Bayesian Sensitivity Analysis for Non-ignorable Missing Data in Longitudinal Studies

by

Tian Li

B.Sc., University of Manchester, 2014

Project Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Statistics and Actuarial Science
Faculty of Science

© Tian Li 2017
SIMON FRASER UNIVERSITY
Spring 2017

All rights reserved.
However, in accordance with the *Copyright Act of Canada*, this work may be reproduced without authorization under the conditions for "Fair Dealing." Therefore, limited reproduction of this work for the purposes of private study, research, education, satire, parody, criticism, review and news reporting is likely to be in accordance with the law, particularly if cited appropriately.

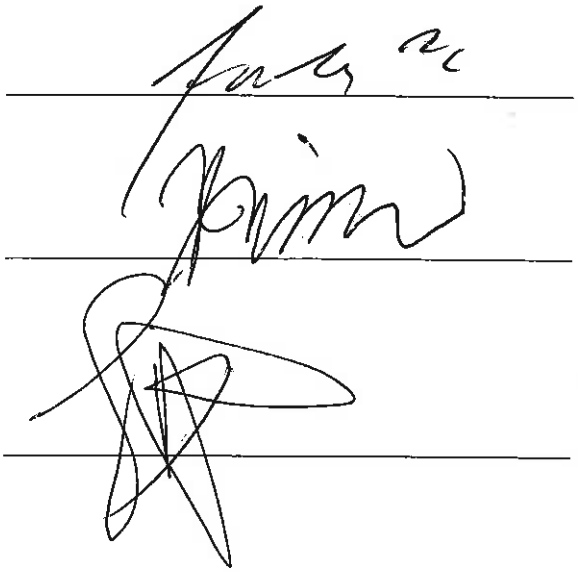
Approval

Name:	**Tian Li**
Degree:	**Master of Science (Statistics)**
Title:	***Bayesian Sensitivity Analysis for Non-ignorable Missing Data in Longitudinal Studies***
Examining Committee:	**Chair: Jinko Graham** Professor

Lawrence McCandless
Senior Supervisor
Associate Professor
Faculty of Health Sciences
&
Associate Member
Statistics and Actuarial Science

Joan Hu
Supervisor
Professor

Julian Somers
Internal Examiner
Associate Professor
Faculty of Health Sciences



Date Defended: April 13 2017

Abstract

The use of Bayesian statistical methods to handle missing data in biomedical studies has become popular in recent years. In this thesis, we propose a novel Bayesian sensitivity analysis (BSA) model that accounts for the influences of missing outcome data on the estimation of treatment effects in randomized control trials with non-ignorable missing data. We implement the method using the probabilistic programming language Stan, and apply it to data from the Vancouver At Home (VAH) Study, which is a randomized control trial that provided housing to homeless people with mental illness. We compare the results of BSA to those from an existing Bayesian longitudinal model that ignores missingness in the outcome. Furthermore, we demonstrate in a simulation study that, when a diffuse conservative prior that describes a range of assumptions about the bias effect is used, BSA credible intervals have greater length and higher coverage rate of the target parameters than existing methods, and that sensitivity increases as the percentage of missingness increases.

Keywords: Bayesian methods; longitudinal analysis; missing data; sensitivity analysis; Simon Fraser University; Vancouver At Home study

Dedication

This thesis is dedicated to **all my friends**, including (in alphabetical order of the last letter of the surname)

Sarshar Hosseinnia

Sagar Mehta

Hassan Kulmie

Vasilis Se

Hossein Sharifi

Qingyuan Feng

Kenneth Chung

Lillian Y L.

Michael McGovern

Angel Granados

Michael James Rogers

Angus Lockhart

Jaypratap Naidu

who have laughed, wept, walked, run and played FIFA with me over the years, and from whom I have learnt far more than I could ever have hoped for.

Gracias, amigos.

Acknowledgements

I would like to express my utmost gratitude to the following individuals and organization:

my supervisor, **Professor Lawrence McCandless**, for the conception of this thesis and for his guidance, support and inspiration throughout my MSc career at SFU that culminated in this final piece of work, which would not have been presentable without his important suggestions and edits,

the examining committee members, **Professor Joan Hu**, **Professor Julian Sommers** and **Professor Jinko Graham**, for their time, patience and insightful feedback,

the creators of R, **Dr. Robert Gentleman** and **Professor Ross Ihaka**, and **many other contributors**, for developing such an amazing statistical language,

the founder of RStudio, **J. J. Allaire**, and the **RStudio team**, for making it so much easier to code in said language,

and last but not least, my parents, **Chunling Wang** and **Yuexue Li**, for unreservedly supporting everything I do.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In longitudinal studies, repeated observations of the same subjects are taken over a period of time. In an ideal world, researchers would be able to monitor every subject at every scheduled point. However, just as most things in life, studies do not always go according to plan and missing observations to various extents do occur against the best wishes of everyone involved.

Numerous methods have been devised to deal with the issue of missing data [8][16][17][28]. In this thesis, we propose a novel Bayesian sensitivity analysis (BSA) specifically for non-ignorable missing data, also known as missing not at random (MNAR) data, and apply this method to data obtained from a study.

In the rest of this chapter, we conduct a review of different types of missing data and their occurrence in longitudinal studies, as well as existing methods in the literature for sensitivity analysis of MNAR data. Chapter 2 discusses an example of a real life longitudinal study with such a problematic type of missing data. It concerns a randomized controlled trial that provided housing to homeless people with mental illness from Vancouver (Patterson et al. [25]). In Chapter 3 we review an existing standard Bayesian random effects model that ignores missing data, which can be compared to the novel BSA method we introduce in Chapter 4. Chapter 5 details a simulation study on the performance of the new method in comparison with the existing method. Discussions are presented in the last chapter.

1.1 Review of Missing Data

Missing data are a common occurrence in research in vast majority of scientific disciplines [8]. Governments, firms and organizations may withhold or fail to report key statistics of

national or commercial interest, which is mostly applicable to studies in social sciences [23]. When data are collected using surveys and interviews, non-response may be a cause of concern, where participants fail to respond to one or more items, or an entire survey altogether (although such cases can be safely ignored) for various reasons. Questions on certain subjects of a private nature, such as level of income, may elicit a higher rate of non-response in a particular group of participants [33].

Missingness may also be inadvertently caused by researchers in the form of human errors, as a consequence of mistakes in data collection and data entry [1]. Nevertheless, this type of missing data can be relatively easy to rectify if the original measurements or observations are available.

Various problems can arise due to missing data. A significant lack of data translates to a reduction in sample size and lower statistical power than intended. In addition to complicating otherwise straightforward statistical analyses, incomplete data may also cause bias in the estimation of model parameters, potentially rendering the conclusions invalid [32]. Bias can occur when the observed data are not representative of the entire population under study, for example, when study participants that have complete observations are healthier and more affluent than the others.

1.2 Missing Data in Longitudinal Studies

Missing data are a critical issue in longitudinal studies. In fact, it has been said that "in longitudinal studies in health sciences, missing data are the rule, not the exception" [10]. The most common type of missingness here is attrition, or dropout. As the name implies, such incidents take place when participants drop out of the study before its completion, resulting in missing observations.

The choice of methods for handling missing data in longitudinal studies is primarily dependent on the pattern and mechanism of missingness, which we outline in the following subsections.

1.2.1 Patterns of Missing Data

Two patterns of missing data exist in longitudinal studies. A monotone missing pattern occurs when a study participant is absent for measurement or observation at a particular point and all points afterwards [17]. That is, once they fail to show up, they are never heard from again.

On the other hand, data are non-monotone missing, or intermittent missing, if observations on a participant are made after they miss a previous data collection point [17]. It is possible that they become missing again, either temporarily or permanently. These participants are not technically considered dropouts since they did not leave of the study forever at the first instance of absence.

In reality, a strictly monotone missing pattern is an uncommon occurrence [16], since it is unlikely that all participants would be able to adhere to the study schedule without missing any intermittent points due to personal or other reasons.

1.2.2 Classification of Missing Data

In a landmark paper on missing data [27], Rubin classifies them into three categories, missing completely at random (MCAR), missing at random (MAR) and neither. The last category was later named as missing not at random (MNAR) [28].

Data are MCAR if the probability of failing to observe a value is independent of any observed or unobserved values of the response variable, or any other observed values [28]. A hypothetical example could be that, in a study on the effect of diet on cholesterol level, participants roll a dice to decide whether to attend a measurement session. Under the MCAR assumption, the observed data can be considered as a random sample of the complete data and there is no bias in the parameter estimates.

Data are said to be MAR if the probability of failing to observe a value is independent of any unobserved values of the response variable, but dependent on observed values of the response variable or some other variables [28]. In this regard, it is perhaps more intuitive to interpret this type of data as "missing conditionally at random". In our hypothetical study, MAR would occur if participants with a lower measured cholesterol level in a session are more likely to miss subsequent sessions, or if participants on a specific type of diet are more inclined to be absent.

By definition, if data are MCAR, they are also MAR [4]. Both MCAR and MAR are ignorable within the likelihood and Bayesian frameworks, whereas in frequentist framework, ignorability is only applicable to MCAR [27].

When the probability of failing to observe a value is dependent on the missing value itself, it is known as MNAR, or informative missing [28]. Failure of participants to turn up for a session in the cholesterol study should be attributed to MNAR provided that the missingness is related to their cholesterol level in that very session, MNAR is non-ignorable

because there is no information on the influence of the missing data, requiring the missing mechanism to be modelled. Treatment of MNAR data is the focus of this thesis.

It is very difficult to ascertain the missing data mechanism in any given study, although several methods have been developed to test for MCAR assumption by Little [22], Listing and Schlittgen [20][21] and Diggle [7], among others. Enders [8] discusses two procedures that distinguish between MCAR and MAR, which rely on the assessment of the independence of the missing indicator and the observed covariates using tools such as logistic regression. Notwithstanding, in general there is no way to determine whether MAR or MNAR exists for they rely on information that is missing, unless follow-up data are obtained from non-respondents for verification [29].

1.3 Review of Existing Methods in the Literature for Sensitivity Analysis of MNAR Missing Data

Popular and well-documented methods for handling missing data include multiple imputation, maximum likelihood estimation, complete case analysis and Bayesian methods [17]. Application of these methods to MNAR data requires specification of the missing data mechanism, which calls for the use of sensitivity analysis to assess the sensitivity of model-based inferences to the unverifiable MAR assumption.

Limited resources on sensitivity analysis for non-ignorable missing data in longitudinal studies exist in the statistical literature, probably as a consequence of the highly speculative nature of such analyses. Of those available, the textbook of Daniels and Hogan [6] describe the procedures applied to two longitudinal studies.

The first example is the Growth Hormone study conducted by Kiel et al. [19]. It is a controlled trial of longitudinal data that investigates the effect of growth hormone and exercise on changes in quadriceps strength, with missing outcome data. Daniel and Hogan [6] limit the discussion to two study arms, exercise plus placebo versus exercise plus growth hormone. In a similar approach to ours (to be introduced later), they construct a pattern mixture model with a large number of sensitivity parameters, which are eventually reduced to a subset of the intercept sensitivity parameters that are allowed to vary in order to account for the deviation from the MAR assumption [6]. Subsequently they carry out a sensitivity analysis examining the posterior inferences about the treatment effect by summarizing the posterior mean and posterior probability over a domain that is calibrated with relevant posterior distributions under MAR.

The second example described by Daniels and Hogan [6] concerns the OASIS trial. This was a randomized controlled trial that compared the effect of standard versus enhanced counselling interventions on smoking cessation rates among alcoholics, with missing binary outcomes. Daniel and Hogan fit both parametric selection models and pattern mixture models with elicited informative priors on the odds ratio at each stage of assessment. Departure from MAR was realized by defining a pair of log odds ratio parameters. They concluded that sensitivity analysis was inappropriate in the case of parametric selection models due to identifiability and that the pattern mixture models fit well because of easily separable parameter space [6].

Elsewhere in the literature, Kenward [18] pedagogically illustrates the use of a sensitivity analysis to examine the effect of the distributional assumptions on the estimation of dropouts using the outcome-based selection model proposed by Diggle and Kenward [7]. In a later paper, Verbeke et al. [35] presents a local influence approach based on the work of Diggle and Kenward to sensitivity analysis for MNAR data, adopted on the concept of individual-specific infinitesimal perturbations around the MAR model. It involves the assignment of a perturbation within the linear predictor of the model to the potentially unobservable measurements.

A different strain of the local influence approach is proposed by Ganjali and Rezaei [12], who utilize a generalized Heckman model to assess the influence of a small perturbation of elements of the covariance structure on the likelihood. It functions as a global sensitivity analysis for cross-sectional and longitudinal data with two periods. The authors suggest the use of normal curvature for longitudinal data with more periods.

Troxel et al. [34] propose a measure of local sensitivity based on a Taylor series approximation to the non-ignorable likelihood, evaluated at the parameter estimates under the ignorability assumption. An index of sensitivity to non-ignorability is derived from the approximated likelihood, which allows researchers to evaluate the need for more elaborate sensitivity analysis or MNAR modelling.

Chapter 2

Data Example

To motivate our discussion of missing data in longitudinal studies, we consider the dataset described by Patterson et al. [25]. The data concern 297 homeless people from Vancouver, British Columbia who participated in the Vancouver At Home (VAH) Study between 2009 and 2013. The VAH Study was a randomized control trial in which homeless participants with mental illness were randomly allocated to receive housing with supports (treatment) or no housing (control), and then followed prospectively to collect information about health outcomes and service use with repeated measurements over time [31].

2.1 Background: Homelessness in Canada and the At Home / Chez Soi Study

Over the last three decades, homelessness has emerged to be one of the most prominent social issues in Canada [36]. According to the Canadian Observatory of Homelessness, it is defined as "the situation of an individual or family without stable, permanent, appropriate housing, or the immediate prospect, means and ability of acquiring it" [2]. A report by the same organization estimates that in 2016, at least 235,000 Canadians were subject to homelessness at some point during the previous year and that 35,000 Canadians were homeless on any given night [11]. In particular, as one of the largest cities in Canada, Vancouver too has seen a steadily growing population of homeless individuals, most of whom are concentrated in Downtown Eastside [5]. As of 2011, 2650 people experienced homelessness in Metro Vancouver, compared to 1121 in 2002 [15].

Studies have demonstrated that homeless people are particularly susceptible to mental illnesses [9] and substance addiction [24], which doubtlessly have detrimental effects on

their already difficult and stressful circumstances. Many of these vulnerable individuals are unable to receive adequate healthcare and social support due to limited funding and investments in community-based mental health programs and affordable housing [14].

To combat homelessness, many policy experts recommend a "Housing First" approach. Housing First provides homeless people with immediate access to subsidized housing, together with supports [14]. No pre-conditions, such as bringing substance abuse under control or being stabilized on medications are imposed. The premise of Housing First is that people should be more capable of moving forward with their lives if they are first provided with housing.

To better understand the impact of a Housing First approach to tackling homelessness in a Canadian context, in 2008 the Mental Health Commission of Canada undertook a $110 million national study called the At Home / Chez Soi Study [36]. This project recruited 2500 participants over four years in five Canadian cities, namely Moncton, Montreal, Toronto, Vancouver and Winnipeg.

2.2 The Vancouver at Home (VAH) Dataset

For this MSc thesis, we will analyze data in the VAH Study that covers specifically the city of Vancouver. In other words, we will focus on data from the Vancouver portion of the nationwide At Home / Chez Soi Study. A complete description of the study protocol is given by Somers et al. [31].

We consider a dataset consisting of the $n = 297$ high-needs (HN) homeless individuals participating in the VAH. Participants were 19 years of age or older, and homelessness was defined as having no fixed place to sleep for more than 7 nights with little likelihood of obtaining accommodation. High need individuals, as defined by Somers et al. [31], had severe mental illness combined with criminal justice involvement, substance dependence or other factors.

In the VAH Study, the 297 participants were randomly assigned to one of three study groups:

1) Independent Housing First with Assertive Community Treatment, which consisted of scattered subsized rental housing around the city

2) Congregate Housing First, where all participants were housed together in a single building in downtown Vancouver, or

3) Treatment-as-Usual (TAU), a control group

People in the TAU group received no further housing or support services from the study apart from the existing services for homeless individuals with mental illnesses in Vancouver.

The participants were randomly assigned to the three treatment groups and then followed prospectively for up to two years. See Figure B.1 for a diagram of the treatment assignment. Data were collected by interview, and each participant was interviewed up to 5 times: once at baselined prior to randomization, then then up to four more times at 6-month intervals.

Several study hypotheses were formulated, among which was that Housing First would have a positive influence on the quality of life (QoL) of the homeless individuals with mental illness as compared to TAU. Thus, the dependent variable in our analysis was the QoL score (see below for details).

The previous analysis by Patterson et al. [25] concluded that the QoL of participants in HF improved significantly more than that of participants in TAU at both 6 and 12 months post baseline. Still, an important potential limitation was that a small proportion of study data was missing as some participants failed to attend one or more scheduled interviews. We are unable to ascertain its impact on the results, especially as participants in TAU were established to be more likely to drop out of the study.

2.3 Basic Descriptive Statistics and Pattern of Missing Data

The objectives of this project are to examine the effects of missingness in the response variable, QoL scores, in the VAH study, and additionally to develop an effective method to explore the sensitivity of analysis results to missing data.

Before conducting any analyses, we begin with the simplifying assumption that the first two Housing First treatment groups (scattered-site housing versus congregated housing) are merged into a single treatment group, which we henceforth call "HF" for Housing First, to facilitate comparison with the control group TAU. This is due to the consideration that the first two groups both involved the provision of housing and there were no statistically significant differences in the measurements of QoL in both groups across the entire study period. Thus, we assume that there are only two arms in the randomized trial: HF (treatment) versus TAU (control). The limitations are discussed further in the Discussion (Section 6.1).

There are a total of 11 variables in the dataset,

`id`	A unique and de-identified id for participant
`visit.number`	The visit number
`visit.type`	A description of the visit type (baseline, 6 months etc.)
`visit.date`	Month and year of visit
`csi`	Colorado Symptom Index (see below)
`qol`	Quality of life score (see below)
`male`	Indicator variable, 1 if participant is male
`age.ord`	Age as an ordinal variable with 3 categories
`num.health.ind`	Total number of health conditions
`hf`	Indicator variable, 1 if Housing First
`total.num.visits`	Total number of visits

We begin with presenting the descriptive statistics of the baseline characteristics of the 297 participants in the study, as shown in Table B.1. A total of 198 (66.7%) individuals were allocated to the HF group and 99 (33.3%) to the TAU group. A total of 72.4% of the participants were male (n = 212), reflecting the actual predominance of males in the homeless population in Vancouver. The largest age group represented was 25-40 years of age (n = 179, 61.1%), followed by over 40 (n = 90, 30.7%) and below 25 (n = 24, 8.2%).

All participants were homeless with mental illness at baseline. Consequently, they had a median of 4 chronic health conditions (interquartile range (IQR) 2-7), which were defined as serious health problems such as diabetes that lasted longer than 6 months. The median Colorado Symptom Index (CSI) score was 30 (IQR 21-41). CSI is a continuous measure of mental health symptoms based on 14 questions with Likert scale 1-5 and higher values represent worse mental health. If more than 50% of the questions were completed, the remainder were imputed with the arithmetic average of the completed questions, otherwise the CSI scores were recorded as missing.

The dependent variable in this MSc thesis is QoL. The median QoL at baseline was 87 (IQR 70-102). The QoL metric adopted in the study was the Quality of Life Interview 20 (QOLI-20), which measures 20 subjective items in 6 subscales: family, finances, leisure, living situation, safety and social. Additionally, there is a global item that assesses an individual's overall satisfaction with life. As some participants did not complete all 20 QoL questions, two approaches were used to handle the missing responses. If more than 50% of the questions were completed, the remainder were imputed with the arithmetic average of the completed questions. Otherwise their QoL scores were recorded as missing.

Table B.3 presents the mean QoL scores (standard deviation (SD)) in HF versus TAU at baseline and at the 4 subsequent 6-month visits. p-values to test for differences were calculated using a 2-sample t-test. As expected, there was no significant difference in the QoL at baseline (p = 0.8819) because the participants were randomly assigned to both

groups and HF cannot affect the outcome at baseline. However, significant differences were found for 6 months ($p = 0.0028$), 12 months ($p = 0.0116$) and a slightly weaker one for 24 months ($p = 0.0808$), indicating higher mean QoL scores in the HF group.

Figure B.1 shows the mean of the QoL trajectories in HF and TAU. It is clear that there is an increasing trend in both groups and the HF group sees a faster increase overall. However, an interesting curiosity is that the difference between the curves tends to diminish over time. Thus, although HF has a positive impact on QoL compared to TAU, the magnitude of the treatment affect appears to be greatest earlier in the follow-up period. The increase of the mean QoL score in TAU was also observed in Patterson et al. [25]. A likely explanation for this interesting finding is that the participants were in such poor health at the time of recruitment that their QoL scores improved even in the absence of provision of housing.

To illustrate the longitudinal nature of the data, Figure B.2 and Figure B.3 present individual QoL trajectories for a random sample of 30 participants from both groups. Figures B.2 and B.3 illustrate the within-subject and between-subject variability in QoL scores over time.

An important concern in the VAH Study, which is the basis of this thesis, is missing data. All participants were scheduled to be interviewed a total of 5 times (baseline, 6, 12, 18 and 24 months). Although data were collected for all 297 at baseline, the number of participants that were revisited at 6 months was reduced to 270, which further declined to 264, 247, 231 at 12, 18 and 24 months, as shown in Table B.2. While a number of dropouts occurred, some individuals remained in the study but skipped one or more intermediate revisits.

Moreover, TAU participants were less likely to participate in follow-up interviews. Out of all QoL measurements at 4 revisits, 9.7% were missing in the HF group ($n = 75$). At 21.5%, the percentage of missingness in the TAU group is slightly more than twice that in HF. ($n = 85$). Clearly and unsurprisingly, participants that were not favoured by the god of probability in the random treatment assignment had little incentive to remain in the study because they did not get housing.

The key scientific question is therefore whether the excess loss to follow-up in the TAU group may have biased the analysis findings. For instance, if healthy TAU participants were less likely to be lost to follow-up, this could in theory bias the mean QoL trajectory curve in the TAU group and decrease the treatment effect. In effect, there would be a "selective attrition" where the TAU group comprised predominantly healthy individuals (with high QoL scores), which would describe a MNAR scenario. However, we have no way to confirm this as the available data cannot answer this question.

Chapter 3

A Longitudinal Analysis that Ignores Missing Data in the Outcome Variable Quality of Life

Recall that the data consists of longitudinal measurements of Quality of Life Scores in $n = 297$ participants, who were followed prospectively for 24 months. Each participant was randomly allocated to either treatment (Housing First (HF)) or control (Treatment as Usual (TAU)). They were then interviewed up to 5 times (baseline, 6 months, 12 months, 18 months and 24 months), and detailed data were recorded. The dependent variable in the analysis was repeated measures of Quality of Life (QoL).

Previously, longitudinal data analyses of the VAH data were conducted in a paper by Patterson et al. [25]. The authors used a linear mixed effects regression to model the association between the different types of HF and the normally distributed outcome QoL. In the regression analysis, the authors included time (discrete 6 month intervals) and interaction terms between time and study arm, which capture the treatment effects. Furthermore, in the multivariable model the authors adjusted for baseline covariates including age, gender and other variables such as housing status at baseline and duration of previous homelessness. Patterson et al. [25] found that HF was associated with significantly greater QoL scores as compared to TAU.

The statistical issue that motivates this research project is the *intermittent* missing data of QoL measurements. As described in Chapter 2, participants assigned to TAU were less likely to be interviewed in the follow-up period because of a lack of incentive to participate in the study. Yet such missingness does not constitute loss to follow-up as some participants skipped one or more interviews but were interviewed again later in the study. The concern

here is then how this might have affected the results. For example, if TAU participants with worse health were more likely to be lost, the QoL trajectory in TAU may have been biased as a result of attrition of the sickest patients.

In this section of the thesis, we begin by replicating the linear mixed effects analysis of Patterson et al. [25] using Bayesian methods implemented in the software STAN. The analysis results will serve as a point of comparison with the subsequent analyses where we model the missing QoL score directly using a non-ignorable missing data model.

3.1 Model

Building upon the analysis of Patterson et al. [25], we present a Bayesian linear mixed effect models to account for correlation in repeated measures of QoL scores in the VAH Study. For ease of reference, this model will be named as the *"naïve"* model because it naïvely ignores the role of missing data in the analysis.

3.1.1 Variables and Notation

Let Y_{ij} be the quality of life score for i^{th} participant in the j^{th} record, where $i = 1, 2, \ldots, 297$ and $j = 1, 2, 3, 4, 5$ represents baseline and first to fourth visit respectively.

Let X_i be an indicator variable for the group allocation of the i^{th} participant, such that

$$\begin{aligned} X_i &= 1 \text{ if } i^{th} \text{ participant is in the HF group} \\ &= 0 \text{ if } i^{th} \text{ participant is in the TAU group} \end{aligned}$$

Note that in the VAH study the treatment allocation was fixed over time. Despite that, the TAU participants were not prevented from find housing on their own. Consequently, a limitation (discussed in Chapter 6) is that the TAU individuals might in fact have obtained housing through other means.

To model time (i.e. follow-up visit), for each participant in the j^{th} record, we create a vector $\widetilde{\boldsymbol{V_j}}$ of length 4 to represent the number of visit, such that

$$\begin{aligned}\widetilde{\boldsymbol{V_j}} &= [0,0,0,0] \text{ for baseline } (0^{th} visit),\\ &= [1,0,0,0] \text{ for the first visit,}\\ &= [0,1,0,0] \text{ for the second visit,}\\ &= [0,0,1,0] \text{ for the third visit,}\\ &= [0,0,0,1] \text{ for the fourth visit.}\end{aligned}$$

Therefore the data consist of $(X_i, \widetilde{\boldsymbol{V_j}}, Y_{ij})$, where X_i and $\widetilde{\boldsymbol{V_j}}$ are always observed, but the outcome variable Y_{ij} is sometimes missing for certain combinations of i and j and those missing values are simply ignored in this model.

3.1.2 Model for QoL that Ignores Missing Data (Naïve)

We present the naïve model here, which ignores missing data. The underlying assumption in this model is that the QoL score of each participant at any time is affected only by the group allocation and time of visit. We model Y_{ij} using the following linear mixed effects model

$$Y_{ij}|X_i, \widetilde{\boldsymbol{V_j}} \sim \boldsymbol{N}(\theta_i + \widetilde{\boldsymbol{V_j}}^{\mathsf{T}}\widetilde{\boldsymbol{\beta}}_{\boldsymbol{v}} + (\widetilde{\boldsymbol{V_j}}^{\mathsf{T}}\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vx}})X_i, \sigma^2) \tag{3.1}$$

for all pairs (i,j) such that $Y_{i,j}$ is observed. In the dataset there were 1305 QoL scores available for analysis, although we would expect a total of $297 \times 5 = 1485$ observations.

In the model, the quantity θ_i is a random effect that can be interpreted as the mean QoL score for individual i at baseline if they were assigned to TAU. We assign a model for the random effects as $\boldsymbol{N}(\mu_\theta, \delta_\theta^2)$, where μ_θ captures the mean and δ_θ^2 governs the heterogeneity.

The vector $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{v}} = [\beta_{v1}, \beta_{v2}, \beta_{v3}, \beta_{v4}]$ of time effects models how the mean of the QoL trajectory changes over time in the TAU group. The vector $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vx}} = [\beta_{vx1}, \beta_{vx2}, \beta_{vx3}, \beta_{vx4}]$ consists of treatment-by-period interactions. Finally, σ is the residual standard deviation of the QoL scores that is not explained by the model.

For simplicity, Equation (3.1) does not include any covariates (e.g. age and gender) and in principle they could be easily added to the model.

The interaction $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vx}}$ are the main target of inference in the analysis because they describe the treatment effect. To illustrate, the expected QoL score for participants in the TAU group, marginalizing over the random effects θ_i, can be expressed as

$$E[Y|X=0,\widetilde{\boldsymbol{V}}] = \mu_\theta + \widetilde{\boldsymbol{V}}^{\mathsf{T}}(\widetilde{\boldsymbol{\beta}}_{\boldsymbol{v}}) \tag{3.2}$$

whereas the expected QoL score for participants in the Housing First group is given by

$$E[Y|X=1,\widetilde{\boldsymbol{V}}] = \mu_\theta + \widetilde{\boldsymbol{V}}^{\mathsf{T}}(\widetilde{\boldsymbol{\beta}}_{\boldsymbol{v}} + \widetilde{\boldsymbol{\beta}}_{\boldsymbol{vx}}) \tag{3.3}$$

Consequently, the vector of four treatment effects at times 6, 12, 18 and 24 months is then

$$E[Y|X=1,\widetilde{\boldsymbol{V}}] - E[Y|X=0,\widetilde{\boldsymbol{V}}] = \widetilde{\boldsymbol{V}}^{\mathsf{T}}(\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vx}}) \tag{3.4}$$

Note that the treatment effect at time zero (baseline) is set to exactly zero in equation (3.1) since, by definition, when participants were assigned to treatment at time zero, the causal effect of treatment must be zero.

3.2 Prior Distributions

There are five parameters in the aforementioned model, namely $\mu_\theta, \sigma^2_\theta, \widetilde{\boldsymbol{\beta}}_{\boldsymbol{v}}, \widetilde{\boldsymbol{\beta}}_{\boldsymbol{vx}}, \sigma^2$. Following Gelman et al. [13], the following priors were chosen for the parameters,

$$\begin{aligned}
\mu_\theta &\sim \boldsymbol{N}(0, 100) \\
\sigma^2_\theta &\sim \boldsymbol{N}(0, 100000)^+ \\
\widetilde{\boldsymbol{\beta}}_{\boldsymbol{v}} &\sim \boldsymbol{N}\left[0, \begin{pmatrix} 100 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 \\ 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 100 \end{pmatrix}\right] \\
\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vx}} &\sim \boldsymbol{N}\left[0, \begin{pmatrix} 100 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 \\ 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 100 \end{pmatrix}\right] \\
\sigma^2 &\sim \boldsymbol{N}(0, 100000)^+
\end{aligned}$$

where $N(a, b)^+$ is a normal distribution with mean a and variance b that is truncated to be strictly positive. An $\boldsymbol{N}(0, 100)$ prior is assigned to the random effect, time effect and treatment by period interaction to reflect the realistic variation in QoL scores (in the range [20, 140] in the dataset). However, a diffuse prior $\boldsymbol{N}(0, 100000)$ is used for the variances σ_θ^2 and σ^2 due to a lack of prior information, indicating that a large range of values are plausible.

3.3 Computations Using Stan

Stan is a probabilistic programming language for Bayesian statistical inference [3]. Written in C++, it is used to specify statistical models and implements Markov Chain Monte Carlo (MCMC) methods, gradient-based variational Bayesian methods and gradient-based optimization for penalized maximum likelihood estimation. Stan is realized in R using the `rstan` package.

In the analysis, we defined the list of data,

`int n;`	(number of participants)
`int nk;`	(number of data records)
`int id[nk];`	(vector of participant ids of length 5n)
`real y[nk];`	(vector of observed QoL scores of length nobs)
`real x[nk];`	(vector of group allocations of length 5n)
`matrix[nk, 4] v;`	(matrix of visit numbers)

and the list of the parameters,

`real theta[n];`	(random effects in linear mixed effects model)
`real mu_theta;`	(mean of the random effects)
`vector[4] betav;`	(time effect)
`vector[4] betaxv;`	(time effect)
`real<lower=0> sigma_theta;`	(sd of the random effects)
`real<lower=0> sigma;`	(sd of the QoL)

as well as the priors specified in the previous section.

The `y` values were modelled in the following *for* loop,

```
for (i in 1:(nmis))
y[i] ~ normal(v[i] * betav + (v[i] * betavx) * x[i] + theta[id[i]],
sigma);
```

Accordingly, we analyzed the VAH dataset and ran the Hamiltonian Monte Carlo algorithm for 2000 iterations with a burn-in of 1000. Sample convergence was assessed using the effective sample size and scale reduction factor, which are automatically generated in Stan.

3.4 Results

The posterior means and 95% posterior credible intervals for all model parameters from the Stan computations are presented in Table B.4. We observe that all of the posterior coefficient estimates of time effects, $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{v}} = [\beta_{v1}, \beta_{v2}, \beta_{v3}, \beta_{v4}]$, are positive, which confirms the findings of Patterson et al. and in Chapter 2 (Table B.3 and Figure B.2) that, within the TAU group, we witness a dramatic upsurge in the QoL scores over time. For example, the posterior mean of β_{v4} is 15.0, which indicates that QoL scores rose by an average of 15 points from baseline.

An important observation is that the QoL scores tended to increase to a large extent initially and then level out. This confirms a well known result in the VAH study that all participants tended to improve dramatically after study enrollment, irrespective of whether they were allocated to HF or TAU. However, this increase occurred primarily during the early period of the study.

The important part of Table B.4 concerns the treatment by period interaction coefficients $\beta_{vx1}, \beta_{vx2}, \beta_{vx3}, \beta_{vx4}$, which are the treatment effect at each of the four follow-up times points. For example, β_{vx1} was estimated as 8.4, which means that at 6 months the QoL trajectory was on average 8.4 points higher in the HF group as compared to TAU. Furthermore, this difference was statistically significant in the sense that the 95% credible interval did not include zero.

While the time effect increases steadily from 6 months to 24 months, the treatment by period interaction, with a posterior mean of $\beta_{vx1} = 8.4$ at 6 months, dips at 12 and 18 months ($\beta_{vx2} = 5.9$, $\beta_{vx3} = 0.7$). A small increase is seen at 24 months ($\beta_{vx4} = 4.3$) but this, along with β_{vx3}, is deemed insignificant due to the inclusion of 0 in its 95% high probability density (HPD) credible interval.

We conclude that although HF had a large effect on QoL scores during the first 12 months of follow-up, it nearly vanished during the second year. This finding was not reported in the linear mixed effects model of Patterson et al. [25] as it analyzed only one year of follow-up. Essentially, the improvement in outcomes for the TAU group was so substantial that it erased much of the anticipated benefit of HF. We emphasize however that the treatment

effect estimates in Table B.4 were still strictly positive at all time points, and this points to an overall conclusion of benefit albeit with larger uncertainty.

Chapter 4

Bayesian Sensitivity Analysis (BSA) for Non-ignorable Missing Data

4.1 Model

It was shown in the previous chapter that some of the QoL scores were missing. A natural question of interest is whether the inclusion of these missing data would distort the conclusions drawn from the analyses using only existing data, and if so, the extent of influence. The key issue is whether the data are missing at random or non-ignorable missing. However, this cannot be ascertained from observed data because we cannot tell how the missing data differs from the observed data.

To better understand the impact of non-ignorable missing data, we approach the problem from a sensitivity analysis perspective [6]. We propose a novel methodology called "Bayesian sensitivity analysis" (BSA) to explore sensitivity to nonignorable missing data. The idea is to propose a model for the complete data (observed and unobserved) that is indexed by non-identifiable *sensitivity parameters* that describe how the missingness is non-ignorable.

Although the data do not tell us much about the sensitivity parameters, they can still be manipulated as part of a sensitivity analysis, allowing us to examine whether the analysis results of Chapter 3 are robust to different and potentially extreme assumptions about non-ignorability. Moreover, it is straightforward to consider the sensitivity analysis within a full Bayesian analysis framework where we assign prior probability distributions to the sensitivity parameters. However, as the resulting model is not identifiable (i.e. the data

cannot distinguish between different models even asymptotically), the Bayesian method has unusual statistical properties.

We begin this chapter with proposing a modified pattern mixture model that accounts for the effect of missingness on the observations, following Hogan et al. [6].

4.1.1 Missing Data Model

Let Y_{ij} be the quality of life score for i^{th} participant in the j^{th} record, where i = 1, 2, 3, ..., 297 and j = 1, 2, 3, 4, 5 represents baseline and first to fourth visit respectively. In the ideal scenario where there is no missing data, there would be $297 \times 5 = 1485$ observations.

Let X_i be an indicator variable for the group allocation of the i^{th} participant, such that

$$\begin{aligned} X_i &= 1 \text{ if } i^{th} \text{ participant is in the Housing First group} \\ &= 0 \text{ if } i^{th} \text{ participant is in the Treatment as Usual group} \end{aligned}$$

For each participant in the j^{th} record, we create a vector $\widetilde{\boldsymbol{V_j}}$ of length 4 to represent the number of visit, such that

$$\begin{aligned} \widetilde{\boldsymbol{V_j}} &= [0,0,0,0] \text{ for baseline } (0^{th} visit), \\ &= [1,0,0,0] \text{ for the first visit}, \\ &= [0,1,0,0] \text{ for the second visit}, \\ &= [0,0,1,0] \text{ for the third visit}, \\ &= [0,0,0,1] \text{ for the fourth visit} \end{aligned}$$

Let I_{ij} be an indicator variable for missingness in the outcome Y_{ij}, such that

$$\begin{aligned} I_{ij} &= 1 \text{ if } Y_{ij} \text{ is missing}, \\ &= 0 \text{ if } Y_{ij} \text{ is observed} \end{aligned}$$

Note that the quantities I_{ij}, X_i, and $\widetilde{\boldsymbol{V_j}}$ are always observed for all possible combinations of i and j.

4.1.2 Pattern Mixture Model for Missing Data

It is clear from the study that the QoL score of each participant at a specific time is affected by the group allocation and time of visit. To model the missing data, we additionally allow QoL to depend directly on the missing indicator variable. By definition of conditional probability, we factorized the conditional distribution of Y_{ij}, I_i given $X_i, \widetilde{\boldsymbol{V_j}}$ as

$$P(Y_{ij}, I_{ij}|X_i, \widetilde{\boldsymbol{V_j}}) = P(Y_{ij}|X_i, \widetilde{\boldsymbol{V_j}}, I_{ij})P(I_{ij}|X_i, \widetilde{\boldsymbol{V_j}}) \tag{4.1}$$

where $P(Y_{ij}|X_i, \widetilde{\boldsymbol{V_j}}, I_{ij})$ is the pattern mixture and $P(I_{ij}|X_i, \widetilde{\boldsymbol{V_j}})$ is the model for indicator variable of missingness. The phrase "pattern mixture" is used to indicate that the marginal model for the complete data, $P(Y_{ij}|X_i, \widetilde{\boldsymbol{V_j}})$, is a mixture with two different components. This differs from a selection model in that the later specifies the joint distribution of Y_{ij} and I_{ij} through models for the marginal distribution of Y_{ij}, $P(Y_{ij}|X_i, \widetilde{\boldsymbol{V_j}})$, and the conditional distribution of I_{ij} given Y_{ij}, $P(I_{ij}|X_i, \widetilde{\boldsymbol{V_j}}, Y_{ij})$ [6].

Since Y_{ij} consists of the random effect, time effect, treatment by period interaction and bias due to missingness, it can be approximated by a Normal distribution,

$$Y_{ij}|X_i, \widetilde{\boldsymbol{V_j}}, I_{ij} \sim \boldsymbol{N}(\theta_i + \widetilde{\boldsymbol{V_j}}^{\mathsf{T}}\widetilde{\boldsymbol{\beta}}_{\boldsymbol{v}} + (\widetilde{\boldsymbol{V_j}}^{\mathsf{T}}\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vx}})X_i + (\widetilde{\boldsymbol{V_j}}^{\mathsf{T}}\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vm}})I_{ij}, \sigma^2) \tag{4.2}$$

where θ_i is the random effect with the distribution $\boldsymbol{N}(\mu_\theta, \delta_\theta^2)$, $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{v}} = [\beta_{v1}, \beta_{v2}, \beta_{v3}, \beta_{v4}]$ is a vector of time effects, $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vx}} = [\beta_{vx1}, \beta_{vx2}, \beta_{vx3}, \beta_{vx4}]$ is a vector of treatment by period interactions, σ is the standard deviation, and finally the quantity $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vm}} = [\beta_{vm1}, \beta_{vm2}, \beta_{vm3}, \beta_{vm4}]$ is a vector of "bias parameters" that describe the influence of the missing indicator I_{ij} on the QoL score.

The quantity $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vm}}$ can be interpreted as the difference in QoL score for the same participant at each visit in the observed case and in the unobserved case, holding all other variables constant. For instance, $\beta_{vm1} = 0$ means that at 6 months the mean QoL scores for missing participants is identical to that of participants with an observed QoL score (no bias, hence the data are Missing at Random). Conversely, if $\beta_{vm1} = -10$, QoL scores for missing participants are on average 10 points lower than observed participants.

To further simplify the BSA method, we assumed the bias parameters are the same for all four visits, i.e.

$$\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vm}} = [\beta_{vm}, \beta_{vm}, \beta_{vm}, \beta_{vm}].$$

Consequently, the user needs to specify only a single parameter β_{vm} in order to undertake the sensitivity analysis. In principle, this assumption could be weakened if we had reason to

believe that the impact of attrition on the study changed over time. However, a disadvantage is that it would require the user to specify a total of four non-identifiable bias parameters rather than just one.

Note that Equation (4.2) omits interactions between the missingness indicator I_{ij} and treatment. This means that we assume that, at each time point, the difference in mean QoL for observed versus unobserved participants does not depend on the assigned treatment. Thus there is a "uniform impact of missingness" that is identical in TAU versus HF. In principle the model could be extended, although with further complications of requiring more bias parameters.

To complete the pattern-mixture model specification we require a model for the missing data indicator variable, which we construct as a logistic regression model

$$P(I_{ij}|X_i, \widetilde{\boldsymbol{V_j}}) = P(I_{ij}|X_i) = \delta_{X_i} \tag{4.3}$$

Hence in the missing data model, there are two missing data proportions: δ_1 and δ_0, which are the proportions of data missing when $X = 1$ or 0. Note that this model allows the missingness to depend on the treatment assignment (e.g. allows higher missing data in the TAU group), however it assumes that the missingness does not change over time. This assumption is a little unrealistic because the missing data are likely to be more common during later follow-up (see Table B.2). As I_{ij} and X_i are fully observed for all participants, we can assess the adequacy of the logistic regression model.

For the sensitivity analysis, we used different values of β_{vm} to assess the influence of the MNAR assumption on the estimated treatment by period interaction.

4.1.3 Parametrization of Effect of HF on QoL for the BSA Model

In the dataset, we are able to observe the conditional probability of QoL among observed participants given group allocation and visit number, $P(Y_{ij}|X_i, \widetilde{\boldsymbol{V_j}}, I_{ij} = 0)$ and the probability of missingness in either group, $P(I_{ij}|X_i)$. Therefore we can calculate the *observed* treatment effect, which is $E[Y|X = 1, \widetilde{\boldsymbol{V}}, I = 0] - E[Y|X = 0, \widetilde{\boldsymbol{V}}, I = 0]$.

However, our goal is to find out the *marginal* treatment effect taking into account both observed and unobserved cases, which is given by $E[Y|X = 1, \widetilde{\boldsymbol{V}}] - E[Y|X = 0, \widetilde{\boldsymbol{V}}]$, where

$$E[Y|X, \widetilde{\boldsymbol{V}}] = E[Y|X, \widetilde{\boldsymbol{V}}, I = 1]P(I = 1|X, \widetilde{\boldsymbol{V}}) + E[Y|X, \widetilde{\boldsymbol{V}}, I = 0]P(I = 0|X, \widetilde{\boldsymbol{V}}). \tag{4.4}$$

Let $\delta_1 = P(I_{ij} = 1|X = 1)$ be the probability of missingness in the HF group, and $\delta_0 = P(I_{ij} = 1|X = 0)$ be the probability of missingness in the TAU group. Then the expected QoL score for the TAU group is

$$\begin{aligned} E[Y|X = 0, \widetilde{\boldsymbol{V}}] &= E[Y|X = 0, \widetilde{\boldsymbol{V}}, I = 1]\delta_0 + E[Y|X = 0, \widetilde{\boldsymbol{V}}, I = 0](1 - \delta_0) \\ &= (\mu_0 + \widetilde{\boldsymbol{V}}^{\mathsf{T}}(\widetilde{\boldsymbol{\beta}}_{\boldsymbol{v}} + \widetilde{\boldsymbol{\beta}}_{\boldsymbol{vm}}))\delta_0 + (\mu_0 + \widetilde{\boldsymbol{V}}^{\mathsf{T}}(\widetilde{\boldsymbol{\beta}}_{\boldsymbol{v}}))(1 - \delta_0) \\ &= \mu_0 + \widetilde{\boldsymbol{V}}^{\mathsf{T}}(\widetilde{\boldsymbol{\beta}}_{\boldsymbol{v}} + \delta_0\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vm}})) \end{aligned} \tag{4.5}$$

where $\delta_0\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vm}}$ is the bias shift.

On the other hand, the expected QoL score for the Housing First group is

$$\begin{aligned} E[Y|X = 1, \widetilde{\boldsymbol{V}}] &= E[Y|X = 1, \widetilde{\boldsymbol{V}}, I = 1]\delta_1 + E[Y|X = 1, \widetilde{\boldsymbol{V}}, I = 0](1 - \delta_1) \\ &= (\mu_0 + \widetilde{\boldsymbol{V}}^{\mathsf{T}}(\widetilde{\boldsymbol{\beta}}_{\boldsymbol{v}} + \widetilde{\boldsymbol{\beta}}_{\boldsymbol{vx}} + \widetilde{\boldsymbol{\beta}}_{\boldsymbol{vm}}))\delta_1 + (\mu_0 + \widetilde{\boldsymbol{V}}^{\mathsf{T}}(\widetilde{\boldsymbol{\beta}}_{\boldsymbol{v}} + \widetilde{\boldsymbol{\beta}}_{\boldsymbol{vx}}))(1 - \delta_1) \\ &= \mu_0 + \widetilde{\boldsymbol{V}}^{\mathsf{T}}(\widetilde{\boldsymbol{\beta}}_{\boldsymbol{v}} + \widetilde{\boldsymbol{\beta}}_{\boldsymbol{vx}} + \delta_1\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vm}})) \end{aligned} \tag{4.6}$$

where $\delta_1\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vm}}$ is the bias shift.

Therefore the overall treatment effect can be shown to be

$$E[Y|X = 1, \widetilde{\boldsymbol{V}}] - E[Y|X = 0, \widetilde{\boldsymbol{V}}] = \widetilde{\boldsymbol{V}}^{\mathsf{T}}(\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vx}} + (\delta_1 - \delta_0)\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vm}}). \tag{4.7}$$

This gives a simple analytical formula for exploring sensitivity to non-ignorable missing data. We see that $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vx}}$ is the observed treatment effect and additionally, the quantity $(\delta_1 - \delta_0)\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vm}}$ is the bias. If the probability of missingness is different in both groups (i.e. if $\delta_0 \neq \delta_1$) and the bias parameter $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vm}} \neq 0$, there is a bias due to missingness.

4.2 Prior Distributions

There are six parameters in the aforementioned model, namely $\mu_\theta, \sigma^2_\theta, \widetilde{\boldsymbol{\beta}}_{\boldsymbol{v}}, \widetilde{\boldsymbol{\beta}}_{\boldsymbol{vx}}, \beta_{vm}, \sigma^2$.

Building on Chapter 3 for the Bayesian linear effects model, we assign the following prior distributions for the parameters,

$$
\begin{aligned}
\mu_\theta &\sim \boldsymbol{N}(0, 100) \\
\sigma^2_\theta &\sim \boldsymbol{N}(0, 100000)^+ \\
\widetilde{\boldsymbol{\beta}}_{\boldsymbol{v}} &\sim \boldsymbol{N}\left[0, \begin{pmatrix} 100 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 \\ 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 100 \end{pmatrix}\right] \\
\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vx}} &\sim \boldsymbol{N}\left[0, \begin{pmatrix} 100 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 \\ 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 100 \end{pmatrix}\right] \\
\sigma^2 &\sim \boldsymbol{N}(0, 100000)^+.
\end{aligned}
\tag{4.8}
$$

For β_{vm}, we used two different approaches to assign a prior distribution. First, we assigned specific fixed values to β_{vm} (so that β_{vm} is not a random variable). Seven values ranging from -30 to 30 were selected to cover a broad range of possible values of β_{vm}. Thereafter, we repeated the analysis and assigned a prior $\beta_{vm} \sim \text{Uniform}(-30, 30)$. Thus the first analysis is a classic "sensitivity analysis" in the sense that the bias parameter β_{vm} is fixed at a specific value that ranges over a grid in order to study the sensitivity analysis of the results, whereas the second analysis is a classic Bayesian sensitivity analysis where we model our prior beliefs about β_{vm} as a uniform distribution.

Since the QoL scores are in the range [20, 140], it would be reasonable to assign a $\boldsymbol{N}(0, 100)$ prior to the random effect, time effect and treatment by period interaction. However, there is no information on the variances σ^2_θ and σ^2. As such, a diffuse prior $\boldsymbol{N}(0, 100000)$ is used.

A crucial aspect of the BSA method is justifying the range of values for β_{vm}. Our approach allowed β_{vm} to vary between -30 and 30. As mentioned above *the data should reveal nothing* about the true value of the β_{vm}, even as n $\to \infty$. Thus choosing a prior is highly speculative and this prior will strongly influence the results. As described in section 4.1.2, β_{vm} can be interpreted as the difference in the QoL score of the same participant at each visit in the observed case versus the unobserved case, holding all other variables constant. Based on a review of the literature, we feel that $\pm$ 30 is the maximum plausible range for β_{vm} in the VAH data.

4.3 Computations Using Stan

In the analysis using the `rstan` package, the list of data was first defined,

`int n;`	(number of participants)
`int nobs;`	(number of observed records)
`int nmis;`	(number of missing records)
`int id[nobs+nmis];`	(vector of participant ids of length 5n)
`real yobs[nobs];`	(vector of observed QoL scores of length nobs)
`real x[nobs+nmis];`	(vector of group allocations of length 5n)
`matrix[nobs+nmis, 4] v;`	(matrix of visit numbers)
`real betavm;`	(bias due to missingness)

followed by list of the parameters,

`real theta[n];`	(random effects in linear mixed effects model)
`real mu_theta;`	(mean of the random effects)
`vector[4] betav;`	(time effect)
`vector[4] betaxv;`	(time effect)
`real<lower=0> sigma_theta;`	(sd of the random effects)
`real<lower=0> sigma;`	(sd of the QoL)
`real ymis[nmis]`	(missing data to be imputed)

and the priors specified in the previous section.

The missing data was imputed in the following *for* loop,

```
for (i in 1:(nmis))
ymis[i] ~ normal(v[i+nobs] * (betav + betavm) + (v[i+nobs] * betavx)
* x[i+nobs] + theta[id[i+nobs]], sigma);
```

with the observed data in another loop,

```
for (i in 1:(nobs))
yobs[i] ~ normal(v[i] * betav + (v[i] * betavx) * x[i] + theta[id[i]],
sigma);
```

We ran the Hamiltonian Monte Carlo algorithm for 2000 iterations with a burn-in of 1000. Sample convergence was once again assessed using the effective sample size and scale reduction factor, which are automatically generated in Stan.

4.4 Results

4.4.1 Sensitivity Analysis Where β_{vm} is Fixed over a Specific Grid of Values

Sensitivity analysis results with fixed β_{vm} over a specific grid of values produced estimates of $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vx}}$ parameters as displayed in Table B.5. As β_{vm} decreases from 30 (mean QoL for missing participants 30 points higher than observed participants) to -30 (mean QoL for missing participants 30 points lower than observed participants), we witness a clear monotonic increase in the estimates of all four treatment effect parameters $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vx}}$.

Upon closer inspection we can see that when β_{vm} is positive (QoL for missing higher than observed), there is no or little significant treatment effect at all four time points. Since the percentage of missingness is higher in the TAU group, higher QoL scores for missing participants would result in a smaller actual difference between HF and TAU than observed, thus overriding any perceived treatment effect.

On the contrary, negative β_{vm} values (QoL for missing lower than observed) signify that the actual difference in QoL between HF and TAU is larger than observed, thereby amplifying the treatment effect. Naturally, its estimates are much more significant.

The case of $\beta_{vm} = 0$ is trivial, as we were simply assuming that there is no difference in QoL between missing and observed participants (i.e. MAR), which is essentially equivalent to ignoring the missing data. As such, the estimates should be very similar to those we obtained using the naïve model (Table B.4).

It should also be noted that even at an extremely low value of β_{vm} = -30, the treatment effect at 18 months β_{vx3} is still non-significant, and that at 24 months β_{vx4} is only slightly significant. Whereas β_{vx1} stays significant throughout the range of β_{vm} values. These observations provide good evidence that the treatment effect is highly non-sensitive to changes in the bias parameter β_{vm}.

4.4.2 Bayesian Sensitivity Analysis Where β_{vm} is a Random Variable with Prior Distribution

We now present the BSA results, wherein we assign a prior probability distribution to β_{vm}. The results from Stan computations using a Uniform prior for β_{vm} are summarized in Table B.6. The key observation is that the posterior means and 95% HPD credible intervals of $\mu_\theta, \sigma^2_\theta, \widetilde{\boldsymbol{\beta}}_{\boldsymbol{v}}, \widetilde{\boldsymbol{\beta}}_{\boldsymbol{vx}}$ and σ^2 are very similar to their counterparts in the naïve model (Table B.4). In other words, the BSA model gives very similar inferences to the naïve model.

This tells us that the analysis results in the VAH dataset are robust to even quite extreme assumptions about bias from non-ignorable missing data.

Table B.4 also shows that β_{vm} has a posterior mean of -4.3 and 95% HPD credible interval of (-29.8, 23.1). The considerable width of this credible interval ensures that it encompasses the more extreme values of β_{vm} and reinforces our view that $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{v}}, \widetilde{\boldsymbol{\beta}}_{\boldsymbol{vx}}$ are non-sensitive to drastic changes in the bias due to missingness.

We theorize that the lack of any real difference in the BSA versus naïve analysis results can be explained by the fact that the difference in missingness between the two groups is relatively small (9.7% in HF, 21.5% in TAU, with a difference of 10.8%) and that β_{vm} is small in magnitude. This is to say that since only 10% to 20% of the QoL scores are missing, even extreme assumptions about MNAR data do not unduly influence the analysis results about the effect of HF on QoL.

Chapter 5

Simulations

5.1 Generating Simulated Datasets

The main comparison that is important in this thesis is Table B.4 versus Table B.6. Results in Table B.4 ignores the problem of missing data, whereas Table B.6 does a BSA that incorporates very extreme assumptions about how detrimental the missing data are. However, it is interesting to note that the results in Table B.4 and B.6 are very similar. In Table B.6 we see that the 95% HPD CIs are only somewhat wider. This is reassuring because it means that even if the missing data were very different from the observed data, it would not change the overall conclusions in the analysis of the VAH data.

We suspect that one reason that the analysis results are insensitive to missing data is that the proportion of missing QoL scores was only 9.7% in the HF group and 21.8% in the TAU group. In order to better understand the behaviour of BSA, we conducted a simulation study where we applied the BSA method to simulated data (i.e. synthetic data generated using a computer). This will better illustrate BSA in extreme situations.

In this section, we outline the procedures for generating simulated datasets using appropriate choices of parameters.

In the simulation, the number of participants `n` and number of visits `k` were defined to be 300 and 5 respectively. A vector `id` was then created, which contained the id numbers from 1 to 300 repeated five times element-wise.

We drew the group allocation of each participant from a Binomial distribution with parameters (1, 0.5), which was stored in the vector `x`. This ensures that each participant

has a equal probability of being assigned to the treatment group (HF) or the control group (TAU), as is the case in the actual randomization procedure.

Corresponding to each element in the vector `id`, a visit number, which ranges from 0 to 4 (with 0 as the baseline) was created and stored in the vector `t`. This was used to construct a design matrix of dummy variables, `v`, where the rows represent participants at each visit and columns represent the visit number (1 to 4). An element (5x+a, y) in the matrix takes the value 1 if the $(a+1)^{th}$ entry of participant x was recorded at visit number y and 0 otherwise, with $0 \leqslant a \leqslant 4$.

A mean QoL score of 70 at baseline `beta.0` was used, as we felt it is a realistic estimate as suggested by the VAH data.

The random effects `theta.i` was generated using the Normal distribution N(0, 64), with an estimated standard deviation of 8. For simplicity, we fixed the time effect `beta.v` to be (10, 10, 20, 20) for each visit, with later visits resulting in a higher effect. On the other hand, the treatment by period interaction `beta.vx = (10, 10, 10, 10)` would be constant throughout all four visits.

The probability of missingness in the HF and TAU group was determined to be 0.25 and 0.75 respectively in a hypothetical scenario where participants in TAU are significantly more likely to drop out of the study. Clearly, most if not all studies in real life would not see such an extreme rate of attrition on accounts of protocol restrictions,

Last but not least, we assigned a value of -5 to the bias parameter `beta.vm`, which we believe is a reasonable estimate of the true value in most situations.

Having determined the individual parameters, we sampled 1500 `y` values from the distribution specified in Equation (4.2), which complete one simulated dataset. The same parameters above were used to generate 100 independent simulated datasets.

5.2 Analysis of Simulated Datasets

In order to compare the performance of estimation of the actual treatment effects, we coded both the naïve model and the BSA model, as described in Chapter 3 and 4, using the Stan language realized by the `rstan` package. Thereafter we applied them to the 100 simulated datasets and obtained the 95% HPD credible intervals of the treatment effect β_{vx}.

Subsequently, we calculated the average length of the 100 credible intervals and their coverage rate of the true β_{vx} values (10, 10, 10, 10). The coverage rate is defined as the proportion of times that the credible intervals contain the true β_{vx} values.

5.3 Results

Table B.7 offers clear evidence that the BSA model produces considerably wider credible intervals (average length of 23.1) of treatment effects than the naïve model (average length of 4.5). Consequently, the true values of β_{vx} are covered approximately twice as often by the BSA credible intervals (87%, 89%, 90%, 89%) as by the naïve credible intervals (45%, 46%, 39%, 38%). The narrow length and low coverage rate of the credible intervals produced by the naïve model substantiates the notion that it is ill-suited to handle datasets with significant amount of missingness (in this case, 50% in total), since simple omission of the missing data could result in a drastic reduction in the variance of posterior β_{vx} and a decisive bias in the analysis results.

The consistent performance of the BSA model in this simulated study with a large proportion of missing data suggests that it is a better choice than the existing naïve model when the bias is truly present because its credible intervals are able to cover the true values of β_{vx} much more frequently. This would be more apparent when the extent of missingness is less extreme, as is the case in many studies in real life.

Chapter 6

Discussion

In this thesis, we proposed a novel Bayesian Sensitivity Analysis method to explore sensitivity to nonignorable missing data for the outcome variable. In summary, we utilized a modified pattern mixture model for the complete data including observed and unobserved information that are indexed by non-identifiable sensitivity parameters that accounts for the effect of missingness on the observations. To use the method, the analyst must specify different values of the sensitivity parameter $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vm}}$, which can be interpreted as the average difference in the mean QoL scores for missing versus observed participants. When each component of $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vm}}$ is equal to zero, the the missingness is ignorable (MAR). Conversely, a negative value such as -20 means that the missing individuals tend to have *lower* QoL scores. Alternatively, a prior probability distribution dependent on the analyst's prior beliefs about bias should be assigned to $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vm}}$. Informative priors can be constructed from literature reviews and expert opinions on the subject matter, while weakly informative and non-informative priors can be modelled using distributions such as Normal and Uniform. Analysis results would be presented as a table (e.g. Table B.5 and Table B.6).

The BSA method provides a major advantage over the naïve Bayesian longitudinal analysis that ignores the effect of missing data in that it produces reliable estimates of the sensitivity parameters with credible intervals as an evidence of the robustness of the other model parameters. Furthermore, in cases where there is a large difference in the frequency of missingness between the treatment and the control groups, then the credible intervals of the treatment effect parameters for BSA are markedly wider and provide much higher coverage of the true values than those for the naïve method. This is an important characteristic that enables us to ascertain the true value of treatment effect in the presence of missing data with greater confidence.

However, for the BSA model careful selection of a prior on the sensitivity parameter $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vm}}$ is required to conduct a meaningful analysis, as inappropriate priors would result in confidence intervals of extreme widths of the treatment effect. Furthermore, the non-identifiability of the model dictates that the true values of its underlying parameters cannot be theoretically calculated and the data usually reveal very little about the true values of $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vm}}$. Consequently, the analyst must carefully choose the prior distribution because it may greatly influence the analysis results. This could constitute a key challenge if there is insufficient information to determine suitable choices. For example, in the VAH dataset there no information in the literature that helps us estimate the true value of $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vm}}$.

Curiously, non-informative priors for $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vm}}$ do not exist because the model is non-identifiable. So for example, if we assign a prior of $\text{Unif}(-10^{10}, 10^{10})$ to $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vm}}$, the resulting posterior on the treatment effects would be infinitely wide.

Another disadvantage of the BSA is that, as are many Bayesian methods that involve Monte Carlo simulations, it could be computationally intensive, especially when a large number of missing observations exist and need to be estimated. It is also known that there is considerable computational difficulty in Bayesian MCMC simulation for non-identifiable models [30], as a result of correlated parameters in and irregular shape of the posterior distribution (Figure B.5) [26].

We applied the BSA method to the VAH data to estimate the effect of the HF intervention on QoL of the homeless participants. In particular, we used BSA to explore sensitivity of the analysis results to different assumptions about non-ignorable missing data (i.e. different values of $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vm}}$). We found that there is remarkably little sensitivity to assumptions about missing data. In other words, missingness is not a great concern as it had little effect on the analysis outcomes, which is doubtlessly good news to the researchers. Table B.5 shows that there is a slight difference in overall conclusion only when the value of $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vm}}$ approaches +30 or -30, corresponding to the extreme assumption that the QoL scores in missing participants are 30 points higher (or lower) than otherwise similar observed participants, which is not reasonable. The rather low level of sensitivity is attributable to the fact that the missingness only 10 to 15 percent.

Perhaps the most curious result of the VAH study is the diminution of the effect of Housing First in the later stages of the study, which was so apparent as to rendering the effect at 18 and 24 months insignificant. In other words, we show that providing supporting housing to homeless people has little or no impact on QoL beyond the first year. This controversial finding was not previously shown in Patterson et al. [25], since they had access to the first 12 months of data only. In addition, given the lack of sensitivity, we are able to rule out missing data as a plausible explanation for this anomaly.

6.1 Limitations

Due to limited time and scope of this thesis, several important statistical issues that arose from the study design, data collection and analysis are not addressed.

The response variable in the study, Quality of Life, is a composite sum of scores for 20 question items. While some of the responses to the questions were missing (see Section B.1 for details), they were not considered in the analysis as another aspect of missing data. For our analysis, we followed the same technique of Patterson et al. [25] that used mean substitution for the individual question item provided that no more than half of the items were missing (and otherwise the total QoL score was taken as missing). The effect of ignoring uncertainty in the imputed question items should be investigated further, however it is beyond the scope of this thesis.

In reality, participants were not interviewed at exact 6 months intervals owing to various reasons. The interview timings were grouped as such for the sake of convenience, even though we do have access to the data about the exact month and year of revisits. An analysis using these data would allow us to evaluate the treatment effect at any point within 24 months, which may be useful information in studies of participants that drop out of the HF scheme. Further research should examine the benefits and disadvantages of modelling QoL using smooth non-parametric curves over time. Suitable approaches are described in Chapter 19 of Fitzmaurice et al. [10] entitled "Smoothing longitudinal data using semi-parametric regression models".

Another issue regarding the study design is that TAU participants were not explicitly prevented from find housing on their own. In other words, even though homeless individuals randomly assigned to TAU did not get housing through the VAH study, they were not being prevented from finding housing elsewhere. Such occurrences may have been difficult to avoid due to ethical concerns. This may have resulted in a form of crossover in some cases. However, the VAH Study includes a "housing stability" variable that was measured at each interview to look at where the participant was living. In future analysis this variable would be good to incorporate this variable into the study as a control variable.

One of the fundamental assumptions we made in the analysis, as discussed in Section B.1, is the amalgamation of the two modalities of HF in the high needs group. Treating them as separate treatment groups would have given us more insight on the benefit of each specific type of HF.

A final limitation is that in the Patterson et al. study [25], the authors imputed the missing quality of life values using the multivariate imputation by chained equation (MICE) method and proceeded to analyze the entire dataset with a model that does not consider

the effects of any missing data, while we simply ignored these missing responses in our implementation of the naïve method. Moreover, the analysis of Patterson et al. adjusted for several predictors of QoL, such as age and gender. Such covariate adjustments did not exist in our analysis. To give a fairer comparison of BSA versus the naïve Bayesian longitudinal analysis, it would be useful to additionally apply MICE to the VAH dataset.

6.2 Future Work

The simulation study was somewhat limited because we generated data where $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vm}}$ was only fixed at -5. A better alternative could be "Bayesian" simulations where data are generated by randomly sampling values of $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vm}}$ from some parameter sampling distribution, with the advantage of showing "average" coverage rate and comparison with the naïve method over an entire distribution range of $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{vm}}$ values. Nevertheless, the choice of a suitable parameter generating distribution would be another concern.

The BSA model proposed in this thesis is a rather basic one, as it includes only the most important variables, specifically time of visit, treatment effect, effect due to missingness and random effect. Further modifications could be made to the BSA model to produce more accurate analysis results, such as the incorporation of the exact timing of the QoL measurements rather than just 6 months interval, as well as other meaningful covariates like age and gender.

Bibliography

[1] H. J. Adèr and G. J. Mellenbergh. *Advising on research methods: A consultant's companion*. Johannes van Kessel Publishing., 2008.

[2] Canadian Observatory on Homelessness. *Canadian Definition of Homelessness*. Technical report, Canadian Observatory on Homelessness, 2012.

[3] B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. A. Brubaker, J. Guo, P. Li, and A. Riddell. Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1), Jan 2017.

[4] D. Curran, M. Bacchi, S. Schmitz, G. Molenberghs, and R. Sylvester. Identifying the types of missingness in quality of life data from clinical trials. *Statistics in Medicine*, 17(5-7):739–756, 1998.

[5] L. B. Currie, A. Moniruzzaman, M. L. Patterson, and J. M. Somers. *At Home/Chez Soi Project: Vancouver Site Final Report*. Technical report, Mental Health Commission of Canada, 2014.

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

[7] P. J. Diggle. Testing for random dropouts in repeated measurement data. *Biometrics*, 1255–1258, 1989.

[8] C. K. Enders. *Applied missing data analysis*. Guilford Press, 2010.

[9] S. Fazel, V. Khosla, H. Doll, and J. Geddes. The prevalence of mental disorders among the homeless in western countries: Systematic review and meta-regression analysis. *PLOS Medicine*, 5(12):e225, Dec 2008.

[10] G. M. Fitzmaurice, N. M. Laird, and J. H. Ware. *Applied longitudinal analysis*, volume 998. John Wiley & Sons, 2012.

[11] S. Gaetz, E. Dej, T. Richter, and M. Redman. *The State of Homelessness in Canada 2016*. Technical report, Canadian Observatory on Homelessness, 2016.

[12] M. Ganjali and M. Rezaei. An influence approach for sensitivity analysis of non-random dropout based on the covariance structure. *Iranian Journal of Science and Technology (Sciences)*, 29(2):287–294, 2005.

[13] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian data analysis*, volume 2. Chapman & Hall/CRC Boca Raton, FL, USA, 2014.

[14] P. N. Goering, D. L. Streiner, C. Adair, T. Aubry, J. Barker, J. Distasio, S. W. Hwang, J. Komaroff, E. Latimer, J. Somers, and D. M. Zabkiewicz. The at home/chez soi trial protocol: a pragmatic, multi-site, randomised controlled trial of a housing first intervention for homeless individuals with mental illness in five canadian cities. *BMJ Open*, 1(2):e000323, Jan 2011.

[15] Greater Vancouver Regional Steering Committee on Homelessness. *Results of the 2011 Metro Vancouver Homeless Count.* Technical report, Greater Vancouver Regional Steering Committee on Homelessness, Feb 2012.

[16] N. J. Horton and K. P. Kleinman. Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *The American Statistician*, 61(1):79–90, 2007.

[17] J. G. Ibrahim and G. Molenberghs. Missing data methods in longitudinal studies: a review. *Test*, 18(1):1–43, 2009.

[18] M. G. Kenward. Selection models for repeated measurements with non-random dropout: an illustration of sensitivity. *Statistics in Medicine*, 17(23):2723–2732, 1998.

[19] D. P. Kiel, J. Puhl, C. J. Rosen, K. Berg, J. B. Murphy, and D. B. MacLean. Lack of an association between insulin-like growth factor-i and body composition, muscle strength, physical performance or self-reported mobility among older persons with functional limitations. *Journal of the American Geriatrics Society*, 46(7):822–828, 1998.

[20] J. Listing and R. Schlittgen. Tests if dropouts are missed at random. *Biometrical Journal*, 40(8):929–935, 1998.

[21] J. Listing and R. Schlittgen. A nonparametric test for random dropouts. *Biometrical Journal*, 45(1):113–127, 2003.

[22] R. J. Little. A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83(404):1198–1202, 1988.

[23] R. J. A. Little and D. B. Rubin. The analysis of social science data with missing values. *Sociological Methods & Research*, 18(2-3):292–326, 1989.

[24] C. S. North, K. M. Eyrich, D. E. Pollio, and E. L. Spitznagel. Are rates of psychiatric disorders in the homeless population changing? *American Journal of Public Health*, 94(1):103–108, Jan 2004.

[25] M. Patterson, A. Moniruzzaman, A. Palepu, D. Zabkiewicz, C. J. Frankish, M. Krausz, and J. M. Somers. Housing first improves subjective quality of life among homeless adults with mental illness: 12-month findings from a randomized controlled trial in vancouver, british columbia. *Social Psychiatry and Psychiatric Epidemiology*, 48(8):1245–1259, Aug 2013.

[26] B. Rannala. Identifiability of parameters in mcmc bayesian inference of phylogeny. *Systematic Biology*, 51(5):754–760, 2002.

[27] D. B. Rubin. Inference and missing data. *Biometrika*, 581–592, 1976.

[28] D. B. Rubin and R. J. Little. Statistical analysis with missing data. *Hoboken, NJ: J Wiley & Sons*, 2002.

[29] J. L. Schafer and J. W. Graham. Missing data: our view of the state of the art. *Psychological Methods*, 7(2):147, 2002.

[30] M. M. Shariati, I. Korsgaard, and D. Sorensen. Identifiability of parameters and behaviour of mcmc chains: a case study using the reaction norm model. *Journal of Animal Breeding and Genetics*, 126(2):92–102, 2009.

[31] J. M. Somers, M. L. Patterson, A. Moniruzzaman, L. Currie, S. N. Rezansoff, A. Palepu, and K. Fryer. Vancouver at home: pragmatic randomized trials investigating housing first for homeless and mentally ill adults. *Trials*, 14(1):365, Nov 2013.

[32] J. A. Sterne, I. R. White, J. B. Carlin, M. Spratt, P. Royston, M. G. Kenward, A. M. Wood, and J. R. Carpenter. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *British Medical Journal*, 338:b2393, 2009.

[33] D. Tomaskovic-Devey, J. Leiter, and S. Thompson. Organizational survey nonresponse. *Administrative Science Quarterly*, 439–457, 1994.

[34] A. B. Troxel, G. Ma, and D. F. Heitjan. An index of local sensitivity to nonignorability. *Statistica Sinica*, 1221–1237, 2004.

[35] G. Verbeke, G. Molenberghs, H. Thijs, E. Lesaffre, and M. G. Kenward. Sensitivity analysis for nonrandom dropout: a local influence approach. *Biometrics*, 57(1):7–14, 2001.

[36] D. M. Zabkiewicz, M. Patterson, J. Frankish, and J. M. Somers. The vancouver at home study: Overview and methods of a housing first trial among individuals who are homeless and living with mental illness. *Journal of Clinical Trials*, 2(4):123, Oct 2012.

Appendix A

Code

A.1 Bayesian Linear Mixed Effects Model that Ignores Missing Data

```
## Stan model specification

lme_code <- '

data {
  int n;      // number of participants
  int nk;     // number of records
  int id[nk]; // participant id
  real y[nk]; // QoL
  real x[nk]; // group assignment indicator
  matrix[nk, 4] v;
  // design matrix of dummy variables to indicate visit time
}

parameters {
  real theta[n];    // random effects
  real mu_theta;    // mean of the random effects
  vector[4] betav;  // time effect
  vector[4] betaxv; // treatment effect

  real<lower=0> sigma_theta;  // sd of the random effects
  real<lower=0> sigma;        // sd of the QoL
}

model {
  mu_theta ~ normal(0, 100); // priors
  betav ~ normal(0, 100);
  betaxv ~ normal(0, 100);
  sigma_theta ~ normal(0,100);
  sigma ~ normal(0, 100);
  theta ~ normal(mu_theta, sigma_theta);
```

```
  for (i in 1:nk)
  y[i] ~ normal(v[i] *betav + (v[i] * betaxv) * x[i] + theta[id[i]], sigma);
  // model QoL
}
'

## Translate Stan model specification to C++ code
tcode <- stanc(model_code = lme_code, model_name = `mymodel', verbose = TRUE)

## Construct Stan model
tmodel <- stan_model(stanc_ret = tcode, verbose = FALSE)

## Specify list of data
ah.data <- list(id, y, x, v, n, nk)

## Sample from Stan model
tsamples <- sampling(tmodel, data = ah.data, chains = 1, iter = 2000,
thin = 1, verbose = TRUE)

## Extract MCMC samples
MCMC.samples <- extract(tsamples)
```

A.2 Bayesian Sensitivity Analysis for Non-ignorable Missing Data

```
## Stan model specification

bsa_code <- '

data {
  int n;              // number of participants
  int nobs;           // number of observed records
  int nmis;           // number of missing records
  int id[nobs+nmis]; // participant id
  real yobs[nobs];   // observed QoL

  real x[nobs+nmis]; // group assignment indicator
  int m[nobs+nmis];  // missing data indicator
  matrix[nobs+nmis, 4] v;
  // design matrix of dummy variables to indicate visit time
}

parameters {
  real theta[n];    // random effects
  real mu_theta;    // mean of the random effects
  vector[4] betav;  // time effect
  vector[4] betavx; // treatment effect

  real<lower=0> sigma_theta; // sd of the random effects
  real<lower=0> sigma;       // sd of the QoL
  real ymis[nmis];           // missing QoL
```

```
  real gamm0;  // logistic regression parameter
  real gammx;  // logistic regression parameter

  real<lower=-30, upper=30> betavm; // bias due to missingness
}

model {
  mu_theta ~ normal(0, 100); // priors
  betav ~ normal(0, 100);
  betavx ~ normal(0, 100);
  sigma_theta ~ normal(0, 100000);
  sigma ~ normal(0, 100000);
  theta ~ normal(mu_theta, sigma_theta);

  gamm0 ~ normal(0, 100);
  gammx ~ normal(0, 100);

  for (i in 1:(nmis))
  ymis[i] ~ normal(v[i+nobs] * (betav + betavm) + (v[i+nobs] * betavx) *
  x[i+nobs] + theta[id[i+nobs]], sigma);
  // imputing the missing data using the true value of betavm

  for (i in 1:(nobs))
  yobs[i] ~ normal(v[i] * betav + (v[i] * betavx) * x[i] +
  theta[id[i]], sigma);
  // analysis of the observed data

  for (i in 1:(nmis+nobs))
  m[i] ~ bernoulli_logit(gamm0 + gammx * x[i]);
  // logistic regression for estimating the proportion missing parameters
}
'

## Translate Stan model specification to C++ code
tcode <- stanc(model_code = bsa_code, model_name = `mymodel', verbose = TRUE)

## Construct Stan model
tmodel <- stan_model(stanc_ret = tcode, verbose = FALSE)

## Specify list of data
ah.data <- list(id, yobs, x, v, m, n, nobs, nmis)

## Sample from Stan model
tsamples <- sampling(tmodel, data = ah.data, chains = 1, iter = 2000,
thin = 1, verbose = TRUE)

## Extract MCMC samples
MCMC.samples <- extract(tsamples)
```

Appendix B

Tables and Figures

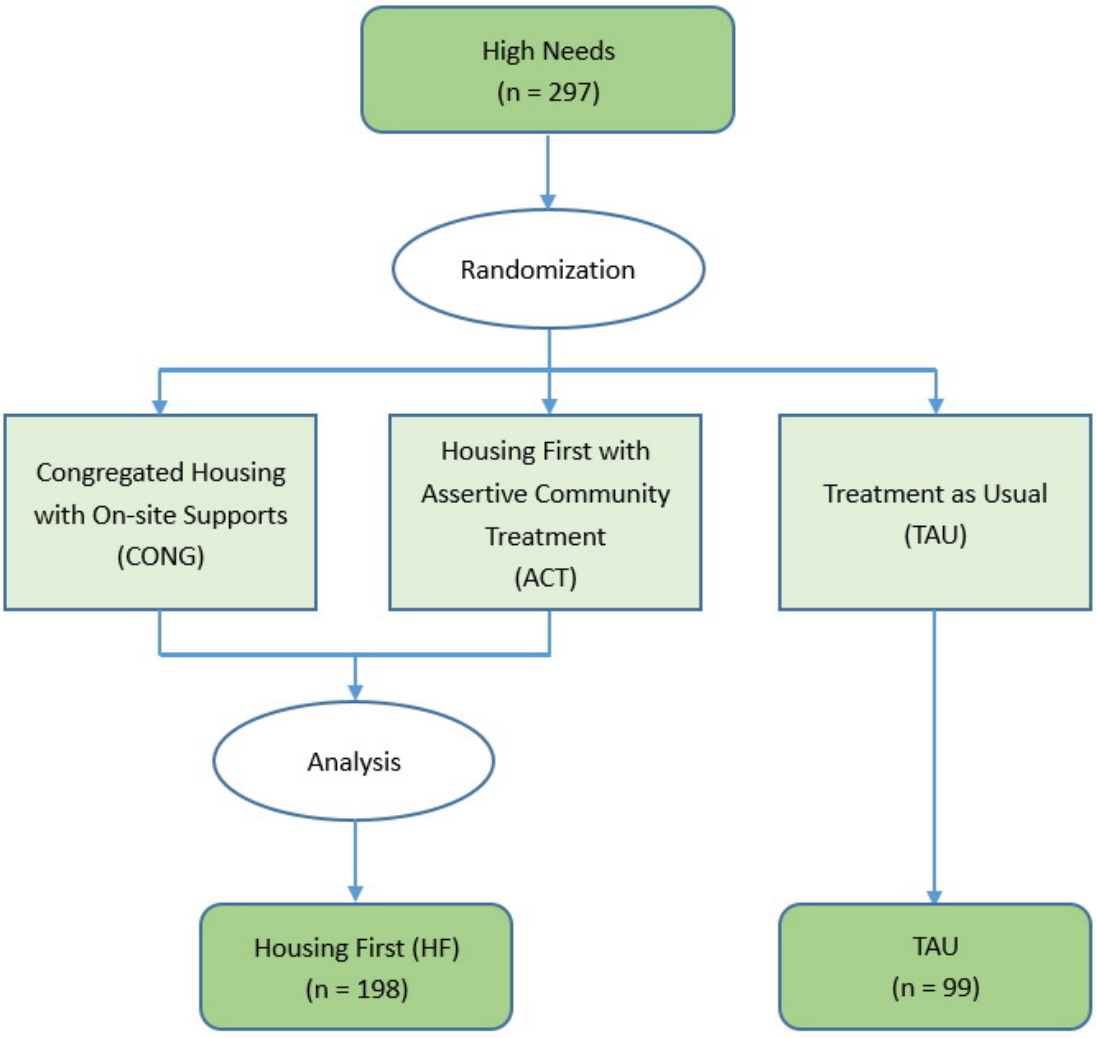


Figure B.1: Group allocation of high needs participants in the At Home study

Variable		Count (Percentages)	Median (IQR)
Housing First (versus TAU)		198 (66.7%)	-
Male		212 (72.4%)	-
Age			
	<25	24 (8.2%)	-
	25-40	179 (61.1%)	-
	>40	90 (30.7%)	-
# Health Conditions		-	4 (2-7)
QoL		-	87 (70-102)
CSI		-	30 (21-41)

Table B.1: Descriptive statistics of the Vancouver at Home dataset of n=297 homeless individuals with mental illness

Visit	Baseline	6 months	12 months	18 months	24 months
Number of participants	297	270	264	247	231

Table B.2: Number of participants at baseline and each revisit

Visit Time	HF Mean ± SD	TAU Mean ± SD	p-value
Baseline	73.32 ± 21.6	74.71 ± 21.5	0.8819
6 Months	89.71 ± 22.3	80.53 ± 25.7	0.0028
12 Months	90.83 ± 23.3	83.34 ± 19.5	0.0116
18 Months	89.05 ± 24.1	87.06 ± 19.2	0.5308
24 Months	93.75 ± 24.3	87.84 ± 20.8	0.0808

Table B.3: QoL mean scores at baseline and 6, 12, 18 and 24 months followup by study arm

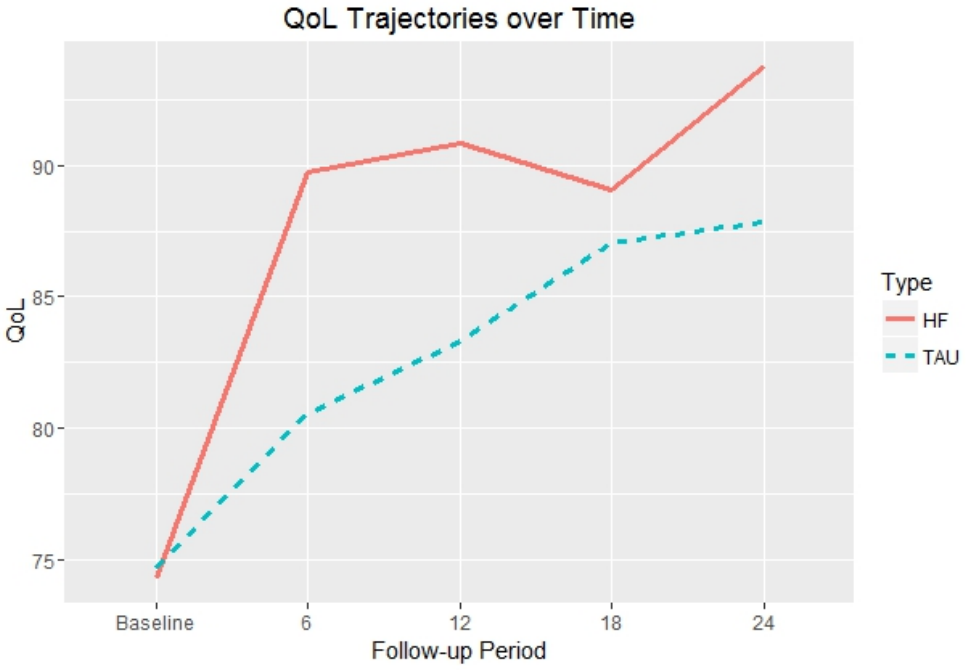


Figure B.2: Average QoL trajectories in HF and TAU groups

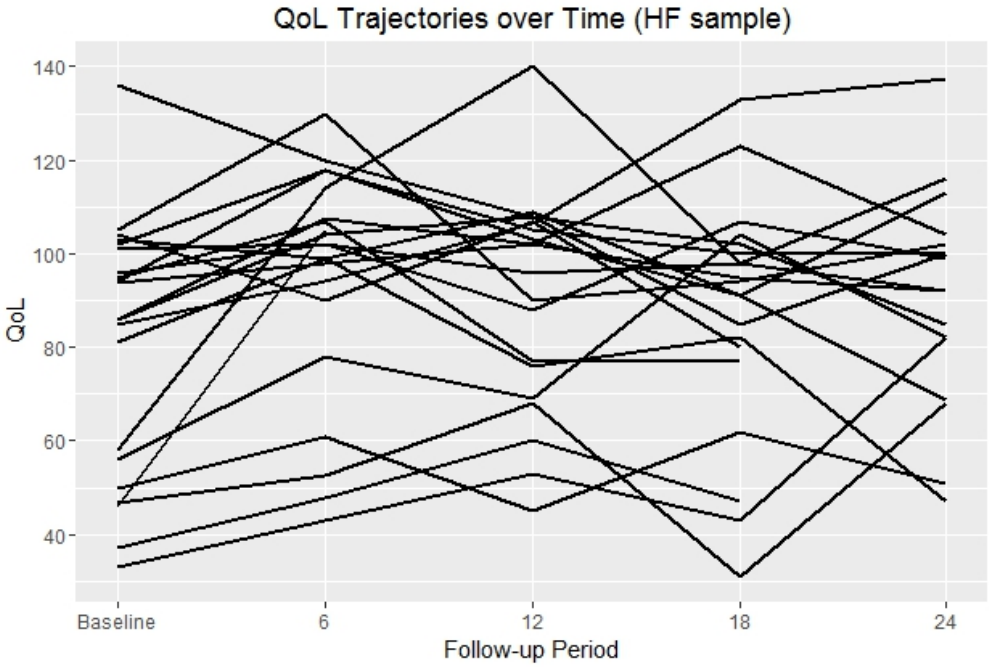


Figure B.3: Individual QoL trajectories for a random sample of size 30 generated from the HF group

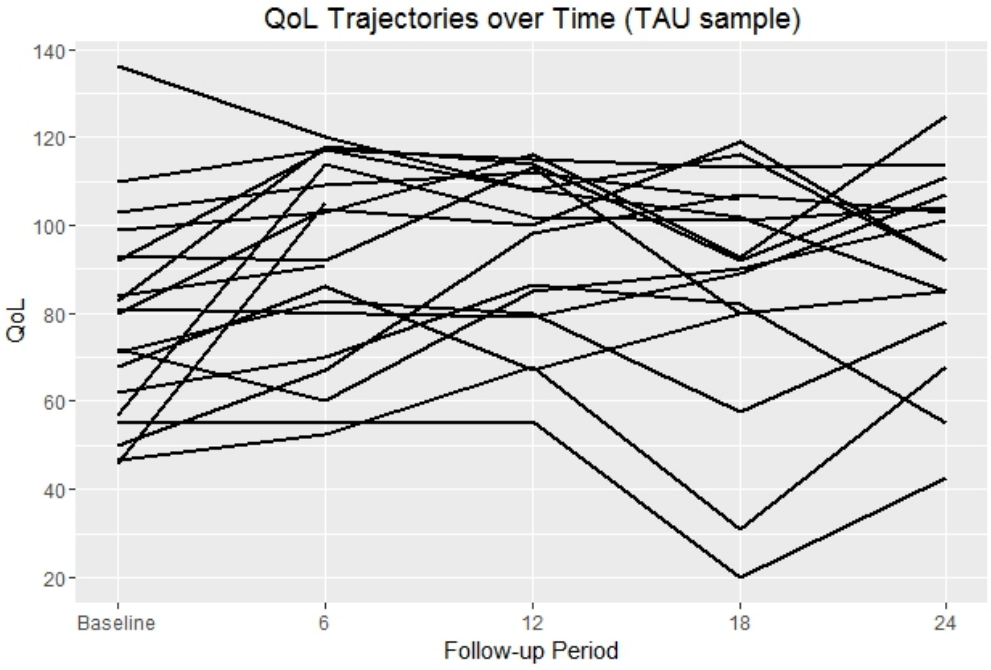


Figure B.4: Individual QoL trajectories for a random sample of size 30 generated from the TAU group

Variable		Posterior Mean	95% HPD CI
Time			
	Baseline		
	6 Months β_{v1}	6.5	(2.6, 11.9)
	12 Months β_{v2}	10.3	(5.9, 15.1)
	18 Months β_{v3}	14.0	(9.4, 18.9)
	24 Months β_{v4}	15.0	(10.4, 19.8)
Time × Housing First			
	HF × 6 Months β_{vx1}	8.4	(2.9, 13.7)
	HF × 12 Months β_{vx2}	5.9	(0.7, 10.9)
	HF × 18 Months β_{vx3}	0.7	(-4.8, 6.0)
	HF × 24 Months β_{vx4}	4.3	(-1.5, 9.5)
μ_θ		74.5	(71.8, 77.0)
σ_θ		15.6	(13.8, 17.3)
σ		16.6	(15.9, 17.3)

Table B.4: Posterior mean and 95% HPD credible interval of variables in a traditional Bayesian longitudinal model that ignores missing data (naïve model)

β_{vm}	β_{vx1}	β_{vx2}	β_{vx3}	β_{vx4}
30	5.5 (0.8, 11.1)	3.0 (-1.8, 8.8)	-2.1 (-7.3, 3.0)	1.1 (-4.5, 6.9)
20	6.2 (0.9, 11.8)	3.6 (-1.7, 9.7)	-1.5 (-7.1, 4.3)	1.7 (-4.1, 7.4)
10	7.3 (2.1, 12.2)	4.9 (-0.5, 9.9)	-0.3 (-5.7, 5.2)	3.0 (-2.3, 8.8)
0	8.4 (3.4, 13.9)	5.8 (0.8, 11.4)	0.6 (-4.2, 6.7)	4.1 (-1.6, 9.2)
-10	9.2 (4.4, 14.7)	6.9 (1.8, 11.8)	1.8 (-3.3, 6.8)	4.9 (-0.3, 10.1)
-20	10.3 (5.3, 15.2)	7.6 (2.6, 12.9)	2.4 (-2.9, 8.5)	5.8 (0.3, 11.4)
-30	11.1 (5.7, 15.9)	8.7 (3.2, 13.9)	3.5 (-2.2, 8.8)	6.9 (0.7, 12.3)

Table B.5: Posterior mean and 95% HPD credible interval of overall treatment effect for β_{vm} = {-30, -20, -10, 0, 10, 20, 30} with varied γ_0 and γ_x

Variable		Posterior Mean	95% HPD CI
Time			
	Baseline		
	6 Months β_{v1}	6.9	(2.5, 11.0)
	12 Months β_{v2}	10.6	(6.2, 14.9)
	18 Months β_{v3}	14.2	(9.5, 18.6)
	24 Months β_{v4}	15.4	(10.6, 20.8)
Time × Housing First			
	HF × 6 Months β_{vx1}	8.4	(2.3, 13.8)
	HF × 12 Months β_{vx2}	6.1	(0.8, 12.0)
	HF × 18 Months β_{vx3}	1.0	(-4.8, 7.6)
	HF × 24 Months β_{vx4}	4.2	(-2.5, 10.5)
μ_θ		74.4	(71.9, 77.0)
σ_θ		15.5	(13.9, 17.1)
σ		16.6	(15.9, 17.3)
β_{vm}		-4.3	(-29.8, 23.1)

Table B.6: Posterior mean and 95% HPD credible interval of model parameters for Bayesian sensitivity analysis (BSA) applied to the At Home data using a uniform (-30, 30) prior distribution on the bias parameter β_{vm}

Method	β_{vx1}		β_{vx2}		β_{vx3}		β_{vx4}	
	Cov.	Length	Cov.	Length	Cov.	Length	Cov.	Length
Naïve	45%	4.5	46%	4.5	39%	4.5	38%	4.5
BSA	87%	23.1	89%	23.1	90%	23.2	89%	23.1

Table B.7: Coverage and length of 95% HPD credible intervals for the naïve and BSA methods obtained from the simulation study

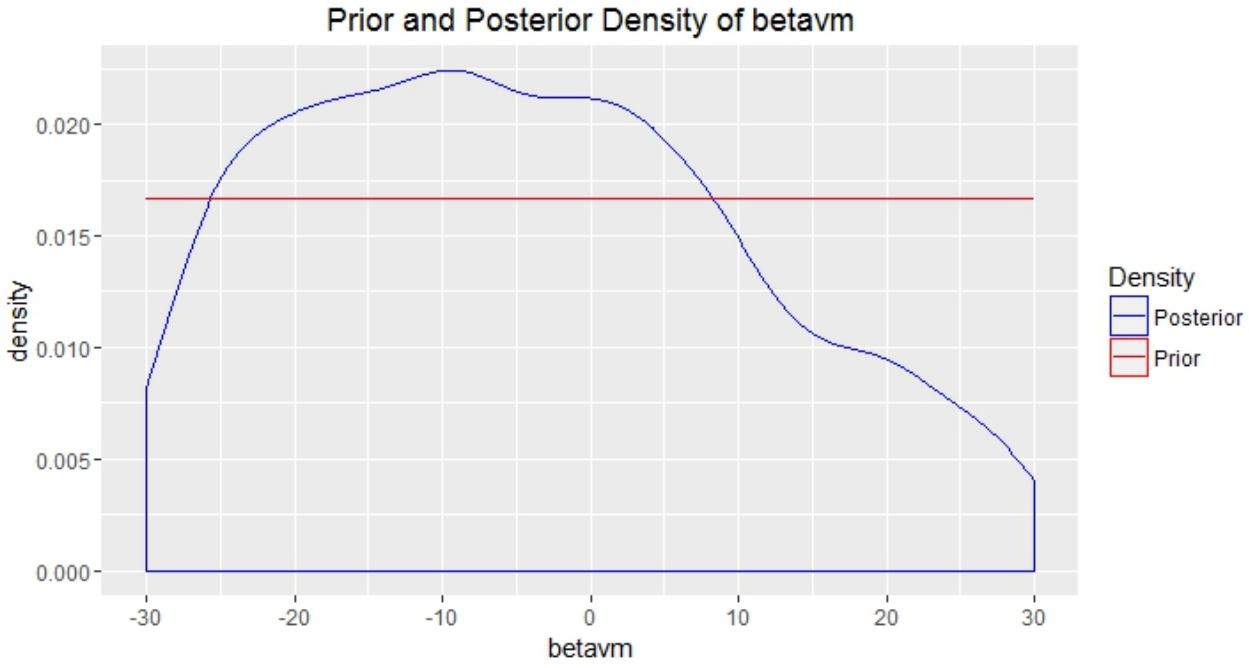


Figure B.5: Prior and posterior density of β_{vm} obtained using Bayesian sensitivity analysis