} B_1 & B_2 \\ B_3 & B_4\end{vmatrix} = |B_1||B_4 - B_3B_1^{-1}B_2|$ for matrices B_1, B_2, B_3, B_4 of dimension $k\times k$, $k\times l$, $l\times k$ and $l\times l$, respectively. Notice, $P_A = A(A^\top A)^{-1}A^\top$ is the projection from $\mathbb{R}^n$ to $M^\perp$ and $P_X = X(X^\top X)^{-1}X^\top$ is the projection from $\mathbb{R}^n$ to M. Hence, $P_X = I_n - P_A$. This implies that (A.4) can be written as:

$$\begin{aligned}
\left|DD^\top\right| &= \left|A^\top A\right|\left|X^\top V^{-1}(I_n - P_A)V^{-1}X\right| \\
&= \left|A^\top A\right|\left|X^\top V^{-1}P_XV^{-1}X\right| \\
&= \left|A^\top A\right|\left|X^\top V^{-1}X(X^\top X)^{-1}X^\top V^{-1}X\right| \\
&= \left|A^\top A\right|\left|X^\top V^{-1}X\right|\left|(X^\top X)^{-1}\right|\left|X^\top V^{-1}X\right| \\
&= \left|A^\top A\right|\left|X^\top X\right|^{-1}\left|X^\top V^{-1}X\right|^2,
\end{aligned} \tag{A.5}$$

where it has been applied that $|B_1B_2| = |B_1||B_2|$ for $k\times k$ matrices. From (A.5) it follows that (A.3) is equivalent to:

$$\log|V| + \log\left|A^\top A\right| + 2\log\left|X^\top V^{-1}X\right| - \log\left|X^\top X\right| = \log\left|A^\top VA\right| + \log\left|X^\top V^{-1}X\right|,$$

which proves the first identity in Proposition A.1.2.

Since $\mathbf{Y}$ is normal distributed, $\widetilde{\mathbf{Y}}$ is also normal distributed. The density of $\widetilde{\mathbf{Y}}$ can be written as:

$$f_{\widetilde{\mathbf{Y}}}(\widetilde{\mathbf{y}}) = (2\pi)^{-n/2}\left|\operatorname{Var}\left[\widetilde{\mathbf{Y}}\right]\right|^{-1/2}\exp\left(-\frac{1}{2}\left(\widetilde{\mathbf{y}} - \mathrm{E}[\widetilde{\mathbf{Y}}]\right)^\top \operatorname{Var}\left[\widetilde{\mathbf{Y}}\right]^{-1}\left(\widetilde{\mathbf{y}} - \mathrm{E}[\widetilde{\mathbf{Y}}]\right)\right).$$

Based on Proposition A.1.1, the densities of $\widetilde{\mathbf{Y}}$ and $\mathbf{Y}$ are related in the following way:

$$\begin{aligned}
&(2\pi)^{-n/2}\left|DVD^\top\right|^{-1/2}\exp\left(-\frac{1}{2}\left(\widetilde{\mathbf{y}} - \mathrm{E}[\widetilde{\mathbf{Y}}]\right)^\top\left(DVD^\top\right)^{-1}\left(\widetilde{\mathbf{y}} - \mathrm{E}[\widetilde{\mathbf{Y}}]\right)\right) \\
&\quad = |D|^{-1}(2\pi)^{-n/2}|V|^{-1/2}\exp\left(-\frac{1}{2}\left(\mathbf{y} - \mathrm{E}[\mathbf{Y}]\right)^\top V^{-1}\left(\mathbf{y} - \mathrm{E}[\mathbf{Y}]\right)\right),
\end{aligned} \tag{A.6}$$

where $\operatorname{Var}[\mathbf{Y}] = V$ and $\operatorname{Var}\left[\widetilde{\mathbf{Y}}\right] = DVD^\top$ have been applied. Remark, $|DVD^\top|^{-1/2} = |D|^{-1}|V|^{-1/2}$. This implies that (A.6) is equivalent to:

$$\left(\widetilde{\mathbf{y}} - \mathrm{E}[\widetilde{\mathbf{Y}}]\right)^\top\left(DVD^\top\right)^{-1}\left(\widetilde{\mathbf{y}} - \mathrm{E}[\widetilde{\mathbf{Y}}]\right) = \left(\mathbf{y} - \mathrm{E}[\mathbf{Y}]\right)^\top V^{-1}\left(\mathbf{y} - \mathrm{E}[\mathbf{Y}]\right). \tag{A.7}$$

By applying (A.1), (A.2) and $\widetilde{\mathbf{y}} = D\mathbf{y}$, the left hand side of (A.7) can be written as:

$$\mathbf{y}^\top A(A^\top V A)^{-1}A^\top \mathbf{y} + \left(X^\top V^{-1}(\mathbf{y} - X\boldsymbol{\beta})\right)^\top \left(X^\top V^{-1}X\right)^{-1}\left(X^\top V^{-1}(\mathbf{y} - X\boldsymbol{\beta})\right). \quad \text{(A.8)}$$

Notice, when the parameter $\boldsymbol{\beta}$ is replaced by $\widehat{\boldsymbol{\beta}} = (X^\top V^{-1}X)^{-1}X^\top V^{-1}\mathbf{y}$ the second term in (A.8) cancels out. This implies:

$$\mathbf{y}^\top A\left(A^\top V A\right)^{-1}A^\top \mathbf{y} = \left(\mathbf{y} - X\widehat{\boldsymbol{\beta}}\right)^\top V^{-1}\left(\mathbf{y} - X\widehat{\boldsymbol{\beta}}\right),$$

which follows from (A.7) and (A.8). ■

B. The Delta Method

In this appendix, the Delta method is presented. The appendix is based on [Bishop et al., 2007]. Let $\mathbf{0}_k$ denote the k dimensional zero vector.

Proposition B.1.1
Let n be the number of observations and $\widehat{\boldsymbol{\beta}}$ be a p dimensional vector of estimators for which the following property holds:

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right) \xrightarrow{d} N_p\left(\mathbf{0}_p, \Sigma\right),$$

where $\boldsymbol{\beta}^*$ consists of the true parameter values and $\xrightarrow{d}$ denotes *convergence in distribution*. Furthermore, let $g\colon \mathbb{R}^p \to \mathbb{R}^q$ be a function which is differentiable at $\boldsymbol{\beta} = \boldsymbol{\beta}^*$. Then:

$$\sqrt{n}\left(g\left(\widehat{\boldsymbol{\beta}}\right) - g(\boldsymbol{\beta}^*)\right) \xrightarrow{d} N_q\left(\mathbf{0}_q, J_g(\boldsymbol{\beta}^*)\ \Sigma\ J_g(\boldsymbol{\beta}^*)^\top\right),$$

where $J_g(\boldsymbol{\beta}^*)$ denotes the Jacobian matrix of g evaluated at $\boldsymbol{\beta} = \boldsymbol{\beta}^*$.

Proof
Only the mean and variance of $\sqrt{n}\left(g\left(\widehat{\boldsymbol{\beta}}\right) - g(\boldsymbol{\beta}^*)\right)$ are proved in this thesis. Consider the following first order Taylor expansion:

$$g\left(\widehat{\boldsymbol{\beta}}\right) \approx g(\boldsymbol{\beta}^*) + J_g(\boldsymbol{\beta}^*)\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right). \tag{B.1}$$

Using (B.1), the mean value of $\sqrt{n}\left(g\left(\widehat{\boldsymbol{\beta}}\right) - g(\boldsymbol{\beta}^*)\right)$ can be approximated by:

$$\begin{aligned}
\mathrm{E}\left[\sqrt{n}\left(g\left(\widehat{\boldsymbol{\beta}}\right) - g(\boldsymbol{\beta}^*)\right)\right] &\approx \mathrm{E}\left[\sqrt{n}\ J_g(\boldsymbol{\beta}^*)\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right)\right] \\
&= J_g(\boldsymbol{\beta}^*)\mathrm{E}\left[\sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right)\right] \\
&= J_g(\boldsymbol{\beta}^*)\ \mathbf{0}_p \\
&= \mathbf{0}_q,
\end{aligned}$$

where it has been applied that $J_g(\boldsymbol{\beta}^*)$ is fixed and $\mathrm{E}\left[\sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right)\right] = \mathbf{0}_p$. The variance of $\sqrt{n}\left(g\left(\widehat{\boldsymbol{\beta}}\right) - g(\boldsymbol{\beta}^*)\right)$ can be approximated by:

$$\begin{aligned}
\mathrm{Var}\left[\sqrt{n}\left(g\left(\widehat{\boldsymbol{\beta}}\right) - g(\boldsymbol{\beta}^*)\right)\right] &\approx \mathrm{Var}\left[\sqrt{n}\ J_g(\boldsymbol{\beta}^*)\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right)\right] \\
&= J_g(\boldsymbol{\beta}^*)\ \mathrm{Var}\left[\sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right)\right] J_g(\boldsymbol{\beta}^*)^\top \\
&= J_g(\boldsymbol{\beta}^*)\ \Sigma\ J_g(\boldsymbol{\beta}^*)^\top
\end{aligned}$$

where (B.1), the fact that $J_g(\boldsymbol{\beta}^*)$ is fixed and $\mathrm{Var}[AX] = A\mathrm{Var}[X]A^\top$ have been applied.■

C. R-Code

C.1 R-Code: Datasets

```
#HbA1c datasets
library(haven)
library(reshape2)

#Load datasets
setwd("/Volumes/aerx-2076/derived")

#Subject info
subjinfo <- read_sas("subjinfo.sas7bdat")

#Escape therapy
escape <- read_sas("escape.sas7bdat")

#Escape therapy dataset only including subjects
#who use escape therapy for at least 14 consecutive days
escape_14 <- escape[(escape$ESP_DUR>=14 & !is.na(escape$ESP_
   DUR)),]

#Subjects without any stop date for use of escape therapy is
   added
escape_14 <- rbind(escape_14,escape[escape$ES_CONT=="Y",])

#Escape therapy dataset is ordered according to the start
   date
escape_14 <- escape_14[order(escape_14$ES_START),]

#Only first period for each subject is needed
escape_14 <- escape_14[!duplicated(escape_14$SUBJ_ID),]

#Efficacy
efficacy <- read_sas("hba1c_fpg.sas7bdat")

#Efficacy dataset only including HbA1c measurements
HbA1c <- efficacy[which(efficacy$LPARM=="HBA1C_BLOOD" &
   efficacy$EXPOSED=="Yes" & !is.na(efficacy$BASELINE)),]

visits <- c("VISIT 1", "VISIT 7", "VISIT 9", "VISIT 11", "
   VISIT 12", "VISIT 13")
HbA1c <- HbA1c[HbA1c$VISIT %in% visits,]
HbA1c$VISIT <- gsub("VISIT ", "", HbA1c$VISIT)

#Merging of HbA1c and subject info
```

```
38 HbA1c_merge <- merge(HbA1c, subjinfo[,c("SUBJ_ID", "SITE_ID",
       "TRT_DAYS", "DI_RE_ID")],
39                              by="SUBJ_ID", all.x=TRUE)
40
41 columns <- c("SUBJ_ID", "TREATMEN", "SITE_ID", "SEX_ID", "AGE
       ", "TRT_DAYS", "DI_RE_ID",
42                   "VISIT", "VALUE", "COLL_DT")
43
44 #Selecting columns and renaming some variables
45 HbA1c_merge <- HbA1c_merge[,columns]
46 names(HbA1c_merge)[names(HbA1c_merge) == "TREATMEN"] <- "
       TREAT"
47 names(HbA1c_merge)[names(HbA1c_merge) == "DI_RE_ID"] <- "
       WITHDRAWAL_REASON"
48 names(HbA1c_merge)[names(HbA1c_merge) == "VALUE"] <- "HbA1c"
49
50 #Columns as factors
51 factors <- c("SUBJ_ID", "TREAT", "SITE_ID", "SEX_ID", "
       WITHDRAWAL_REASON")
52 HbA1c_merge[,factors] <- lapply(HbA1c_merge[,factors], as.
       factor)
53
54 #relevel TREAT
55 HbA1c_merge$TREAT <- relevel(HbA1c_merge$TREAT, "Insulin
       Aspart")
56
57 #Wide format of HbA1c
58 HbA1c_reshape <- reshape(HbA1c_merge,
59                          idvar=c("SUBJ_ID", "TREAT", "SITE_ID
                             ", "SEX_ID", "AGE", "TRT_DAYS", "
                             WITHDRAWAL_REASON"),
60                          timevar = "VISIT", direction = "wide
                             ")
61
62 names(HbA1c_reshape) <- gsub("HbA1c.", "", names(HbA1c_
       reshape))
63
64 columns_wide <- c("SUBJ_ID", "TREAT", "SITE_ID", "SEX_ID", "
       AGE", "TRT_DAYS", "WITHDRAWAL_REASON",
65               "1", "7", "9", "11", "12", "13",
66               "COLL_DT.1", "COLL_DT.7", "COLL_DT.9", "COLL_DT
                    .11", "COLL_DT.12", "COLL_DT.13")
67
68 #Selecting columns and renaming some variables
69 HbA1c_wide <- HbA1c_reshape[,columns_wide]
70 names(HbA1c_wide)[names(HbA1c_wide) == "1"] <- "BASELINE"
71 names(HbA1c_wide)[names(HbA1c_wide) == "COLL_DT.1"] <- "COLL_
       DT.BL"
72
```

```
73 #Add column with use of escape therapy start date
74 HbA1c_wide <- merge(HbA1c_wide, escape_14[,c("SUBJ_ID","ES_
     START")], all.x=TRUE)
75
76 visits <- c("VISIT_7", "VISIT_9", "VISIT_11", "VISIT_12", "
     VISIT_13")
77 coll_dt <- c("COLL_DT.7", "COLL_DT.9", "COLL_DT.11", "COLL_DT
     .12", "COLL_DT.13")
78 escape_info <- c("ES_14_DAYS_VISIT_7", "ES_14_DAYS_VISIT_9",
     "ES_14_DAYS_VISIT_11", "ES_14_DAYS_VISIT_12", "ES_14_DAYS_
     VISIT_13")
79
80 #A column is added for each visit which equals "Yes" if the
     intercurrent event
81 #"use of escape therapy for at least 14 days" has occurred
82 HbA1c_wide[,escape_info] <- "No"
83
84 for(i in 1:length(visits)){
85   #An indication of the subjects for which the intercurrent
       event has occurred at visit i
86   escape_use <- which(difftime(strptime(HbA1c_wide[,coll_dt[i
       ]], format = "%Y-%m-%d", tz="GMT"),
87                                strptime(HbA1c_wide$ES_START,
                                     format = "%Y-%m-%d", tz="
                                    GMT"),units="days")>=0)
88
89   #Use of escape therapy is set to "Yes" for all subsequent
       visits
90   HbA1c_wide[escape_use, escape_info[i:length(escape_info)]]
       <- "Yes"
91 }
92
93 #Creating a long format of the data
94 HbA1c_long_visits <- melt(HbA1c_wide[,!(names(HbA1c_wide) %in
     % c(coll_dt, escape_info))], id.vars=c("SUBJ_ID", "TREAT",
      "SITE_ID", "SEX_ID", "AGE", "TRT_DAYS", "WITHDRAWAL_
     REASON", "BASELINE", "COLL_DT.BL", "ES_START"),
95                   value.name = "HbA1c", variable.name = "
                        VISIT")
96
97 HbA1c_long_coll_dt <- melt(HbA1c_wide[,names(HbA1c_wide) %in%
      c("SUBJ_ID", coll_dt)], id.vars="SUBJ_ID",
98                   value.name = "COLL_DT", variable.name = "
                        VISIT")
99 HbA1c_long_coll_dt$VISIT <- gsub("COLL_DT.", "", HbA1c_long_
     coll_dt$VISIT)
100
101 HbA1c_long_escape_info <- melt(HbA1c_wide[,names(HbA1c_wide)
     %in% c("SUBJ_ID", escape_info)], id.vars="SUBJ_ID",
```

```
102                              value.name = "ES_14_DAYS",
                                    variable.name = "VISIT")
103 HbA1c_long_escape_info$VISIT <- gsub("ES_14_DAYS_VISIT_", "",
        HbA1c_long_escape_info$VISIT)
104
105 HbA1c_long_merge <- merge(HbA1c_long_visits, HbA1c_long_coll_
    dt, by=c("SUBJ_ID","VISIT"))
106 HbA1c_long <- merge(HbA1c_long_merge, HbA1c_long_escape_info,
        by=c("SUBJ_ID","VISIT"))
107
108 #The data is ordered(long format)
109 HbA1c_long <- HbA1c_long[order(HbA1c_long$SUBJ_ID,HbA1c_long$
    VISIT),]
110 HbA1c_long$VISIT <- factor(HbA1c_long$VISIT, levels=c("7","9"
    ,"11","12","13"))
111
112 #Renaming the the visit variables in the wide format
113 names(HbA1c_wide)[names(HbA1c_wide) %in% c("7","9","11","12",
    "13")] <- c("VISIT_7", "VISIT_9", "VISIT_11", "VISIT_12",
    "VISIT_13")
114
115 #Datasets consisting of the change from baseline in HbA1c are
        generated
116 change_HbA1c_wide <- HbA1c_wide
117 change_HbA1c_wide[,visits] <- change_HbA1c_wide[,visits] -
    change_HbA1c_wide$BASELINE
118
119 change_HbA1c_long <- HbA1c_long
120 change_HbA1c_long$HbA1c <- change_HbA1c_long$HbA1c - change_
    HbA1c_long$BASELINE
121
122 #Remove row.names
123 row.names(HbA1c_wide) <- NULL
124 row.names(change_HbA1c_wide) <- NULL
125 row.names(HbA1c_long) <- NULL
126 row.names(change_HbA1c_long) <- NULL
127
128 #Save datasets
129 setwd("/Volumes/aerx-2076")
130 save(HbA1c_wide,file="HbA1c_wide.RData")
131 save(change_HbA1c_wide,file="change_HbA1c_wide.RData")
132 save(HbA1c_long,file="HbA1c_long.RData")
133 save(change_HbA1c_long,file="change_HbA1c_long.RData")
```

C.2 R-Code: Copy Reference

```
#Multiple Imputation using the copy reference approach
library(mice)

#Function to extract parameters from imputation step in mice
mice.impute.norm.extract <- function (y, ry, x, ...){
  x <- cbind(1, as.matrix(x))
  parm <- .norm.draw(y, ry, x, ...)
  beta_imp_models[[length(beta_imp_models)+1]] <<- parm$beta
  sigma_imp_models[[length(sigma_imp_models)+1]] <<- parm$
     sigma
  return(x[!ry, ] %*% parm$beta + rnorm(sum(!ry)) * parm$
     sigma)
}

#Function to impute missing data using the copy reference
   approach
MI_copy_ref <- function(data, trt.var, copy.trt ,ref.trt,
    meth, pred, post, vis, where.copy=NULL, where.ref=NULL, M,
     seed){

  if(!all(meth %in% c("","norm.extract"))){
    stop("The only method allowed is norm.extract")
  }

  data_ref <- data[which(data[,trt.var]==ref.trt),]
  data_copy <- data[which(data[,trt.var]==copy.trt),]

  if(is.null(where.ref)){
    where.ref <- is.na(data_ref)
  }
  if(is.null(where.copy)){
    where.copy <- is.na(data_copy)
  }

  beta_imp_models <<- list()
  sigma_imp_models <<- list()

  imp_ref <- mice(data_ref, meth = meth, pred = pred, post =
     post, vis = vis, where = where.ref, m = M, maxit = 1,
     seed = seed)

  imp_copy <- rep(list(cbind(".imp" = 0, data_copy)),(M+1))
  vis.sequence <- names(vis)
  for(j in 1:M){
    cat(c("Imputation ",j,": "))
    for (i in 1:length(vis.sequence)) {
      cat(c(" ",vis.sequence[i]))
```

```
41        data_imp <- imp_copy[[(j+1)]][,-1]
42        p <- mice:::padModel(data = data_imp, method = meth,
              predictorMatrix = pred, visitSequence = vis,
43                             form = vector("character", length
                                   = ncol(data)), post = post,
                                   nvar=ncol(data))
44        predictors <- p$predictorMatrix[vis.sequence[i], ] == 1
45        x <- cbind(1,as.matrix(p$data[, predictors, drop=FALSE
              ]))
46
47        where.copy_visit <- which(where.copy[, vis.sequence[i
              ]])
48        imp_copy[[(j+1)]][where.copy_visit, vis.sequence[i]] <-
              x[where.copy_visit,] %*% beta_imp_models[[i+length(
              vis.sequence)*(j-1)]] + rnorm(length(where.copy_
              visit)) * sigma_imp_models[[i+length(vis.sequence)*(
              j-1)]]
49      }
50      imp_copy[[(j+1)]][,".imp"] <- j
51      cat("\n")
52    }
53    imp_long_copy <- do.call(rbind,imp_copy)
54
55    imp_long_ref <- complete(imp_ref, action="long",incl=TRUE)
56
57    #Column 2 (".id") is removed
58    imp_long <- rbind(imp_long_ref[,-2],imp_long_copy)
59
60    imp_long <- imp_long[order(imp_long$.imp,imp_long$SUBJ_ID)
          ,]
61
62    #Transform imp_long into a mids object
63    imp_long_mids <- as.mids(imp_long)
64
65    return(imp_long_mids)
66  }
```

C.3 R-Code: Jump to Reference

```
#Multiple Imputation using the jump to reference approach
library(mice)

#Function to extract parameters from imputation step in mice
mice.impute.norm.extract <- function (y, ry, x, ...){
  x <- cbind(1, as.matrix(x))
  parm <- .norm.draw(y, ry, x, ...)
  beta_imp_models[[length(beta_imp_models)+1]] <<- parm$beta
  sigma_imp_models[[length(sigma_imp_models)+1]] <<- parm$
     sigma
  return(x[!ry, ] %*% parm$beta + rnorm(sum(!ry)) * parm$
     sigma)
}

#Function to impute missing data using the jump to reference
   approach
MI_jump_to_ref <- function(data, trt.var, jump.trt, ref.trt,
    meth, pred, post, vis, where.jump=NULL, where.ref=NULL, M,
     seed){

  if(!all(meth %in% c("","norm.extract"))){
    stop("The only method allowed is norm.extract")
  }

  data_ref <- data[which(data[,trt.var]==ref.trt),]
  data_jump <- data[which(data[,trt.var]==jump.trt),]

  if(is.null(where.ref)){
    where.ref <- is.na(data_ref)
  }
  if(is.null(where.jump)){
    where.jump <- is.na(data_jump)
  }

  beta_imp_models <<- list()
  sigma_imp_models <<- list()

  imp_ref <- mice(data_ref, meth = meth, pred = pred, post =
     post, vis = vis, where = where.ref, m = M, maxit = 1,
     seed = seed)

  imp_jump <- rep(list(cbind(".imp" = 0, data_jump)),(M+1))
  vis.sequence <- names(vis)
  p <- mice:::padModel(data = data_jump, method = meth,
     predictorMatrix = pred, visitSequence = vis,
                       form = vector("character", length =
                          ncol(data)), post = post, nvar=ncol
```

```
                              (data))
39  factor.var <- rownames(p$categories[which(p$categories[,"is
      .factor"]==TRUE),])
40  for(j in 1:M){
41    cat(c("Imputation ",j,": "))
42    imp_data_ref <- complete(imp_ref,j)
43
44    for (i in 1:length(vis.sequence)) {
45      cat(c(" ",vis.sequence[i]))
46      where.jump_visit <- where.jump[, vis.sequence[i]]
47
48      if(identical(factor.var,character(0))){
49        subgroup_vis_ref <- imp_data_ref[,vis.sequence[i]]
50        mean_subgroup_vis_ref <- mean(subgroup_vis_ref)
51        imp_jump[[(j+1)]][where.jump_visit, vis.sequence[i]]
            <- rep(mean_subgroup_vis_ref,sum(where.jump_visit)
            ) + rnorm(sum(where.jump_visit)) * sigma_imp_
            models[[i+length(vis.sequence)*(j-1)]]
52      }
53      else{
54        subgroup_vis_ref <- split(imp_data_ref[,vis.sequence[
            i]],interaction(imp_data_ref[,factor.var]))
55        mean_subgroup_vis_ref <- lapply(subgroup_vis_ref,mean
            )
56        subgroup_indicators_jump <- model.matrix(~-1+
            interaction(data_jump[, factor.var]))
57        colnames(subgroup_indicators_jump) <- gsub("
            interaction(data_jump[, factor.var])","",colnames(
            subgroup_indicators_jump),fixed=T)
58
59        for(k in 1:length(mean_subgroup_vis_ref)){
60          subgroup_indicators_jump_where.jump <- which((
              subgroup_indicators_jump[, names(mean_subgroup_
              vis_ref)[k]]*where.jump_visit) == 1)
61          imp_jump[[(j+1)]][subgroup_indicators_jump_where.
              jump, vis.sequence[i]] <- rep(mean_subgroup_vis_
              ref[[k]],length(subgroup_indicators_jump_where.
              jump)) + rnorm(length(subgroup_indicators_jump_
              where.jump)) * sigma_imp_models[[i+length(vis.
              sequence)*(j-1)]]
62        }
63      }
64
65      if(any(is.na(imp_jump[[(j+1)]][where.jump[,vis.sequence
          [i]], vis.sequence[i]]))){
66        stop("The subgroup does not match across treatments")
67      }
68    }
69    imp_jump[[(j+1)]][,".imp"] <- j
```

```
70      cat("\n")
71    }
72    imp_long_jump <- do.call(rbind,imp_jump)
73
74    imp_long_ref <- complete(imp_ref, action="long",incl=TRUE)
75
76    #Column 2 (".id") is removed
77    imp_long <- rbind(imp_long_ref[,-2],imp_long_jump)
78
79    imp_long <- imp_long[order(imp_long$.imp,imp_long$SUBJ_ID)
          ,]
80
81    #Transform imp_long into a mids object
82    imp_long_mids <- as.mids(imp_long)
83
84    return(imp_long_mids)
85  }
```

C.4 R-Code: Mean Plots

```
#Mean plots
library(ggplot2)
library(grid)

#Load datasets
setwd("/Volumes/aerx-2076")
HbA1c_wide <- get(load("HbA1c_wide.RData"))
change_HbA1c_wide <- get(load("change_HbA1c_wide.RData"))

#Directory for plots
setwd("/Users/Kristoffer/Dropbox/Speciale/Latex/figures")

#######################################
####  Mean change from baseline   ####
#######################################

#Visits for change from baseline
visits <- c("VISIT_7","VISIT_9","VISIT_11","VISIT_12","VISIT_
   13")

#Mean change from baseline for each treatment across visits
mean_change_visit_AERx <- as.matrix(sapply(change_HbA1c_wide[
   change_HbA1c_wide$TREAT == "AERx",visits],mean,na.rm=TRUE)
   )
mean_change_visit_Asp <- as.matrix(sapply(change_HbA1c_wide[
   change_HbA1c_wide$TREAT == "Insulin Aspart",visits],mean,
   na.rm=TRUE))

#Standard errors related to the means of the change from
   baseline
se_mean_change_visit_AERx <- as.matrix(sapply(change_HbA1c_
   wide[change_HbA1c_wide$TREAT == "AERx",visits],function(x)
    sqrt(var(x,na.rm=TRUE)/sum(!is.na(x)))))
se_mean_change_visit_Asp <- as.matrix(sapply(change_HbA1c_
   wide[change_HbA1c_wide$TREAT == "Insulin Aspart",visits],
   function(x) sqrt(var(x,na.rm=TRUE)/sum(!is.na(x)))))

#Data frame
mean_change_visit <- as.data.frame(cbind(rbind(unname(mean_
   change_visit_AERx), unname(mean_change_visit_Asp)), rbind(
   unname(se_mean_change_visit_AERx),unname(se_mean_change_
   visit_Asp))))
colnames(mean_change_visit) <- c("mean_change_HbA1c","se_mean
   _change_HbA1c")
mean_change_visit$TREAT <- as.factor(c(rep("AERx",length(
   visits)),rep("Insulin Aspart",length(visits))))
```

```
32 mean_change_visit$VISIT <- factor(c(rep(visits,2)),levels=
       visits)
33
34 #Plot of mean change from baseline in HbA1c for each
       treatment
35 mean.change.baseline.treat <- ggplot(mean_change_visit, aes(x
       = VISIT, y = mean_change_HbA1c, ymin = mean_change_HbA1c-
       se_mean_change_HbA1c, ymax = mean_change_HbA1c+se_mean_
       change_HbA1c, colour = TREAT, group=TREAT)) +
36   geom_line(size=2) +
37   geom_errorbar(size=1,width=0.1) +
38   scale_x_discrete(labels=c("VISIT_7" = "Visit 7", "VISIT_9"
       = "Visit 9", "VISIT_11" = "Visit 11", "VISIT_12" = "
       Visit 12", "VISIT_13" = "Visit 13")) +
39   labs(y="Mean change from baseline in HbA1c", x="") +
40   guides(colour=guide_legend(title="Treatment")) + ylim(c
       (-1,0)) +
41   theme(text = element_text(size=26), axis.title.x =element_
       text(margin=margin(0,0,30,0)))
42
43 #Footnote
44 footnote <- "The error bars indicates the mean plus/minus the
       related standard error."
45
46 #Export, plot of mean change from baseline in HbA1c for each
       treatment
47 fileName.change.baseline.treat <- "mean_change_from_baseline_
       treat.png"
48 png(fileName.change.baseline.treat, width=1000, height=600,
       units="px", pointsize = 16)
49 mean.change.baseline.treat
50 grid.text(footnote,unit(.015, 'npc'),unit(.04, 'npc'), just=c
       ("left"), gp=gpar(fontsize=12, font = 3))
51 dev.off()
52
53 ######################################
54 ####          Mean HbA1c          ####
55 ######################################
56
57 #Visits for HbA1c
58 visits <- c("BASELINE","VISIT_7","VISIT_9","VISIT_11","VISIT_
       12","VISIT_13")
59
60 #Mean HbA1c for each treatment across visits
61 mean_visit_AERx <- as.matrix(sapply(HbA1c_wide[HbA1c_wide$
       TREAT == "AERx",visits],mean,na.rm=TRUE))
62 mean_visit_Asp <- as.matrix(sapply(HbA1c_wide[HbA1c_wide$
       TREAT == "Insulin Aspart",visits],mean,na.rm=TRUE))
63
```

```
64 #Standard errors related to the means of HbA1c
65 se_mean_visit_AERx <- as.matrix(sapply(HbA1c_wide[HbA1c_wide$
     TREAT == "AERx",visits],function(x) sqrt(var(x,na.rm=TRUE)
     /sum(!is.na(x)))))
66 se_mean_visit_Asp <- as.matrix(sapply(HbA1c_wide[HbA1c_wide$
     TREAT == "Insulin Aspart",visits],function(x) sqrt(var(x,
     na.rm=TRUE)/sum(!is.na(x)))))
67
68 #Data frame
69 mean_visit <- as.data.frame(cbind(rbind(unname(mean_visit_
     AERx), unname(mean_visit_Asp)), rbind(unname(se_mean_visit
     _AERx),unname(se_mean_visit_Asp))))
70 colnames(mean_visit) <- c("mean_HbA1c","se_mean_HbA1c")
71 mean_visit$TREAT <- as.factor(c(rep("AERx",length(visits)),
     rep("Insulin Aspart",length(visits))))
72 mean_visit$VISIT <- factor(c(rep(visits,2)),levels=visits)
73
74 #Plot of mean HbA1c for each treatment
75 mean.HbA1c.treat <- ggplot(mean_visit, aes(x = VISIT, y =
     mean_HbA1c, ymin = mean_HbA1c-se_mean_HbA1c, ymax = mean_
     HbA1c+se_mean_HbA1c, colour = TREAT, group=TREAT)) +
76   geom_line(size=2) +
77   geom_errorbar(size=1,width=0.1) +
78   scale_x_discrete(labels=c("BASELINE" = "Baseline", "VISIT_7
       " = "Visit 7", "VISIT_9" = "Visit 9", "VISIT_11" = "
       Visit 11", "VISIT_12" = "Visit 12", "VISIT_13" = "Visit
       13")) +
79   labs(y="Mean HbA1c", x="") + guides(colour=guide_legend(
       title="Treatment")) +
80   theme(text = element_text(size=26), axis.title.x =element_
       text(margin=margin(0,0,30,0)))
81
82 #Export, plot of mean HbA1c for each treatment
83 fileName.HbA1c.treat <- "mean_HbA1c_treat.png"
84 png(fileName.HbA1c.treat, width=1000, height=600, units="px",
      pointsize = 16)
85 mean.HbA1c.treat
86 grid.text(footnote,unit(.015, 'npc'),unit(.04, 'npc'), just=c
     ("left"), gp=gpar(fontsize=12, font = 3))
87 dev.off()
```

C.5 R-Code: Treatment Policy Estimand

```
#Treatment Policy Estimand
library(lme4)
library(lmerTest)
library(Deriv)
library(mice)
library(VIM)
library(ggplot2)
library(qqplotr)
library(gridExtra)
library(forestplot)

#Load dataset
setwd("/Volumes/aerx-2076")
change_HbA1c_wide <- get(load("change_HbA1c_wide.RData"))
change_HbA1c_long <- get(load("change_HbA1c_long.RData"))

#HbA1c visits
visits <- c("VISIT_7","VISIT_9","VISIT_11","VISIT_12","VISIT_13")

#Non-inferiority margin
non_inf_margin <- 0.4

#Directory for plots
setwd("/Users/Kristoffer/Dropbox/Speciale/Latex/figures")

#Export, plot of missing data pattern
fileName.mis.pattern <- "mis_data_pattern_treatpol.png"
png(fileName.mis.pattern, width=1000, height=600, units="px", pointsize = 16)
aggr(change_HbA1c_wide[,c("BASELINE",visits)], col=c('navyblue','red'), numbers=TRUE,
     sortVars=FALSE, gap=3, cex.lab=1.6, cex.axis=1.2, cex.numbers=1.1,
     labels=c("Baseline","Visit 7","Visit 9","Visit 11","Visit 12","Visit 13"),
     ylab=c("Histogram of Missing Data","Missing Data Pattern"), oma=c(6,5,2,1))
dev.off()

################################
####    Selection Model     ####
################################

#Releveling VISIT to get the treatment difference at visit 13
change_HbA1c_long$VISIT <- relevel(change_HbA1c_long$VISIT,"13")
```

```
41
42 #MMRM based on the selection model approach
43 mmrm_sel <- lmer(HbA1c~TREAT*VISIT+BASELINE*VISIT+(1|SUBJ_ID)
   ,data=change_HbA1c_long)
44 coef_sel <- fixef(mmrm_sel)
45 vcov_sel <- vcov(mmrm_sel)
46
47 #The estimated treatment difference at visit 13
48 est_sel <- coef_sel["TREATAERx"]
49
50 #The std. error of the estimator
51 se_sel <- sqrt(vcov_sel["TREATAERx","TREATAERx"])
52
53 #The degrees of freedom(Satterthwaite)
54 df_sel <- summary(mmrm_sel)$coefficients["TREATAERx","df"]
55
56 #Confidence interval
57 lower_ci_sel <- est_sel + qt(0.025, df_sel)*se_sel
58 upper_ci_sel <- est_sel + qt(0.975, df_sel)*se_sel
59
60 #Test for non-inferiority
61 t_sel <- unname((est_sel-non_inf_margin)/se_sel)
62 p_sel <- pt(t_sel,df_sel) #OBS: pt() calculates P(X \leq x)
   by default
63
64 ##################################
65 ####  Pattern-Mixture Model ####
66 ##################################
67
68 #Subjects who completes visit 13(Completers)
69 completers <- which(!is.na(change_HbA1c_wide$VISIT_13))
70 comp_flag <- rep("No",nrow(change_HbA1c_wide))
71 comp_flag[completers] <- "Yes"
72 change_HbA1c_long$COMP_FLAG <- factor(rep(comp_flag,1,each=
   length(visits)),levels=c("Yes","No"))
73
74 #Number of subjects who do not complete visit 13(Non-
   completers)
75 num_non_comp <- sum(comp_flag=="No")
76
77 #Total number of subjects
78 num_subj <- length(comp_flag)
79
80 #The proportion of non-completers
81 prop_non <- num_non_comp/num_subj
82
83 #The estimated variance related to the proportion
84 var_prop_non <- prop_non*(1-prop_non)/num_subj
85
```

```
86 #MMRM based on the pattern-mixture model approach
87 mmrm_pm <- lmer(HbA1c~TREAT*VISIT+TREAT*COMP_FLAG+BASELINE*
      VISIT+(1|SUBJ_ID),data=change_HbA1c_long)
88 coef_pm <- fixef(mmrm_pm)
89 vcov_pm <- vcov(mmrm_pm)
90
91 #The estimated treatment difference at visit 13 for
      completers
92 TREAT_comp <- coef_pm["TREATAERx"]
93
94 #The estimated difference between the treatment differences
95 #for completers and non-completers
96 TREAT_diff <- coef_pm["TREATAERx:COMP_FLAGNo"]
97
98 #Function to estimate the average treatment difference across
       completers and non-completers
99 est_func_pm <- function(x_1,x_2,pi){
100   x_1 + pi*x_2
101 }
102
103 #The estimated average treatment difference across completers
       and non-completers
104 est_pm <- est_func_pm(TREAT_comp, TREAT_diff, prop_non)
105
106 #The transposed Jacobian matrix of est_func_pm
107 t_jacobian_pm <- Deriv(est_func_pm)
108
109 #Calculation of the variance and std. error of the estimator
110 #based on the Delta method
111 est_jacobian_pm <- t(t_jacobian_pm(TREAT_comp, TREAT_diff,
      prop_non))
112 Sigma_pm <- as.matrix(bdiag(vcov_pm[c("TREATAERx","TREATAERx:
      COMP_FLAGNo"),
113                                     c("TREATAERx","TREATAERx:
                                           COMP_FLAGNo")],
114                             var_prop_non))
115 var_pm <- as.vector(est_jacobian_pm %*% Sigma_pm %*% t(est_
      jacobian_pm))
116 se_pm <- sqrt(var_pm)
117
118 #Confidence interval
119 lower_ci_pm <- est_pm + qnorm(0.025)*se_pm
120 upper_ci_pm <- est_pm + qnorm(0.975)*se_pm
121
122 #Test for non-inferiority
123 z_pm <- unname((est_pm-non_inf_margin)/se_pm)
124 p_pm <- pnorm(z_pm) #OBS: pnorm() calculates P(X \leq x) by
      default
125
```

```
126 ################################
127 ####  Multiple Imputation   ####
128 ################################
129
130 #Dry run(to create the needed elements for imputation)
131 ini <- mice(change_HbA1c_wide, m=0,max=0)
132
133 #Predictor matrix
134 pred.matrix <- ini$predictorMatrix
135 pred.matrix[,] <- 0
136
137 #Previous measurements as predictors
138 for(i in 2:length(visits)){
139   pred.matrix[visits[i],visits[1:(i-1)]] <- 1
140 }
141
142 pred.matrix[visits,"BASELINE"] <- 1
143
144 #Post-processes
145 post.process <- ini$post
146 #Round imputations to one decimal
147 post.process[visits] <- "imp[[j]][,i] <- round(imp[[j]][,i
    ],1)"
148
149 #Visit sequence
150 vis.seq <- ini$visitSequence
151 vis.seq <- vis.seq[visits]
152
153 #Number of imputations
154 m <- 1000
155
156 #seed
157 imp.seed <- 100
158
159 ################################
160 ####     Copy Reference     ####
161 ################################
162
163 #Load functions to do MI using the copy reference approach
164 source("/Users/Kristoffer/Dropbox/Speciale/Latex/code/MI_copy
    _ref.R")
165
166 #Method
167 method <- ini$method
168 method[] <- ""
169 method[visits] <- "norm.extract"
170
171 #MI using copy reference
```

```
172 imp_copy_ref <- MI_copy_ref(change_HbA1c_wide, trt.var = "
    TREAT", copy.trt="AERx", ref.trt = "Insulin Aspart", meth
    = method, pred = pred.matrix, post = post.process, vis =
    vis.seq, M = m, seed = imp.seed)
173
174 #Analysis of each imputed dataset using an ANCOVA
175 ana_copy <- with(imp_copy_ref, lm(VISIT_13~TREAT+BASELINE))
176
177 #The m analysis are pooled
178 pool_copy <- pool(ana_copy)
179
180 #The estimated treatment difference at visit 13 and std.
    error
181 est_copy <- summary(pool_copy)["TREAT2","est"]
182 se_copy <- summary(pool_copy)["TREAT2","se"]
183
184 #The degrees of freedom related to the treatment difference
185 df_copy <-  summary(pool_copy)["TREAT2","df"]
186
187 #Confidence interval
188 lower_ci_copy <- summary(pool_copy)["TREAT2","lo 95"]
189 upper_ci_copy <- summary(pool_copy)["TREAT2","hi 95"]
190
191 #Test for non-inferiority
192 t_copy <- (est_copy-non_inf_margin)/se_copy
193 p_copy <- pt(t_copy,df_copy) #OBS: pt() calculates P(X \leq x
    ) by default
194
195 ################################
196 ####    Jump to Reference   ####
197 ################################
198
199 #Load functions to do MI using the jump to reference approach
200 source("/Users/Kristoffer/Dropbox/Speciale/Latex/code/MI_jump
    _to_ref.R")
201
202 #MI using jump to reference
203 imp_jump_ref <- MI_jump_to_ref(change_HbA1c_wide, trt.var = "
    TREAT", jump.trt="AERx", ref.trt = "Insulin Aspart", meth
    = method, pred = pred.matrix, post = post.process, vis =
    vis.seq, M = m, seed = imp.seed)
204
205 #Analysis of each imputed dataset using an ANCOVA
206 ana_jump <- with(imp_jump_ref, lm(VISIT_13~TREAT+BASELINE))
207
208 #The m analysis are pooled
209 pool_jump <- pool(ana_jump)
210
```

```
211 #The estimated treatment difference at visit 13 and std.
    error
212 est_jump <- summary(pool_jump)["TREAT2","est"]
213 se_jump <- summary(pool_jump)["TREAT2","se"]
214
215 #The degrees of freedom related to the treatment difference
216 df_jump <-  summary(pool_jump)["TREAT2","df"]
217
218 #Confidence interval
219 lower_ci_jump <- summary(pool_jump)["TREAT2","lo 95"]
220 upper_ci_jump <- summary(pool_jump)["TREAT2","hi 95"]
221
222 #Test for non-inferiority
223 t_jump <- (est_jump-non_inf_margin)/se_jump
224 p_jump <- pt(t_jump,df_jump) #OBS: pt() calculates P(X \leq x
    ) by default
225
226 ###################################
227 ####  Tipping Point Analysis  ####
228 ###################################
229
230 #Update method
231 method <- ini$method
232 method[] <- ""
233 method[visits] <- "norm"
234
235 #Multiple imputation for each treatment group
236 imp_AERx_MAR <- mice(change_HbA1c_wide[change_HbA1c_wide$
    TREAT=="AERx",], meth = method, pred = pred.matrix, post =
     post.process, vis = vis.seq, m = m, maxit = 1, seed = imp
    .seed)
237 imp_Asp_MAR <- mice(change_HbA1c_wide[change_HbA1c_wide$TREAT
    =="Insulin Aspart",], meth = method, pred = pred.matrix,
    post = post.process, vis = vis.seq, m = m, maxit = 1, seed
     = imp.seed)
238
239 #The imputed datasets
240 imp_long_AERx <- complete(imp_AERx_MAR, action="long",incl=
    TRUE)
241 imp_long_Asp <- complete(imp_Asp_MAR, action="long",incl=TRUE
    )
242
243 #Shift parameter
244 delta <- seq(0.0,0.014,0.007)
245
246 #Vectors for estimates, CI and p values
247 est_tip <- c()
248 lower_ci_tip <- c()
249 upper_ci_tip <- c()
```

```
250 p_tip <- c()
251
252 for(k in 1:length(delta)){
253   cat(c("Delta[",k,"] \n"))
254
255   #The imputed datasets for TREAT=="AERx"
256   imp_long_AERx_tip <- imp_long_AERx
257
258   #Indicator of the imputed values at the last visit(The
        first rep() accounts for the original data)
259   imputed_last_visit <- c(rep(FALSE,length(imp_AERx_MAR$where
        [,visits[length(visits)]])),unname(rep(imp_AERx_MAR$
        where[,visits[length(visits)]],m)))
260
261   #Add delta[k] to the imputed values at the last visit
262   imp_long_AERx_tip[imputed_last_visit,visits[length(visits)
        ]] <- imp_long_AERx_tip[imputed_last_visit,visits[length
        (visits)]] + delta[k]
263
264   #Combine the imputed datasets across treatments
265   imp_long_tip <- rbind(imp_long_AERx_tip,imp_long_Asp)
266
267   #imp_long_tip is sorted
268   imp_long_tip <- imp_long_tip[order(imp_long_tip$.imp,imp_
        long_tip$SUBJ_ID),]
269
270   #Transform imp_long_tip into a mids object
271   imp_tip <- as.mids(imp_long_tip)
272
273   #Analysis of each imputed dataset using an ANCOVA
274   ana_tip <- with(imp_tip, lm(VISIT_13~TREAT+BASELINE))
275
276   #The m analysis are pooled
277   pool_tip <- pool(ana_tip)
278
279   #The estimated treatment difference at visit 13 and std.
        error
280   est_tip[k] <- summary(pool_tip)["TREAT2","est"]
281   se_tip <- summary(pool_tip)["TREAT2","se"]
282
283   #The degrees of freedom related to the treatment difference
284   df_tip <-  summary(pool_tip)["TREAT2","df"]
285
286   #Confidence interval
287   lower_ci_tip[k] <- summary(pool_tip)["TREAT2","lo 95"]
288   upper_ci_tip[k] <- summary(pool_tip)["TREAT2","hi 95"]
289
290   #Test for non-inferiority
291   t_tip <- (est_tip[k]-non_inf_margin)/se_tip
```

```
292   p_tip[k] <- pt(t_tip,df_tip) #OBS: pt() calculates P(X \leq
          x) by default
293 }
294
295 ################################
296 ####      Forest plot       ####
297 ################################
298
299 #Directory for plots
300 setwd("/Users/Kristoffer/Dropbox/Speciale/Latex/figures")
301
302 tabletext <- list(c(list("", "Primary Analysis:",expression("
    MMRM"["S"]),"Sensitivity Analyses:",expression("MMRM"["PM"
    ]),expression("MI"["ANCOVA"]^"CR"),expression("MI"["ANCOVA
    "]^"J2R"),"Tipping Point Analysis:"),as.list(paste("delta
    =",delta))),
303                   c(list("Estimate",NA,round(est_sel,3),NA,
                          round(est_pm,3),round(est_copy,3),round(
                          est_jump,3),NA),as.list(round(est_tip,3)
                          )),
304                   c(list("95%-CI",NA,paste("[",round(lower_ci
                          _sel,3),";",round(upper_ci_sel,3),"]"),
                          NA,paste("[",round(lower_ci_pm,3),";",
                          round(upper_ci_pm,3),"]"),paste("[",
                          round(lower_ci_copy,3),";",round(upper_
                          ci_copy,3),"]"),paste("[",round(lower_ci
                          _jump,3),";",round(upper_ci_jump,3),"]")
                          ,NA),as.list(paste("[",round(lower_ci_
                          tip,3),";",round(upper_ci_tip,3),"]"))),
305                   c(list("p-value",NA,round(p_sel,4),NA,round
                          (p_pm,4),round(p_copy,4),round(p_jump,4)
                          ,NA),as.list(round(p_tip,4))))
306
307 ci.data <- cbind(
308   mean  = c(NA, NA, est_sel, NA, est_pm, est_copy, est_jump,
          NA, est_tip),
309   lower = c(NA, NA, lower_ci_sel, NA, lower_ci_pm, lower_ci_
          copy, lower_ci_jump, NA, lower_ci_tip),
310   upper = c(NA, NA, upper_ci_sel, NA, upper_ci_pm, upper_ci_
          copy, upper_ci_jump, NA, upper_ci_tip))
311 colnames(ci.data) <- c("mean", "lower", "upper")
312
313 footnote <- expression(italic(atop("H"["0"]~":"~beta["T"["13"
    ]]>=0.4,"H"["A"]~":"~beta["T"["13"]]<0.4)))
314
315 #Export of forestplot
316 fileName.forestplot <- paste("forestplot_estimand_treatpol_
    non_inf_0",10*non_inf_margin,"_nimp_",m,"_seed_",imp.seed,
    ".png",sep="")
```

```
317 png(fileName.forestplot, width=1000, height=600, units="px",
       pointsize = 16)
318 forestplot(tabletext, hrzl_lines = list("2" = gpar(lty=1,lwd
       =3)), txt_gp = fpTxtGp(ticks=gpar(cex=1.2),xlab=gpar(cex
       =1.4),label=gpar(cex=1.3)), ci.data, is.summary=c(TRUE,
       TRUE,FALSE,TRUE,rep(FALSE,4),rep(FALSE,length(delta))),
       grid= structure(c(.4), gp = gpar(lty = 2, lwd=2.5, col = "
       red")), col=fpColors(box="royalblue",line="darkblue",
       summary="royalblue"), ci.vertices = T, boxsize = 0.25,
       xlab="AERx - Insulin Aspart",lwd.ci=2,xticks=seq
       (-0.1,0.5,0.1),title="Test for Non-Inferiority")
319 grid.text(footnote,unit(.015, 'npc'),unit(.075, 'npc'), just=
       c("left"), gp=gpar(fontsize=16, font = 3))
320 dev.off()
321
322 ################################
323 ####        Results         ####
324 ################################
325
326 estimates <- c(est_sel, est_pm, est_copy, est_jump, est_tip)
327 lower_ci <- c(lower_ci_sel, lower_ci_pm, lower_ci_copy, lower
       _ci_jump, lower_ci_tip)
328 upper_ci <- c(upper_ci_sel, upper_ci_pm, upper_ci_copy, upper
       _ci_jump, upper_ci_tip)
329 p_values <- c(p_sel, p_pm, p_copy, p_jump, p_tip)
330
331 results <- cbind(estimates, lower_ci, upper_ci, p_values)
332 names(results) <- c("Estimate", "Lower bound CI", "Upper
       bound CI", "P-value")
333
334 #Save results
335 fileName.results <- paste("results_estimand_treatpol_non_inf_
       0",10*non_inf_margin,"_nimp_",m,"_seed_",imp.seed,".RData"
       ,sep="")
336 save(results, file=fileName.results)
```

C.6 R-Code: Hypothetical Estimand

```
#Hypothetical Estimand
library(lme4)
library(lmerTest)
library(Deriv)
library(mice)
library(VIM)
library(ggplot2)
library(qqplotr)
library(gridExtra)
library(forestplot)

#Load dataset
setwd("/Volumes/aerx-2076")
change_HbA1c_wide <- get(load("change_HbA1c_wide.RData"))
change_HbA1c_long <- get(load("change_HbA1c_long.RData"))

#HbA1c visits
visits <- c("VISIT_7","VISIT_9","VISIT_11","VISIT_12","VISIT_
   13")

#Non-inferiority margin
non_inf_margin <- 0.4

#Measurements are set to NA if the intercurrent event "use of
    escape therapy for at least 14 days" has occurred
change_HbA1c_long[change_HbA1c_long$ES_14_DAYS == "Yes", "
   HbA1c"] <- NA
change_HbA1c_wide[change_HbA1c_wide$ES_14_DAYS_VISIT_7 == "
   Yes", "VISIT_7"] <- NA
change_HbA1c_wide[change_HbA1c_wide$ES_14_DAYS_VISIT_9 == "
   Yes", "VISIT_9"] <- NA
change_HbA1c_wide[change_HbA1c_wide$ES_14_DAYS_VISIT_11 == "
   Yes", "VISIT_11"] <- NA
change_HbA1c_wide[change_HbA1c_wide$ES_14_DAYS_VISIT_12 == "
   Yes", "VISIT_12"] <- NA
change_HbA1c_wide[change_HbA1c_wide$ES_14_DAYS_VISIT_13 == "
   Yes", "VISIT_13"] <- NA

#Directory for plots
setwd("/Users/Kristoffer/Dropbox/Speciale/Latex/figures")

#Export, plot of missing data pattern
fileName.mis.pattern <- "mis_data_pattern_hypothetical.png"
png(fileName.mis.pattern, width=1000, height=600, units="px",
    pointsize = 16)
aggr(change_HbA1c_wide[,c("BASELINE",visits)], col=c('
   navyblue','red'), numbers=TRUE,
```

```
38      sortVars=FALSE, gap=3, cex.lab=1.6, cex.axis=1.2, cex.
           numbers=1.1,
39      labels=c("Baseline","Visit 7","Visit 9","Visit 11","
           Visit 12","Visit 13"),
40      ylab=c("Histogram of Missing Data","Missing Data Pattern
           "), oma=c(6,5,2,1))
41 dev.off()
42
43 ################################
44 ####    Selection Model     ####
45 ################################
46
47 #Releveling VISIT to get the treatment difference at visit 13
48 change_HbA1c_long$VISIT <- relevel(change_HbA1c_long$VISIT,"
      13")
49
50 #MMRM based on the selection model approach
51 mmrm_sel <- lmer(HbA1c~TREAT*VISIT+BASELINE*VISIT+(1|SUBJ_ID)
      ,data=change_HbA1c_long)
52 coef_sel <- fixef(mmrm_sel)
53 vcov_sel <- vcov(mmrm_sel)
54
55 #The estimated treatment difference at visit 13
56 est_sel <- coef_sel["TREATAERx"]
57
58 #The std. error of the estimator
59 se_sel <- sqrt(vcov_sel["TREATAERx","TREATAERx"])
60
61 #The degrees of freedom(Satterthwaite)
62 df_sel <- summary(mmrm_sel)$coefficients["TREATAERx","df"]
63
64 #Confidence interval
65 lower_ci_sel <- est_sel + qt(0.025, df_sel)*se_sel
66 upper_ci_sel <- est_sel + qt(0.975, df_sel)*se_sel
67
68 #Test for non-inferiority
69 t_sel <- unname((est_sel-non_inf_margin)/se_sel)
70 p_sel <- pt(t_sel,df_sel) #OBS: pt() calculates P(X \leq x)
      by default
71
72 ################################
73 ####  Pattern-Mixture Model ####
74 ################################
75
76 #Subjects who completes visit 13(Completers)
77 completers <- which(!is.na(change_HbA1c_wide$VISIT_13))
78 comp_flag <- rep("No",nrow(change_HbA1c_wide))
79 comp_flag[completers] <- "Yes"
```

```
80 change_HbA1c_long$COMP_FLAG <- factor(rep(comp_flag,1,each=
      length(visits)),levels=c("Yes","No"))
81
82 #Number of subjects who do not complete visit 13(Non-
      completers)
83 num_non_comp <- sum(comp_flag=="No")
84
85 #Total number of subjects
86 num_subj <- length(comp_flag)
87
88 #The proportion of non-completers
89 prop_non <- num_non_comp/num_subj
90
91 #The estimated variance related to the proportion
92 var_prop_non <- prop_non*(1-prop_non)/num_subj
93
94 #MMRM based on the pattern-mixture model approach
95 mmrm_pm <- lmer(HbA1c~TREAT*VISIT+TREAT*COMP_FLAG+BASELINE*
      VISIT+(1|SUBJ_ID),data=change_HbA1c_long)
96 coef_pm <- fixef(mmrm_pm)
97 vcov_pm <- vcov(mmrm_pm)
98
99 #The estimated treatment difference at visit 13 for
      completers
100 TREAT_comp <- coef_pm["TREATAERx"]
101
102 #The estimated difference between the treatment differences
103 #for completers and non-completers
104 TREAT_diff <- coef_pm["TREATAERx:COMP_FLAGNo"]
105
106 #Function to estimate the average treatment difference across
       completers and non-completers
107 est_func_pm <- function(x_1,x_2,pi){
108   x_1 + pi*x_2
109 }
110
111 #The estimated average treatment difference across completers
       and non-completers
112 est_pm <- est_func_pm(TREAT_comp, TREAT_diff, prop_non)
113
114 #The transposed Jacobian matrix of est_func_pm
115 t_jacobian_pm <- Deriv(est_func_pm)
116
117 #Calculation of the variance and std. error of the estimator
118 #based on the Delta method
119 est_jacobian_pm <- t(t_jacobian_pm(TREAT_comp, TREAT_diff,
      prop_non))
120 Sigma_pm <- as.matrix(bdiag(vcov_pm[c("TREATAERx","TREATAERx:
      COMP_FLAGNo"),
```

```
121                                     c("TREATAERx","TREATAERx:
                                           COMP_FLAGNo")],
122                               var_prop_non))
123 var_pm <- as.vector(est_jacobian_pm %*% Sigma_pm %*% t(est_
       jacobian_pm))
124 se_pm <- sqrt(var_pm)
125
126 #Confidence interval
127 lower_ci_pm <- est_pm + qnorm(0.025)*se_pm
128 upper_ci_pm <- est_pm + qnorm(0.975)*se_pm
129
130 #Test for non-inferiority
131 z_pm <- unname((est_pm-non_inf_margin)/se_pm)
132 p_pm <- pnorm(z_pm) #OBS: pnorm() calculates P(X \leq x) by
       default
133
134 ################################
135 ####  Multiple Imputation   ####
136 ################################
137
138 #Dry run(to create the needed elements for imputation)
139 ini <- mice(change_HbA1c_wide, m=0,max=0)
140
141 #Predictor matrix
142 pred.matrix <- ini$predictorMatrix
143 pred.matrix[,] <- 0
144
145 #Previous measurements as predictors
146 for(i in 2:length(visits)){
147   pred.matrix[visits[i],visits[1:(i-1)]] <- 1
148 }
149
150 pred.matrix[visits,"BASELINE"] <- 1
151
152 #Post-processes
153 post.process <- ini$post
154 #Round imputations to one decimal
155 post.process[visits] <- "imp[[j]][,i] <- round(imp[[j]][,i
       ],1)"
156
157 #Visit sequence
158 vis.seq <- ini$visitSequence
159 vis.seq <- vis.seq[visits]
160
161 #Number of imputations
162 m <- 1000
163
164 #seed
165 imp.seed <- 100
```

```
166
167 ################################
168 ####     Copy Reference     ####
169 ################################
170
171 #Load functions to do MI using the copy reference approach
172 source("/Users/Kristoffer/Dropbox/Speciale/Latex/code/MI_copy
    _ref.R")
173
174 #Method
175 method <- ini$method
176 method[] <- ""
177 method[visits] <- "norm.extract"
178
179 #MI using copy reference
180 imp_copy_ref <- MI_copy_ref(change_HbA1c_wide, trt.var = "
    TREAT", copy.trt="AERx", ref.trt = "Insulin Aspart", meth
    = method, pred = pred.matrix, post = post.process, vis =
    vis.seq, M = m, seed = imp.seed)
181
182 #Analysis of each imputed dataset using an ANCOVA
183 ana_copy <- with(imp_copy_ref, lm(VISIT_13~TREAT+BASELINE))
184
185 #The m analysis are pooled
186 pool_copy <- pool(ana_copy)
187
188 #The estimated treatment difference at visit 13 and std.
    error
189 est_copy <- summary(pool_copy)["TREAT2","est"]
190 se_copy <- summary(pool_copy)["TREAT2","se"]
191
192 #The degrees of freedom related to the treatment difference
193 df_copy <-  summary(pool_copy)["TREAT2","df"]
194
195 #Confidence interval
196 lower_ci_copy <- summary(pool_copy)["TREAT2","lo 95"]
197 upper_ci_copy <- summary(pool_copy)["TREAT2","hi 95"]
198
199 #Test for non-inferiority
200 t_copy <- (est_copy-non_inf_margin)/se_copy
201 p_copy <- pt(t_copy,df_copy) #OBS: pt() calculates P(X \leq x
    ) by default
202
203 ################################
204 ####    Jump to Reference   ####
205 ################################
206
207 #Load functions to do MI using the jump to reference approach
```

```
208 source("/Users/Kristoffer/Dropbox/Speciale/Latex/code/MI_jump
    _to_ref.R")
209
210 #MI using jump to reference
211 imp_jump_ref <- MI_jump_to_ref(change_HbA1c_wide, trt.var = "
    TREAT", jump.trt="AERx", ref.trt = "Insulin Aspart", meth
    = method, pred = pred.matrix, post = post.process, vis =
    vis.seq, M = m, seed = imp.seed)
212
213 #Analysis of each imputed dataset using an ANCOVA
214 ana_jump <- with(imp_jump_ref, lm(VISIT_13~TREAT+BASELINE))
215
216 #The m analysis are pooled
217 pool_jump <- pool(ana_jump)
218
219 #The estimated treatment difference at visit 13 and std.
    error
220 est_jump <- summary(pool_jump)["TREAT2","est"]
221 se_jump <- summary(pool_jump)["TREAT2","se"]
222
223 #The degrees of freedom related to the treatment difference
224 df_jump <-  summary(pool_jump)["TREAT2","df"]
225
226 #Confidence interval
227 lower_ci_jump <- summary(pool_jump)["TREAT2","lo 95"]
228 upper_ci_jump <- summary(pool_jump)["TREAT2","hi 95"]
229
230 #Test for non-inferiority
231 t_jump <- (est_jump-non_inf_margin)/se_jump
232 p_jump <- pt(t_jump,df_jump) #OBS: pt() calculates P(X \leq x
    ) by default
233
234 ###################################
235 ####  Tipping Point Analysis  ####
236 ###################################
237
238 #Update method
239 method <- ini$method
240 method[] <- ""
241 method[visits] <- "norm"
242
243 #Multiple imputation for each treatment group
244 imp_AERx_MAR <- mice(change_HbA1c_wide[change_HbA1c_wide$
    TREAT=="AERx",], meth = method, pred = pred.matrix, post =
     post.process, vis = vis.seq, m = m, maxit = 1, seed = imp
    .seed)
245 imp_Asp_MAR <- mice(change_HbA1c_wide[change_HbA1c_wide$TREAT
    =="Insulin Aspart",], meth = method, pred = pred.matrix,
    post = post.process, vis = vis.seq, m = m, maxit = 1, seed
```

```
    = imp.seed)
246
247 #The imputed datasets
248 imp_long_AERx <- complete(imp_AERx_MAR, action="long",incl=
    TRUE)
249 imp_long_Asp <- complete(imp_Asp_MAR, action="long",incl=TRUE
    )
250
251 #Shift parameter
252 delta <- seq(0.0,-0.054,-0.027)
253
254 #Vectors for estimates, CI and p values
255 est_tip <- c()
256 lower_ci_tip <- c()
257 upper_ci_tip <- c()
258 p_tip <- c()
259
260 for(k in 1:length(delta)){
261   cat(c("Delta[",k,"] \n"))
262
263   #The imputed datasets for TREAT=="AERx"
264   imp_long_AERx_tip <- imp_long_AERx
265
266   #Indicator of the imputed values at the last visit(The
        first rep() accounts for the original data)
267   imputed_last_visit <- c(rep(FALSE,length(imp_AERx_MAR$where
        [,visits[length(visits)]])),unname(rep(imp_AERx_MAR$
        where[,visits[length(visits)]],m)))
268
269   #Add delta[k] to the imputed values at the last visit
270   imp_long_AERx_tip[imputed_last_visit,visits[length(visits)
        ]] <- imp_long_AERx_tip[imputed_last_visit,visits[length
        (visits)]] + delta[k]
271
272   #Combine the imputed datasets across treatments
273   imp_long_tip <- rbind(imp_long_AERx_tip,imp_long_Asp)
274
275   #imp_long_tip is sorted
276   imp_long_tip <- imp_long_tip[order(imp_long_tip$.imp,imp_
        long_tip$SUBJ_ID),]
277
278   #Transform imp_long_tip into a mids object
279   imp_tip <- as.mids(imp_long_tip)
280
281   #Analysis of each imputed dataset using an ANCOVA
282   ana_tip <- with(imp_tip, lm(VISIT_13~TREAT+BASELINE))
283
284   #The m analysis are pooled
285   pool_tip <- pool(ana_tip)
```

```
286
287   #The estimated treatment difference at visit 13 and std.
          error
288   est_tip[k] <- summary(pool_tip)["TREAT2","est"]
289   se_tip <- summary(pool_tip)["TREAT2","se"]
290
291   #The degrees of freedom related to the treatment difference
292   df_tip <-  summary(pool_tip)["TREAT2","df"]
293
294   #Confidence interval
295   lower_ci_tip[k] <- summary(pool_tip)["TREAT2","lo 95"]
296   upper_ci_tip[k] <- summary(pool_tip)["TREAT2","hi 95"]
297
298   #Test for non-inferiority
299   t_tip <- (est_tip[k]-non_inf_margin)/se_tip
300   p_tip[k] <- pt(t_tip,df_tip) #OBS: pt() calculates P(X \leq
          x) by default
301 }
302
303 ################################
304 ####      Forest plot       ####
305 ################################
306
307 #Directory for plots
308 setwd("/Users/Kristoffer/Dropbox/Speciale/Latex/figures")
309
310 tabletext <- list(c(list("", "Primary Analysis:",expression("
       MMRM"["S"]),"Sensitivity Analyses:",expression("MMRM"["PM"
       ]),expression("MI"["ANCOVA"]^"CR"),expression("MI"["ANCOVA
       "]^"J2R"),"Tipping Point Analysis:"),as.list(paste("delta
       =",delta))),
311                   c(list("Estimate",NA,round(est_sel,3),NA,
                          round(est_pm,3),round(est_copy,3),round(
                          est_jump,3),NA),as.list(round(est_tip,3)
                          )),
312                   c(list("95%-CI",NA,paste("[",round(lower_ci
                          _sel,3),";",round(upper_ci_sel,3),"]"),
                          NA,paste("[",round(lower_ci_pm,3),";",
                          round(upper_ci_pm,3),"]"),paste("[",
                          round(lower_ci_copy,3),";",round(upper_
                          ci_copy,3),"]"),paste("[",round(lower_ci
                          _jump,3),";",round(upper_ci_jump,3),"]")
                          ,NA),as.list(paste("[",round(lower_ci_
                          tip,3),";",round(upper_ci_tip,3),"]"))),
313                   c(list("p-value",NA,round(p_sel,4),NA,round
                          (p_pm,4),round(p_copy,4),round(p_jump,4)
                          ,NA),as.list(round(p_tip,4))))
314
315 ci.data <- cbind(
```

```
316   mean  = c(NA, NA, est_sel, NA, est_pm, est_copy, est_jump,
          NA, est_tip),
317   lower = c(NA, NA, lower_ci_sel, NA, lower_ci_pm, lower_ci_
          copy, lower_ci_jump, NA, lower_ci_tip),
318   upper = c(NA, NA, upper_ci_sel, NA, upper_ci_pm, upper_ci_
          copy, upper_ci_jump, NA, upper_ci_tip))
319 colnames(ci.data) <- c("mean", "lower", "upper")
320
321 footnote <- expression(italic(atop("H"["0"]~":"~beta["T"["13"
      ]]>=0.4,"H"["A"]~":"~beta["T"["13"]]<0.4)))
322
323 #Export of forestplot
324 fileName.forestplot <- paste("forestplot_estimand_
      hypothetical_non_inf_0",10*non_inf_margin,"_nimp_",m,"_
      seed_",imp.seed,".png",sep="")
325 png(fileName.forestplot, width=1000, height=600, units="px",
      pointsize = 16)
326 forestplot(tabletext, hrzl_lines = list("2" = gpar(lty=1,lwd
      =3)), txt_gp = fpTxtGp(ticks=gpar(cex=1.2),xlab=gpar(cex
      =1.4),label=gpar(cex=1.3)), ci.data, is.summary=c(TRUE,
      TRUE,FALSE,TRUE,rep(FALSE,4),rep(FALSE,length(delta))),
      grid= structure(c(.4), gp = gpar(lty = 2, lwd=2.5, col = "
      red")), col=fpColors(box="royalblue",line="darkblue",
      summary="royalblue"), ci.vertices = T, boxsize = 0.25,
      xlab="AERx - Insulin Aspart",lwd.ci=2,xticks=seq
      (-0.1,0.5,0.1),title="Test for Non-Inferiority")
327 grid.text(footnote,unit(.015, 'npc'),unit(.075, 'npc'), just=
      c("left"), gp=gpar(fontsize=16, font = 3))
328 dev.off()
329
330 ##################################
331 ####         Results          ####
332 ##################################
333
334 estimates <- c(est_sel, est_pm, est_copy, est_jump, est_tip)
335 lower_ci <- c(lower_ci_sel, lower_ci_pm, lower_ci_copy, lower
      _ci_jump, lower_ci_tip)
336 upper_ci <- c(upper_ci_sel, upper_ci_pm, upper_ci_copy, upper
      _ci_jump, upper_ci_tip)
337 p_values <- c(p_sel, p_pm, p_copy, p_jump, p_tip)
338
339 results <- cbind(estimates, lower_ci, upper_ci, p_values)
340 names(results) <- c("Estimate", "Lower bound CI", "Upper
      bound CI", "P-value")
341
342 #Save results
343 fileName.results <- paste("results_estimand_hypothetical_non_
      inf_0",10*non_inf_margin,"_nimp_",m,"_seed_",imp.seed,".
      RData",sep="")
```

```
344 save(results, file=fileName.results)
```

C.7 R-Code: Composite Estimand

```
#Composite Estimand
library(mice)
library(ggplot2)
library(qqplotr)
library(gridExtra)
library(forestplot)

#Load dataset
setwd("/Volumes/aerx-2076")
HbA1c_wide <- get(load("HbA1c_wide.RData"))

#HbA1c visits
visits <- c("VISIT_7","VISIT_9","VISIT_11","VISIT_12","VISIT_
   13")

#HbA1c success criteria
HbA1c_success <- 7.0

#Response variable
HbA1c_wide$response <- factor(rep(NA,nrow(HbA1c_wide)),levels
   =c("Failure","Success"))
HbA1c_wide[(!is.na(HbA1c_wide$VISIT_13) & HbA1c_wide$VISIT_13
    <= HbA1c_success & HbA1c_wide$ES_14_DAYS_VISIT_13 == "No"
   ),"response"] <- "Success"
HbA1c_wide[(!is.na(HbA1c_wide$VISIT_13) & (HbA1c_wide$VISIT_
   13 > HbA1c_success | HbA1c_wide$ES_14_DAYS_VISIT_13 == "
   Yes")),"response"] <- "Failure"

################################
####  Multiple Imputation   ####
################################

#Dry run(to create the needed elements for imputation)
ini <- mice(HbA1c_wide, m=0,max=0)

#Predictor matrix
pred.matrix <- ini$predictorMatrix
pred.matrix[,] <- 0

#Previous measurements as predictors
for(i in 2:length(visits)){
  pred.matrix[visits[i],visits[1:(i-1)]] <- 1
}

pred.matrix[visits,"BASELINE"] <- 1

#Post-processes
```

```
post.process <- ini$post
#Round imputations to one decimal
post.process[visits] <- "imp[[j]][,i] <- round(imp[[j]][,i
   ],1)"

#Method
method <- ini$method
method[] <- ""
method[visits] <- "norm"

#Visit sequence
vis.seq <- ini$visitSequence
vis.seq <- vis.seq[visits]

#Number of imputations
m <- 1000

#seed
imp.seed <- 100

#Multiple imputation for each treatment group
imp_AERx_MAR <- mice(HbA1c_wide[HbA1c_wide$TREAT=="AERx",],
   meth = method, pred = pred.matrix, post = post.process,
   vis = vis.seq, m = m, maxit = 1, seed = imp.seed)
imp_Asp_MAR <- mice(HbA1c_wide[HbA1c_wide$TREAT=="Insulin
   Aspart",], meth = method, pred = pred.matrix, post = post.
   process, vis = vis.seq, m = m, maxit = 1, seed = imp.seed)

#The imputed datasets
imp_long_AERx <- complete(imp_AERx_MAR, action="long",incl=
   TRUE)
imp_long_Asp <- complete(imp_Asp_MAR, action="long",incl=TRUE
   )

#Bind the two datasets
imp_long_MAR <- rbind(imp_long_AERx,imp_long_Asp)

#Response variable
imp_long_MAR[(imp_long_MAR$.imp != 0 & imp_long_MAR$VISIT_13
   <= HbA1c_success & imp_long_MAR$ES_14_DAYS_VISIT_13 == "No
   "), "response"] <- "Success"
imp_long_MAR[(imp_long_MAR$.imp != 0 & (imp_long_MAR$VISIT_13
    > HbA1c_success | imp_long_MAR$ES_14_DAYS_VISIT_13 == "
   Yes")), "response"] <- "Failure"

#imp_long is sorted
imp_long_MAR <- imp_long_MAR[order(imp_long_MAR$.imp,imp_long
   _MAR$SUBJ_ID),]

```

```
79 #Transform imp_long into a mids object
80 imp_MAR <- as.mids(imp_long_MAR)
81
82 #Analysis of each imputed dataset using a logistic regression
83 ana_MAR <- with(imp_MAR, glm(response~TREAT+BASELINE, family=
       "binomial"))
84
85 #The m analysis are pooled
86 pool_MAR <- pool(ana_MAR)
87
88 #The estimated log(odds ratio) and std. error
89 est_MAR <- summary(pool_MAR)["TREAT2","est"]
90 se_MAR <- summary(pool_MAR)["TREAT2","se"]
91
92 #The degrees of freedom related to the estimated log(odds
       ratio)
93 df_MAR <-  summary(pool_MAR)["TREAT2","df"]
94
95 #Confidence interval
96 lower_ci_MAR <- summary(pool_MAR)["TREAT2","lo 95"]
97 upper_ci_MAR <- summary(pool_MAR)["TREAT2","hi 95"]
98
99 #Test for log(odds ratio) to be greater than zero
100 t_MAR <- est_MAR/se_MAR
101 p_MAR <- pt(t_MAR,df_MAR,lower.tail = FALSE) #OBS: pt()
       calculates P(X \leq x) by default
102
103 #Transform to odds ratio
104 est_OR_MAR <- exp(est_MAR)
105 lower_ci_OR_MAR <- exp(lower_ci_MAR)
106 upper_ci_OR_MAR <- exp(upper_ci_MAR)
107
108 ################################
109 ####     Copy Reference     ####
110 ################################
111
112 #Load functions to do MI using the copy reference approach
113 source("/Users/Kristoffer/Dropbox/Speciale/Latex/code/MI_copy
       _ref.R")
114
115 #Method
116 method <- ini$method
117 method[] <- ""
118 method[visits] <- "norm.extract"
119
120 #MI using copy reference
121 imp_copy_ref <- MI_copy_ref(HbA1c_wide, trt.var = "TREAT",
       copy.trt="AERx", ref.trt = "Insulin Aspart", meth = method
       , pred = pred.matrix, post = post.process, vis = vis.seq,
```

```
   M = m, seed = imp.seed)

#The imputed datasets
imp_long_copy <- complete(imp_copy_ref, action="long",incl=
   TRUE)

#Response variable
imp_long_copy[(imp_long_copy$.imp != 0 & imp_long_copy$VISIT_
   13 <= HbA1c_success & imp_long_copy$ES_14_DAYS_VISIT_13 ==
    "No"), "response"] <- "Success"
imp_long_copy[(imp_long_copy$.imp != 0 & (imp_long_copy$VISIT
   _13 > HbA1c_success | imp_long_copy$ES_14_DAYS_VISIT_13 ==
    "Yes")), "response"] <- "Failure"

#imp_long_copy is sorted
imp_long_copy <- imp_long_copy[order(imp_long_copy$.imp,imp_
   long_copy$SUBJ_ID),]

#Transform imp_long_copy into a mids object
imp_copy <- as.mids(imp_long_copy)

#Analysis of each imputed dataset using a logistic regression
ana_copy <- with(imp_copy, glm(response~TREAT+BASELINE,
   family="binomial"))

#The m analysis are pooled
pool_copy <- pool(ana_copy)

#The estimated log(odds ratio) and std. error
est_copy <- summary(pool_copy)["TREAT2","est"]
se_copy <- summary(pool_copy)["TREAT2","se"]

#The degrees of freedom related to the estimated log(odds
   ratio)
df_copy <-  summary(pool_copy)["TREAT2","df"]

#Confidence interval
lower_ci_copy <- summary(pool_copy)["TREAT2","lo 95"]
upper_ci_copy <- summary(pool_copy)["TREAT2","hi 95"]

#Test for log(odds ratio) to be greater than zero
t_copy <- est_copy/se_copy
p_copy <- pt(t_copy,df_copy,lower.tail = FALSE) #OBS: pt()
   calculates P(X \leq x) by default

#Transform to odds ratio
est_OR_copy <- exp(est_copy)
lower_ci_OR_copy <- exp(lower_ci_copy)
upper_ci_OR_copy <- exp(upper_ci_copy)
```

```
161
162 ################################
163 ####    Jump to Reference   ####
164 ################################
165
166 #Load functions to do MI using the jump to reference approach
167 source("/Users/Kristoffer/Dropbox/Speciale/Latex/code/MI_jump
    _to_ref.R")
168
169 #MI using jump to reference
170 imp_jump_ref <- MI_jump_to_ref(HbA1c_wide, trt.var = "TREAT",
     jump.trt="AERx", ref.trt = "Insulin Aspart", meth =
    method, pred = pred.matrix, post = post.process, vis = vis
    .seq, M = m, seed = imp.seed)
171
172 #The imputed datasets
173 imp_long_jump <- complete(imp_jump_ref, action="long",incl=
    TRUE)
174
175 #Response variable
176 imp_long_jump[(imp_long_jump$.imp != 0 & imp_long_jump$VISIT_
    13 <= HbA1c_success & imp_long_jump$ES_14_DAYS_VISIT_13 ==
     "No"), "response"] <- "Success"
177 imp_long_jump[(imp_long_jump$.imp != 0 & (imp_long_jump$VISIT
    _13 > HbA1c_success | imp_long_jump$ES_14_DAYS_VISIT_13 ==
     "Yes")), "response"] <- "Failure"
178
179 #imp_long_jump is sorted
180 imp_long_jump <- imp_long_jump[order(imp_long_jump$.imp,imp_
    long_jump$SUBJ_ID),]
181
182 #Transform imp_long_jump into a mids object
183 imp_jump <- as.mids(imp_long_jump)
184
185 #Analysis of each imputed dataset using a logistic regression
186 ana_jump <- with(imp_jump, glm(response~TREAT+BASELINE,
    family="binomial"))
187
188 #The m analysis are pooled
189 pool_jump <- pool(ana_jump)
190
191 #The estimated log(odds ratio) and std. error
192 est_jump <- summary(pool_jump)["TREAT2","est"]
193 se_jump <- summary(pool_jump)["TREAT2","se"]
194
195 #The degrees of freedom related to the estimated log(odds
    ratio)
196 df_jump <-  summary(pool_jump)["TREAT2","df"]
197
```

```r
#Confidence interval
lower_ci_jump <- summary(pool_jump)["TREAT2","lo 95"]
upper_ci_jump <- summary(pool_jump)["TREAT2","hi 95"]

#Test for log(odds ratio) to be greater than zero
t_jump <- est_jump/se_jump
p_jump <- pt(t_jump,df_jump,lower.tail = FALSE) #OBS: pt()
   calculates P(X \leq x) by default

#Transform to odds ratio
est_OR_jump <- exp(est_jump)
lower_ci_OR_jump <- exp(lower_ci_jump)
upper_ci_OR_jump <- exp(upper_ci_jump)

###################################
####  Tipping Point Analysis  ####
###################################

#Shift parameter
delta <- seq(-1.0,-1.2,-0.1)

#Vectors for estimates, CI and p values
est_OR_tip <- c()
lower_ci_OR_tip <- c()
upper_ci_OR_tip <- c()
p_tip <- c()

for(k in 1:length(delta)){
  cat(c("Delta[",k,"] \n"))

  #The imputed datasets for TREAT=="AERx"
  imp_long_AERx_tip <- imp_long_AERx

  #Indicator of the imputed values at the last visit(The
     first rep() accounts for the original data)
  imputed_last_visit <- c(rep(FALSE,length(imp_AERx_MAR$where
     [,visits[length(visits)]])),unname(rep(imp_AERx_MAR$
     where[,visits[length(visits)]],m)))

  #Add delta[k] to the imputed values at the last visit
  imp_long_AERx_tip[imputed_last_visit,visits[length(visits)
     ]] <- imp_long_AERx_tip[imputed_last_visit,visits[length
     (visits)]] + delta[k]

  #Combine the imputed datasets across treatments
  imp_long_tip <- rbind(imp_long_AERx_tip,imp_long_Asp)

  #imp_long_tip is sorted
```

```
240   imp_long_tip <- imp_long_tip[order(imp_long_tip$.imp,imp_
        long_tip$SUBJ_ID),]
241
242   #Response variable
243   imp_long_tip[(imp_long_tip$.imp != 0 & imp_long_tip$VISIT_
        13 <= HbA1c_success & imp_long_tip$ES_14_DAYS_VISIT_13
        == "No"), "response"] <- "Success"
244   imp_long_tip[(imp_long_tip$.imp != 0 & (imp_long_tip$VISIT_
        13 > HbA1c_success | imp_long_tip$ES_14_DAYS_VISIT_13 ==
         "Yes")), "response"] <- "Failure"
245
246   #Transform imp_long_tip into a mids object
247   imp_tip <- as.mids(imp_long_tip)
248
249   #Analysis of each imputed dataset using a logistic
        regression
250   ana_tip <- with(imp_tip, glm(response~TREAT+BASELINE,
        family="binomial"))
251
252   #The m analysis are pooled
253   pool_tip <- pool(ana_tip)
254
255   #The estimated log(odds ratio) and std. error
256   est_tip <- summary(pool_tip)["TREAT2","est"]
257   se_tip <- summary(pool_tip)["TREAT2","se"]
258
259   #The degrees of freedom related to the estimated log(odds
        ratio)
260   df_tip <-  summary(pool_tip)["TREAT2","df"]
261
262   #Confidence interval
263   lower_ci_tip <- summary(pool_tip)["TREAT2","lo 95"]
264   upper_ci_tip <- summary(pool_tip)["TREAT2","hi 95"]
265
266   #Test for log(odds ratio) to be greater than zero
267   t_tip <- est_tip/se_tip
268   p_tip[k] <- pt(t_tip,df_tip,lower.tail = FALSE) #OBS: pt()
        calculates P(X \leq x) by default
269
270   #Transform to odds ratio
271   est_OR_tip[k] <- exp(est_tip)
272   lower_ci_OR_tip[k] <- exp(lower_ci_tip)
273   upper_ci_OR_tip[k] <- exp(upper_ci_tip)
274 }
275
276 ################################
277 ####      Forest plot       ####
278 ################################
279
```

```
280 #Directory for plots
281 setwd("/Users/Kristoffer/Dropbox/Speciale/Latex/figures")
282
283 tabletext <- list(c(list("", "Primary Analysis:",expression("
    MI"["LogReg"]^"MAR"),"Sensitivity Analyses:",expression("
    MI"["LogReg"]^"CR"),expression("MI"["LogReg"]^"J2R"),"
    Tipping Point Analysis:"),as.list(paste("delta =",delta)))
    ,
284                   c(list("Estimate",NA,round(est_OR_MAR,3),NA
                        ,round(est_OR_copy,3),round(est_OR_jump
                        ,3),NA),as.list(round(est_OR_tip,3))),
285                   c(list("95%-CI",NA,paste("[",round(lower_ci
                        _OR_MAR,3),";",round(upper_ci_OR_MAR,3),
                        "]"),NA,paste("[",round(lower_ci_OR_copy
                        ,3),";",round(upper_ci_OR_copy,3),"]"),
                        paste("[",round(lower_ci_OR_jump,3),";",
                        round(upper_ci_OR_jump,3),"]"),NA),as.
                        list(paste("[",round(lower_ci_OR_tip,3),
                        ";",round(upper_ci_OR_tip,3),"]"))),
286                   c(list("p-value",NA,round(p_MAR,4),NA,round
                        (p_copy,4),round(p_jump,4),NA),as.list(
                        round(p_tip,4))))
287
288 ci.data <- cbind(
289   mean  = c(NA, NA, est_OR_MAR, NA, est_OR_copy, est_OR_jump,
          NA, est_OR_tip),
290   lower = c(NA, NA, lower_ci_OR_MAR, NA, lower_ci_OR_copy,
        lower_ci_OR_jump, NA, lower_ci_OR_tip),
291   upper = c(NA, NA, upper_ci_OR_MAR, NA, upper_ci_OR_copy,
        upper_ci_OR_jump, NA, upper_ci_OR_tip))
292 colnames(ci.data) <- c("mean", "lower", "upper")
293
294 footnote <- expression(italic(atop("H"["0"]~":"~"log("~"OR"["
    S"]~")"<=0,"H"["A"]~":"~"log("~"OR"["S"]~")">0)))
295
296 #Export of forestplot
297 fileName.forestplot <- paste("forestplot_estimand_composite_
    nimp_",m,"_seed_",imp.seed,".png",sep="")
298 png(fileName.forestplot, width=1000, height=600, units="px",
    pointsize = 16)
299 forestplot(tabletext, hrzl_lines = list("2" = gpar(lty=1,lwd
    =3)), txt_gp = fpTxtGp(ticks=gpar(cex=1.2),xlab=gpar(cex
    =1.4),label=gpar(cex=1.3)), ci.data, is.summary=c(TRUE,
    TRUE,FALSE,TRUE,rep(FALSE,3),rep(FALSE,length(delta))),
    grid= structure(c(1), gp = gpar(lty = 2, lwd=2.5, col = "
    red")), col=fpColors(box="royalblue",line="darkblue",
    summary="royalblue"), ci.vertices = T, boxsize = 0.25,
    xlab="Odds ratio, AERx/Aspart",lwd.ci=2,title="Test for
    Odds Ratio to Be Greater Than One")
```

```
300 grid.text(footnote,unit(.015, 'npc'),unit(.075, 'npc'), just=
    c("left"), gp=gpar(fontsize=16, font = 3))
301 dev.off()
302
303 ################################
304 ####         Results          ####
305 ################################
306
307 estimates <- c(est_OR_MAR, est_OR_copy, est_OR_jump, est_OR_
    tip)
308 lower_ci <- c(lower_ci_OR_MAR, lower_ci_OR_copy, lower_ci_OR_
    jump, lower_ci_OR_tip)
309 upper_ci <- c(upper_ci_OR_MAR, upper_ci_OR_copy, upper_ci_OR_
    jump, upper_ci_OR_tip)
310 p_values <- c(p_MAR, p_copy, p_jump, p_tip)
311
312 results <- cbind(estimates, lower_ci, upper_ci, p_values)
313 names(results) <- c("Estimate", "Lower bound CI", "Upper
    bound CI", "P-value")
314
315 #Save results
316 fileName.results <- paste("results_estimand_composite_nimp_",
    m,"_seed_",imp.seed,".RData",sep="")
317 save(results, file=fileName.results)
```

C.8 R-Code: While on Treatment Estimand

```r
#While on Treatment Estimand
library(mice)
library(ggplot2)
library(qqplotr)
library(gridExtra)
library(forestplot)

#Load dataset
setwd("/Volumes/aerx-2076")
change_HbA1c_wide <- get(load("change_HbA1c_wide.RData"))

#HbA1c visits
visits <- c("VISIT_7","VISIT_9","VISIT_11","VISIT_12","VISIT_13")

#Indicators for use of escape therapy for at least 14 consecutive days at each visit
ES_14_DAYS <- c("ES_14_DAYS_VISIT_7","ES_14_DAYS_VISIT_9","ES_14_DAYS_VISIT_11","ES_14_DAYS_VISIT_12","ES_14_DAYS_VISIT_13")

#Non-inferiority margin
non_inf_margin <- 0.4

#Measurements are set to NA if the intercurrent event "use of escape therapy for at least 14 days" has occurred
change_HbA1c_wide[change_HbA1c_wide$ES_14_DAYS_VISIT_7 == "Yes", "VISIT_7"] <- NA
change_HbA1c_wide[change_HbA1c_wide$ES_14_DAYS_VISIT_9 == "Yes", "VISIT_9"] <- NA
change_HbA1c_wide[change_HbA1c_wide$ES_14_DAYS_VISIT_11 == "Yes", "VISIT_11"] <- NA
change_HbA1c_wide[change_HbA1c_wide$ES_14_DAYS_VISIT_12 == "Yes", "VISIT_12"] <- NA
change_HbA1c_wide[change_HbA1c_wide$ES_14_DAYS_VISIT_13 == "Yes", "VISIT_13"] <- NA

################################
####  Multiple Imputation   ####
################################

#Dry run(to create the needed elements for imputation)
ini <- mice(change_HbA1c_wide, m=0,max=0)

#Predictor matrix
pred.matrix <- ini$predictorMatrix
pred.matrix[,] <- 0
```

```
38
39 #Previous measurements as predictors
40 for(i in 2:length(visits)){
41   pred.matrix[visits[i],visits[1:(i-1)]] <- 1
42 }
43
44 pred.matrix[visits,"BASELINE"] <- 1
45
46 #Post-processes
47 post.process <- ini$post
48 #Round imputations to one decimal
49 post.process[visits] <- "imp[[j]][,i] <- round(imp[[j]][,i
     ],1)"
50
51 #Method
52 method <- ini$method
53 method[] <- ""
54 method[visits] <- "norm"
55
56 #Visit sequence
57 vis.seq <- ini$visitSequence
58 vis.seq <- vis.seq[visits]
59
60 #Where to impute
61 where <- ini$where
62 where_AERx <- where[change_HbA1c_wide$TREAT== "AERx",]
63 where_Asp <- where[change_HbA1c_wide$TREAT== "Insulin Aspart"
     ,]
64 for(i in 1:length(visits)){
65   es_use_AERx <- change_HbA1c_wide[change_HbA1c_wide$TREAT==
       AERx",ES_14_DAYS[i]]=="Yes"
66   es_use_Asp <- change_HbA1c_wide[change_HbA1c_wide$TREAT==
       Insulin Aspart",ES_14_DAYS[i]]=="Yes"
67   where_AERx[es_use_AERx,visits[i]] <- FALSE
68   where_Asp[es_use_Asp,visits[i]] <- FALSE
69 }
70
71 #Number of imputations
72 m <- 1000
73
74 #seed
75 imp.seed <- 100
76
77 #Multiple imputation for each treatment group
78 imp_AERx_MAR <- mice(change_HbA1c_wide[change_HbA1c_wide$
     TREAT=="AERx",], meth = method, pred = pred.matrix, post =
      post.process, vis = vis.seq, where=where_AERx, m = m,
     maxit = 1, seed = imp.seed)
```

```
imp_Asp_MAR <- mice(change_HbA1c_wide[change_HbA1c_wide$TREAT
   =="Insulin Aspart",], meth = method, pred = pred.matrix,
   post = post.process, vis = vis.seq, where=where_Asp, m = m
   , maxit = 1, seed = imp.seed)

#The imputed datasets
imp_long_AERx <- complete(imp_AERx_MAR, action="long",incl=
   TRUE)
imp_long_Asp <- complete(imp_Asp_MAR, action="long",incl=TRUE
   )

#Bind the two datasets
imp_long_MAR <- rbind(imp_long_AERx,imp_long_Asp)

#imp_long_MAR is sorted
imp_long_MAR <- imp_long_MAR[order(imp_long_MAR$.imp,imp_long
   _MAR$SUBJ_ID),]

#The last value before an eventually occurrence of the
   intercurrent event is
#carried forward to visit 13
for(i in 1:(length(visits)-1)){
  #Missing values at visit 13 are located
  na_visit_13 <- is.na(imp_long_MAR$VISIT_13)

  #The original data have to be unchanged, which is done by
  na_visit_13[1:dim(change_HbA1c_wide)[1]] <- FALSE

  #The last value before an eventually occurrence of the
     intercurrent event is
  #carried forward to visit 13 for all imputed datasets
  imp_long_MAR[na_visit_13,"VISIT_13"] <- imp_long_MAR[na_
     visit_13,visits[(length(visits)-i)]]
}

#Transform imp_long_MAR into a mids object
imp_MAR <- as.mids(imp_long_MAR)

#Analysis of each imputed dataset using an ANCOVA
ana_MAR <- with(imp_MAR, lm(VISIT_13~TREAT+BASELINE))

#The m analysis are pooled
pool_MAR <- pool(ana_MAR)

#The estimated treatment difference at visit 13 and std.
   error
est_MAR <- summary(pool_MAR)["TREAT2","est"]
se_MAR <- summary(pool_MAR)["TREAT2","se"]

```

```
#The degrees of freedom related to the treatment difference
df_MAR <-  summary(pool_MAR)["TREAT2","df"]

#Confidence interval
lower_ci_MAR <- summary(pool_MAR)["TREAT2","lo 95"]
upper_ci_MAR <- summary(pool_MAR)["TREAT2","hi 95"]

#Test for non-inferiority
t_MAR <- (est_MAR-non_inf_margin)/se_MAR
p_MAR <- pt(t_MAR,df_MAR) #OBS: pt() calculates P(X \leq x) by default

################################
####     Copy Reference     ####
################################

#Load functions to do MI using the copy reference approach
source("/Users/Kristoffer/Dropbox/Speciale/Latex/code/MI_copy_ref.R")

#Method
method <- ini$method
method[] <- ""
method[visits] <- "norm.extract"

#MI using copy reference
imp_copy_ref <- MI_copy_ref(change_HbA1c_wide, trt.var = "TREAT", copy.trt="AERx", ref.trt = "Insulin Aspart", meth = method, pred = pred.matrix, post = post.process, vis = vis.seq, where.copy = where_AERx, where.ref = where_Asp, M = m, seed = imp.seed)

#The imputed datasets
imp_long_copy_ref <- complete(imp_copy_ref, action="long", incl=TRUE)

#The last value before an eventually occurrence of the intercurrent event is
#carried forward to visit 13
for(i in 1:(length(visits)-1)){
  #Missing values at visit 13 are located
  na_visit_13 <- is.na(imp_long_copy_ref$VISIT_13)

  #The original data have to be unchanged, which is done by
  na_visit_13[1:dim(change_HbA1c_wide)[1]] <- FALSE

  #The last value before an eventually occurrence of the intercurrent event is
  #carried forward to visit 13 for all imputed datasets
```

```
158   imp_long_copy_ref[na_visit_13,"VISIT_13"] <- imp_long_copy_
        ref[na_visit_13,visits[(length(visits)-i)]]
159 }
160
161 #Transform imp_long_copy_ref into a mids object
162 imp_copy_ref <- as.mids(imp_long_copy_ref)
163
164 #Analysis of each imputed dataset using an ANCOVA
165 ana_copy <- with(imp_copy_ref, lm(VISIT_13~TREAT+BASELINE))
166
167 #The m analysis are pooled
168 pool_copy <- pool(ana_copy)
169
170 #The estimated treatment difference at visit 13 and std.
      error
171 est_copy <- summary(pool_copy)["TREAT2","est"]
172 se_copy <- summary(pool_copy)["TREAT2","se"]
173
174 #The degrees of freedom related to the treatment difference
175 df_copy <-  summary(pool_copy)["TREAT2","df"]
176
177 #Confidence interval
178 lower_ci_copy <- summary(pool_copy)["TREAT2","lo 95"]
179 upper_ci_copy <- summary(pool_copy)["TREAT2","hi 95"]
180
181 #Test for non-inferiority
182 t_copy <- (est_copy-non_inf_margin)/se_copy
183 p_copy <- pt(t_copy,df_copy) #OBS: pt() calculates P(X \leq x
      ) by default
184
185 ################################
186 ####    Jump to Reference   ####
187 ################################
188
189 #Load functions to do MI using the jump to reference approach
190 source("/Users/Kristoffer/Dropbox/Speciale/Latex/code/MI_jump
      _to_ref.R")
191
192 #MI using jump to reference
193 imp_jump_ref <- MI_jump_to_ref(change_HbA1c_wide, trt.var = "
      TREAT", jump.trt="AERx", ref.trt = "Insulin Aspart", meth
      = method, pred = pred.matrix, post = post.process, vis =
      vis.seq, where.jump = where_AERx, where.ref = where_Asp, M
       = m, seed = imp.seed)
194
195 #The imputed datasets
196 imp_long_jump_ref <- complete(imp_jump_ref, action="long",
      incl=TRUE)
197
```

```
198 #The last value before an eventually occurrence of the
    intercurrent event is
199 #carried forward to visit 13
200 for(i in 1:(length(visits)-1)){
201   #Missing values at visit 13 are located
202   na_visit_13 <- is.na(imp_long_jump_ref$VISIT_13)
203
204   #The original data have to be unchanged, which is done by
205   na_visit_13[1:dim(change_HbA1c_wide)[1]] <- FALSE
206
207   #The last value before an eventually occurrence of the
      intercurrent event is
208   #carried forward to visit 13 for all imputed datasets
209   imp_long_jump_ref[na_visit_13,"VISIT_13"] <- imp_long_jump_
      ref[na_visit_13,visits[(length(visits)-i)]]
210 }
211
212 #Transform imp_long_jump_ref into a mids object
213 imp_jump_ref <- as.mids(imp_long_jump_ref)
214
215 #Analysis of each imputed dataset using an ANCOVA
216 ana_jump <- with(imp_jump_ref, lm(VISIT_13~TREAT+BASELINE))
217
218 #The m analysis are pooled
219 pool_jump <- pool(ana_jump)
220
221 #The estimated treatment difference at visit 13 and std.
    error
222 est_jump <- summary(pool_jump)["TREAT2","est"]
223 se_jump <- summary(pool_jump)["TREAT2","se"]
224
225 #The degrees of freedom related to the treatment difference
226 df_jump <-  summary(pool_jump)["TREAT2","df"]
227
228 #Confidence interval
229 lower_ci_jump <- summary(pool_jump)["TREAT2","lo 95"]
230 upper_ci_jump <- summary(pool_jump)["TREAT2","hi 95"]
231
232 #Test for non-inferiority
233 t_jump <- (est_jump-non_inf_margin)/se_jump
234 p_jump <- pt(t_jump,df_jump) #OBS: pt() calculates P(X \leq x
    ) by default
235
236 ###################################
237 ####  Tipping Point Analysis  ####
238 ###################################
239
240 #Shift parameter
241 delta <- seq(-0.065,-0.067,-0.001)
```

```
242
243 #Vectors for estimates, CI and p values
244 est_tip <- c()
245 lower_ci_tip <- c()
246 upper_ci_tip <- c()
247 p_tip <- c()
248
249 for(k in 1:length(delta)){
250   cat(c("Delta[",k,"] \n"))
251
252   #The imputed datasets for TREAT=="AERx"
253   imp_long_AERx_tip <- imp_long_AERx
254
255   #Indicator of the imputed values at the last visit(The
        first rep() accounts for the original data)
256   imputed_last_visit <- c(rep(FALSE,length(where_AERx[,visits
        [length(visits)]])),unname(rep(where_AERx[,visits[length
        (visits)]],m)))
257
258   #Add delta[k] to the imputed values at the last visit
259   imp_long_AERx_tip[imputed_last_visit,visits[length(visits)
        ]] <- imp_long_AERx_tip[imputed_last_visit,visits[length
        (visits)]] + delta[k]
260
261   #Combine the imputed datasets across treatments
262   imp_long_tip <- rbind(imp_long_AERx_tip,imp_long_Asp)
263
264   #imp_long_tip is sorted
265   imp_long_tip <- imp_long_tip[order(imp_long_tip$.imp,imp_
        long_tip$SUBJ_ID),]
266
267   #The last value before an eventually occurrence of the
        intercurrent event is
268   #carried forward to visit 13
269   for(i in 1:(length(visits)-1)){
270     #Missing values at visit 13 are located
271     na_visit_13 <- is.na(imp_long_tip$VISIT_13)
272
273     #The original data have to be unchanged, which is done by
274     na_visit_13[1:dim(change_HbA1c_wide)[1]] <- FALSE
275
276     #The last value before an eventually occurrence of the
          intercurrent event is
277     #carried forward to visit 13 for all imputed datasets
278     imp_long_tip[na_visit_13,"VISIT_13"] <- imp_long_tip[na_
          visit_13,visits[(length(visits)-i)]]
279   }
280
281   #Transform imp_long_tip into a mids object
```

```
282   imp_tip <- as.mids(imp_long_tip)
283
284   #Analysis of each imputed dataset using an ANCOVA
285   ana_tip <- with(imp_tip, lm(VISIT_13~TREAT+BASELINE))
286
287   #The m analysis are pooled
288   pool_tip <- pool(ana_tip)
289
290   #The estimated treatment difference at visit 13 and std.
        error
291   est_tip[k] <- summary(pool_tip)["TREAT2","est"]
292   se_tip <- summary(pool_tip)["TREAT2","se"]
293
294   #The degrees of freedom related to the treatment difference
295   df_tip <-  summary(pool_tip)["TREAT2","df"]
296
297   #Confidence interval
298   lower_ci_tip[k] <- summary(pool_tip)["TREAT2","lo 95"]
299   upper_ci_tip[k] <- summary(pool_tip)["TREAT2","hi 95"]
300
301   #Test for non-inferiority
302   t_tip <- (est_tip[k]-non_inf_margin)/se_tip
303   p_tip[k] <- pt(t_tip,df_tip) #OBS: pt() calculates P(X \leq
        x) by default
304 }
305
306 ################################
307 ####      Forest plot       ####
308 ################################
309
310 #Directory for plots
311 setwd("/Users/Kristoffer/Dropbox/Speciale/Latex/figures")
312
313 tabletext <- list(c(list("", "Primary Analysis:",expression("
      MI"["ANCOVA"]^"MAR"),"Sensitivity Analyses:",expression("
      MI"["ANCOVA"]^"CR"),expression("MI"["ANCOVA"]^"J2R"),"
      Tipping Point Analysis:"),as.list(paste("delta =",delta)))
      ,
314                   c(list("Estimate",NA,round(est_MAR,3),NA,
                        round(est_copy,3),round(est_jump,3),NA),
                        as.list(round(est_tip,3))),
315                   c(list("95%-CI",NA,paste("[",round(lower_ci
                        _MAR,3),";",round(upper_ci_MAR,3),"]"),
                        NA,paste("[",round(lower_ci_copy,3),";",
                        round(upper_ci_copy,3),"]"),paste("[",
                        round(lower_ci_jump,3),";",round(upper_
                        ci_jump,3),"]"),NA),as.list(paste("[",
                        round(lower_ci_tip,3),";",round(upper_ci
                        _tip,3),"]"))),
```

```
316                   c(list("p-value",NA,round(p_MAR,4),NA,round
                        (p_copy,4),round(p_jump,4),NA),as.list(
                        round(p_tip,4))))
317
318 ci.data <- cbind(
319   mean  = c(NA, NA, est_MAR, NA, est_copy, est_jump, NA, est_
        tip),
320   lower = c(NA, NA, lower_ci_MAR, NA, lower_ci_copy, lower_ci
        _jump, NA, lower_ci_tip),
321   upper = c(NA, NA, upper_ci_MAR, NA, upper_ci_copy, upper_ci
        _jump, NA, upper_ci_tip))
322 colnames(ci.data) <- c("mean", "lower", "upper")
323
324 footnote <- expression(italic(atop("H"["0"]~":"~beta["T"
    ]>=0.4,"H"["A"]~":"~beta["T"]<0.4)))
325
326 #Export of forestplot
327 fileName.forestplot <- paste("forestplot_estimand_while_on_
    treatment_non_inf_0",10*non_inf_margin,"_nimp_",m,"_seed_"
    ,imp.seed,".png",sep="")
328 png(fileName.forestplot, width=1000, height=600, units="px",
    pointsize = 16)
329 forestplot(tabletext, hrzl_lines = list("2" = gpar(lty=1,lwd
    =3)), txt_gp = fpTxtGp(ticks=gpar(cex=1.2),xlab=gpar(cex
    =1.4),label=gpar(cex=1.3)), ci.data, is.summary=c(TRUE,
    TRUE,FALSE,TRUE,rep(FALSE,3),rep(FALSE,length(delta))),
    grid= structure(c(.4), gp = gpar(lty = 2, lwd=2.5, col = "
    red")), col=fpColors(box="royalblue",line="darkblue",
    summary="royalblue"), ci.vertices = T, boxsize = 0.25,
    xlab="AERx - Insulin Aspart",lwd.ci=2,xticks=seq
    (-0.1,0.5,0.1),title="Test for Non-Inferiority")
330 grid.text(footnote,unit(.015, 'npc'),unit(.075, 'npc'), just=
    c("left"), gp=gpar(fontsize=16, font = 3))
331 dev.off()
332
333 ################################
334 ####        Results         ####
335 ################################
336
337 estimates <- c(est_MAR, est_copy, est_jump, est_tip)
338 lower_ci <- c(lower_ci_MAR, lower_ci_copy, lower_ci_jump,
    lower_ci_tip)
339 upper_ci <- c(upper_ci_MAR, upper_ci_copy, upper_ci_jump,
    upper_ci_tip)
340 p_values <- c(p_MAR, p_copy, p_jump, p_tip)
341
342 results <- cbind(estimates, lower_ci, upper_ci, p_values)
343 names(results) <- c("Estimate", "Lower bound CI", "Upper
    bound CI", "P-value")
```

```
344
345 #Save results
346 fileName.results <- paste("results_estimand_while_on_
      treatment_non_inf_0",10*non_inf_margin,"_nimp_",m,"_seed_"
      ,imp.seed,".RData",sep="")
347 save(results, file=fileName.results)
```

Bibliography

Akacha, M., Bretz, F., and Ruberg, S. (2016). *Estimands in clinical trials - broadening the perspective. Statistics in Medicine*, 36(1):5–19.

Barnard, J. and Rubin, D. B. (1999). *Small-Sample Degrees of Freedom with Multiple Imputation. Biometrika*, 86(4):948–955.

Bishop, Y. M., Fienberg, S. E., and Holland, P. W. (2007). *Discrete Multivariate Analysis.* Springer.

Brown, H. and Prescott, R. (1999). *Applied Mixed Models in Medicine.* Wiley.

Carpenter, J. R. and Kenward, M. G. (2013). *Multiple Imputation and its Application.* Wiley, 1. edition.

Davis, C. S. (2002). *Statistical Methods for the Analysis of Repeated Measurements.* Springer.

Diggle, P. J., Liang, K.-Y., and Zeger, S. L. (1994). *Analysis of Longitudinal Data.* Oxford University Press.

EMA (2010). *Guideline on Missing Data in Confirmatory Clinical Trials.* `http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/09/WC500096793.pdf`.

EMA (2018). *Guideline on clinical investigation of medicinal products in the treatment or prevention of diabetes mellitus.* `http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2018/02/WC500243464.pdf`.

FDA (2008). *Guidance for Industry, Diabetes Mellitus: Developing Drugs and Therapeutic Biologics for Treatment and Prevention.* `https://www.fda.gov/downloads/Drugs/../Guidances/ucm071624.pdf`.

Harrell, F. E. (2001). *Regression Modeling Strategies.* Springer.

ICH (1998). *ICH Harmonised Tripartite Guideline, Statistical Principles for Clinical Trials E9.* `http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf`.

ICH (2000). *Choice of Control Group and Related Issues in Clinical Trials E10.* `http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E10/Step4/E10_Guideline.pdf`.

ICH (2017). *ICH Harmonised Guideline, Estimands and Sensitivity Analysis in Clinical Trials E9(R1).* `http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/E9-R1EWG_Step2_Guideline_2017_0616.pdf`.

LaMotte, L. R. (2007). *A Direct Derivation of the REML Likelihood Function. Statistical Papers*, 48(2):321–327.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data.* Wiley, 2. edition.

Lu, K. (2017). Number of imputations needed to stabilize estimated treatment difference in longitudinal data analysis. *Statistical Methods in Medical Research*, 26(2):674–690.

National Research Council (2010). *The Prevention and Treatment of Missing Data in Clinical Trials*. The National Academies Press.

Novo Nordisk (2008). *Novo Nordisk refocuses its activities within inhaled insulin and discontinues the development of AERx*. `https://www.novonordisk.com/content/Denmark/HQ/www-novonordisk-com/en_gb/home/media/news-details.1182031.html`.

Olofsson, P. and Andersson, M. (2012). *Probability, statistics, and stochastic processes*. Wiley, 2. edition.

Rubin, D. B. (2004). *Multiple Imputation for Nonresponse in Surveys*. Wiley.