Pattern-mixture sensitivity analysis in longitudinal trials with drop-out

George Vamvakas

Thesis submitted in accordance with the requirements for the degree of Master of Philosophy of the University of London

February 2015

Department of Medical Statistics
Faculty of Epidemiology and Population Health
LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Funded by: MRC Industrial CASE studentship

Contents

Abstract

The occurrence of missing data due to protocol deviations is inevitable in clinical trials. When missing data exist, analyses rely on assumptions about the behaviour of the individuals after dropping out. As a result, sensitivity analysis, which is now advocated by regulatory bodies, should be performed to explore the robustness of the inference to those assumptions. These assumptions should be relevant to the estimand of the study and be practically accessible by all parties.

The aim of this document is twofold: to assess the statistical validity of a new method for sensitivity analysis, and apply this method to a published Alzheimer's study. At the beginning of the thesis, a description of the Alzheimer's study and issues with missing data encountered therein, take place. This study was mainly set up to investigate the effect rosiglitazone, as an adjunct therapy in Alzheimer's patients. Two different doses of the drug were compared to placebo. The study suffered from a moderate degree of missing data in each treatment arm.

The thesis proceeds with a critique on the per-protocol and intention-to-treat estimands, and revisits their meaning when missing data occur. Two new estimands are introduced, which are particularly amenable to studies with missing data. They are termed *de-jure* and *de-facto*. Following that, the main methods for dealing with missing data are introduced, with a particular emphasis on multiple imputation, and how it can easily incorporate missing not at random (MNAR) analyses.

A thorough presentation of the new methodology is given. This is built around a set of assumptions, that reflect possible distributional behaviours of the subjects after protocol deviation. The assumptions are Randomised-treatment Missing at Random (MAR), Jump to Reference, Last Mean Carried Forward, Copy Increments in Reference, and Copy Reference. Estimation and inference is achieved via multiple imputation, and it is shown how the predictive distribution of the imputation model can be constructed from parameters borrowed from an MAR model and manipulated in a pattern-mixture model approach, to obtain the five assumptions for the

unobserved component of groups of individuals.

A number of simulations whose aim is to explore the statistical properties of the new method, are carried out. The simulated datasets, which are based on the parameters from the Alzheimer's study, focus on the estimator bias, the size and power of the methods, the bias of the variance estimator, and coverage. The results obtained from the simulations show the method has sensible properties; no bias for the estimator was detectable and the sizes and power of the methods agreed closely with their theoretical equivalents. The main result however, pertains to Rubin's variance estimator, which proves to appropriately reflect the loss of information from missing data. It is therefore argued, it is the right estimator to use in this setting.

The results from the application of the proposed method on the Alzheimer's dataset are presented in tables. Inferences from the sensitivity analysis assumptions were consistent with those from the original MAR analyses. The comparison between high dose rosiglitazone and placebo did not show any evidence in favour of the treatment under any sensitivity assumption. Some evidence of treatment difference existed when the low dose treatment was compared to placebo. This finding though, should be interpreted with caution, as the differences were obtained from analyses not subjected to the rigorous inferential process that was used in the Harringtons study. It was further argued, this finding might have been due to chance, and it was not replicated in a different study.

1 Introduction

Randomised clinical trials (RCTs) remain the gold standard method for exploring the efficacy, effectiveness and safety of medicinal products and interventions. Missing data usually occur in these studies either due to poor treatment compliance or withdrawal from the intervention or loss to follow-up. A number of approaches to tackle the problem of missing data exist, which can be broadly distinguished into simple *ad-hoc* methods and *principled* methods. Ad-hoc methods include i) the *completers* analysis, where only subjects with no missing values in any variable are retained, ii) imputation of simple mean, where missing values are replaced with the arithmetic average of the observed data for that variable, iii) regression imputation, which replaces the missing values with predicted values from a regression of the missing variable on complete variables, and iv) last observation carried forward (LOCF), where missing values are replaced by the last observed value of that variable. The principled methods include Maximum Likelihood (ML), Genearalised Estimating Equations (GEEs), Multiple Imputation (MI) and methods based on the Expectation-Maximisation (EM) algorithm and its derivatives.

Generally, the objective of statistical analyses is to make inferences that apply to the population targeted by the complete sample. With missing data the aim remains the same, but the methodology is generally more complex. Since, it is not possible to know the data of subjects after drop-out or the definitive reason of the data being missing only by looking at the observed data, the analyses relies inevitably on statistical assumptions about the behaviour of individuals after drop-out. As such, tackling the problem of missing data can be thought of as a two-step process. The first step is to make sensible assumptions about the distributional behaviour of the missing data. Then the second step would be to use an appropriate method to draw valid inferences under these assumptions. The process of making assumptions is therefore separate from the statistical methodology used for estimation and inference.

Note that the ad-hoc methods do not take this principled approach. Instead, they

create a single dataset, with no missing values, which is analysed as if it were the true complete dataset. Then they seek to justify the results, which is often not appropriate as they make strong restrictive assumptions which are hard to justify. On the other hand, the principled methods do not attempt to replace a missing value directly. They combine available information from the observed data with assumptions about the statistical distribution of the missing data and then employ a method that is inferentially valid under these assumptions.

In the presence of missing data, in order to explore the sensitivity of the inferences, a set of plausible assumptions about the missing data distribution need be formulated, and a flexible method for parameters estimation should be used, which can be applied under the main assumptions. The assumptions should be relevant and accessible to all interested parties. Multiple imputation can be used for parameter estimation, as it proves to be a sufficiently flexible approach. The need for sensitivity analysis is widely acknowledged by researchers (Kenward, 1998; Molenberghs, et al., 2004; Carpenter and Kenward, 2007; Carpenter et al., 2013; Daniels and Hogan, 2008) and regulatory bodies (ICH E9 Expert Working Group, 1999; CHMP, 2001).

Taking a parametric approach, three classes of models have been developed within which sensitivity analyses can be accommodated. These are: selection models (Diggle and Kenward, 1994), shared-parameters models (Wu and Carroll, 1998) and pattern-mixture models (Little, 1993; Little, 1994). With selection models, the data, either missing or not, are being weighted (or selected) through the probability of being observed. Selection models describe 'a unit's self-selection mechanism to either continue or leave the study' (Molenberghs and Kenward, 2007). They can be thought of as asking the question 'what is the probability of a subject missing in the next visit?'. The pattern-mixture models, where 'pattern' in this thesis will refer to a separate response distribution, allow for a different response model for each pattern of missingness. The observed data are a mixture of patterns, weighted by the proportion of missing data in each drop-out pattern (Molenberghs and Kenward, 2007). Similarly to pattern-mixture models, shared-parameter models allow for a different response model for each pattern of missingness, but they also introduce

latent variables, upon which the response and the drop-out pattern is conditionally independent (Molenberghs and Kenward, 2007). The mathematical forms of these models are presented in Chapter 3.5.

This thesis has the following aims: first, a recent proposal for sensitivity analysis by Carpenter, Roger and Kenward (CRK) (2013) will be demonstrated showing how it provides a natural route for sensitivity analysis in our multivariate dataset. Then, the properties of this method will be explored via simulation based on an Alzheimer's study and finally, the method will be applied to the Alzheimer's study and discuss the results.

1.1 Outline of the thesis

Chapter 2 introduces the Alzheimer's study carried out by GSK. Its primary analysis is described and the power calculations are presented. Chapter 3 develops the notation that will be used throughout the thesis, followed by a short review of Rubin's missing data taxonomy. In the same chapter the meaning of the most common clinical trial estimands is expounded, for both complete and missing data analyses. In addition, the main methods for the analysis of missing data in longitudinal studies are described. The chapter concludes with a review of sensitivity analysis models, and it makes a justificantion for the use of pattern mixture models in this setting.

In Chapter 4, CRK's proposal for sensitivity analysis is formally illustrated, with some analytical details on five different sensitivity assumptions. Chapter 5 includes simulation studies for the evaluation of the proposed method, as well as a discussion of the results. In Chapter 6 the Alzheimer's study is revised with a view to identifying the predictors of withdrawal and the study is analysed under the proposed sensitivity analysis method in Chapter 7. There, the results from the analyses are tabulated and contrasted with those from the original study. Finally, in Chapter 8, a concluding discussion takes place around the practical and methodological points raised in the thesis.

2 Alzheimer's disease study

In this Chapter, a Randomised Controlled Trial on patients with Alzheimer's disease is introduced to motivate the material. This study has been selected because it is very illustrative for the kinds of data problems with which this thesis is concerned and because it is reanalysed using the proposed method for sensitivity analysis. The disease, the primary research goals and outcomes, and the key complications presented by missing data are described in this chapter.

The disease

Alzheimer's disease (AD) is a physical condition which changes the way brain works. Patients suffering from the disease develop protein 'plaques' and 'tangles' in the structure of the brain, leading to the death of brain cells. Also, a general lack of the acetylcholine chemical in the brain means nerve messages aren't passed on properly. Over time, as more and more areas of the brain become damaged, the symptoms of the disease get progressively worse. Clinical symptoms include confusion and forgetfulness, mood swings and difficulty carrying out everyday activities. No single cause of the disease has been identified. A combination of factors such as age, genes, environmental factors, lifestyle and overall general health is likely to be conducive to the development of the disease (alzheimers.org.uk).

With respect to genetics, apolipoprotein E (APOE) allele plays an important role in AD. A number of studies confirmed the pivotal role of the allele as a strong genetic risk factor for AD (Corder E.H., et al. (1993); Strittmatter W.J., et al. (1993)). Out of all different combinations of the APOE copies, people with two $\varepsilon 4$ alleles have the highest chance to develop AD, with up to 20 times the risk compared to other combinations. The $\varepsilon 3/\varepsilon 4$ genotype is at increased risk, albeit not to the extend of those with $\varepsilon 4/\varepsilon 4$ are. The genotype $\varepsilon 2/\varepsilon 3$ is considered at lower risk for AD, and people with $\varepsilon 3/\varepsilon 3$ and $\varepsilon 2/\varepsilon 4$ are at normal risk (Blennow K., de Leon M.J., Zetterberg H. (2006)).

Currently, there is no cure for AD and all treatments are directed at alleviating some

of the symptoms or slowing down the disease progression. In the U.K., the National Institute for Health and Clinical Excellence recommends the drugs donepezil, galantamine and rivastigmine as an option for people in the mild-to-moderate stages of the disease. They work by maintaining existing supplies of acetylcholine. For people in moderate and severe stages the only drug recommended is memantine (nice.org.uk/ta217). These drugs may stabilise some of the symptoms for a limited period, typically 6-12 months or longer (alzheimers.org.uk).

The study

The data analysed in this thesis come from the AVA102672 GSK study, which was a published Phase III, multi-national, multi-center, randomised, double-blind, placebo controlled, parallel-group clinical trial, investigating the efficacy and safety of rosiglitazone extended release tablets (RSG XR) as adjunctive therapy in mild to moderate AD patients already being treated with donepezil. The motivation behind this study was a 'pharmacogenetic sub-group analysis which detected a statistically significant interaction between RSG treatment and APOE $\varepsilon 4$ allele status, suggesting that RSG may be effective in a genetically defined subset of individuals with mild-to-moderate AD (Harrington et al., 2011).

The primary research objectives of the study were a) to investigate the add-on effects of daily dosing for 48 weeks with RSG XR versus placebo on cognitive function in donepezil-treated subjects with mild to moderate Alzheimers disease, as a function of APOE 4 status and b) to investigate the add-on effects of daily dosing for 48 weeks with RSG XR versus placebo on overall clinical response in donepezil-treated subjects with mild to moderate Alzheimers disease, as a function of APOE 4 status.

Eligible subjects for the study were those between 50 and 90 years old, who were diagnosed with mild to moderate AD and had a Mini-Mental State Examination (MMSE) score (Folstein et al. (1975)) of 10 to 26 at screening. Also, subjects should have received at least 6 months of ongoing donepezil for AD, with stable dosing for at least 2 months prior to enrollment.

The primary outcomes of the study were two. The first outcome was change from baseline in Alzheimers Disease Assessment Scale cognitive subscale (ADAS-Cog) total score, and the second outcome was change from baseline in Clinical Dementia Rating scale - Sum of Boxes (CDR-SB).

ADAS-cog is the standard primary neuropsychological measure for AD trials covering several cognitive domains, including memory, language, and praxis (Rockwood et al., 2007). Total scores in the ADAS-Cog test range from 0 to 70, with higher scores ($\geq$ 18) indicating greater cognitive impairment (Rosen et al., 1984). The CDR-SB score is another instrument widely used in AD trials as a global measure of disease progression. The scores range from 0 to 18 with higher scores indicating more impairment (Coley et al. 2011). Currently, there is no consensus as to the choice of the most appropriate outcomes instrument in AD trials (Coley et al. 2011). ADAS-Cog is the most commonly used instrument, but its known shortcomings (insensitivity to very mild impairments, difficulties in determining the clinical relevance of changes, high within-subject variability resulting in large sample sizes (Coley et al. (2011)) has led some investigators examine other potential candidate tools for measurement in dementia. Coley et al. 2011 after comparing ADAS-Cog and CDR-SB on structural and convergent validity, responsiveness and sample size, and data quality, asserted that CDR-SB is 'a promising candidate for a sole primary endpoint for AD trials'. All analyses of Alzheimer's data in this document concentrate on the ADAS-Cog outcome measure only.

The actual duration of the study was 58 weeks; this comprised a 4-week screening phase, a 48-week double-blind treatment phase and a 6-week single-blind placebo treatment phase, at the end of which all patients completing the 48-week double-blind period were admitted receiving placebo daily as well as their regular donepezil regime.

The scores from the ADAS-Cog tool were used as continuous variables. Under the 48-week double-blind phase, which formed the primary phase for analyses, there were 6 time points: baseline, and week 8, 16, 24, 36, and 48. The two primary

comparisons of interest were the effect of RSG at 2mg and 8mg compared to placebo, on both ADAS-Cog and CDR-SB, at the end of the study treatment for the 'Full population'. The Full population consisted of all subjects randomised to treatment, who had taken at least one dose of study medication and who had at least one post baseline efficacy assessment. Comparisons were also conducted for two further populations. These were: 1) APOE $\varepsilon 4$ negative subjects (i.e., subjects with $\varepsilon 2/\varepsilon 2$, $\varepsilon 2/\varepsilon 3$ or $\varepsilon 3/\varepsilon 3$ genotype) and 2) all subjects except $\varepsilon 4/\varepsilon 4$'s, which comprised APOE $\varepsilon 4$ negative subjects and $\varepsilon 4$ heterozygote subjects.

In order to be able to determine effects at the APOE $\varepsilon 4$ stratum level the sample size calculations were based on powering at this level. Initially, within an individual APOE stratum, 522 evaluable subjects (174 per treatment arm) were required to provide a 90% power to detect a difference of 2 points between placebo and active treatment in the change from baseline in ADAS-Cog score. The significance level was set at 5% assuming a standard deviation (SD) of 5.74. However, a drop-out rate of 10% was allowed, which meant 193 subjects per treatment arm per stratum, or 579 subjects within each stratum. The final eligible sample size in the study ended up being 1496 subjects. This was obtained because the rate of recruitment in the two APOE strata did not stop until both strata reached the target of 579 subjects.

Change from baseline in ADAS-Cog and CDR-SB were analysed using a mixed model for repeated measures (MMRM) with restricted maximum likelihood estimation and an unstructured covariance matrix. Inferences were based in least squares mean change from baseline at week 48, obtained from the MMRM model. For the APOE $\varepsilon 4$ negative subgroup, the MMRM model included treatment, visit, treatment by visit interaction, country, and continuous fixed covariates of baseline endpoint score, baseline endpoint score by visit, screening Mini Mental State Examination (MMSE), screening MMSE by visit, and baseline BMI (the MMSE was also used as an outcome in secondary analyses; it ranges from 0-30, with higher values indicating less cognitive impairment). Analyses for the APOE negatives combined with $\varepsilon 4$ heterozygotes and Full populations also utilised the above MMRM model structure with an addition of a fixed categorical covariate APOE $\varepsilon 4$ gene copies ($\varepsilon 4$ negatives,

$\varepsilon 4$ heterozygotes, $\varepsilon 4$ homozygotes).

Based on the study's protocol, in order to proceed and draw inferences for a comparison between the low dose treatment and placebo, a statistical significance between the high dose treatment and placebo should have been observed first, on both co-primary measures. Given that, a hierarchical testing procedure was implemented to preserve the Type I error, with reduced α levels, for the individual tests. The procedure started by examining the efficacy of 8mg RSG XR first, in the 'All except $\varepsilon 4/\varepsilon 4$ genotype' and 'APOE $\varepsilon 4$ negative' populations, with an α level set at 1% and 4%, respectively.

2.1 Missing data in the Alzheimer study

This section presents the patterns of missing data encountered in the Full population in the Alzheimer's study, for the double-blind phase, that formed the primary analysis in the GSK study. Data were missing because patients withdrew, giving rise to a monotonic pattern of missingness, with a few instances of non-monotone missing data, as shown below. Monotone missing data imply that if a subject drops out at a given time point, all responses up until drop out are observed and all their subsequent responses are missing.

Monotone missing data for *placebo*						Number of subjects (%)
X	X	X	X	X	X	357 (71.4)
X	X	X	X	X	·	25 (5.03)
X	X	X	X	·	·	26 (5.23)
X	X	X	·	·	·	21 (4.23)
X	X	·	·	·	·	14 (2.82)
X	·	·	·	·	·	37 (7.44)
·	·	·	·	·	·	1 (0.20)

Non-monotone missing data for *placebo*						Number of subjects (%)
X	X	X	X	·	X	2 (0.40)
X	X	X	·	X	X	3 (0.60)
X	X	·	X	X	X	2 (0.40)
X	·	X	X	X	X	2 (0.40)
X	X	X	·	X	·	1 (0.20)
X	·	X	X	·	·	1 (0.20)
X	·	·	X	X	X	1 (0.20)
X	·	·	X	·	X	1 (0.20)
X	·	·	X	·	·	1 (0.20)
X	X	X	·	·	X	1 (0.20)
X	X	·	X	X	·	1 (0.20)

Fig 1: Monotone and non-monotone missing data patterns for Placebo

Monotone missing data for *2mg RSG XR*						Number of subjects (%)
X	X	X	X	X	X	387 (77.86)
X	X	X	X	X	·	15 (3.04)
X	X	X	X	·	·	25 (5.06)
X	X	X	·	·	·	14 (2.83)
X	X	·	·	·	·	18 (3.64)
X	·	·	·	·	·	19 (3.85)
·	·	·	·	·	·	1 (0.20)

Non-monotone missing data for *2mg RSG XR*						Number of subjects (%)
X	X	X	X	·	X	3 (0.61)
X	X	X	·	X	X	4 (0.81)
X	X	·	X	X	X	2 (0.40)
X	·	X	X	X	X	4 (0.81)
X	X	X	·	·	X	1 (0.20)
X	X	·	X	X	·	1 (0.20)

Fig 2: Monotone and non-monotone missing data patterns for 2mg RSG XR

Monotone missing data for *8mg RSG XR*						Number of subjects (%)
X	X	X	X	X	X	335 (67.13)
X	X	X	X	X	·	31 (6.24)
X	X	X	X	·	·	28 (5.63)
X	X	X	·	·	·	23 (4.63)
X	X	·	·	·	·	29 (5.84)
X	·	·	·	·	·	36 (7.24)
·	·	·	·	·	·	1 (0.20)

Non-monotone missing data for *8mg RSG XR*						Number of subjects (%)
X	X	X	X	·	X	1 (0.20)
X	X	X	·	X	X	4 (0.80)
X	X	·	X	X	X	1 (0.20)
X	·	X	X	X	X	4 (0.80)
X	X	X	·	·	X	1 (0.20)
X	X	·	X	·	·	1 (0.20)
X	·	X	·	·	·	1 (0.20)
X	X	·	X	X	·	1 (0.20)

Fig 3: Monotone and non-monotone missing data patterns for 8mg RSG XR

At the start of the trial, there were 500, 497 and 499 patients randomised to placebo, 2mg treatment and 8mg treatment, respectively. By the final time point (week 48), only 357 (71.4%) remained in the placebo group, 387 (77.86%) in the 2mg group, and 335 (67.13%) in the 8mg group. Hence, by the end of the trial, the 2mg RSG XR arm had the lowest proportion of missing data and the 8mg RSG XR arm the highest.

The trellis plots below display the patterns for all patients with monotone missing data in each treatment group. The first pattern of monotone missing data, shown in the upper left plot, display those patients for whom only baseline measurements are observed; that is, they dropped out after baseline. The second pattern, which corresponds to the upper middle plot, show those patients that dropped out after time 1. The same logic applies to the rest of the patterns. Overall, with six time points, there were six patterns of missing data for each treatment group. The bold line represents the group average for each pattern.

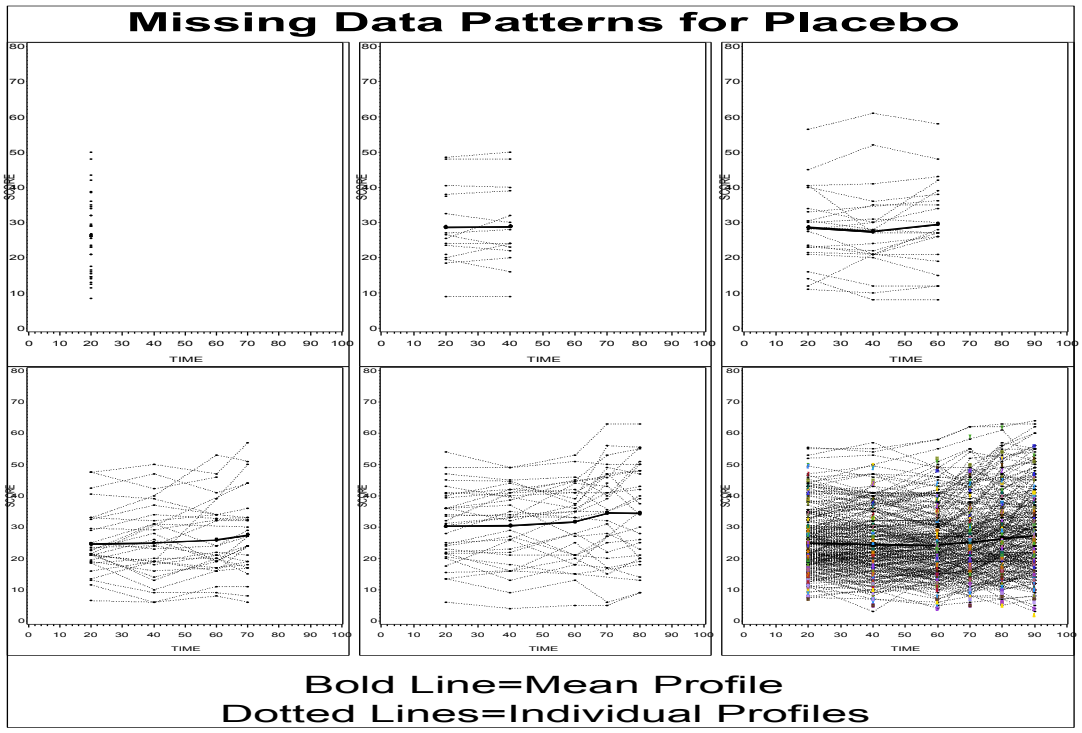

Fig 4: Monotone missing data patterns in the Placebo arm.

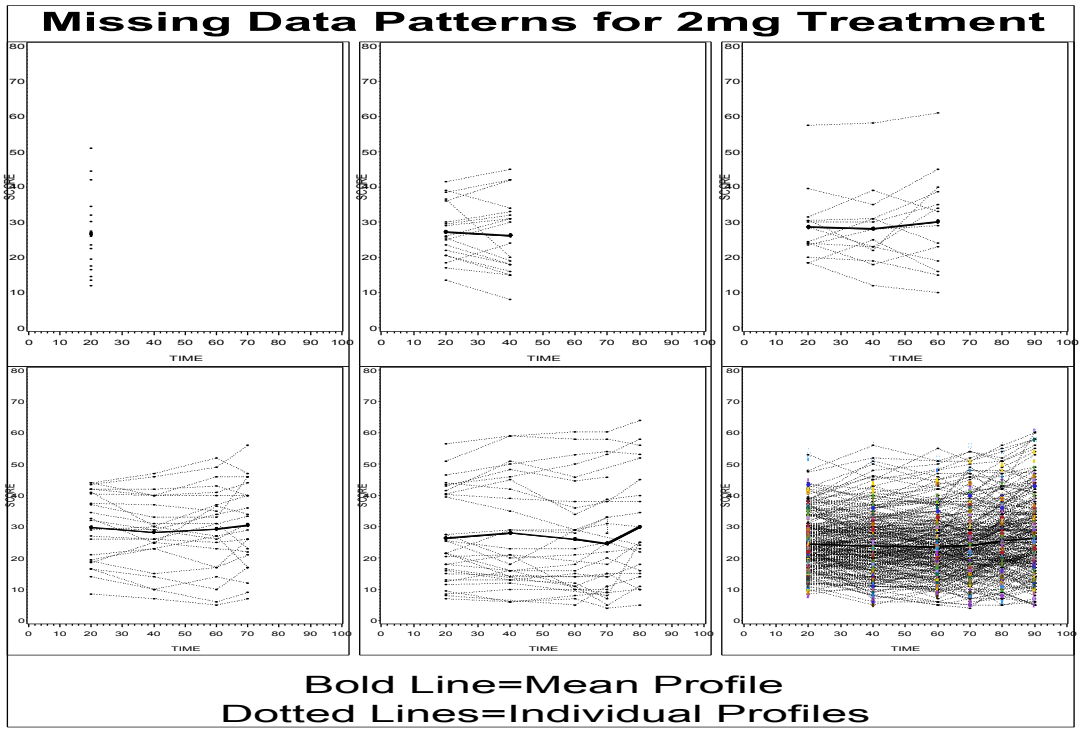

Fig 5: Monotone missing data patterns in the 2mg RSG XR arm.

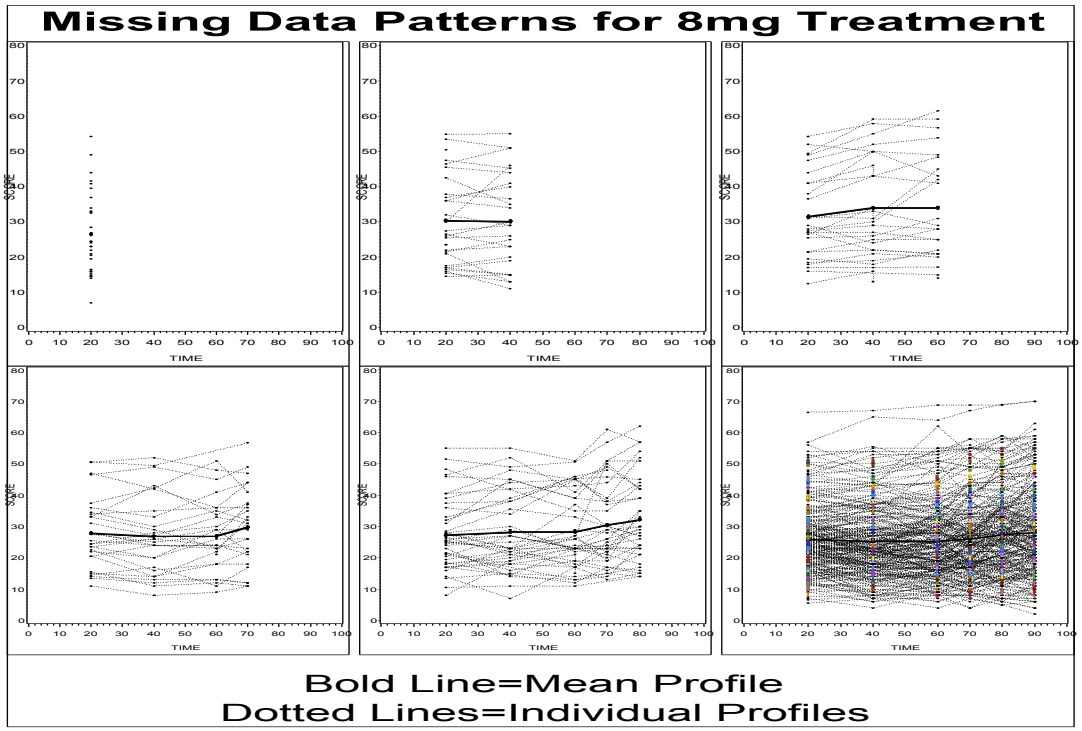

Fig 6: Monotone missing data patterns in the 8mg RSG XR arm.

In Figure 4, where the monotone patterns for the placebo arm are displayed, subjects exhibit a similar degree of variability between their baseline scores and those at the final time points, across all patterns of missing data. The general impression from

the patterns is that the patients' condition remained relatively stable; the mean profiles seem to be following a straight line too. In Figure 5, no discernible effect of the treatment can be asserted. The picture is roughly the same, as in the previous trellis plot, with respect to the means and variability between starting and final time points. Even at higher doses a treatment effect cannot be ascertained. In Figure 6, by examining the means and comparing the variability of the scores at the final time point to that at the start, is that patients' condition remained stable.

If the data were complete, an ANCOVA model could be used to estimate treatment effects at the final time point, with treatment as the class variable and baseline as the covariate. Inferences from the ANCOVA model however cannot be justified, unless some assumptions about the missing responses are made; for example, unless the missing data are missing at random given baseline and treatment, analysing the dataset as if it were complete would lead to wrong inferences. So, what would an analysis with an ANCOVA model mean? Would it reflect what is seen in practice? It is true, that patients in real life after stopping treatment, might seek out and receive alternative therapeutic regimes, or cease to receive treatment altogether, or maybe carry on with the same treatment. As we shall see later, more sophisticated analyses are needed to be put in place that would capture this reality. These analyses would allow for a careful consideration of plausible assumptions about the distributional behaviour of the subjects after drop-out.

2.2 Summary

The chapter describes a published trial on Alzheimer's disease that is used to motivate this thesis. Alzheimer's is a chronic disease, and affects how brain works. A number of risk factors are associated with the disease, such as age, lifestyle, overall health, and the APOE allele, with homozygous people for $\varepsilon 4$ at increased risk. Currently, there is no cure for AD. The trial looked at the efficacy of daily intake of RSG XR, as an adjunctive therapy, in people with mild to moderate Alzheimer's, as a means to slow down progression. The effect of two doses, a low and a high dose

of RSG XR, were evaluated against placebo and the study was stratified by APOE allele status. The stratification was done because there was evidence of drug efficacy in a genetic subset of patients. The primary outcomes were change from baseline in ADAS-Cog and CDR-SB at week 48, and measurements were taken at 6 time points. The main tool for conducting the comparisons was an MMRM model. The final sample size of the study was 1496 individuals, and there was approximately 20% to 30% drop-out in each treatment arm by the final time point.

3 Background theory and methods

As stated earlier, incomplete data issues complicate the analysis of clinical and other studies. In this chapter, the main estimands used in clinical trials and a number of methods for dealing with the problem of missing data will be discussed. Two new estimands will be presented which have been developed specifically to address the confusion over what is estimated when a trial suffers from missing data. In later chapters, we will see how methods presented here can come together to tackle different assumptions about the unobserved values.

3.1 Notation

The analyses described herein assume the existence of monotone missing longitudinal data patterns, which is also referred to as missing data with drop-out. Monotone missing data implies that the absence of a measurement at time $t = 0, ..., T$, with $t = 0$ denoting time at baseline measurement, implies that all subsequent measurements for subject i are absent. Therefore, for the special case of longitudinal data with drop-out, let $D_i = 0, ..., d_i, ..., T_i$, denote the last observed time before subject i drops out. It is also possible for a subject to drop-out at the final time point, that is $d_i = T_i$. It is assumed that baseline is always observed. Hence, for $D_i = 0$, all measurements for the specific subject are missing except baseline. Let $\mathbf{y}_i = (y_{i1}, ..., y_{iT})'$ denote the full $T \times 1$ vector of responses for subject i. With monotone missing data, $\mathbf{y}_i$ can be separated into $\mathbf{y}_{i,o} = (y_{i,0}, ..., y_{i,d_i})'$ and $\mathbf{y}_{i,m} = (y_{i,d_i+1}, ..., y_{i,T_i})'$, where the first d_i responses are observed and the rest $T_i - d_i$ are missing. Also, $\mathbf{y}_O = (\mathbf{y}'_{1,o}, ..., \mathbf{y}'_{n,o})'$ and $\mathbf{y}_M = (\mathbf{y}'_{1,m}, ..., \mathbf{y}'_{n,m})'$ represent the vectors of all observed and missing data, respectively. Finally, let $\mathbf{g} = \{g_i = j\}$ denote the group of treatment for subject i in treatment group j, $\boldsymbol{\theta}$ a parameter vector that describes the measurement process and $\boldsymbol{\phi}$ a parameter vector that describes the drop-out process. Finally, let z_{ik} be the i^{th} subject's value of the k^{th} baseline covariate.

3.2 The drop-out mechanisms

Little and Rubin (2002) classified 3 distinct mechanisms for missing data: *missing completely at random* (MCAR), *missing at random* (MAR) and *missing not at random* (MNAR). These mechanisms were first developed in a survey setting. By analogy, in a longitudinal data setting, with a fully observed matrix of baseline covariates Z_k, Verbeke and Molenberghs (2000) show that, MCAR is expressed as $f(D_i|\mathbf{y}_{i,o}, \mathbf{y}_{i,m}, \mathbf{Z}_{i,k}, \boldsymbol{\phi}) = f(D_i|\boldsymbol{\phi})$, where the probability of a subject dropping out is a constant. A different way to think of this mechanism, is the probability of a subject dropping out is independent of all individual characteristics $\mathbf{y}_i$,$\mathbf{Z}_{i,k}$. MAR is expressed as $f(D_i|\mathbf{y}_{i,o}, \mathbf{y}_{i,m}, \mathbf{Z}_i, \boldsymbol{\phi}) = f(D_i|\mathbf{y}_{i,o}, \mathbf{Z}_i, \boldsymbol{\phi})$, where the probability of drop-out for subject i may depend on data observed prior to drop-out but, conditionally on these observed data, not on the unobserved data. Hence, the probability of drop-out is conditionally independent of the unobserved component $\mathbf{y}_{i,m}$, after having conditioned on the observed $\mathbf{y}_{i,o}$ Finally, if neither of the previous assumptions hold, then $f(D_i|\mathbf{y}_{i,o}, \mathbf{y}_{i,m}, \mathbf{Z}_i, \boldsymbol{\phi})$ does not simplify; the probability of drop-out depends on the missing data, even after having conditioned on the observed data. The data in this case are said to be MNAR.

3.3 Common clinical trial estimands

In general, according to Rubin (1996) the term *estimand* can be defined as "the quantity of scientific interest that can be calculated in the population and does not change its value depending on the data collection design used to measure it". Careful description of the population when defining an estimand should be made.

A 'per-protocol' estimand seeks to estimate the effect of a new treatment only from participants who conformed with the protocol requirements. These patients form the per-protocol population. The analysis is typically restricted only to those subjects who adhered to the clinical trial terms, such as eligibility, interventions, and outcome assessment as stipulated in the protocol (National Research Council, 2010). According to published guidelines (ICH E9 Expert Working Group, 1999) the use

of this set of population, "may maximize the opportunity for a new treatment to show additional efficacy in the analysis, and most closely reflects the scientific model underlying the protocol". On the negative side however, exclusion of patients and restriction of analyses to the per-protocol population hampers the examination of the drug's practical value. Also, it may cause severe biases, as adherence of the subjects to the study protocol may be related to treatment assignment and outcome.

Contrary to the per-protocol analyses, the intention-to-treat (ITT) analysis describes the principle of analysing data from *all* participants randomised to treatment, irrespective of the level of treatment received or protocol adherence (Hill, 1961). Little and Yau (1996), argues that an ITT analysis has two principles. Firstly, all randomised subjects should be included in the analysis. Secondly, subjects should be analysed as part of the treatment group to which they were originally assigned to, no matter if they actually received a different treatment during the course of the trial. The population under the ITT principle is called the ITT population. Ideally, an ITT analysis aims to analyse the full set of individuals enrolled in the study. Sometimes though, failure to satisfy major entry criteria (eligibility violations), or take at least one dose of trial medication coupled with the lack of any data post randomisation, may sometimes lead to exclusion of patients from the ITT population. No analysis should be considered complete however, unless all biases arising from these exclusions are addressed. For instance, in the case patients were excluded due to failure to take any trial medication, it can be argued that the ITT principle can still be valid if this was not influenced by knowledge of the allocated treatment (ICH E9 Expert Working Group, 1999).

Another type of analysis the makes use of the full set of randomised individuals in a study is the as-treated analysis. Under an as-treated estimand, all randomised subjects are included in the analysis, but these subjects, are grouped and analysed according to the treatment they actually received, even if the initial randomisation arm was different (Piantadosi's, 1997). By doing so, an as-treated analysis does not preserve the initial randomisation of patients. As a result, selection bias in the treatment effect can be introduced; the treatment groups can be systematically dif-

ferent and hence no longer directly comparable, leading to the emergence of spurious effects.

The rationale behind an ITT analysis is to preserve randomisation: groups must be alike in all important aspects and only differ with respect to treatment. Tsiatis (1990) explains that with randomisation, comparable treatment groups that are on average similar, can be constructed, since prognostic factors will be averaged out in the two different groups by the chance mechanism. Therefore, maintenance of the initial randomisation would prevent bias. As Little and Yau (1996) point out, contrary to the as-treated analysis where, when subjects take a different treatment from that assigned to is more prone to selection bias, an ITT analysis can mirror what actually occurs in clinical practice. For example, if a very effective drug is avoided by patients due to side effects, this negative feature is taken into account by an ITT analysis. An as-treated analysis, by focusing only on subjects who took the treatment ignores this fact (Little and Yau, 1996). The as-treated estimand will not be considered in this document, as selection bias and the methodology for dealing with it, is a different topic.

Carpenter et al. (2013) embody the principles of the per-protocol and ITT analyses into two new estimands. The *de-jure* estimand and the *de-facto* estimand. They use the term *de-jure*, to address questions such as "what the expected treatment effect be in the target population of eligible patients if the treatment and control were taken as specified in the protocol". Under this estimand it is assumed that the targeted population commits no protocol deviation either at baseline or post-randomisation. On the other hand, they term *de-facto* an estimand that seeks to answer questions about"what would be the effect seen in practice if this treatment was assigned to the target population of eligible patients as defined by the trial inclusion criteria". A de-facto estimand attempts to represent actual practice that is observed in clinical trials.

As it will become obvious in the next section, these estimands are amenable to situations where drop-out exists.

3.3.1 Implications for the occurence of missing data

To begin with, an implication of the MAR assumption needs to be stated. Molenberghs et al. (1998) proved that if the MAR assumption holds, within the context of longitudinal data with drop-out, then:

$$f(y_{i,d+1}|y_{i,0},...,y_{i,d},D=d)=f(y_{i,d+1}|y_{i,0},...,y_{i,d},D\geq d+1) \tag{1}$$

hence, the conditional distribution of the missing responses Y_{d+1} given the observed past responses, for those whose follow-up terminates at some time prior to $d+1$, that is $D<d+1$, can be estimated from the conditional distribution of individuals who have observed responses at or beyond $d+1$.

Equation (1) implicitly assumes that the conditional distribution of the subjects with missing responses is equivalent to the conditional distribution of the subjects with observed responses who share the same 'history', where history typically includes the treatment group. For instance, let us assume a clinical trial with two treatment groups g_1 and g_2, with responses measured across 3 time points. Further, assume a subset of individuals in group 1 drops out after time 2. Then the conditional distribution of the drop-outs, that is $f(y_{i,3}|y_{i,0},y_{i,1},y_{i,2},g_{i,1},D=2)$ can be estimated by the conditional distribution of the completers who have the same $y_{i,0}$, $y_{i,1}$, $y_{i,2}$ and $g_{i,1}$ values, that is $f(y_{i,3}|y_{i,0},y_{i,1},y_{i,2},g_{i,1},D=3)$. In practice, hardly ever will two patients have the same $y_{i,0},y_{i,1}$, and $y_{i,2}$ values, and therefore a statistical model is needed to estimate the conditional distributions.

The MAR assumption is sensible when we wish to answer per-protocol questions, that is, to estimate the distribution of the subjects' responses assuming they continued to adhere to the protocol (Carpenter and Kenward, 2007), since, under MAR, the future statistical behaviour of a subject, conditional on the history, is the same whether the subject drops out or not in the future. This assumes that all data prior to drop-out are from patients who are complying with the protocol.

As previously stated, with no missing data, the ITT principle dictates the inclusion

of all subjects, irrespective of protocol adherence. Consequently, if data are available on subjects after protocol violation, then an ITT analysis can still be conducted. When drop-out exists however, ITT requires a method of dealing with the missing data in order to preserve the principle of analysing the full set of randomised individuals. According to Nich and Carroll (2002), and Hollis and Campbell (1999), in the presence of missing data, evidence of confusion over the directives and assumptions governing an ITT analysis appears to be widespread in the literature. These authors concluded the ITT approach was often inadequately described and inadequately applied, with the handling of missing data being the main problem.

It is true subjects that drop out might seek out and receive a treatment other than that stated in their protocol. This imply a move away from the MAR assumption, since according to Carpenter and Kenward (2007), the outcomes for patients who discontinue treatment are likely to be different from those who continue with their randomised treatment. Thus an ITT analysis, when a non-trivial proportion of patients discontinue treatment and we don't have outcome data for them after drop-out, may well imply an MNAR missingness mechanism: the probability of drop-out depends on the missing data, even after having noted the initial randomised treatment, as well as baseline covariates and pre-drop-out data. There can be no definitive ITT analysis, since assumptions should be put in place about the post-treatment/post-withdrawal data. Thus in practice, we wish to make a primary assumption and explore robustness of inferences to other plausible assumptions. For this reason, some authors argue that an ITT analysis with drop-out, should be seen in the context of a sensitivity analysis (Little and Yau, 1996; Heumann, 2000; Carpenter and Kenward, 2007; Carpenter et al., 2013). Little and Yau (1996) and Carpenter et al. (2013) developed such methodology.

To avoid confusion of what is meant by per-protocol and ITT estimands when drop-out exists, the new set of estimands, that is de-jure and de-facto can be used. The mathematical definitions of these are: let $\widetilde{f}_{act}$ denote the joint probability distribution (pdf) of baseline and post-randomisation responses for patients randomised to the *active* treatment, without deviating from protocol, and analogously $\widetilde{f}_{cont}$ for

patients who receive the control treatment. Then a de-jure estimand is defined as:

$$E_{\widetilde{f}_{act}}(g(\mathbf{Y})) - E_{\widetilde{f}_{cont}}(g(\mathbf{Y}))$$

where $g(.)$ is any suitable function and $\mathbf{Y}$ is the response profile. In this thesis, $g(y)$ refers to the last scheduled observation, so that if the observed data come from $\widetilde{f}_{act}$ and $\widetilde{f}_{cont}$, then we can use regression of the final response on treatment group and baseline to estimate this quantity (Carpenter et al. 2013).

The mathematical definition of a de-facto estimand is given by:

$$E_{f_{act}}(g(\mathbf{Y})) - E_{f_{cont}}(g(\mathbf{Y}))$$

where now f_{act} and f_{cont} refer to the pdfs "of baseline and post-randomisation responses that would be seen in the context of interest among patients randomised to the active arm" (Carpenter et al. 2013).

Under the de-facto estimand deviation usually includes instances of unblinding and loss to follow-up, but not moving to partial compliance with treatment and withdrawal from treatment. Under the de-jure estimand, by deviation is typically meant any instance of unblinding, for example of treatment allocation, moving to partial compliance with treatment, withdrawal from treatment following an adverse event and loss to follow-up.

In practice, patients after deviation will either switch treatments and continued to be followed-up or withdraw entirely with no further contribution of data. As will become obvious when the formal presentation of the new sensitivity method takes place, to address a de-jure or a de-facto estimand, knowledge of which treatment is being used after deviation and the reason why a patient withdraw can be readily used.

Carpenter et al. (2013) also explain, through the use of the proposed estimands, confusion over a per-protocol estimand, which emanates from the fact that is sometimes associated with an 'on-treatment' estimand which, in turn, does not specify

the extent of treatment compliance, can be avoided. Similarly, uncertainty caused by the use of the ITT estimand, which sometimes alludes to a specific population or an estimation method, can be prevented.

3.4 Methods for analysing longitudinal studies with missing data

3.4.1 Simple methods

This section presents some basic points on simple methods for dealing with missing data. A thorough discussion of such simple methods, can be found in a monograph by Carpenter and Kenward (2007) and downloaded from www.missingdata.org.uk.

One of the most commonly used imputation methods in longitudinal datasets with drop-out is the LOCF (Kenward and Molenberghs, 2009). It replaces missing values with the last observed value. There is widespread agreement amongst researchers LOCF should be avoided, as it is prone to bias, and distorts the variance and correlation structures (Yau and Little, 1996; Carpenter and Kenward, 2007; Mallinckrodt, et al., 2003; Mallinckrodt, et al., 2004; Lane, 2008, Carpenter, et al., 2004; Beunckens et al., 2005; Verbeke and Molenberghs, 2000; Molenberghs et al., 2004; Siddiqui et al., 2009). As Kenward and Molenberghs (2009) show, this method, that gained prominence due to its simplicity, can only be justified under stringent assumptions (see also Shao and Zhong, 2003 and Carpenter et al., 2004).

All simple methods, apart from Completers Analysis, allow all cases to be analysed, and the treatment effects are estimated with cases assigned to their randomised arm. As noted by many authors, they suffer from severe drawbacks (Carpenter and Kenward, 2007; Little and Rubin, 2002). The main process under which simple methods operate is: they complete the dataset first, and then look at the assumption the method makes. It turns out in the case of such methods these assumptions are implausible. This is the reason why they are not considered in this thesis.

the approaches considered herein are such that an estimand is chosen, assumptions

about the missing data are put into place in an accessible way, and valid inferences under these assumptions are obtained.

3.4.2 Principled methods

Principled methods refer to the set of procedures for analysing missing data that, contrary to simple methods, choose an estimand, and make assumptions about the missing component of group of individuals. The analysis and inference under these methods are usually conducted within a frequentist or Bayesian framework (Kenward and Carpernter, 2007). Principled methods do not attempt to replace the missing data. As previously stated, they combine available information from the observed data with assumptions about the missing data, in order to generate statistical information about both the missing values and the process that caused the missing data (Kenward and Carpernter, 2007).

Three commonly used methods for the analysis of clinical trials with longitudinal follow-up and missing data are reviewed, focusing on the assumptions they make about the missing data mechanism and hence their suitability for use in sensitivity analysis.

3.4.2.1 Generalised Estimating Equations

With repeated measurements, when the response vector is Normal, only the specification of the first two moments suffices to fully determine the likelihood. With discrete data however, the additional assumptions about higher-order moments as well as the existence of many nuisance parameters that often make the likelihood intractable give rise to the use of the Generalised Estimating Equations (GEEs), as opposed to likelihood-based estimation methods (Diggle et al., 2002).

GEEs were originally developed by Liang and Zeger (1986) for clustered and repeated data. They require only the correct specification of marginal distributions, and they make assumptions about the association structure of responses either within

a subject or cluster (Molenberghs and Verbeke, 2005) ignoring the higher-order moments, while obtaining valid inferences with reasonable efficiency (Molenberghs and Kenward, 2007). Molenberghs and Verbeke (2005) explain the GEEs estimate the parameter values associated with the mean of the individuals' responses vector, and express the assumptions about the association structure in terms of marginal correlations. The standard errors of the GEEs are calculated in a robust way that adjusts any incorrect assumptions about the covariance structure.

GEEs are the popular choice of estimation for the population-averaged (PA) class of models. Usually, with longitudinal data the primary interest lies on marginal inferences, such as treatment difference at the final time point. In PA models the outcomes are conditioned on covariates, but not on any other measurements or latent covariates (Kenward and Molenberghs, 2007). The parameter estimates from these models denote the change in the average response for a unit increase in a covariate across the population (Ballinger, 2004).

With missing data, analysis of PA models is complicated by the fact that GEEs are valid under MCAR but not MAR, since marginal means and variances are not appropriate estimators of MAR (Carpenter and Kenward, 2007); as it was previously shown in(1), the MAR assumption is a conditional statement. On the other hand, GEEs use marginal distributions, and provide marginal estimates for the parameters. For this reason, it is not sensible to use them directly for parameter estimation when there are MAR missing data and the goal is to phrase population averaged statements. Also, the fact that PA models do not specify the joint distribution of the outcomes for each subject (in particular the dependence structure is left unspecified) makes them unsuitable for use as sequential imputation models (Molenberghs and Kenward, 2007). Under MAR, likelihood based methods use the covariance matrix to correct for the bias caused by the missing data (Brown and Prescott, 2006) and this correction is not available with GEEs as the correlation structure is only a 'working approximation'. Carpenter and Kenward (2007) illustrate the steps for a strategy when the aim is to obtain population averaged treatment effects from longitudinal discrete outcomes with missing values. In short, a subject-specific model and be

fitted as an imputation model, and then a population-average model can be fitted for the analysis of the imputed datasets, combining the estimates using multiple imputation rules; these will be described later on in the thesis. GEEs will not be pursued further in this thesis; herein, the response vector is assumed to be Normal where population averaged and subject-specific (or conditional) treatment estimates coincide.

3.4.2.2 Maximum likelihood

Maximum likelihood (ML), as opposed to the simple ad-hoc methods, analyses all available data without the need to discard subjects with incomplete sequences, or impute missing observations. ML based analyses provide consistent estimators under both the MCAR and MAR mechanisms (Kenward and Molenberghs, 2009). In particular, a key result of Rubin (1976) pertains to likelihood analyses when missing data are MAR. Suppose subject i withdraws at time d_i. Then their contribution to the likelihood is:

$$\int f(\mathbf{y}_i, d_i; \boldsymbol{\theta}, \boldsymbol{\phi}) d\mathbf{y}_{i,m} =$$
$$\int f(d_i|\mathbf{y}_i; \boldsymbol{\phi}) f(\mathbf{y}_i; \boldsymbol{\theta}) d\mathbf{y}_{i,m} =$$
$$\int f(d_i|\mathbf{y}_{i,o}, \mathbf{y}_{i,m}; \boldsymbol{\phi}) f(\mathbf{y}_{i,o}, \mathbf{y}_{i,m}; \boldsymbol{\theta}) d\mathbf{y}_{i,m} =$$
$$\int f(d_i|\mathbf{y}_{i,o}; \boldsymbol{\phi}) f(\mathbf{y}_{i,o}, \mathbf{y}_{i,m}; \boldsymbol{\theta}) d\mathbf{y}_{i,m} = \text{, under MAR}$$
$$f(d_i|\mathbf{y}_{i,o}; \boldsymbol{\phi}) \int f(\mathbf{y}_{i,o}, \mathbf{y}_{i,m}; \boldsymbol{\theta}) d\mathbf{y}_{i,m} =$$
$$f(d_i|\mathbf{y}_{i,o}; \boldsymbol{\phi}) f(\mathbf{y}_{i,o}; \boldsymbol{\theta})$$

Since we are interested in estimating $\boldsymbol{\theta}$, it follows subject i's contribution to the likelihood is simply the marginal density of their observed data, $f(\mathbf{y}_{i,o}|\boldsymbol{\theta})$, which is straightforward to calculate for the multivariate Normal model. Inferences therefore can be based solely on this marginal observed data density provided $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ are independent too, a requirement referred to as *distinctness* by Little and Rubin

(2002). That is, knowing the values of $\boldsymbol{\theta}$ does not provide any additional information about $\boldsymbol{\phi}$, and vice versa. So, the drop-out indicator D_i can be ignored, since it does not depend on the missing data, and the missing values are treated as unknown random variables to be averaged over (Little and Rubin, 2002). As a result, analyses with MAR data can be simplified, since there is no need to model $\mathbf{D}$. If patients have no data after deviation, where deviation refers to deviations from protocol, and all the data come from on-protocol patients, then the MAR analysis through ML estimates the de-jure effect.

Carpenter and Kenward (2007) provide a thorough illustration of ML methods in analyses with missing data. The choice of covariates in such models is important. Choosing the appropriate covariates for inclusion increases the plausibility of the missing data being MAR, and hence reduces biases, as well as improving the precision of the estimates (Carpenter and Kenward, 2007; White et al., 2011). In practice, it is advisable to include all covariates that have a significant association with the incomplete variable, as well as the chance of withdrawal (Carpenter and Kenward, 2007; White et al., 2011; Collins et al., 2001). This point will be revisited in the discussion about choosing covariates for the imputation model within a multiple imputation settng. Carpenter and Kenward (2007) advocate for the use of logistic and survival analysis to identifying key predictors of withdrawal.

Usually, in clinical trials, missingness is mainly encountered in the response variable, but sometimes predictors of the response/withdrawal (that would be advisable include in the model), may also be partially incomplete (Kenward and Carpenter, 2007). Methods such as the expectation-maximisation algorithm are suited to this problem. Alternatively, assuming that these incomplete covariates can be modelled using a Normal distribution, Carpenter and Kenward (2007) illustrate a method which treats them as additional responses. At this point, the authors draw a distinction between obtaining a treatment effect conditional on the predictor (which is now in the response vector) to be included in the analysis, and a treatment effect marginal to this predictor. This distinction becomes important in situations where it is inappropriate to obtain treatment effects conditional on a specific variable, for

example with variables measured after randomisation, since they can be correlated with treatment. The authors show that to obtain conditional effects, the predictor in the response vector of the joint Normal distribution must have the same mean across the treatment groups, whereas to obtain marginal effects the predictor should be assigned a different mean in each treatment group (Carpenter and Kenward, 2007; pages 55-67). The authors argue however, if the predictor is measured at baseline, it tends to have relatively few missing values, and it is best included as a covariate (given that we wish to adjust the treatment estimate for it).

When the outcomes are Normally distributed analysis of longitudinal data with drop-out is carried out through the use of linear mixed models. The preferred method of estimation in such models is the restricted maximum likelihood, which controls for the downwardly biased estimate of variance in the profile likelihood (Diggle et al., 2002; Verbeke and Molenberghs, 2000; Brown and Prescott, 2006). In addition, when fitting a linear model, Carpenter and Kenward (2007) advocate for the use of an unstructured covariance, the use of a separate covariance matrix in each arm, and the adjustment to the standard errors and degrees of freedom derived by Kenward and Roger (1997).

3.4.2.3 Multiple imputation

Multiple imputation (MI) was developed by Rubin (1987). It is known as the method for fitting models to partially observed data. MI is a Bayesian procedure: the imputed values are draws from the posterior predictive distribution of the missing data given the observed data and the parameters $f(\mathbf{y}_M|\mathbf{y}_O, \boldsymbol{\theta})$. The parameters are draws from the posterior distribution $f(\boldsymbol{\theta}|\mathbf{y}_O, \mathbf{y}_M)$, in which, as it can be seen, they are not treated as fixed. This is the key to how MI handles uncertainty about the parameters in the imputation model (Little and Rubin, 2002). It implies that both missing data and parameters have distributions, where neither the missing data nor the parameters are held fixed.

The model which is used to produce the imputations is called the *imputation* model,

whereas the respective model for the analysis of the imputed datasets is called the *substantive* or analysis model. In practice, the imputation model can either be based on Bayesian distributions or an approximation to these distributions. In the latter case, caution should be exercised to reflect the uncertainty in both the parameters and missing data (Carpenter and Kenward, 2007).

The generic algorithm for the implementation of MI is: each missing datum in the vector $\mathbf{y}_{i,m} = (y_{i,d_i+1}, ..., y_{i,T_i})'$ is replaced by a number of simulated values, say $y^{(1)}_{i,d+1}, y^{(2)}_{i,d+1}, ..., y^{(k)}_{i,d+1}$, creating every time a number of different datasets that contain both the original observed data and the imputed ones. Each of the k imputed datasets is then analysed by standard, complete-data methods, as if they were complete datasets, in order to obtain estimates of the parameters of interest and their standard errors. Finally, the k different estimates and their standard errors are combined under certain rules to form one inference (Little and Rubin, 2002; p. 85-87), as shown next.

Rubin's combining rules

With no missing data the mean and variance from the posterior distribution would be $E(\theta|Y)$ and $var(\theta|Y)$. However, with missing data the simulated posterior mean and variance from each imputed dataset k are:

$$E_\theta(\theta|Y_o, Y^k_m) = \hat{\theta}, \text{ and} \quad (2)$$

$$var(\theta|Y_o, Y^k_m) = V_k \quad (3)$$

calculated using the completed imputed dataset.

Little and Rubin (2002; p. 209-211), after relating the obseved data posterior distribution to the complete data posterior distribution, that is $p(\theta|Y_o) = \int p(\theta|Y_o, Y_m)$ $p(Y_m|Y_o)dY_m$, they show the overall posterior mean across the datasets is the average

of (2), that is:

$$E_{Y_m|Y_o}(E_\theta(\theta|Y_o, Y_m^k)) \tag{4}$$

Similarly, the overall posterior variance across the imputations is the sum of the average of the posterior within-variance estimate (3) across the K imputed and the between-imputation variances:

$$E_{Y_m|Y_o}(var_\theta(\theta|Y_o, Y_m^k)) + var_{Y_m|Y_o}(E_\theta(\theta|Y_o, Y_m^k)) \tag{5}$$

After some manipulation, both (4) and (5) can be shown to equal $E(\theta|Y_o)$ and $var(\theta|Y_o)$, the posterior mean and variance if no missing data existed.

In practice, by the law of large numbers, as $K \to \infty$,
$E_{Y_m|Y_o}(E_\theta(\theta|Y_o, Y_m^k))$, $E_{Y_m|Y_o}(var_\theta(\theta|Y_o, Y_m^k))$, and $var_{Y_m|Y_o}(E_\theta(\theta|Y_o, Y_m^k))$ can be approximated by,

$$\bar{\theta} = k^{-1} \sum_{t=1}^{k} \hat{\theta}^k \tag{6}$$

$$\bar{V} = k^{-1} \sum_{t=1}^{k} V_k \tag{7}$$

$$B = (k-1)^{-1} \sum_{t=1}^{k} (\hat{\theta}^k - \bar{\theta})^2 \tag{8}$$

respectively (Schafer, 1997). Hence, (6) is the MI estimate of θ, which is simply the average of the estimates across the imputed datasets. Equation (7) provides an estimate of the variability that would be obtained from a single complete dataset, whereas (8) represents the variability across the imputations (with no missing data $B = 0$). To approximate (5) the sum of the equations (7) and (8) would be normally used. However, to allow for the extra uncertainty from using a finite number of imputations K, an improved formula is used especially when the number of imputations is small (Molenberghs and Kenward, 2007):

$$T = \bar{V} + (1 + k^{-1})B_k \tag{9}$$

As previously stated, missing data and the parameters can be draws either from fully Bayesian distributions or approximations to these. In sufficiently large samples, Carpenter and Kenward (2007) describe an algorithm to conduct MI based on regression, by approximations to the posterior distribution of the parameters. There, they use the sampling distribution of the parameters, and the large sample covariance matrix (the inverse of the information matrix), as in large samples these tend to approximate the posterior distribution (Kenward and Carpenter, 2009). These algorithms can be applied to many settings and is simple to extend to longitudinal data with drop-out, where a *sequential* imputation procedure can be used (see for example Kenward and Carpenter, 2009). Approximate MI draws can also be done through a fully Bayesian process, usually through the Markov Chain Monte Carlo (MCMC) algorithm (Little and Rubin, 2002; Carpenter and Kenward, 2013). Allison (2002) provides an algorithm for using regression within an MCMC algorithm and for general forms of missing data Schafer (1997) provides a comprehensive description of fully Bayesian imputations.

Multiple imputation and the maximum likelihood

Carpenter and Kenward (2007) show that results from 200 analysed imputed datasets (in the case of estimating a single parameter and its variance) produce nearly identical estimates to mixed effects models: in their analyses they fitted a mixed model in order to estimate 3 year treatment effect conditional on baseline and two other covariates, exercising caution to develop an imputation model that uses the same structure as the mixed model. ML methods and MI would produce asymptotically the same results, provided, according to Collins et al. (2001), that:

> *the user of the ML procedure and the imputer use the same set of input data, their models apply equivalent distributional assumptions to the variables and the relationships among them, the sample size is large, and the number of imputations, is sufficiently large*

The asymptotic equivalence of likelihood-based methods and Bayesian procedures is based on the fact that as more data arrive, the posterior distribution of the parameters approaches multivariate normality (Gelman et al., 2003). Gelman et al. (2003) explain that in large samples, the importance of the prior distribution diminishes, as the sample size increases; it is getting dominated by the likelihood and as such, "the mode of the likelihood (the maximum likelihood estimator) and the inverse of the curvature of the likelihood (the information based covariance matrix) can be used to obtain the required moments (Carpenter and Kenward, 2013). Hence, with large sample sizes, the choice of the prior distribution that accurately reflects all available information "should not be of a concern" (Gelman et al., 2003).

The structure of the imputation model and further remarks

Within MI, the substantive model is constructed separately from the imputation model. White et al. (2011) argues that all variables that appear in the substantive model should exist in the imputation model too. It is important to preserve all relationships between the variables, so that biases in the analysis model would be reduced. In addition, the imputation model should be at least as functionally complex as the substantive model, in terms of interactions and/or higher order terms used in the substantive model (Schafer, 1997).

MI permits additional covariates to be included in the imputation model which may not be desirable to include in the substantive model. If the aim for instance, is to obtain a treatment effect conditional on a baseline predictor, but marginal to a post-randomisation variable, this can simply be done by including both predictors in the imputation model, but only the baseline covariate in the substantive model. As White et al. (2011) explain, "the imputation model should include every variable that both predicts the incomplete variable and predicts whether the incomplete variable is missing". This way, the MAR assumption would become more plausible, the imputations can be improved and the standard errors of the estimates for the analysis model can be reduced. Collins et al. (2001) term these predictors *auxiliary* variables. If a variable is just predictive of deviation and not predictive of the

incomplete variable, then there is no gain in including it in the imputation model, since the justification for its inclusion is based on the fact that it may (or may not) have a true association with the incomplete variable that failed to reach statistical significance (White et al. 2011).

Meng (1994) calls the substantive and the imputation models *uncongenial*, when the former does not correspond to the latter in terms of the structure of the predictive distribution $f(\mathbf{y}_M|\mathbf{y}_O)$ each model implies. Uncongeniality is an important feature of the MI process, as it makes it very amenable to sensitivity analysis. Molenberghs and Kenward (2007) explain, this can be achieved for example, by introducing an explicit MNAR model into the imputation to modify the future behaviour of drop-outs conditional on the past. Whereas under MAR such conditional behaviour is implied to be the same between the completers and those who drop-out, under MNAR such a behaviour may differ. Thus, the imputation model can be allowed to accommodate different distributional assumptions from those implied by the substantive model (Little and Yau, (1996); Kenward et al. (2003); Carpenter et al. (2013)) addressing in this manner de-facto questions.

Based on Schafer (1997), the issue of uncongeniality can be split into two types. Firstly, when the substantive model 'assumes more', in terms of estimating less parameters than the imputation model. Then, as Schafer (1997) explains, if the assumptions of the substantive model were true, inferences from the imputation model will be unbiased. The variance estimates however would tend to be a little conservative, because the imputation model will reflect an extra degree of uncertainty due to the fact that it estimates more parameters. Also, interval estimates will tend to be somewhat wider than what they would have been, if the imputation model estimated the same number of parameters as the substantive model (Schafer, 1997). This type of uncongeniality will be explored later on in this thesis.

The second type of uncongeniality arises when the imputation model assumes more than the substantive model, that is, the substantive model is more elaborate. Fay (1992) explored the properties of MI under this type of uncongeniality, given the

assumptions made by the imputation model are true. Fay compared $E(T_\infty|\mathbf{X},\mathbf{y})$ to $var(\bar{\theta}_\infty|\mathbf{X},\mathbf{y})$ over repeated realisations of the sampling and imputation procedure, and showed that when uncongeniality exists $E(T_\infty|\mathbf{X},\mathbf{y}) \geq var(\bar{\theta}_\infty|\mathbf{X},\mathbf{y})$, as opposed to $E(T_\infty|\mathbf{X},\mathbf{y}) = var(\bar{\theta}_\infty|\mathbf{X},\mathbf{y})$, which holds when the two models apply the same assumptions. Carpenter and Kenward (2013) explain this type of uncongeniality should be avoided, "as it generally results in both inconsistent estimators of the substantive model parameters and invalidity of the Rubin's variance estimator".

3.5 Multiple imputation and MNAR models

This section focused on the incorporation of MNAR assumptions within the MI setting. Implicit up to this point was the fact that MI is valid under MAR. That is, D_i can be ignored in the conditional distribution $f(\mathbf{y}_{i,m}|\mathbf{y}_{i,o}, D_i)$ to give $f(\mathbf{y}_{i,m}|\mathbf{y}_{i,o})$ since:

$$f(\mathbf{y}_{i,m}|\mathbf{y}_{i,o}, D_i) = \frac{f(\mathbf{y}_{i,m}, \mathbf{y}_{i,o}, D_i)}{f(\mathbf{y}_{i,o}, D_i)} =$$
$$\frac{f(D_i|\mathbf{y}_{i,m}, \mathbf{y}_{i,o}) f(\mathbf{y}_{i,m}, \mathbf{y}_{i,o})}{f(D_i|\mathbf{y}_{i,o}) f(\mathbf{y}_{i,o})} = \frac{f(D_i|\mathbf{y}_{i,o}) f(\mathbf{y}_{i,m}, \mathbf{y}_{i,o})}{f(D_i|\mathbf{y}_{i,o}) f(\mathbf{y}_{i,o})} =$$
$$\frac{f(\mathbf{y}_{i,m}, \mathbf{y}_{i,o})}{f(\mathbf{y}_{i,o})} = f(\mathbf{y}_{i,m}|\mathbf{y}_{i,o})$$

As such, it is suitable to address de-jure questions. Under MNAR however, D_i cannot be ignored, since the reason for dropping out could be different for each individual or, more typically, group of individuals. In order to reflect this, $f(\mathbf{y}_{i,m}|\mathbf{y}_{i,o}, D_i)$, that forms the part of the imputation model, where the missing data from are drawn from, can be modified, so that it represents assumptions specific to the treatment arm, drop-out time-point and possibly other characteristics. This in turn, entails the need to consider MNAR models, which retain D_i. There are three main classes of such models; selection models, shared-parameter models and pattern-mixture models. Starting from the full density, where $\mathbf{Z}_i$ and $\mathbf{W}_i$ denote design matrices for the measurement and missingness mechanisms respectively, the factorisations each class of models implies is shown below.

The factorisation under selection models is:

$$f(\mathbf{y}_i, D_i|\mathbf{Z}_i, \mathbf{W}_i; \boldsymbol{\theta}, \boldsymbol{\phi}) = f(\mathbf{y}_i|\mathbf{Z}_i; \boldsymbol{\theta}) f(D_i|\mathbf{y}_i, \mathbf{W}_i; \boldsymbol{\phi}) \tag{10}$$

The first factor is the marginal density of the measurement process, and the second one is the density of the missingness process conditional on the response. It can be seen, the full vector of responses $\mathbf{y}_i = (y_{i1}, ..., y_{iT})'$ for subject i, either missing or not, is being weighted (or selected) through the probability of being observed. Kenward and Molenberghs (2007) explain that Rubin's classification of missing value processes, introduced in Section 3.2 is most naturally expressed within the selection models framework, based on factorisation (10). For instance, under MAR, the joint distribution for a selection model would be $f(\mathbf{y}_{i,o}|\mathbf{Z}_i; \boldsymbol{\theta}) f(D_i|\mathbf{y}_{i,o}, \mathbf{W}_i; \boldsymbol{\phi})$, as shown in section 3.4.2.2.

The factorisation that characterises the shared-parameter models is:

$$f(\mathbf{y}_i, D_i, \mathbf{b}_i|\mathbf{Z}_i, \mathbf{W}_i, \mathbf{Q}_i; \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\xi}) = f(\mathbf{y}_i|\mathbf{Z}_i, \mathbf{b}_i; \boldsymbol{\theta}) f(D_i|\mathbf{W}_i, \mathbf{b}_i; \boldsymbol{\phi}) f(\mathbf{b}_i|\mathbf{Q}_i; \boldsymbol{\xi})$$

Shared-parameter models allow for a different response model for each pattern of missingness, and they also introduce latent variables, upon which the response and the drop-out pattern is conditionally independent (Molenberghs and Kenward, 2007). The main characteristic of these models is that a single set of parameters $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{V})$ which denotes a vector of subject specific random effects, are shared between both factors in the joint distribution of $\mathbf{y}_i$ and D_i. The factorisation follows from the assumption that the drop-out mechanism D_i is independent of both $\mathbf{y}_{i,o}$ and $\mathbf{y}_{i,m}$, conditionally on the random effects (Daniels and Hogan, 2008).

Finally, the factorisation of the joint distribution under the PM models is:

$$f(\mathbf{y}_i, D_i|\mathbf{Z}_i, \mathbf{W}_i; \boldsymbol{\theta}, \boldsymbol{\phi}) = f(\mathbf{y}_i|D_i, \mathbf{Z}_i; \boldsymbol{\theta}) f(D_i|\mathbf{W}_i; \boldsymbol{\phi}) \tag{11}$$

The pattern-mixture models, where 'pattern' in this thesis refers to a separate response distribution, allow for a different response model for each pattern of miss-

ingness. Here, the data $\mathbf{y} = (y_1, ..., y_T)'$ for a group of subjects, are a mixture of patterns, weighted by the proportion of missing data in each drop-out pattern (Molenberghs and Kenward, 2007).

There are no specific guidelines in the literature with regards to the choice among those models. Each approach has its advantages and disadvantages (Molenberghs and Kenward, 2007; Daniels and Hogan, 2008). Selection models treat the probability of drop-out as dependent on the response variable. Carpenter et al. (2002) explain that in a trials context, patient drop-out would depend on treatment response, and they argue in their asthma study, selection models are deemed the most appropriate choice for analysis, since drop-out can be explained by a steady deterioration of the health status of the patients. With respect to the shared model parameters, formulation of the MAR assumption and consequently deviations from it, as Daniels and Hogan (2002) note, is hampered by the fact that the random effects structure make it difficult to separate parameters indexing $f(\mathbf{y}_{i,m}|\mathbf{y}_{i,o}, d_i)$ from those indexing $f(\mathbf{y}_{i,o}, d_i)$. The focus of this thesis however, will be on PM models. This is due to their flexibility to formulate different patterns of response for those who drop-out and those who do not (Carpenter et al., 2002), since the probability distribution of the response depends on the drop-out status.

With PM models, the conditional distribution of the measurements $\mathbf{y}_i$ given the drop-out pattern is combined with the marginal distribution of the drop-out variable, which can depend on covariates but not on measurements. As stated before, PM models allow the distribution of $\mathbf{y}_i$ to differ for each pattern of missing data. For example, given the patterns in section 2.1, the PM factorisation (11), for drop-out pattern 2, in the placebo group, would be:

$$\begin{aligned} f(\mathbf{y}_i|D_i = 2, \mathbf{Z}_i, g_{i,pl}; \boldsymbol{\theta}) f(D_i = 2|\mathbf{W}_i; \boldsymbol{\phi}) = \\ \prod_{t=3}^{6} f(y_{i,t}|y_{i,1}, ..., y_{i,t-1}, D_i = 2, \mathbf{Z}_i, g_{i,pl}; \boldsymbol{\theta}) \\ \times f(y_{i,1}, y_{i,2}|D_i = 2, \mathbf{Z}_i, g_{i,pl}; \boldsymbol{\theta}) \\ \times f(D_i = 2|\mathbf{W}_i; \boldsymbol{\phi}) \end{aligned} \tag{12}$$

Within each pattern, the parameters of interest are estimated. With Normally distributed data, when pattern-specific estimates are of no interest, marginal estimates can be obtained by averaging over the distribution of D.

An important issue with PM models is that, by construction, they are under-identified. These unidentified parameters of the conditional distributions of the incomplete patterns need be estimated nonetheless. Many authors addressed this issue (Little, 1993; Little, 1994; Molenberghs et al., 1998; Thijs et al., 2002). These authors constructed restrictions according to which, the unidentified parameters are set equal to a linear combination of their equivalent counterparts from identified patterns. Kenward et al. (2003) call these restrictions *interior family constraints*. In their paper they develop further restrictions, namely *non-future dependence missing value*.

Pattern-mixture models under-identifiability can also be dealt with through model simplification. This approach was used by Hogan and Laird (1997); a linear mixed model can be fitted in each of the patterns separately, replacing unobserved time coefficients with their equivalent from observed patterns. Moreover, as Verbeke and Molenberghs (2000) and Daniels and Hogan (2008) explain, one can also perform model simplification by fitting a pattern factor in a linear mixed model.

The MAR assumption motivated by equation (1) is a good starting point for sensitivity analysis with PM models. According to this, under the interior family constraints for example, the unidentified components of the means and variances of the conditional distribution (12) would be identified as: $f(y_{i,3}|y_{i,1}, y_{i,2}, D_i = 2, \mathbf{Z}_i, g_{i,pl}; \boldsymbol{\theta}) = f(y_{i,3}|y_{i,1}, y_{i,2}, D_i \geq 3, \mathbf{Z}_i, g_{i,pl}; \boldsymbol{\theta})$, $f(y_{i,4}|y_{i,1}, y_{i,2}, y_{i,3}, D_i = 2, \mathbf{Z}_i, g_{i,pl}; \boldsymbol{\theta}) = f(y_{i,4}|y_{i,1}, y_{i,2}, y_{i,3}, D_i \geq 4, \mathbf{Z}_i, g_{i,pl}; \boldsymbol{\theta})$ and so on. The analysis then may proceed by constructing MNAR models that match exactly the MAR model in its observed components, but differ in the unobserved. In other words, the unidentified components of the conditional means and variances would be identified according to appropriate MNAR assumptions imposed by researchers. This gives a great advantage to PM models, since they can accommodate assumptions that deviate from MAR in a

transparent way.

Alternatively, as White et al. (2007) and Daniels and Hogan (2008) show, one could add a *sensitivity parameter*, δ, to the unidentified parameters of the conditional mean and variance; this parameter measures the degree of deviation from MAR. For example, the intercept of the conditional mean of (12), $b_o^{(D_i=2)} = \bar{y}_{i,3}^{(D_i=2)} - b_1^{(D_i=2)}\bar{y}_{i,2}^{(D_i=2)} - b_2^{(D_i=2)}\bar{y}_{i,1}^{(D_i=2)}$ is equated to $\bar{y}_{i,3}^{(D_i\geq 3)} - b_1^{(D_i\geq 3)}\bar{y}_{i,2}^{(D_i\geq 3)} - b_2^{(D_i\geq 3)}\bar{y}_{i,1}^{(D_i\geq 3)} + \delta$. It follows that under MAR $\delta = 0$. The details of this approach, which can be carried out within a Bayesian framework, where a prior on δ can be assigned to construct the posterior distribution from which the parameters are drawn, are provided in Daniels and Hogan (2008, ch. 9).

3.5.0.4 Pattern-mixture imputation models

MNAR PM models can serve as imputation models. These will almost always be uncongenial with the substantive model used at the end. Thijs et al. (2002) and Kenward et al. (2003) describe the steps for drawing imputations from conditional distributions, identified using interior family constraints and non-future dependence missing value constraints, respectively. Following identification, the conditional distributions are then used for imputing missing data. Implicit in the identifications used by the interior family constraints and non-future dependence missing value constraints is the assumption that information is borrowed from within the same treatment arm.

Little and Yau (1996) and Carpenter et al. (2013) presented an alternative strategy for the use of PM models within MI. They developed sensitivity analyses wherein the unidentified conditional distribution can be identified via components borrowed from different treatment arms.

These sensitivity analyses are amenable to an ITT estimand, which requires the drop-out population, that after dropping out may have switched treatments, should be analysed as part of the treatment regime they were initially assigned to. Little and

Yau (1996) modelled the behaviour of the drop-out population as imputations that conditioned on actual or assumed treatments received; the subjects who dropped out differed from the completers in their compliance status, where compliance under MAR, was assumed to be the same in both groups. This was then followed by a classical ITT analysis based on the treatment 'as randomised'. Under this setup, the imputation and substantive MAR models differed only with respect to the unobserved behaviour of the drop-outs.

3.6 Summary

This chapter presented the 3 main missing data assumptions, MCAR, MAR, and MNAR within a longitudinal data setting, and introduced two new estimands, namely de-jure and de-facto. These estimands are intended to improve upon the meaning of traditional estimands, such as per-protocol and ITT, as they are particularly suited to situations where missing data occur. As such, a de-jure estimand seeks to answer questions that pertain to the expected treatment effect in the population, if treatments were taken as specified in the protocol, whereas a de-facto estimand would address questions that pertain to the treatment effect seen in practice by the eligible population.

The most popular principled methods for the analysis of longitudinal data were presented. These were GEEs, maximum likelihood based methods, and multiple imputation. Analyses under GEEs are only valid if the missing data are MCAR. For this reason, they cannot be used for a direct parameter estimation when the data are MAR. However, if the aim is to obtain marginal inferences, multiple imputation can still be performed using a population-average model in the final analysis stage. With MAR data, both likelihood based approaches and MI are suitable for analysis of longitudinal datasets. Also, both likelihood based approaches and MI address de-jure questions and asymptotically provide the same results. Both methods allow for the incorporation of auxiliary variables to increase the possibility of obtaining MAR missing data, even if it is not desired to condition effects on specific auxiliary

variables; likelihood based models can achieve this by assigning the auxiliary variable a different mean in each group, whereas MI by including the auxiliary variable in the imputation model, but not in the substantive model.

The great advantage of MI over likelihood however, is its flexibility to have an uncongenial imputation model, in terms of allowing for different structures between the imputation and substantive models. This flexibility permits MNAR models to enter the analysis of incomplete datasets and make various statistical assumptions about the behaviour of the missing components. Three classes of MNAR models exist; selection models, shared-parameter models and pattern-mixture models. Pattern-mixture models are very amenable to MI; they form the conditional predictive distribution of the imputation model, which missing data are drawn from, and allow for the construction of assumptions about the incomplete components in a transparent way. The next chapter presents a new method that shows how MI incorporate PM structures into its imputation stage, and develop in this way various assumptions about the statistical behaviour of groups of individuals with missing data.

In conclusion, if a patient's or group of patients' post-deviation conditional distribution given their pre-deviation data is estimated from the available data in their treatment arm, then a de-jure MAR assumption is being addressed (the probability of deviation, given pre-deviation data, does not depend on post-deviation data). In such a setting, both maximum likelihood and MI produce valid estimates. However, to answer de-facto questions, where departures from the MAR assumption are assumed, MI is the natural choice of estimation, because the imputation model, which represents the conditional post-deviation distribution, is allowed to be estimated from a different treatment arm.

4 Current research

The new sensitivity analysis developed by Carpenter J., Roger J. and Kenward M. (CRK) (2013) is presented in this chapter. The method explicitly incorporates a model for subjects' post-deviation outcome data. This model will typically be based on an assumption about the treatment patients receive post-deviation, and can be extended to include post-deviation data where there are available. The key feature of the CRK method is that, for the construction of the post-deviation distribution, information can be borrowed from other patient groups, which may receive different treatments to those received by the drop-out patients pre-deviation; this is different to the methods proposed by White et al. (2007) and Daniels and Hogan (2008) that make use of a sensitivity parameter. The CRK method utilises simulated parameters initially estimated from an MAR model, and then manipulates these parameters to construct conditional distributions that would correspond to imputing under an MAR or a transparent set of various MNAR assumptions. As a result, this framework allows to make a wide range of assumptions about post-deviation given pre-deviation data, which are relevant to both de-jure and de-facto estimands, and then use multiple imputation as a convenient tool for estimation and inference.

4.1 The method

Consider a clinical trial, which gives rise to a quantitative outcome that is measured longitudinally. Also, assume this can be modelled by the multivariate Normal distribution, that is:

$$\mathbf{y}_i = \mathbf{X}_i\mathbf{b} + \boldsymbol{\varepsilon}_i, \qquad \boldsymbol{\varepsilon}_i \sim MVN(\mathbf{0}, \mathbf{V}_i) \tag{13}$$

where $\mathbf{y}_i$ is a $(T_i \times 1)$ response vector for subject i, $\mathbf{X}_i$ a $(T_i \times p)$ design matrix of p covariates, $\mathbf{b}$ a $(p \times 1)$ vector of parameters, $\boldsymbol{\varepsilon}_i^{'} = (\varepsilon_{i1}, ..., \varepsilon_{iT_i})$ the residual, with $\boldsymbol{\Sigma}_i$ the positive definite covariance matrix for $\boldsymbol{\varepsilon}_i$. With a balanced design, the subscript i from T and $\boldsymbol{\Sigma}$ can be dropped.

The aim of an analysis, if there were no missing data, would be to estimate the treatment effect at time T. In order to do this, either an ANCOVA model using as outcomes the measurements at the final time point, or a mixed-effects regression model can be fitted. With missing data, model (13) would give equivalent estimates and standard errors, provided time is included as a categorical variable, a treatment by time and baseline by time interaction is fitted, and a common unstructured covariance matrix to the treatment arms is used. The treatment by time interaction allows to examine the outcome profile across time within each treatment arm. As Siddiqui et al. (2009) note, the advantage of this model specification, is that "it provides the direct estimates and statistical test of least square mean (LSMEAN) differences of the treatment groups at the study endpoint, as well as at each scheduled study time point with respect to the primary efficacy measure".With missing data, since estimation with model (13) is likelihood based, a de-jure question is addressed (see previous chapter).

The approach proceeds as follows. At the beginning, model (13) is fitted through REML. Initial values for the means, variances and covariances are obtained, for each treatment separately. These values are then used to initialise the MI process, which is carried out through the MCMC algorithm (Carpenter and Kenward, 2013, Appendix A); a sequence of parameter values is drawn, whose stationary distribution is the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y})$ ($\boldsymbol{\theta}$ denotes the means and covariance parameters of model (13)). After discarding the early values of the chain, the 'burn-in', separately for each treatment arm, a vector of means and a covariance matrix are obtained. The posterior has a Normal multivariate likelihood. Following Schafer (1997), an improper prior distribution is assigned to the mean and a Jeffreys' prior to the covariance matrix. With two treatments, for instance, an active treatment $j = A$ and a reference treatment $j = R$, there will therefore be two vectors of posterior

means and two posterior covariance matrices:

$$\widetilde{\boldsymbol{\mu}}'_{i,A,t} = (\mu_{i,A,0}, ..., \mu_{i,A,T}) \quad , \quad \widetilde{\boldsymbol{\mu}}'_{i,R,t} = (\mu_{i,R,0}, ..., \mu_{i,R,T}) \tag{14}$$

$$\widetilde{\boldsymbol{\Sigma}}_{i,A} = \begin{bmatrix} A_{i,11} & A_{i,12} & ... & A_{i,1T} \\ \vdots & \vdots & \ddots & \vdots \\ A_{i,T1} & A_{i,T1} & ... & A_{i,TT} \end{bmatrix} \quad , \quad \widetilde{\boldsymbol{\Sigma}}_{i,R} = \begin{bmatrix} R_{i,11} & R_{i,12} & ... & R_{i,1T} \\ \vdots & \vdots & \ddots & \vdots \\ R_{i,T1} & R_{i,T1} & ... & R_{i,TT} \end{bmatrix} \tag{15}$$

Missing data are then drawn from the posterior predictive distribution:

$$f(\mathbf{y}_M|\mathbf{y}_O, \mathbf{D}, \boldsymbol{\theta}) = \prod_{i=1}^{n} f(\mathbf{y}_{i,m}|\mathbf{y}_{i,o}, D_i; \boldsymbol{\theta}) \tag{16}$$

where the parameter vector $\boldsymbol{\theta}$ uses the imputed parameters from (14) and (15). For each patient who deviates, their posterior predictive distribution of post-deviation given pre-deviation responses, (16), is defined using one of the rules described in Section 4.1.1. MI proceeds sequentially and each subject of the sample leads to a separate imputed dataset. The substantive model, used to analyse the imputed datasets, is an ANCOVA model.

4.1.1 Assumptions

This exposition assumes two treatment groups (for simplicity), but the ideas can be extended to multi-arm studies. Before constructing the conditional distributions for the imputation models, the time of drop-out, $D_i = d$, for each individual, and for each treatment arm is noted. The vectors and matrices in (14) and (15) can be partitioned, according to d_i, as:

$$\widetilde{\boldsymbol{\mu}}'_{i,A,T} = (\widetilde{\mu}_{i,A,0}, ..., \widetilde{\mu}_{i,A,d_i}, \widetilde{\mu}_{i,A,d_i+1}, ..., \widetilde{\mu}_{i,A,T}) \tag{17}$$

$$\widetilde{\boldsymbol{\mu}}'_{i,R,T} = (\widetilde{\mu}_{i,R,0}, ..., \widetilde{\mu}_{i,R,d_i}, \widetilde{\mu}_{i,R,d_i+1}, ..., \widetilde{\mu}_{i,R,T}) \tag{18}$$

$$\widetilde{\boldsymbol{\Sigma}}_{i,A} = \left[\begin{array}{ccc|ccc} \widetilde{A}_{i,11} & \dots & \widetilde{A}_{i,1d_i} & \widetilde{A}_{i,1d_i+1} & \dots & \widetilde{A}_{i,1T} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \widetilde{A}_{i,d_i1} & \dots & \widetilde{A}_{i,d_id_i} & \widetilde{A}_{i,d_id_i+1} & \dots & \widetilde{A}_{i,d_iT} \\ \hline \widetilde{A}_{i,d_i+11} & \dots & \widetilde{A}_{i,d_i+1d_i} & \widetilde{A}_{i,d_i+1d_i+1} & \dots & \widetilde{A}_{i,d_i+1T} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \widetilde{A}_{i,T1} & \dots & \widetilde{A}_{i,Td_i} & \widetilde{A}_{i,Td_i+1} & \dots & \widetilde{A}_{i,TT} \end{array}\right] = \left[\begin{array}{c|c} \widetilde{\mathbf{A}}_{i,11} & \widetilde{\mathbf{A}}_{i,12} \\ \hline \widetilde{\mathbf{A}}_{i,21} & \widetilde{\mathbf{A}}_{i,22} \end{array}\right] \quad (19)$$

$$\widetilde{\boldsymbol{\Sigma}}_{i,R} = \left[\begin{array}{ccc|ccc} \widetilde{R}_{i,11} & \dots & \widetilde{R}_{i,1d_i} & \widetilde{R}_{i,1d_i+1} & \dots & \widetilde{R}_{i,1T} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \widetilde{R}_{i,d_i1} & \dots & \widetilde{R}_{i,d_id_i} & \widetilde{R}_{i,d_id_i+1} & \dots & \widetilde{R}_{i,d_iT} \\ \hline \widetilde{R}_{i,d_i+11} & \dots & \widetilde{R}_{i,d_i+1d_i} & \widetilde{R}_{i,d_i+1d_i+1} & \dots & \widetilde{R}_{i,d_i+1T} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \widetilde{R}_{i,T1} & \dots & \widetilde{R}_{i,Td_i} & \widetilde{R}_{i,Td_i+1} & \dots & \widetilde{R}_{i,TT} \end{array}\right] = \left[\begin{array}{c|c} \widetilde{\mathbf{R}}_{i,11} & \widetilde{\mathbf{R}}_{i,12} \\ \hline \widetilde{\mathbf{R}}_{i,21} & \widetilde{\mathbf{R}}_{i,22} \end{array}\right] \quad (20)$$

Hence, the joint distribution

$$N\left(\left(\begin{array}{c} \widetilde{\boldsymbol{\mu}}_{i,j,o} \\ \widetilde{\boldsymbol{\mu}}_{i,j,m} \end{array}\right), \left(\begin{array}{c|c} \widetilde{\mathbf{V}}_{i,11} & \widetilde{\mathbf{V}}_{i,12} \\ \hline \widetilde{\mathbf{V}}_{i,21} & \widetilde{\mathbf{V}}_{i,22} \end{array}\right)\right) \quad (21)$$

for subject i under treatment j, is pieced together according to the assumptions proposed by Carpenter et al. (2013), in order to construct the conditional distributions for the post-deviation given pre-deviation data for each patient. The authors explain that each option addresses either a de-jure or de-facto estimand. The nomination of the 'Reference' treatment below is left at the researcher's discretion and depends entirely upon the assumptions they wish to apply.

Randomised-arm MAR (MAR) The subject's observed and missing responses are multivariate normal with mean and covariance from their randomised arm. Under this assumption, for each subject that deviates at any time point, the unobserved components of their conditional distribution of post- given pre-deviation data, bor-

rows information from within the same treatment group. Therefore, a de-jure estimand is addressed.

For instance, for a patient that is randomised to treatment group A their conditional mean will be constructed from components from (17). With regards to the covariance matrix a combination of marginal and conditional components is used: for pre-deviation time-points, that is $t_i = 1, ..., d_i$, the marginal component of (19) is used, whereas for post-deviation time-points, $t_i = d_i + 1, ..., T$, the conditional components of (19) are used.

First, the conditional covariance matrix can be obtained after sweeping the symmetric matrix (19) on positions $1, 2, ..., d_i$, using the SWEEP operator described by Goodnight (1979):

$$SWP[1, 2, ..., d_i]\widetilde{\mathbf{\Sigma}}_{i,A} = \left[\begin{array}{c|c} -\widetilde{\mathbf{A}}_{i,11}^{-1} & \widetilde{\mathbf{A}}_{i,11}^{-1}\widetilde{\mathbf{A}}_{i,12} \\ \hline \widetilde{\mathbf{A}}_{i,21}\widetilde{\mathbf{A}}_{i,11}^{-1} & \widetilde{\mathbf{A}}_{i,22} - \widetilde{\mathbf{A}}_{i,21}\widetilde{\mathbf{A}}_{i,11}^{-1}\widetilde{\mathbf{A}}_{i,12} \end{array}\right]$$

where $\widetilde{\mathbf{A}}_{i,22} - \widetilde{\mathbf{A}}_{i,21}\widetilde{\mathbf{A}}_{i,11}^{-1}\widetilde{\mathbf{A}}_{i,12}$ can be used to compute the conditional covariance matrix of $\mathbf{y}_{i,m}$ given $\mathbf{y}_{i,o}$. Likewise, $\widetilde{\mathbf{A}}_{i,21}\widetilde{\mathbf{A}}_{i,11}^{-1}$ can be used for the conditional coefficients.

The new covariance matrix will therefore be constructed from the following constraints:

$$\begin{aligned} \widetilde{\mathbf{V}}_{i,11} &= \widetilde{\mathbf{A}}_{i,11} \\ \widetilde{\mathbf{V}}_{i,21}\widetilde{\mathbf{V}}_{i,11}^{-1} &= \widetilde{\mathbf{A}}_{i,21}\widetilde{\mathbf{A}}_{i,11}^{-1} \\ \widetilde{\mathbf{V}}_{i,22} - \widetilde{\mathbf{V}}_{i,21}\widetilde{\mathbf{V}}_{i,11}^{-1}\widetilde{\mathbf{V}}_{i,12} &= \widetilde{\mathbf{A}}_{i,22} - \widetilde{\mathbf{A}}_{i,21}\widetilde{\mathbf{A}}_{i,11}^{-1}\widetilde{\mathbf{A}}_{i,12} \end{aligned} \tag{22}$$

where the marginal components are:

$$\widetilde{\mathbf{V}}_{i,11} = \widetilde{\mathbf{A}}_{i,11} \tag{23}$$

$$\widetilde{\mathbf{V}}_{i,21} = \widetilde{\mathbf{A}}_{i,21}\widetilde{\mathbf{A}}_{i,11}^{-1}\widetilde{\mathbf{V}}_{i,11} = \widetilde{\mathbf{A}}_{i,21} \tag{24}$$

For instance, for subjects that belong to treatment group A their conditional distributions at time point T given the past will be:

$$y_{i,T}|y_{i,0}, y^*_{i,1}, ..., y^*_{i,T-1}, D_i = T-1, g_i = A \sim$$
$$N(\widetilde{\mu}_{i,A,T} + \widetilde{\mathbf{v}}_{i,21}\widetilde{\mathbf{V}}^{-1}_{i,11}(\mathbf{y}_{i,A} - \widetilde{\boldsymbol{\mu}}_{i,A}), \widetilde{V}_{i,22} - \widetilde{\mathbf{v}}_{i,21}\widetilde{\mathbf{V}}^{-1}_{i,11}\widetilde{\mathbf{v}}_{i,12}) \tag{25}$$

where $(\mathbf{y}_{i,A} - \widetilde{\boldsymbol{\mu}}_{i,A})$ is a $(T-1 \times 1)$ vector, and $\widetilde{\mathbf{V}}^{-1}_{i,11}$ is the generalised inverse of $\widetilde{\mathbf{V}}_{i,11}$ from (23). Following identification of these distributions the MI process will proceed sequentially to fill-in the missing values for each individual. Details of the how a sequential imputation algorithm proceeds can be found in Kenward and Carpenter (2009).

Jump to reference (J2R)

Here, the missing components of the conditional distributions for subjects in the active group A are taken from the reference group R, addressing, this way, a de-facto estimand. Missing components for distributions in the reference group are imputed under the randomised-arm MAR assumption. For example, the conditional distribution for a subject that belongs to treatment A and switches to treatment R after time $T-1$, will be:

$$y_{i,T}|y_{i,0}, y^*_{i,1}, ..., y^*_{i,T-1}, D_i = T-1, g_i = A \sim$$
$$N(\widetilde{\mu}_{i,R,T} + \widetilde{\mathbf{v}}_{i,21}\widetilde{\mathbf{V}}^{-1}_{i,11}(\mathbf{y}_{i,A} - \widetilde{\boldsymbol{\mu}}_{i,A}), \widetilde{V}_{i,22} - \widetilde{\mathbf{v}}_{i,21}\widetilde{\mathbf{V}}^{-1}_{i,11}\widetilde{\mathbf{v}}_{i,12}) \tag{26}$$

where now the conditional variance is:

$$\widetilde{V}_{i,22} - \widetilde{\mathbf{v}}_{i,21}\widetilde{\mathbf{V}}^{-1}_{i,11}\widetilde{\mathbf{v}}_{i,12} = \widetilde{R}_{i,22} - \widetilde{\mathbf{r}}_{i,21}\widetilde{\mathbf{R}}^{-1}_{i,11}\widetilde{\mathbf{r}}_{i,12}$$

and

$$\begin{aligned}\widetilde{\mathbf{V}}_{i,11} &= \widetilde{\mathbf{A}}_{i,11} \\ \widetilde{\mathbf{v}}_{i,21} &= \widetilde{\mathbf{r}}_{i,21}\widetilde{\mathbf{R}}^{-1}_{i,11}\widetilde{\mathbf{V}}_{i,11} = \widetilde{\mathbf{r}}_{i,21}\widetilde{\mathbf{R}}^{-1}_{i,11}\widetilde{\mathbf{A}}_{i,11}\end{aligned}$$

This set up is based on the idea that we want the new covariance matrix to match that from the active arm for the pre-deviation measurements and the reference arm for the conditional components for the post-deviation given the pre-deviation measurements.

Last mean carried forward (LMCF)

Under this assumption the marginal mean of the subject's distribution stays constant after deviation, at the marginal mean of their randomised treatment arm. Also, the covariance matrix remains that of the same treatment. Since, it is assumed after deviation the patient is off treatment (maintaining however a certain level of benefit achieved through treatment), a de-facto estimand is addressed. For example for a subject in treatment A, it is assumed that the marginal means stay constant after deviation:

$$\widetilde{\boldsymbol{\mu}}'_{i,A,t} = (\widetilde{\mu}_{i,A,0}, ..., \widetilde{\mu}_{i,A,d_i-1}, \widetilde{\mu}_{i,A,d_i-1}, ..., \widetilde{\mu}_{i,A,d_i-1}) \tag{27}$$

and therefore, the conditional distribution would be:

$$\begin{gathered} y_{i,T}|y_{i,0}, y^*_{i,1}, ..., y^*_{i,T-1}, D_i = T-1, g_i = A \sim \\ N(\widetilde{\mu}_{i,A,T-1} + \widetilde{\mathbf{v}}_{i,21}\widetilde{\mathbf{V}}^{-1}_{i,11}(\mathbf{y}_{i,A} - \widetilde{\boldsymbol{\mu}}_{i,A}), \widetilde{V}_{i,22} - \widetilde{\mathbf{v}}_{i,21}\widetilde{\mathbf{V}}^{-1}_{i,11}\widetilde{\mathbf{v}}_{i,12}) \end{gathered} \tag{28}$$

where $(\mathbf{y}_{i,A} - \widetilde{\boldsymbol{\mu}}_{i,A})$ is again a $(T-1 \times 1)$ vector of pre-deviation residuals. The conditional covariance matrix and the marginal components are same as in (22), (23) and (24). It should be noted, LMCF should not be confused with LOCF; LMCF does not impute the same single value for each of the post-deviation values. Rather, data are imputed from a conditional distribution with fixed post-withdrawal marginal means.

Copy increments in reference (CiR)

After deviation, the subjects' mean increments copy those from the reference group.

This assumption seeks to answer a de-facto question. Specifically, if the reference is the control arm, then the patient's mean profile following deviation, tracks that of the mean profile in the control arm, but starting from the benefit already obtained from the treatment. For a subject in treatment A the conditional distribution would be:

$$y_{i,T}|y_{i,0}, y^*_{i,1}, ..., y^*_{i,T-1}, D_i = T-1, g_i = A \sim$$
$$N(\widetilde{\mu}_{i,A,T-1} + (\widetilde{\mu}_{i,R,T} - \widetilde{\mu}_{i,R,T-1}) + \widetilde{\mathbf{v}}_{i,21}\widetilde{\mathbf{V}}^{-1}_{i,11}(\mathbf{y}_{i,A} - \widetilde{\boldsymbol{\mu}}_{i,A}), \widetilde{V}_{i,22} - \widetilde{\mathbf{v}}_{i,21}\widetilde{\mathbf{V}}^{-1}_{i,11}\widetilde{\mathbf{v}}_{i,12}) \quad (29)$$

The conditional variances are constructed in the same fashion as those under J2R. For subjects that belong to the reference group their missing data are imputed as per the randomised-arm MAR assumption.

Copy reference (CR)

As the name of the assumption implies, when subjects drop-out of the study the means and the covariance matrix of their response distribution, both before and after deviation, are replaced entirely with those from the reference arm. Hence, a de-facto question is addressed. Specifically, if the reference group is assumed to be the control, then this assumption mimics the case where those deviating do not respond to treatment. For subjects who belong in the reference arm their imputation model is that under the randomised arm MAR assumption. For instance, according to this assumption, for a subject in treatment A its conditional distribution at time T given the past, would be:

$$y_{i,T}|y_{i,0}, y^*_{i,1}, ..., y^*_{i,T-1}, D_i = T-1, g_i = A \sim$$
$$N(\widetilde{\mu}_{i,R,T} + \widetilde{\mathbf{v}}_{i,21}\widetilde{\mathbf{V}}^{-1}_{i,11}(\mathbf{y}_{i,A} - \widetilde{\boldsymbol{\mu}}_{i,R}), \widetilde{V}_{i,22} - \widetilde{\mathbf{v}}_{i,21}\widetilde{\mathbf{V}}^{-1}_{i,11}\widetilde{\mathbf{v}}_{i,12}) \quad (30)$$

Under this assumption, both the mean and the covariance structures come from the reference arm, irrespective of deviation time. Note how the residuals $(\mathbf{y}_{i,A} - \widetilde{\boldsymbol{\mu}}_{i,R})$, are measured from mean $\widetilde{\boldsymbol{\mu}}_{i,R}$ for the reference arm, rather than that for the subjects'

own arm. The covariance is therefore:

$$\widetilde{V}_{i,22} - \widetilde{\mathbf{v}}_{i,21}\widetilde{\mathbf{V}}_{i,11}^{-1}\widetilde{\mathbf{v}}_{i,12} = \widetilde{R}_{i,22} - \widetilde{\mathbf{r}}_{i,21}\widetilde{\mathbf{R}}_{i,11}^{-1}\widetilde{\mathbf{r}}_{i,12}$$

$$\widetilde{\mathbf{V}}_{i,11} = \widetilde{\mathbf{R}}_{i,11}$$

$$\widetilde{\mathbf{v}}_{i,21} = \widetilde{\mathbf{r}}_{i,21}\widetilde{\mathbf{R}}_{i,11}^{-1}\widetilde{\mathbf{V}}_{i,11} = \widetilde{\mathbf{r}}_{i,21}$$

4.1.2 Summary

What has been described so far is a method for conducting sensitivity analysis, that injects new ideas about the handling of the distributional behaviour of the subjects after deviation. The method is based on MI and estimation of the parameters proceeds within the MI paradigm too. The method takes advantage of MI's flexibility to incorporate pattern-mixture constructs into its imputation stage, and construct various MNAR assumptions about the unseen component of groups of individuals. As such, five different assumptions were developed; Randomised-arm MAR, Jump to Reference, Last Mean Carried Forward, Copy Increments in Reference, and Copy Reference. It was shown these assumptions were build by components drawn from an MAR model, within a Bayesian context.

However, as the conditional predictive distributions are pattern-mixture models and the substantive models are ANCOVA on the final time point, the imputation and analysis models are uncongenial. Hence, there is a need to explore the statistical properties of the MI variance estimator in this new setting. In the next section, a simulation study is presented which has been developed for this purpose.

5 Evaluation of the properties of sensitivity analysis

5.1 Introduction

This chapter is concerned with the establishment of the statistical properties of the CRK method. To this end, the properties of the estimators and the corresponding hypotheses tests were examined, using simulation. The simulations were based on the Alzheimer study described in Chapter 2. More specifically, they make use of the information on the sample size, parameter estimates under MAR, and drop-out patterns reported there.

5.2 Description of the simulations

Simulated data were constructed for two treatment arms, a placebo and an active treatment, and two scenarios; one where the null hypothesis of no treatment effect is true, and another one where the alternative hypothesis of treatment effect is true. As per the Alzheimer's study, a baseline and 5 subsequent time points were created.

Choice of parameters for the simulation study

The sample size of each treatment arm was similar to that in the Alzheimer's study. Specifically, the simulation was powered at 0.9, for an ADAS-cog score difference of 2 at the final time point, a significance level of 0.05 and a standard deviation of 5.74 per treatment group, as reported in the Alzheimer's study. Following this, the sample size was calculated as follows:

$$\begin{aligned} N &= \frac{2(z_{(1-\alpha/2)} + z_{(1-\beta)})var(Y_6|Y_1)}{(\mu_2 - \mu_1)^2} \\ &= \frac{2 \times (1.96 + 1.28)^2 \times 5.74^2}{2^2} \simeq 174 \end{aligned}$$

per treatment arm.

In order to choose the means for the MVN distribution, where the simulated variates would be drawn from, the following was done: a regression model, which included only the variable time (as a factor), was initially fitted to the baseline and the 5 post-randomisation measurements of ADAS-cog scores from the placebo arm of the APOE negative stratum. The model used an unstructured mean and covariance matrix. Under the null hypothesis, the LSMEANS output from this model populated the means for both the placebo and treatment arms. Under the alternative hypothesis, the marginal means for the active arm were constructed in such a way that would allow them to differ from the placebo means by a factor of 0.5 ADAS-cog score for time points 1 and 2, 1 ADAS-cog score for time points 3 and 4, and 2 ADAS-cog scores for the final time point 5. Baseline means were the same for both treatment arms. These differences were broadly consistent with the means reported in the Alzheimer's study for placebo and 2mg RSG XR, under the on treatment MAR assumption.

With respect to the covariance structure used in the MVN distribution for generating the simulated data, it is noted the treatment arms shared the same covariance matrix. In order to preserve the same sample size with the Alzheimer's study, the marginal covariance matrix Σ_{pl}, which was taken from the same regression model previously used to estimate the mean response at each time point, had to agree with the standard deviation for the final time point reported in the Alzheimer's study. It was noted that since the estimate of the standard deviation reported in the Alzheimer's study was a conditional value, a parameter c had to be found such that $c\Sigma_{pl}$ had conditional variance of the final time point given baseline of 5.74^2. Therefore,

$$\begin{aligned} 5.74^2 &= c(\Sigma_{pl,66} - \Sigma_{pl,61}(\Sigma_{pl,11})^{-1}\Sigma_{pl,16}) \\ c &= \frac{5.74^2}{(\Sigma_{pl,66} - \Sigma_{pl,61}(\Sigma_{pl,11})^{-1}\Sigma_{pl,16})} \end{aligned}$$

where $\Sigma_{pl,11}$ denotes the marginal variance of baseline placebo, $\Sigma_{pl,16}$ the marginal covariance of placebo for the final time point, and baseline and $\Sigma_{pl,66}$ the marginal

variance of placebo at the final time point. The resulting covariance matrix used in the simulations was therefore $c\Sigma_{pl}$.

Having estimated the means and covariance parameters at each time point for each treatment arm, 1000 hypothetical datasets under the null hypothesis and another 1000 datasets under the alternative hypothesis were created. Each dataset contained 348 individuals, and 6 time points. The Normal variates that populated the datasets were drawn using the RANDNORMAL() function in SAS, that took a vector of means for the 6 time points and the matrix of variance/covariance estimates, for each treatment arm separately.

The original set of 2000 datasets, Set 1, did not have any missing data. At later steps, Set 1 was manipulated, so that two other sets of datasets, Set 2 and Set 3, would be constructed. In Set 2, some variates were deleted and then filled back in, using the assumptions implied by the new sensitivity analysis. Set 3 composed of datasets with missing data. As a result, Sets 1 and 2 were analysed as 'complete' datasets, and Set 3 as incomplete datasets. It should be kept in mind that datasets whose missing values were filled back in, as well as datasets with incomplete follow-up used the same pre-deviation data taken from the original set of complete datasets. Complete datasets were analysed with either an MMRM or an ANCOVA model, whereas incomplete datasets were analysed with an ANCOVA model, after having been imputed by the new sensitivity analysis. Under the alternative hypothesis, the MMRM model was fitted assuming either a common covariance matrix in the two treatment groups (consistent with the data generating mechanism), or a separate covariance matrix in each treatment group. Further details on the 3 sets of datasets are given below.

Complete follow-up

With respect to the original set of datasets, estimates from treatment difference bias, MMRM/ANCOVA or Rubin's estimator variance, sampling variance, power or size, and coverage for the final time point over 1000 replications, for each hypothesis,

were obtained. Analyses were firstly run under a mixed MMRM model, and then under an ANCOVA model on the final time point. The MMRM model took the form:

$$Y_{ij} = \sum_{t=1}^{5} 1(t = j)\{\beta_{0j} + \beta_{1j} y_{i0} + \beta_{2j} g_i + \varepsilon_{ij}\} \tag{31}$$

where $\varepsilon_i = (\varepsilon_{i1}, ..., \varepsilon_{i5})^T$ is multivariate Normal with an unstructured covariance matrix. The justification for the functional form of the model is that, with no missing data, the estimates and standard errors of β_{05}, β_{15} and β_{25} will be the same as from an ANCOVA model.

With respect to Set 2 of the simulated datasets, missing values were created and were then replaced with data assuming either MAR or one of the 4 MNAR assumptions, namely J2R, LMCF, CiR, and CiR, encountered in section 4.1. This way, fully observed datasets under each of the different assumptions were obtained and analysed with both the MMRM and the ANCOVA models. Details on how missing data were created in the datasets are given in the next section. The post-deviation data were imputed by random draws from the appropriate conditional Normal distribution of each individual. For example, under MAR, this distribution resembles equation (25). As a result, the post-deviation data are stochastically equivalent to the original values before deleting them. With respect to the MNAR assumptions, the data were imputed by draws from the corresponding conditional Normal distributions, whose means resemble equations (26), (28), (29), and (30). All means and covariance parameters used in the construction of the various conditional Normal distributions were obtained from the marginal parameters of the data generating mechanism, described earlier.

In general, post-deviation data in Set 2 were put back in as though the various post-deviation assumptions were true. The purpose for creating this set of datasets, was to explore the behaviour of the ANCOVA and MMRM models, assuming one were fully able to observe the post-deviation data under these assumptions. Also, the results from these datasets provide a theoretical benchmark against which, results

from analyses that imply the same assumptions and applied to incomplete datasets can be judged.

Incomplete follow-up

The second phase of simulations involved creating some monotone missing data in the datasets; a stochastic process was used to allow for a random determination of a deviation time for each patient, after which all data were deleted. The missing data were created under a pattern-mixture approach. As such, there was a fixed proportion of individuals missing the final time point only, another fixed proportion of individuals missing the last two time points, and so on. Data were deleted only for the final 4 time points, leaving baseline and time point 1 fully observed across all subjects. These were MCAR missing data; each subject was assigned a constant uniform probability and based on this, they were allocated to one of the missing data patterns. The proportion of missing data in each pattern was based on the Alzheimer's study. These proportions however, were slightly adjusted to allow for the fact that datasets did not include non-monotone patterns. A higher proportion of individuals with missing data was also allowed. Hence, under the low proportion (based on the Alzheimer's study), each dataset contained approximately 20% of individuals that had incomplete follow-up, whereas under the high proportion, half of the subjects in each dataset had incomplete follow-up. The table below shows the percentages of missing data by pattern of missingness:

Patterns	**Missing measurement**	**Low %**	**High %**
D=5	No missing data	81	50
D=4	At time point 6	5.75	14.94
D=3	At time points 5 and 6	5.78	14.95
D=2	At time points 4, 5 and 6	4.6	10.34
D=1	At time points 3, 4, 5 and 6	2.87	9.77

Table 1: Patterns for the low and high proportions of missing data.

The incomplete datasets of Set 3, were firstly imputed by SAS macros, written by James Roger, that implement the pattern-mixture sensitivity analyses of section 4.1.1. The datasets that were fed into the macros contained the following variables: baseline measurement, outcome measurements for the subsequent 5 times, a time factor and a treatment factor. Under the alternative hypothesis, the macros were run with both single and separate covariance matrices for the treatment groups. For each dataset, the MCMC algorithm produced 100 imputed datasets. The thinning value for the algorithm was set to 500. For the simulations, the placebo treatment was chosen as the Reference arm in all assumptions.

The simulations focused on inference for the treatment difference at the final time point. All 1000 null-hypothesis, as well as the 1000 alternative-hypothesis datasets were fed into the macros, under each sensitivity analysis assumption. In turn, the bias $E(\bar{\theta}_m|\mathbf{y})$ of the treatment difference $\bar{\theta} = E(y_{i,6,2}|y_{i,0,2}, g_i, T_t) - E(y_{i,6,1}|y_{i,0,1}, g_i, T_t))$ was estimated, as well as the expectation of the Rubin's Rules (or imputation) variance $E(T_m|\mathbf{y})$ returned by the program, over the m iterations (or generated datasets). The latter was compared to the sampling variance of the imputation estimator $\bar{\theta}$, that is: $s^2 = \sum_{m=1}^{1000}(\bar{\theta}_m - \bar{\bar{\theta}})^2/(m-1)$. Under the null hypothesis the size of the hypothesis test was calculated, and under the alternative hypothesis the power of the methods was obtained. Finally, the coverage of the nominal 95% confidence intervals was recorded.

5.3 Results from the simulations

The tables below show the results from the simulations. They are grouped according to whether the analyses used fully observed datasets or partially observed datasets. Results from both the null hypothesis and alternative hypothesis are presented, for both the low and high proportion of missing values.

Fully observed datasets - Null hyp. - low %	**True difference**	**Mean difference at time 6 (C.I.)**	**Bias**	**Size**	**M-A/Imput. variance**	**Sampling variance**	**Coverage**
1) No missing data - analyse with MMRM	0	0.0010 (-0.0378, 0.0399)	0.0010	0.063	0.3755	0.3924	0.937
2) No missing data - analyse with ANCOVA	0	0.0064 (-0.0331, 0.0459)	0.0010	0.058	0.3752	0.4071	0.942
3) Make missing - put data back under MAR - analyse with MMRM	0	0.0094 (-0.0295, 0.0482)	0.0094	0.052	0.3750	0.3928	0.946
4) Make missing - put data back under MAR - analyse with ANCOVA	0	0.0094 (-0.0295, 0.0482)	0.0094	0.052	0.3750	0.3928	0.9468
5) Make missing - put data back under J2R - analyse with MMRM	0	-0.0006 (-0.0394, 0.0383)	-0.0006	0.059	0.3803	0.3934	0.941
6) Make missing - put data back under LMCF - analyse with MMRM	0	0.0159 (-0.0229, 0.0549)	0.0159	0.055	0.3826	0.3929	0.944
7) Make missing - put data back under CR - analyse with MMRM	0	0.0007 (-0.0373, 0.0387)	0.0007	0.054	0.3752	0.3759	0.946
8) Make missing - put data back under CiR - analyse with MMRM	0	-0.0220 (-0.0612, 0.0172)	-0.0220	0.064	0.3753	0.4001	0.936

Table 2: Table of results from methods applied onto 1000 **fully** observed simulated datasets under the **null** hypothesis and with a **low** proportion of missing data (approximately 20% of individuals in each dataset). 'M-A/Imput.' refers to either MMRM-ANCOVA or Imputation variance estimator.

Partially observed datasets - Null hyp. - low %	**True difference**	**Mean difference at time 6 (C.I.)**	**Bias**	**Size**	**M-A/Imput. variance**	**Sampling variance**	**Coverage**
9) Make missing - analyse with MMRM	0	-0.0022 (-0.0439, 0.0394)	-0.0022	0.059	0.4255	0.4520	0.94
10) Make missing - analyse with ANCOVA	0	-0.0050 (-0.0484, 0.0384)	-0.0050	0.059	0.4634	0.4912	0.941
11) Make missing - analyse with MACRO MAR	0	-0.0085 (-0.0501, 0.0333)	-0.0085	0.06	0.4231	0.4530	0.94
12) Make missing - analyse with MACRO J2R	0	-0.0088 (-0.0424, 0.0249)	-0.0088	0.02	0.4199	0.2962	0.979
13) Make missing - analyse with MACRO LMCF	0	-0.0061 (-0.0435, 0.0312)	-0.0061	0.037	0.4227	0.3632	0.963
14) Make missing - analyse with MACRO CR	0	-0.0070 (-0.0439, 0.0299)	-0.0070	0.038	0.4150	0.3545	0.961
15) Make missing - analyse with MACRO CiR	0	-0.0062 (-0.0435, 0.0311)	-0.0062	0.039	0.4153	0.3630	0.961

Table 3: Table of results from methods applied onto 1000 **partially** observed simulated datasets under the **null** hypothesis and with a **low** proportion of missing data (approximately 20% of individuals in each dataset). Table of results from methods applied onto 1000 partially observed simulated datasets. The MCMC macro, produced 100 imputations for each incomplete dataset with a thinning value of 500. 'M-A/Imput.' refers to either MMRM-ANCOVA or Imputation variance estimator.

Fully observed datasets - alternative hyp. - low %	**True difference**	**Mean difference at time 6 (C.I.)**	**Bias**	**Power**	**M-A/Imput. variance**	**Sampling variance**	**Coverage**
16) No missing data - analyse with MMRM single covariance mat.	-2	-1.9899 (-2.0285, -1.951)	0.0100	0.902	0.3769	0.3879	0.951
17) No missing data - analyse with MMRM separate cov. mat.	-2	-1.9898 (-2.0285, -1.951)	0.0101	0.902	0.3769	0.3880	0.951
18) No missing data - analyse with ANCOVA	-2	-1.9899 (-2.0285, -1.951)	0.0100	0.902	0.3769	0.3879	0.951
19) Make missing - put data back under MAR - analyse with MMRM - single cov. mat.	-2	-1.9990 (-2.0377, -1.9604)	0.0009	0.895	0.3765	0.3895	0.942
20) Make missing - put data back under MAR - analyse with MMRM - separate cov. mat.	-2	-1.9969 (-2.0350, -1.9587)	0.0030	0.901	0.3766	0.3786	0.951
21) Make missing - put data back under MAR - analyse with ANCOVA	-2	-2.0018 (-2.039, -1.9636)	-0.0018	0.906	0.3774	0.3785	0.949

Table 4: Table of results from methods applied onto 1000 **fully** observed simulated datasets under the **alternative** hypothesis and with a **low** proportion of missing data (approximately 20% of individuals in each dataset). 'M-A/Imput.' refers to either MMRM-ANCOVA or Imputation variance estimator.

Fully observed datasets - alternative hyp. - low %	**True difference**	**Mean difference at time 6 (C.I.)**	**Bias**	**Power**	**M-A/Imput. variance**	**Sampling variance**	**Coverage**
22) Make missing - put data back under J2R - analyse with MMRM - single cov. mat.	-1.62	-1.6199 (-1.6587, -1.5811)	0.00002	0.748	0.3802	0.3918	0.903
23) Make missing - put data back under J2R - analyse with MMRM - separate cov. mat.	-1.62	-1.6206 (-1.6593, -1.5819)	-0.0006	0.737	0.3803	0.3889	0.896
24) Make missing - put data back under LMCF - analyse with MMRM - single cov. mat.	-1.7726	-1.7737 (-1.8120, -1.7355)	-0.0011	0.812	0.3816	0.3810	0.939
25) Make missing - put data back under LMCF - analyse with MMRM - separate cov. mat.	-1.7726	-1.7753 (-1.8147, -1.7359)	-0.0026	0.803	0.3809	0.4042	0.931
26) Make missing - put data back under CR - analyse with MMRM - single cov. mat.	-1.7535	-1.7540 (-1.7925, -1.7155)	-0.0005	0.816	0.3788	0.3852	0.936
27) Make missing - put data back under CR - analyse with MMRM - separate cov. mat.	-1.7535	-1.7587 (-1.7971, -1.7203)	-0.0052	0.8	0.3790	0.3841	0.933
28) Make missing - put data back under CiR - analyse with MMRM - single cov. mat.	-1.7727	-1.7702 (-1.8080, -1.7325)	0.0023	0.823	0.3783	0.3707	0.934
29) Make missing - put data back under CiR - analyse with MMRM - separate cov. mat.	-1.7727	-1.7619 (-1.8005, -1.7233)	0.0107	0.817	0.3791	0.3873	0.93

Table 5: Table of results from methods applied onto 1000 **fully** observed simulated datasets under the **alternative** hypothesis and with a **low** proportion of missing data (approximately 20% of individuals in each dataset). 'M-A/Imput.' refers to either MMRM-ANCOVA or Imputation variance estimator.

Partially observed datasets alternative hyp. - low %	**True difference**	**Mean difference at time 6 (C.I.)**	**Bias**	**Power**	**M-A/Imput. variance**	**Sampling variance**	**Coverage**
30) Make missing - analyse with MMRM single covariance mat.	-2	-2.0010 (-2.0423, -1.9596)	-0.0010	0.857	0.4273	0.4454	0.94
31) Make missing - analyse with MMRM separate cov. mat.	-2	-2.0009 (-2.0423, -1.9596)	-0.0009	0.856	0.4278	0.4455	0.938
32) Make missing - analyse with ANCOVA	-2	-1.9975 (-2.0402, -1.9547)	0.0024	0.825	0.4651	0.4749	0.946
33) Make missing - analyse with MACRO MAR single covariance mat.	-2	-1.9917 (-2.0331, -1.9503)	0.0082	0.855	0.4273	0.4460	0.938
34) Make missing - analyse with MACRO MAR separate covariance mat.	-2	-2.0102 (-2.0515, -1.9688)	-0.0082	0.862	0.4283	0.4452	0.939

Table 6: Table of results from methods applied onto 1000 **partially** observed simulated datasets under the **alternative** hypothesis and with a **low** proportion of missing data (approximately 20% of individuals in each dataset). Table of results from methods applied onto 1000 partially observed simulated datasets. The MCMC macro, produced 100 imputations for each incomplete dataset with a thinning value of 500. 'M-A/Imput.' refers to either MMRM-ANCOVA or Imputation variance estimator.

Partially observed datasets alternative hyp. - low %	**True difference**	**Mean difference at time 6 (C.I.)**	**Bias**	**Power**	**M-A/Imput. variance**	**Sampling variance**	**Coverage**
35) Make missing - analyse with MACRO J2R single covariance mat.	-1.62	-1.6132 (-1.6467, -1.5797)	0.0067	0.72	0.4276	0.2920	0.976
36) Make missing - analyse with MACRO J2R separate covariance mat.	-1.62	-1.6294 (-1.6629, -1.5958)	-0.0094	0.734	0.4254	0.2933	0.977
37) Make missing - analyse with MACRO LMCF single covariance mat.	-1.7726	-1.7639 (-1.8010, -1.7269)	0.0086	0.792	0.4237	0.3571	0.964
38) Make missing - analyse with MACRO LMCF separate covariance mat.	-1.7726	-1.7660 (-1.8031, -1.7290)	0.0065	0.796	0.4216	0.3571	0.964
39) Make missing - analyse with MACRO CR single covariance mat.	-1.7535	-1.7600 (-1.7966, -1.7234)	-0.0065	0.798	0.4184	0.3483	0.978
40) Make missing - analyse with MACRO CR separate covariance mat.	-1.7535	-1.7615 (-1.7981, -1.7248)	-0.0079	0.798	0.4188	0.3495	0.962
41) Make missing - analyse with MACRO CiR single covariance mat.	-1.7727	-1.7643 (-1.8013, -1.7273)	0.0082	0.795	0.4208	0.3568	0.964
42) Make missing - analyse with MACRO CiR separate covariance mat.	-1.7727	-1.7662 (-1.8033, -1.7291)	0.0064	0.794	0.4211	0.3582	0.963

Table 7: Table of results from methods applied onto 1000 **partially** observed simulated datasets under the **alternative** hypothesis and with a **low** proportion of missing data (approximately 20% of individuals in each dataset). Table of results from methods applied onto 1000 partially observed simulated datasets. The MCMC macro, produced 100 imputations for each incomplete dataset with a thinning value of 500. 'M-A/Imput.' refers to either MMRM-ANCOVA or Imputation variance estimator.

Fully observed datasets - Null hyp. - high %	**True difference**	**Mean difference at time 6 (C.I.)**	**Bias**	**Size**	**M-A/Imput. variance**	**Sampling variance**	**Coverage**
43) Make missing - put data back under MAR - analyse with MMRM	0	0.0129 (-0.0254, 0.0513)	0.0129	0.055	0.3752	0.3837	0.944
44) Make missing - put data back under MAR - analyse with ANCOVA	0	0.0156 (-0.0237, 0.0549)	0.0156	0.058	0.3757	0.4038	0.941
45) Make missing - put data back under J2R - analyse with MMRM	0	-0.0011 (-0.0406, 0.0383)	-0.0011	0.06	0.3763	0.4060	0.939
46) Make missing - put data back under LMCF - analyse with MMRM	0	-0.0134 (-0.0521, 0.0252)	-0.0134	0.049	0.3874	0.3900	0.95
47) Make missing - put data back under CR - analyse with MMRM	0	-0.0124 (-0.0509, 0.0260)	-0.0124	0.041	0.3763	0.3865	0.959
48) Make missing - put data back under CiR - analyse with MMRM	0	-0.0035 (-0.0424, 0.0353)	-0.0035	0.054	0.3754	0.3942	0.944

Table 8: Table of results from methods applied onto 1000 **fully** observed simulated datasets under the **null** hypothesis and with a **high** proportion of missing data (approximately 20% of individuals in each dataset). 'M-A/Imput.' refers to either MMRM-ANCOVA or Imputation variance estimator.

Partially observed datasets - Null hyp. - high %	**True difference**	**Mean difference at time 6 (C.I.)**	**Bias**	**Size**	**M-A/Imput. variance**	**Sampling variance**	**Coverage**
49) Make missing - analyse with MMRM	0	-0.0104 (-0.0596, 0.0386)	-0.0104	0.062	0.5777	0.6289	0.938
50) Make missing - analyse with ANCOVA	0	-0.0259 (-0.0815, 0.0297)	-0.0259	0.058	0.7517	0.8050	0.941
51) Make missing - analyse with MACRO MAR	0	0.0212 (-0.0279, 0.0704)	0.0212	0.0660	0.5542	0.6299	0.9299
52) Make missing - analyse with MACRO J2R	0	-0.0251 (-0.0498, -0.0004)	-0.0251	0	0.5286	0.1586	1
53) Make missing - analyse with MACRO LMCF	0	0.0315 (-0.0030, 0.0661)	0.0315	0.0200	0.4798	0.3116	0.9799
54) Make missing - analyse with MACRO CR	0	-0.0203 (-0.0536, 0.0128)	-0.0203	0.014	0.4956	0.2877	0.986
55) Make missing - analyse with MACRO CiR	0	-0.0195 (-0.0541, 0.0150)	-0.0195	0.016	0.4976	0.31117	0.984

Table 9: Table of results from methods applied onto 1000 **partially** observed simulated datasets under the **null** hypothesis and with a **high** proportion of missing data (approximately 20% of individuals in each dataset). Table of results from methods applied onto 1000 partially observed simulated datasets. The MCMC macro, produced 100 imputations for each incomplete dataset with a thinning value of 500. 'M-A/Imput.' refers to either MMRM-ANCOVA or Imputation variance estimator.

Fully observed datasets - Alternative hyp. - high %	**True difference**	**Mean difference at time 6 (C.I.)**	**Bias**	**Power**	**M-A/Imput. variance**	**Sampling variance**	**Coverage**
56) Make missing - put data back under MAR - analyse with MMRM	-2	-2.0083 (-2.0465, -1.9701)	-0.0083	0.905	0.3774	0.3791	0.947
57) Make missing - put data back under MAR - analyse with ANCOVA	-2	-1.9866 (-2.0252, -1.9480)	0.0133	0.891	0.3767	0.3878	0.937
58) Make missing - put data back under J2R - analyse with MMRM	-1	-0.9907 (-1.0283, -0.9531)	0.0092	0.357	0.3824	0.3679	0.955
59) Make missing - put data back under LMCF - analyse with MMRM	-1.3995	-1.3920 (-1.4291, -1.3549)	0.0074	0.605	0.3841	0.3578	0.952
60) Make missing - put data back under CR - analyse with MMRM	-1.3462	-1.3408 (-1.3783, -1.3032)	0.0053	0.584	0.3800	0.3669	0.954
61) Make missing - put data back under CiR - analyse with MMRM	-1.3994	-1.3873 (-1.4251, -1.3495)	0.0120	0.626	0.3796	0.3722	0.949

Table 10: Table of results from methods applied onto 1000 **fully** observed simulated datasets under the **alternative** hypothesis and with a **high** proportion of missing data (approximately 20% of individuals in each dataset). 'M-A/Imput.' refers to either MMRM-ANCOVA or Imputation variance estimator.

Partially observed datasets - Alternative hyp. - high %	**True difference**	**Mean difference at time 6 (C.I.)**	**Bias**	**Power**	**M-A/Imput. variance**	**Sampling variance**	**Coverage**
62) Make missing - analyse with MMRM	-2	-1.9879 (-2.0347, -1.9410)	0.0121	0.73	0.5803	0.5713	0.946
63) Make missing - analyse with ANCOVA	-2	-1.9770 (-2.0294, -1.9247)	0.0229	0.617	0.7549	0.7126	0.952
64) Make missing - analyse with MACRO MAR	-2	-2.0392 (-2.0861, -1.9924)	-0.0392	0.751	0.5765	0.5715	0.943
65) Make missing - analyse with MACRO J2R	-1	-1.0130 (-1.0364, -0.9895)	-0.0130	0.125	0.5361	0.1436	0.991
66) Make missing - analyse with MACRO LMCF	-1.3995	-1.3604 (-1.3933, -1.3274)	0.0390	0.483	0.47709	0.2825	0.987
67) Make missing - analyse with MACRO CR	-1.3462	-1.3550 (-1.3866, -1.3234)	-0.0088	0.457	0.5017	0.2599	0.992
68) Make missing - analyse with MACRO CiR	-1.3994	-1.4117 (-1.4446, -1.3787)	-0.0123	0.497	0.5022	0.2825	0.991

Table 11: Table of results from methods applied onto 1000 **partially** observed simulated datasets under the **alternative** hypothesis and with a **high** proportion of missing data (approximately 20% of individuals in each dataset). Table of results from methods applied onto 1000 partially observed simulated datasets. The MCMC macro, produced 100 imputations for each incomplete dataset with a thinning value of 500. 'M-A/Imput.' refers to either MMRM-ANCOVA or Imputation variance estimator.

With no missing data, the MMRM and the ANCOVA models in Table 2, models 1 and 2, and Table 4, models 16-18, returned almost identical results. There was no estimator bias and the type I error, the power and the coverage achieved the nominal levels, as they should by design. The MMRM/ANCOVA variance returned from these models was also unbiased; it agreed very closely with the sampling variance.

With partially observed datasets and no imputation, shown in Tables 3, 6, 9, and 11, models 9-10, 30-32, 49-50, and 62-63, the ANCOVA model was less efficient compared to the MMRM model. Under the alternative hypothesis, with partially observed datasets and no imputation, the ANCOVA model was both less efficient and more underpowered than the MMRM model. This was the case under both the low and the high proportion of missing data (Tables 6 and 11). The MMRM model performed better because, as Brown and Prescott (2006) explain, it uses a covariance matrix which is specified to describe the within-subject correlations across the time points. So, observations at each time point influence estimates of treatment effects at every other time point. Therefore, the observed values of subjects who drop-out of the study will nevertheless be taken into account at later time points.

With both a low and a high proportion of missing data, no estimator bias was detected with any of the models applied to datasets that create missing data and then filled back in. In Tables 2, 4, 5, 8, and 10, whenever there are fully observed datasets, whose missing values are put back in according to the sensitivity analysis assumptions, it can be seen the sampling and MMRM/ANCOVA variances agree well. In Table 4, the results from the MAR models 16-18 that were fitted to datasets with no missing data, agreed very closely with the MAR models 19-21, that were fitted to datasets whose missing data were filled back in. A loss of power can be observed in Table 6, for the MAR models 30-32, fitted on datasets with missing data, when compared to table 4, models 16-18; the power appears to be around 85% when there were missing data, compared to 90% when datasets were complete. This was natural, since with missing data there is always loss of information.

More precisely, with respect to type II error, it is worth pointing out how power

decreases for the MMRM model from 90% in model 16 with no missing data, to 85% in model 30 with a low rate of missing data, to 73% in model 62 with a high rate of missing data. The power from the pattern-mixture methods (models 33-42 for the low proportion of missing values and 64-68 for the high proportion of missing data) was generally lower than the models which used the true means and covariances, fitted in datasets whose missing data were put back in (methods 19-29 and 56-61, respectively). Among the MACRO models, those under the MNAR assumptions had a lower power when compared to the MACRO MAR models. The MACRO MAR achieved values closer to 0.9, and agreed very well with the MMRM analyses for the incomplete datasets, whereas the greatest drop in power can be observed with the MACRO J2R model, especially with a high proportion of missing data, method 65.

With respect to the MACRO MAR models, the results show models 11, 33, and 34 agreed closely with those from the MMRM models 9, 30, and 31 under both the null and alternative hypotheses. This was also true with a high proportion of missing data and expected, as these analyses were carried out under the MAR assumption. In addition, the rate of coverage as well as the actual rejection rate were very close to the nominal level.

When post-deviation data are missing, the appropriate variance estimator is the one that reflects the loss of information, by returning a variance estimate greater than the corresponding variance estimate that would have been obtained, if data were fully observed. It can be seen from Table 3, under the MAR assumption, models 9-11, both variance estimators reflect the loss of information by returning higher variance estimates when compared to the respective variance estimates in Table 2, models 1-4, fitted in fully observed datasets. However, under the MNAR assumptions, only Rubin's Rules variance estimator reflect appropriately the loss of information, since it appears inflated compared to the respective variance estimate in Table 2. Therefore, the sampling variances in Tables 3 and 9, models 12-15, and 52-55, which are less than the respective ones in Tables 2 and 8, models 5-8, and 45-48, are inappropriate. At the same time, since Rubin's variance estimator returns

greater values than the empirical estimator of the sampling variances does, the CIs will have a frequency coverage greater than the nominal $100(1-0.05)\%$ (Table 3, models 12-15).

The same is true under the alternative hypothesis as well, for both low and high rates of missing data. It can be seen from Table 7, and under the MNAR models 35-42, how the value from Rubin's variance estimator is the appropriate estimate to use; a comparison of the variance estimates between Tables 7 and 5 implies Rubin's variance estimator appropriately account for the loss of information, as opposed to the sampling variance estimate, which is decreased (also seen when comparing models 65 to 58, 66 to 59, 67 to 60 and 68 to 61). Moreover, with a high rate of missing data in Tables 8 and 9 that refer to datasets under the null hypothesis, it becomes discernible how the increase in Rubin's variance has a noticeable impact on the size as well (MACRO MNAR methods J2R, LMCF, CR and CiR, table 9, models 52-55).

The discrepancies that were found between the sampling variances of the MNAR macros 12-15, 35-42, 52-55, and 65-68, and the sampling variances of the respective models 5-8, 22-29, 45-48, and 58-61 under the fully observed datasets, is due to two reasons. Firstly, with respect to the models in the fully observed datasets (5-8, 22-29, 45-48 and 58-61), under the various assumptions, the post deviation distribution in each arm has a mixture of means (with the same variance about each mean), so that the overall variance about a common mean in each arm is increased. The second reason is the existence of a correlation that is induced by estimating the conditional means for the missing responses using data from the reference treatment. To show that, let the variance of the treatment difference be expressed as:

$$\text{var}(A-R) = \text{var}(A) + \text{var}(R) - 2\text{cov}(A,R)$$

With no missing data $\text{cov}(A,R) = 0$. However, with missing data there exists a correlation in the estimation of the variance of the treatment difference, because the two treatments shared the same conditional variances used in the distribution from

which the missing data were drawn. An analytic expression of the existence of this covariance that shows that the variance of the estimated treatment difference with dropout is less than the variance of the treatment difference with fully observed data is presented in Appendix 2.

On a different note, the discrepancy between Rubin's variance and the sampling variance estimates, seen in the MACRO MNAR models 12-15 and 35-42, for the low proportion of missing data, and models 52-55 and 65-68, for the high proportion of missing data, can be attributed to the first form of uncongeniality discussed in section 3.4.2.3; the MNAR pattern-mixture models contain a 'richer' structure in terms of the estimated parameters in each treatment group at each time, to that in the substantive ANCOVA model used for the final analyses. On the other hand, under MAR, both imputation and substantive models imply the same structure; there, the two variances are similar, reflecting the fact that the imputation and substantive models are congenial.

With regards to coverage, all results from models 12-15, 35-42, 52-55 and 65-68 were now somewhat overestimated compared to the MAR assumption based models. This was due to the increase in Rubin's Rules variance estimator. The results across the datasets with a low proportion of missing data showed that models with separate covariances did not yield different values to those with single covariance. For this reason, models with separate covariances were not used under the high proportion of missing data datasets.

As far as the alternative hypothesis is concerned, in order to calculate the biases and the rates of coverage, the true value of the difference of the estimates at the final time point, implied by the pattern-mixture models, was used. The derivation of the true values is presented in Appendix 1. It can be seen the true values for the MACRO MNAR models are all different from -2.

The results show that with a high proportion of missing data, with a sample size of 348 and with 100 imputations, a very small bias was introduced in the MACRO J2R and LMCF, methods 52 and 66, as the values of the true difference was slightly

outside their confidence interval. However, when these simulations were rerun with 250 imputations, the bias in both cases disappeared (J2R difference estimate (C.I.): -0.0178 (-0.0425,0.0067), LMCF difference estimate (C.I.):-1.3855 (-1.4138, -1.3590)). The disappearance of the bias that was due to Monte Carlo error, agrees with published research that advices on running more imputations as the number of missing values increases (for example, White et al. (2011)). It is worth noting that the estimates for the rest of the quantities returned after 250 imputations, remained the same.

5.4 Summary

The statistical properties of the proposed sensitivity analysis method described in Chapter 4 were explored in this chapter using simulation. A number of hypothetical datasets were constructed, populated with Normal variates, based on information from the Alzheimer's study presented in Chapter 2. Six time point included and 174 subjects were allowed per treatment arm. Datasets were created both under the null and alternative hypothesis, allowing for both a low and a high proportion of missing observations. As well as incomplete datasets, fully observed datasets were created after having deleted missing data, by putting back in values drawn by models that were based on the MNAR assumptions described in Chapter 4. These datasets were analysed with MMRM models. The creation of these kind of fully observed datasets was done in order to facilitate comparisons of the results with those obtained when incomplete datasets were analysed by the equivalent assumptions of the new sensitivity method.

The results from the simulations show that the new method is a sensible tool for sensitivity analyses. No evidence of bias for the treatment difference estimator was detected either under de-jure or de-facto questions. The size, the power, the sampling variance as well as Rubin's variance of the MACRO MAR methods were similar to those obtained from the MMRM model when fitted to incomplete datasets, as expected. With respect to the MACRO MNAR models however, there were some

noticeable discrepancies between the estimates returned by the sampling and Rubin's variance. When those variances were compared to their equivalent estimates, from models fitted in datasets whose missing values had been populated based on the corresponding assumptions, it became obvious that Rubin's variance was always increased, whereas the estimator of the sampling variance consistently returned decreased estimates. Therefore, Rubin's estimator of variance is the appropriate one in this setting, as it properly reflected the loss of information from missing data. Also, in terms of size and power the estimates returned from the MACRO MNAR models were slightly lower, but in accordance with their theoretical equivalent from the fully observed datasets.

After running simulations to examine the statistical properties of the CRK method, the Alzheimer's study will be analysed using the MNAR assumptions of the proposed method. This will test the robustness of the MAR results obtained in the original study to the various deviations. The next chapter attempts to identify the predictors that drive the missingness in the Alzheimer's dataset, and Chapter 7 presents the results from the sensitivity analyses.

6 Missing data in the Alzheimer's dataset

In this chapter, the Alzheimer's study is revisited. The primary analysis, carried out for the ADAS-cog score, is repeated, and then the new sensitivity analysis methodology is applied to examine the robustness of the inferences under a range of assumptions about the statistical post-deviation behaviour of the subjects in the dataset.

As previously mentioned, in order to reduce biases in the imputation model and the standard errors of the estimates of the substantive model, the imputation model may be extended by the inclusion of additional auxiliary variables. At the beginning, following Carpenter and Kenward (2007), logistic regression and survival analyses were used to identify key independent predictors of deviation in the Alzheimer's dataset, discussed in Chapter 2. Both analyses deployed a backward and forward selection procedure. A significance level of 10% was adopted. The Full analysis population of all randomised patients was used. All baseline variables in the dataset were considered. The exception was 'child bearing potential' which was highly correlated with age and sex.

Firstly, unadjusted log odds and log hazard ratios were calculated. With regards to the logistic regression, the response variable was chosen to be the binary indicator of whether a subject completed the study or not. A subject would be deemed to have completed the study, if they were present at the final time point. With regards to the survival analysis, the time-to-event variable was defined as the time up to completion or withdrawal, measured in days. As withdrawal was considered to be the event in the time-to-event analysis, all patients who completed the study were censored (coded as 0 in the program), otherwise they had the event (coded as 1 in the program).

The results from the likelihood ratio tests for the effect of each variable in the logistic models showed that the following variables were significant at the 10% level: treatment (p-value=0.0021), age (p-value=<.0001), country (p-value=<.0001), race (p-value=0.0146), disease history of relatives (p-value=0.0183), time to diagnosis

(p-value=0.0033). The log odds for 2mg RSG XR and 8mg RSG XR were -0.3699 (s.e.=0.1619) and 0.1765 (s.e.=0.1510), implying that patients who received 2mg RSG XR as opposed to placebo were less likely to withdraw, and patients who received 8mg RSG XR as opposed to placebo were more likely to withdraw. With respect to 'country', the results showed that the log odds ratio for 'USA+CANADA' was standing out at 1.0664 (s.e.=0.2950), suggesting that withdrawal in this category was mainly driven by the combination of these two countries.

The picture from the unadjusted hazard ratios was the same. The variables that were found to be significant predictors of withdrawal were: treatment (p-value=0.0020), age (p-value <.0001), country (p-value <.0001), race (p-value=0.0032), disease history of relatives (p-value=0.0130), time to diagnosis (p-value=0.0032). However, with the time-to-event analysis, baseline mmse score was also found to be significant at the 10% level (p-value=0.060).

Having noted the unadjusted results above, full forward and backward analyses were conducted. The results are displayed in Tables 12 and 13. Both logistic and survival forward analyses started by including 'country' as the first variable. This was because it returned the lowest p-value in the unadjusted analyses. Regarding the logistic analyses, both forward and backward selection processes identified treatment, country, age, and baseline mmse score as predictors. Under the logistic models, it can be seen that the forward selection process included 'time to diagnosis', while the backward selection did not. The term that appears in brackets in the logistic forward selection process, denotes the next most suitable candidate term to be added on to the process, whereas the bracketed term in the logistic backward selection process denotes the last term omitted by the process.

Regarding the results from the survival analyses, the predictors common to both forward and backward selection processes were again treatment, country, age, and baseline mmse score. The term 'race' is in brackets in the forward selection process, because it became non-significant after the inclusion of the final term 'disease history of relatives', as the association between 'race' and 'disease history of relatives' was

significant: χ^2 p-value=0.04. The bracketed term in the survival backward selection process was the last one omitted by the process.

For both logistic and survival forward analyses, the displayed estimates and p-values are those obtained after fitting the model with all variables in. The reference group for the categorical covariates in Tables 12 and 13 are: *Placebo* for *Treatment*, *Argentina* & *Brazil* & *Chile* & *Mexico* & *India* for *Country*, *Positives* for *APOE Status* and finally, *African American* for *Race*. The *Negatives* category in the *APOE status* includes the following allele combinations: $\varepsilon 3.\varepsilon 3$, $\varepsilon 2.\varepsilon 3$ and $\varepsilon 2.\varepsilon 2$. The *Positives* category includes: $\varepsilon 3.\varepsilon 4$, $\varepsilon 2.\varepsilon 4$ and $\varepsilon 4.\varepsilon 4$.

Forward Logistic			Backward Logistic		
Predictors	Estimate (s.e.)	p-value	Predictors	Estimate (s.e.)	p-value
Treatment		0.002	**Treatment**		0.014
2 mg RSG XR	-0.415 (0.167)		2 mg RSG XR	-0.314 (0.180)	
8 mg RSG XR	0.155 (0.156)		8 mg RSG XR	0.203 (0.172)	
Country		<.001	**Country**		<.001
HUN/POL/CZE/SUI	-0.343 (0.323)		HUN/POL/CZE/SUI	-0.402 (0.335)	
GRE/ESP/POR	-0.656 (0.344)		GRE/ESP/POR	-0.747 (0.375)	
USA/CAN	0.804 (0.244)		USA/CAN	0.732 (0.259)	
JPN	-0.614 (0.343)		JPN	-0.714 (0.349)	
FRA	0.001 (0.295)		FRA	-0.009 (0.324)	
GER	0.339 (0.261)		GER	0.367 (0.289)	
ITA	0.193 (0.265)		ITA	0.063 (0.281)	
AUT	-0.214 (0.339)		AUT	-0.339 (0.366)	
Age	0.032 (0.009)	<.001	**Age**	0.028 (0.010)	0.003
MMSE score	-0.031 (0.016)	0.058	**MMSE score**	-0.034 (0.017)	0.053
Time to diag.	0.073 (0.041)	0.078	(**Relatives hist.**)		(0.139)
			No	-0.340 (0.227)	
(**APOE status**)		(0.116)			
Negatives	0.211 (0.134)				

Table 12: Results of forward and backward stepwise logistic analyses. HUN; Hungary, Pol; Poland, CZE; Czech Republic, SUI; Switzerland, GRE; Greece, ESP; Spain, POR; Portugal, CAN; Canada, JPN; Japan, FRA; France, GER; Germany, ITA; Italy, AUT; Austria.

Forward Survival			Backward Survival		
Predictors	Estimate (s.e.)	p-value	Predictors	Estimate (s.e.)	p-value
Country		<.0001	**Country**		<.0001
HUN/POL/CZE/SUI	-0.305 (0.316)		HUN/POL/CZE/SUI	-0.368 (0.301)	
GRE/ESP/POR	-0.604 (0.355)		GRE/ESP/POR	-0.664 (0.342)	
USA/CAN	0.654 (0.236)		USA/CAN	0.572 (0.217)	
JPN	0.514 (1.064)		JPN	-0.733 (0.320)	
FRA	0.002 (0.295)		FRA	-0.045 (0.281)	
GER	0.347 (0.263)		GER	0.303 (0.246)	
ITA	0.111 (0.262)		ITA	0.049 (0.244)	
AUT	-0.283 (0.339)		AUT	-0.342 (0.326)	
Treatment		0.0159	**Treatment**		0.0137
2 mg RSG XR	-0.282 (0.157)		2 mg RSG XR	-0.287 (0.156)	
8 mg RSG XR	0.157 (0.144)		8 mg RSG XR	0.159 (0.144)	
Age	0.026 (0.008)	0.0017	**Age**	0.026 (0.008)	0.0015
MMSE score	-0.027 (0.015)	0.0675	**MMSE score**	-0.026 (0.015)	0.0731
Relatives hist.		0.0853	**(Relatives hist.)**		(0.1231)
No	-0.315 (0.183)		No	-0.282 (0.183)	
(Race)		(0.4098)			
American Indian/ Alaskan Native	1.530 (1.121)				
Central/S. Asian	0.821 (1.439)				
Asian/S.E. Asian	1.551 (1.441)				
White/Arabic/ North African	2.432 (1.241)				
White/Caucasian/ European	1.190 (1.012)				

Table 13: Results of forward and backward stepwise analyses. HUN; Hungary, Pol; Poland, CZE; Czech Republic, SUI; Switzerland, GRE; Greece, ESP; Spain, POR; Portugal, CAN; Canada, JPN; Japan, FRA; France, GER; Germany, ITA; Italy, AUT; Austria.

Following on from the discussion in Chapter 3, after having identified the key baseline predictors of deviation, important baseline predictors of the incomplete outcome variable in the Alzheimer's study were also sought. In order to do this, an MMRM model was fitted to the full analysis set. This model had the same structure as in section 5.2 that is, it regressed ADAS-cog score on full time-baseline and time-treatment interactions plus a baseline variable. For instance, if the aim was to check whether the covariate 'country' was a significant predictor of the outcome (or whether it had any significant association with the outcome), then the MMRM model would include full time-baseline and time-treatment interactions plus 'country'. As a result, a number of MMRM models were fitted that included a different baseline variable each time. For the comparison between placebo and 2mg RSG XR, the covariates that significantly predicted the incomplete outcome were found to be: country (p-value= <0.001), mmse score (p-value= <0.001), race (p-value=0.052), ethnicity (being hispanic/latino or not) (p-value=0.036) and time to diagnosis (p-value=0.070). For the comparison between placebo and 8mg RSG XR the corresponding significant predictors were: country (p-value= <0.0001), ethnicity (being hispanic/latino or not) (p-value=0.013) and mmse score (p-value= <0.0001).

The results from the analyses so far suggest that country, mmse score, ethnicity and time to diagnosis are all potentially useful auxiliary variables. All these variables (except ethnicity) were found to predict both outcome and deviation. However, as the purpose was to make inferences about treatment effect, these variables were further tested to see if adjusting for them, in the MMRM model 31, would alter the estimated treatment effect at the final time point. The justification for doing this is, if these variables were not to change the inference about the treatment effect on the outcome under an MAR model, then there is no reason to believe they would change this inference when fitted as auxiliary variables in an MAR imputation analysis either. So, the additional work of including them in the sensitivity analysis is unlikely to be worthwhile.

To this end, the impact of all covariates in the dataset (not just the auxiliary variables previously identified) on treatment was tested, by calculating confidence inter-

vals around the treatment estimates obtained firstly, from the original MMRM model 31, and secondly, from the ones that included an additional variable, and checking whether these intervals overlap. If the intervals obtained from the MMRM models with the additional variable overlap with that from the original MMRM model, that would mean the treatment effect on the outcome had not changed. The treatment effect from the original model 31 was also compared against two more elaborate models: i) a MMRM model as reported in the Harrington et al. study, which included the additional variables of country, mmse score, mmse by time, APOE copies, and bmi, and ii) a model (termed 'combo' in the figures below) which utilised the auxiliary variables previously found to be significant that is, country, mmse score, race, ethnicity and time to diagnosis for the comparison between 2mg RSG XR arm and placebo, and country, mmse score and ethnicity for the comparison between 8mg RSG XR arm and placebo.

In Figures 1 and 2 the original treatment estimate is plotted along with its 95% C.I., together with the corresponding results after adjusting for each of the covariates in turn.

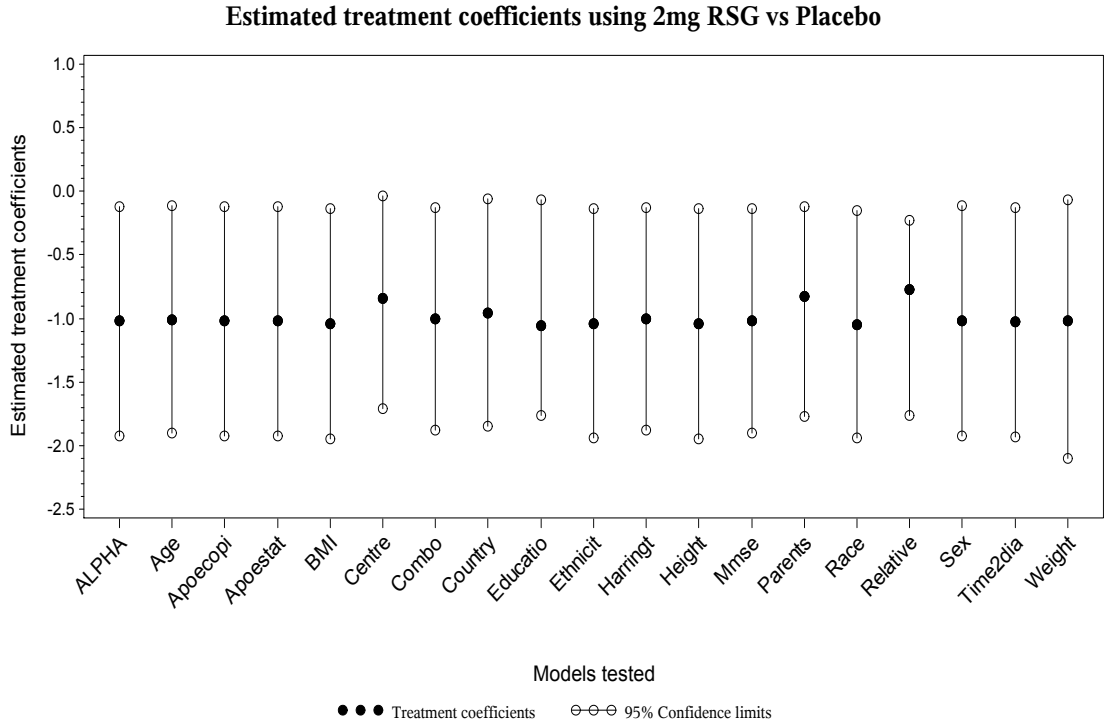

Figure 1: Points represents treatment coefficients from the original MMRM model (Alpha) and the MMRM models with the additional variable. Lines represent the corresponding confidence intervals. The labels along the x-axis refer to the additional variable included in the MMRM models. The model from the Harrington study is termed 'Harrington', whereas the model including all the significant predictors of the outcome is termed 'Combo'.

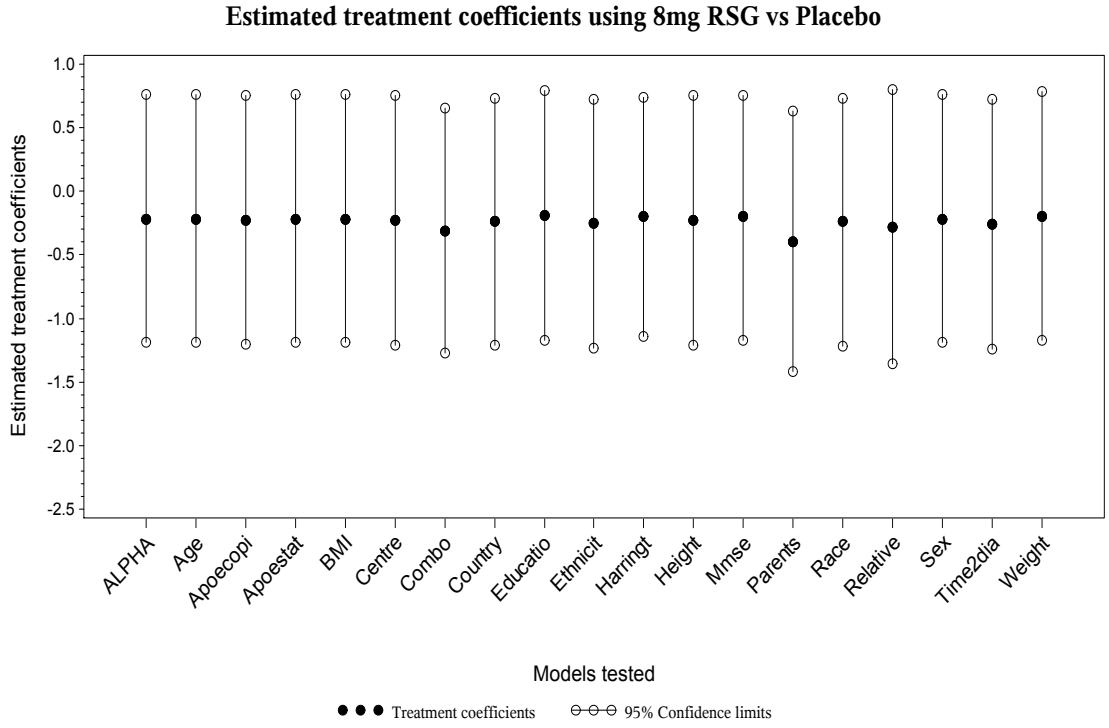

Figure 2: Points represents treatment coefficients from the original MMRM model (Alpha) and the MMRM models with the additional variable. Lines represent the corresponding confidence intervals. The labels along the x-axis refer to the additional variable included in the MMRM models. The model from the Harrington study is termed 'Harrington', whereas the model including all the significant predictors of the outcome is termed 'Combo'.

Considering the comparison between 2mg RSG XR and placebo, it can be seen that none of the treatment estimates, in models with country, mmse score, race, ethnicity and time to diagnosis, are significantly different to the treatment estimate obtained from the original MMRM model. Actually, none of the covariates (or combination of covariates) change dramatically the effect of treatment. The same conclusions can be drawn from the comparison between 8mg RSG XR and placebo. Appendix C presents the estimates of the treatment effects, after fitting the additional covariates, as well as the estimates of the covariates themselves.

6.1 Summary

This chapter was concerned with finding suitable auxiliary variables for the imputation analysis of the Alzheimer's data. The aim was to find predictors of deviation, as well as predictors of the incomplete outcome, and finally test whether these would

alter the inference around the treatment estimate. At the beginning, for the identification of predictors of deviation, unadjusted logistic and survival analyses were conducted, and the significant covariates were noted. These were then put into a logistic and a survival model, and stepwise processes for the elimination of non-significant covariates were carried out. Pooling the results from both logistic and survival analyses together, the identified predictors of deviation were: country, age, mmse score, time-to-diagnosis, and relatives history of the disease. Moreover, the predictors of outcome, after fitting an MMRM model, were found to be: country, mmse score, race, ethnicity, and time-to-diagnosis for the low dose comparison, and country, ethnicity, and mmse score for the high dose comparison. It was noted however from further analyses, that none of these, nor any other baseline variables altered the treatment estimate significantly. As such, it was decided not to use any of these covariates as auxiliary variables.

7 Sensitivity analysis for the Alzheimer's study

In this chapter, a sensitivity analysis for the treatment difference at the final visit in the Alzheimer's study, introduced in Chapter 2, is performed and discussed. The aim is to apply the new sensitivity analysis to explore the robustness of these conclusions to various assumptions about the post-deviation behaviour.

Analyses are performed for all three analysis populations in the Alzheimer's study; APOE $\varepsilon 4$ negative patients, all patients except those who are homozygous for the $\varepsilon 4$ genotype, and the full patient population. The Harrington study, showed there was no significant difference at the final time point between 8 mg RSG XR and placebo in change from baseline in any of these three analysis populations. As a result, the authors did not proceed to formally test the comparison between 2 mg RSG XR and placebo within the hierarchical process. They offer however results from 'exploratory' analyses between 2 mg RSG XR and placebo; there, they state 'a small to moderate potential benefit of 2 mg RSG XR was suggested' (APOE $\varepsilon 4$ negative: -1.3 ADAS-cog score points; $p = 0.049$, All except $\varepsilon 4/\varepsilon 4$ genotype: -1.0 ADAS-cog score points; $p = 0.035$, Full population: -1.0 ADAS-cog score points; $p = 0.02$). The authors note however in the document, 'these results should be interpreted with caution because they were not statistically significant in light of the hierarchical procedure which was employed to control type I error over multiple statistical tests'. In the same study, ADAS-cog scores in the placebo group declined over time. At the final time they had declined by 3.4 points since baseline.

As the proportion of individuals with missing data was approximately 20%, for each sensitivity analysis 100 imputations were created updating the sampler 500 times between each imputed dataset. All MAR and the four MNAR assumptions were deployed. The MAR analysis target the de-jure estimand, in other words, what treatment effect is expected if post-deviation patients continued to adhere to treatment as specified in the protocol. The J2R assumption addresses the de-facto question that post-deviation patients would receive a different treatment from the one they were randomised to. Usually, patients with chronic conditions who receive

placebo would switch to an active treatment. Therefore, under the placebo vs 2mg RSG XR comparison, J2R would allow placebo patients who deviate, to jump to 2mg RSG XR, whereas under the placebo vs 8mg RSG XR comparison, placebo patients who deviate jump to 8mg RSG XR. However, in this setting this assumption is rather unfair to the treatments, and are not useful for evaluating whether the drugs work. Instead, it shows the effect of a pragmatic post-deviation switch to active treatment for the placebo patients.

The LMCF assumption assumes that patients after deviation did not receive any treatment, but that their condition remained stable around their treatment group mean at deviation. This is a plausible assumption, if it is believed the treatment has managed to control the course of the Alzheimer's disease.

The CiR assumption uses the placebo treatment as the reference treatment. Under CiR, patients on the active treatment follow their randomised profile prior to deviation, but after deviation the marginal means in the conditional distribution change according to changes observed in the placebo group. This assumption is suited to situations where it is believed that patients cease receiving treatment after drop-out and as in the case of the Alzheimer's disease, it is generally known that without any treatment, patients display a steady deterioration. Under CR, when imputing post deviation data for an 8mg say, patient who deviates, the joint distribution of their pre- and post-deviation data is replaced entirely with those from the placebo arm. For the Alzheimer's disease data, the placebo treatment was used as reference, reflecting this way, a situation where someone on an active treatment does not respond to it at all.

To run the imputations, three subjects with missing baseline measurements were deleted. Moreover, the LMCF and CiR assumptions, required that the first measurement after baseline be fully observed in order to run. As a result, all subjects that were missing their first ADAS-cog measurement were deleted, that is 20 patients for the comparison between placebo and 2mg RSG XR and 47 patients for the comparison between placebo and 8mg RSG XR.

Following the conclusions of the previous chapter, auxiliary variables were not included in the imputation model for the sensitivity analyses. Hence, the structure of the imputation model involved only time as a categorical factor, treatment, baseline, a baseline by time and a treatment by time interaction.

The results from the SAS MACROS for the treatment difference at the final time point, under the three analysis populations are shown in Tables 14-19. Under the 'J2R' method the '2mg RSG XR' and '8mg RSG XR' treatments act as the 'reference' arm, where 'placebo' patients 'jump to', whereas the reference arm under 'CR' and 'CiR' is placebo. The results from the Harrington's study are also displayed.

Analysis	2mg RSG XR vs Placebo		Harrington's Result	
	Estimate (St. error)	p-value	Estimate	95% C.I.
MAR	-1.3006 (0.6837)	0.0580	-1.3	(-2.7, -0.0)
J2R	-1.0631 (0.6911)	0.1250		
LMCF	-1.1301 (0.6956)	0.1052		
CR	-1.1721 (0.6804)	0.0859		
CiR	-1.2669 (0.6909)	0.0676		

Table 14: Results from the APOE $\varepsilon 4$ negative subpopulation, for the 2mg RSG XR vs Placebo comparison.

Analysis	2mg RSG XR vs Placebo		Harrington's Result	
	Estimate (St. error)	p-value	Estimate	95% C.I.
MAR	-0.9903 (0.4670)	0.0440	-1.0	(-1.9, -0.1)
J2R	-0.7398 (0.4689)	0.1151		
LMCF	-0.8919 (0.4689)	0.0576		
CR	-0.9029 (0.4686)	0.0544		
CiR	-0.9698 (0.4657)	0.0377		

Table 15: Results from the All except $\varepsilon 4/\varepsilon 4$ genotype subpopulation, for the 2mg RSG XR vs Placebo comparison.

Analysis	2mg RSG XR vs Placebo		Harrington's Result	
	Estimate (St. error)	p-value	Estimate	95% C.I.
MAR	-1.0122 (0.4575)	0.0272	-1.0	(-1.9, -0.2)
J2R	-0.7881 (0.4512)	0.0811		
LMCF	-0.9935 (0.4608)	0.0314		
CR	-0.9394 (0.4573)	0.0403		
CiR	-0.9932 (0.4574)	0.0302		

Table 16: Results from the Full population, for the 2mg RSG XR vs Placebo comparison

Analysis	8mg RSG XR vs Placebo		Harrington's Result	
	Estimate (St. error)	p-value	Estimate	95% C.I.
MAR	-0.1835 (0.4941)	0.7105	-0.2	(-1.7, 1.3)
J2R	-0.1252 (0.4965)	0.8002		
LMCF	-0.3196 (0.4910)	0.5154		
CR	-0.1722 (0.4926)	0.7268		
CiR	-0.1367 (0.5000)	0.7847		

Table 17: Results from the APOE $\varepsilon 4$ negative subpopulation, for the 8mg RSG XR vs Placebo comparison.

Analysis	8mg RSG XR vs Placebo		Harrington's Result	
	Estimate (St. error)	p-value	Estimate	95% C.I.
MAR	-0.0823 (0.5189)	0.8740	0.0	(-1.0, 1.0)
J2R	-0.0183 (0.5414)	0.9730		
LMCF	-0.2446 (0.5243)	0.6408		
CR	-0.0805 (0.5251)	0.8783		
CiR	-0.0075 (0.5273)	0.9887		

Table 18: Results from the All except $\varepsilon 4/\varepsilon 4$ genotype subpopulation, for the 8mg RSG XR vs Placebo comparison.

Analysis	8mg RSG XR vs Placebo		Harrington's Result	
	Estimate (St. error)	p-value	Estimate	95% C.I.
MAR	-0.2290 (0.5027)	0.6488	-0.2	(-1.2, 0.7)
J2R	-0.1797 (0.4919)	0.7150		
LMCF	-0.3349 (0.4957)	0.4996		
CR	-0.1124 (0.4996)	0.8221		
CiR	-0.1330 (0.4963)	0.7888		

Table 19: Results from the Full population, for the 8mg RSG XR vs Placebo comparison.

Under the MAR assumption the results above agree with those found in the Harrington study: the high dose treatment did not show a statistically significant effect when compared to placebo, in any population. This was the case for all comparisons under all analysis populations, and under all MNAR assumptions.

With regards to the low dose comparison, it can be seen, under MAR, the results across all populations agree again with those from the Harrington study. The 2mg RSG XR treatment showed an effect at a 5% level at the final visit. Under J2R, the 2mg RSG XR dose had a significant effect at the 10% level under the 'Full population', but not under the other two populations. Generally, the differences were more pronounced in the 'Full population', since this dataset had the highest number of data, and hence the most information. With regards to the LMCF assumption, a significant effect at the 10% level in favour of the 2mg RSG XR across all populations exists, and under the 'Full population' the same treatment is showing an effect at the 5% level. Under CR, the comparison shows a significant difference at the 10% level in all populations, as does under CiR, as well. With regards to the latter assumption, 2mg RSG XR exhibits a difference at the 5% level under the 'all except $\varepsilon 4$ genotype homozygous', and the 'Full population' populations.

In general, the results above imply that inferences from the initial MAR methods are robust to the sensitivity analysis assumptions. Especially, in the case of the high dose treatment, inferences are uniform across all analyses. The size of the differences between the low dose RSG XR and placebo, seen in Harrington's study, was maintained in the results presented here too. These differences were described as small to moderate. They were obtained from analyses that were not subjected to the more rigorous inferential process, which controlled for type I error, since 8 RSG XR was not found statistically different from placebo. This, coupled with the fact the treatments were taken on top of the main therapy, makes it hard to assert the clinical superiority of the low-dose treatment. In Harrington's study, the authors claim the difference seen in favour of the low dose treatment, but not in favour of the high dose treatment, was "unexpected". Scientifically, "the effect of RSG XR in AD is thought to be via enhanced glucose uptake into the brain", and as such the high dose treatment is expected to confer more of a benefit. It was further argued, the small supremacy of 2mg RSG XR over placebo, may well be due to chance, as this finding was not replicated in a different study.

7.1 Summary

In this chapter, the new MI sensitivity analysis was applied to the Alzheimer's dataset. The analyses populations were split according to criteria set in the original Alzheimer's study, and both de-jure and de-facto estimands were estimated under the 5 available assumptions, the method allowed at the time of writing. None of the results from the sensitivity analysis showed any difference between 8mg RSG XR and placebo. On the other hand, there was a general tendency in favour of the low dose RSG XR when compared to placebo, but this did not meet a clinically meaningful cut-off point, and has not been adjusted for rigorous control of type I error. The MAR findings which imply the high dose treatment is not any different from placebo, is robust to de-facto sensitivity analyses. Also, it is fairly robust under the low dose treatment comparisons, and agrees very closely with the CiR assumption, which is perhaps one of the most plausible assumptions in this setting.

8 Discussion

This thesis considered the analysis of clinical trials with continuous longitudinal outcomes, when not all patients adhere to the protocol. It is widely acknowledged that in this setting, inferences depends on untestable assumptions, so that sensitivity analyses, as a means to testing inferences from such assumptions, is vital. However, accessible and relevant methods for conducting sensitivity analyses in this setting are lacking. The aim of this thesis was to evaluate and then apply to an Alzheimer's trial a new approach to sensitivity analysis, which frames accessible and relevant assumptions in which, post-deviation, patients's data are imputed by reference to another treatment arm, or group, in the study.

In the beginning, the data from an Alzheimer's study were introduced. These data served to motivate this thesis. They came from a randomised controlled trial, which was set up to examine whether the daily addition of rosiglitazone extended release tablets to donepezil treated Alzheimer's patients for 48 weeks, could slow the development of the disease, as measured by ADAS-cog scores. The analysis was done on three separate populations; the 'APOE $\varepsilon 4$ negative' subpopulation, the 'All except $\varepsilon 4/\varepsilon 4$ genotype' subpopulation and the 'Full population'. The study suffered from a non trivial proportion of drop-out, with 29%, 22%, and 33% missing data at the final visit in the placebo, 2mg RSG XR, and 8mg RSG XR groups, respectively. As the main tool for analysis in the study was an MMRM model, which implies an MAR assumption, this thesis investigated the robustness of inference from these data to different assumptions about patients' behaviour post-deviation.

In Chapter 3, notation and basic principles of missing data are introduced, within the longitudinal data setting. The occurrence of missing data in clinical trials necessitates a careful consideration of the population which the treatment effect is estimated for, and for this reason two new estimands were introduced by Carpenter et al. (2013); de-jure and de-facto. A de-jure estimand answers questions such as, what the treatment effect be on average in the target population of eligible patients, if treatment and control were taken as specified in the protocol. On the other hand,

a de-facto estimand aims to answer questions about what would be the effect seen in practice, if the treatment was assigned to the target population of eligible patients, as defined in the trial's inclusion criteria. The new estimands clarify the target of estimation, by carefully describing the population under which the analysis is conducted. Within this framework, it was explained that the analysis of longitudinal per-protocol measurements addresses a de-jure question, and that a traditional ITT analysis should best be seen within an MNAR framework, which seeks to answer 'de-facto' questions.

In order to analyse a trial and make inferences for these estimands, a primary analysis assumption should be identified followed by a series of alternative assumptions to assess the robustness of the inferences. Therefore, in Chapter 4, a description of the CRK sensitivity analysis, which makes assumptions about post-deviation behaviour under de-jure and de-facto estimands by reference to other patient group in the study, was presented. It was shown this approach was based on MI for parameter estimation and inference, and that the incorporation of the different de-facto assumptions was achieved via the manipulation of the predictive distribution of the imputation model.

The first main contribution of this thesis was to thoroughly explore the statistical properties of the new method via simulation. To this end, a series of simulation studies were set up based on the motivating Alzheimer's data. Two different sets of simulated data were created; one under the null hypothesis and another one under the alternative hypothesis. Within each set, datasets were divided into datasets with no missing data, datasets whose missing data were filled back in using the assumptions implied by the new sensitivity analysis, and finally datasets with missing data. Moreover, datasets with missing data were allowed to have a small or a high proportion of missing values (20% and 50%). Datasets with complete follow-up were analysed with either an MMRM or an ANCOVA model, whereas datasets with incomplete follow-up were analysed with an ANCOVA model after having been imputed by the new sensitivity analysis. In all datasets, the bias, size or power, Rubin's and empirical variances, as well as the coverage were calculated. The principal

results were as follows:

Under the MAR assumption, the inferences drawn from the MACRO MAR models, agreed closely with those from the MMRM analyses on datasets with missing data. This was true for both the null and alternative hypotheses datasets, as well as for the low and high rates of missing data datasets. No evidence of bias for the treatment difference estimator was found. Under the MAR assumption, no bias was found for Rubin's Rules variance either. This variance was very similar to the empirical variance; the very small observed differences between the two variance estimates were due to Monte Carlo variability. Under the null hypothesis, the MAR method attained coverages and sizes very close to the nominal levels for both low and high proportions of missing data. Under the alternative hypothesis, the power values for the new method under the MAR assumption, were a little lower than 90%, reflecting the loss of information when missing data exist. It however remained close to the power returned by the MMRM analyses, for both low and high missing data proportions. In relation to coverage, the MACRO MAR methods, for either hypotheses and proportions of missing data, achieved the nominal levels.

With respect to the MNAR methods, no estimator bias was found. In order to estimate the various quantities in the simulations under the alternative hypotheses, the true expected value for the difference between the two treatments, for each MNAR scenario, had to be calculated. This theoretical value, that was operating within a pattern-mixture approach, is different to the original value for treatment differences that was allowed in the simulations by construction.

In the null hypothesis datasets, the sizes attained by the MNAR models of the new method were similar to the equivalent sizes obtained from fully observed datasets. With a large rate of missing data however there were some noticeable differences. This resulted by the fact that Rubin's variance increased in the absence of fully observed dataset. In the alternative hypothesis datasets, the power returned by the MNAR models of the new method agreed too with the powers from the fully observed datasets, but this agreement started to disappear with a higher amount of

missing data in the datasets.

It was noted, when missing data exist the appropriate variance estimator should reflect the loss of information in the dataset, and hence, should be greater than the respective variance estimate if data are complete. Although under MAR, both Rubin's and sampling variance estimators were able to reflect this loss of information (they both increased compared to the respective variance estimates under fully observed scenarios), it was only Rubin's variance estimator that appropriately accounted for it under the MNAR assumptions, as well. By contrast, under the MNAR scenarios, the sampling variance was decreasing.

It was also shown that the discrepancies between the sampling variances of the MACRO MNAR models and the sampling variances of the respective MNAR models under fully observed datasets, were attributed to two operations; first was the fact that in the fully observed datasets the variance about a common mean, derived from a mixture of means, is increased and second was a correlation that enters the formula for the variance of the treatment difference. This correlation exists due to the fact the method is using information from the reference arm to impute the active arm. The use of a greater number of imputations did not alleviate this problem.

All the above, suggest the new method is a valid technique for conducting sensitivity analyses with missing data, with sensible statistical properties. It returns unbiased point estimates with inflated standard errors that appropriately take into account the loss of information. Inherent in principled analyses with missing data is the fact that a power to detect an effect is reduced and this was also reflected in this investigation.

The new sensitivity analysis method was applied to a published Alzheimer's study. Analyses were conducted in order to identify auxiliary variables for the imputation models. The results showed that none of the tested covariates would alter the inference for treatment, and therefore it was decided the imputation models should run with a baseline by time and treatment by time interaction only. Under the de-jure estimand, no evidence of treatment difference was discernible from the comparison

between the high dose treatment and placebo. Some evidence of treatment difference existed when the low dose treatment was compared to placebo. This finding though, should be interpreted with caution, as the differences were obtained from analyses not subjected to the rigorous inferential process that was used in the Harrington's study, and, as the authors argued, the finding may have been due to chance, and it was not replicated in a different study..

Under the de-facto estimands, the results showed that inferences, under the same treatment comparison, were broadly similar to those obtained from the de-jure analyses. This showed robustness of the primary inference to different assumptions about patients' behaviour after deviation. Given that gradually the condition of Alzheimer's patients who stop taking treatment deteriorates, the most plausible de-facto assumption was deemed to be CiR. The CiR assumption agreed closely with inferences from the MAR assumption. It was clear that any comparison between 8mg RSG XR and placebo did not provide any evidence towards the superiority of the active treatment over placebo.

In summary, this thesis was set up to investigate the statistical behaviour of the new method for sensitivity analysis proposed by Carpenter et al. (2013), and to apply the method onto a published Alzheimer's clinical trial, with a non-trivial proportion of missing data. In order to explore the statistical properties of the method, a number of simulations were conducted, with fully observed datasets that had missing data and were filled in assuming 'J2R', 'LMCF', 'CR', and 'CiR', and with datasets whose post-deviation data were missing, and imputed under the same assumptions using MI. It was shown the sampling variance of the estimator was much smaller than that from Rubin's rules. The sampling variance under imputation was also much smaller than the sampling variance obtained, if post-deviation data were observed, something that was not true with Rubin's rules variance. The latter was always higher with missing data rather than with fully observed datasets, reflecting the loss of information due to missing data. Therefore, in this context, Rubin's rules variance estimates are preferable. The second part of the thesis demonstrated the application of the methodology in an Alzheimer's study. The results corroborated

the original findings of the de-jure analyses, since these were fairly robust to all the de-facto analyses.

A key theoretical challenge is yet to be explored. This is to relate Rubin's variance estimator to the proportion of information lost by not seeing the post-deviation data, and hence give the new method a stronger theoretical justification.

In conclusion, the new sensitivity analysis proposed by Carpenter et al. (2013) is a valid method which permits the construction of transparent, relevant and accessible assumptions for the post-deviation behaviour of the subjects in longitudinal settings. Its validity was demonstrated by simulation and its applicability by real trial data.

A Appendix

True pattern-mixture treatment differnce under the alternative hypothesis

This section illustrates how the true value at the final time-point under the alternative hypothesis is obtained, according to the various constraints implied by the sensitivity assumptions in section 4.1.1. The means for the Reference and Active treatments are:

$$\begin{aligned}\text{Reference} &= 25.2569,\quad -0.1545,\quad -0.05384,\quad 0.9923,\quad 1.3188,\quad 2.5282\\ \text{Active} &= 25.2569,\quad -0.6545,\quad -0.55384,\quad -0.0077,\quad 0.3188,\quad 0.5282\end{aligned}$$

Before applying the assumptions, the conditional distributions of the final time point given the previous time points, split by the various patterns, are presented. The patterns are defined by $D \in \{1,2,3,4,5\}$. Distributions with non-identified components (whose parameters do not appear in the observed data likelihood) are marked with ☒. For simplicity, notation on subject i has been suppressed.

☒ $f(y_6|y_1,y_2,y_3,y_4,y_5,g=j,D=1) \sim$
$N(b_0^{(D=1)} + b_1^{(D=1)}y_1 + b_2^{(D=1)}y_2 + b_3^{(D=1)}y_3 + b_4^{(D=1)}y_4 + b_5^{(D=1)}y_5, \tau^{2(D=1)})$

☒ $f(y_6|y_1,y_2,y_3,y_4,y_5,g=j,D=2) \sim$
$N(b_0^{(D=2)} + b_1^{(D=2)}y_1 + b_2^{(D=2)}y_2 + b_3^{(D=2)}y_3 + b_4^{(D=2)}y_4 + b_5^{(D=2)}y_5, \tau^{2(D=2)})$

☒ $f(y_6|y_1,y_2,y_3,y_4,y_5,g=j,D=3) \sim$
$N(b_0^{(D=3)} + b_1^{(D=3)}y_1 + b_2^{(D=3)}y_2 + b_3^{(D=3)}y_3 + b_4^{(D=3)}y_4 + b_5^{(D=3)}y_5, \tau^{2(D=3)})$

☒ $f(y_6|y_1,y_2,y_3,y_4,y_5,g=j,D=4) \sim$
$N(b_0^{(D=4)} + b_1^{(D=4)}y_1 + b_2^{(D=4)}y_2 + b_3^{(D=4)}y_3 + b_4^{(D=4)}y_4 + b_5^{(D=4)}y_5, \tau^{2(D=4)})$

✓ $f(y_6|y_1,y_2,y_3,y_4,y_5,g=j,D=5) \sim$
$N(b_0^{(D=5)} + b_1^{(D=5)}y_1 + b_2^{(D=5)}y_2 + b_3^{(D=5)}y_3 + b_4^{(D=5)}y_4 + b_5^{(D=5)}y_5, \tau^{2(D=5)})$

where τ^2 denotes the conditional variance.

The conditional mean of the un-identified distribution for pattern 1 can be expressed as:

$$\begin{aligned}
\mu_{6,j}^{(D=1)} &= E_{y_6}(y_6|D=1, g=j) \\
&= E_{y_5}[E_{y_4}[E_{y_3}[E_{y_2}[E_{y_1}[E_{y_6}(y_6|y_1, y_2, y_3, y_4, y_5, D=1, g=j)|D=1, g=j]]]]] \\
&= E_{y_5}[E_{y_4}[E_{y_3}[E_{y_2}[E_{y_1}[b_{0,j}^{(D=1)} + b_{1,j}^{(D=1)} y_1 + b_{2,j}^{(D=1)} y_2 + b_{3,j}^{(D=1)} y_3 \\
&+ b_{4,j}^{(D=1)} y_4 + b_{5,j}^{(D=1)} y_5|D=1, g=j]]]]] \\
&= b_{0,j}^{(D=1)} + b_{1,j}^{(D=1)} \mu_{1,j}^{(D=1)} + b_{2,j}^{(D=1)} \mu_{2,j}^{(D=1)} + b_{3,j}^{(D=1)} \mu_{3,j}^{(D=1)} + b_{4,j}^{(D=1)} \mu_{4,j}^{(D=1)} + b_{5,j}^{(D=1)} \mu_{5,j}^{(D=1)}
\end{aligned}$$

where the intercept:

$$b_{0,j}^{(D=1)} = \mu_{6,j}^{(D=1)} - b_{1,j}^{(D=1)} \mu_{1,j}^{(D=1)} - b_{2,j}^{(D=1)} \mu_{2,j}^{(D=1)} - b_{3,j}^{(D=1)} \mu_{3,j}^{(D=1)} - b_{4,j}^{(D=1)} \mu_{4,j}^{(D=1)} - b_{5,j}^{(D=1)} \mu_{5,j}^{(D=1)}$$

Similarly, for the rest of the patterns:

$$\begin{aligned}
\mu_{6,j}^{(D=2)} &= b_{0,j}^{(D=2)} + b_{1,j}^{(D=2)} \mu_{1,j}^{(D=2)} + b_{2,j}^{(D=2)} \mu_{2,j}^{(D=2)} + b_{3,j}^{(D=2)} \mu_{3,j}^{(D=2)} + b_{4,j}^{(D=2)} \mu_{4,j}^{(D=2)} + b_{5,j}^{(D=2)} \mu_{5,j}^{(D=2)} \\
&\vdots \\
\mu_{6,j}^{(D=5)} &= b_{0,j}^{(D=5)} + b_{1,j}^{(D=5)} \mu_{1,j}^{(D=5)} + b_{2,j}^{(D=5)} \mu_{2,j}^{(D=5)} + b_{3,j}^{(D=5)} \mu_{3,j}^{(D=5)} + b_{4,j}^{(D=5)} \mu_{4,j}^{(D=5)} + b_{5,j}^{(D=5)} \mu_{5,j}^{(D=5)}
\end{aligned}$$

The marginal mean for each treatment is given by:

$$\begin{aligned}
\mu_{6,j} &= E(y_6|g=j) = E_D[E_{y_6}(y_6|D=d, g=j)|g=j] \\
&= \sum_{D=1}^{5} E_{y_6}(y_6|D=d, g=j) \pi_D^{g=j} \qquad (32)
\end{aligned}$$

So, the true marginal difference between the Active treatment and the Reference treatment, at the final time point, is:

$$\mu_{6,A} - \mu_{6,R} \qquad (33)$$

Randomised arm MAR

The population means across the different patterns for each treatment arm are equal. Under the MAR assumption the marginal population treatment difference was calculated as:

$$
\begin{aligned}
&E_{y_6}(y_6|D=1, g=j, MAR) \\
&\quad = b_{0,j}^{(D=5)} + b_{1,j}^{(D=5)}\mu_{1,j}^{(D=1)} + b_{2,j}^{(D=5)}\mu_{2,j}^{(D=1)} + b_{3,j}^{(D=5)}\mu_{3,j}^{(D=1)} + b_{4,j}^{(D=5)}\mu_{4,j}^{(D=1)} + b_{5,j}^{(D=5)}\mu_{5,j}^{(D=1)} \\
&\quad = \mu_{6,j}^{(D=5)} + b_{1,j}^{(D=5)}(\mu_{1,j}^{(D=1)} - \mu_{1,j}^{(D=5)}) + b_{2,j}^{(D=5)}(\mu_{2,j}^{(D=1)} - \mu_{2,j}^{(D=5)}) + b_{3,j}^{(D=5)}(\mu_{3,j}^{(D=1)} - \mu_{3,j}^{(D=5)}) \\
&\quad + b_{4,j}^{(D=5)}(\mu_{4,j}^{(D=1)} - \mu_{4,j}^{(D=5)}) + b_{5,j}^{(D=5)}(\mu_{5,j}^{(D=1)} - \mu_{5,j}^{(D=5)}) \\
&\quad = \mu_{6,j}^{(D=5)} \\
&\quad \vdots \\
&E_{y_6}(y_6|D=4, g=j, MAR) \\
&\quad = b_{0,j}^{(D=5)} + b_{1,j}^{(D=5)}\mu_{1,j}^{(D=4)} + b_{2,j}^{(D=5)}\mu_{2,j}^{(D=4)} + b_{3,j}^{(D=5)}\mu_{3,j}^{(D=4)} + b_{4,j}^{(D=5)}\mu_{4,j}^{(D=4)} + b_{5,j}^{(D=5)}\mu_{5,j}^{(D=4)} \\
&\quad = \mu_{6,j}^{(D=5)} + b_{1,j}^{(D=5)}(\mu_{1,j}^{(D=4)} - \mu_{1,j}^{(D=5)}) + b_{2,j}^{(D=5)}(\mu_{2,j}^{(D=4)} - \mu_{2,j}^{(D=5)}) + b_{3,j}^{(D=5)}(\mu_{3,j}^{(D=4)} - \mu_{3,j}^{(D=5)}) \\
&\quad + b_{4,j}^{(D=5)}(\mu_{4,j}^{(D=4)} - \mu_{4,j}^{(D=5)}) + b_{5,j}^{(D=5)}(\mu_{5,j}^{(D=4)} - \mu_{5,j}^{(D=5)}) \\
&\quad = \mu_{6,j}^{(D=5)}
\end{aligned}
$$

Following (32), for the low proportion of missing data we have:

$$
\begin{aligned}
\mu_{6,R} &= \mu_{6,R}^{(D=5)}\pi_1 + \mu_{6,R}^{(D=5)}\pi_2 + \mu_{6,R}^{(D=5)}\pi_3 + \mu_{6,R}^{(D=5)}\pi_4 + \mu_{6,R}^{(D=5)}\pi_5 \\
&= (2.5282 \times 0.0287) + (2.5282 \times 0.0460) + (2.5282 \times 0.0578) \\
&\quad + (2.5282 \times 0.0575) + (2.5282 \times 0.81) \\
&= 2.5282 \\
\mu_{6,A} &= \mu_{6,A}^{(D=5)}\pi_1 + \mu_{6,A}^{(D=5)}\pi_2 + \mu_{6,A}^{(D=5)}\pi_3 + \mu_{6,A}^{(D=5)}\pi_4 + \mu_{6,A}^{(D=5)}\pi_5 \\
&= (0.5282 \times 0.0287) + (0.5282 \times 0.0460) + (0.5282 \times 0.0578) \\
&\quad + (0.5282 \times 0.0575) + (0.5282 \times 0.81) \\
&= 0.5282
\end{aligned}
$$

and for the high proportion:

$$\begin{aligned}
\mu_{6,R} &= \mu_{6,R}^{(D=5)}\pi_1 + \mu_{6,R}^{(D=5)}\pi_2 + \mu_{6,R}^{(D=5)}\pi_3 + \mu_{6,R}^{(D=5)}\pi_4 + \mu_{6,R}^{(D=5)}\pi_5 \\
&= (2.5282 \times 0.0977) + (2.5282 \times 0.1034) + (2.5282 \times 0.1495) \\
&+ (2.5282 \times 0.1494) + (2.5282 \times 0.5) \\
&= 2.5282 \\
\mu_{6,A} &= \mu_{6,A}^{(D=5)}\pi_1 + \mu_{6,A}^{(D=5)}\pi_2 + \mu_{6,A}^{(D=5)}\pi_3 + \mu_{6,A}^{(D=5)}\pi_4 + \mu_{6,A}^{(D=5)}\pi_5 \\
&= (0.5282 \times 0.0977) + (0.5282 \times 0.1034) + (0.5282 \times 0.1495) \\
&+ (0.5282 \times 0.1494) + (0.5282 \times 0.5) \\
&= 0.5282
\end{aligned}$$

Following (33), the true difference in both cases is -2.

Jump to reference

Under J2R, the conditional means for the Active treatment take the following form:

$$\begin{aligned}
&E_{y_6}(y_6|D=1, g=A, J2R) \\
&\quad= \mu_{6,R}^{(D=5)} + b_{1,A\cdot R}^{(D=5)}(\mu_{1,A}^{(D=1)} - \mu_{1,A}^{(D=5)}) + b_{2,A\cdot R}^{(D=5)}(\mu_{2,A}^{(D=1)} - \mu_{2,A}^{(D=5)}) + b_{3,A\cdot R}^{(D=5)}(\mu_{3,A}^{(D=1)} - \mu_{3,A}^{(D=5)}) \\
&\quad+ b_{4,A\cdot R}^{(D=5)}(\mu_{4,A}^{(D=1)} - \mu_{4,A}^{(D=5)}) + b_{5,A\cdot R}^{(D=5)}(\mu_{5,A}^{(D=1)} - \mu_{5,A}^{(D=5)}) \\
&\quad= \mu_{6,R}^{(D=5)} \\
&\quad\vdots \\
&E_{y_6}(y_6|D=4, g=A, J2R) \\
&\quad= \mu_{6,R}^{(D=5)} + b_{1,A\cdot R}^{(D=5)}(\mu_{1,A}^{(D=4)} - \mu_{1,A}^{(D=5)}) + b_{2,A\cdot R}^{(D=5)}(\mu_{2,A}^{(D=4)} - \mu_{2,A}^{(D=5)}) + b_{3,A\cdot R}^{(D=5)}(\mu_{3,A}^{(D=4)} - \mu_{3,A}^{(D=5)}) \\
&\quad+ b_{4,A\cdot R}^{(D=5)}(\mu_{4,A}^{(D=4)} - \mu_{4,A}^{(D=5)}) + b_{5,A\cdot R}^{(D=5)}(\mu_{5,A}^{(D=4)} - \mu_{5,A}^{(D=5)}) \\
&\quad= \mu_{6,R}^{(D=5)}
\end{aligned}$$

since the true means within each treatment were the same across the patterns. The composite coefficients $b_{A\cdot R}$ formed from components of the Active and Reference treatments. However in this case, their construction was straightforward since a single covariance matrix was used

to generate the data for both Active and Reference treatments.

Therefore, for the low proportion of missing data, we have:

$$\begin{aligned}
\mu_{6,R} &= \mu_{6,R}^{(D=5)}\pi_1 + \mu_{6,R}^{(D=5)}\pi_2 + \mu_{6,R}^{(D=5)}\pi_3 + \mu_{6,R}^{(D=5)}\pi_4 + \mu_{6,R}^{(D=5)}\pi_5 \\
&= (2.5282 \times 0.0287) + (2.5282 \times 0.0460) + (2.5282 \times 0.0578) \\
&+ (2.5282 \times 0.0575) + (2.5282 \times 0.81) \\
&= 2.5282 \\
\mu_{6,A} &= \mu_{6,R}^{(D=5)}\pi_1 + \mu_{6,R}^{(D=5)}\pi_2 + \mu_{6,R}^{(D=5)}\pi_3 + \mu_{6,R}^{(D=5)}\pi_4 + \mu_{6,A}^{(D=5)}\pi_5 \\
&= (2.5282 \times 0.0287) + (2.5282 \times 0.0460) + (2.5282 \times 0.0578) \\
&+ (2.5282 \times 0.0575) + (0.5282 \times 0.81) \\
&= 0.9082
\end{aligned}$$

Hence, the true difference is -1.62. For the high proportion of missing data we have:

$$\begin{aligned}
\mu_{6,R} &= \mu_{6,R}^{(D=5)}\pi_1 + \mu_{6,R}^{(D=5)}\pi_2 + \mu_{6,R}^{(D=5)}\pi_3 + \mu_{6,R}^{(D=5)}\pi_4 + \mu_{6,R}^{(D=5)}\pi_5 \\
&= (2.5282 \times 0.0977) + (2.5282 \times 0.1034) + (2.5282 \times 0.1495) \\
&+ (2.5282 \times 0.1494) + (2.5282 \times 0.5) \\
&= 2.5282 \\
\mu_{6,A} &= \mu_{6,R}^{(D=5)}\pi_1 + \mu_{6,R}^{(D=5)}\pi_2 + \mu_{6,R}^{(D=5)}\pi_3 + \mu_{6,R}^{(D=5)}\pi_4 + \mu_{6,A}^{(D=5)}\pi_5 \\
&= (2.5282 \times 0.0977) + (2.5282 \times 0.1034) + (2.5282 \times 0.1495) \\
&+ (2.5282 \times 0.1494) + (0.5282 \times 0.5) \\
&= 1.5282
\end{aligned}$$

Thus, the true difference is -1.

Last mean carried forward

This assumptions implies:

$$
\begin{aligned}
&E_{y_6}(y_6|D=1, g=j, LMCF) \\
&\qquad = \mu_{2,j}^{(D\geq 1)} + b_{1,j}^{(D\geq 1)}(\mu_{1,j}^{(D=1)} - \mu_{1,j}^{(D\geq 1)}) + b_{2,j}^{(D\geq 1)}(\mu_{2,j}^{(D=1)} - \mu_{2,j}^{(D\geq 1)}) \\
&\qquad + b_{3,j}^{(D\geq 1)}(\mu_{3,j}^{(D=1)} - \mu_{3,j}^{(D\geq 1)}) + b_{4,j}^{(D\geq 1)}(\mu_{4,j}^{(D=1)} - \mu_{4,j}^{(D\geq 1)}) + b_{5,j}^{(D\geq 1)}(\mu_{5,j}^{(D=1)} - \mu_{5,j}^{(D\geq 1)}) \\
&\qquad = \mu_{2,j}^{(D\geq 1)} \\
&\qquad \vdots \\
&E_{y_6}(y_6|D=4, g=j, LMCF) \\
&\qquad = \mu_{5,j}^{(D\geq 4)} + b_{1,j}^{(D\geq 4)}(\mu_{1,j}^{(D=4)} - \mu_{1,j}^{(D\geq 4)}) + b_{2,j}^{(D\geq 4)}(\mu_{2,j}^{(D=4)} - \mu_{2,j}^{(D\geq 4)}) \\
&\qquad + b_{3,j}^{(D\geq 4)}(\mu_{3,j}^{(D=4)} - \mu_{3,j}^{(D\geq 4)}) + b_{4,j}^{(D\geq 4)}(\mu_{4,j}^{(D=4)} - \mu_{4,j}^{(D\geq 4)}) + b_{5,j}^{(D\geq 4)}(\mu_{5,j}^{(D=4)} - \mu_{5,j}^{(D\geq 4)}) \\
&\qquad = \mu_{5,j}^{(D\geq 4)}
\end{aligned}
$$

since the true means within each treatment were the same across the patterns.

For the low proportion of missing data, we obtain:

$$
\begin{aligned}
\mu_{6,R} &= \mu_{2,R}^{(D\geq 1)}\pi_1 + \mu_{3,R}^{(D\geq 2)}\pi_2 + \mu_{4,R}^{(D\geq 3)}\pi_3 + \mu_{5,R}^{(D\geq 4)}\pi_4 + \mu_{6,R}^{(D=5)}\pi_5 \\
&= (-0.1545 \times 0.0287) - (0.0538 \times 0.0460) + (0.9923 \times 0.0578) \\
&\quad + (1.3188 \times 0.0575) + (2.5282 \times 0.81) \\
&= 2.1741 \\
\mu_{6,A} &= \mu_{2,A}^{(D\geq 1)}\pi_1 + \mu_{3,A}^{(D\geq 2)}\pi_2 + \mu_{4,A}^{(D\geq 3)}\pi_3 + \mu_{5,A}^{(D\geq 4)}\pi_4 + \mu_{6,A}^{(D=5)}\pi_5 \\
&= (-0.6545 \times 0.0287) - (0.5538 \times 0.0460) - (0.0077 \times 0.0578) \\
&\quad + (0.3188 \times 0.0575) + (0.5282 \times 0.81) \\
&= 0.4015
\end{aligned}
$$

Hence, the true difference is -1.7726. For the respective difference under the high proportion of missing data, we have:

$$\begin{aligned}
\mu_{6,R} &= \mu_{2,R}^{(D\geq 1)}\pi_1 + \mu_{3,R}^{(D\geq 2)}\pi_2 + \mu_{4,R}^{(D\geq 3)}\pi_3 + \mu_{5,R}^{(D\geq 4)}\pi_4 + \mu_{6,R}^{(D=5)}\pi_5 \\
&= (-0.1545 \times 0.0977) - (0.0538 \times 0.1034) + (0.9923 \times 0.1495) \\
&+ (1.3188 \times 0.1494) + (2.5282 \times 0.5) \\
&= 1.5888 \\
\mu_{6,A} &= \mu_{2,A}^{(D\geq 1)}\pi_1 + \mu_{3,A}^{(D\geq 2)}\pi_2 + \mu_{4,A}^{(D\geq 3)}\pi_3 + \mu_{5,A}^{(D\geq 4)}\pi_4 + \mu_{6,A}^{(D=5)}\pi_5 \\
&= (-0.6545 \times 0.0977) - (0.5538 \times 0.1034) - (0.0077 \times 0.1495) \\
&+ (0.3188 \times 0.1494) + (0.5282 \times 0.5) \\
&= 0.1894
\end{aligned}$$

Hence, the true difference is -1.3995.

Copy reference

Under CR, the residuals component of the conditional means is measured from the mean of the reference arm, rather than that of the subjects' own arm. This means that the non-identified conditional distributions of the Active treatment will be formed as follows:

$$\begin{aligned}
&E_{y_6}(y_6|D=d, g=A, CR) \\
&= E_{y_5}[E_{y_4}[E_{y_3}[E_{y_2}[E_{y_1}[E_{y_6}(y_6|y_1,y_2,y_3,y_4,y_5, D=d, g=A, CR)|D=d, g=A, CR]]]]] \\
&= E_{y_5}[E_{y_4}[E_{y_3}[E_{y_2}[E_{y_1}[b_{0,j} + b_{1,j}y_1 + b_{2,j}y_2 + b_{3,j}y_3 \\
&+ b_{4,j}y_4 + b_{5,j}y_5|D=d, g=A, CR]]]]] \\
&= b_{0,R} + b_{1,R}\mu_{1,A} + b_{2,R}\mu_{2,A} + b_{3,R}\mu_{3,A} + b_{4,R}\mu_{4,A} + b_{5,R}\mu_{5,A}
\end{aligned}$$

where the intercept:

$$b_{0,R} = \mu_{6,R} - b_{1,R}\mu_{1,R} - b_{2,R}\mu_{2,R} - b_{3,R}\mu_{3,R} - b_{4,R}\mu_{4,R} - b_{5,R}\mu_{5,R}$$

So, we get:

$$
\begin{aligned}
&E_{y_6}(y_6|D=1,g=A,CR)\\
&\quad=\mu_{6,R}^{(D=5)}+b_{1,R}^{(D=5)}(\mu_{1,A}^{(D=1)}-\mu_{1,R}^{(D=5)})+b_{2,R}^{(D=5)}(\mu_{2,A}^{(D=1)}-\mu_{2,R}^{(D=5)})\\
&\quad+b_{3,R}^{(D=5)}(\underline{\mu_{3,A}}^{(D=1)}-\mu_{3,R}^{(D=5)})+b_{4,R}^{(D=5)}(\underline{\mu_{4,A}}^{(D=1)}-\mu_{4,R}^{(D=5)})+b_{5,R}^{(D=5)}(\underline{\mu_{5,A}}^{(D=1)}-\mu_{5,R}^{(D=5)})\\
&E_{y_6}(y_6|D=2,g=A,CR)\\
&\quad=\mu_{6,R}^{(D=5)}+b_{1,R}^{(D=5)}(\mu_{1,A}^{(D=2)}-\mu_{1,R}^{(D=5)})+b_{2,R}^{(D=5)}(\mu_{2,A}^{(D=2)}-\mu_{2,R}^{(D=5)})\\
&\quad+b_{3,R}^{(D=5)}(\mu_{3,A}^{(D=2)}-\mu_{3,R}^{(D=5)})+b_{4,R}^{(D=5)}(\underline{\mu_{4,A}}^{(D=2)}-\mu_{4,R}^{(D=5)})+b_{5,R}^{(D=5)}(\underline{\mu_{5,A}}^{(D=2)}-\mu_{5,R}^{(D=5)})\\
&E_{y_6}(y_6|D=3,g=A,CR)\\
&\quad=\mu_{6,R}^{(D=5)}+b_{1,R}^{(D=5)}(\mu_{1,A}^{(D=3)}-\mu_{1,R}^{(D=5)})+b_{2,R}^{(D=5)}(\mu_{2,A}^{(D=3)}-\mu_{2,R}^{(D=5)})\\
&\quad+b_{3,R}^{(D=5)}(\mu_{3,A}^{(D=3)}-\mu_{3,R}^{(D=5)})+b_{4,R}^{(D=5)}(\mu_{4,A}^{(D=3)}-\mu_{4,R}^{(D=5)})+b_{5,R}^{(D=5)}(\underline{\mu_{5,A}}^{(D=3)}-\mu_{5,R}^{(D=5)})\\
&E_{y_6}(y_6|D=4,g=A,CR)\\
&\quad=\mu_{6,R}^{(D=5)}+b_{1,R}^{(D=5)}(\mu_{1,A}^{(D=4)}-\mu_{1,R}^{(D=5)})+b_{2,R}^{(D=5)}(\mu_{2,A}^{(D=4)}-\mu_{2,R}^{(D=5)})\\
&\quad+b_{3,R}^{(D=5)}(\mu_{3,A}^{(D=4)}-\mu_{3,R}^{(D=5)})+b_{4,R}^{(D=5)}(\mu_{4,A}^{(D=4)}-\mu_{4,R}^{(D=5)})+b_{5,R}^{(D=5)}(\mu_{5,A}^{(D=4)}-\mu_{5,R}^{(D=5)})
\end{aligned}
$$

Note that the underlined means are non-identifiable and therefore they were equated to:

$$
\begin{aligned}
\mu_{3,A}^{(D=1)}&=\mu_{3,R}^{D\geq 2}+b_{1,R}^{(D\geq 2)}(\mu_{1,A}^{(D=1)}-\mu_{1,R}^{(D\geq 2)})+b_{2,R}^{(D\geq 2)}(\mu_{2,A}^{(D=1)}-\mu_{2,R}^{(D\geq 2)})\\
\mu_{4,A}^{(D=1)}&=\mu_{4,R}^{D\geq 3}+b_{1,R}^{(D\geq 3)}(\mu_{1,A}^{(D=1)}-\mu_{1,R}^{(D\geq 3)})+b_{2,R}^{(D\geq 3)}(\mu_{2,A}^{(D=1)}-\mu_{2,R}^{(D\geq 3)})+b_{3,R}^{(D\geq 3)}(\mu_{3,A}^{(D=1)}-\mu_{3,R}^{(D\geq 3)})\\
\mu_{5,A}^{(D=1)}&=\mu_{5,R}^{D\geq 4}+b_{1,R}^{(D\geq 4)}(\mu_{1,A}^{(D=1)}-\mu_{1,R}^{(D\geq 4)})+b_{2,R}^{(D\geq 4)}(\mu_{2,A}^{(D=1)}-\mu_{2,R}^{(D\geq 4)})+b_{3,R}^{(D\geq 4)}(\mu_{3,A}^{(D=1)}-\mu_{3,R}^{(D\geq 4)})\\
&\quad+b_{4,R}^{(D\geq 4)}(\mu_{4,A}^{(D=1)}-\mu_{4,R}^{(D\geq 4)})\\
\mu_{4,A}^{(D=2)}&=\mu_{4,R}^{D\geq 3}+b_{1,R}^{(D\geq 3)}(\mu_{1,A}^{(D=2)}-\mu_{1,R}^{(D\geq 3)})+b_{2,R}^{(D\geq 3)}(\mu_{2,A}^{(D=2)}-\mu_{2,R}^{(D\geq 3)})+b_{3,R}^{(D\geq 3)}(\mu_{3,A}^{(D=2)}-\mu_{3,R}^{(D\geq 3)})\\
\mu_{5,A}^{(D=2)}&=\mu_{5,R}^{D\geq 4}+b_{1,R}^{(D\geq 4)}(\mu_{1,A}^{(D=2)}-\mu_{1,R}^{(D\geq 4)})+b_{2,R}^{(D\geq 4)}(\mu_{2,A}^{(D=2)}-\mu_{2,R}^{(D\geq 4)})+b_{3,R}^{(D\geq 4)}(\mu_{3,A}^{(D=2)}-\mu_{3,R}^{(D\geq 4)})\\
&\quad+b_{4,R}^{(D\geq 4)}(\mu_{4,A}^{(D=2)}-\mu_{4,R}^{(D\geq 4)})\\
\mu_{5,A}^{(D=3)}&=\mu_{5,R}^{D\geq 4}+b_{1,R}^{(D\geq 4)}(\mu_{1,A}^{(D=3)}-\mu_{1,R}^{(D\geq 4)})+b_{2,R}^{(D\geq 4)}(\mu_{2,A}^{(D=3)}-\mu_{2,R}^{(D\geq 4)})+b_{3,R}^{(D\geq 4)}(\mu_{3,A}^{(D=3)}-\mu_{3,R}^{(D\geq 4)})\\
&\quad+b_{4,R}^{(D\geq 4)}(\mu_{4,A}^{(D=3)}-\mu_{4,R}^{(D\geq 4)})
\end{aligned}
$$

The conditional coefficients for the regressions, were estimated using the sweep operator, based on Dempster's (1969) algorithm. In order to estimate the conditional coefficients, a matrix was formed whose elements comprised the true marginal means, variances and covariances of the reference group. The augmented matrix took the form:

$$\mathbf{B} = \begin{bmatrix} -1 & \mu_1 & \mu_2 & \mu_3 & \mu_4 & \mu_5 & \mu_6 \\ \mu_1 & \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} & \sigma_{16} \\ \mu_2 & \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} & \sigma_{25} & \sigma_{26} \\ \mu_3 & \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} & \sigma_{35} & \sigma_{36} \\ \mu_4 & \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} & \sigma_{45} & \sigma_{46} \\ \mu_5 & \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_{55} & \sigma_{56} \\ \mu_6 & \sigma_{61} & \sigma_{62} & \sigma_{63} & \sigma_{64} & \sigma_{65} & \sigma_{66} \end{bmatrix}$$

Giving:

$b_{31\cdot 12} = SWEEP[2,3]B_{23} = -0.0135$ $\quad$ $b_{51\cdot 1234} = SWEEP[2,3,4,5]B_{25} = 0.1015$

$b_{32\cdot 12} = SWEEP[2,3]B_{33} = 0.5074$ $\quad$ $b_{52\cdot 1234} = SWEEP[2,3,4,5]B_{35} = 0.1062$

$b_{41\cdot 123} = SWEEP[2,3,4]B_{24} = 0.0535$ $\quad$ $b_{53\cdot 1234} = SWEEP[2,3,4,5]B_{45} = 0.2890$

$b_{42\cdot 123} = SWEEP[2,3,4]B_{34} = 0.3184$ $\quad$ $b_{54\cdot 1234} = SWEEP[2,3,4,5]B_{55} = 0.4941$

$b_{43\cdot 123} = SWEEP[2,3,4]B_{44} = 0.6182$

As a result,

$\mu_{3,A}^{(D=1)} = -0.3075$ $\quad$ $\mu_{4,A}^{(D=2)} = 0.524$ $\quad$ $\mu_{5,A}^{(D=3)} = 0.6271$

$\mu_{4,A}^{(D=1)} = 0.6763$ $\quad$ $\mu_{5,A}^{(D=2)} = 0.8898$

$\mu_{5,A}^{(D=1)} = 1.0362$

and therefore, the conditional means at the final time point were:

$$\mu_{6,A}^{(D=1)} = 2.2101, \qquad \mu_{6,A}^{(D=2)} = 2.0591, \qquad \mu_{6,A}^{(D=3)} = 1.7364, \qquad \mu_{6,A}^{(D=3)} = 1.5362$$

Therefore, for the low proportion of missing data, we have:

$$\begin{aligned}
\mu_{6,R} &= (2.5282 \times 0.0287) + (2.5282 \times 0.0460) + (2.5282 \times 0.0578) \\
&\quad + (2.5282 \times 0.0575) + (2.5282 \times 0.81) \\
&= 2.5282 \\
\mu_{6,A} &= (2.2101 \times 0.0287) + (2.0591 \times 0.0460) + (1.7364 \times 0.0578) \\
&\quad + (1.5362 \times 0.0575) + (0.5282 \times 0.81) \\
&= 0.7747
\end{aligned}$$

Hence, the true difference is -1.7535. For the high proportion of missing data we have:

$$\begin{aligned}
\mu_{6,R} &= (2.5282 \times 0.0977) + (2.5282 \times 0.1034) + (2.5282 \times 0.1495) \\
&\quad + (2.5282 \times 0.1494) + (2.5282 \times 0.5) \\
&= 2.5282 \\
\mu_{6,A} &= (2.2101 \times 0.0977) + (2.0591 \times 0.1034) + (1.7364 \times 0.1495) \\
&\quad + (1.5362 \times 0.1494) + (0.5282 \times 0.5) \\
&= 1.1820
\end{aligned}$$

Thus, the true difference is -1.3462.

Copy increments in reference

This assumption implies that the conditional means of the Active treatment would be formed as:

$$
\begin{aligned}
&E_{y_6}(y_6|D=1, g=A, CiR) \\
&\quad = \mu_{2,A}^{(D=1)} + (\mu_{6,R}^{(D=5)} - \mu_{2,R}^{(D=1)}) + b_{1,A\cdot R}^{(D=5)}(\mu_{1,A}^{(D=1)} - \mu_{1,A}^{(D=5)}) + b_{2,A\cdot R}^{(D=5)}(\mu_{2,A}^{(D=1)} - \mu_{2,A}^{(D=5)}) \\
&\quad + b_{3,A\cdot R}^{(D=5)}(\mu_{3,A}^{(D=1)} - \mu_{3,A}^{(D=5)}) + b_{4,A\cdot R}^{(D=5)}(\mu_{4,A}^{(D=1)} - \mu_{4,A}^{(D=5)}) + b_{5,A\cdot R}^{(D=5)}(\mu_{5,A}^{(D=1)} - \mu_{5,A}^{(D=5)}) \\
&\quad = \mu_{2,A}^{(D=1)} + (\mu_{6,R}^{(D=5)} - \mu_{2,R}^{(D=1)}) \\
&\quad \vdots \\
&E_{y_6}(y_6|D=4, g=A, CiR) \\
&\quad = \mu_{5,A}^{(D=4)} + (\mu_{6,R}^{(D=5)} - \mu_{5,R}^{(D=4)}) + b_{1,A\cdot R}^{(D=5)}(\mu_{1,A}^{(D=4)} - \mu_{1,A}^{(D=5)}) + b_{2,A\cdot R}^{(D=5)}(\mu_{2,A}^{(D=4)} - \mu_{2,A}^{(D=5)}) \\
&\quad + b_{3,A\cdot R}^{(D=5)}(\mu_{3,A}^{(D=4)} - \mu_{3,A}^{(D=5)}) + b_{4,A\cdot R}^{(D=5)}(\mu_{4,A}^{(D=4)} - \mu_{4,A}^{(D=5)}) + b_{5,A\cdot R}^{(D=5)}(\mu_{5,A}^{(D=4)} - \mu_{5,A}^{(D=5)}) \\
&\quad = \mu_{5,A}^{(D=4)} + (\mu_{6,R}^{(D=5)} - \mu_{5,R}^{(D=4)})
\end{aligned}
$$

Therefore, for the low proportion of missing data, we have:

$$
\begin{aligned}
\mu_{6,R} &= (2.5282 \times 0.0287) + (2.5282 \times 0.0460) + (2.5282 \times 0.0578) \\
&\quad + (2.5282 \times 0.0575) + (2.5282 \times 0.81) \\
&= 2.5282 \\
\mu_{6,A} &= (2.0282 \times 0.0287) + (2.0282 \times 0.0460) + (1.5282 \times 0.0578) \\
&\quad + (1.5282 \times 0.0575) + (0.5282 \times 0.81) \\
&= 0.7555
\end{aligned}
$$

Hence, the true difference is -1.7726. For the high proportion of missing data we have:

$$\begin{aligned}
\mu_{6,R} &= (2.5282 \times 0.0977) + (2.5282 \times 0.1034) + (2.5282 \times 0.1495) \\
&\quad + (2.5282 \times 0.1494) + (2.5282 \times 0.5) \\
&= 2.5282 \\
\mu_{6,A} &= (2.0282 \times 0.0977) + (2.0282 \times 0.1034) + (1.5282 \times 0.1495) \\
&\quad + (1.5282 \times 0.1494) + (0.5282 \times 0.5) \\
&= 1.1288
\end{aligned}$$

Thus, the true difference is -1.3994.

B Appendix

Proof of covariance for treatment difference

Consider a trial with 100 patients in each of the two arms, baseline and a single follow-up, common covariance matrix and normally distributed outcome. Suppose 50 patients drop-out of the active arm, and none from the placebo arm. Further, suppose we use either Jump to Placebo or Copy Placebo. These two assumptions agree in the case of a single follow-up time, as baseline is assumed to be the same on average for the two arms. The respective means for placebo and the active treatment can be expressed as:

$$\hat{\mu}_R = \frac{1}{100}\sum_{i=1}^{100} y_{i,1,R},$$
$$\hat{\mu}_A = \frac{1}{100}\left(\sum_{i=1}^{50} y^o_{i,1,A} + \sum_{i=51}^{100} E\left(y^m_{i,1,A}|y^o_{i,0,A}, J2R/CR\right)\right)$$

Hence, the covariance of the two means equals:

$$
\begin{aligned}
&\text{cov}(\hat{\mu}_R, \hat{\mu}_A) \\
&= \text{cov}\left(\frac{1}{100}\sum_{i=1}^{100} y_{i,1,R}, \frac{1}{100}\left(\sum_{i=1}^{50} y^o_{i,A} + \sum_{i=51}^{100} E\left(y^m_{i,1,A} | y^o_{i,0,A}, CR/J2R\right)\right)\right) \\
&= \text{cov}\left(\frac{1}{100}\sum_{i=1}^{100} y_{i,1,R}, \frac{1}{100}\left(\sum_{i=1}^{50} y^o_{i,A} + \sum_{i=51}^{100} \left(\hat{\beta}_{0,R} + \hat{\beta}_{1,R} y^o_{i,0,A}\right)\right)\right) \\
&= \text{cov}\left(\frac{1}{100}\sum_{i=1}^{100} y_{i,1,R}, \frac{1}{100}\left(\sum_{i=1}^{50} y^o_{i,A} + \sum_{i=51}^{100} \left(\bar{y}_{1,R} + \hat{\beta}_{1,R}\left(y_{i,0,A} - \bar{y}_{0,R}\right)\right)\right)\right) \\
&= \text{cov}\left(\frac{1}{100}\sum_{i=1}^{100} y_{i,1,R}, \frac{1}{100}\left(\sum_{i=1}^{50} y^o_{i,A} + \sum_{i=51}^{100} \left(\bar{y}_{1,R} + \frac{\sum_{i=51}^{100}(y_{i,0,R} - \bar{y}_{0,R})(y_{i,1,R} - \bar{y}_{1,R})}{\sum_{i=51}^{100}(y_{i,0,R} - \bar{y}_{0,R})^2}\left(y_{i,0,A} - \bar{y}_{0,R}\right)\right)\right)\right) \\
&= \text{cov}\left(\frac{1}{100}\sum_{i=1}^{100} y_{i,1,R}, \frac{50}{100}\bar{y}_{1,R}\right) + \\
&\quad \text{cov}\left(\frac{1}{100}\sum_{i=1}^{100} y_{i,1,R}, \frac{1}{100}\sum_{i=51}^{100} \frac{\sum_{i=51}^{100}(y_{i,0,R} - \bar{y}_{0,R})(y_{i,1,R} - \bar{y}_{1,R})}{\sum_{i=51}^{100}(y_{i,0,R} - \bar{y}_{0,R})^2}\left(y_{i,0,A} - \bar{y}_{0,R}\right)\right) \\
&= \text{cov}\left(\frac{1}{100}\sum_{i=1}^{100} y_{i,1,R}, \frac{50}{100}\frac{1}{50}\sum_{i=51}^{100} y_{i,1,R}\right) + \\
&\quad \frac{1}{200}\frac{1}{\sum_{i=51}^{100}(y_{i,0,R} - \bar{y}_{0,R})^2}\text{cov}\left(\sum_{i=1}^{100} y_{i,1,R}, \sum_{i=51}^{100}(y_{i,0,R} - \bar{y}_{0,R}) y_{i,1,R}\left(y_{i,0,A} - \bar{y}_{0,R}\right)\right) \\
&= \frac{1}{100}\frac{1}{100}\text{cov}\left(\sum_{i=1}^{100} y_{i,1,R}, \sum_{i=51}^{100} y_{i,1,R}\right) + \frac{1}{200}\frac{1}{\sum_{i=51}^{100}(y_{i,0,R} - \bar{y}_{0,R})^2}\text{cov}(y_{1,1,R} + ... + y_{100,1,R}, \\
&\quad (y_{51,0,R} - \bar{y}_{0,R}) y_{51,1,R}(y_{51,0,A} - \bar{y}_{0,R}) + ... + (y_{100,0,R} - \bar{y}_{0,R}) y_{100,1,R}(y_{100,0,A} - \bar{y}_{0,R})) \\
&= \frac{1}{100}\frac{1}{100}50\text{var}(y_{1,R}) + \frac{1}{200}\frac{1}{\sum_{i=51}^{100}(y_{i,0,R} - \bar{y}_{0,R})^2}[(y_{51,0,R} - \bar{y}_{0,R})(y_{51,0,A} - \bar{y}_{0,R})\text{var}\left(y_{51,1,R}\right) + ... \\
&\quad + (y_{100,0,R} - \bar{y}_{0,R})(y_{100,0,A} - \bar{y}_{0,R})\text{var}\left(y_{100,1,R}\right)] > 0
\end{aligned}
$$

C Appendix

The following two tables present the treatment estimates from fitting the MMRM model after the incorporation of each predictor in turn, as well as the estimate of the respective predictor variable itself. The reference groups of the factor variables are: *Argentina* & *Brazil* & *Chile* & *Mexico* & *India* for *Country*, *Positives* for *APOE Status* and finally, *African American* for *Race*. The *Negatives* category in the *APOE status* includes the following allele combinations: $\varepsilon 3.\varepsilon 3$, $\varepsilon 2.\varepsilon 3$ and $\varepsilon 2.\varepsilon 2$. The *Positives* category includes: $\varepsilon 3.\varepsilon 4$, $\varepsilon 2.\varepsilon 4$ and $\varepsilon 4.\varepsilon 4$. The reference category for *APOE copies* is 0. The allele combinations that form the categories in *APOE copies* are: $\varepsilon 4.\varepsilon 4$ coded as 0, $\varepsilon 3.\varepsilon 4$ and $\varepsilon 2.\varepsilon 4$ coded as 1, and $\varepsilon 3.\varepsilon 3$, $\varepsilon 2.\varepsilon 3$ and $\varepsilon 2.\varepsilon 2$ coded as 2.

Estimated Effect 2mg RSG XR vs Placebo		**Coefficient Estimate for additional variables included**		
Estimate (s.e.)	p-value	Predictors	Estimate (s.e.)	p-value
-1.019 (0.459)	0.026	None	-	-
-0.960 (0.457)	0.036	**Country**		0.005
		HUN/POL/CZE/SUI	0.434 (0.605)	
		GRE/ESP/POR	1.001 (0.612)	
		USA/CAN	0.529 (0.520)	
		JPN	1.399 (0.608)	
		FRA	0.643 (0.601)	
		GER	-0.126 (0.543)	
		ITA	1.062 (0.553)	
		AUT	-0.956 (0.617)	
-1.006 (0.457)	0.026	**Age**	-0.030 (0.016)	0.074
-1.015 (0.459)	0.027	**Gender**		0.692
		Male	-0.107 (0.272)	
-1.018 (0.449)	0.023	**MMSE score**	-0.316 (0.048)	<.0001
-1.046 (0.457)	0.022	**Race**		0.051
		American Indian/ Alaskan Native	-3.454 (2.267)	
		Central/S. Asian	-4.323 (2.557)	
		Asian/S.E. Asian	1.196 (2.580)	
		White/Arabic/ North African	-2.152 (2.998)	
		White/Caucasian/ European	-2.551 (1.837)	
-1.039 (0.457)	0.023	**Hispanic/Latino**		0.036
		Not	0.715 (0.341)	
-0.836 (0.447)	0.061	**Centre**		<.001
-1.017 (0.459)	0.027	**Weight**	-0.003 (0.009)	0.733
-1.043 (0.461)	0.024	**Height**	0.004 (0.013)	0.793
-1.043 (0.460)	0.023	**BMI**	-0.018 (0.032)	0.569
-1.058 (0.458)	0.021	**Educ. years**	-0.011 (0.032)	0.715
-0.828 (0.482)	0.086	**Parents hist.**		0.308
		No	-0.314 (0.308)	
-0.768 (0.507)	0.130	**Relatives hist.**		0.556
		No	-0.290 (0.492)	
-1.031 (0.460)	0.025	**Time to diagnosis**	0.160 (0.088)	0.069
-1.018 (0.459)	0.026	**APOE copies**		0.689
		1	-0.399 (0.469)	
		2	-0.368 (0.473)	
-1.019 (0.459)	0.026	**APOE status**		0.889
		Negatives	-0.037 (0.271)	

Estimated Effect 8mg RSG XR vs Placebo		**Coefficient Estimate for additional variables included**		
Estimate (s.e.)	p-value	Predictors	Estimate (s.e.)	p-value
-0.215 (0.497)	0.664	None	-	-
-0.238 (0.493)	0.630	**Country**		<.001
		HUN/POL/CZE/SUI	0.917 (0.591)	
		GRE/ESP/POR	1.513 (0.594)	
		USA/CAN	1.254 (0.511)	
		JPN	1.485 (0.597)	
		FRA	0.994 (0.594)	
		GER	0.481 (0.538)	
		ITA	1.098 (0.539)	
		AUT	-1.325 (0.643)	
-0.217 (0.498)	0.662	**Age**	0.009 (0.017)	0.596
-0.215 (0.497)	0.664	**Gender**		0.934
		Male	0.022 (0.274)	
-0.201 (0.492)	0.684	**MMSE score**	-0.238 (0.049)	<.001
-0.243 (0.496)	0.624	**Race**		0.439
		American Indian/ Alaskan Native	-2.711 (2.574)	
		Central/S. Asian	-4.114 (3.016)	
		Asian/S.E. Asian	0.178 (3.602)	
		White/Arabic/ North African	-4.083 (3.273)	
		White/Caucasian/ European	-2.339 (2.282)	
-0.254 (0.496)	0.608	**Hispanic/Latino**		0.013
		Not	0.848 (0.342)	
-0.230 (0.500)	0.645	**Centre**		<.001
-0.197 (0.497)	0.692	**Weight**	-0.017 (0.009)	0.055
-0.231 (0.498)	0.642	**Height**	-0.001 (0.013)	0.922
-0.216 (0.497)	0.663	**BMI**	-0.066 (0.031)	0.035
-0.191 (0.501)	0.702	**Educ. years**	-0.035 (0.033)	0.285
-0.395 (0.523)	0.451	**Parents hist.**		0.241
		No	-0.361 (0.308)	
-0.279 (0.550)	0.611	**Relatives hist.**		0.036
		No	-1.039 (0.496)	
-0.258 (0.498)	0.603	**Time to diagnosis**	0.104 (0.085)	0.219
-0.225 (0.497)	0.650	**APOE copies**		0.416
		1	0.288 (0.285)	
		2	0.503 (0.438)	
-0.217 (0.497)	0.661	**APOE status**		0.219
		Negatives	-0.332 (0.270)	

References

Allison, P. (2002) *Missing Data (Quantitative Applications in the Social Sciences)*, SAGE Publications, USA

Alzheimer's Society (2011) What is Alzheimer's disease, cited at: *www.alzheimers.org.uk/site/scripts/documents_info.php?documentID* = 100 (accessed on 10/07/2013)

Ballinger, G. (2004) Using Generalized Estimating Equations for Longitudinal Data Analysis, *Organizational Research Methods*, 7: 127

Beunckens, C., Molenberghs, G., Kenward, M. (2005) Direct likelihood analysis versus simple forms of imputation for missing data in randomised clinical trials, *Clin. Trials*, 2:379

Blennow K., de Leon M.J., Zetterberg H. (2006) Alzheimer's Disease *Lancet*, 368(9533):387403.

Brown, H., Prescott, R. (2006) *Applied Mixed Models*, Wiley, England

Carpenter, J., Roger, J., Kenward, M. (2013) Analysis of longitudinal trials with protocol deviation: a Framework for Relevant, Accessible Assumptions, and Inference via Multiple Imputation, *Journal of Biopharmaceutical Statistics*, 23(6): 1352-71

Carpenter, J., Kenward, M. (2013) *Multiple Imputation and its Applications*, Wiley, U.K.

Carpenter, J., Kenward, M., Evans, S., White, I. (2004) Last observation carry-forward and last observation analysis. Letter to the Editor, *Stat. Med.*, 23(10): 3241-44

Carpenter, J., Kenward, M. (2007) Missing data in randomised controlled trials - a practical guide, *http://www.pcpoh.bham.ac.uk/publichealth/methodology/projects/RM03_JH17_MK.shtml*

Carpenter, J., Pocock, S., Lamm, C.J. (2002) Coping with missing data in clinical trials: A model-based approach applied to asthma trials, Statistics in Medicine, 21: 1043-1066

Coley, N., Andrieua, S., Jarosc, M., Weinerd, M., Cedarbaumc, J., Vellas, B. (2011) Suitability of the Clinical Dementia Rating-Sum of Boxes as a single primary endpoint for Alzheimers disease trials, *Alzheimers & Dementia*, 7: 602610

Collins, L., Schafer, J., Kam, C. (2001) A Comparison of Inclusive and Restrictive Strategies in Modern Missing Data Procedures, *Psychological Methods*, 6(4): 330-351

Committee for Proprietary Medicinal Products (CPMP) (2001) Points to consider on missing data, European Agency for the Evaluation of Medicinal Products, London, cited at: http://www.emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003641.pdf (accessed on: 14/05/2012)

Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, Pericak-Vance MA (1993). Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families, *Science*, 261 (5123): 9213.

Daniels, M.J., Hogan, J.W. (2008) *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*, Chapman and Hall/CRC, USA

Dempster, A.P. (1969) *Elements of Continuous Multivariate Analysis*, Addison-Wesley, Reading, MA

Diggle, P.J., Heagerty, P., Liang, K., Zeger, S.L. (2002) *Analysis of Longitudinal Data*, OSSS,

New York

Diggle, P., Kenward, M.G. (1994) Informative Drop-out in Longitudinal Data Analysis, *Appl. Statist.*, 43(1): 49-93

Fay, R.E. (1992) When Are Inferences from Multiple Imputation Valid? in *Proceedings of the 1991 Annual Research Conference, U.S. Bureau of the Census*

Folstein M.F., Folstein S.E., McHugh P.R. (1975) "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician, *J. Psychiatr. Res.* 12: 189-198

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B. (2003) *Bayesian Data Analysis*, Chapman and Hall/CRC, Florida, USA

Goodnight, J.H. (1979) A Tutorial on the SWEEP Operator, *The American Statistician*, 33(3): 149-158

Harrington, C., Sawchak, S., Chiang, C., Davies, J., Donovan, C., Saunders, A.M., Irizarry, M., Jeter, M., Zvartau-Hind, M., van Dyck., C.H., Gold, M. (2011) Rosiglitazone Does Not Improve Cognition or Global Function when Used as Adjunctive Therapy to AChE Inhibitors in Mild-to-Moderate Alzheimer's Disease: Two Phase 3 Studies, *Current Alzheimer Research*, 8: 592-606

Heumann (2000) Intention-to-treat with drop-out, cited at: http://epub.ub.uni-muenchen.de/ (accessed on: 25/11/2011)

Hill, A.B. (1961) *Principles of Medical Statistics*, Oxford University Press, New York

Hogan, J.W., Laird, N.M. (1997) Mixture models for the joint distribution of repeated measures and event times, *Statistics in Medicine*, 16(3): 239-257

Hollis, S., Campbell, F. (1999) What is meant by intention to treat analysis? Survey of published randomised controlled trials, *British Medical Journal*, 319: 670-4

ICH E9 Expert Working Group (1999) Statistical Principles for Clinical Trials: step 5: note for guidance on statistical principles for clinical trials, cited at: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf (accessed on: 14/05/2012)

Kenward, M.G., Molenberghs, G. (2009) Last observation carried forward: a crystal ball?, *Journal of Biopharmaceutical Statistics*, 19: 1-17

Kenward, M., Molenberghs, G., Thijs, H. (2003) Pattern-mixture models with proper time dependence, *Biometrika*, 90(1): 53-71

Kenward, M., Carpenter, J., (2009) Multiple Imputation. *In: Fitzmaurice, G., Davidian, M., Verbeke, G., Molenberghs, G. (eds.) Longitudinal Data Analysis*, Chapman and Hall/CRC, USA

Kenward, M., Carpenter, J., (2007) Multiple Imputation: current perspectives, *Statistical methods in medical research*, 16: 199-218

Kenward, M. (1998) Selection models for repeated measurements with non-random dropout: an illustration of sensitivity, *Statistics in Medicine*, 17(23): 27232732

Kenward, M., Roger, J. (1997) Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood, *Biometrics*, 53: 983-997

Lane, P. (2008) Handling drop-out in longitudinal clinical trials: a comparison of the LOCF and MMRM approaches, *Pharmaceutical Statistics*, 7: 93-106

Liang, K., Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models, *Biometrika*, 73(1): 13-22

Little, R., Rubin, D. (2002) *Statistical analysis with missing data*, Wiley, USA

Little, R., Yau, L. (1996) Intent-to-treat Analysis for Longitudinal Studies with Drop-outs, *Biometrics*, 52(4): 1324-1333

Little, R. (1994) A class of pattern-mixture models for normal incomplete data, *Biometrika*, 81(3): 471-83

Little, R.J.A. (1993) Pattern-mixture models for multivariate incomplete data, *Journal of the American Statistical Association*, 88(421): 125-134

Mallinckrodt, C.H., et al. (2003) Assessing and interpreting treatment effects in longitudinal clinical trials with missing data, *Biol. Psychiatry*, 53: 754-60

Mallinckrodt, C.H., et al. (2004) The effect of correlation structure on treatment contrasts estimated from incomplete clinical trial data with likelihood-based repeated measures compared with last observation carried forward ANOVA, *Clinical Trials*, 1: 477

Meng, X. (1994) Multiple-Imputation Inferences with Uncongenial Sources of Input, *Statistical Science*, 9(4): 538-573

Missing data website, cited at: http://missingdata.lshtm.ac.uk (accessed on: 12/5/2012)

Molenberghs, G., Kenward, M., Mallinckrodt, G., Carroll, R. (2004) Analysing incomplete longitudinal clinical trial data, *Biostatistics*, 5(3): 445-464

Molenberghs, G., Kenward, M. (2007) *Missing data in clinical trials*, Wiley, England

Molenberghs, G., Michiels, B., Kenward, M., Diggle, P. (1998) Monotone missing data and pattern-mixture models, *Statistica Neerlandica*, 52(2): 153-161

Molenberghs, G., Verbeke, G. (2005) *Models for Discrete Longitudinal Data*, Springer Series in Statistics, New York

National Institute for Health and Clinical Excellence (2006) Donepezil, galantamine, rivastigmine and memantine for the treatment of Alzheimer's disease, cited at: guidance.nice.org.uk/TA217/Gui (accessed on: 10/07/2013)

National Research Council (2010) The Prevention and Treatment of Missing Data in Clinical Trials. *Panel on Handling Missing Data in Clinical Trials. Committee on National Statistics, Division of Behavioral and Social Sciences and Education.* The National Academies Press, Washington, DC

Neuhaus, J.M., Kalbfleisch, J.D., Hauch, W.W. (1991) A comparison of cluster-specific and population averaged approaches for analyzing correlated binary data, *International Statistical Review*, 59: 25-36

Nich, C., Carroll, K.M. (2002) Intention-to-treat meets missing data: implications of alternate strategies for analyzing clinical trials data, *Drug and Alcohol Dependence*, 68: 121-130

Piantadosi, S. (1997) *Clinical Trials: a methodologic perspective*, John Wiley and Sons, USA

Rockwood, K., Fay, S., Gorman, M., Carver, D., Graham, J.E. (2007) The clinical meaningfulness of ADAS-Cog changes in Alzheimer's disease patients treated with donepezil in an open-label trial, *BMC Neurology*, 7(26): 1-8

Rosen W.G., Mohs, R.C., Davis, K.L. (1984) A new rating scale for Alzheimer's disease, *Am. J. Psychiatry*, 141(11): 1356-64

Rubin, D.B. (1996) Multiple imputation after 18+ years (with discussion), *Journal of the American Statistical Association*, 91: 473-489

Rubin, D.B. (1987) *Multiple Imputation for Nonresponse in Surveys*, John Wiley and Sons, USA

Rubin, D.B. (1976) Inference and missing data, *Biometrika*, 63(3): 581-92

SAS, The MI procedure, cited at: http://support.sas.com/rnd/app/papers/miv802.pdf (accessed on: 01/06/2012)

Schafer, J.L. (1997) *Analysis of incomplete multivariate data*, Chapman and Hall/CRC, USA

Shao,J., Zhong, B. (2003) Last observation carry-forward and last observation analysis, *Statistics in Medicine*, 22: 2429-441

Siddiqui, O., Hung, H.M.J., O'Neil, R. (2009) MMRM vs LOCF: a comprehensive comparison based on simulation study and 25 NDA datasets, *Journal of Biopharmaceutical Statistics*, 19: 227-246

Strittmatter WJ, Saunders AM, Schmechel D, Pericak-Vance M, Enghild J, Salvesen GS, Roses AD (1993) Apolipoprotein E: High avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease, *Proceedings of the National Academy of Sciences of the United States of America*, 90 (5): 19771981

Thijs, H., Molenberghs, G., Michiels, B., Verbeke, G., Curran, D. (2002) Strategies to fit pattern-mixture models, *Biostatistics*, 3(2): 245-265

Tsiatis, A. (1990) Intent-to-Treat Analysis, *Journal of Acquired Immune Deficiency Syndromes*, 3: S1-S138

Verbeke, G., Molenberghs, G. (2000) *Linear mixed models for longitudinal data*, Springer, New York

White, I., Carpenter, J., Evans, S., Schroter, S. (2007) Eliciting and using expert opinions about dropout bias in randomised controlled trials, *Clin. Trials*, 4(2): 125-139

White, I., Royston, P., Wood, A.M. (2011) Multiple imputation using chained equations: Issues and guidance for practice, *Statistics in Medicine*, 30, 377-399

Wikipedia (2013) Apolipoprotein E, cited at: *www.wikipedia.org/wiki/Apolipoprotein_E* (accessed on: 10/07/2013)

Wu, M.C. and Carroll, R.J. (1988) Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process, *Biometrics*, 44, 175-188.

Zeger, S.L., Liang, K., Albert, P.S. (1988) Models for longitudinal data: a generalized estimating equation approach, *Biometrics*, 44: 1049-1060