Longitudinal Data Analysis for Improving Patient Outcomes Following Hip Replacement Surgery

Bethany Phipps

January 18, 2019

Thesis submitted for the degree of
Master of Philosophy
in
Statistics
at The University of Adelaide
Faculty of Engineering, Computer and Mathematical Sciences
School of Mathematical Sciences

THE UNIVERSITY
of ADELAIDE

Contents

Abstract

Joint replacement surgery leaves many patients with postoperative pain and function limitations for extensive periods of time after surgery. This research will predict the likelihood of poor symptomatic recovery following surgery using preoperative patient data, including data on patient age, sex and comorbidities. The dataset to be analysed is total hip replacement data collected between 1989 and 2013 at the Royal Adelaide Hospital. Using the pain and function data collected repeatedly after surgery, longitudinal data analysis will be explored. The mortality information in the data will be used to explore the survival probability of patients based on different predictors using survival analysis. Repeated pain and function outcomes are modelled using mixed-effects modelling. The joint modelling of both survival and longitudinal models will be developed. Prediction methods surrounding these models will be compared to help assess the potential benefits of total hip replacement surgery for patients prior to surgery.

Signed Statement

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Signed: .... Date: ....

Acknowledgements

I would like to begin by thanking my supervisors, Professor Patricia Solomon, Dr Simon Tuke, Dr Tyman Stanford and Associate Professor Gary Glonek. Thank you for your support, whether you were with me for the whole journey or only a small part of it. Specifically, thank you to Professor Patricia Solomon for all you time and direction and Dr Simon Tuke for taking me on and being a pleasure to work with. Also thank you to Dr Melissa Humphries for helping edit my work. Thank you all for the little ways you took care of me and boosted my confidence so that I could complete this.

Thank you to the staff at the Royal Adelaide Hospital, particularly Kerry Costi and Professor Bogdan Solomon, for sharing the data and answering all my questions. Seeing a total hip replacement up close was a truly remarkable experience and I am grateful that you made it possible.

To my parents, thank you for being excited by my love of maths. I practised all my presentations on you and not only were you attentive but you helped me believe that my work was important. Your pride in my achievements is the greatest gift I could ask for. Thank you for making me feel as though there is nothing I cannot achieve. To my grandparents and siblings, Esther, Ben, Ethan, Sam and Jacob, thank you for all the laughs over family meals which kept me sane.

To my closest friend, Kiara, thank you for taking the ups and downs of this journey

with me. Seeing you always helped me forget about the stress and laugh a dangerous amount.

To my partner, Alex, for his endless belief in me. Thank you for supporting me through this and for always getting me Haigh's when the research seemed overwhelming. Whenever I began to doubt myself you could always pull me back up. Your love went a long way.

Dedication

For my parents. Thank you for always being a source of inspiration, encouragement and love.

Chapter 1

Introduction

Total hip replacement (THR) surgery is a procedure in which damaged bone and cartilage is removed and replaced with prosthetic components. This surgery is often done when other treatment options do not provide adequate pain relief. After surgery a patient returns for multiple follow-up appointments where their joint pain and function is measured with a clinician. These measurements are obtained by answering questions about the level of pain the patient experiences and through a clinician assessing the patient's range of motion. Each patient is intended to have multiple follow-ups after surgery; therefore their pain and function is measured repeatedly through time. This type of data is called longitudinal data as it involves repeated patient measurements after surgery. This is the type of THR data that we received from the Royal Adelaide Hospital and will be explored in this thesis.

In this case, the longitudinal data is collected retrospectively as we are using historical THR data. The dataset contains not only the patient's repeated pain and function measurements but also many variables such as age and gender. These variables contain enough information to help us predict an individual patient's pain and function at some

point after surgery. For example, if a patient is a particular gender or age this may help us predict that their pain and function at 24 months after surgery will be poor. Therefore, we wish to use patient measurements of joint pain and function to predict postoperative patient pain and function based on preoperative factors, such as age and gender.

The pain and function measure that we use is the Harris hip score (HHS). The HHS is a clinician-based outcome tool used to evaluate patient recovery following THR surgery by combining measures of pain and function into a single score [12]. It is a score between 0 and 100, where higher scores are associated with less pain and better function and lower scores are associated with more pain and reduced function. A HHS of less than 70 is classified as poor pain and function [12]. This score is the combined measurement of patient pain and function that is used in this study.

We use statistical modelling techniques to determine ways to predict postoperative patient HHS based on preoperative factors. There are two main modelling techniques that are explored in this thesis; mixed-effects models and joint models. Mixed-effects models are a type of statistical model that is used to analyse correlated data, such as repeated measurements. They include both fixed effects and random effects, where fixed effects are the parameters associated with an entire population and random effects are the parameters associated with the individual within the population. The main advantage of this model is its ability to separate age and cohort effects. This means that it can distinguish changes over time within patients from differences among patients at the baseline level. In turn, this model describes the relationship between a response variable, such as the HHS, and some covariates, such as age and gender.

Along with a longitudinal response variable, this dataset also includes a time-to-event outcome, patient death. Instead of analysing these two outcomes separately we can analyse them together within a joint model framework. This allows us to account for the

effect of the HHS, the longitudinal outcome, as a time-dependent covariate. This model describes the relationship between a longitudinal response variable and some covariates, as well as the relationship between the hazard of survival and some covariates, including the longitudinal response variable.

A key difference between this study and previous studies into THR is the use of subjective reports of symptomatic relief and patient satisfaction, namely the HHS [6]. From the literature, on average between 25% and 50% of patients report persistent postoperative pain and function limitations at one to two years after joint replacement surgery by which time there should be resolution of symptoms [21, 7]. Within Australia alone approximately 590,000 THR procedures took place in 2016, and this is only increasing each year [6]. This is a large number of patients who are not experiencing the expected results of THR surgery. Therefore, research that focuses on improving symptomatic relief for patients after THR surgery, by focusing on measures such as the HHS, is crucial. Otherwise, the intended advantages of THR, such as reducing individual, social and health care burdens, cannot be fully realised.

This study also acknowledges the increased risk of adverse patient outcomes after a THR due to comorbidity and multi-morbidity, which is not always addressed. Two or more coexisting medical conditions or diseases that are additional to the initial diagnosis for surgery are called comorbidities and multi-morbidities [1]. In 2009, 49% of people within Australia aged 65 to 74 had five or more long-term chronic health conditions and this prevalence increased with age [23]. This means that patients presenting for THR surgery are likely to have multi-morbidities. So the management of patient care for patients with comorbidities needs to be addressed. This study clearly addresses the role of comorbidities in patient recovery after THR surgery.

Firstly, in Chapter 2, we present a summary of the THR dataset that will be analysed.

Data cleaning and missing data procedures are explained. We use the data summaries to determine characteristics of THR patients. An understanding of the limitations of the dataset prepare us for preliminary analysis. This analysis begins with linear regression, explored in Chapter 3. We show that this type of model, one with only fixed effects, is not appropriate in the context of our dataset as it does not address the correlation between repeated measures within one patient. Therefore, this chapter provides motivation for longitudinal data analysis, specifically mixed-effects modelling.

In Chapter 4 we present an introduction to mixed-effects modelling, subsequently using the THR dataset, and a model selection process, to create an optimal model. We find that variables such as age, gender and number of comorbidities significantly contribute to a patient's pain and function after surgery. We discuss methods of prediction relating to mixed-effects models. For new patients that are not within the estimation model frame the use of best linear unbiased predictions (BLUPs) of the random effects of patients with similar characteristics is implemented.

For some THR patients a HHS measurement is taken at the point of surgery. This means that the change in HHS after surgery can be determined. In Chapter 5, we continue with mixed-effects modelling but the response variable is instead the change in HHS. The predictor variables within this model vary significantly from the model in the previous chapter. Through prediction we can assess whether a patient is likely to experience any pain and function improvements at 24 months after surgery.

In Chapter 6 we introduce survival analysis using patient survival information, such as date of death. We produce a Cox model to show that the number of comorbidities that a patient suffers from is associated with their risk of survival [9]. The time-dependent model is shown to be inappropriate for modelling the biomarker HHS as a predictor of survival. This forms the motivation for joint modelling.

In Chapter 7 we provide an overview of modelling both the survival and longitudinal outcomes through joint modelling. We use this model framework to create a joint model where survival status and HHS are both outcomes. We compare these results to the optimal mixed-effects model in Chapter 4 as well as the HHS predictions. Specific patient predictions, relating to both the HHS and survival outcome, are determined using dynamic prediction.

Some THR patients undergo surgery on both their left and right hip. Often one side is operated on first and then the second surgery is performed at a later time. In Chapter 8 we apply joint modelling where the survival outcome is second surgery. Therefore, allowing us to assess the risk of a patient having to undergo a THR on their other side after their initial surgery. This showed that patients who experience worse pain and less function, after their first THR, are more likely to undergo a second surgery.

To address the problems associated with combining both pain and function, two important measures, into one score we then separated them into separate scores. These two outcomes, pain score and function score, were then analysed independently. Joint modelling with multivariate longitudinal outcomes was then introduced and applied within the context of pain score and function score being separate longitudinal outcomes. Both pain score and function score interacted differently with the survival sub-model showing the importance of addressing two important measures separately.

In summary, this research explores THR surgery and develops statistical predictive algorithms to determine the likelihood of recovery and satisfaction with surgery within the first two years following surgery. Not only will this enable patients to make an informed decision before choosing to undergo surgery but it will also lead to optimising preoperative patient education as part of a greater holistic approach.

Chapter 2

Data Summary

This thesis will analyse the clinical and health outcomes of data collected prospectively for 3049 primary total hip replacement (THR) procedures between 1989 and 2013 at the Royal Adelaide Hospital (RAH). The study will utilise data maintained as part of the Orthopaedic and Traumas Joint Replacement Database, which includes data on patient socio-demographics, and death status cross-matched against the South Australian Births, Deaths and Marriages Database. Clinical outcomes in the database include doctor-assessed joint range of motion. Patient recorded outcomes include joint pain and function measured at multiple points after surgery.

The dataset was in the form of a relational database (a database structured to recognise the relations between stored data) and was subsequently converted to a flat file (removes internal hierarchy) for the purposes of this study. Before analysing the dataset it was important to identify erroneous data entries, so they could be removed and not affect further analysis. Therefore, we spent time initially reformatting the data, addressing free text fields and performing general data cleaning.

This Chapter will begin by explaining both the variables and the characteristics of

the cleaned, and partially imputed, THR dataset that is used in later analysis. This is done to provide clear context to the research problem. Following this, information about the data cleaning procedures and imputation methods used will be explained. Therefore, clarifying the decisions that were made when preparing the dataset for analysis. All data analysis was completed using R [13].

2.1 The Data

The dataset received from the RAH contained information on THR patients with surgery dates evenly spanned over a 25 year period, from the start of 1989 to the end of December 2013. We were given 92 variables, which included patient age, gender, lifestyle factors, operation details, comorbidities, medications and longitudinal pain and function data following surgery (see Appendix A.1 for a full description of the variables). The cleaned dataset, which will be summarised in this section, is reduced to 2280 patients as a result of data cleaning and missingness (see Sections 2.3 and 2.4). Table 2.1 shows descriptive data, frequency and percentages of patient demographics, for THR patients. There are more females than males, with 1341 (58.8%) females. Whether the operation is performed on the left or right side is almost balanced with 450 (19.7%) patients who had both left and right hip replacements. Figure 2.1 is a histogram of the age of THR patients at the point of surgery. The solid black line marks the mean age of 67 years. The majority of the patients (65%) were aged between 60-80 years old, at the time of THR surgery. The youngest patient is 16 years old and the oldest is 100 years old.

Osteoarthritis (OA) is a disease that results in the degeneration of joint cartilage and bone [1]. Rheumatoid Arthritis (RA) is a chronic disease that causes inflammation in the joints resulting in deformity and immobility [1]. A high proportion of patients that

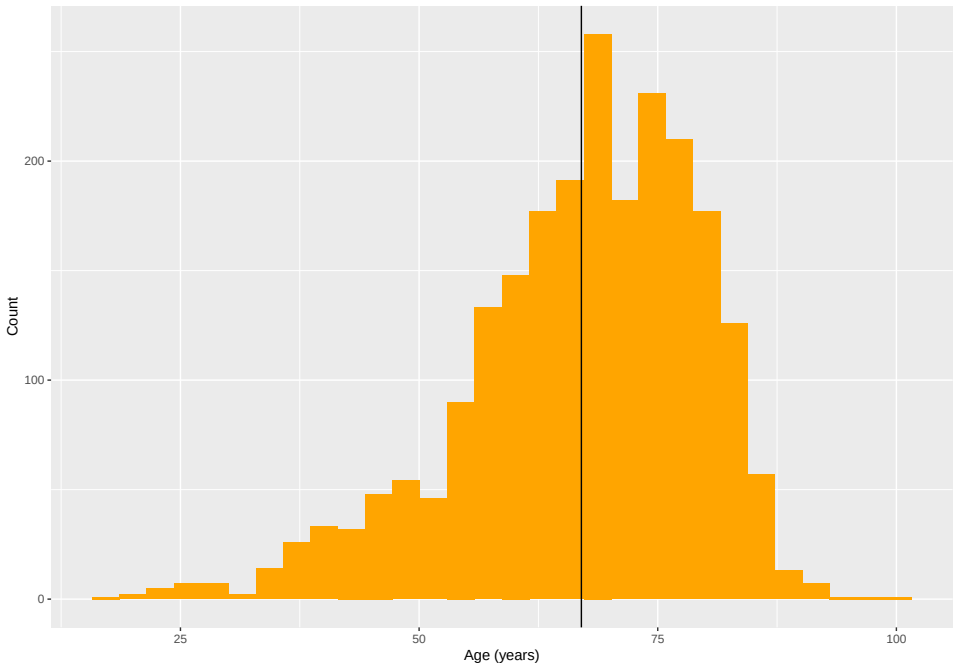

Figure 2.1: Histogram of the age of patients at the point of surgery. The black solid line refers to the mean age of these patients, 67.2 years (n=2280).

present for THR suffer from either OA or RA. OA is the most common form of arthritis that these patients suffer from when presenting for surgery, although this is not necessarily the primary reason for surgery (see Table 2.1). Avascular Necrosis (AVN) is a disease which occurs when a blood supply failure results in the death of bone tissue [1]. No patients that presented for THR surgery suffered from both OA and RA. More patients in the dataset suffer from AVN than RA but the majority of patients suffer from OA (87.9%).

Many patients that present for THR surgery also suffer from comorbidities. Comorbidity is the presence of one or more additional diseases or disorders co-occurring with a primary disease or disorder [1]. Approximately 70% of patients within the dataset have comorbidities. Table 2.1 shows descriptive data on the ten most prevalent comorbidities

Patient Demographics	
n of patients	2280
Mean (SD) age (yrs)	67.2 (12.3)
Mean (SD) follow-up time (months)	36.8 (49.5)
Mean (SD) weight (kgs)	78.6 (17.7)
Mean (SD) height (cms)	165.4 (12.1)
Gender (n,%)	
Female	1341 (58.8)
Male	939 (41.2)
Operation Side (n,%)	
Left	1029 (45.5)
Right	1232 (54.5)
Diagnosis	
Osteoarthritis (n,%)	1734 (87.9)
AVN (n,%)	182 (9.2)
Rheumatoid arthritis (n,%)	57 (2.9)
Comorbidity (n,%)	
Other	1163 (51.0)
Hypertension	596 (26.0)
Asthma	503 (22.1)
NIDDM[1]- stable	140 (6.1)
Angina - stable	109 (4.8)
Prostatism	87 (3.8)
COAD[2]	84 (3.7)
CCF[3]	55 (2.4)
Current malignancy	38 (1.7)
Hypothyroidism	34 (1.5)
Number of Comorbidities (n,%)	
None	704 (30.9)
One	745 (32.7)
Two	497 (21.8)
Three	240 (10.5)
Four or more	94 (4.1)
Taking medication (n,%)	1829 (80.2)

[1] Non-Insulin-Dependent Diabetes Mellitus.
[2] Chronic Obstructive Airways Disease.
[3] Congestive Cardiac Failure.

Table 2.1: Frequency (and percentages) of patient demographics for THR patients (n=2280). Only the ten most prevalent comorbidities shown in table.

in the dataset. The two most common comorbidities are hypertension (26%), abnormally high blood pressure, and asthma (22%). The number of comorbidities that a patient suffers from is also shown in Table 2.1. Approximately 33% of patients suffer from one comorbidity, while only 22% suffer from two comorbidities. Due to these comorbidities, and pain after THR surgery, 80% of patients take medication.

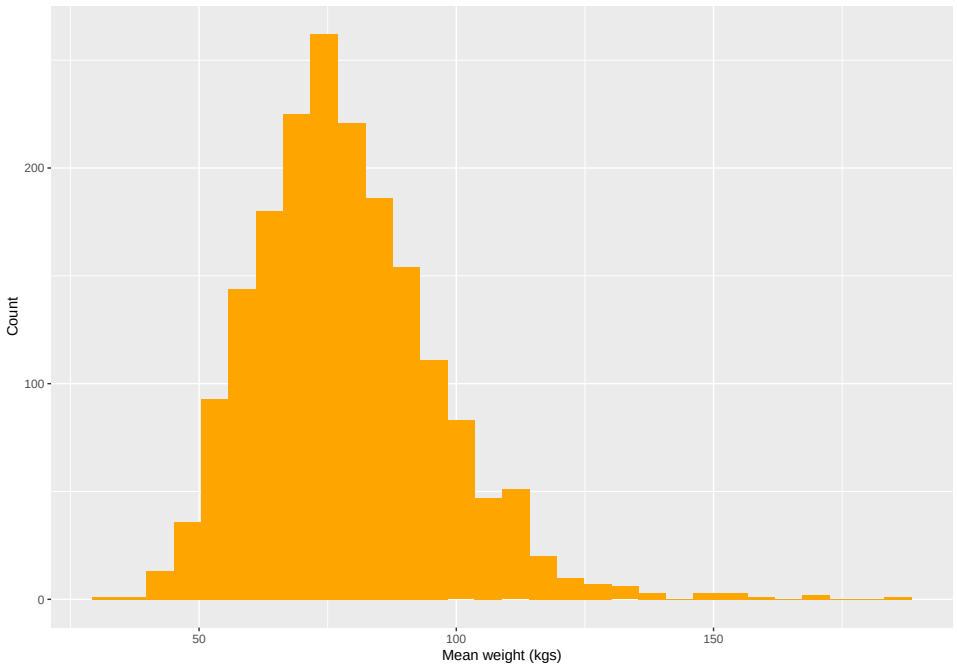

Figure 2.2: Histogram of the weight of patients at the point of surgery. The x-axis is the imputed weight measurement, which is the mean weight for each patient (n=1663).

The distribution of the imputed weight and height variables is also shown in Table 2.1. This application of imputation, due to missing data, is explained in Section 2.4. The mean weight of patients is approximately 79 kg (sd=17.7kg) while the mean height is 165 cm (sd=12.1cm). Figure 2.2 is a histogram of weight where the x-axis is the imputed weight variable, which is the mean weight for each patient. There is a right skew in the histogram as there are some patients with heavier weights, while the majority of patients

(64%) weigh between 60-90 kg.

The dataset contained the date of death (DOD) for approximately half of the patients (47%). This information is useful for examining whether there is a relationship between operation date and death when considering other factors such as patient age, gender, operation details and comorbidities. This is explored through survival analysis (see Chapter 6). Half of the patients with DOD's die within 8 years following their operation date. This may be due to hospital records losing track of patient information, such as their DOD, after long periods of time. Therefore, those patients that die within eight years are recorded whereas those that die well after this time are not recorded due to loss of contact, creating a bias in the data.

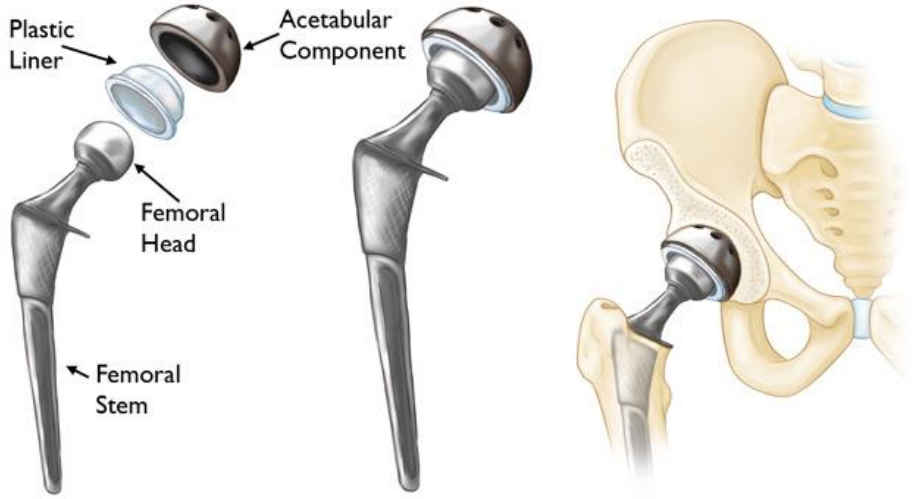

Figure 2.3: Diagram of Acetabulum and Femoral implants used for THR surgery [2].

In THR surgery a patient receives an acetabulum and a femoral implant. These implants are artificial prothesis that replace the ball and socket bones in the hip joint. Figure 2.3 shows the components of the acetabulum and femoral implants and their position within the hip following surgery. There are different brands associated with each

implant and the prevalence of each brand for both the acetabulum and the femoral are shown in Table 2.2. There are specific combinations of acetabulum and femoral implants that are used, such as Exeter and Exeter for the acetabulum and femoral implant respectively. Only the most common implant combinations are shown in Table 2.2. We see that Trilogy and CPT is the most common type of acetabulum and femoral implant combination, as 44% of patients use this.

Type of Implant (Acetabulum & Femoral)	**Number of Patients** (%) n = 2280
Trilogy & CPT	1004 (44.0)
Exeter & Exeter	346 (15.2)
Vitalock & Exeter	215 (9.4)
Charnley & Charnely	182 (8.0)
PCA & PCA	108 (4.7)

Table 2.2: Table of the types of Acetabulum and Femoral Implants (n=2280).

After surgery, prosthetic hips can sometimes wear out or cause other problems and a further operation is required to replace the implant. This is called a revision. As expected the survival of the acetabulum and femoral are almost identically distributed. This is because most procedures replace both the acetabulum and femoral implants simultaneously. Though revision is often used as a measure of success after THR, where lack of need for a revision is a positive outcome, we can also look at the pain and function of a patient as indicative of the benefits of the surgery. To measure the pain and function following THR we can use the HHS.

2.2 Harris Hip Score

The Harris hip score (HHS) is a clinician-based outcome tool used to evaluate patient recovery following THR surgery [12]. It is a score between 0 and 100, where higher scores

are associated with less pain and better function and lower scores are associated with more pain and reduced function. A HHS of less than 70 is classified as poor pain and function [12]. Figure 2.4 shows a histogram of the HHS for all patient follow-up measurements. It is clear that this variable is bounded at 100 and is left skewed. The majority of patients have a HHS of 70 and above and therefore do not have poor pain and function.

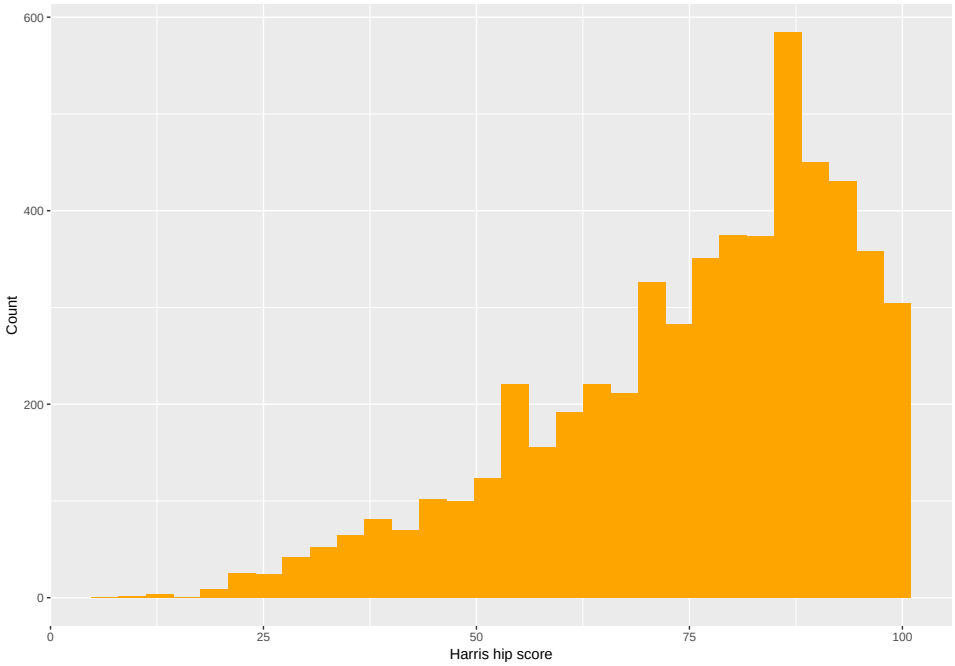

Figure 2.4: Histogram of the Harris hip score of patients at each follow-up (n=1905).

Following surgery each patient completes an outpatient form at each follow-up visit. The outpatient forms contain a series of questions relating to the mobility and pain of their joint(s). For example, the data contains the range of motion in internal and external rotation of the hip, if there is a limp and what causes it, leg length and leg power. These fields are then used to calculate the Harris hip score (see Appendix A.2 for full description of the HHS calculation process). It is intended that each patient completes this outpatient form at 3, 12 and 24 months after surgery, depending on the precise timing of their return

clinic visits. The time between the date of surgery and the date of an outpatient form being completed is called the 'follow-up time'.

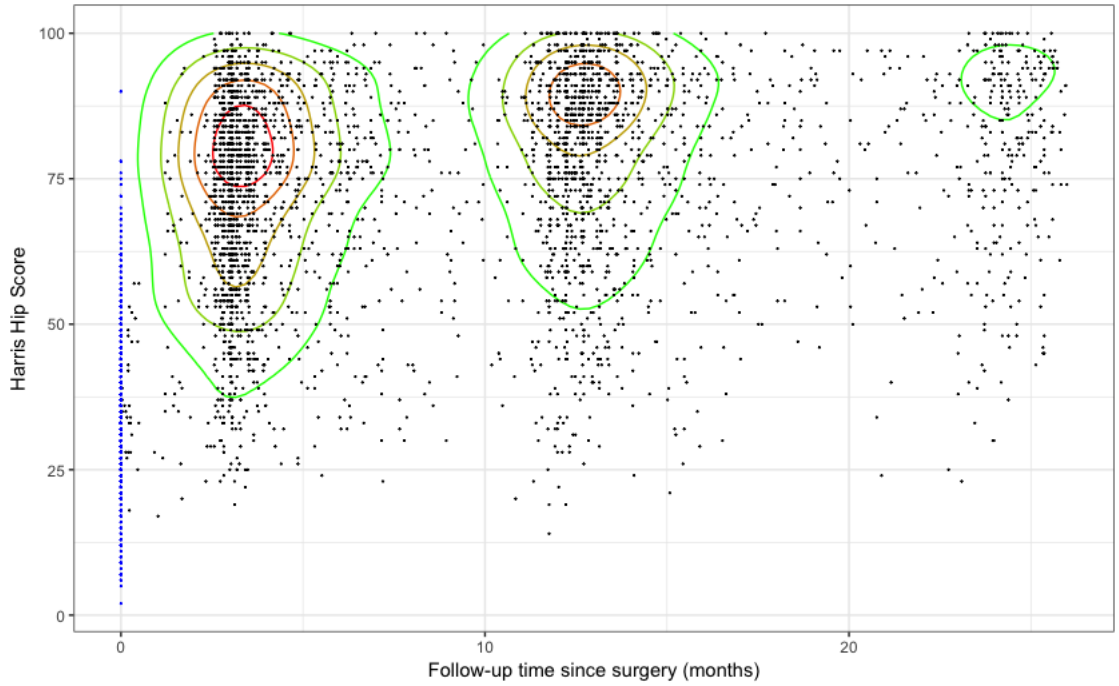

Figure 2.5: Distribution of the HHS for non-negative and non-zero follow-up times. Follow-up times greater than 26 months are excluded. Points labelled blue show the HHS measurements at the time of surgery (follow-up time is zero). The coloured lines display the density of the points where red indicates high density and green indicates low density (n=1905).

Figure 2.5 shows the distribution of the HHS for non-negative and non-zero follow-up times, where follow-up times greater than 26 months are excluded. The x-axis is the follow-up time in months and the y-axis is the HHS. Points labelled blue show the HHS measurements at the time of surgery (follow-up time is zero). The coloured lines display the density of the points where red indicates high density and green indicates low density. It is clear that over time, the follow-up data becomes more sparse. Nevertheless, we see that the 3, 12 and 24 month scheduled follow-up times are directly reflected in the data.

Table 2.3 shows the follow-up distribution of THR patients for the first six follow-ups. We see that 84% of patients have a minimum of at least one assessment during the postoperative period with 64%, 41% and 26% assessed 2, 3 and 4 times respectively. Some patients have as many as 10 follow-up assessments with the maximum being a patient with 14 follow-up assessments. Almost 50% of the patients within the dataset have at least three follow-up measurements, which encourages the exploration of a quadratic term for time (months) in later models.

Follow-up	**Number of patients (%)**	**Mean (months), (SD)**
First follow-up	1905 (83.6)	5.7, (16.6)
Second follow-up	1464 (64.2)	12.3, (20.1)
Third follow-up	942 (41.3)	27.2, (27.6)
Fourth follow-up	588 (25.8)	50.0, (34.7)
Fifth follow-up	338 (14.8)	71.6, (40.3)
Sixth follow-up	175 (7.7)	94.2, (45.0)

Table 2.3: Table of the follow-up distribution of THR patients for the first six follow-ups (n=1905).

The HHSs given within the data were used primarily throughout analysis as the longitudinal outcome. However, there are obvious issues with combining two important measures such as pain and function into a single variable. In Chapter 8 we will address this explicitly by developing an algorithm which can decompose HHS into its pain and function components for use, separately in a multivariate longitudinal model.

At this stage, with a greater understanding of the THR dataset, we can now explain the process that was used to clean this dataset. Understanding this process is important as it makes clear the limitations of this dataset and the decisions that were made to address these limitations.

2.3 Data Cleaning

Data cleaning is the process of detecting inaccurate records from a dataset and then determining whether to modify or delete these entries. For example, a patient that was entered as 6 years of age is not a real measurement as patients this age do not undergo THR surgery at the RAH. In this case you could either check the patient records and correct the age, or if that information is not available, remove the age or the entire patient from the dataset. In our case, we did not have access to records that would enable us to correct inaccurate details, so the following decisions were made.

- The patient that was 6 years of age was removed.
- The weight of patients was recorded in kilograms, therefore entries that were greater than 200kg or less than 30kg were removed.
- Heights that were entered as greater than 210cm or less than 60cm were removed.

We also decided to remove the four patients from the dataset who had undergone a THR on both sides within one operation, a bilateral THR. These specific patients were removed due to the potential difference in recovery from undergoing a bilateral THR compared to a regular THR.

Another facet of data cleaning is simplifying and relabelling variables. Variables may not be inaccurate but may be in an inappropriate form. An example of which in our data was the year of operation being entered inconsistently. The year of operation was sometimes entered as two digits, such as 93, but needed to be converted to four digits, 1993.

New variables were also created using natural language processing, which is the application of computational techniques to the analysis of natural language. For example,

within the dataset there was a variable that specified patient medications. The information was entered in the following form for five patients:

```
[1] Aspirin, Voltaren, Frusemide, Provastatin...
[2] Aspirin, Warfarin, Tramadol, Ibuprofen
[3] ATROBEL,CALTRATE
[4] Aspirin, Zimsta, Coversyl, Simvastatin...
[5] ATRVNT, VNTLN, BEDOFORTE, ARUDIS, LASIX...
```

We created a variable that identified if a patient was taking a strong opioid by using natural language processing to mark when the words morphine, tramadol, endone and others were mentioned. Therefore, the second patient in the above list would be identified as using strong opioid medication due to the mention of tramadol. This was also completed for other medication groups such as paracetamol and weak opioids. Whether a patient smoked marijuana was also determined using this method on a social drug specification variable. Variables such as the hip side, implant brand and place of residence were converted to numeric factors.

The dataset specified separately whether a patient smoked and how many cigarettes a day they would have. These two variables were combined to create one variable that identified patients as either non-smokers, light smokers or heavy smokers. When relabelling this variable we made the assumption that a patient with an unspecified number of cigarettes per day, but that is specified as being a smoker, is a light smoker. A similar process was completed for the relabelling of the alcohol status variable. These variable relabelings are outlined in Appendix A.3.

Another new variable was created to combine patient information about comorbidities. The dataset contained a binary variable that identified whether a patient suffered from

a comorbidity and also a variable that stated the specific comorbidity, number coded from 1 to 45, where 14 is 'other'. For example, one patient is listed as suffering from stable insulin-dependent diabetes mellitus, prostatism, asthma, hypertension and 'other'. Using this information we created a new variable, with an upper limit of 'four or more', which stated the number of comorbidities a patient suffers from at their first follow-up. Therefore, the above patient was marked as suffering from four or more comorbidities.

After detecting and addressing inaccurate records from the dataset the next step was to develop protocols for dealing with missing data.

2.4 Missing Data

Often observations within electronic health records are not complete. For example, a patient may have their age recorded but not their place of residence. This empty field is called missingness or missing data. There are different ways that we can view missingness that affect how we address entries not being observed. For example, a patient may not have their height or weight measured at the time of surgery. Therefore, we can view this missingness as underlying values that could have been observed but were not, potentially due to the patient's request or lack of time. We do not question whether a patient has a height or weight just that it was not measured. Therefore, it is more natural to treat these unobserved values as missing [19]. Likewise for variables such as place of residence we do not assume that a patient has no place of residence even if this is unobserved.

In other cases it may not be as clear. If a patient has no entry in relation to the type of medication that they take, does this imply that a patient is not on medication or that this entry is missing? In this research, following consultation with the RAH, it is assumed that missing entries for patient medication are far more likely to imply that

a patient does not take that form of medication. We also assume that patients with no comorbidity information do not suffer from comorbidities. The variable which recorded a patient's previous surgery contained a large amount of missing data due to the inherent difficulty in accessing that information. We therefore excluded the entire variable from our analysis.

Some variables, such as the Harris hip score, were imperative to our analysis. Using some sort of numerical imputation of these missing values was not deemed appropriate due to the vital contribution they brought to the analysis. Patient observations that had missing values in this variable, or with missing follow-up times, were therefore excluded from the analysis. The effect that this had on the number of patients is shown in Figure 2.6, which is a flow chart of the effect of missingness on the number of patients within the dataset. This reduced the dataset to 1905 patients. For variables such as smoker and alcohol status we also did not assume that an unobserved value means that a patient does not smoke or drink. As a result missing entries in the variables smoker and alcohol status were replaced with NA but were still included in the analysis.

2.4.1 Imputation

Imputation is the process of filling in missing values and then using the resultant completed data for analysis [19]. When considering imputation of some of the missingness within the dataset there were some important considerations. We should consider the intended use of the model, which in this case is the prediction of a patient's pain and function after surgery. Previous research has shown that weight and height are all important predictors in the context of THR surgery [4, 16, 8]. For this reason, the missing values in these variables were addressed using imputation techniques that are outlined in this section.

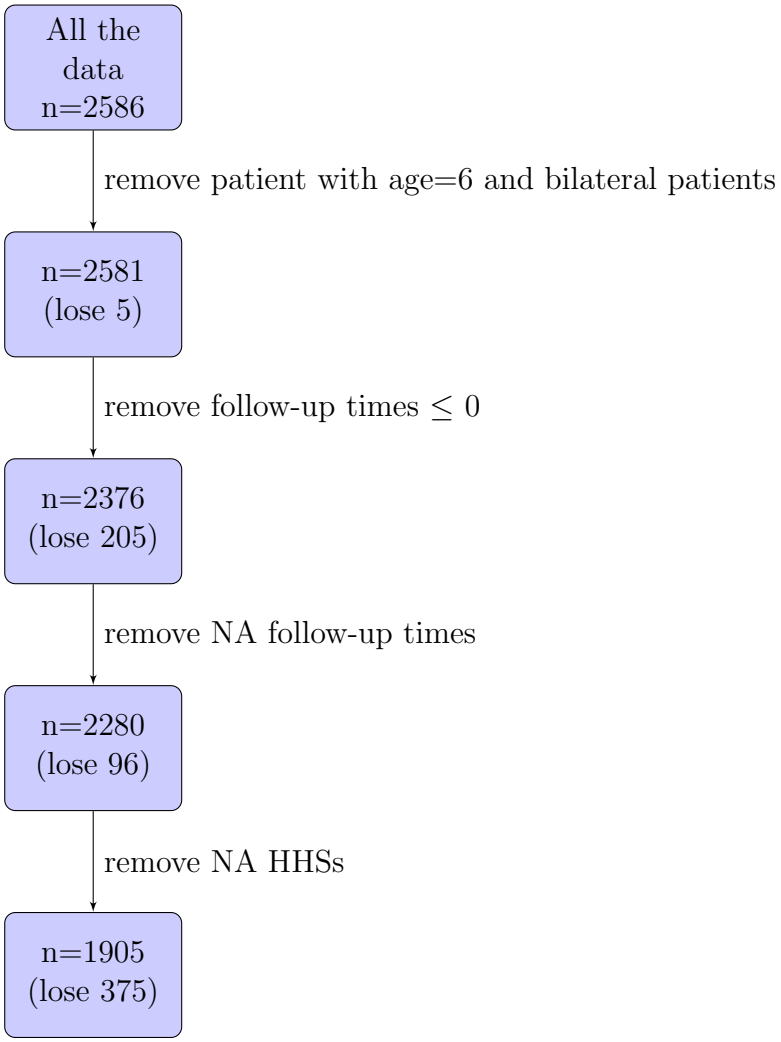

Figure 2.6: Flow chart of the effect of missingness on the number of patients to be analysed.

For weight and height, it was assumed that the missingness is completely random. We acknowledge that variables with lots of missing data points would be expected to end up with larger error terms than those with fewer missing data points, so the ability to detect significant relations to those variables would be limited accordingly. For these reasons we did the following exploration to decide whether the missingness within the weight and height variables should be addressed using imputation or not. Specifically

mean imputation, where means from subgroups of the sample are substituted [19]. In this case, we explore substituting in means at a patient level.

Figure 2.7 shows the missingness in the covariates weight, height, implant brand, place of residence, smoker level and alcohol level. The histogram in Figure 2.7 displays the proportion of missing data within each of the covariates. The variables height and weight contain the most missingness with approximately 70% of observations being unobserved. The plot on the right, in Figure 2.7, shows the most frequent combinations of missingness across the covariates. We see that approximately 32% of observations within the dataset do not have a measured weight, height, alcohol or smoker status and place of residence. The covariates age, gender, OA, RA, AVN, specific medications and comorbidities are not included due to their lack of missingness. Other variables that are not included here are not considered as important, such as date of previous surgery.

Since weight and height are potentially important variables, as discussed above, we need to address their high missingness. We decided to calculate the mean weight of each patient and use mean imputation to address the missingness. Though there are some patients that have significant weight loss or gain the majority of patients lie between a weight of 50-100 kg and maintain a fairly fixed weight throughout the follow-up period. The change in height over time is slightly negative for some patients but overall fairly insignificant and approximately constant.

Figure 2.8 shows a comparison between the distribution of the response variable, HHS, for missing and non-missing covariates. Where the x-axis outlines the type of covariate and the y-axis is the HHS. This shows that the distribution of HHS does not appear to be dependent on missingness, including missingness in weight and height. Similarly the age and gender distributions are approximately the same across the missingness in weight and height. Within this context using the mean of previous weight and height measurements

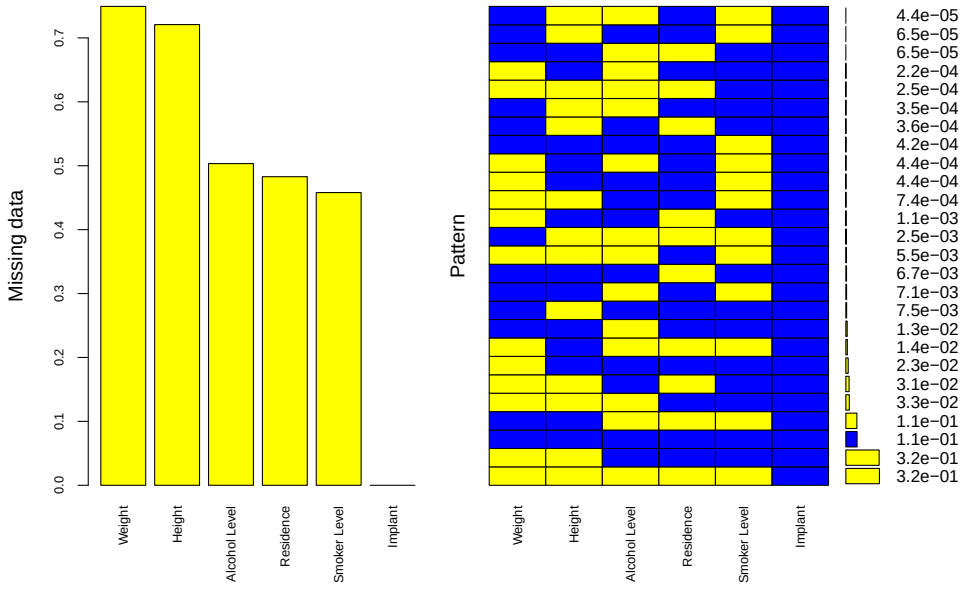

Figure 2.7: Missingness in the covariates weight, height, implant brand, place of residence, smoker level and alcohol level. Left: Histogram displaying the proportion of missing data within each of the covariates. Right: Plot showing the most frequent combinations of missingness across the covariates; e.g. approximately 32% of observations within the dataset do not have a measured weight, height, alcohol or smoker status and place of residence (n=1905).

to impute missing observations is appropriate.

Since weight is a variable that is calculated at each follow-up we impute the mean weight of each patient for each follow-up. If a patient has three follow-ups and their weight is unobserved for their final follow-up, then all weight measurements for that patient are the mean of their previously observed weights. A similar method is used for the variable height. This lowers the missingness as there are 1290 patients (57% of patients) that have at least one weight measurement with a corresponding valid HHS and 69% have at least one observed weight with or without a corresponding valid HHS. Therefore, we impute the mean weights and heights before removing the missingness in the HHS so as to avoid losing information relating to a patient's weight and height. This decision helped us retain

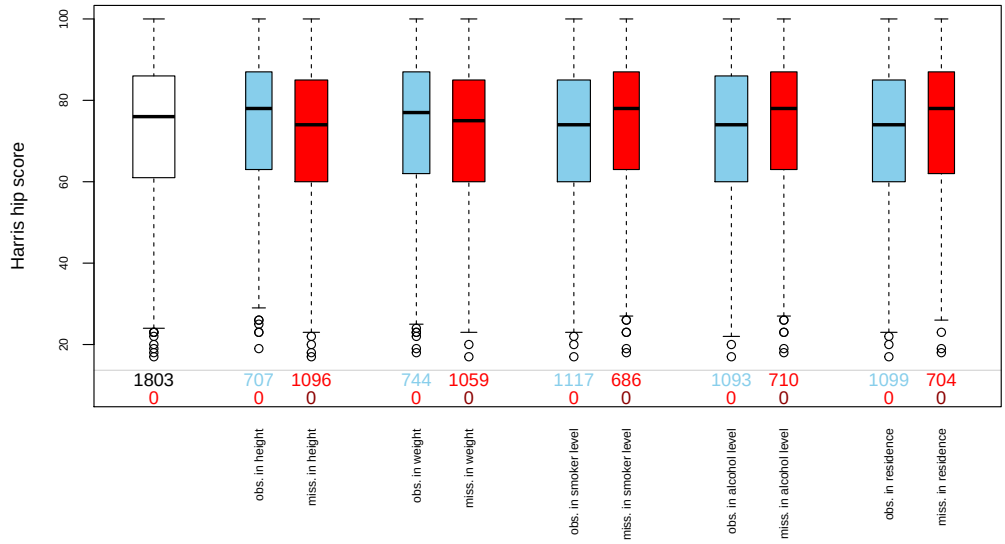

Figure 2.8: Comparison of the distribution of Harris hip score for missing and non-missing covariates weight, height, implant brand, place of residence, smoker level and alcohol level (n=1905).

and effectively use patient weight and height information.

This chapter displayed the importance of data cleaning and addressing missing data before proceeding to analysis. Specifically using mean imputation for the weight and height variables. We note that multiple imputation was not used and further work into this method would be valuable. Summaries of the cleaned data, described earlier in this chapter, show the demographic of THR patients, such as age and gender. Now that we understand the nature of the data and have addressed the missingness and erroneous data entries we can approach longitudinal data analysis to predict patient postoperative pain and function based on preoperative factors.

Chapter 3

Linear Modelling

The first statistical method that we used to explore the total hip replacement (THR) dataset was linear modelling. Therefore, in this chapter we fit a linear model with independent, homoscedastic residual errors where Harris hip score (HHS) is the response variable. However, before we explore linear modelling we present the bivariate analyses of the THR dataset as these results form the basis of feature selection for the linear model in this chapter. Therefore, this chapter begins with a summary of the bivariate analysis and is followed by an exploration of linear modelling.

In view of the structure of this dataset, and of the results shown in the bivariate analysis presented at the start of this chapter, the assumptions of independence and homoscedasticity of the HHS measurements are not appropriate. Therefore, this chapter provides the motivation for mixed-effects modelling used in Chapter 4. Nevertheless, the linear modelling results are still outlined below as an important part of the analysis that follows.

3.1 Bivariate Analysis

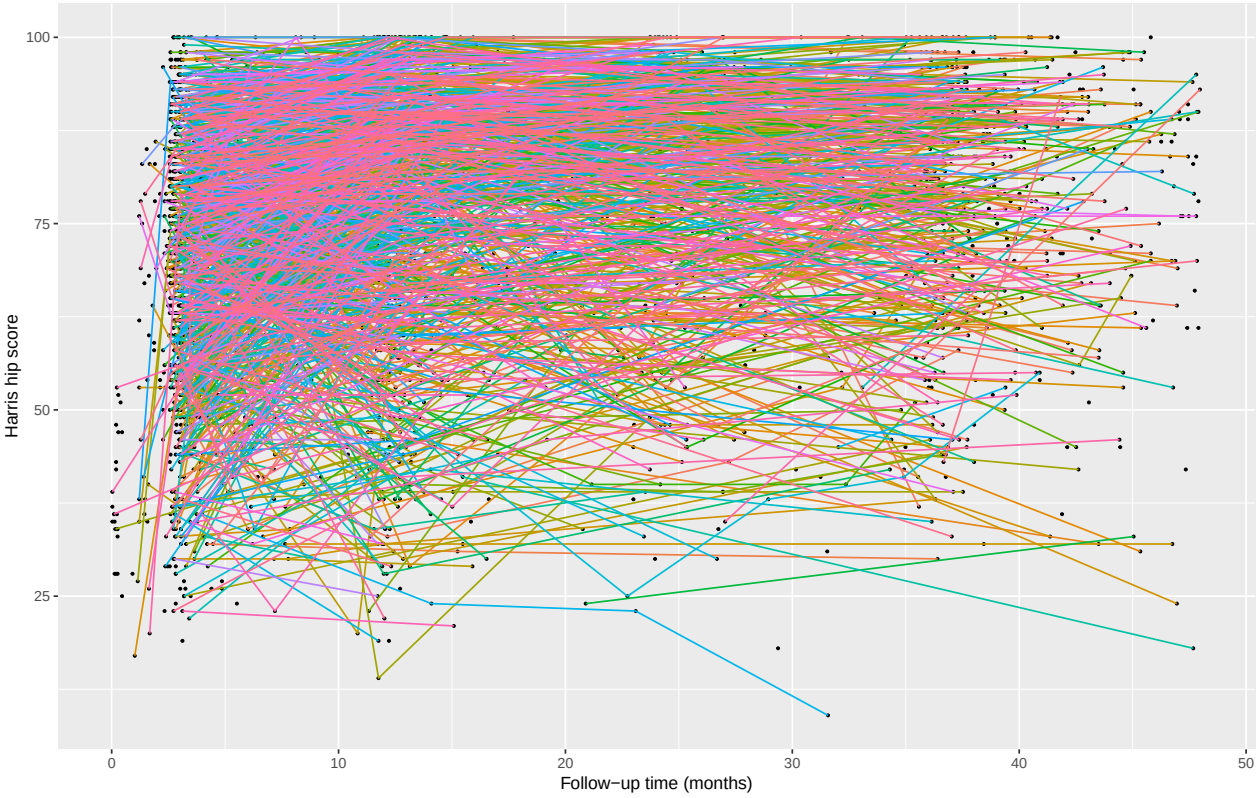

Figure 3.1: Plot of the Harris hip score at each follow-up time for all patients. The points mark each observation and each line connects the observations between one patients. The follow-up times are restricted to 48 months (n=1720).

In this section, we present the bivariate analyses of the THR dataset. These results are the basis of feature selection for the linear model in this chapter, and mixed-effects models in Chapter 4.

Since we are interested in the HHS of patients over time we begin by examining the relationship between these variables, follow-up time and HHS. Figure 3.1 displays the HHS on the y-axis and the follow-up time on the x-axis, for all patients. The points mark each observation and each line connects observations between one patient. Points that are not

connected to lines represent patients with only one HHS observation. The follow-up times are restricted to 48 months. It is clear that there is great deviation between patient HHS trajectories. While some patients appear to have a fairly constant HHS, other patients are extremely variable. It is also clear that the time at which follow-up's occur is variable, especially for follow-up times greater than 24 months.

We can likewise explore the relationship between other explanatory variables and HHS. Out of the large number of variables within this dataset a subset of variables was selected, to explore through analysis. This was done with the help of clinicians. These variables were age, gender, follow-up time, side of hip operated on, mean weight, mean height, implant brand, osteoarthritis status, rheumatoid arthritis status, AVN status, place of residence, specific medications, number of comorbidities, smoker status and alcohol status. These were selected as they were of particular interest to the clinicians but also due to practical reasons, such as low missingness.

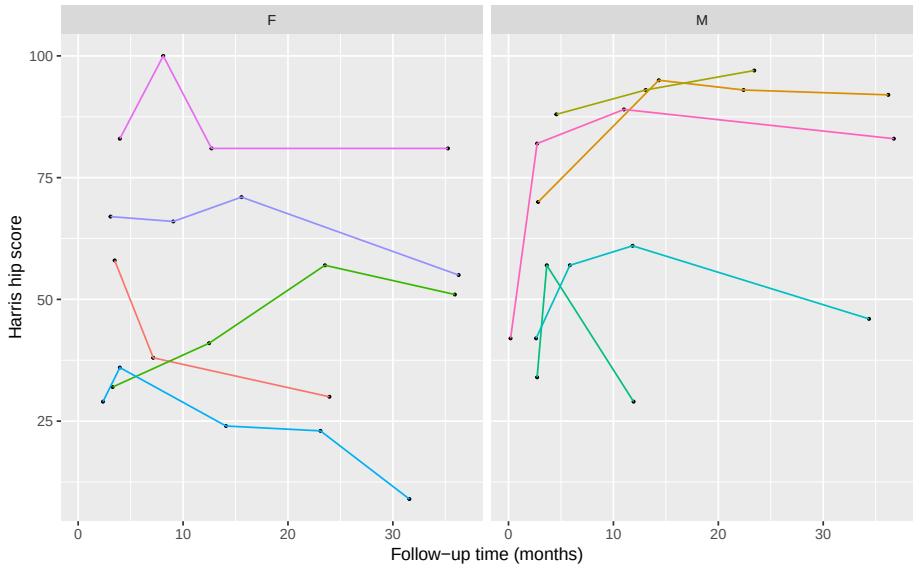

Figure 3.2: Plot of the Harris hip score at each follow-up time for five selected male and female patients. Follow-up times are restricted to 48 months (n=10).

We are interested in the effect of these preoperative factors on HHS over time. Thus in Figure 3.2 we plot HHS against follow-up time for several selected patients from both genders. From this figure we see that in general, HHS tends to slightly decrease with time after 12-24 months. Some patients display a linear change in HHS over time but there are also patients for whom individual profiles strongly deviate from a linear trend. HHS measurements adjacent in time are fairly well correlated. The HHS measured at the patient's first follow-up appears to, at least partly, determine the overall level of the patient's later HHS measurements.

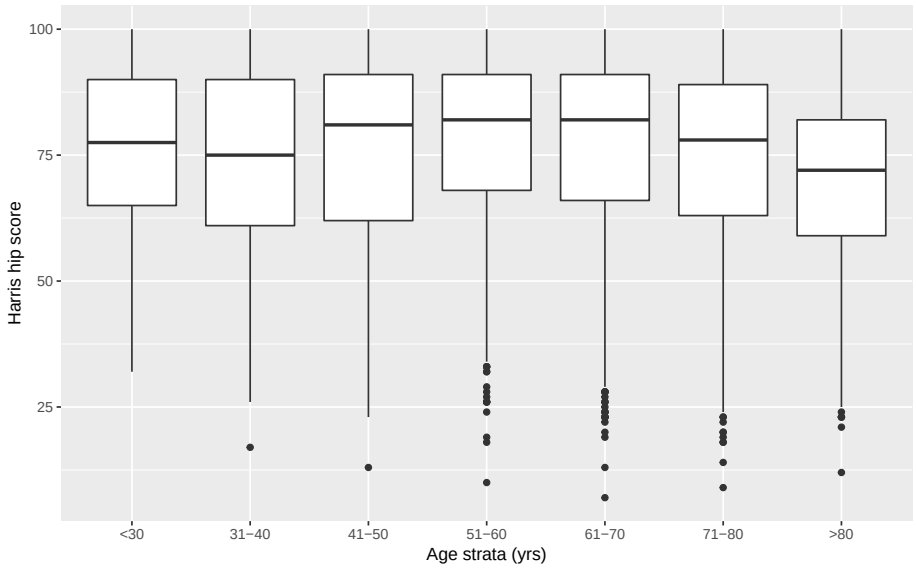

Figure 3.3: Side-by-side boxplot of the Harris hip score for each age strata (n=1720).

We can also investigate the medians and spread of HHS measurements for different variables through boxplots. Figure 3.3 presents a side-by-side boxplot of the HHS of patients within different age bands. We see that patients who are older than 80 years have a lower median HHS. Patients associated with the highest median HHS are aged between 40 to 70 years old. As there are less patients within the dataset who are older than 80 years old, a possible explanation could be that patients only underwent surgery at

this age if their pain and function was particularly poor. We note that a disadvantage of this plot, and all boxplots within this section, is that they do not reflect the longitudinal structure of the dataset. Nevertheless, they can still provide valuable insight into the nature of the dataset.

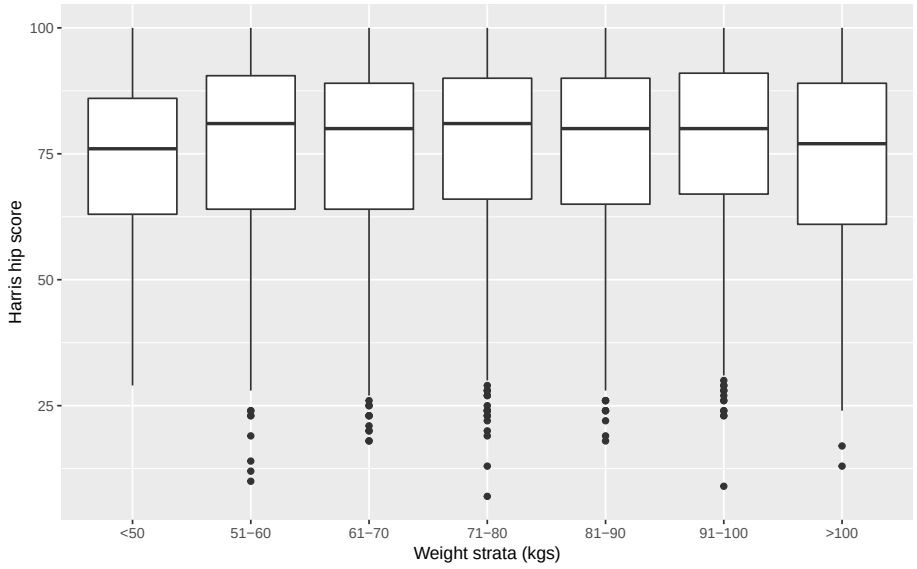

Figure 3.4: Side-by-side boxplot of the Harris hip score for each weight strata (n=1720).

Figure 3.4 shows a side-by-side boxplot of HHS measurements separated by patient weight. Patients with weights heavier than 100kg have a slightly lower median HHS. Likewise, patients with weights lower than 50kg also have a lower median HHS. This supports the clinicians' assessment that the heaviest patients are often associated with poor pain and function recovery after surgery. Overall, there does not appear to be a significant difference between the distributions of HHS for the weight strata between 50kg and 100kg, which all have an approximate median HHS of 80.

Avascular necrosis (AVN), also called osteonecrosis or bone infarction, is death of bone tissue due to interruption of the blood supply [1]. Figure 3.5 shows side-by-side boxplots of the HHS for patients with and without AVN, where 1 indicates the patient does not

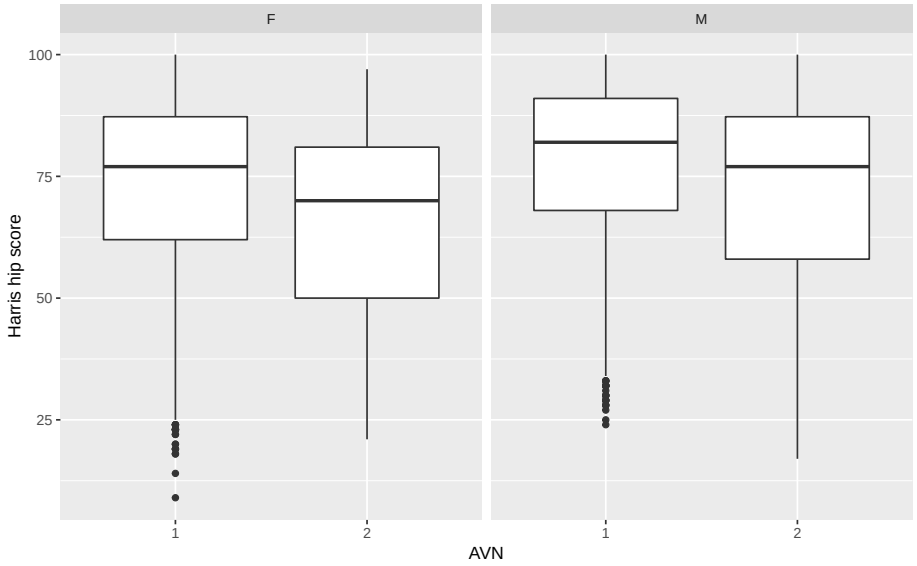

Figure 3.5: Side-by-side boxplot of the Harris hip score for patients with and without AVN, where 1 indicates the patient does not suffer from AVN and 2 indicates the patient does suffer from AVN. Left: patients who are female. Right: patients who are male (n=1720).

suffer from AVN and 2 indicates the patient does suffer from AVN. The left plot is of female patients and the right plot is of male patients. Within this dataset, patients with AVN appear to have a lower HHS than patients who do not suffer from AVN. This is true across both male and female patients. Interestingly, female patients who do not suffer from AVN have a similar median HHS to male patients who do suffer from AVN. This suggests that males are associated with a higher HHS overall, supported by Figure 3.2. To check this possibility, we look at the overall mean of HHSs for each gender. It shows that males have a mean HHS of 77.6 (sd = 17.2) while females have a mean HHS of 73.7 (sd = 18.1), where the p-value for the t-test comparing male and female HHSs is < 0.0001. This further confirms that males are associated with a higher HHS.

The number of comorbidities that a patient suffers from is of particular interest within this study. It is hypothesised that suffering from more comorbidities is associated with a

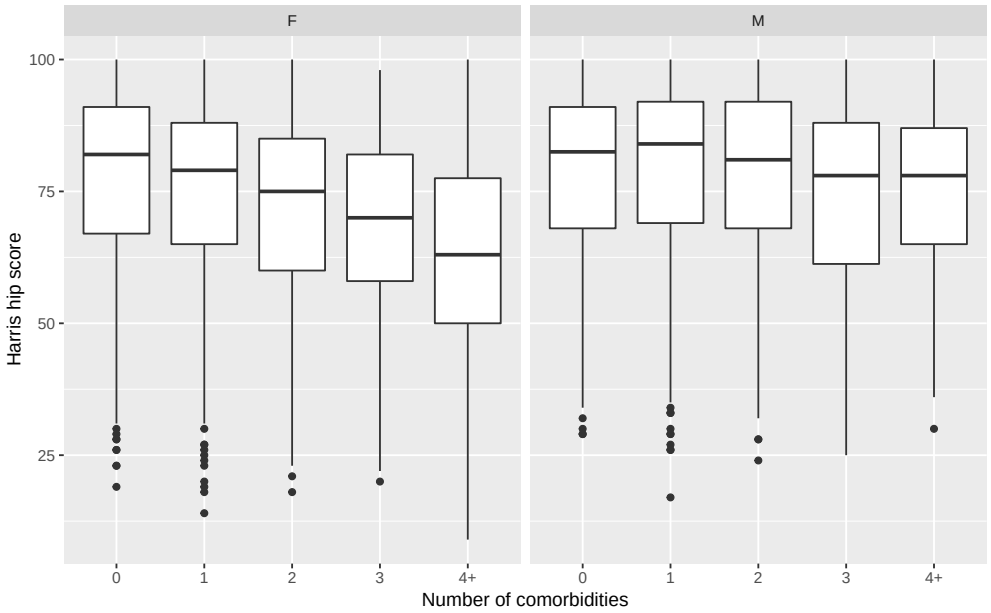

Figure 3.6: Side-by-side boxplot of the Harris hip score for patients with a differing number of comorbidities. Left: patients who are female. Right: patients who are male (n=1720).

lower HHS. Figure 3.6 compares the distribution of HHS with the number of comorbidities that a patient suffers from. The left plot is of female patients and the right plot is of male patients. The side-by-side boxplot of HHS for females shows a consistent decrease in the median HHS as the number of comorbidities increases. This supports the hypothesis that suffering from less comorbidities increases your association with good pain and function recovery after THR surgery. In contrast, the boxplots of HHS for male patients shows a less consistent, but still slight, decrease in HHS as the number of comorbidities increases. An exception to this is the median HHS for males with one comorbidity, which is greater than the median HHS for male patients with no comorbidities. Overall, Figure 3.6 shows that, particularly for females, there is a potential relationship between the number of comorbidities that a patient suffers from and their HHS after surgery. For males this relationship is less distinct.

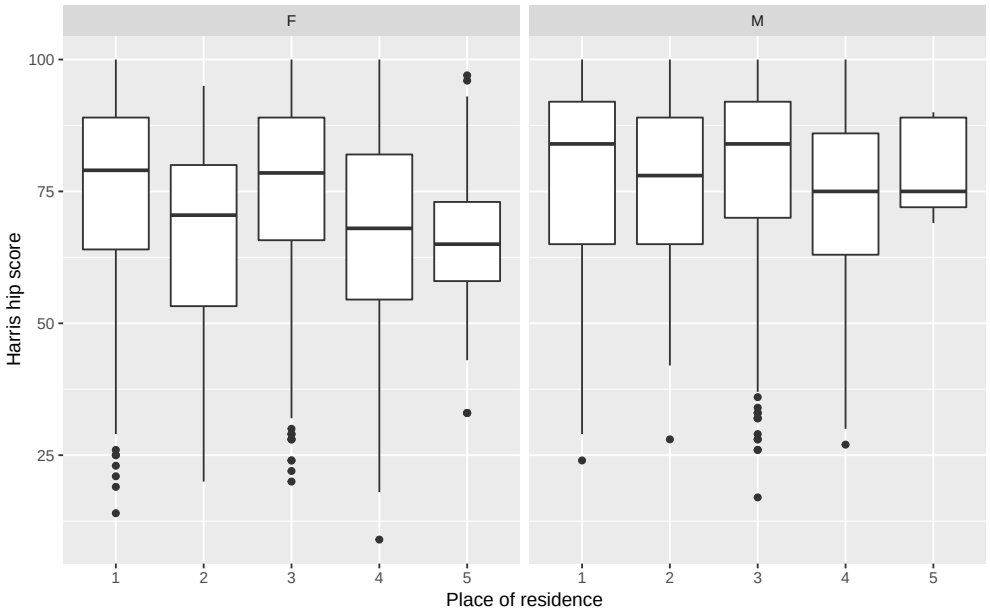

Figure 3.7: Side-by-side boxplot of the Harris hip score for patients separated by their place of residence: 1 indicates the patient lives at home alone and is independent, 2 indicates they live at home alone and with external support, 3 indicates they live at home with others and are independent, 4 indicates they live at home with others and with support and 5 indicates they live in another situation. Left: patients who are female. Right: patients who are male (n=1720).

Figure 3.7 shows a side-by-side boxplot of HHS measurements separated by a patient's place of residence. Within this variable, 1 indicates the patient lives at home alone and is independent, 2 indicates they live at home alone and with external support, 3 indicates they live at home with others and are independent, 4 indicates they live at home with others and with support and 5 indicates they live in another situation (e.g. nursing home). The left plot is of female patients and the right plot is of male patients. It is clear that patients who require external support (levels 2 and 4) have lower median HHSs. Likewise patients whose place of residence is recorded as 'other', such as nursing homes, are also associated with lower median HHSs for both males and females. Again, females appear to have a lower HHS than males across each different option for place of residence.

The results of this bivariate analysis identify some of the variables that may be predictive of HHS after surgery, such as age, gender and number of comorbidities that a patient suffers from. Therefore, using the variables identified in this section, we can now build models that predict HHS.

3.2 Linear Modelling

Linear models allow us to describe a continuous response variable, in this case the HHS, as a function of predictor variables. This description can then help us make pain and function predictions for THR patients. Firstly, we fit a linear model with independent, homoscedastic residual errors to the HHS from the THR dataset. In view of the structure of this data, and of the results shown in the exploratory analysis presented earlier in this chapter, the assumptions of independence and homoscedasticity of the HHS measurements are not appropriate. The structure of the data will be properly taken into account, using more advanced linear models, in Chapter 4. Nevertheless, the results of this linear modelling are still outlined below as an important part of the analysis that follows.

Linear, or fixed effects, models make inferences about the population from which the subjects of the dataset are drawn, in this case the patients. However, these models do not account for the correlation within patient measurements. The linear model can be written as:

$$\begin{aligned} \mathbf{R}_i &= \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \\ \boldsymbol{\epsilon}_i &\sim N(\mathbf{0}, \boldsymbol{\sigma}\boldsymbol{I}), \end{aligned} \tag{3.1}$$

where $\mathbf{R}_i$ is the vector of observed HHSs for patient i, $\boldsymbol{\beta}$ is the fixed-effects parameters for the covariates used in constructing the design matrix $\mathbf{X}_i$, and $\boldsymbol{\epsilon}_i$ are independent,

normally distributed error terms. Therefore, we consider the following model for HHS:

$$\begin{aligned} HHS_{ij} =& \beta_0 + \beta_1 \text{Sex}_i + \beta_2 \text{Age}_i + \beta_3 \text{FollowupTime}_{ij} + \beta_4 \text{FollowupTime}_{ij}^2 + \\ & \beta_5 \text{Side}_i + \beta_6 \text{Weight}_i + \beta_7 \text{Height}_i + \beta_8 \text{ImplantBrand}_i + \\ & \beta_9 \text{Osteoarthritis}_i + \beta_{10} \text{RheumatoidArthritis}_i + \beta_{11} \text{AVN}_i + \\ & \beta_{12} \text{PlaceOfResidence}_i + \beta_{13} \text{Paracetamol}_i + \\ & \beta_{14} \text{NSAIDSNonselectCoxInhibitor}_i + \beta_{15} \text{NSAIDSCox2Inhibitor}_i + \\ & \beta_{16} \text{OpioidStrong}_i + \beta_{17} \text{NumberOfComorbs}_i + \\ & \beta_{18} \text{SmokerStatus}_i + \beta_{19} \text{AlcoholStatus}_i + \epsilon_{ij}. \end{aligned} \tag{3.2}$$

In the model specified in Equation 3.2, HHS_{ij} is the value of HHS measured for patient i at observation number j. In the explanatory (fixed) part of the model, Sex_i is gender, Age_i is the age of patient i at the point of surgery, FollowupTime_{ij} is the follow-up time for observation j, Side_i is the operation side, Weight_i is the mean weight for patient i, Height_i is the mean height for patient i, ImplantBrand_i is the implant brand, Osteoarthritis_i is baseline osteoarthritis status, $\text{RheumatoidArthritis}_i$ is baseline rheumatoid arthritis status, AVN_i is baseline AVN status, $\text{PlaceOfResidence}_i$ is place of residence at point of surgery, NumberOfComorbs_i is the baseline value of the number of comorbidities patient i suffers from, SmokerStatus_i is the baseline smoker status and AlcoholStatus_i is the baseline alcohol status. We note that Paracetamol_{ij}, $\text{NSAIDSNonselectCoxInhibitor}_{ij}$, OpioidStrong_{ij} and $\text{NSAIDSCox2Inhibitor}_{ij}$ are the status of specific medications that are defined in Appendix A.3. The β_k, where $k = 1, \ldots, 19$, are the variable specific effects, with β_0 denoting the intercept.

The variables Sex, Side, ImplantBrand, Osteoarthritis, RheumatoidArthritis, AVN, PlaceOfResidence, Paracetamol, NSAIDSNonselectCoxInhibitor, NSAIDSCox2Inhibitor

and OpioidStrong are all nominal categorical variables. All of these nominal categorical variables have two levels except for ImplantBrand, which has three levels, and PlaceOfResidence, which has five. SmokerStatus, AlcoholStatus, NumberOfComorbs are ordinal categorical variables with three, three and five levels respectively. Age, FollowupTime, Weight and Height are all continuous quantitative variables.

The random part of the model shown in Equation 3.2 is the residual random error, ϵ_{ij}. We assume that errors for each observation are independent and normally distributed with mean 0 and constant variance σ^2. Obviously these assumptions are not correct due to the longitudinal structure of the dataset. Thus, this analysis is only an illustration of the nature of the dataset.

To further simplify the model shown in Equation 3.2 we test the hypothesis of whether there is an effect for each variable using analysis of variance with a backward algorithm and an F-test p-value as the heuristic. This performs F-tests for continuous covariates and groups of contrasts for factor variables included in the model. Using a step-wise process we remove the least significant variable until the p-value of the F-test for each variable within the model is statistically significant at the 5% level. The resulting simplified model determined through this step-wise process is shown in Equation 3.3.

$$\begin{aligned} HHS_{ij} =& \beta_0 + \beta_1 \text{Sex}_i + \beta_2 \text{Age}_i + \beta_3 \text{FollowupTime}_{ij} + \beta_4 \text{FollowupTime}_{ij}^2 + \\ & \beta_5 \text{Weight}_i + \beta_6 \text{Osteoarthritis}_i + \beta_7 \text{AVN}_i + \beta_8 \text{PlaceOfResidence}_i + \\ & \beta_9 \text{OpioidStrong}_i + \beta_{10} \text{NumberOfComorbs}_i + \beta_{11} \text{SmokerStatus}_i + \\ & \beta_{12} \text{AlcoholStatus}_i + \epsilon_{ij}. \end{aligned} \tag{3.3}$$

Each nominal categorical and ordinal categorical variable within the model has a reference level that we need to acknowledge to make appropriate inferences about the

estimated coefficients. The reference level for the variables Osteoarthritis and AVN is not suffering from these diseases. The reference level for OpioidStrong, SmokerStatus and AlcoholStatus is not taking any opioids, not smoking and not drinking, respectively. The reference level for Gender is female, for PlaceOfResidence is independent and home alone and for NumberOfComorbs is suffering from no comorbidities. This means that, for example, the coefficient for Gender, β_1, indicates the mean difference of the HHS between males and females, adjusted for all other covariates in the model. This is detailed in Appendix A.1 and A.3.

Table 3.1 shows the parameters estimates for the optimal model shown in Equation 3.3. The positive estimates for the coefficients indicate a favourable effect on HHS. For example, being male is associated with a higher, and therefore better, HHS, adjusted on the other covariates (we note that each covariate effect has to be interpreted as being adjusted on the other covariates, though this is not always stated). In contrast, the negative estimated coefficient relating to the number of comorbidities a patient suffers from, indicates a worse HHS for patients with a large number of comorbidities. However, as was discussed earlier, the model does not take into account the correlations within patient observations. Thus it should not be used as a basis for inference.

The misspecification of this model is reflected in Figure 3.8, which is a side-by-side boxplot of the studentized residuals for ten randomly selected patients. A studentized residual is the residual divided by an estimated of its standard deviation. The boxplots of the residuals for the model (Equation 3.2) illustrates the basic problem with ignoring the correlation within patients measurements. These correlations are incorporated into the residuals, as there is heterogeneity in the residuals. This leads to an inflated estimate of the within patient variability [11].

We also discovered through this process that the response variable, HHS, requires a

	Parameter	Estimate	SE	p-value
Fixed Effect Estimates	Intercept (mean change in HHS at 0 months)	84.97	3.75	<.0005
	Slope (months after surgery)	1.03	0.12	<.0005
	Quadratic (months)^2	-0.02	0.003	<.0005
	Sex (if male)	5.23	0.88	<.0005
	Age (yrs)	-0.16	0.04	<.0005
	Weight (kgs)	-0.10	0.02	<.0005
	AVN	-4.39	1.78	0.01
	OA	4.06	1.22	0.0009
	Number of Comorbidities (zero)			
	One	0.40	1.15	0.72
	Two	-2.59	1.22	0.03
	Three	-4.08	1.39	0.003
	Four or more	-7.20	3.89	<.0005
	Place of Residence (home alone & indep)			
	Home alone & external support	-4.44	1.68	0.008
	Home with others & independent	-1.00	0.99	0.31
	Home with others & support	-7.40	1.25	<.0005
	Other	-7.80	2.82	0.005
	Smoker (non-smoker)			
	Moderate	-5.19	1.39	<.0005
	Heavy	3.63	3.53	0.30
	Alcohol (non-drinker)			
	Moderate	1.77	0.82	0.03
	Heavy	4.83	3.93	0.22
	σ	16.52		

Table 3.1: Model parameters for linear model HHS (Equation 3.3); Number of observations: 1905. Number of patients: 744.

transformation to better address the assumptions of linear models. HHS is a variable that is bounded between 0 and 100, shown in Figure 2.4 in Chapter 2. To account for this bounding we applied multiple transformations to the variable HHS (e.g. a log transformation, which did not work). The logit transformation was chosen as the most

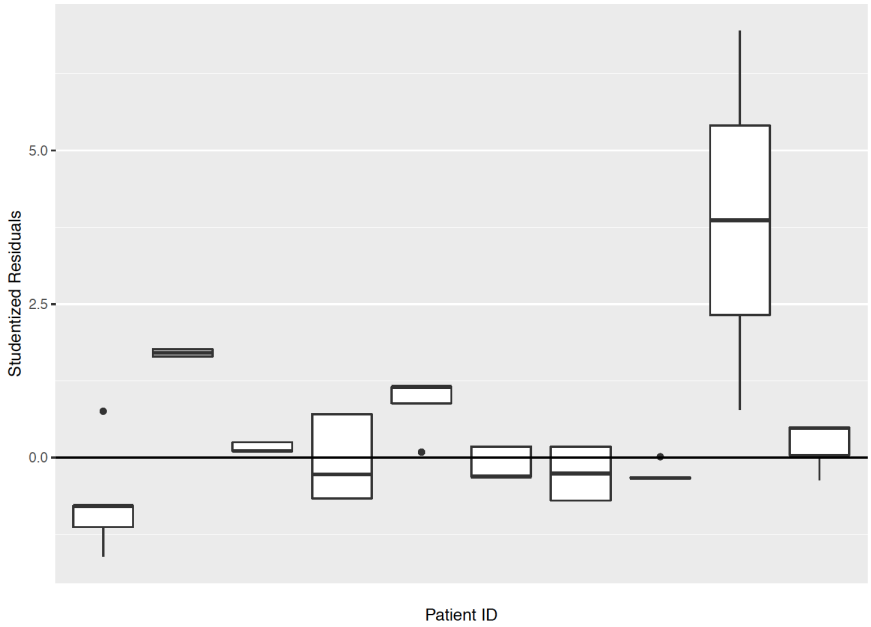

Figure 3.8: Side-by-side boxplot of the studentized residuals for ten randomly selected patients (n=10).

appropriate transformation as it increased the normality of the residuals and converted the HHS to lie between $-\infty$ and ∞ instead of 0 and 100. The logit transformation is outlined in Equation 3.4. We added a 0.1 offset to account for HHS measurements that are 100. This offset stops these observations from being undefined under transformation.

$$\text{logit}HHS = \log\left(\frac{HHS + 0.1}{100.1 - HHS}\right) \tag{3.4}$$

Figure 3.9 displays a histogram of the transformed HHS measurements, using the logit transformation (Equation 3.4). We see that the transformed HHS measurements that are approximately 6.9 are outliers in this histogram, while the rest of the measurements are fairly normally distributed. These measurements of 6.9 correspond to a HHS measurement of exactly 100. Therefore, we see poor behaviour from HHS measurements of 100, which

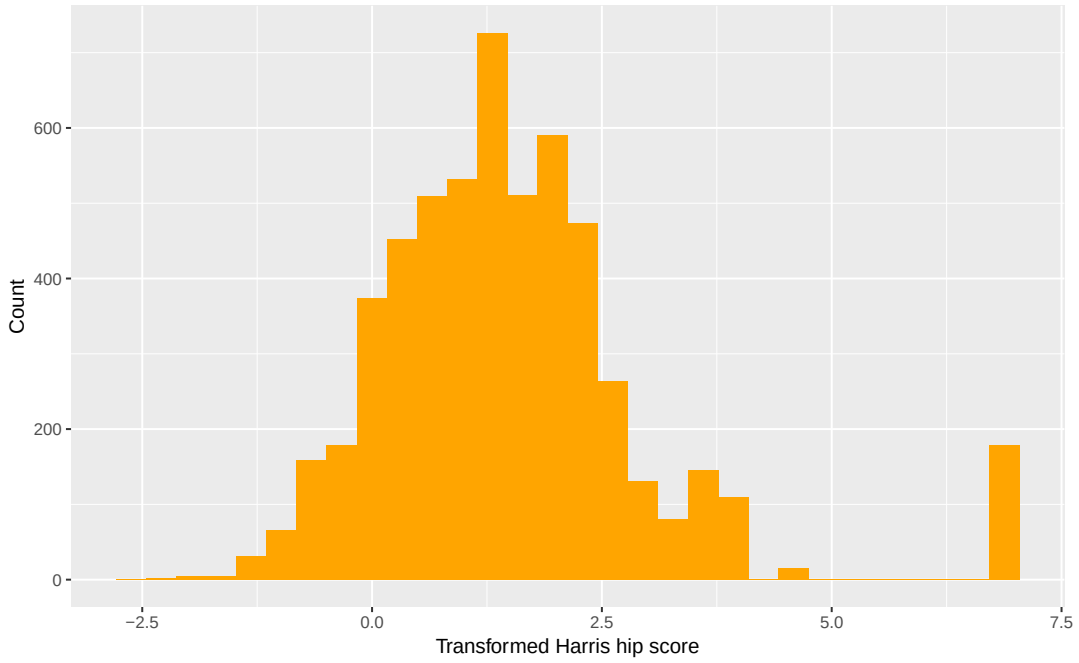

Figure 3.9: Histogram of the transformed Harris hip score of patients at each follow-up. The transformation used is the logit transformation outlined in Equation 3.4 (n=1720).

is a problem that will be explored in further analysis. Nevertheless, this transformation does significantly improve the behaviour of the residuals to fit the assumptions of the linear model. Therefore, it will be applied in further analysis.

Now that we have a greater understanding of the relationship between predictor variables and the HHS, as well as an appropriate transformation for the response variable, we progress beyond a linear model to a mixed-effects model. A model, that in contrast, appropriately takes into account the repeated measures structure of the dataset.

Chapter 4

Mixed-Effects Modelling

An alternative approach to the modelling of the total hip replacement (THR) dataset, which accounts for the correlation between the observations, is to use linear mixed-effects models. This method will be defined and explored in this chapter. Using this approach, the hierarchical structure of the data is addressed through the inclusion of random effects. These random effects describe the contribution of the variability at different levels of the hierarchy within the dataset. For example, random effects can describe the contribution of the Harris hip score (HHS) variability at an individual patient level as well as total variability across all patients.

In this chapter we begin by considering a random intercept model and a random intercept and slope model. We also address the use of restricted maximum likelihood (REML) or maximum likelihood (ML) estimation. A summary of the analyses is then provided. Following this, methods of HHS prediction are explored to help determine which prediction methods are optimal depending on the amount of patient information that is available.

The linear mixed-effects model is such that it combines both random and fixed effects.

It can be written as,

$$\begin{aligned} &\mathbf{R}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \\ &\mathbf{b}_i \sim \mathbf{N}(\mathbf{0}, \mathbf{D}), \\ &\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma_i}), \\ &\mathbf{b_1}, \ldots, \mathbf{b_N}, \boldsymbol{\epsilon_1}, \ldots, \boldsymbol{\epsilon_N} \quad \text{are independent,} \end{aligned} \tag{4.1}$$

where $\mathbf{R}_i$ is the vector of continuous responses, $\mathbf{X}_i\boldsymbol{\beta}$ is the fixed component and $\mathbf{Z}_i\mathbf{b}_i$ is the random component, where i denotes the subject, in this case the patient. The vector of the residual errors for patient i is $\boldsymbol{\epsilon}_i$. There is no index for the vector of parameters $\boldsymbol{\beta}$ as it is the same for all patients. It is the fixed-effects parameters for the covariates used in constructing the design matrix $\mathbf{X}_i$. Similarly $\mathbf{Z}_i$ and $\mathbf{b}_i$ are the matrix of covariates and the corresponding vector of random effects for q covariates,

$$\mathbf{Z}_i = \begin{bmatrix} z_{i1}^{(1)} & z_{i1}^{(2)} & \ldots & z_{i1}^{(q)} \\ \vdots & \vdots & \vdots & \\ z_{in_i}^{(1)} & z_{in_i}^{(2)} & \ldots & z_{in_i}^{(q)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{z}_i^{(1)} & \boldsymbol{z}_i^{(2)} & \ldots & \boldsymbol{z}_i^{(q)} \end{bmatrix}, \quad \mathbf{b}_i = \begin{bmatrix} b_{i1} \\ \vdots \\ b_{iq} \end{bmatrix}.$$

This representation, shown in Equation 4.1, is in the general form. A more specific form of this model could be the random intercept mixed-effects model. Where i is the subject number, n_i is the number of observations for that subject, j is the number of explanatory variables included and R_{i1} is the HHS for the i-th patient's first follow-up.

The random intercept model for the i-th patient can be written as

$$\begin{bmatrix} R_{i1} \\ R_{i2} \\ \vdots \\ R_{in_i} \end{bmatrix} = \begin{bmatrix} 1 & age_i & \dots & time_{i1} \\ 1 & age_i & \dots & time_{i2} \\ \vdots & \vdots & & \\ 1 & age_i & \dots & time_{in_i} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_j \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} b_i \end{bmatrix} + \begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{in} \end{bmatrix}$$

where *age* is not time dependent, as it is the baseline age, and *time* refers to the follow-up time of each HHS observation. The β_j are the variable specific effects, with α denoting the fixed component of the intercept. The random part of the model is the residual random error, ϵ_{ij} and the random intercept, b_i. We assume that both the random intercepts and the residual random errors are independent and normally distributed, as discussed with Equation 4.1. We note that even though the within patient form of the model appears aliased, across all patients this is not the case. Therefore, using this model the i-th patient's first follow-up can be written as,

$$R_{i1} = \alpha + \beta_1 age_i + \ldots + \beta_j time_{i1} + b_i + \epsilon_{i1},$$

where $\alpha + b_i$ is the random intercept, while the slopes remain fixed. This random intercept model effectively incorporates correlations between observations for each patient. Let's suppose there is a relationship between HHS and follow-up time which is different for each patient. This is when we can include not only random intercepts but also random slopes. In a random slopes and intercepts model both the intercept and the slope have random components for each patient. Or more specifically, the individual random intercept and slope capture the individual deviations from the mean baseline value and

the mean slope, respectively. This can be written as,

$$\begin{bmatrix} R_{i1} \\ R_{i2} \\ \vdots \\ R_{in_i} \end{bmatrix} = \begin{bmatrix} 1 & age_i & \dots & time_{i1} \\ 1 & age_i & \dots & time_{i2} \\ \vdots & \vdots & & \\ 1 & age_i & \dots & time_{in_i} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_j \end{bmatrix} + \begin{bmatrix} 1 & time_{i1} \\ 1 & time_{i2} \\ \vdots & \vdots \\ 1 & time_{in_i} \end{bmatrix} \begin{bmatrix} b_{i0} \\ b_{i1} \end{bmatrix} + \begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{in} \end{bmatrix},$$

where *time*, representing follow-up time, is now included in the random effects design matrix. Therefore, using this model the i-th patient's first follow-up can be written as,

$$R_{i1} = \alpha + \beta_1 age_i + \ldots + \beta_j time_{i1} + b_{i0} + b_{i1} time_{i1} + \epsilon_{i1},$$

where the random intercept is $\alpha + b_{i0}$. The coefficient of the i-th patient's follow-up time is $\beta_j + b_{i1}$, which includes the random slope b_{i1}. We note that correlation between the random intercept and slope will be allowed. The only explanatory variable that will be modelled as potentially having a random coefficient will be the follow-up time associated with patient observations. This is because it is feasible that there is a relationship between HHS and follow-up time that is different for each patient. In Chapter 8 we will also consider the mixed-effects models outlined above, where the response variable R_{ij} is the pain or function score, instead of the HHS.

Common heuristics used for mixed-effects models are Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC). These two terms are a function of the likelihood value (a measure of fit) and the number of parameters in the model (a measure

of complexity). The AIC and BIC are defined as,

$$\begin{aligned} AIC &= -2 \times L(\boldsymbol{\theta}) + 2 \times p, \\ BIC &= -2 \times L(\boldsymbol{\theta}) + 2 \times p \times \ln n^*, \end{aligned} \tag{4.2}$$

where L is the likelihood and p is the number of parameters in $\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ denotes the unknown parameters. Depending on what method we use to perform maximum likelihood estimation we obtain different AIC and BIC calculations. For ML we let $n^* = n$ and for REML, $n^* = n - p$ [33].

REML and ML are two functions that can be used to perform maximum likelihood estimation. The limitation of ML is that the estimator for the variance is biased by a factor of $(n-p)/2$, for p explanatory variables [33]. This is because ML ignores the fact that within mixed-effects models we also estimate the intercept and the slope. REML is used due to its ability to correct the estimator for the variance. The log-likelihood function for REML is written as,

$$L_{REML}(\boldsymbol{\theta}) = |\sum_{i=1}^{n} \boldsymbol{X}_i' \times V_i^{-1} \times \boldsymbol{X_i}|^{-0.5} \times L_{ML}(\boldsymbol{\theta}), \tag{4.3}$$

where $\boldsymbol{X}_i$ is the design matrix for the fixed effects, V_i is the variance of the normally distributed responses and $L_{ML}(\boldsymbol{\theta})$ is the log-likelihood function for ML. Though these two estimators often give similar results they do differ substantially [10]. As we see from Equation 4.2 and 4.3 an AIC calculated using REML is not comparable to an AIC calculated using ML and likewise for BIC. This means that at different stages in the model selection process REML and ML will be applied to determine the optimal fixed and random structure.

4.1 Model Selection

In earlier sections we discussed that although the original dataset contained many variables this was narrowed to 18 explanatory variables, with the aid of the clinicians as shown in Chapter 3. The strategy used for model selection is the top-down strategy [33] and is as follows:

- Begin with a model that contains all explanatory variables. A quadratic term for the variable follow-up time is included, as mentioned in Chapter 2. Due to the large number of explanatory variables no interactions are included.

- Using this model, determine the optimal random components structure. This decreases the risk of the random components having any information that we want in the fixed components part of the model. Since ML cannot be used to compare two models with nested random structures, as the variable estimators are biased, we must use REML estimators. In this step the REML likelihood ratio test, AIC and BIC are used to determine the optimal random components structure, whether to include random intercepts and slopes.

- After determining the optimal random structure we find the optimal fixed structure. To compare models with nested fixed effects and the same random structure, ML estimation must be used. Both the F-statistics and the t-statistics can be used to compare models with differing fixed structures. The least significant variables are removed until all fixed effects are significant at the 5% level.

- Now that we have obtained the optimal random effects and fixed effects structure we have the final model. Lastly, the final mixed-effects model is calculated using REML estimation.

Using this strategy a full model, included all explanatory variables, was fitted. To compare random effects structures a random intercept, random slope and intercept and generalised least squares model were compared. The models were estimated with REML to allow the likelihood ratio test to be applied to help compare the models. The random intercept model is written as,

$$\begin{aligned}\text{logit}(HHS)_{ij} =&\alpha + \beta_1 \text{Sex}_i + \beta_2 \text{Age}_i + \beta_3 \text{FollowupTime}_{ij} + \beta_4 \times \text{FollowupTime}_{ij}^2 + \\ &\beta_5 \text{Side}_i + \beta_6 \text{Weight}_i + \beta_7 \text{Height}_i + \beta_8 \text{ImplantBrand}_i + \\ &\beta_9 \text{Osteoarthritis}_i + \beta_{10} \text{RheumatoidArthritis}_i + \beta_{11} \text{AVN}_i + \\ &\beta_{12} \text{PlaceOfResidence}_i + \beta_{13} \text{Paracetamol}_i + \\ &\beta_{14} \text{NSAIDSNonselectCoxInhibitor}_i + \beta_{15} \text{NSAIDSCox2Inhibitor}_i + \\ &\beta_{16} \text{OpioidStrong}_i + \beta_{17} \text{NumberOfComorbs}_i + \\ &\beta_{18} \text{SmokerStatus}_i + \beta_{19} \text{AlcoholStatus}_i + b_i + \epsilon_{ij},\end{aligned} \tag{4.4}$$

where $\text{logit}(HHS)_{ij}$ is the logit transformed HHS of observation j for the i-th patient. The explanatory variables used in Equation 4.4 have been previously outlined in Chapter 3. The random effects, b_i, are normally distributed: $\text{N}(0, d^2)$. The errors, ϵ_{ij}, are also normally distributed. Again we note that all explanatory variables, excluding follow-up time, are not time-dependent. Therefore, they are all baseline variables or averaged variables, in the case of weight and height. Since the model in Equation 4.4 is a random intercept model it allows for a random shift around the intercept resulting in fitted lines that are parallel to the population fitted line.

Let's suppose that alternatively there is a relationship between HHS and follow-up time which is different for each patient. In this case, we can apply the mixed-effects model with not only a random intercept, but also a random slope. The random intercept

and slope model is specified as,

$$\begin{aligned}\text{logit}(HHS)_{ij} =& \alpha + \beta_1 \text{Sex}_i + \beta_2 \text{Age}_i + \beta_3 \text{FollowupTime}_{ij} + \beta_4 \text{FollowupTime}_{ij}^2 + \\ & \beta_5 \text{Side}_i + \beta_6 \text{Weight}_i + \beta_7 \text{Height}_i + \beta_8 \text{ImplantBrand}_i + \\ & \beta_9 \text{Osteoarthritis}_i + \beta_{10} \text{RheumatoidArthritis}_i + \beta_{11} \text{AVN}_i + \\ & \beta_{12} \text{PlaceOfResidence}_i + \beta_{13} \text{Paracetamol}_i + \\ & \beta_{14} \text{NSAIDSNonselectCoxInhibitor}_i + \beta_{15} \text{NSAIDSCox2Inhibitor}_i + \\ & \beta_{16} \text{OpioidStrong}_i + \beta_{17} \text{NumberOfComorbs}_i + \\ & \beta_{18} \text{SmokerStatus}_i + \beta_{19} \text{AlcoholStatus}_i + b_{i0} + b_{i1} \text{FollowupTime}_{ij} + \epsilon_{ij}\end{aligned} \tag{4.5}$$

where the term b_{i0} is the random intercept and b_{i1} is the random slope for follow-up time. Comparing the two models shown in Equations 4.4 and 4.5, along with a generalised least squares model, allows us to determine the optimal random effects structure.

The AIC values for the random intercept, random intercept and slope, and generalised least squares model are 6288.1, 6260.8 and 6637.5 respectively. The likelihood ratio test comparing the random intercept model and the random intercept and slope model indicates that the model with the random slope and intercept is preferable, with $L = 30.90$ ($df = 34, p < 0.001$). Therefore, the optimal random effects structure is a random slope and intercept, as shown in Equation 4.5.

Next we check whether we have homogeneity of variance for the model shown in Equation 4.5. Figure 4.1 displays diagnostic plots for this model. This includes scatterplots of the residuals against fitted values, and the residuals against the continuous quantitative explanatory variables, such as age and weight. Also included are the side-by-side boxplots of residuals against the explanatory variables, such as gender, sum of comorbidities and smoker status. The fitted against residuals plot shows a distinct linear line in the top right

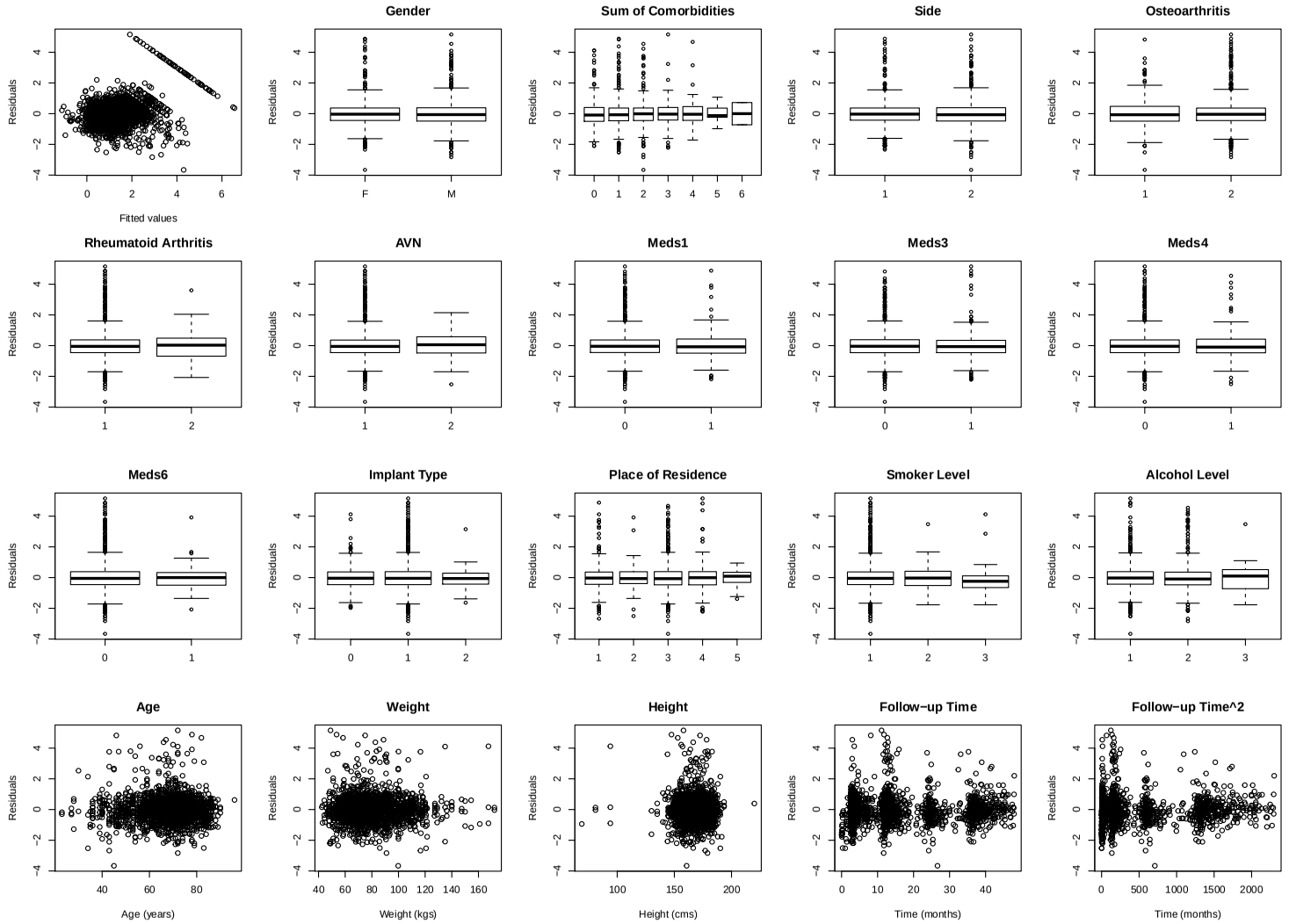

Figure 4.1: Diagnostic plots for the random intercept and slope mixed-effects model defined in Equation 4.5. This includes scatterplots of the residuals against fitted values, and the residuals against the continuous explanatory variables, such as age and weight. Also included are the side-by-side boxplots of residuals against the explanatory variables, such as gender, sum of comorbidities and smoker status (n=744).

of the plot. We can again isolate these residuals as being the residuals for observations of HHS that are 100. Nevertheless, the remainder of the residuals, for all other values of HHS, fit the assumptions of the model. Overall these plots do not show any clear violation of homogeneity.

To calculate the optimal model for the explanatory variables we fit the same model, Equation 4.5, but with ML and use the likelihood ratio test to determine the least significant term. This term is then removed from the model. This process is repeated until all the explanatory variables within the model are significant at the 5% level. The model that is selected is the optimal model and is written as,

$$\begin{aligned}\text{logit}(HHS)_{ij} =& \alpha + \beta_1 \text{Sex}_i + \beta_2 \text{Age}_i + \beta_3 \text{FollowupTime}_{ij} + \\ & \beta_4 \text{FollowupTime}_{ij}^2 + \beta_5 \text{Weight}_i + \beta_6 \text{NumberOfComorbs}_i + \\ & \beta_7 \text{AVN}_i + \beta_8 \text{PlaceOfResidence}_i + \beta_9 \text{SmokerStatus}_i + \\ & b_{i0} + b_{i1} \text{FollowupTime}_{ij} + \epsilon_{ij}.\end{aligned} \tag{4.6}$$

The estimated parameters are calculated using REML and are shown in Table 4.1, including the components of variance estimators. The slope for gender is 0.40, which indicates that the logit HHS for an observation from a male patient is estimated to be 0.40 higher than a female patient, adjusting for all the other covariates in the model. Indicating that being male is associated with less pain and better function after THR surgery. Having two, three or four or more comorbidities, instead of no comorbidities, is associated with a lower HHS. Living home alone with external support is associated with a lower HHS than patients who live independently. Likewise living at home with others with external support or living in a nursing home (part of the category 'other') is associated with a lower HHS than living independently.

Table 4.1 also shows that the slope for the follow-up time is 0.08 and the slope for the quadratic term for follow-up time is -0.002. Therefore, if the follow-up time is between 0 and 40 months then time is associated with a higher HHS. In contrast, if the follow-up time is greater than 40 months than time is associated with a lower HHS. This indicates that more time after surgery is associated with less pain and better function, up until 40 months.

The coefficient for moderate smoker status is -0.42. This means that the logit HHS for a patient who moderately smokes is estimated to be 0.42 lower than a patient who does not smoke at all, adjusting for all the other covariates in the model. This indicates that patients who do not smoke are associated with less pain and better function, after THR surgery, compared to patients who moderately smoke. Patients with a heavier weight are associated with a lower HHS, as are patients who suffer from AVN (avascular necrosis).

These results suggest that to increase a patient's chance of less pain and better function after surgery it may be beneficial to decrease weight and be a non-smoker. In regards to the results for the variable place of residence, this is most likely indicative of a patient's frailty. For example, if a patient is home with others or needs support then they are more likely to be frail than a patient who lives independently. Dependent and frail patients are associated with a lower HHS after THR surgery, compared to patients who live independently.

In relation to the components of variance, shown in Table 4.1, we see that the slope and intercept variance estimates are significantly lower than the within patient residual. The random intercept, b_{i0}, is normally distributed with mean 0 and variance 0.72^2. This is less than the residual term, ϵ_{ij}, which has variance 0.96^2. Using these values we can determine that the correlation between observations from the same patient is $0.72^2/(0.72^2+0.96^2) = 0.36$. Also we see that the correlation between the random intercept and slope estimates

	Parameter	**Estimate**	**p-value**
Fixed Effect Estimates	Intercept (mean change in logit HHS at 0 months)	2.53	<.0005
	Slope (months after surgery)	0.08	<.0005
	Quadratic (months)^2	-0.002	<.0005
	Sex	0.40	<.0005
	Age	-0.01	0.001
	Weight	-0.007	<.0005
	AVN	-0.41	0.005
	Number of Comorbidities		
	One	-0.02	0.89
	Two	-0.25	0.04
	Three	-0.37	0.007
	Four or more	-0.35	0.04
	Place of Residence		
	Home alone & external support	-0.37	0.02
	Home with others & independent	-0.06	0.57
	Home with others & support	-0.56	<.0005
	Other	-0.56	0.05
	Smoker (non-smoker)		
	Moderate	-0.42	0.001
	Heavy	0.26	0.36
		Standard Deviation	
Component of Variance Estimates	Within Patient Residual	0.96	
	Between Patient Residual (intercept)	0.72	
	Between Patient Residual (slope)	0.01	
	corr(b_{i0}, b_{i1})	0.99	

Table 4.1: Model parameters for HHS; Number of observations: 2049. Number of patients: 808.

is high. This is due to there being little variation within the grouping (random) variables.

To further validate the model heterogeneity was examined. Figure 4.2 displays diag-

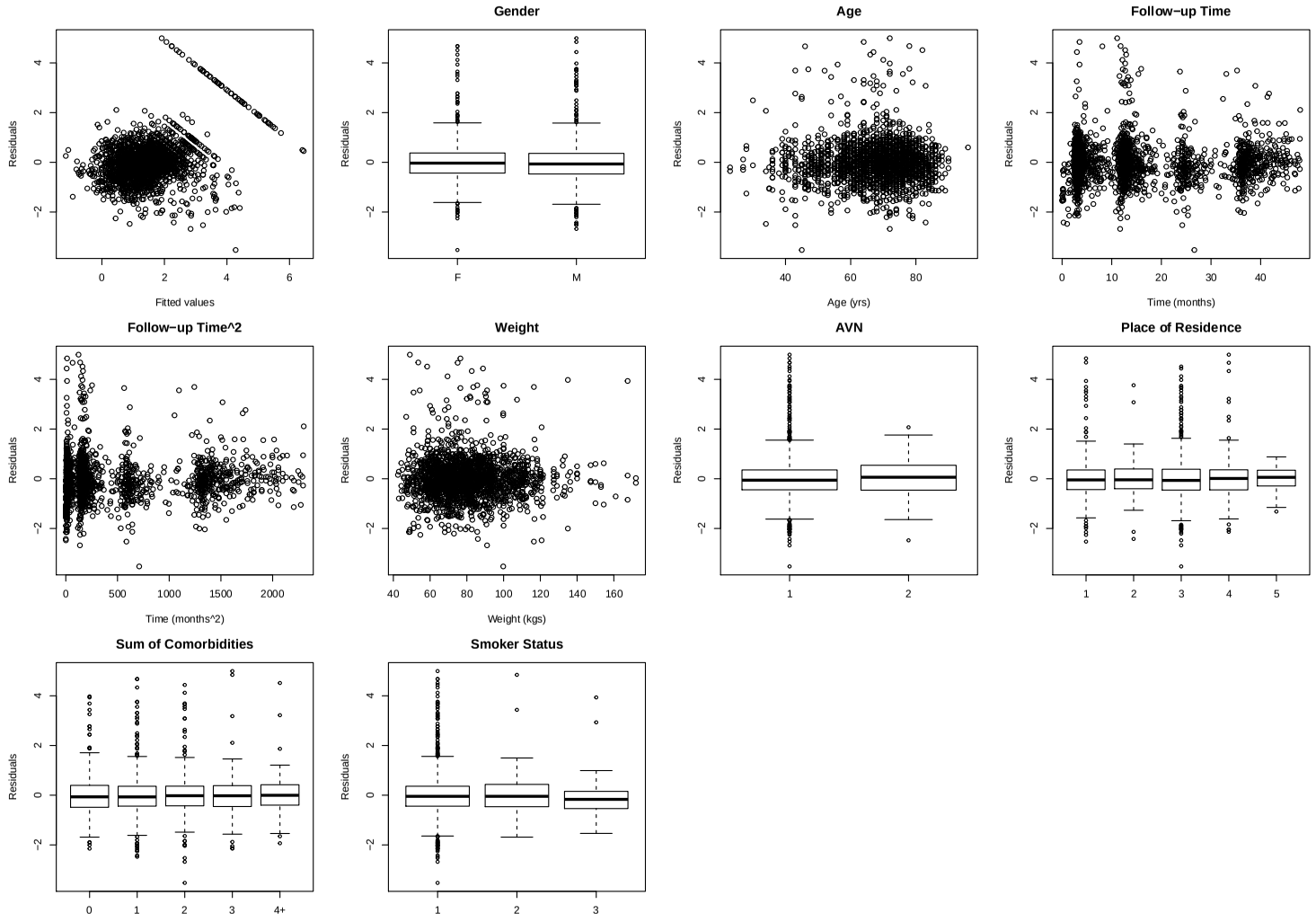

Figure 4.2: Diagnostic plots for the random intercept and slope mixed-effects model defined in Equation 4.6. This includes scatterplots of the residuals against fitted values, and the residuals against the continuous explanatory variables, such as age and weight. Also included are the side-by-side boxplots of residuals against the explanatory variables, such as gender, sum of comorbidities and smoker status (n=808).

nostic plots for the random intercept and slope mixed-effects model defined in Equation 4.6. This includes scatterplots of the residuals against fitted values, and the residuals against the continuous quantitative explanatory variables, such as age and weight. Also included are the side-by-side boxplots of residuals against the explanatory variables, such as gender, sum of comorbidities and smoker status. This figure shows that the assumption of homogeneity is satisfied, as there does not appear to be any clear trends within the residuals of the model.

Our optimal model at this stage contains the nominal categorical variables place of residence, AVN and gender. The ordinal categorical variables within the model are the number of comorbidities and smoking status. The continuous quantitative explanatory variables are the baseline age, baseline weight and follow-up time of a patient. Since there is no clear pattern if the residuals are plotted against follow-up time, then independence is assumed, see Figure 4.2.

The plot of the residuals against fitted values for the model described in Equation 4.6 is shown in Figure 4.3. The blue points are HHS measurements that are 100. The red points are the HHS measurements that are not 100. The darker shaded points indicate areas of higher density while the lighter shade indicates lower density. Therefore, this model fails to effectively predict values about the upper bound of the HHS. Instead the model consistently underestimates these observations as it does not predict any value that is over the HHS upper bound of 100. Nevertheless, the remainder of the residuals are fairly normally distributed. To check further if the assumptions of the model were satisfied we explored the residual distributions for each level of the explanatory variables within the model. We stratify these analyses by sex for further insights into the residual distributions.

Figure 4.4 is a plot of the residuals for each gender and AVN status. 1 indicates not

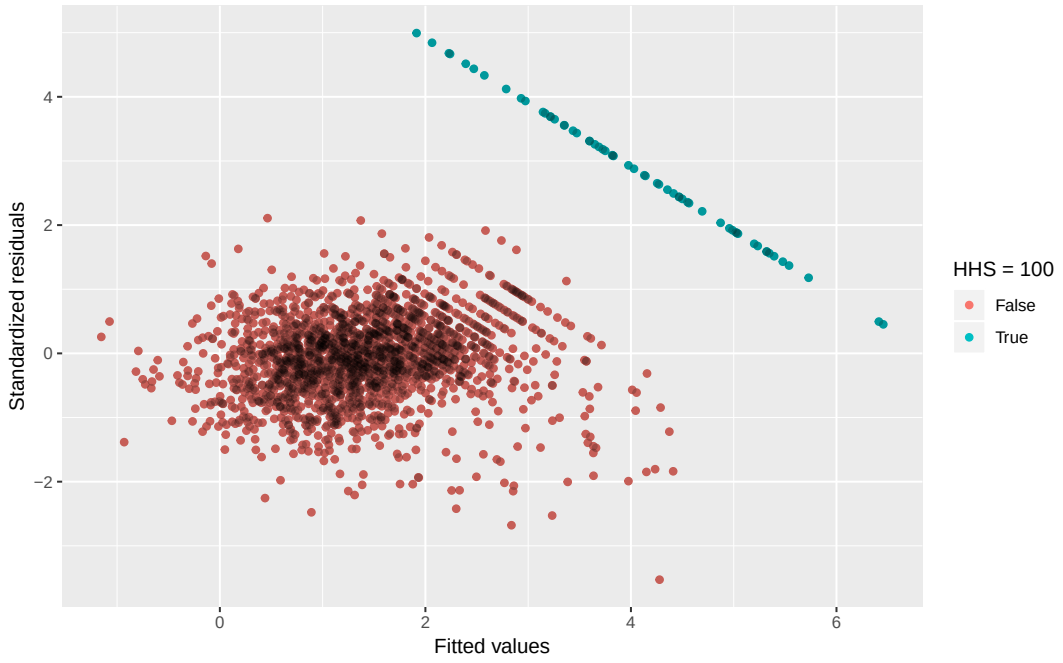

Figure 4.3: Plot of the standardized residuals against fitted values. The blue points are HHS measurements that are 100. The red points are the HHS measurements that are not 100. The darker shaded points indicate areas of higher density while the lighter shade indicates lower density (n=808).

suffering from AVN while 2 indicates that a patients suffers from AVN. M indicates a male patient while F indicates a female patient. The solid lines are LOESS smoothers to show the overall trend of the residuals. This plot allows for the evaluation of the distribution of the conditional Pearson residuals across each follow-up time for both genders and status of AVN. The variability of the residuals does not seem to vary greatly. Though the plot does reveal some outliers, *i.e.*, residuals larger than four for patients without AVN. These outliers all correspond to HHS measurements of 100, which as we discussed previously, do not behave well in this model. Therefore, these outliers are present in both genders and across all time points up until 24 months when the number of observations begins to decrease across the dataset.

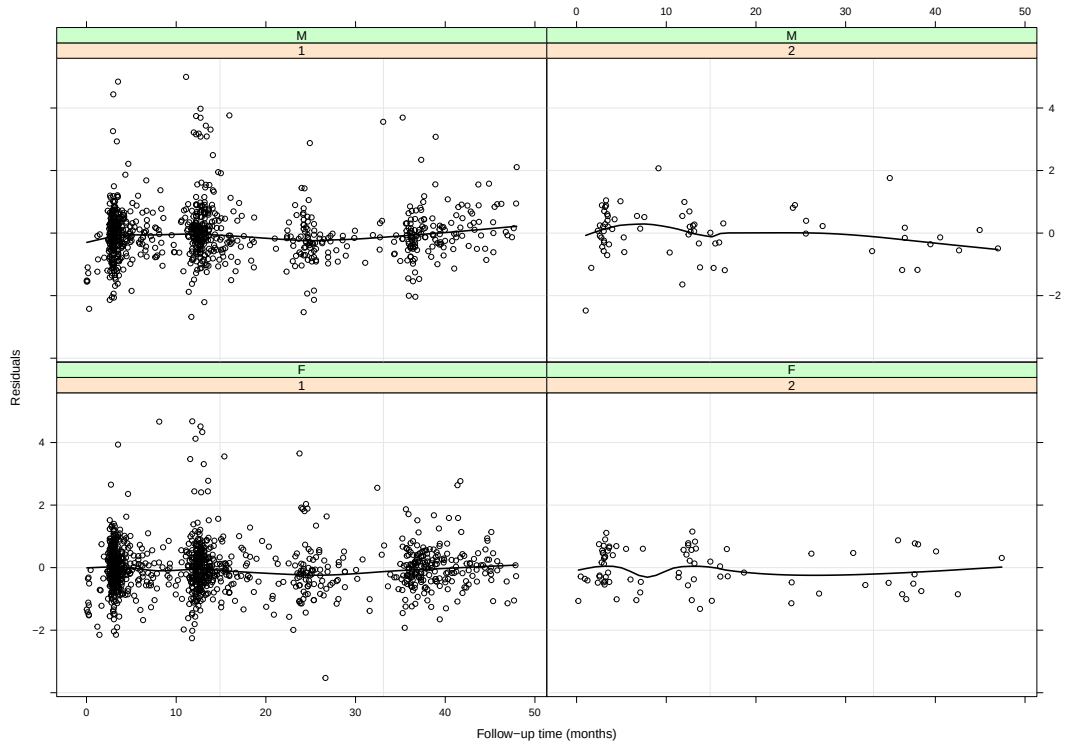

Figure 4.4: Plot of the residuals against time separated by gender and AVN status. 1 indicates not suffering from AVN while 2 indicates that a patients suffers from AVN. M indicates a male patient while F indicates a female patient. The solid lines are LOESS smoothers to show the overall trend of the residuals (n=808).

We also note in Figure 4.4 that we see a clustering of residuals around 3, 12 and 24 months. This is because the clinic normally schedules a follow-up after surgery at 3, 12 and 24 months. Therefore, we would expect to see this clustering in the residuals when they are plotted against time.

Figure 4.5 is a plot of the residuals against time separated by gender and smoker status. 1 indicates non-smoker, 2 indicates light smoker and 3 indicates heavy smoker. The solid lines are LOESS smoothers to show the overall trend of the residuals. This plot reveals outliers with residuals greater than 4, which again corresponds to HHS measurements of 100. There appears to be some non-linearity in the residuals for heavy smokers across

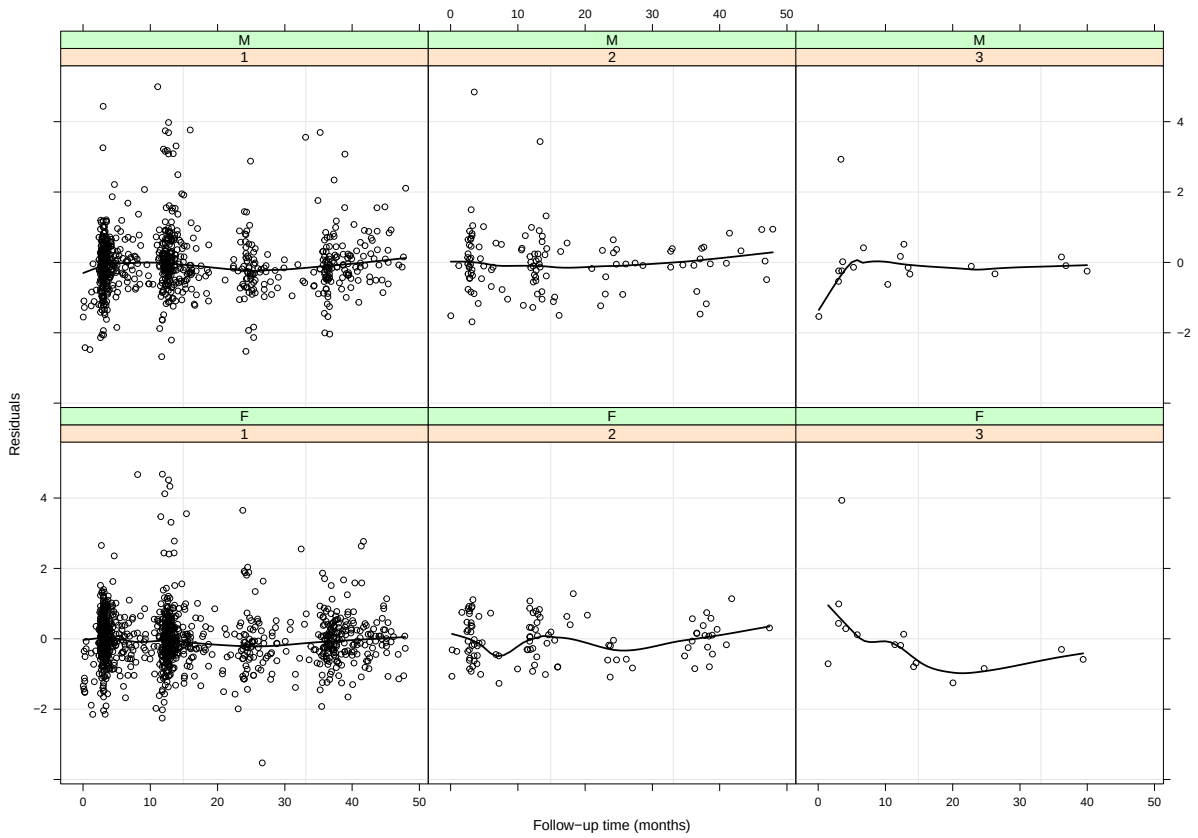

Figure 4.5: Plot of the residuals against time separated by gender and smoker status. 1 indicates non-smoker, 2 indicates light smoker and 3 indicates heavy smoker. The solid lines are LOESS smoothers to show the overall trend of the residuals (n=808).

both genders. This is most likely reflective of the lack of observations within this category. Nevertheless, the residuals appear to be fairly randomly spread across both genders and each smoking status.

Another variable that was of significance within the optimal mixed-effects model was place of residence. Figure 4.6 shows a plot of the residuals against time separated by gender and place of residence. 1 indicates home alone & independent, 2 indicates home alone & external support, 3 indicates home with others & independent, 4 indicates home with others & support and 5 indicates 'other' living situation. The solid lines are LOESS

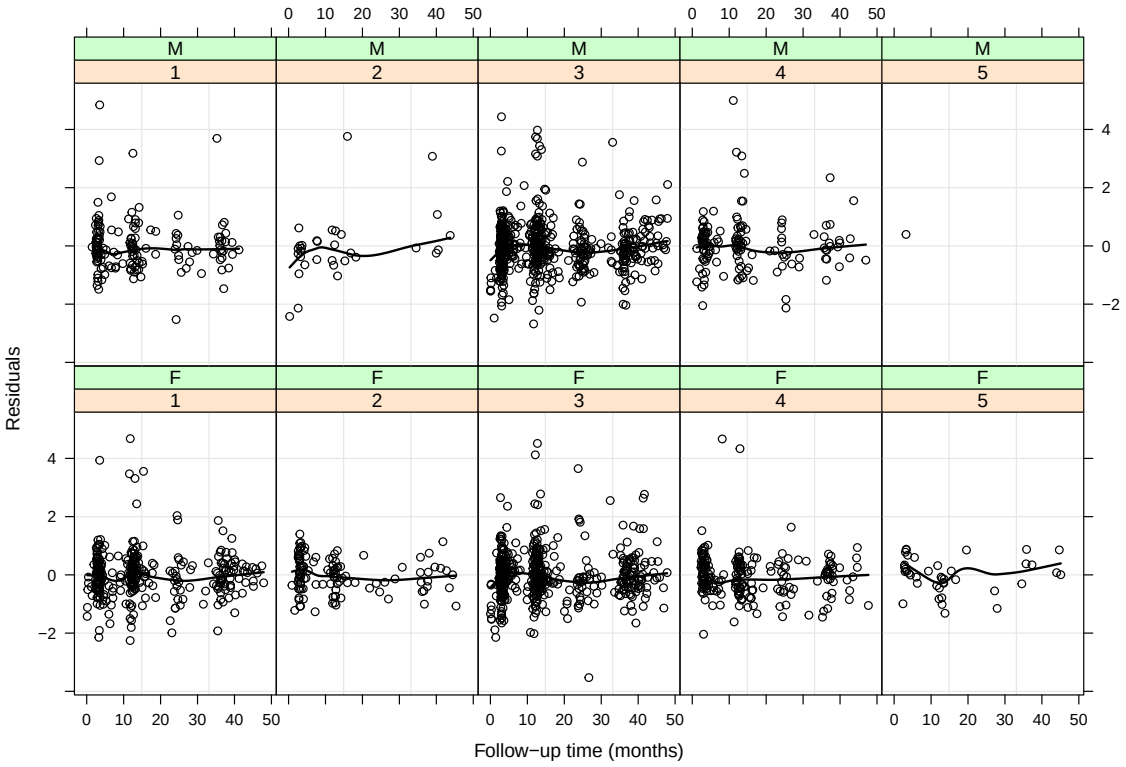

Figure 4.6: Plot of the residuals against time separated by gender and place of residence. 1 indicates home alone & independent, 2 indicates home alone & external support, 3 indicates home with others & independent, 4 indicates home with others & support and 5 indicates other living situation. The solid lines are LOESS smoothers to show the overall trend of the residuals (n=808).

smoothers to show the overall trend of the residuals. Levels 1, 3 and 4 of the place of residence variable show no clear patterns in the residuals. In contrast, we see that level 2, home alone & external support, shows some linear trend for the male patients. But this is also reflective of the lack of observations within these categories, as there are less males than females who are home alone & have external support. Patients who are recorded as having a place of residence that is not specified, those listed as 'other', are few, particularly for males.

Figure 4.7 displays a plot of the residuals versus follow-up time for each sex and

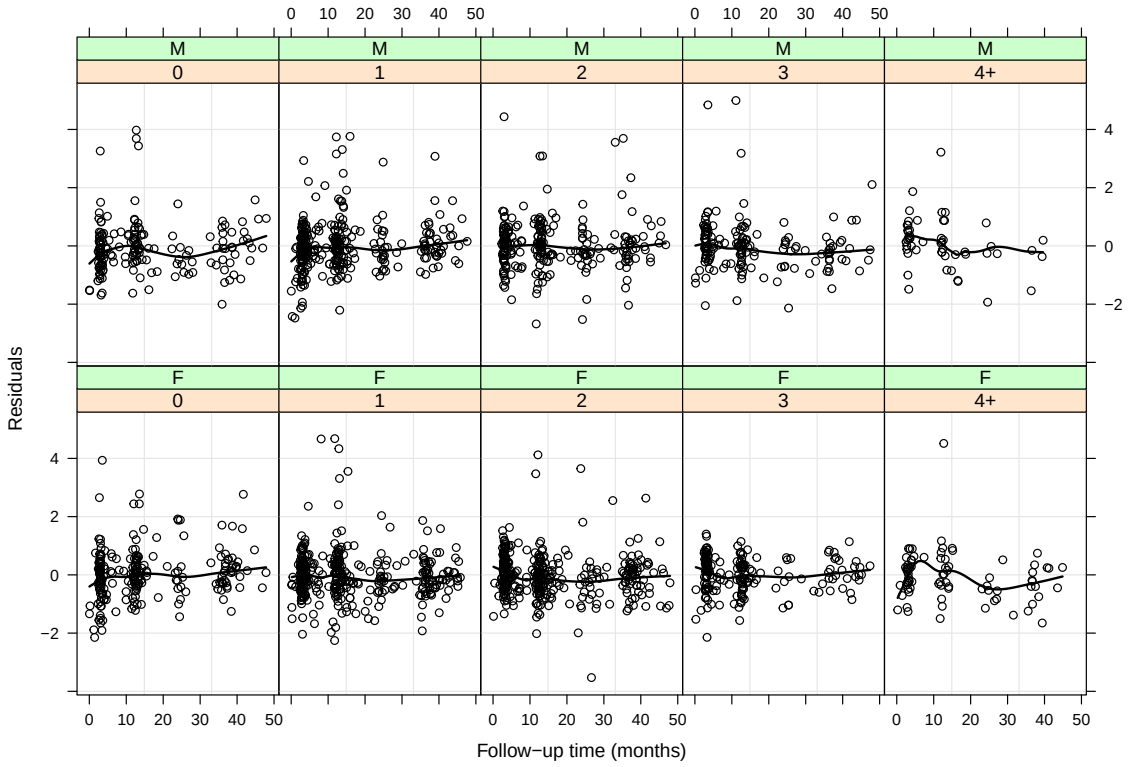

Figure 4.7: Residuals versus follow-up time for each sex and number of comorbidities combination. Where 0 indicates not having any comorbidities, 1 indicates have one comorbidity, etc. A LOESS smoother with a span of 0.5 was fitted to aid visual interpretation (n=808).

number of comorbidities combination. Where 0 indicates not having any comorbidities, 1 indicates have one comorbidity, etc. A LOESS smoother with a span of 0.5 was fitted to aid visual interpretation. We see that across both genders, and different numbers of comorbidities, patient residuals do not show any real pattern. Outliers are still present within the residuals, which are again the HHS measurements of 100. The residuals for females that have four or more comorbidities may potentially not be linear but this is most likely due to the lack of observations, particularly between 3 and 12 months, which may create the non-linear pattern.

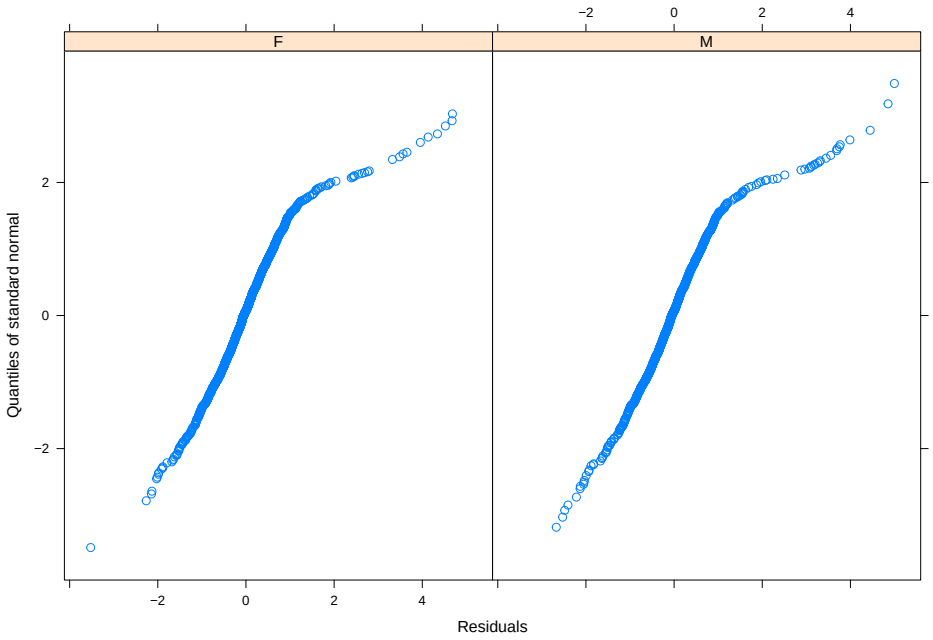

Figure 4.8: Normal Q-Q plots of the conditional Pearson residuals for each gender for the model in Equation 4.6 (n=808).

To further investigate the assumptions of the optimal mixed-effects model normal Q-Q plots were explored. Figure 4.8 shows the normal Q-Q plot of the conditional Pearson residuals separated by gender. The normal Q-Q plots do show some deviations from a linear trend at the right upper tail. This is a reflection of the poor behaviour of the HHS measurements of 100, as discussed before. These observations create a strong deviation from the linear trend we expect to see in a normal Q-Q plot.

We can also look at the normal Q-Q plot of the predicted random effects (random intercept and slope). The effects are estimated using BLUPs (best linear unbiased predictors). These are the conditional expectations of the random effects, given the observed responses and will be discussed further in Section 4.2. The resulting Q-Q plot is shown in Figure 4.9 and again displays a non-linear tail while the majority of the points lie on the linear line. Again these points, on further investigation, are related to HHS measure-

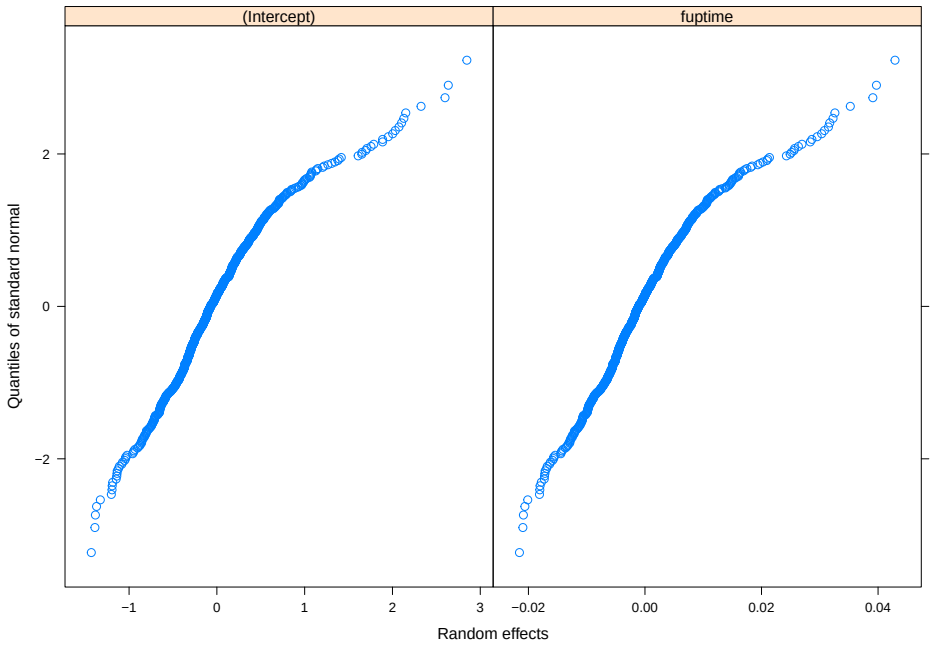

Figure 4.9: The normal Q-Q plot of the predicted random intercepts and slopes for the model in Equation 4.6. Left: normal Q-Q plot of the predicted random intercepts; Right: normal Q-Q plot of the predicted random slopes (n=808).

ments of 100. Nevertheless this plot may not necessarily reflect the true distribution of the random effects. This is because the observed distributions of the conditional expectations of the random effects, given the observed response, does not necessarily reflect the true distribution of the random effects [11]. However they are still helpful at identifying outlying values that warrant further investigation.

Many plots have been shown regarding the relationship between explanatory variables and the residuals. In most cases, we assume independence because we cannot see a clear pattern when the residuals are plotted against follow-up time. However, in some cases there appeared to be a potential pattern in the residuals. Though we suspect this is due to small observation numbers in specific groups we still want to investigate this further.

To do this we fit an additive mixed model, which is written as,

$$\begin{aligned}\text{logit}(HHS)_{ij} =& \alpha + \beta_1 \text{Sex}_i + \beta_2 \text{Age}_i + f(\text{FollowupTime}_{ij}) + \\ & f(\text{FollowupTime}_{ij}^2) + \beta_5 \text{Weight}_i + \beta_6 \text{NumberOfComorbs}_i + \\ & \beta_7 \text{AVN}_i + \beta_8 \text{PlaceOfResidence}_i + \beta_9 \text{SmokerStatus}_i + \\ & b_{i0} + b_{i1} \text{FollowupTime}_{ij} + \epsilon_{ij}.\end{aligned} \tag{4.7}$$

where the term $f(\text{FollowupTime}_{ij})$ and $f(\text{FollowupTime}_{ij}^2)$ are now smoothing splines. It is expected that the shape of the smoother is a straight line if the linear mixed-effects model is appropriate.

Figure 4.10 displays the estimated smoother for the term $f(\text{FollowupTime}_{ij})$ within the additive mixed model shown in Equation 4.7. The solid line is the estimated smoother and the dotted lines are the 95% point-wise confidence bands. The horizontal axis shows the follow-up time in months and the vertical axis the contribution of the smoother to the fitted values.

This figure shows that the relationship between follow-up time and the contribution of the smoother to the fitted values is linear. The estimated regression parameters for the additive mixed model are similar to the linear mixed-effects model. The AIC for this model is 8522 while the AIC for the previous model is 8605. The smoother is significant and has one degree of freedom for the $f(\text{FollowupTime}_{ij})$ term. The degrees of freedom for the $f(\text{FollowupTime}_{ij}^2)$ term is slightly more at 3.5. As a result, there appears to be no pattern in the residuals and this mixed-effects model. This supports the mixed-effects model in Equation 4.6.

We will now look closer at this model's ability to predict HHS observations for patients with varying levels of information available. For example, for patients who have some HHS

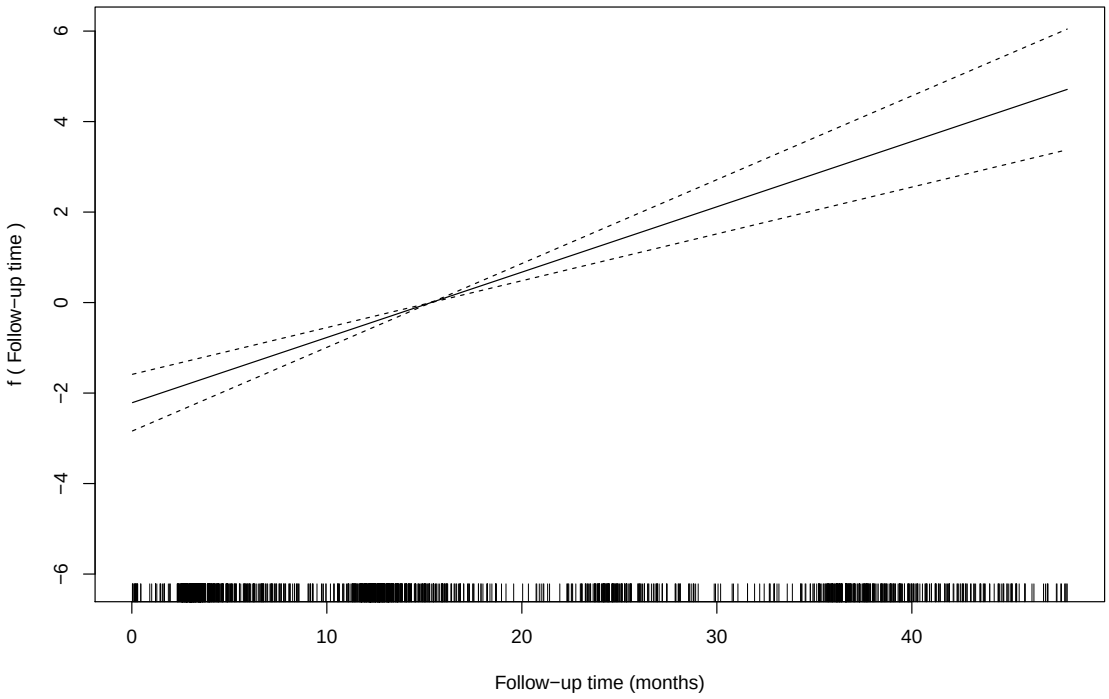

Figure 4.10: Estimated smoother for the additive mixed model. The solid line is the estimated smoother and the dotted lines are the 95% point-wise confidence bands. The horizontal axis shows the follow-up time in months and the vertical axis the contribution of the smoother to the fitted values (n=808).

observations, and patients with only baseline information.

4.2 Prediction and Cross-Validation

In this section, we discuss prediction and cross-validation methods for the optimal mixed-effects model outlined in Equation 4.6. Prediction is important for two reasons. Firstly, it helps us predict what a patient's HHS might be at a certain point after surgery. This includes predicting future observations for patients who already have some measured HHS observations. This also includes predicting the HHS of new patients, who are not within

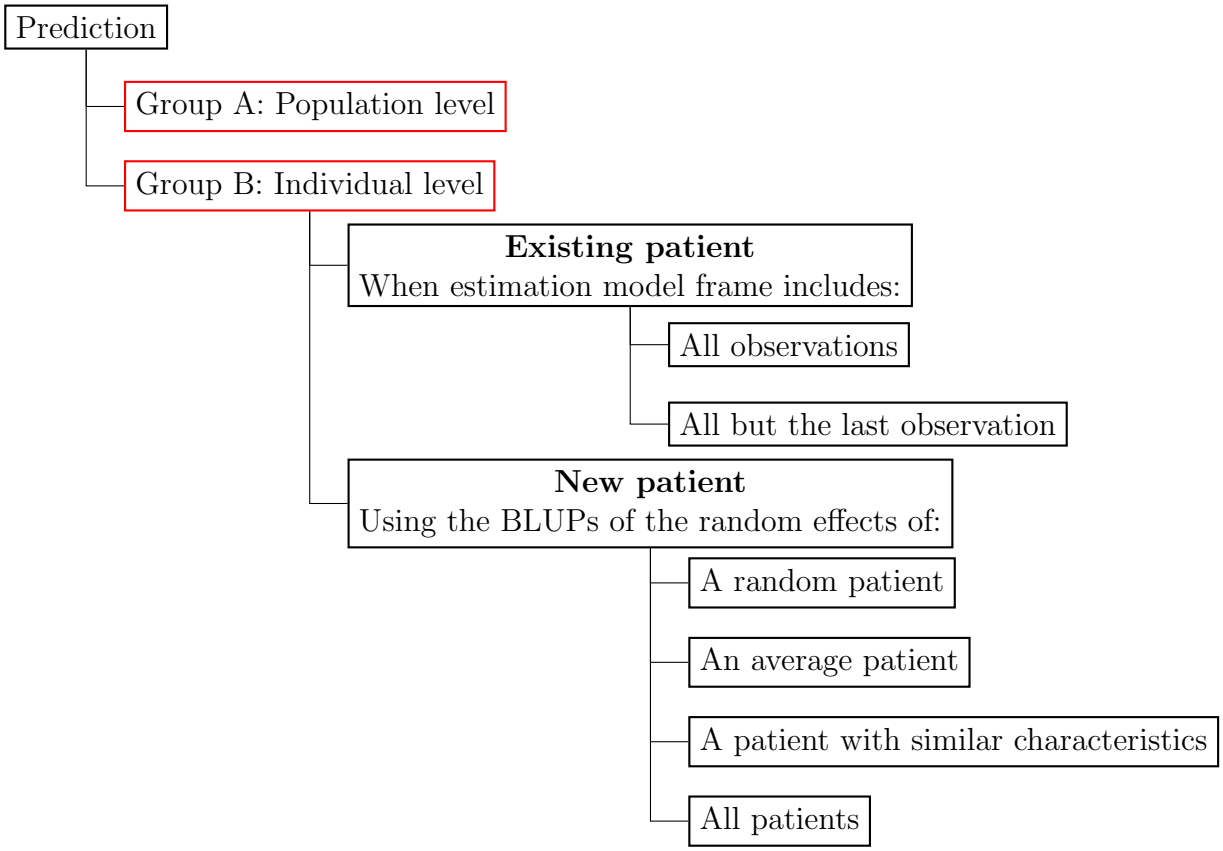

Figure 4.11: Flow chart of prediction methods explored at both a population and individual level.

the estimation model frame. Secondly, prediction allows us to check that the model is appropriate. We can do this through comparison of the predicted values to the true observations, through cross-validation. In addition, an analysis of the variability of the parameter estimates will be explored to further validate the model. Together, this helps us understand the limitations of this model and how effective it is at predicting postoperative patient pain and function based on preoperative factors.

Prediction for mixed-effects models comes with some challenges. One of the advantages of mixed-effects models is the inclusion of a random intercept or slope term to take into account the correlations within patient measurements. This means that we can make

predictions at the patient level, or at the population level. Figure 4.11 displays a flow chart of the different methods of prediction that are explored in this section. These are population and individual level prediction. When making individual patient level predictions there are numerous approaches. Figure 4.11 shows that these approaches are dependent on whether you are predicting the HHS for a patient that exists in the estimation model frame or for a new patient.

Prediction at a population level (Group A in Figure 4.11) estimates the marginal expected value of the response. For example, the marginal expected value of the corresponding response y_h is,

$$E[y_h] = \boldsymbol{x}_h^T \boldsymbol{\beta},$$

where $\boldsymbol{x}_h$ is a vector of the fixed effects and $\boldsymbol{\beta}$ is the estimated coefficients of the fixed effects.

Predicted values at the patient level (Group B in Figure 4.11) are the conditional expectation of the response given the random effects. For example the patient level predictions can be written as,

$$E[y_h(i)|\boldsymbol{b}_i] = \boldsymbol{x}_h^T \boldsymbol{\beta} + \boldsymbol{z}_h(i)^T \boldsymbol{b}_i,$$

where $\boldsymbol{z}_h(i)$ is a vector of the random effects and $\boldsymbol{b}_i$ is the estimated random effects associated with the i-th patient [24].

To help make predictions at both the population and individual level we use conditional estimates and BLUPs. At a population level, the BLUPs of the expected values are obtained by replacing $\boldsymbol{\beta}$ with its conditional estimate, $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$. At an individual level the conditional expectations given the random effects are obtained by making a similar replacement but also replacing the random effects with their BLUPs. For example, the

population and individual level BLUPs corresponding to the expected values in the above examples are:

$$E[y_h] = \boldsymbol{x}_h^T \hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$$
$$E[y_h(i)|\boldsymbol{b}_i] = \boldsymbol{x}_h^T \hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) + \boldsymbol{z}_h(i)^T \hat{\boldsymbol{b}}_i(\boldsymbol{\theta}).$$

The unknown parameters, $\boldsymbol{\theta}$, are replaced by the REML estimate for maximum likelihood, providing us with BLUPs of the expected values [24].

This implies that a patient that the model has trained on has associated BLUPs, which can be extracted from the model and used to make further predictions for that patient. A patient that is not in the estimation model frame does not have this and as a result there are different ways in which we can predict HHS. Therefore, these two situations need to be treated differently, as shown in Figure 4.11 with the separation of new and existing patients.

For existing patients we can compare the predictions made when we have varying amounts of information, as shown in Figure 4.11. We begin by looking at prediction if we have all patient follow-up measurements within the estimation model frame. This is then compared to prediction of a patient's last HHS observation given that their last observation is not within the estimation model frame.

For new patients one method of prediction is to solely use the estimated coefficients for the fixed effects to predict a patient's HHS. Another method is to use the BLUPs of the random effects of a patient that is in the estimation model frame to predict the HHS for a new patient. The patient that is already within the estimation model frame can be selected at random, they can be a representation of an average patient in the population or they can have similar characteristics to the new patient [17]. For example, if a new patient is female and does not smoke then we could find another patient within

the estimation model frame with these characteristics and use the BLUPs of the random effects for prediction.

The last method that will be addressed, as shown in Figure 4.11, is comparing the prediction of a new patient across all possible BLUPs of the random effects within the model. Therefore, presenting the variability of prediction depending on which patient's BLUPs for the random effects are applied.

4.2.1 Group A: Population Level

Firstly we will look at population level prediction. Though this may not be the most accurate method of prediction it can still be useful when comparing two models, through cross-validation. It can help us understand the limitations of the model and inform us of the potential limitations regarding prediction.

One method is to use population level prediction to perform k-fold cross validation. This method of cross-validation takes the dataset and partitions it randomly into k roughly equal groups. In our case, we chose $k = 10$ therefore we partition the rows into 10 groups. Each partition accounts for $\frac{1}{k} = \frac{1}{10}$ of the total dataset. Each group contains 10% of the dataset and becomes a test group. The complementary 90% of the dataset, for each group, trains the model. Therefore, this method takes a random sample of 10% of the dataset and trains the model on the remaining 90%. Then it uses this model to predict on the 10%, using only the fixed effects and therefore predicting population means. Finally, after doing this 10 times, it calculates the mean squared errors (MSE),

$$MSE = \sum_{i=1}^{n} \frac{(observed_i - predicted_i)^2}{n},$$

for each 10 test and training sets, where n is the total number of HHS measurements.

The means of these MSEs are then used to compare models.

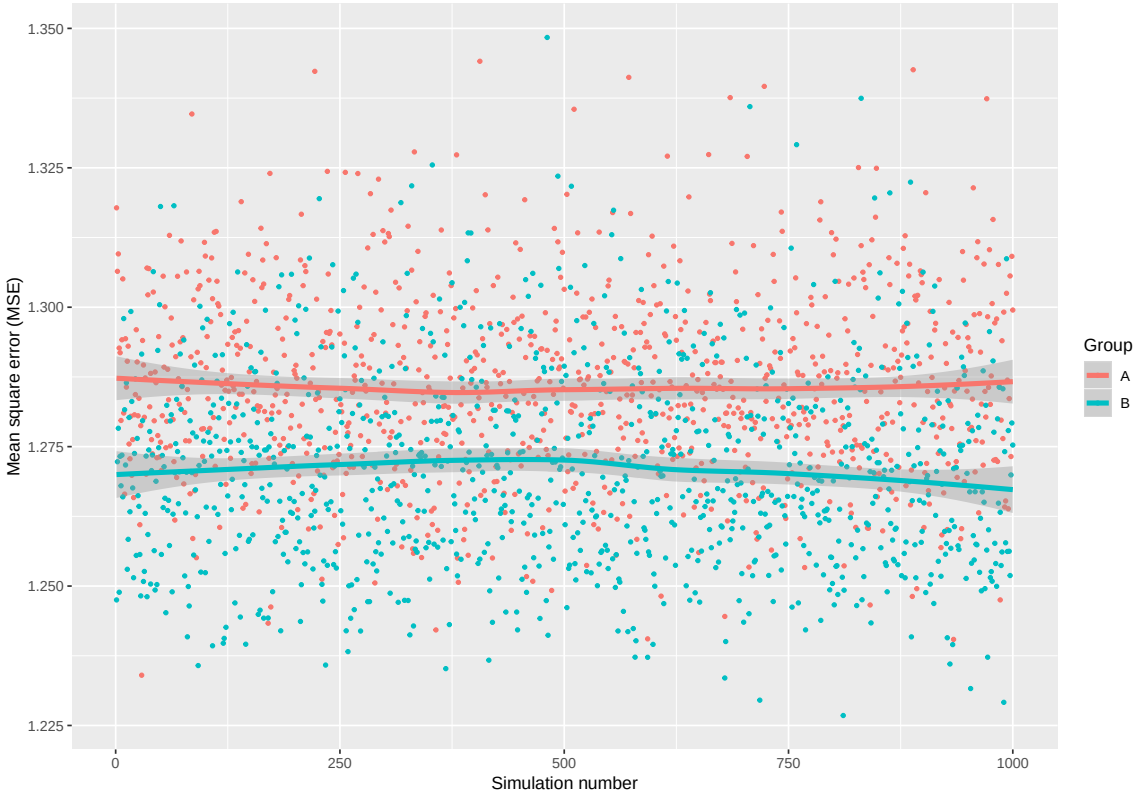

Figure 4.12: Plot of the mean squared error of 1000 simulations of k-fold cross validation. The solid lines are LOESS smoothers to show the overall trend of the MSEs. Group A is the optimal model without the random slopes and group B is the optimal model (Equation 4.6) (n=808).

This process of cross validation was simulated 1000 times for the optimal model. It was also done 1000 times for the optimal model but only with random intercepts. Therefore, comparing the random effect structure. Figure 4.12 is a plot of the mean squared error of 1000 simulations of k-fold cross validation. The solid lines are LOESS smoothers to show the overall trend of the MSEs. Group A is the optimal model without the random slopes and Group B is the optimal model (Equation 4.6). It is clear that Model B, the model with both a random intercept and slope, has a lower MSE. A lower MSE indicates that

the predicted HHS values are closer to the observed values, overall. Therefore, the most appropriate model is still the optimal mixed-effects model (Equation 4.6).

Cross validation was again completed to compare methods of determining the optimal model. This process was simulated 1000 times for a mixed-effects model that was determined through stepwise model selection by AIC. Again we found that the optimal mixed-effects model, defined by Equation 4.6, was the most appropriate as it had a lower MSE.

Figure 4.12 also shows that there is great variability within the MSE of the optimal mixed-effects model. Even though Group B has a lower MSE, indicated by the LOESS smoother, it still has the single highest MSE out of both Model A and Model B. Therefore, it is important to check the variability within the model, such as the variability of the estimates.

Figure 4.13 shows a side-by-side boxplot of the distribution of the estimated coefficients of the optimal model, Equation 4.6, for 1000 exponentially weighted bootstrap simulations. This means that we drew repeatedly from an exponential distribution to create bootstrap weights that were applied to the mixed-effects model. Therefore, random observations were given more weight within the model, for each simulation, to determine if this had an effect on the variability of the coefficient estimates [30]. Each simulation had the same variables present as the optimal model, therefore each coefficient had a value for each cross-validation subsample.

We observe, from Figure 4.13, that the intercept, in particular, varies more than other variables, over the bootstrap sample. Likewise the coefficient for patients who smoke heavily appears fairly variable. Nevertheless, this particular factor level is not significant within the model. We also see that variables such as age and weight have little variability and lie very close to zero. It is clear that the majority of variables are

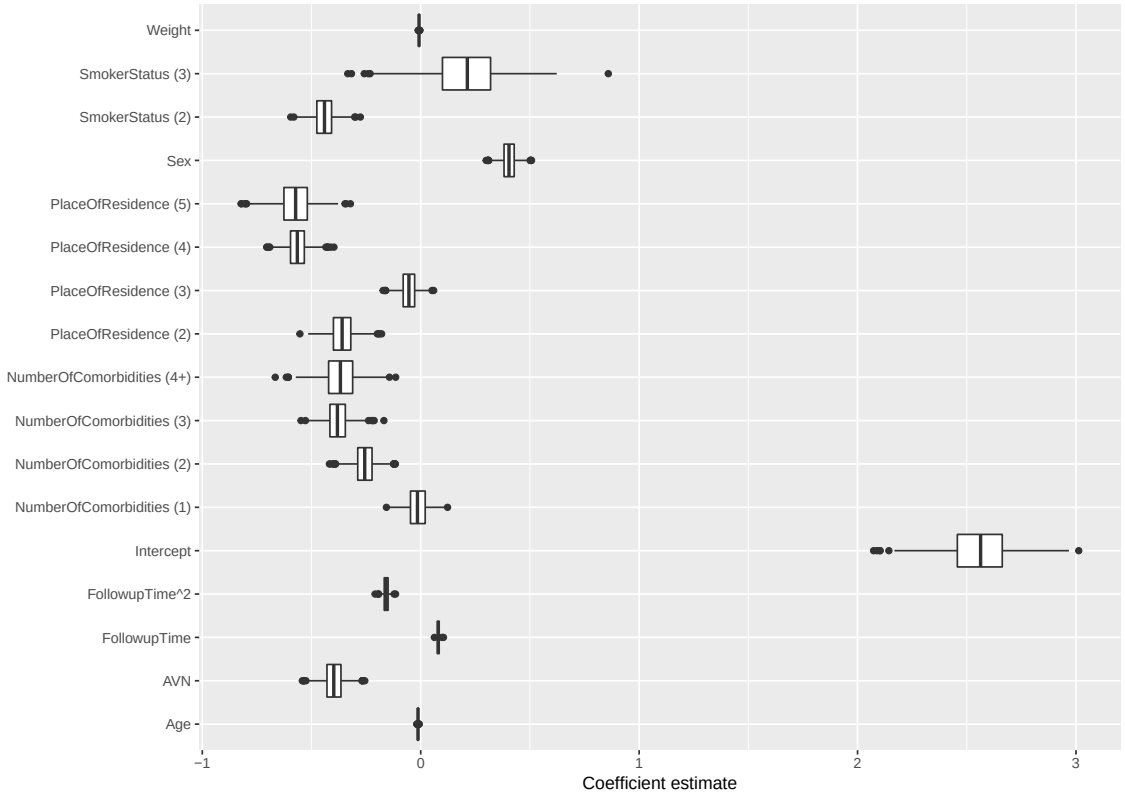

Figure 4.13: Boxplots of the distribution of the estimated coefficients of the optimal model, Equation 4.6, for 1000 exponentially weighted bootstrap simulations (n=808).

associated with a lower HHS, due to the estimated coefficients consistently appearing below zero, particularly the variables place of residence and number of comorbidities.

Though population level prediction can help us cross-validate our model it is also important to compare this to patient level prediction. Specifically comparing the predicted values to the true observations at both the individual and population level.

4.2.2 Group B: Individual Level for Existing Patients

Individual level prediction methods, as shown in Figure 4.11, can be separated by prediction for an existing patient and for a new patient. For an existing patient the estimated

BLUPs of the random effects for that specific patient can be utilized. Whereas, for a new patient this is not possible and another approach needs to be applied. Individual level prediction for a new patient will be discussed later, in Section 4.2.3. In this section we focus on prediction methods for existing patients. The methods addressed in this section are outlined in Figure 4.11 in Group B under existing patients. We begin with prediction for an existing patient, where the estimation model frame includes all of a patient's HHS observations.

Figure 4.14 shows the observed and predicted transformed HHS measurements for 16 randomly selected patients, where all predicted follow-up measurements are in the estimation model frame. The x-axis is the follow-up time and the y-axis is the transformed HHS. The blue points show the observed HHS measurements for that patient. The blue lines show the population level prediction and the pink dashed lines show the individual level prediction, both using the optimal model. Therefore, we can compare the predicted population mean, calculated solely using the estimated coefficients for the fixed effects, and the individual patient predictions.

Due to our assumed structure of the model, the population means are shifted for individual patients by subject specific random intercepts and slopes. Therefore, we see that the slopes and intercepts of the individual profiles are different across subjects. We can also note that the predicted population means, shown in the Figure 4.14, increase very slightly over time.

For some patients the predicted individual profiles deviate from the observed measurements. For example, Subjects 25995 and 20627 show a decrease in the observed values over time, while the predicted individual patterns suggest an increase in HHS over time. Despite this, Figure 4.14 shows that individual level prediction is the most appropriate method of prediction if we wish to get an accurate representation of a specific patient's

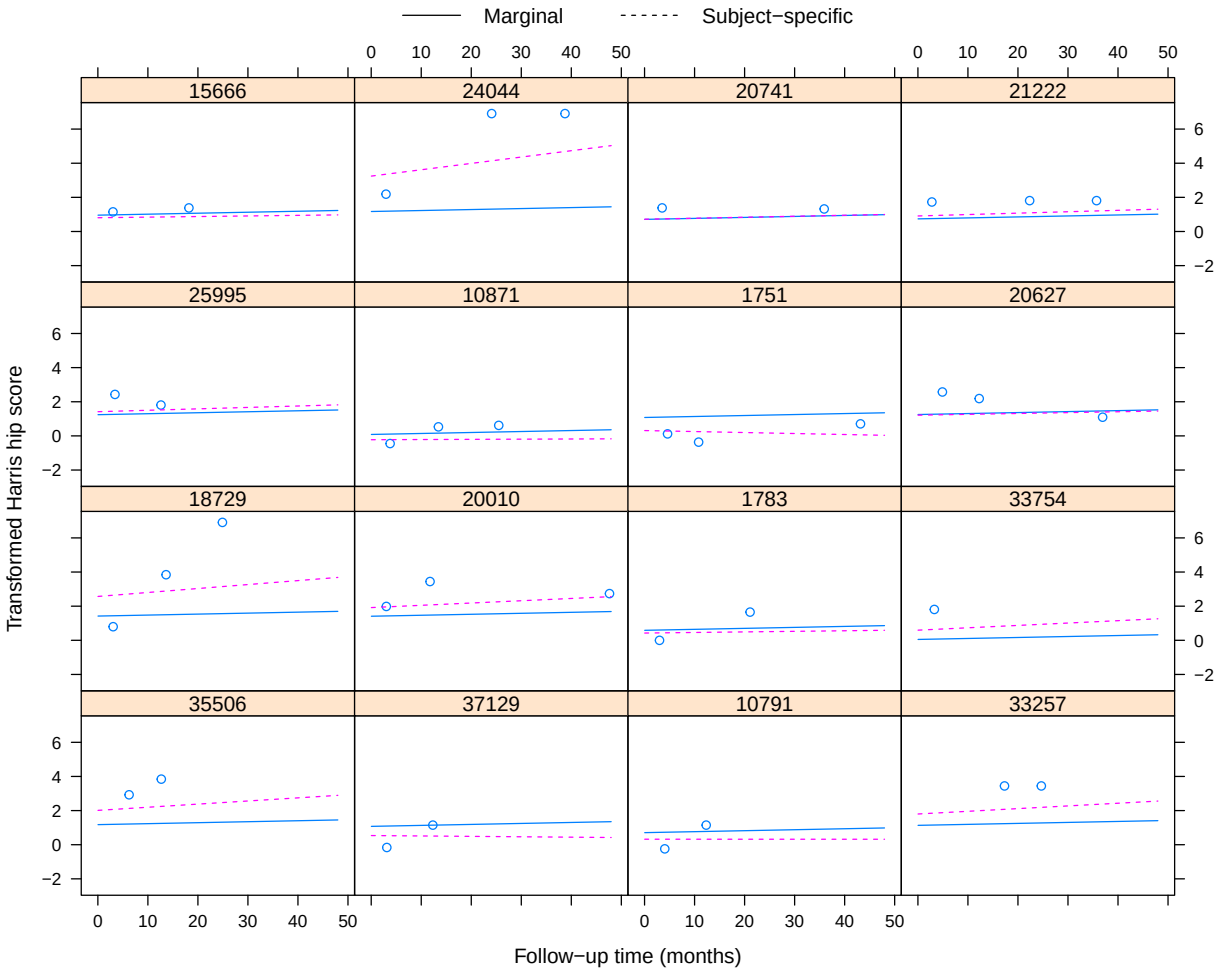

Figure 4.14: Observed and predicted transformed HHS for 16 random patients using the optimal model (Equation 4.6). The blue points show the observed HHS measurements for each patient. The blue lines show the population level prediction and the pink dashed lines show the individual level prediction (n=16).

predicted HHS. This is important because we need the best predictions for individuals so that they can seek the best treatment for their specific condition. Though population level prediction can help us cross-validate our model it is not as capable of making representative patient level predictions.

In the previous case we were comparing predicted values to observations that were in

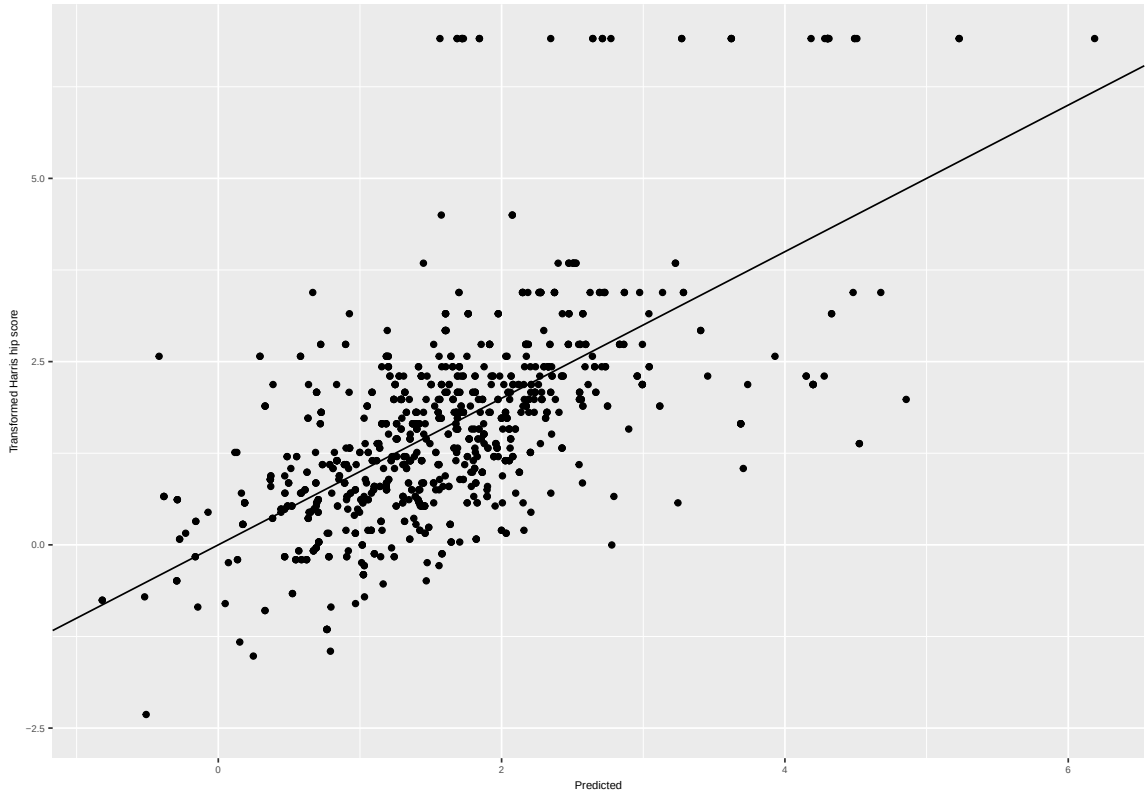

Figure 4.15: Plot of the predicted against the observed HHS values. This is specifically prediction of a patient's last HHS observation given that we only know their previous observations, for 500 random patients.

the estimation model frame. We are also interested in the case where the model predicts future HHS observations that are not in the estimation model frame. Let's consider the case where we have some information, some follow-up observations for a patient so we can utilise BLUPs of the random effects for patient level prediction. Therefore, if we refer to Figure 4.11, this is the case where we have an existing patient but their last observation is not within the estimation model frame.

To do this we remove a random patient's final observation from the dataset and then estimate the model parameters using this subset. Using the new estimated parameters we then predict the HHS measurement at the time of the patient's last observation (the

observation that was removed). These predictions were then compared to the actual HHS observation for that patient, at that time.

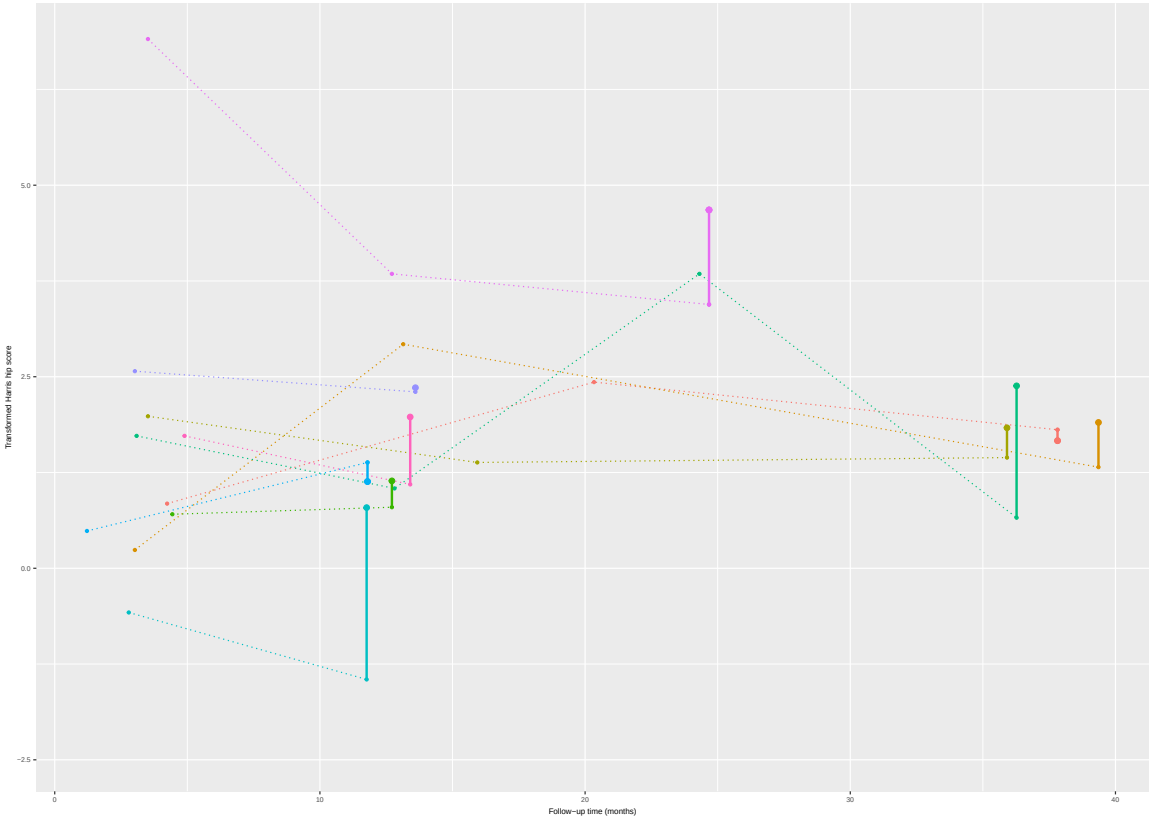

Figure 4.16: Plot of the observed and predicted HHS measurements for 10 randomly selected patients last observation, given that we know the HHS measurements of earlier observations. The x-axis is the follow-up time and the y-axis is the transformed HHS. The dotted lines show the observed patient HHS trajectories. The smaller points are the observed patient HHS measurements. The larger points are the predicted HHS measurements for the patient's last observation. The solid line shows the difference between the predicted and the observed HHS measurements for the patient's last observation (n=10).

Figure 4.15 is a plot of the predicted against the observed HHS values. This is specifically prediction of a patient's final observation given that we only know their previous observations. This plot is the results of 500 simulations, where in each simulation a random patient with more than one observation was used for prediction. As we expect, the

model predicts HHS well in this case, as the points lie close to the reference line. The model still struggles to accurately predict observed HHS measurements of 100, as shown by the horizontal line of outliers at the top of Figure 4.15.

Figure 4.16 is a plot of the observed and predicted HHS measurements for 10 randomly selected patients final observation, given that we know the HHS measurements of their earlier observations. The x-axis is the follow-up time and the y-axis is the transformed HHS. The dotted lines show the observed patient HHS trajectories. The smaller points are the observed patient HHS measurements. The larger points are the predicted HHS measurements for the patient's last observation. The solid line shows the difference between the predicted and the observed HHS measurements for the patient's last observation. This shows that for most patients, within this figure, the predictions are fairly accurate, as shown by the length of the solid lines. The predictions are particularly accurate for patients whose HHS does not deviate greatly over time.

Therefore, we can fairly accurately predict a patient's later observations given that we have preceding observations. Nevertheless, we are most interested in predicting a new patient's HHS trajectory after surgery, based completely on their baseline characteristics. Or in other words, we are interested in predicting postoperative patient pain and function based on preoperative factors, such as age and gender. This is explored in the following section.

4.2.3 Group B: Individual Level for New Patients

To make individual level predictions for new patients we use the BLUPs of the random effects from other patients that are within the estimation model frame. Nevertheless, there are numerous ways in which one can select which patient's BLUPs of the random

effects to apply to a new patient for prediction. Four such methods will be explored in this section and are outlined Figure 4.11. These are, a patient selected at random, an average patient in the population, a patient with similar characteristics to the new patient or a comparison across all patients [17].

The first method is to pick an observation at random and use that patient's estimated random effects to make predictions for a new patient. Nevertheless, selecting a random patient may not be particularly meaningful. Alternatively, we can use the BLUPs of the random effects of an average observation. An average observation is the modal observation for factor variables and the mean value for numeric variables, across all patients. In this case, the random effect terms are set to the level equivalent to the median effect.

Therefore, we predict using an average observation but we modify the variable values to see how prediction varies accordingly. This is due to interest in the amount of variation in predicted HHS trajectories that is caused by changes in the explanatory variables. For example, we can compare the predictions made for an average patient against an average patient who we assume is a smoker (as the modal observation is non-smoker). This comparison shows how the HHS predictions change, for an average patient, if we keep everything but one variable constant.

Figure 4.17 shows the predicted HHS trajectory for different genders and smoking statuses. The transformed HHS is on the y-axis and the follow-up time is on the x-axis. Each plot corresponds to either a male or female patient (indicated by the M and F). The smoking status of the patient is shown by a 1, which indicates a non-smoker, or a 2, which indicates a smoker. The smoker status of 3 (heavy smoker) is not considered here due to it's lack of significance within the optimal model. In this case, we are using an average observation to determine the random effect BLUPs for individual level prediction. The coloured lines display the predicted trajectories for the patient depending on the

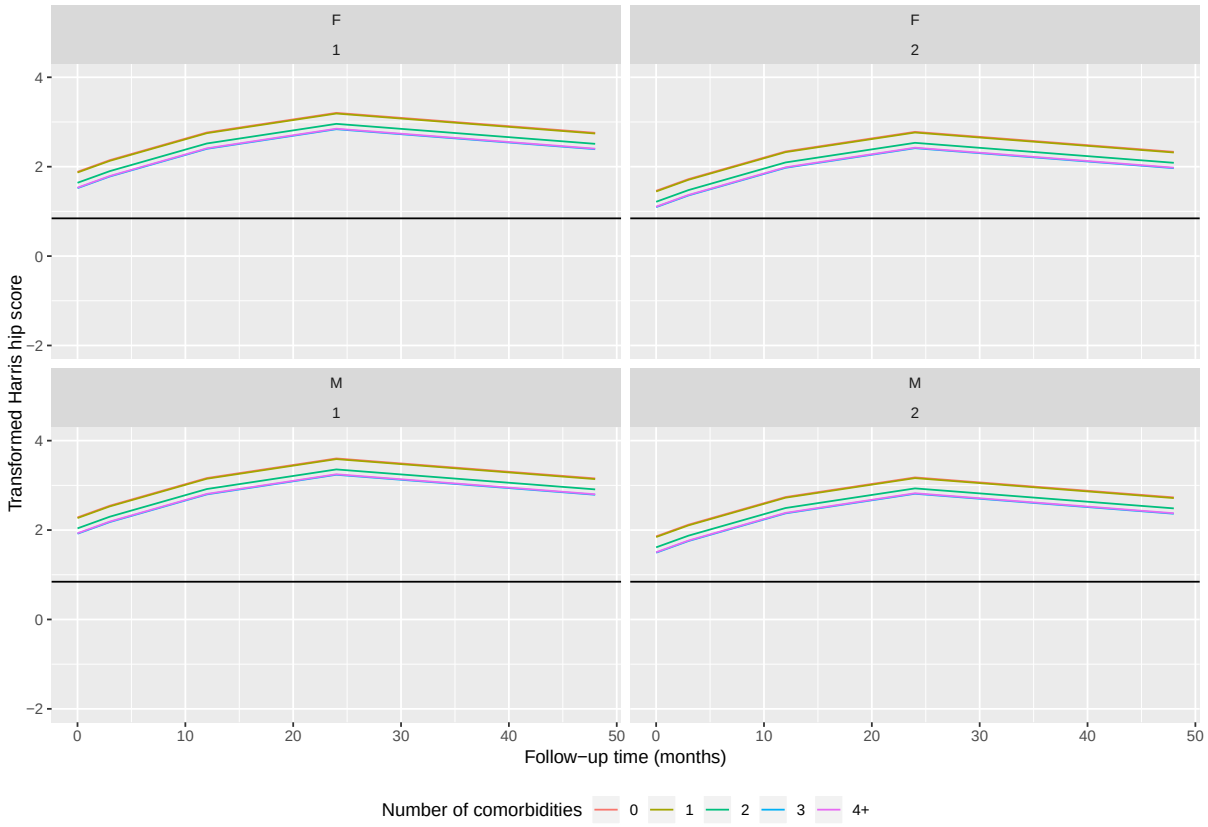

Figure 4.17: Plot of the predicted HHS trajectories against follow-up time using the estimated parameters of an average patient. The characteristics of this patient are altered to compare gender and smoking status. The smoking status 1, indicates a non-smoker, and 2, indicates a smoker. The coloured lines indicate the predicted HHS trajectories given the number of comorbidities the patient suffered from. The solid black line denotes where HHS is equal to 70 and is the line between a poor HHS, below the line, or a good HHS, above the line.

number of comorbidities that a patient suffers from, at the point of surgery. The solid black horizontal line denotes where the HHS equals 70 and is the line between what is considered a poor HHS, below the line, or a good HHS, above the line [12].

From Figure 4.17 we see that in all cases there is a predicted decrease in HHS from 24 months to 48 months. This may be because patients who recover well from the surgery have less follow-up measurements after 24 months, while patients who are still experi-

encing physical limitations are more likely to continue seeking help beyond 24 months after the surgery. Nevertheless we do see a predicted increase in HHS between 0 and 24 months, with the fastest improvement occurring between 0 and 12 months. It is also clear that patients who smoke are predicted to have a lower HHS than patients who are non-smokers. Males are associated with a higher HHS. The number of comorbidities that a patient suffers from does have an effect on their predicted HHS trajectory. Having one or no comorbidities is associated with the highest HHS. Having 3 or 4 or more comorbidities is associated with a lower HHS, and these two levels appear to have fairly identical trajectories.

Not only can we use BLUPs of the random effects of an average patient but we can use BLUPs of the random effects from random patients to understand the variability of this prediction. Figure 4.18 shows the predicted HHS trajectory for four randomly selected patients. Therefore, each patient has individual estimated random effects. This is evident by the different slopes and intercepts that are seen in the four plots. The coloured lines indicate the predicted trajectories given the number of comorbidities the patient suffered from. The solid black line denotes where HHS is equal to 70 and is the line between a poor HHS, below the line, or a good HHS, above the line.

Across all four patients we see that the number of comorbidities does affect the predicted HHS trajectory. Suffering from less comorbidities is associated with less pain and better function. Patient 37129 appears to have the lowest HHS in Figure 4.18. This patient is aged 62 years and is a female of weight 52 kg. They do not suffer from AVN, do not smoke and only suffer from two comorbidities. As a result they live at home with others and are independent. In comparison, the patient with the highest HHS is patient 35506. This patient has the same characteristics as patient 37129 except they are aged 67 years, weigh 61 kg and only suffer from one comorbidity. Nevertheless, there is a stark

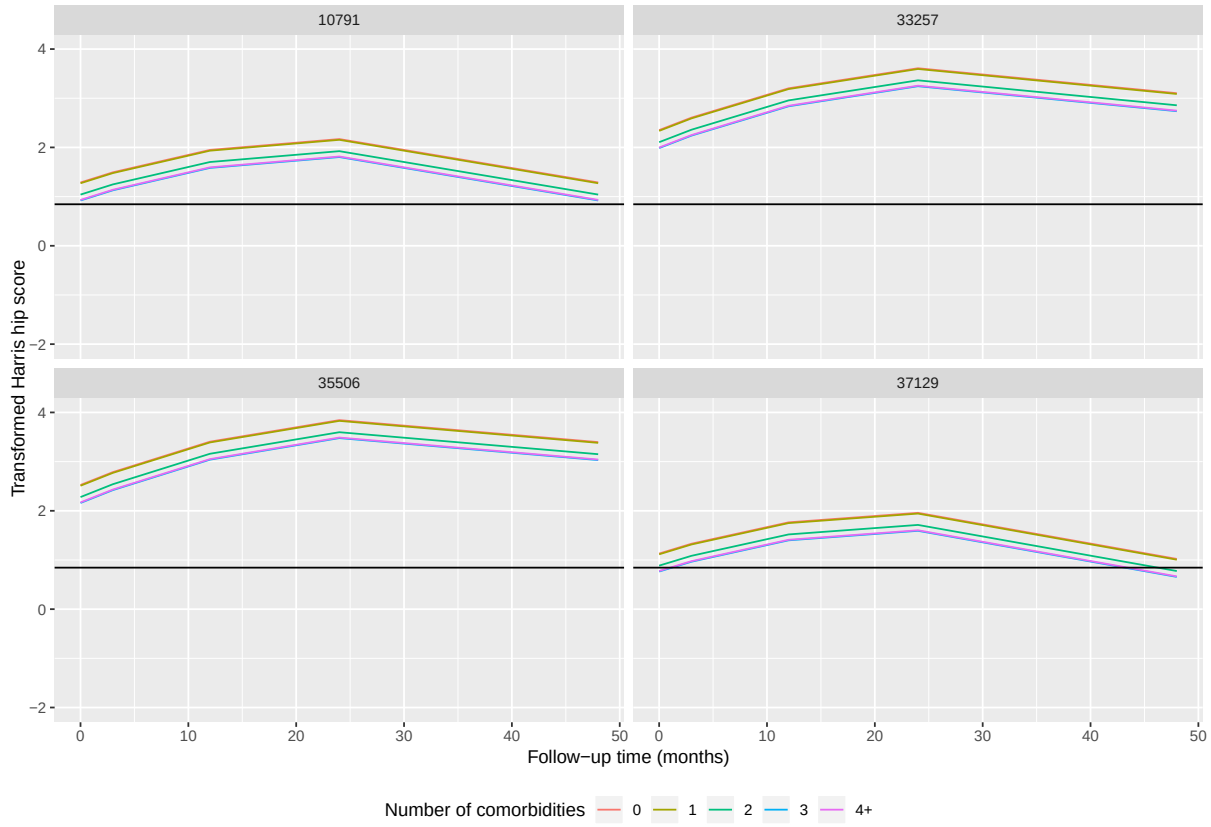

Figure 4.18: Plot of the predicted HHS trajectories against follow-up time using the estimated parameters of four randomly selected patients. The coloured lines indicate the predicted HHS trajectories given the number of comorbidities the patient suffered from. The solid black line denotes where HHS is equal to 70 and is the line between a poor HHS, below the line, or a good HHS, above the line.

difference in their trajectories due to the estimated random effects, which are reflective of the observed HHS measurements.

In Figure 4.18 there is no visible difference between the predicted HHS trajectory for patients with one or two comorbidities. Yet the biggest difference between the patient with the highest and lowest HHS trajectory is the number of comorbidities that they suffer from. If we look closer at the HHS observations attached with patient 37129 and patient 35506 we see that the first HHS observation for patient 37129 is significantly lower

than patient 33257. Therefore, an initial HHS observation can be crucial in obtaining an accurate predicted HHS trajectory. This is because the patient specific estimated random effects that help inform these predictions, make them significantly more representative. Unfortunately, it may not always be possible to obtain an initial HHS measurement that can be used for prediction.

To address the variability in prediction that occurs, depending on how we select the appropriate estimators for the individual random effects, we can look at the variation in prediction across patients. Therefore, we use the third method for individual level prediction of a new patient, shown in Figure 4.11. This method uses the BLUPs of the random effects of patients with the same characteristics as the new patient. This means that we look at how prediction varies if we use all available BLUPs of the random effects, within the subset of patients that have similar characteristics to the new patient.

For example, if a patient presents for a THR they may have the following characteristics:

- Female, 65 years, 70 kg, lives at home and is independent, non-smoker, does not suffer from AVN and suffers from two comorbidities.

We can draw out 100 random patients with similar characteristics and compare the difference in the predicted HHS trajectories to determine the most likely trajectory for that patient. Figure 4.19 is a plot of the predicted HHS trajectories against follow-up time using the estimated parameters of 100 patients with similar characteristics to the specified new patient. By similar characteristics, we mean, they have the same gender, place of residence, smoker status, AVN status and number of comorbidities. The coloured lines show each patient's predicted HHS trajectory. The solid black horizontal line denotes where HHS is equal to 70 and is the line between a poor HHS, below the line, or a good

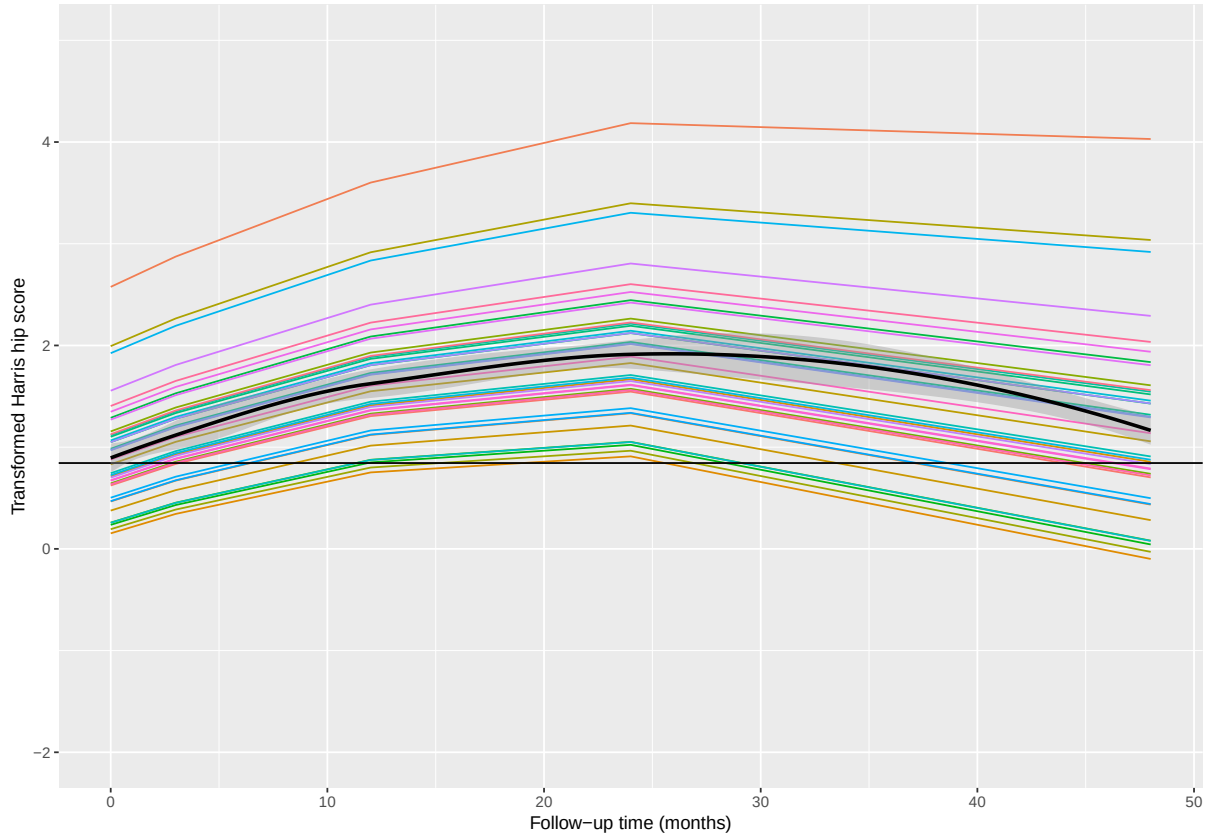

Figure 4.19: Plot of the predicted HHS trajectories against follow-up time using the estimated parameters of 100 patients with similar characteristics to a new patient. The coloured lines show each patient's predicted HHS trajectory. The solid black horizontal line denotes where HHS is equal to 70 and is the line between a poor HHS, below the line, or a good HHS, above the line. The solid black curved line denotes the LOESS smoother for all predicted trajectories.

HHS, above the line. The solid black curved line denotes the LOESS smoother for all predicted HHS trajectories.

Nevertheless, patients with different ages or weights, the variables we did not specify, may be creating the variability in the predicted HHS trajectories. To check this we restrict predictions to patients that are five years older or younger than the new patient and 5 kg heavier or lighter in weight. This restricted the dataset to four patients. We are interested

in seeing if this increased specificity changes the average predicted HHS trajectory.

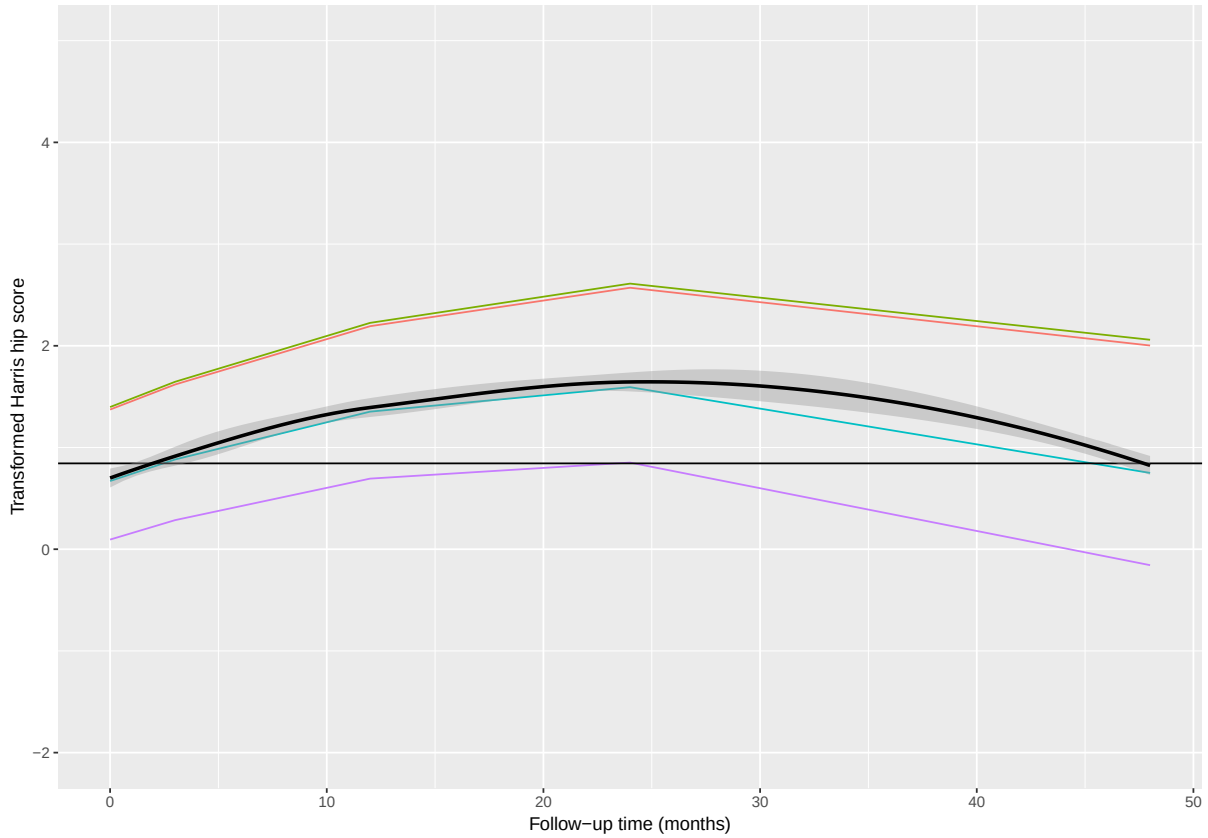

Figure 4.20: Plot of the predicted HHS trajectories against follow-up time using the estimated parameters of 100 patients with similar characteristics to a new patient. The coloured lines show each patient's predicted HHS trajectory. The solid black horizontal line denotes where HHS is equal to 70 and is the line between a poor HHS, below the line, or a good HHS, above the line. The solid black curved line denotes the LOESS smoother for all predicted trajectories.

Figure 4.20 is a plot of the predicted HHS trajectories against follow-up time using the estimated parameters of four patients with almost identical characteristics to a new patient. The coloured lines show each patient's predicted HHS trajectory. The solid black horizontal line denotes where the HHS is equal to 70 and is the line between a poor HHS, below the line, or a good HHS, above the line. The solid black curved line

denotes the LOESS smoother for all predicted trajectories. The smoother line is slightly lower in Figure 4.20 than in Figure 4.19 due to the increased specificity. This shows that selecting for specific characteristics can have an effect on the predicted HHS trajectory. Nevertheless, there is still a significant amount of variability. This variability is largely explained by the inclusion of a random intercept and therefore having at least one HHS observation can significantly improve prediction.

From this exploration it is clear that even among patients with similar characteristics the predicted trajectories were variable. Therefore, to address the variability in predictions that occurs, depending on how we select the appropriate estimators for the individual random effects, we use the last method stated in Figure 4.11. This method examines the variation in prediction across all patients. This shows us how the prediction varies if we use all available BLUPs of the random effects and apply them to a new patient. Furthermore, we can compare these predictions at different factor levels within the explanatory variables.

Figure 4.21 displays the predicted HHS at 24 months after surgery. The x-axis is the patient ID. Each individual patient's BLUPs for the random effects is used but they are assumed to have the characteristics of an average patient. Therefore, comparing the variability in the prediction of the HHS at 24 months if all patients are assumed to have the same characteristics as an average patient. The coloured smoother lines show the average HHS for patients at 24 months given the number of comorbidities that they suffer from. This again confirms that suffering from more comorbidities is associated with a lower HHS. We also see that patients that suffer from no comorbidities have a very similar predicted HHS, at 24 months, to patients who suffer from one comorbidity.

Figure 4.21 shows that the effect of both the intercept and the number of comorbidities slope on each patient simultaneously affect our predicted value. This results in the multi-

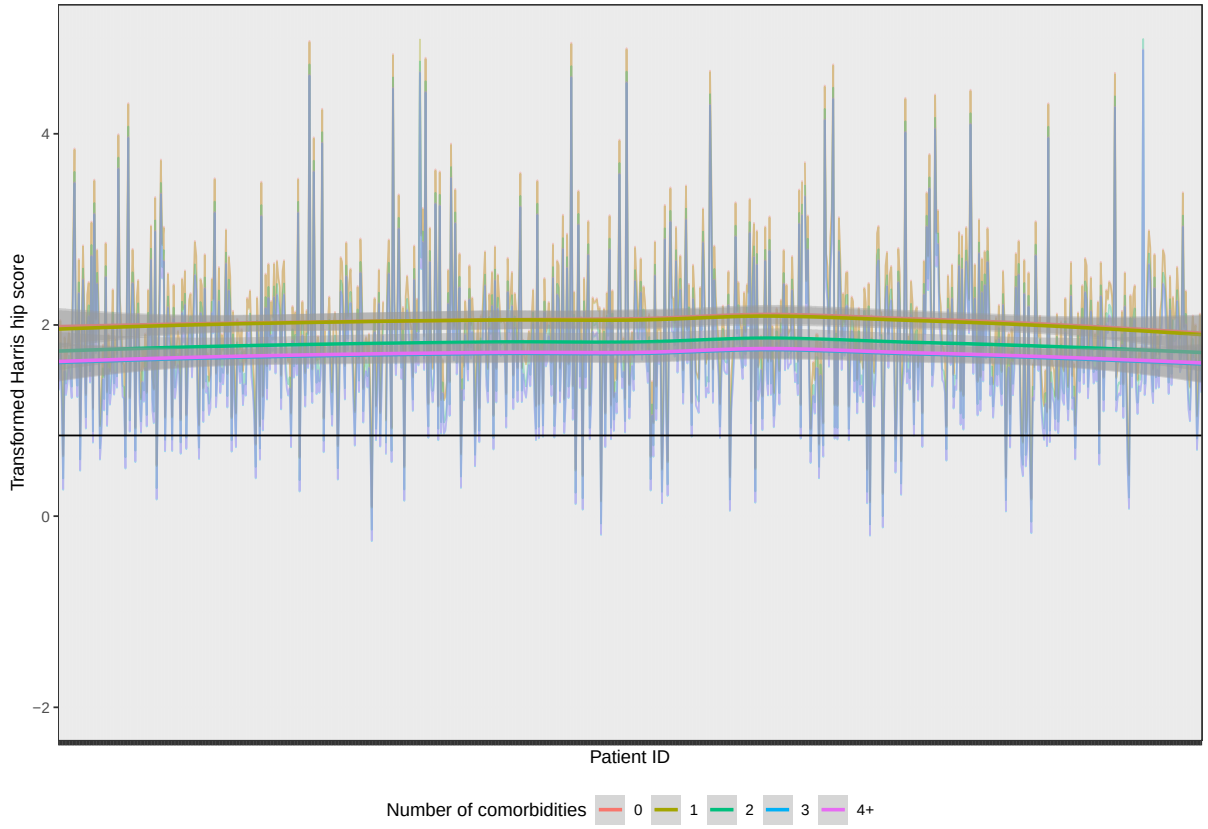

Figure 4.21: Plot of the predicted HHS at 24 months after surgery. The x-axis is the patient ID. Each individual patient BLUPs for the random effects is used but they are assumed to have the characteristics of an average patient (mean and mode of predictor variables). The coloured smoother lines show the average HHS for patients at 24 months given the number of comorbidities that they suffer from. The solid black line separates HHSs that are good, above the line, and poor, below the line.

ple lines for predicted values across all patients not being parallel, instead they resemble noise.

Figure 4.22 shows the predicted HHS at 24 months after surgery. This plot compares the variability in the prediction of the HHS at 24 months if all patients are assumed to have the same characteristics as an average patient. The coloured smoother lines show the average HHS for patients at 24 months given a patient's smoker status. We note that 2

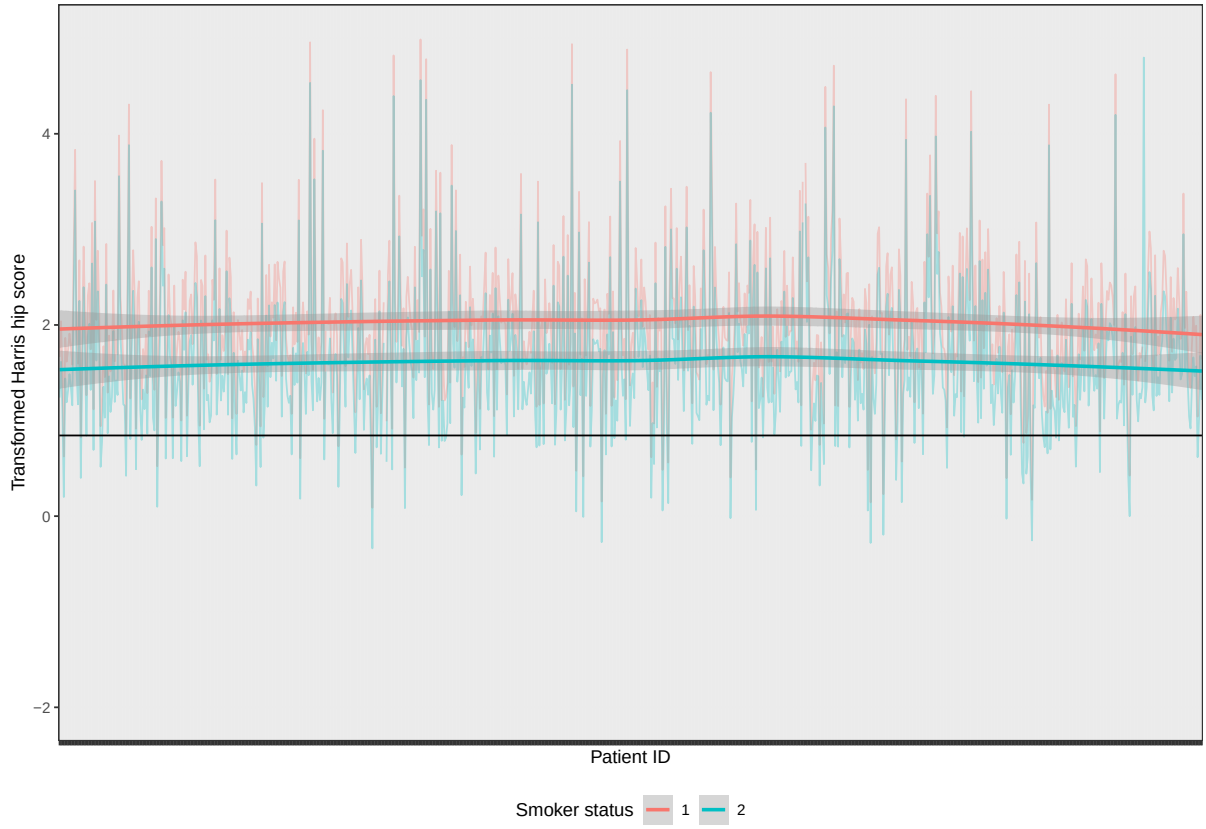

Figure 4.22: Plot of the predicted HHS at 24 months after surgery. The x-axis is the patient ID. Each individual patient's BLUPs for the random effects is used but they are assumed to have the characteristics of an average patient (mean and mode of explanatory variables). The coloured smoother lines show the average HHS for patients at 24 months given a patient's smoker status. The solid black line separates HHSs that are good, above the line, and poor, below the line.

indicates that a patient smokes and 1 indicates that a patient does not smoke. This again confirms that smoking is associated with a lower HHS than non-smoking. The separation between smoker statuses is quite distinct.

Figure 4.23 shows the predicted HHS at 24 months after surgery. This plot compares the variability in the prediction of the HHS at 24 months if all patients are assumed to have the same characteristics as an average patient. The coloured smoother lines show

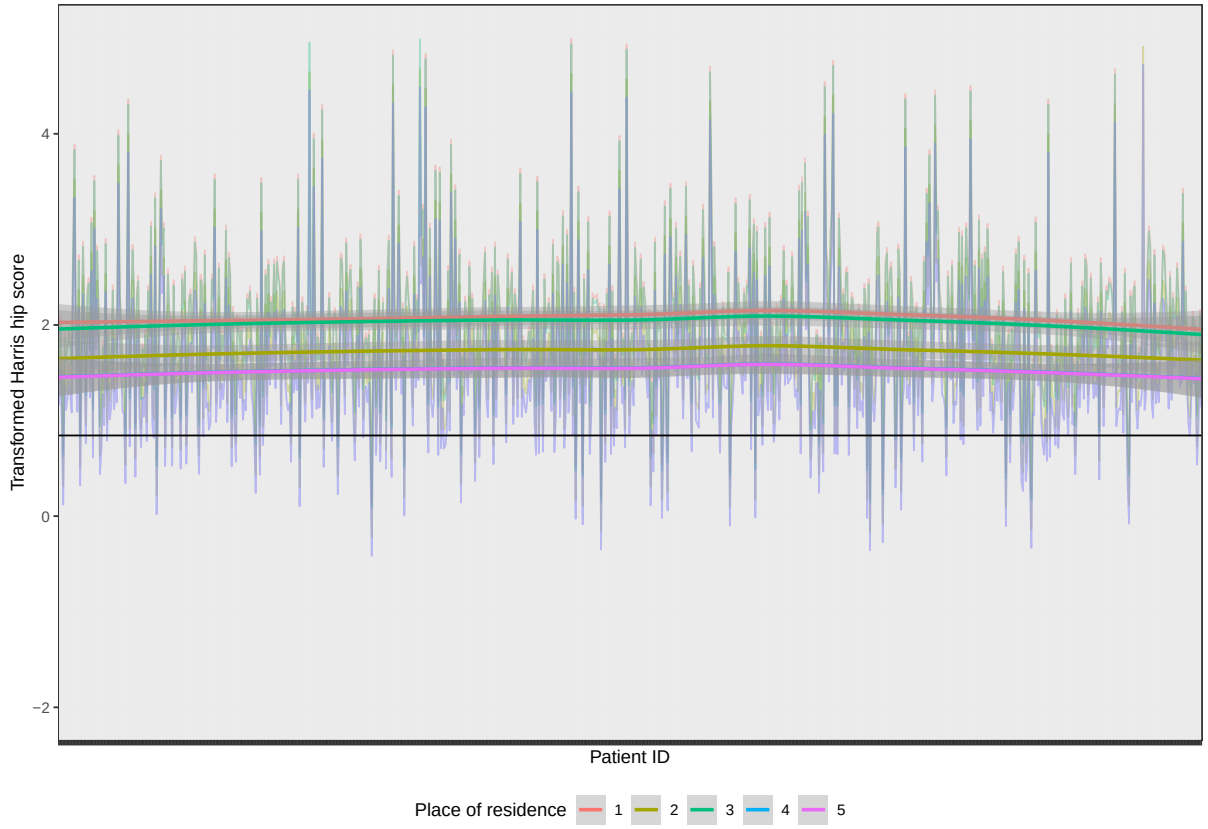

Figure 4.23: Plot of the predicted HHS at 24 months after surgery. The x-axis is the patient ID. Each individual patient's BLUPs for the random effects is used but they are assumed to have the characteristics of an average patient (mean and mode of predictor variables). The coloured smoother lines show the average HHS for patients at 24 months given a patient's place of residence. The solid black line separates HHSs that are good, above the line, and poor, below the line.

the average HHS for patients at 24 months given a patient's place of residence. This figure confirms that patients who need external support (level 2, 4 and 5) are associated with a lower HHS at 24 months. While patients who can live independently (level 1 and 3) are associated with a higher HHS.

We can use this method of prediction to compare the predicted HHS at different follow-up times. Figure 4.24 shows the predicted HHS at 0, 3, 12, 24 and 48 months

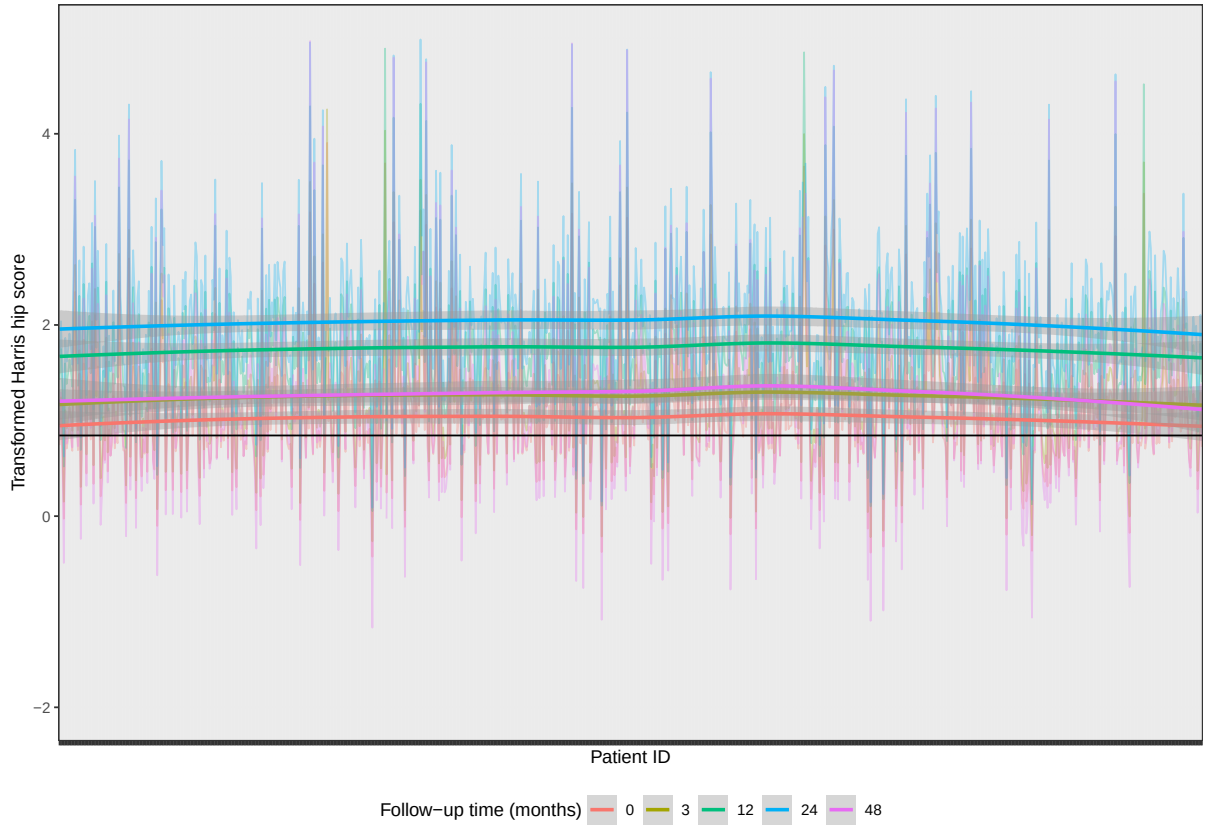

Figure 4.24: Plot of the predicted HHS at 0, 3, 12, 24 and 48 months after surgery. The x-axis is the patient ID. Each individual patient's BLUPs for the random effects is used but they are assumed to have the characteristics of an average patient (mean and mode of predictor variables). The coloured smoother lines show the average HHS for patients at each follow-up time. The solid black line separates HHSs that are good, above the line, and poor, below the line.

after surgery. This plot compares the variability in the prediction of the HHS at these times if all patients are assumed to have the same characteristics as an average patient. The coloured smoother lines show the average HHS for patients at the different follow-up times. Fortunately, this shows that the worst HHS is predicted to occur at 0 months. Likewise the best predicted HHS appears to occur at 24 months. Unfortunately, it appears as though the HHS lowers after 24 months as the predicted HHS at 48 months is similar

to 3 months after surgery. Again this could be because patients who recover well from the surgery have less follow-up measurements after 24 months, creating a bias in the predicted HHS at 48 months.

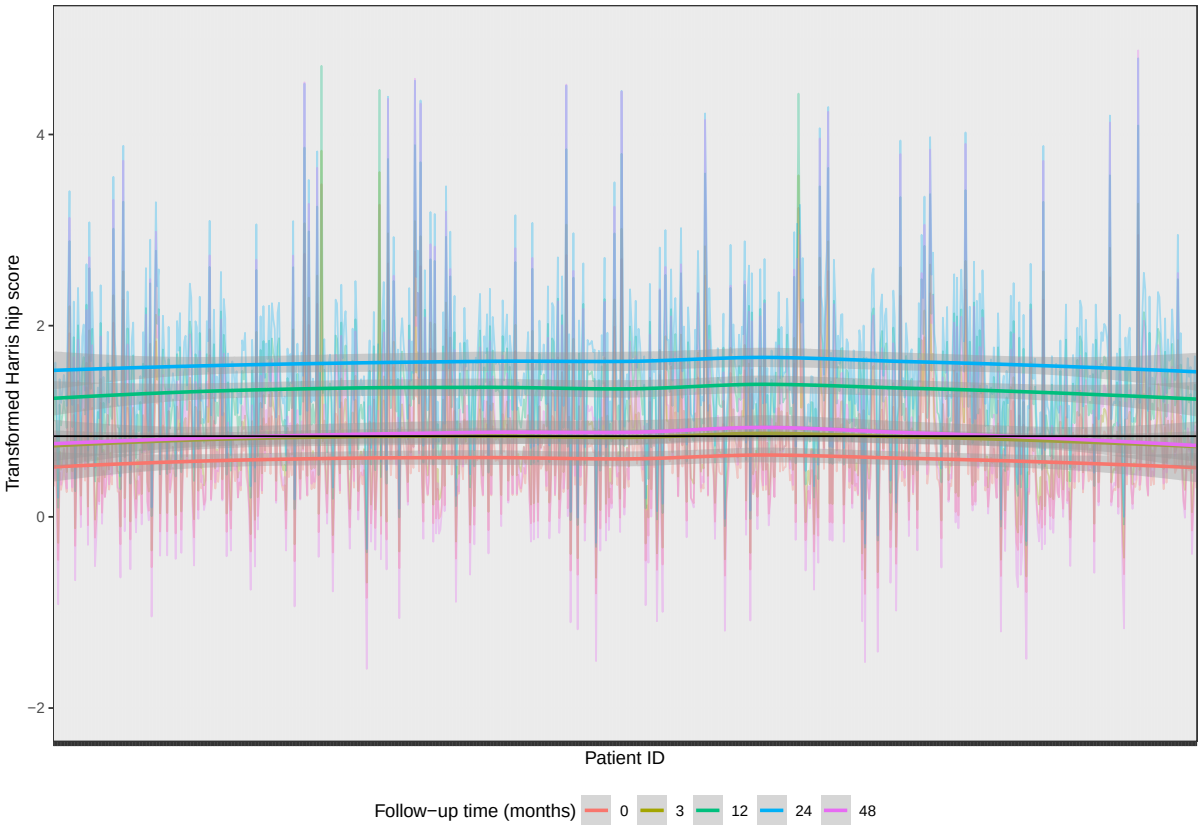

Figure 4.25: Plot of the predicted HHS at 0, 3, 12, 24 and 48 months after surgery. The x-axis is the patient ID. Each individual patient's BLUPs for the random effects is used but they are assumed to have the characteristics of an average patient (mean and mode of predictor variables), except each patient is set to being a smoker. The coloured smoother lines show the average HHS for patients at each follow-up time. The solid black line separates HHSs that are good, above the line, and poor, below the line.

It is also interesting to see how the predicted HHS at different follow-up times is effected by assuming different values for predictor variables. For example, Figure 4.25 shows the predicted HHS at 0, 3, 12, 24 and 48 months after surgery where each patient

is set as smoking. The coloured smoother lines still show the average HHS for patients at the different follow-up times. Again we see that the HHS appears to improve up until 24 months. At this point the HHS is predicted to decrease until at 48 months a patient is back at the same level of pain and function they experienced at 3 months after THR surgery.

Comparing Figure 4.24 and Figure 4.25, it is clear that patients who smoke are associated with a lower HHS at each time point, compared to patients who do not smoke. What is concerning is that at 48 months a patient's HHS is predicted to be on the edge of what is classified as 'good' pain and function. Therefore, it is predicted that this surgery will leave a patient of this demographic with potentially poor pain and function after 48 months.

4.2.4 Summary

On average most patients are predicted to be within the 'good' pain and function bracket (HHS of 70 or above) at 24 months after surgery. Nevertheless, using these various methods of prediction we can get a more accurate representation of a patient's specific HHS trajectory based on their baseline characteristics. Specifically, appropriately using BLUPs for the random effects of patients that are already within the estimation model frame to predict new patient's HHSs. It is also optimal to have at least one HHS measurement to aid in prediction, due to the effect of the intercept, though this is not always practical.

Population level prediction was the fastest and also the simplest method of prediction. Nevertheless, individual level prediction was more representative of the THR dataset. For existing patients, prediction is fairly simple and does not have the same limitations as prediction for new patients. The most representative way to approach prediction for a new

patient is to use the BLUPs of the random effects of patients with similar characteristics while also checking the variability in the prediction over all possible BLUPs of the random effects. This approach to prediction is limited by the sample size. Therefore, if a sample is not large enough to have a number of patients with similar characteristics then using an average patient's BLUPs of the random effects is optimal.

Throughout this chapter we have looked in depth at determining an optimal mixed-effects model and using it for prediction. Nevertheless, we have only looked at mixed-effects models with the response variable of HHS. Within the dataset there is a subset of patients who have a recorded baseline HHS measurement. Using this baseline measurement the change in HHS can be determined for a subset of patients. Therefore, we can look at predicting the change in HHS after surgery based on preoperative factors, such as age and gender.

Chapter 5

Baseline Sub-Study

Since the focus of this work is the analysis of the Harris hip score (HHS), as a measurement of a patient's pain and function, the change in HHS is of interest. In previous analyses, preoperative factors such as age and gender have been used to inform postoperative recovery. Furthermore, the baseline HHS measurements can be used to explore the effect of preoperative factors on the change in a patient's pain and function after surgery. In this chapter we will again used mixed-effects modelling to explore the change in HHS for THR patients.

Our study sample consists only of patients who have a valid baseline HHS measurement. A baseline HHS is classified as a HHS that is measured between zero to 12 months before their date of surgery, inclusive of zero and 12. This definition is based on advice given by clinicians. Even though the baseline measurement is inclusive of up to a year before surgery we note that 91.4% of the baseline measurements occur at time zero. This is because clinicians try and assess the baseline measurements as close to the time of surgery as possible.

The sample size of the sub-study is 897 patients as this is the number of patients with

a valid baseline HHS that we can use to calculate the change in HHS over time. To do this the follow-up HHS measurement was anchored on the baseline HHS. For example, let's say a patient with a baseline HHS of 70 has two follow-up HHS measurements of 60 and 65, respectively. The change in HHS for this patient is 10 and 5, for the first and second observations, respectively. Therefore, by subtracting the follow-up HHS for an individual patient from their specific baseline measurement we have a new response variable, change in HHS.

Patient Demographics	**Baseline**	**Follow-up**
n of patients	897	2280
Mean (SD) age (yrs)	67.2 (12.3)	67.0 (12.3)
Mean (SD) weight (kgs)	80.5 (18.3)	78.6 (17.7)
Mean (SD) height (cms)	166.2 (12.0)	165.4 (12.1)
Female (n,%)	485 (54.1)	1341 (58.8)
Male (n,%)	412 (45.9)	939 (41.2)
Diagnosis (n,%)		
Osteoarthritis	668 (74.5)	1734 (76.1)
AVN	75 (8.4)	182 (8.0)
Rheumatoid arthritis	21 (2.3)	57 (2.5)
Number of Comorbidities (n,%)		
None	177 (19.7)	704 (30.9)
One	263 (29.3)	745 (32.7)
Two	221 (24.6)	497 (21.8)
Three	125 (13.9)	240 (10.5)
Four or more	61 (6.8)	94 (4.1)

Table 5.1: Summary of the distribution of demographic variables for baseline and follow-up data. Where the baseline data consists of observations measured between zero and 12 months before surgery. The Follow-up data consists of observations that occurred after surgery (n = 897).

When we take a subset of the dataset, such as this baseline data, it is important to check if the subset has a particular demographic profile that is not a true reflection of

the dataset as a whole. Table 5.1 shows a comparison of the distribution of demographic variables for both the 897 baseline patients and 2280 patients with follow-up data. This shows that the distribution of demographic variables is relatively consistent across the baseline and the follow-up data, with the mean age being approximately 67 years. There is a slightly lower percentage of females within the baseline sample. The proportion of patients with only one comorbidity is less within the baseline dataset (20%) than the follow-up dataset (31%).

To further check our definition of baseline measurements including a year before surgery a comparison of the demographic characteristics was completed. This compared the baseline data that was collected at the time of surgery and the baseline data that was collected within the year before surgery. There appeared to be no significant difference between the distribution of variables within those two groups. This supports our definition of baseline including a year before surgery.

For example, Figure 5.1 displays a side-by-side boxplot of the distribution of age for the baseline data (≤ 0) and follow-up data (> 0). The baseline data is separated into those at the time of surgery ($= 0$) and those in the year preceding surgery (< 0). Fainter points indicate a low density of outliers while darker points indicate a high density of outliers. The follow-up data appears to have a higher density of outliers that are younger. This is most likely a reflection of the difference in sample size between the follow-up data and baseline data. Therefore, age is distributed similarly across the baseline data, the baseline data subsets and the follow-up data.

After linking the relevant follow-up data to the baseline data the change in HHS was determined for each follow-up. The number of patients within the study was reduced to 798, as all patients were required to have at least one follow-up with a HHS. The distribution of the change in HHS for each follow-up showed that 96.6% of follow-up HHS

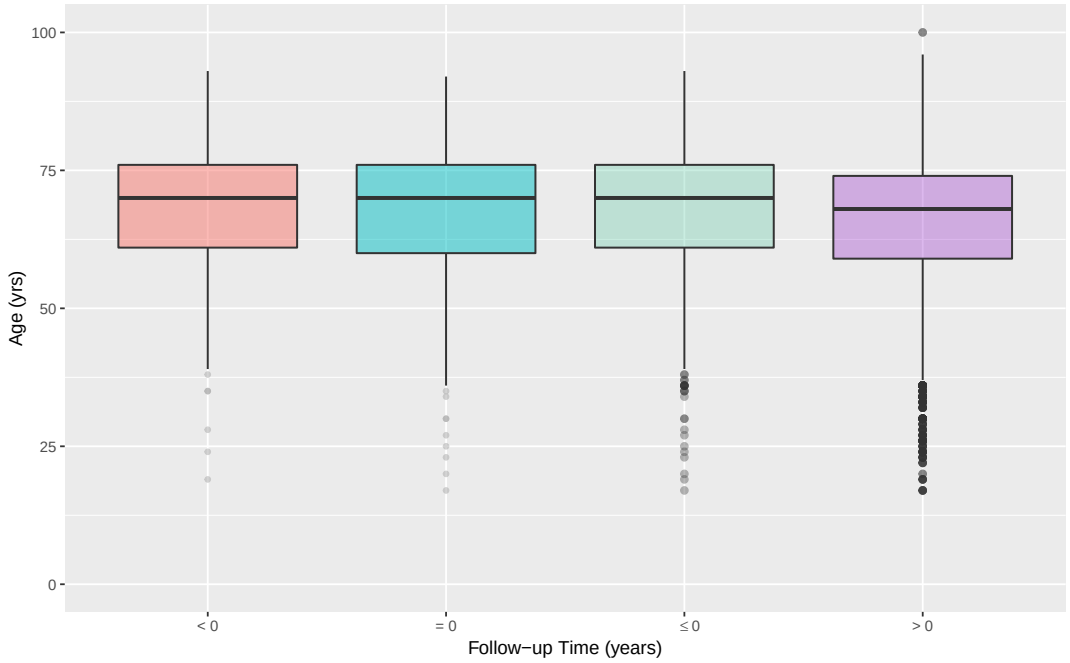

Figure 5.1: Side-by-side boxplot of the distribution of age for the baseline data (≤ 0) and follow-up data (> 0). The baseline data is separated into those at the time of surgery ($= 0$) and those in the year preceding surgery (< 0). Fainter points indicate a low density of outliers while darker points indicate a high density of outliers (n = 897).

measurements are higher than the patient's baseline. The mean change in HHS is 40.14, which is relatively high as the HHS is only between 0 and 100, while the standard deviation is 0.85. Of those patients who have recorded baseline measurements and follow-ups for pain and function, almost all experience an improvement in HHS after surgery.

Figure 5.2 shows the distribution of the response variable, change in HHS. We see that the majority of observations are an improvement from the baseline HHS measurement. This again shows that there are relatively few observations that are worse than the baseline HHS observation. The majority of observations are improvements of around 40-50 points. In the context of the HHS, which is a score between 0 and 100, this is a large change in HHS.

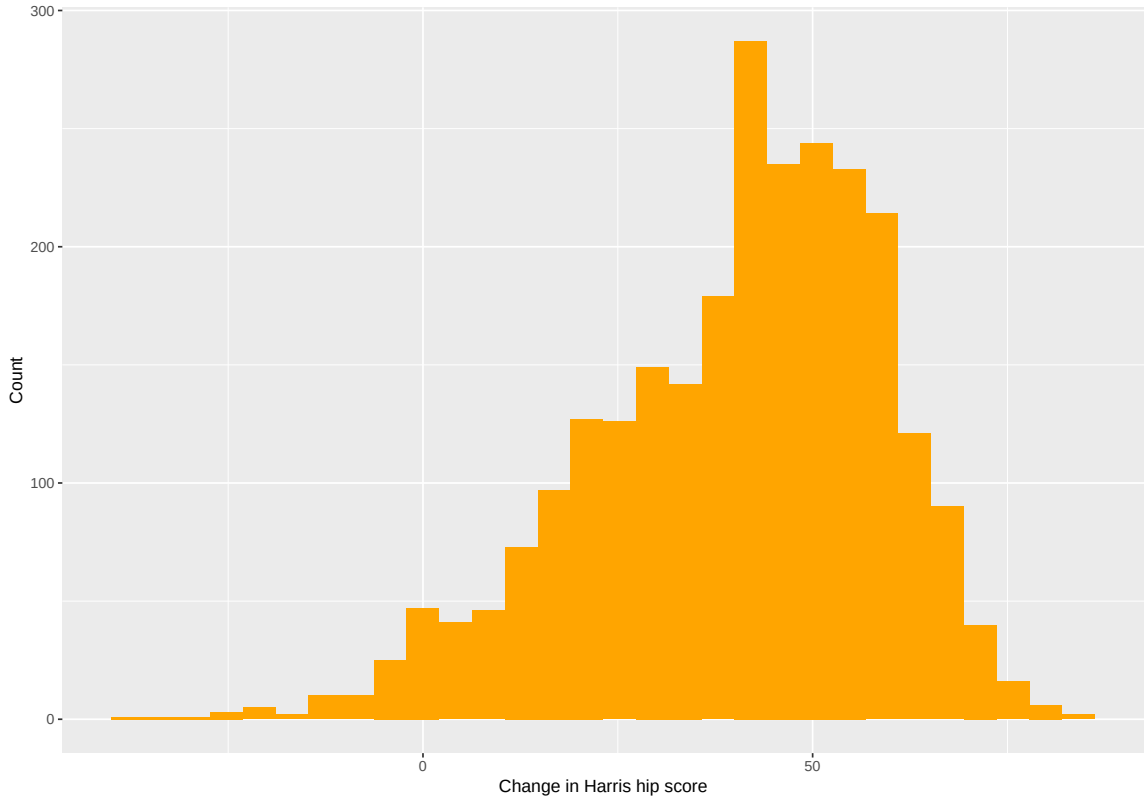

Figure 5.2: Histogram of the change in Harris hip score (n = 897).

Figure 5.3 shows the density plots of the change in HHS (from baseline to follow-up) separated into follow-up time brackets. The distribution of the change in HHS for follow-ups that occurred within 12 months after surgery appears to have a lower peak and be less left skewed than follow-ups that occurred beyond 12 months after surgery. Though there is no stark difference between the first two and the last two time brackets we can still see a slight increase in the peak change in HHS as time after surgery increases. This suggests that patient pain and function is in general improving with time. The follow-up times with a negative change in HHS are fairly evenly spread across the time brackets, with the majority occurring within the first two years after surgery.

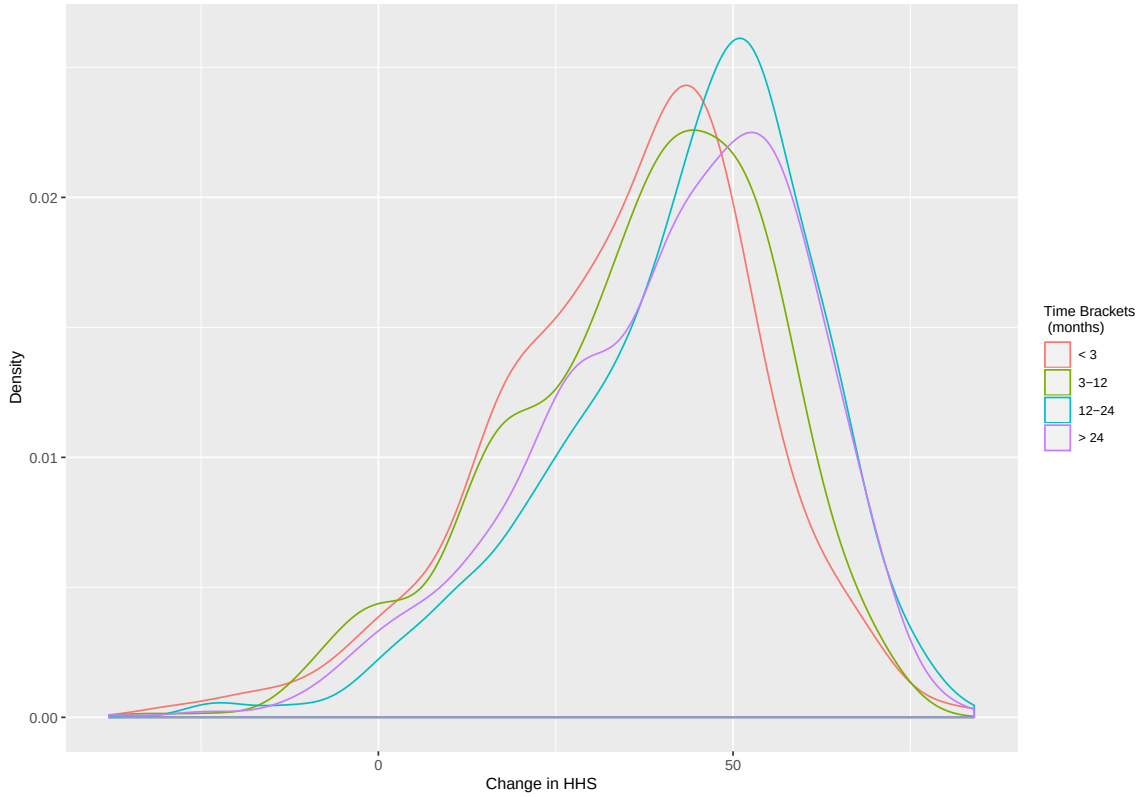

Figure 5.3: Density plots of the change in HHS (from baseline to follow-up) separated into follow-up time brackets (n = 897).

5.1 Mixed-Effects Model

To account for the repeated measurements of the change in HHS over time, application of the class of statistical models known as mixed-effects models is appropriate [10]. These models have been introduced in Chapter 4. Mixed-effects models can handle unbalanced datasets, where each patient can have a varying number of follow-up measurements and varying time between these follow-ups. Importantly, they account for the correlation between repeated HHS measurements within a specific patient through the inclusion of random effects.

Figure 5.4 shows a plot of the change in HHS trajectories over time for all patients

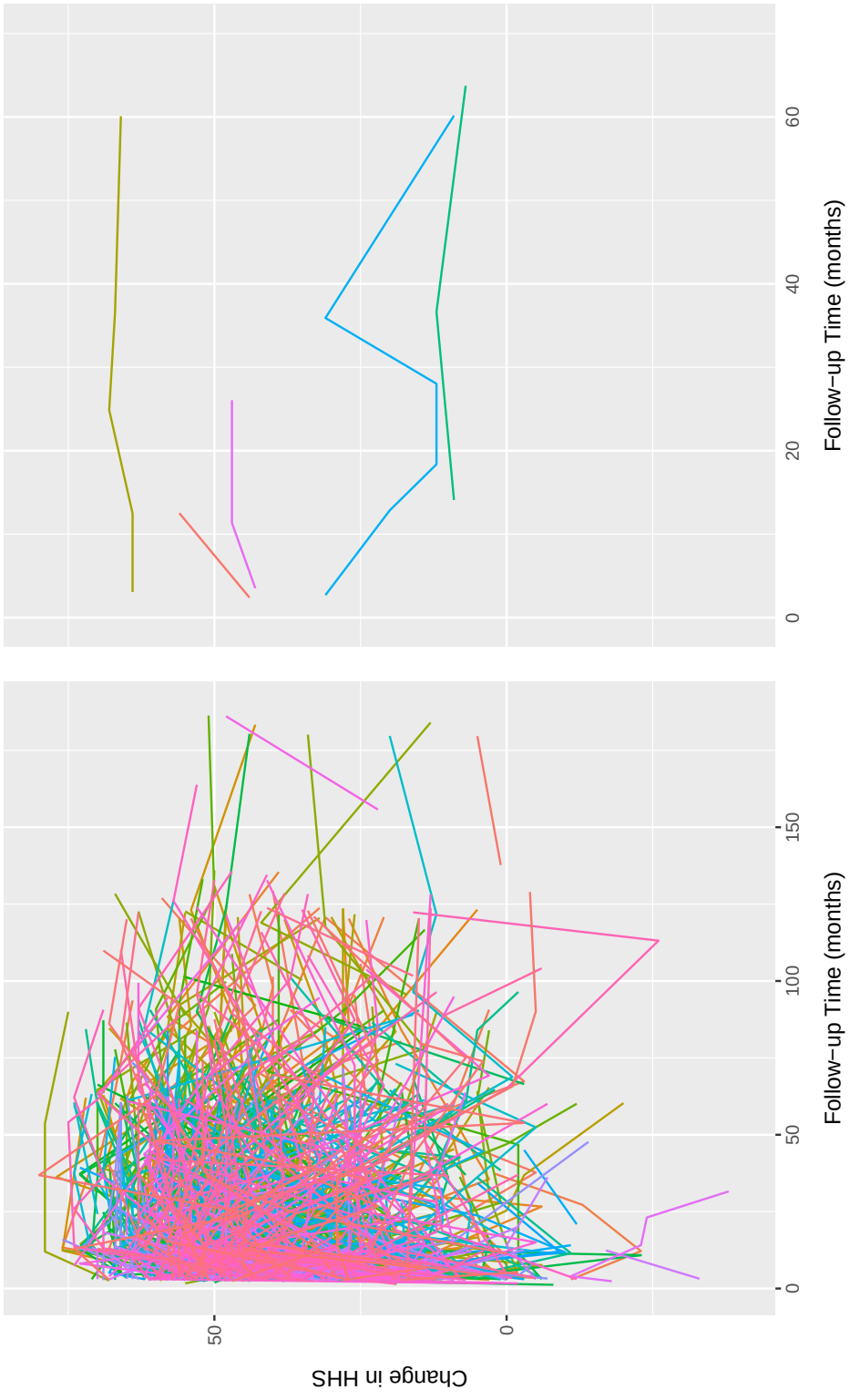

Figure 5.4: Left: Plot of the change in HHS trajectories over time for all patients with a baseline HHS measurement (colours for each patient are random). Right: Plot of the change in HHS trajectories over time for five randomly selected patients (n = 897).

with a baseline HHS measurement. As we would expect we see that patients often have a varying number of follow-ups at non-consistent times. This again encourages the application of mixed-effects models. We also see that there are few patients with a change in HHS that is negative as most patients in this baseline sub-study experience improvements in HHS after THR surgery.

A mixed-effects model was constructed with the change in HHS as the response variable. In this case, we allow for separate average evolutions of change in HHS for each patient, through random effects. We controlled for a collection of variables including, but not limited to, baseline age, gender, follow-up time, place of residence and implant type. On further analysis, there are no significant correlations between any of the predictor variables. The strongest correlation is between weight and age and it is approximately 0.2. This model does not take into account any interactions but it includes a quadratic term for follow-up time. Using the same stepwise process outlined in Chapter 4, predictor variables that were significant (p-value < 0.05) were retained in the model. This model was ultimately fit using restricted maximum likelihood (REML).

In the following analyses, the logit of the change in HHS is not used as it was in previous analyses. This is because the assumption that the variance of the error terms is constant was satisfied on the original HHS scale when we take the differences, which will be discussed further.

The optimal random effects structure included only random intercepts. This was in contrast to the previous mixed-effects model that also included random slopes. The optimal model is written as,

$$\begin{aligned}\text{ChangeInHHS}_{ij} =& \alpha + \beta_1 \text{FollowupTime}_{ij} + \beta_2 \text{FollowupTime}_{ij}^2 + \\ & b_i + \epsilon_{ij}.\end{aligned} \tag{5.1}$$

where ChangeInHHS$_{ij}$ is the change in HHS of observation j from the baseline HHS measurement for the i-th patient. The explanatory variables used in Equation 5.1 are follow-up time and the quadratic term for follow-up time. The random effects, b_i, are normally distributed: $N(0, d^2)$. The errors, ϵ_{ij}, are also normally distributed. Since the optimal model is a random intercept model it allows for a random shift around the intercept resulting in fitted lines that are parallel to the population fitted line.

	Parameter	**Estimate**	**SE**	**p-value**
Fixed Effect Estimates	Intercept (mean change in HHS at 0 months)	34.65	0.81	<.0005
	Slope (months after surgery)	0.72	0.09	<.0005
	Quadratic (months)^2	-0.01	0.002	<.0005
			Standard Deviation	
Component of Variance Estimates	Within Patient Residual	11.66		
	Between Patient Residual (int.)	14.19		

Table 5.2: Model parameters for the optimal model of the change in HHS (Equation 5.1); Number of observations: 779.

The results for the optimal model of the change in HHS are shown in Table 5.2. This model suggests that follow-up time after surgery is strongly associated with the change in HHS. An increase in follow-up time is associated with an increase in the change in HHS. An increase in the quadratic term for follow-up time is associated with a decrease in the change in HHS.

The variability of the intercept across patients has a standard deviation of 14.19, while 11.66 is the standard deviation of the variability within patients. The within patients variability is smaller than the intercept variability, which suggests that the mixed-effects

model is potentially accounting for the correlations within patient repeated measurements.

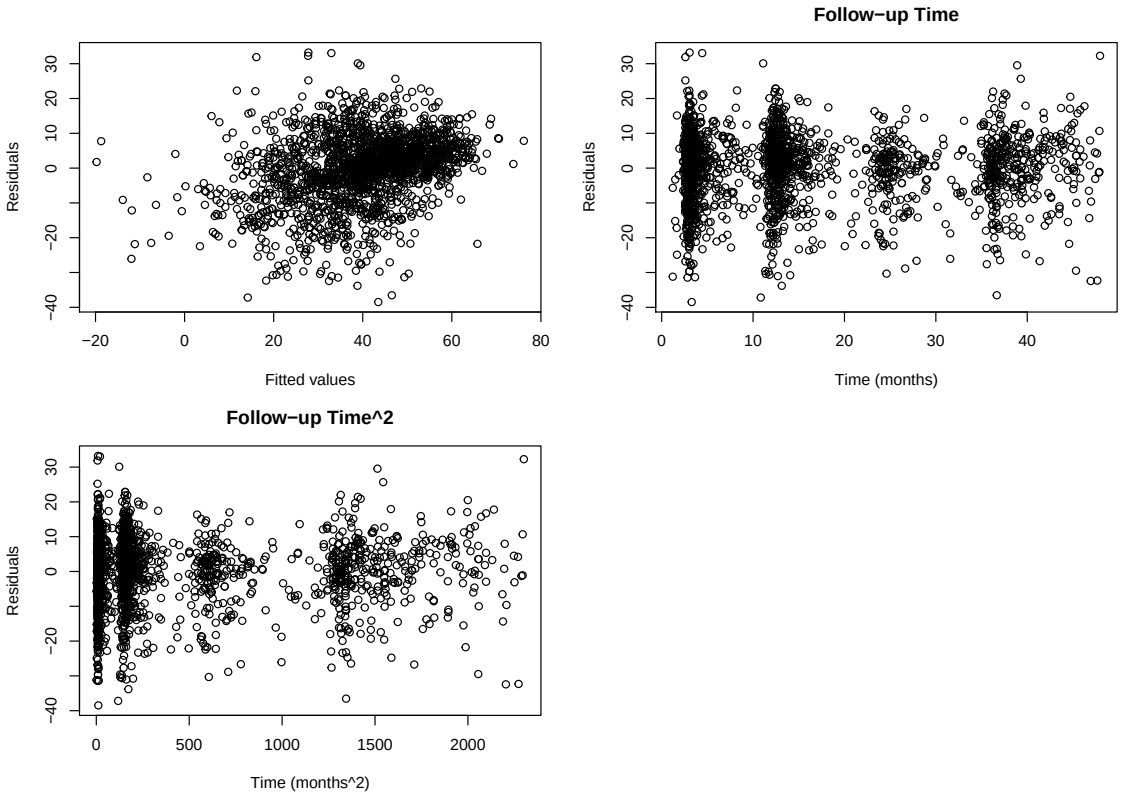

Figure 5.5: Diagnostic plots for the random intercept model defined in Equation 5.1. This includes scatterplots of the residuals against fitted values, and the residuals against follow-up time and follow-up time^2 (n = 779).

To check if the assumption of homogeneity was violated the following diagnostics were explored. Figure 5.5 displays diagnostic plots for the random intercept model defined in Equation 5.1. This includes scatterplots of the residuals against fitted values, and the residuals against the continuous quantitative explanatory variables, follow-up time and follow-up time^2. Unlike previous models, this model does not fail to effectively predict values about the upper bound of the HHS. This is due to the newly defined response variable being change in HHS, instead of the transformed HHS. Figure 5.5 shows that the assumption of homogeneity is satisfied, as there does not appear to be any clear trends

within the residuals of the model.

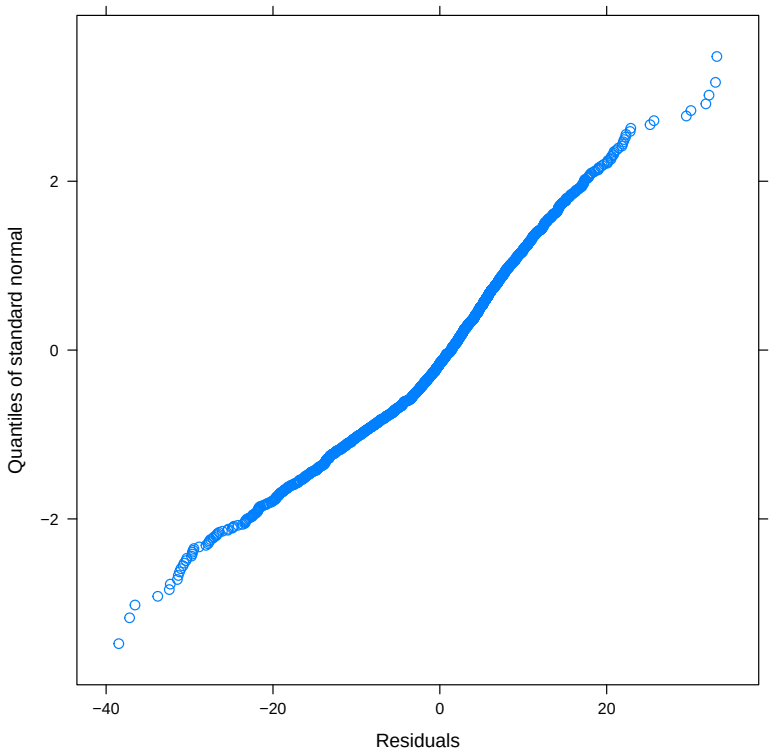

Figure 5.6: Normal Q-Q plots of the conditional Pearson residuals for the model in Equation 5.1 (n = 779).

To further investigate the assumptions of the optimal mixed-effects model normal Q-Q plots were explored. Figure 5.6 shows the normal Q-Q plot of the conditional Pearson residuals. The normal Q-Q plots do show some small deviations from a linear trend at the right upper tail. Nevertheless, these observations do not create a strong deviation from the linear trend we expect to see in a normal Q-Q plot and therefore the assumption of normality is reasonable.

We also look at the normal Q-Q plot of the predicted random effects, in this case the random intercept. The effects are estimated using BLUPs (best linear unbiased predictors), which were discussed in Section 4.2. The resulting Q-Q plot is shown in Figure

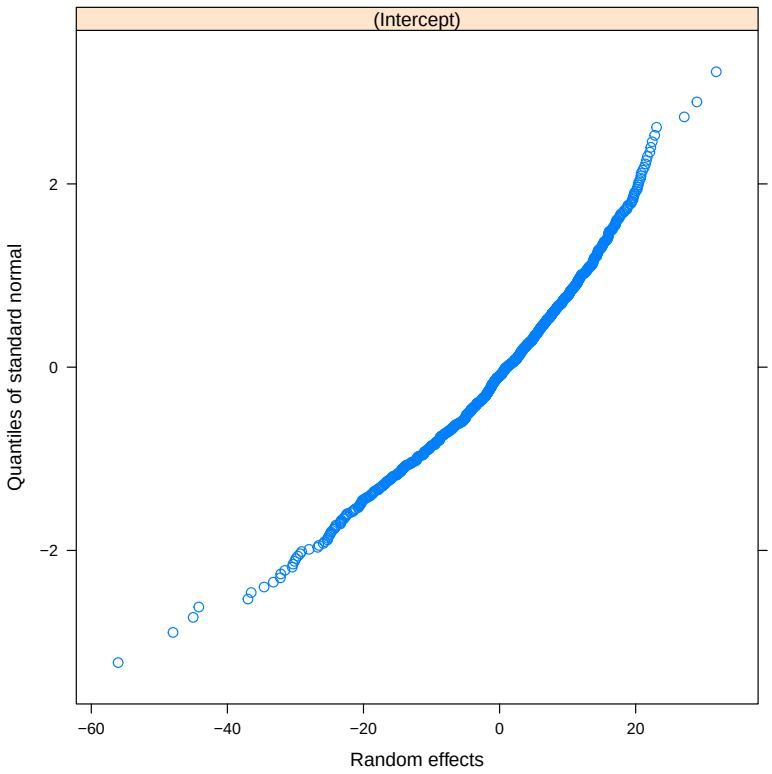

Figure 5.7: The normal Q-Q plot of the predicted random intercepts for the model in Equation 5.1 (n = 779).

5.7 and appears to be slightly curvilinear. However, this figure overall suggests that the random effects are fairly normally distributed.

Since this model appears to satisfy the mixed-effects model assumptions we can now look at it's ability to make meaningful predictions. Though the optimal model within this chapter only contains time variables we can still look at prediction. We are interested in seeing if the predicted trajectories suggest that patient HHS generally improves with time or not. We want to determine which patients are predicted to have an overall decrease in HHS. Specifically the models ability to make predictions about a specific patient's change in HHS after surgery, in comparison to the previous mixed-effects model explored in Chapter 4.

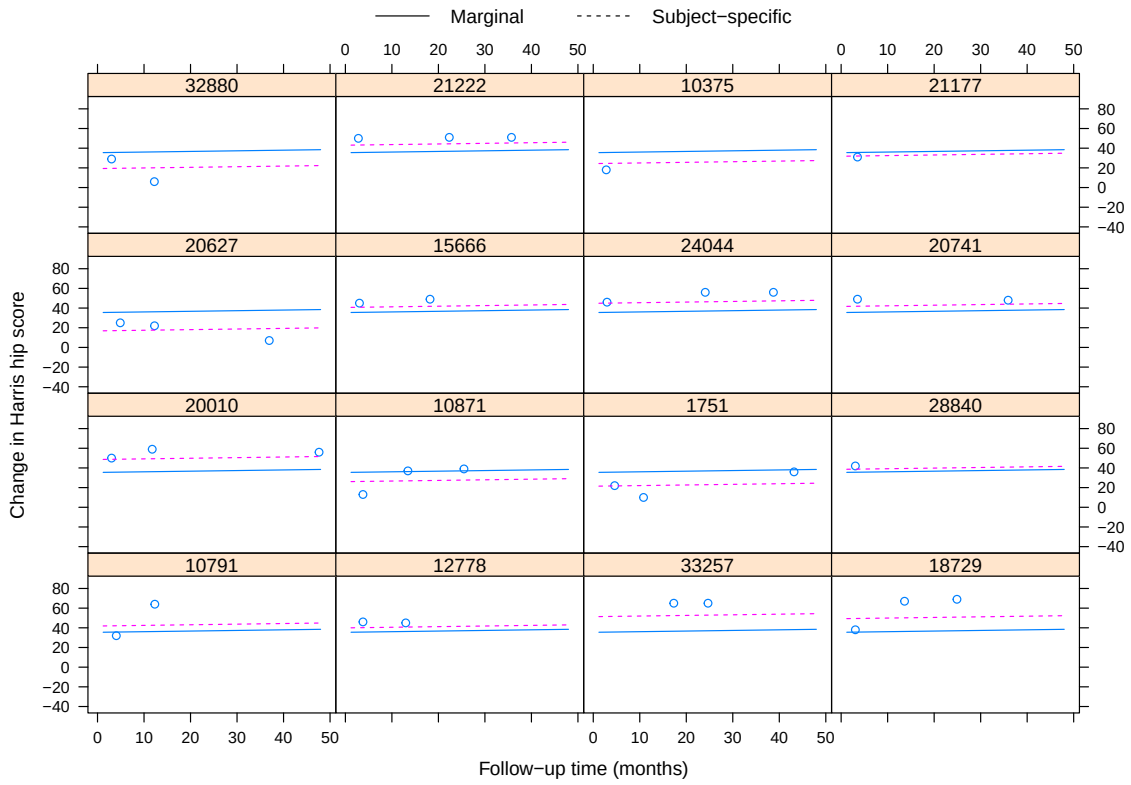

Figure 5.8: Observed and predicted change in HHS for 16 random patients using the optimal model (Equation 5.1). The blue points show the observed change in HHS for each patient. The blue lines show the population level prediction and the pink dashed lines show the individual level prediction.

Figure 5.8 shows the observed and predicted change in HHS for 16 randomly selected patients. The x-axis is the follow-up time and the y-axis is the change in HHS. The blue points show the observed change in HHS for that patient. The blue lines shows the population level prediction and the pink dashed lines shows the individual level prediction, both using the optimal model.

The predicted population means, shown in the Figure 5.8, increase very slightly over time. Due to our assumed structure of the model, the population means are shifted for individual patients by subject specific random intercepts. Therefore, we see that the

intercepts of the individual profiles are different across subjects but the slope remains the same. For some patients the predicted individual profiles deviate from the observed measurements. For example, subjects 20627 shows a decrease in the observed values over time, while the predicted individual patterns suggest an increase in the change in HHS over time. Despite this, Figure 5.8 shows that individual level prediction can give an accurate representation of a specific patient's predicted change in HHS over time.

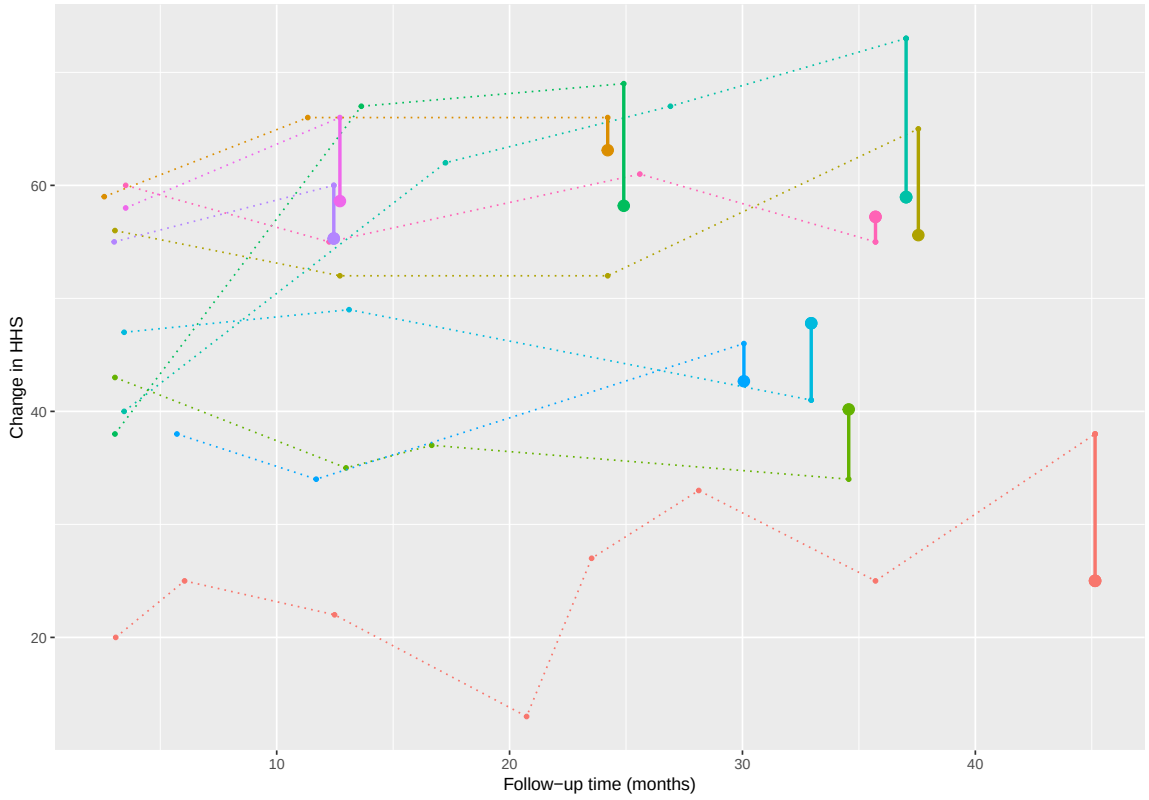

Figure 5.9: Plot of the observed and predicted change in HHS for 10 randomly selected patients last observation, given that we know the change in HHS of earlier observations. The x-axis is the follow-up time and the y-axis is the change in HHS. The dotted lines show the observed change in HHS trajectories. The smaller points are the observed change in HHS. The larger points are the predicted change in HHS for the patient's last observation. The solid line shows the difference between the predicted and the observed change in HHS for the patient's last observation.

Figure 5.9 is a plot of the observed and predicted change in HHS for 10 randomly

selected patients final observation, given that we know the change in HHS of their earlier observations. The x-axis is the follow-up time and the y-axis is the change in HHS. The dotted lines show the observed change in HHS trajectories. The smaller points are the observed change in HHS. The larger points are the predicted change in HHS for the patient's last observation. The solid line shows the difference between the predicted and the observed change in HHS for the patient's last observation.

This shows that for most patients, within this figure, the predictions are fairly accurate, as shown by the length of the solid lines. The predictions are particularly accurate for patients whose HHS does not deviate greatly over time. This is a reflection of the use of random intercepts. Since we do not include random slopes in this model and the only explanatory variable within this model is follow-up time, it does not predict patients with variable HHSs well. Due to the variability of many patient's HHS trajectories within the dataset this is a considerable limitation.

We can also predict a new patient's HHS trajectory after surgery, based completely on their baseline characteristics. Or in other words, we can predict the change in post-operative patient pain and function based solely on follow-up time. To do this we look at two methods. One method is to use the BLUPs, of the random effects, from other patients within the estimation model frame to predict the HHS for a new patient. The other method is comparing the prediction of a new patient using all possible BLUPs, of the random effects, within the model. These methods are explained in more detail in Section 4.2.

Figure 5.10 shows the predicted change in HHS trajectory. The change in HHS is on the y-axis and the follow-up time is on the x-axis. In this case, we are using 12 random observations to determine the random effect BLUPs for individual level prediction. The solid black line denotes predicted individual change in HHS. The red line denotes the

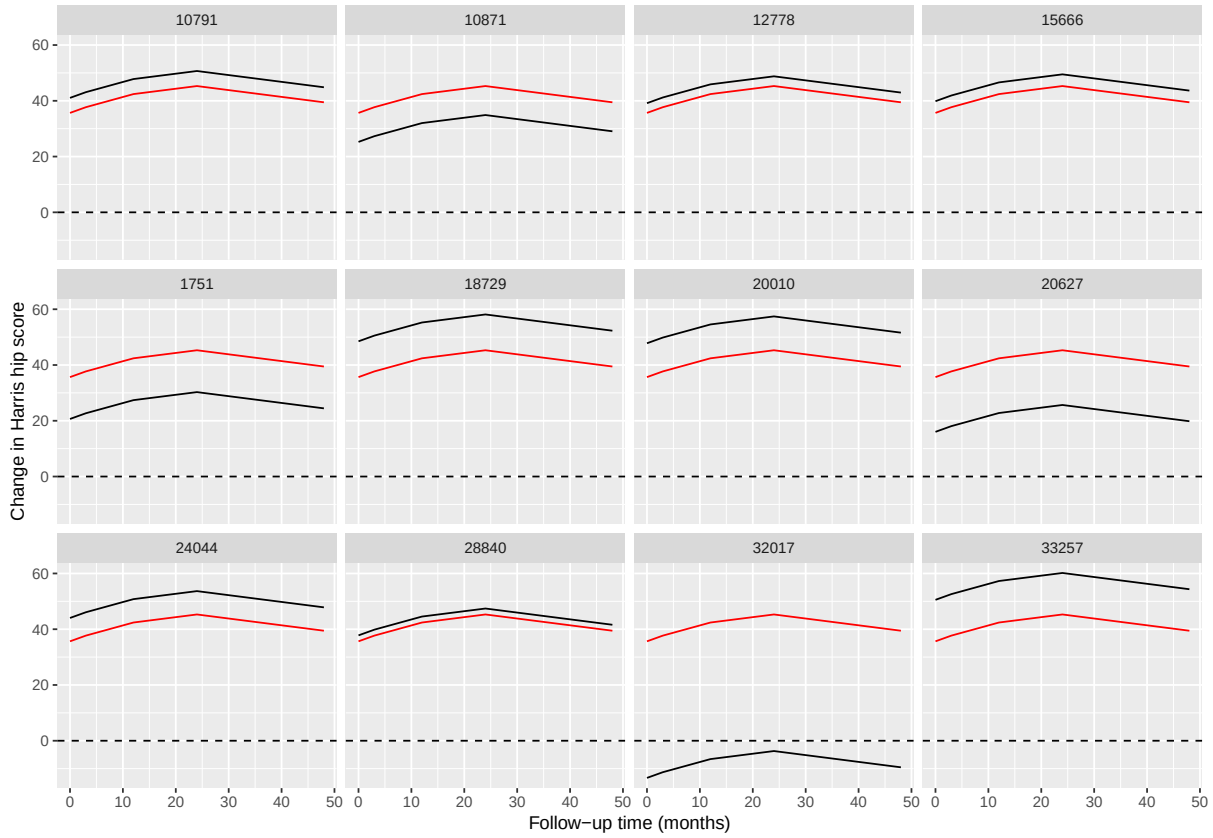

Figure 5.10: Plot of the predicted change in HHS trajectories against follow-up time using the estimated parameters of 16 random patients. The solid black line denotes predicted individual change in HHS. The red line denotes the predicted change in HHS trajectory for an average patient. The dashed line indicates where the change in HHS is zero.

predicted change in HHS trajectory for an average patient. The dashed line indicates where the change in HHS is zero. From Figure 5.10 we see that there is consistently a predicted decrease in the change in HHS from 24 months to 48 months. Also we see a predicted increase in the change in HHS between 0 and 24 months, with the fastest improvement occurring between 0 and 12 months.

Similarly to Chapter 4 we can address the variability in predictions that occurs, depending on how we select the appropriate estimators for the individual random effects, we can look at the variation in prediction across all patients. This shows us how the

prediction varies if we use all available BLUPs of the random effects and apply them to a new patient.

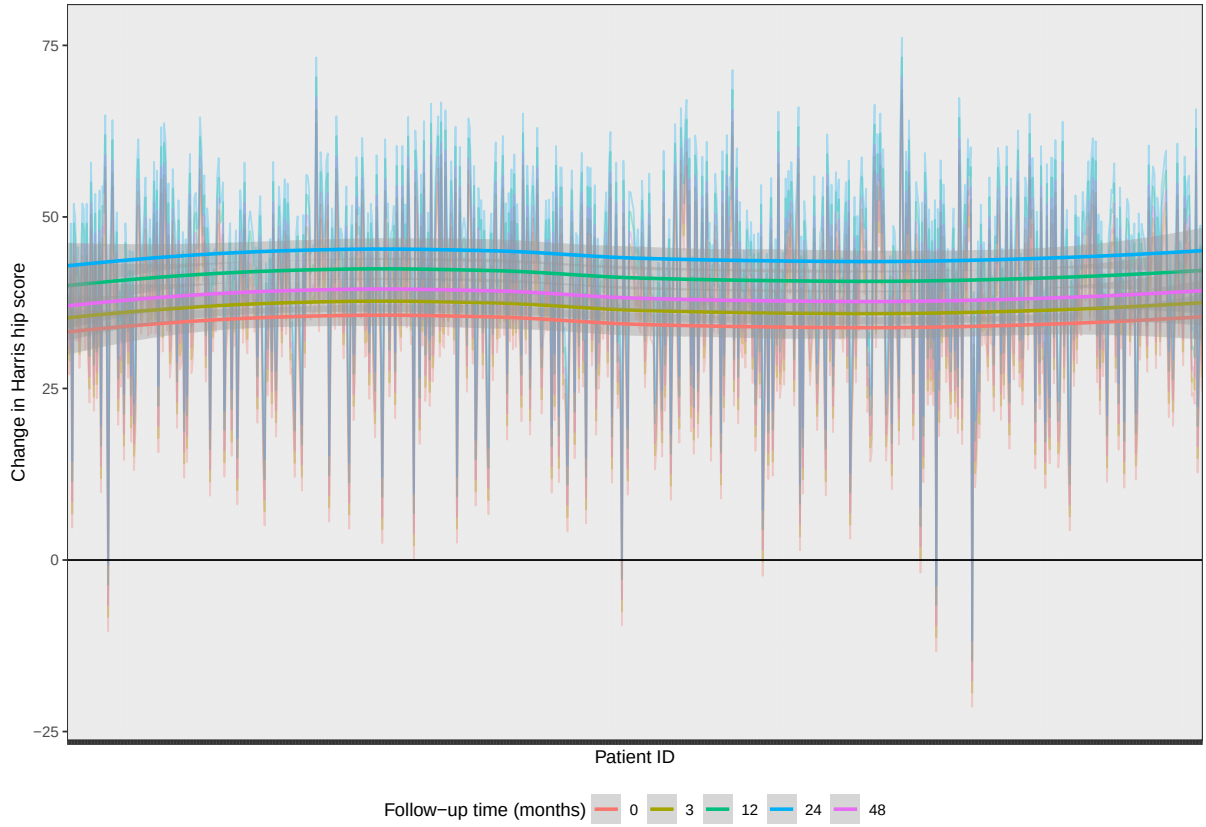

Figure 5.11: Plot of the predicted change in HHS at 0, 3, 12, 24 and 48 months after surgery. The x-axis is the patient ID. Each individual patient's BLUPs for the random effects is used but they are assumed to have the characteristics of an average patient. The coloured smoother lines show the average change in HHS for patients at each follow-up time. The solid black line indicates where the change in HHS is zero (n = 779).

Figure 5.11 displays the predicted change in HHS at 0, 3, 12, 24 and 48 months after surgery. The x-axis is the patient ID. Each individual patient's BLUPs for the random effects is used but they are assumed to have the characteristics of an average patient. The coloured smoother lines show the average change in HHS for patients at each follow-up time. The solid black line indicates where the change in HHS is zero. Fortunately, this

shows that the worst change in HHS is predicted to occur at 0 months. Likewise the best predicted change in HHS appears to occur at 24 months. Unfortunately, it appears as though the change in HHS lowers after 24 months, as the predicted change in HHS at 48 months is similar to 3 months after surgery. Again this could be because patients who recover well from the surgery have less follow-up measurements after 24 months, creating a bias in the predicted change in HHS at 48 months.

Figure 5.11 shows that the effect of the intercept on each patient simultaneously affects our predicted value. This results in the multiple lines for predicted values across all patients not being parallel, instead they resemble noise.

Throughout this chapter we have looked at determining an optimal mixed-effects model for predicting the change in HHS after THR. It is clear that most patients are predicted to see improvements in pain and function at 24 months after surgery. Nevertheless, we have seen that it is also optimal to have at least one HHS measurement to aid in prediction, due to the effect of the intercept, though this is not always practical. Within the dataset we also have access to patient survival information. This information could help inform the predictions of HHS after surgery. Therefore, in the next chapter we will explore survival analysis within the context of the THR dataset.

Chapter 6

Survival Analysis

The total hip replacement (THR) dataset that we received from the Royal Adelaide Hospital (RAH) included, among patient and operation details, the survival outcomes for THR patients. Though the primary focus of my research is analysing the longitudinal pain and function data, survival analysis can still provide crucial insights. Specifically, joint modelling of the survival and longitudinal outcomes has the potential to improve postoperative patient pain and function prediction. This will be discussed further in Chapter 7.

Firstly, it is important to study the survival times and the factors that influence them before exploring joint modelling. This chapter includes preliminary survival analysis to help us understand the nature of survival outcomes within this dataset, to prepare for further modelling. Therefore, the ultimate aim is to pursue joint longitudinal and survival modelling that will build on the survival analysis within this chapter.

Survival analysis is the study of survival times and the factors that influence them [20]. In this case, we will use survival analysis to analyse the time until an event occurs, the event being death. Within the dataset we have a variable that denotes a patient's

survival status (either alive or deceased) and another variable that states a patient's date of death (DOD), if relevant. Therefore, we can calculate the time between the operation date and the DOD. This difference in time can be used as the time until the event occurs within survival modelling.

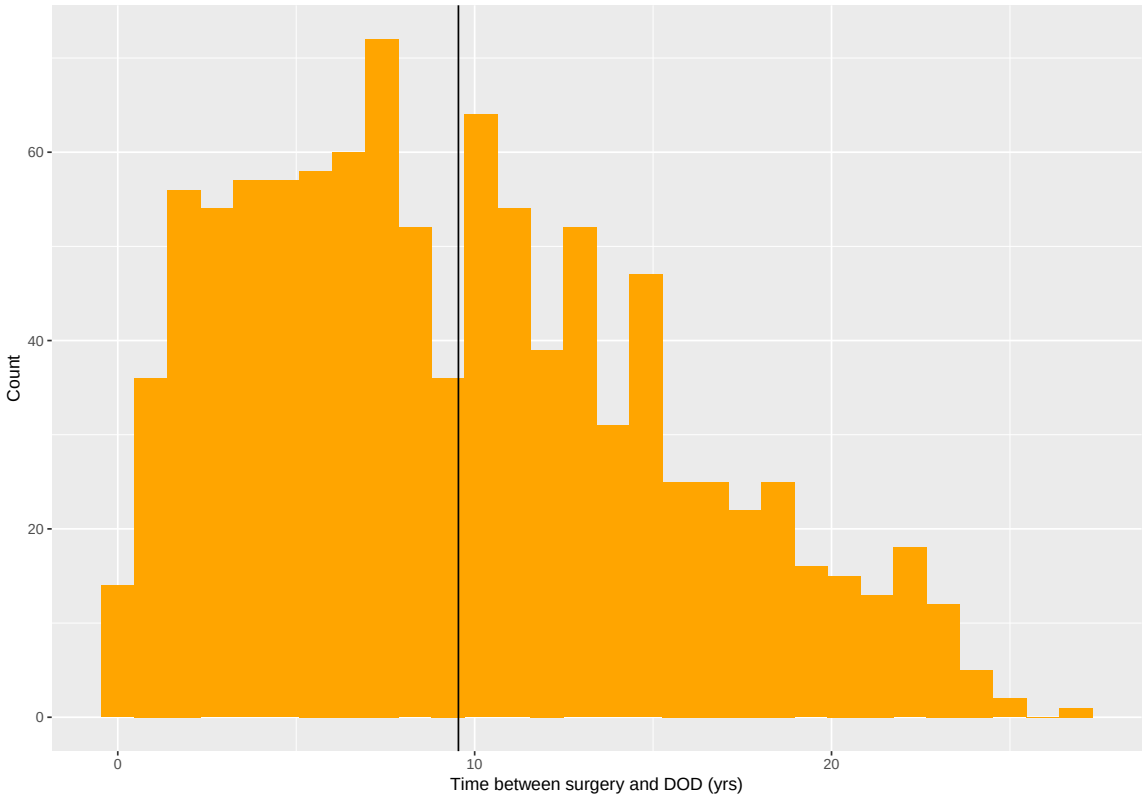

Figure 6.1: Histogram of the time between surgery and DOD for patients who are deceased. The solid black line refers to the mean time between surgery and DOD, which is 9.5 years (n = 2280).

Figure 6.1 displays a histogram of the time between surgery and DOD for patients who are deceased. The solid black line refers to the mean time between surgery and DOD (9.5 years). The distribution appears to be right skewed with the majority of the DODs occurring within 15 years after THR surgery. This is potentially indicative of patients losing contact with the hospital as more time passes after surgery. More likely

it is reflective of the older demographic that undergoes THR surgery and therefore only live for around 15-20 years after surgery.

Patient Demographics	**Survived**	**Deceased**
n of patients	1244	1036
Mean (SD) age (yrs)	63.2 (12.7)	71.6 (9.8)
Mean (SD) weight (kgs)	81.1 (18.4)	75.6 (16.5)
Mean (SD) height (cms)	166.0 (12.9)	165.0 (11.2)
Female (n,%)	734 (59.0)	607 (58.6)
Male (n,%)	510 (41.0)	429 (41.4)
Diagnosis (n,%)		
Osteoarthritis	920 (74.0)	814 (78.6)
AVN	112 (9.8)	70 (6.8)
Rheumatoid arthritis	27 (2.2)	30 (2.9)
Number of Comorbidities (n,%)		
None	402 (32.3)	303 (29.2)
One	389 (31.2)	355 (34.3)
Two	258 (20.7)	239 (23.1)
Three	145 (11.7)	95 (9.2)
Four or more	50 (4.0)	44 (4.2)

Table 6.1: Summary of the distribution of demographic variables for patients separated by survival status (n = 2280).

Table 6.1 shows a summary of the distribution of demographic variables for patients, separated by survival status (alive or deceased). It is clear that there is large proportion of patients within the dataset that are deceased (1036 patients). Therefore, we want to determine if the profile of these patients is significantly different to the profile of patients who are alive. We see that the mean age of deceased patients is greater than patients who are alive and has a smaller standard deviation. There is also a larger proportion of patients with one or two comorbidities within the subset of deceased patients when compared to patients that are alive. Similarly, there is a higher proportion of patients

with no comorbidities for patients that survived, compared to those that are deceased. Interestingly, the mean weight of patients who survived is 81.1 kg, which is higher than the mean weight of patients who are deceased (75.6 kg). Nevertheless, overall the distribution of the demographic variables appears fairly similar across both survival statuses.

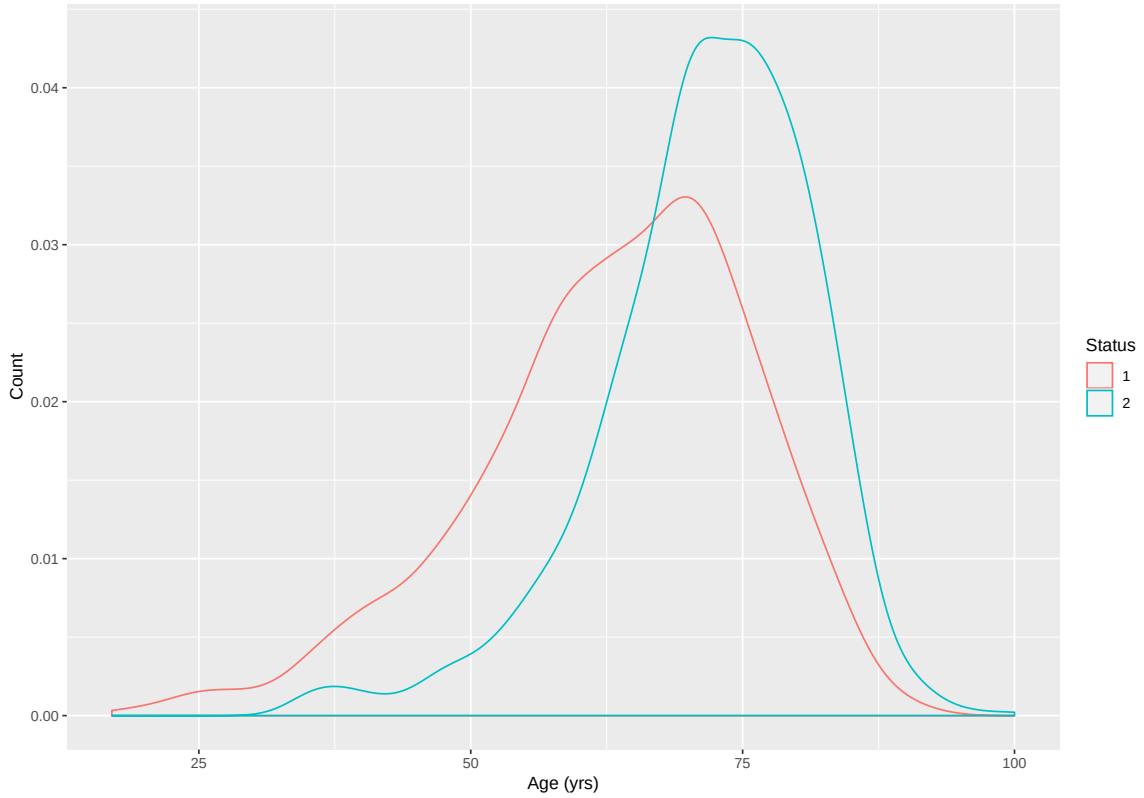

Figure 6.2: Density plot of the age of THR patients separated by survival status. 1 indicates that a patient is alive and 2 indicates that a patient is deceased (n = 2280).

For variables that do show differences in distribution across the two survival groups (alive and deceased) we can explore these differences further. Figure 6.2 shows the density plots of the age of THR patients separated by survival status. 1 indicates that a patient is alive and 2 indicates that a patient is deceased. The distribution of age for deceased patients has a stronger left skew than for patients who are alive. This means that patients who are deceased are more likely to have been older at the point of surgery. This is

intuitive, as we would expect that patients who were younger at the point of surgery are more likely to still be alive.

With this understanding of the survival outcomes within our dataset we can begin to explore survival analysis. In our case, we are analysing the time between when a patient has THR surgery and their DOD. But as we see in Table 6.1 some patients are still alive at the end of the follow-up period, specifically the end of 2013. For these patients we know that they survived up until the end of the follow-up period but we do not know how much longer they survived after that time. These times are said to be right-censored, at the end of 2013, which is the most common type of censoring within survival analysis [20].

Survival analysis methods depend on specifying a survival distribution. This can be done through a survival function and a hazard function. The survival function is the probability of surviving up until a time t and it can be written as,

$$S(t) = P(T > t), \quad 0 < t < \infty.$$

The hazard function is the instantaneous failure rate and is often used to define the survival function. More specifically, the hazard function is the probability that, given a patient is still alive at time t, the event, in this case death, will occur within the next time point. This can be written as,

$$h(t) = \lim_{\delta \to 0} \frac{P(t < T < t + \delta | T > t)}{\delta}$$

There are numerous distributions that can be used to model survival data, making it a parametric model. Nevertheless, it can be hard to determine which parametric model is appropriate, or if there are any. Therefore, non-parametric methods are often used

due to their flexibility. The most common of these is the Kaplan-Meier estimator derived by Kaplan and Meier [15]. It is the product over the failure times of the conditional probabilities of surviving to the next failure time. This can be written as,

$$\hat{S}(t) = \prod_{s<t} \left(1 - \frac{d_i}{r_i}\right),$$

where there are d_i failures among r_i individuals in view at t. Therefore, the Kaplan-Meier estimator tells us the estimated probability of a subject surviving beyond a certain point, t [9].

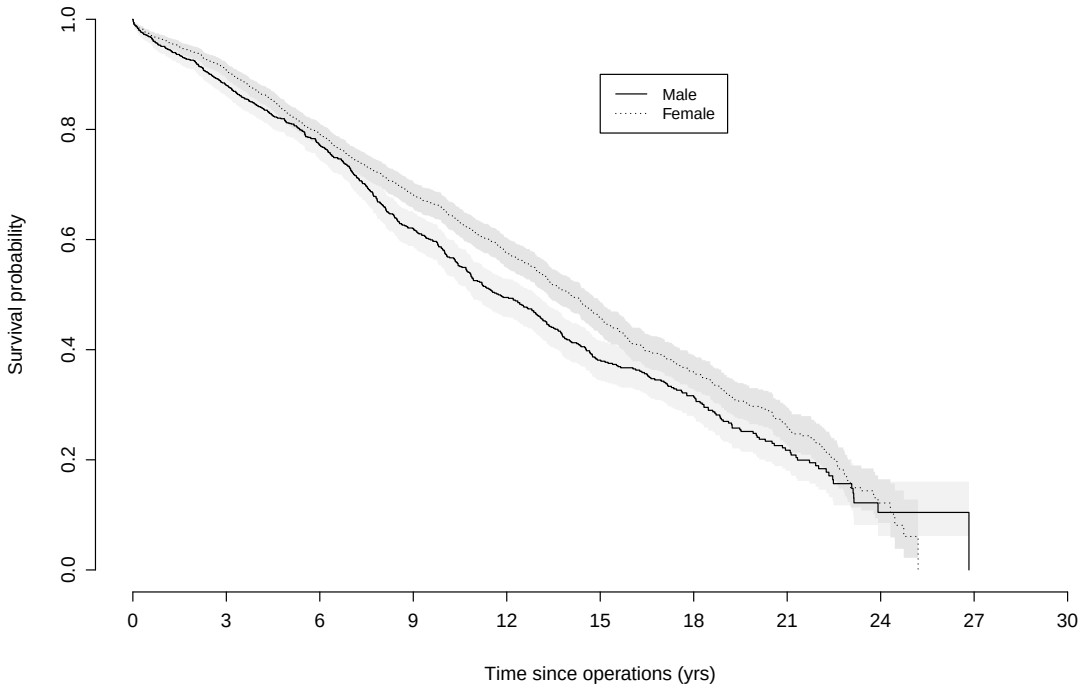

Figure 6.3: Comparison of the Kaplan-Meier survival curves by gender; the dashed line represents females and the solid line represents males. The grey areas are 95% confidence intervals (n = 2280).

Figure 6.3 is a comparison of the estimated Kaplan-Meier survival curves by gender. The dashed line represents females and the solid line represents males. The grey areas are 95% confidence intervals. We can also determine the median lifetime, which for females is 14 years and for males 11.7 years. Therefore, females are associated with a higher survival probability. This is clear from Figure 6.3 as the dashed line lies above the solid line up until about 24 years. At which point males are associated with a higher survival probability.

Therefore, the Kaplan-Meier estimator helps us compare two survival distributions, without assuming a parametric form. Nevertheless, we want to extend this to include covariate information. A common method for doing this is the Cox proportional hazards regression model. To help us define this model we let T_i^* be the true failure time for the i-th individual and C_i be the censoring time. Then $T_i = \min(T_i^*, C_i)$ represents the observed failure time for the i-th patient. Therefore, the Cox proportional hazards model is defined as,

$$\begin{aligned} h_i(t) &= \lim_{dt \to 0} \frac{Pr(t \leq T^* < t + dt | T^* \geq t)}{dt} \\ &= h_0(t) \exp(\gamma^T w_i), \end{aligned}$$

where $t > 0$, w_i is the covariates associated with the hazard, γ is the vector of regression coefficients and $h_0(t)$ is the baseline hazard. This model investigates the effect of several variables on a time-specified event taking place (in this case death, where censoring is an incompletely observed death). This model assumes that the hazard ratio $\frac{h_i(t)}{h_0(t)}$ depends only on covariates, whose value is fixed during follow-up [9]. Therefore, this model assumes that the covariates are not time-dependent.

A time-dependent explanatory variable, which is a variable that is not necessarily constant through the whole study, may sometimes be of use within survival analysis. For example, if one wishes to examine the link between age and second surgery (if they are a bilateral patient) this would be complicated by the fact that study subjects move from their first surgery to their second at different times. So second surgery can be introduced in the statistical model as a time-dependent covariate. This is done by splitting each study subject into two observations, one for their first surgery and one for their second surgery [31].

Time-varying covariates can be added to Cox models by making w_i a function of time, $w_i(t)$. Therefore, the Cox proportional hazards model becomes,

$$h_i(t) = h_0(t) \exp(\gamma^T w_i(t)), t > 0.$$

There are two types of time-dependent covariates, called external or internal covariates. An external covariate is a covariate which is not effected by an event occurring at an earlier time. For example which season of the year we are in or any variable that is predetermined at the start of the study (eg. treatment). In contrast internal covariates are all covariates that are patient measurements, such as biomarkers [3]. For example, within the data we have measurements relating to a patient's Harris hip score (HHS) at different follow-up times. This is an an internal time-varying covariate because in the future if that patient undergoes a second THR their HHS will most likely be affected. Whereas the season of the year with be unaffected, making it an external covariate. Time-dependent Cox models are theoretically valid for external time-varying covariates. It is not always appropriate when considering biomarkers, such as HHS, which will be discussed in Chapter 7. At this point we are only looking at the survival outcomes in terms of the covariates, and the

HHS is being ignored.

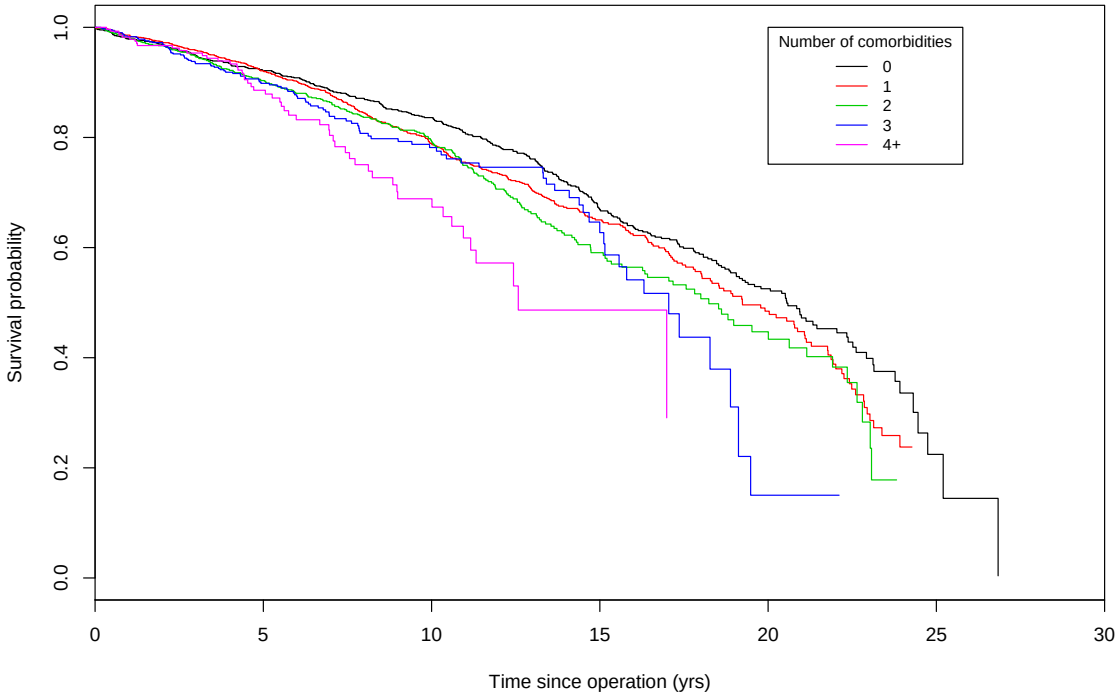

Figure 6.4: Comparison of the number of comorbidities for time-dependent Cox proportional hazards model with parameters: patient was on first surgery, had a revision, 67 years, had OA and was female (n = 2280).

Using a patient's bilateral surgery as a time-dependent covariate we can explore the difference in survival curves for patients that suffer from a varying number of comorbidities. Therefore, a simple time-dependent Cox proportional hazards model was fitted with the covariates age, gender, osteoarthritis (OA) and the number of comorbidities that they suffered from. The time-dependent covariate was whether a patient was on their first or second surgery.

Figure 6.4 shows the results from fitting a time-dependent Cox proportional hazards model plotted at different numbers of comorbidities. The variables are set to the following

values: patient was on first surgery, 67 years, had OA and was female. This figure indicates that suffering from more comorbidities is associated with a lower survival probability. Patients with no comorbidities are associated with a higher survival probability than patients with any other number of comorbidities.

Using this framework for the modelling of survival outcomes we can explore joint modelling. This will rely on the use of the Cox proportional hazards model. Likewise, the longitudinal outcomes analysis presented in Chapter 4 will also be crucial in implementing joint modelling. This is explored with the hypothesis that the inclusion of survival outcomes, jointly modelled with the longitudinal outcomes, will help inform the prediction of pain and function after surgery.

Chapter 7

Joint Modelling

In medical research, longitudinal and time-to-event information are often collected together. For example, in the total hip replacement (THR) dataset, patients who undergo a THR have follow-up Harris hip score (HHS) measurements and during that period death or revision can occur. Another example is a prostate cancer study, where death and metastasis can occur during the follow-up period [25]. While it is possible to analyse the time-to-event and longitudinal outcomes separately, this is not always adequate. There are situations were addressing both outcomes in one analysis is optimal. Sometimes it is more appropriate to address both of these outcomes within a joint model.

7.1 Joint Model

Joint modelling may be useful for a few reasons. Firstly, if we are interested in the association between the longitudinal and survival outcomes. For example, the association between the HHS measurements and the time until death. Another reason may be to include the effect of the time-dependent covariate measured with error in survival analysis.

This means we can include the effect of the HHS, measured with error, in the survival sub-model. Joint modelling can also be used to account for possible outcome dependant drop out in a longitudinal study. Therefore, it can account for the possible dependancy between HHS and survival status [5].

In previous chapters we have addressed both the time-to-event and longitudinal outcomes separately. In this chapter we will address the joint modelling of these outcomes. Joint models are a combination of both mixed-effects models and Cox models. Therefore, what follows is a brief reminder of mixed-effects models and Cox models, followed by the joint modelling framework.

7.1.1 Mixed-Effects Model

Mixed-effects models, addressed in more depth in Chapter 4, allow for random intercepts and slopes for each patient. This allows each patient within the study to have their own evolution in HHS over time. We let $y_i(t)$ denote the follow-up measurements for the i-th subject (i=1,. . .,n) at time t. These measurements are obtained at time points t_{ij}, where j=1,. . ., n_i. Therefore, a general mixed-effects model can be written as

$$y_i(t) = x_i(t)^T\beta + z_i(t)^T b_i + \epsilon_i(t),$$

$$b_i \sim N(0, D),$$

$$\epsilon_i \sim N(0, \sigma^2),$$

where β is the regression coefficients of the design matrix for the fixed effects, x_i^T, and z_i^T is the design matrix for the random effects, b_i. D is the covariance matrix of the random effects, $\epsilon_{ij}(t)$ are the error terms, with variance σ^2. The fixed effects relate to the

population effects while the random effects refer to the subject specific effects.

The parameters for this model are estimated by maximum or restricted maximum likelihood (ML and REML). The random effects are predicted by their conditional expectations given the data, this minimises the mean square prediction error.

7.1.2 Cox Model

Cox models, addressed in Chapter 6, use covariate information to determine the hazard of each subject. Let T_i^* be the true failure time for the i-th individual and C_i be the censoring time. Then $T_i = \min(T_i^*, C_i)$ represents the observed failure time for the i-th patient. The Cox proportional hazards regression model is written as

$$h_i(t) = \lim_{dt \to 0} \frac{Pr(t \leq T^* < t + dt | T^* \geq t)}{dt} = h_0(t) \exp(\gamma^T w_i), t > 0,$$

where w_i is the covariates associated with the hazard, γ is the vector of regression coefficients and $h_0(t)$ is the baseline hazard. This model assumes that the hazard ratio $\frac{h_i(t)}{h_0(t)}$ depends only on covariates, whose value is fixed during follow-up. Therefore, excluding time-varying covariates.

As discussed previously, time-varying covariates can be added to Cox models by making w_i a function of time, $w_i(t)$. Time-dependent Cox models are theoretically valid for external time-varying covariates. It is not always appropriate when considering biomarkers, such as HHS. This is due to an expectation that biomarkers will evolve relatively smoothly over time, which the time-dependent Cox model does not take into account. This creates bias in the estimated effect of the biomarker. This further motivates the use of joint modelling, as HHS is not an appropriate time-dependent covariate within a Cox model. In contrast, the HHS is more appropriately addressed in joint modelling.

As discussed earlier, the idea behind the joint model is to combine the continuous time-to-dropout process with a mixed-effects model for the longitudinal outcome. A commonly used joint model framework can be written as

$$\begin{cases} y_i(t) = x_i(t)^T\beta + z_i(t)^T b_i + \epsilon_i(t) \\ h_i(t) = h_0(t)\exp\left[\gamma^T w_i + \alpha\{x_i(t)^T\beta + z_i(t)^T b_i\}\right], \quad t > 0, \end{cases} \tag{7.1}$$

where α is the effect of the underlying longitudinal outcome to the risk for an event. Therefore, α is of particular interest as it describes how the covariates affect the relative hazard. This model assumes that the risk associated with the outcome dependent dropout is related to the unobserved value of the longitudinal outcome. It also assumes that the random effects account for the correlation between repeated measures and also the association between the longitudinal and survival processes [28]. This survival sub-model is more appropriate than a Cox model with a time-varying covariate as it takes into account the idea that HHS evolves smoothly over time. It is fair to assume that in reality a patient's pain and function would smoothly increase or decrease between visits, unless there was a trauma incident. In contrast, the time-dependent Cox model assumes that the HHS is constant between visits.

An important part of joint modelling is that repeated measurements and survival data are modelled simultaneously. By using the model's assumptions it predicts the values of $y_i(t)$ at all times and therefore estimates the association between the longitudinal and survival processes, α, while making allowance for the measurement error in the observed response [5].

The parameters of this joint model are estimated through maximum likelihood, and random effects are predicted by their conditional expectations given the data. There are

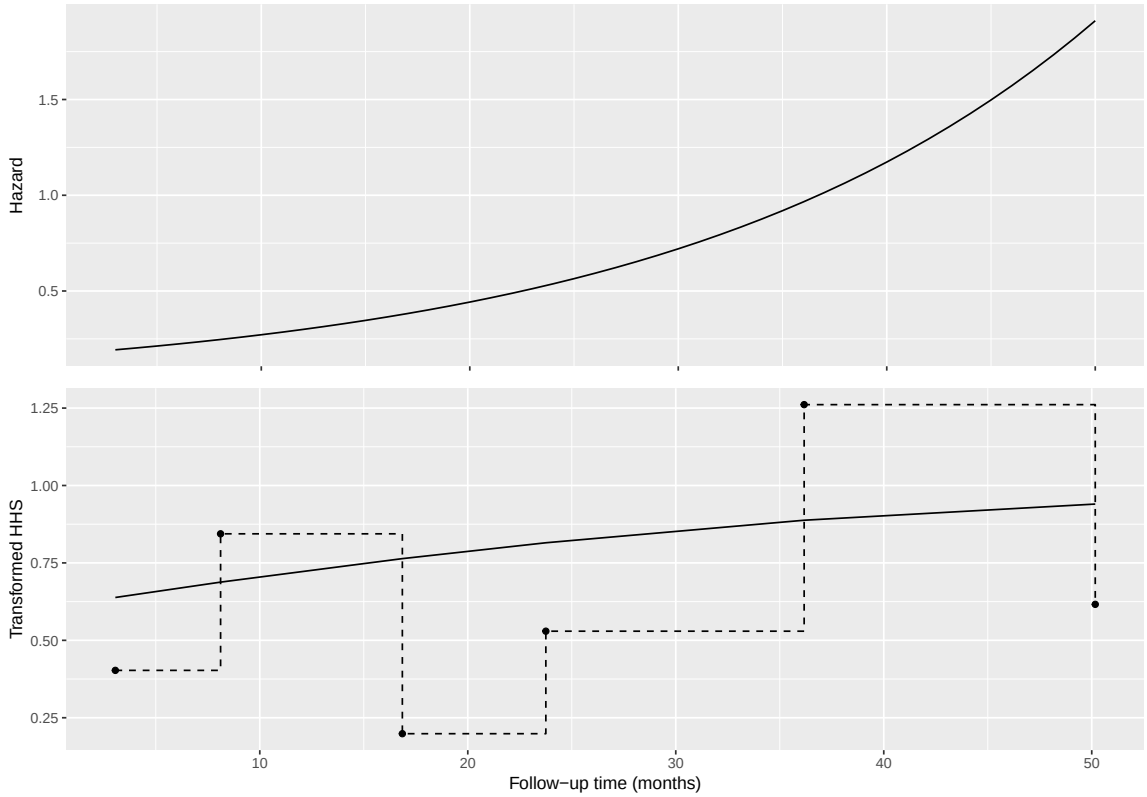

Figure 7.1: Graphical representation of joint models. Top: The hazard function for survival status. Bottom: the dashed line describes a time-dependent covariate as used in the time-dependent Cox model, and the solid line the mixed-effects model reconstruction of the transformed HHS trajectory.

different methods used to maximise the likelihood, one of which will be addressed later in this chapter.

The joint model outlined above uses mixed-effects modelling to estimate biomarkers that evolve smoothly over time. It then uses these estimated biomarkers within a time-dependent Cox model in the place of the observed biomarkers. Figure 7.1 displays this idea as the bottom panel compares the estimated evolution of the transformed HHS using mixed-effects models (solid line) and compares it to the observed transformed HHS as a time-dependent covariate (dashed line).

This model is often used in the context of improving estimation of survival outcomes. Our study is ultimately concerned with how to account for outcome dependent drop out, and it's effect on the longitudinal outcome. This is because our primary concern is still estimating postoperative patient pain and function. Nevertheless, this commonly used joint model framework is a suitable foundation to begin joint modelling the HHS and survival outcomes.

With joint modelling, as with all modelling, comes limitations. The main disadvantage is computation cost and effort and the relatively little software in place to compute these models [5]. This problem is less significant when dealing with smaller datasets with less observations and fewer random effects. There are packages, such as JM [27], that are being developed within Rstudio that are able to compute some of these models. In this chapter we will use this package to perform joint modelling.

7.2 Joint Model Sub-study

Now we explore joint modelling within the context of our THR dataset. Firstly, we created a subset of patients with a valid survival status (whether they were deceased) and a valid date of death, if they were deceased. Furthermore, these patients all needed at least one valid follow-up HHS measurement. After these restrictions the number of patients in the sub-study was 1850. Using t-tests and comparing proportions we could determine that the distribution of gender, age and survival status was approximately the same in the sub-study as in the whole dataset (2280 patients). The response variable, HHS, is transformed using the logit transformation to better satisfy the assumptions of the model, as discussed in Chapter 4.

The longitudinal sub-model of the joint model is a random intercept and slope model,

as this was the optimal random effects structure for the mixed-effects model explored in Chapter 4. Each patient has their own random intercept and random slope. Therefore, we assume there is a relationship between HHS and follow-up time that is different for each patient.

This joint model was fit using the JM package in R [26]. This uses maximum likelihood but requires specification regarding the type of sub-model to be fitted and the method of numerical integration. The type of sub-model to be fitted is a relative risk model with a piecewise-constant baseline risk function. The relative risk model is in the form

$$h_i(t|M_i(t), w_i) = h_0(t) \exp\{\gamma^T w_i + \alpha m_i(t)\},$$

where $M_i(t) = m_i(u), 0 \leq u < t$ is the history of the true unobserved longitudinal process up to time point t, $h_0(.)$ is the baseline risk function and w_i is the vector of baseline covariates with a vector of regression coefficients γ [27]. The piecewise-constant baseline risk function is written as

$$h_0(t) = \sum_{q=1}^{Q} \xi_q I(v_{q-1} < t \leq v_q),$$

where $0 = v_0 < v_1 < \ldots < v_Q$ is a split of the time scale, with v_Q being larger than the maximum observed time, and ξ_q is the value of the hazard in the interval $(v_{q-1}, v_q]$.

To use maximum likelihood estimation we need to maximise the log-likelihood of the joint distribution. To define the log-likelihood of the joint distribution we define the time-to-event and longitudinal outcomes as T_i, δ_i, y_i and recall the assumption that the random effects, b_i, underlie both the longitudinal and survival process. This creates conditional independence as the random effects account for both the association between

the longitudinal and event outcomes and the correlation between repeated measurements in the longitudinal process. Since conditional independence implies that

$$p(A, B|C) = p(A|C)p(B|C).$$

In the context of the probability density function for the joint model we get

$$p(T_i, \delta_i, y_i|b_i; \alpha) = p(T_i, \delta_i|b_i; \alpha)p(y_i|b_i; \delta),$$

$$p(y_i|b_i; \alpha) = \prod_j p(y_i(t_{ij})|b_i; \alpha),$$

where $\alpha = (\alpha_t^T, \alpha_y^T, \alpha_b^T)^T$ is a vector of parameters and the event time outcome, longitudinal outcome and random effects covariance matrix respectively. The variable y_i is the vector of longitudinal responses of the i-th subject. Therefore, using these assumptions we can derive the joint log-likelihood contribution for the i-th subject.

$$\begin{aligned}
\log p(T_i, \delta_i, y_i; \alpha) &= \log \int p(T_i, \delta_i, y_i, b_i; \alpha) db_i \\
&= \log \int p(T_i, \delta_i, y_i|b_i; \alpha)p(b_i; \alpha_b) db_i \\
&= \log \int p(T_i, \delta_i|b_i; \alpha_t, \gamma) \left[\prod_j p\{y_i(t_{ij})|b_i; \alpha_y\}\right] p(b_i; \alpha_b) db_i
\end{aligned}$$

This is the product of the the likelihood of the survival, the univariate normal density for the longitudinal response and the multivariate normal density for random effects, respectively. This is computationally difficult to solve because the integral, with respect to the random effects and survival function, does not have an analytical solution (except in a few cases) [27]. To solve this integral one approximation that can be used is the

Gauss-Hermite integration rule. This rule approximates as shown

$$\int_{-\infty}^{+\infty} \exp\{-x^2\} f(x) dx \approx \sum_{i=1}^{n} w_i f(x_i),$$

where n is the number of sample points, x_i are the roots of the Hermite polynomial $H_n(x)$ and the weightings are $w_i = \frac{2^{n-1} n! \sqrt{\pi}}{n^2 [H_{n-1}(x_i)]^2}$. There are other methods of approximation that can be used such as the Laplace approximation [29] though in this case it is not as appropriate as we do not have high-dimensional random effects structures.

Variable	Estimate	SE	p-value	HR
logit(HHS)	-0.12	0.002	<0.001	0.891
Baseline age (years)	0.05	0.047	<0.001	1.053
Gender	0.36	0.018	<0.001	1.426

Table 7.1: Results for the Cox model, where HHS is only the last observed HHS measurement (n = 1850).

To illustrate the advantage of using a joint modelling approach, first a Cox model was fitted using only the last observed HHS measurement, baseline age and gender as predictors of survival status. The results of this model are shown in Table 7.1. Then a joint model was specified using the framework previously defined (Equation 8.1). The longitudinal sub-model included the predictor variables baseline age, gender, and follow-up time. Moreover, it included random intercepts and slopes, due to the random effects structure of the optimal model in Chapter 4. Though in further exploration we could include different baseline covariates for the Cox model and mixed-effects model, depending on what is of interest. At this stage we simply kept the covariates the same across both models.

Results from this joint model are shown in Table 7.2 and compared to the results of

Variable	Estimate	SE	p-value	
Longitudinal sub-model				
Intercept	1.9841	0.1665	<0.001	
Baseline age (years)	-0.0123	0.0024	<0.001	
Follow-up time	0.0107	0.0010	<0.001	
(Follow-up time)^2	-0.0001	<0.001	<0.001	
Gender	0.3581	0.0572	<0.001	
Variable	Estimate	SE	p-value	HR
Survival sub-model				
logit(HHS)	-0.3117	0.0480	<0.001	0.7322
Baseline age (years)	0.0776	0.0045	<0.001	1.0807
Gender	0.5613	0.0784	<0.001	1.7530

Table 7.2: Results for the joint modelling analysis of the THR dataset. For both sub-models, estimated parameters, standard errors (SE) and p-values are reported (n = 1850).

the Cox model shown in Table 7.1. Both the parameter estimates and the p-values are recorded. In the Cox model for mortality (Table 7.1), older patient age, lower transformed HHS and gender are associated with an increased death hazard. Using the joint model for mortality (Table 7.2), the same predictor variables remain significantly associated with an increased death hazard. Nevertheless, we see that in the joint model a higher coefficient is estimated for gender when compared to the Cox model. This implies that the joint model estimates being male has a larger effect on a patient's survival probability than in the Cox model. By using repeated HHS measurements instead of only the last HHS measurement the model outcomes can change and different conclusions can be drawn. Therefore, accounting for the idea that pain and function are not independent from death, through joint modelling, may allow for more appropriate insights into patient pain and

function after THR.

We again note that this particular joint modelling framework is motivated by the time-to-event point of view. It incorporates the HHS as a time-dependent covariate measured with error in a survival model. Nevertheless, incorporating the association between the survival and longitudinal outcomes still gives us greater insight into the trajectory of patient pain and function after THR. This is crucial as we are more motivated by how best to incorporate the time-to-event information into a longitudinal analysis. Therefore, this joint modelling framework is still informative and will be the framework explored further in this chapter. Though a different approach could be explored in future to further focus on the behaviour of the HHS after surgery.

7.2.1 Diagnostics

To assess the appropriateness of the joint model (Table 7.2) we need to explore the diagnostics. Firstly we look at the residuals for the longitudinal sub-model of the joint model. Due to the nature of the sub-model, it has both subject specific residuals and population averaged residuals. The first of these aims to validate the homoscedasticity and normality assumptions through determining the residuals of the hierarchical model,

$$\begin{cases} y_i(t) = x_i^T(t)\beta + z_i^T(t)b_i + \epsilon_i, \\ \\ b_i \sim N(0, D), \epsilon_i \sim N(0, \sigma^2), \end{cases}$$

with residuals

$$r_i^{ys}(t) = \{y_i(t) - x_i^T(t)\hat{\beta} - z_i^T(t)\hat{b_i}\}$$

that predict the conditional errors $\epsilon_i(t)$. In contrast, the marginal residuals investigate misspecification of the mean structure $x_i^T(t)\beta$ and validate the assumptions for the within-

subject covariance structure. These residuals are based on the marginal model,

$$\begin{cases} y_i(t) = x_i^T(t)\beta + \epsilon_i^*, \\ \\ \epsilon_i^* \sim N(0, Z_i D Z_i^T + \sigma^2 I_{n_i}), \end{cases}$$

with residuals

$$r_i^{ysm} = y_i - X_i\hat{\beta}$$

that predict the marginal error [28]. Both the subject-specific and the marginal residuals are used to validate the assumptions of the longitudinal sub-model within the joint model.

Figure 7.2 shows the subject specific residuals versus the fitted values, the Q-Q plot of the subject specific residuals, and the marginal survival and cumulative risk functions for the event process. The residuals versus fitted plot appears to validate the model's assumptions due to the random scatter of the residuals and their centering around zero. The deviations that form a sharp line in the top right of this plot are again HHS observations that are 100. Likewise, the normal Q-Q plot displays this poor behaviour at the boundary of the HHS with the nonlinearity of the upper tail. Overall, without the poor behaviour at the boundary of the HHS, these residuals support the model's assumptions.

Figure 7.3 shows the marginal standardised residuals versus fitted values for the longitudinal outcome of the THR dataset, which is the transformed HHS. There is a fitted loess curve that suggests that there are more positive residuals for smaller fitted values. Likewise, there appears to be more negative residuals for larger fitted values. This may be due to nonrandom dropout in the HHS measurements caused by death. This implies that patients who have a lower HHS may have a higher death rate. This may be because the residuals of the smaller fitted values are based on patients with, mostly, good HHSs (> 70), creating a systematic trend in the marginal residuals.

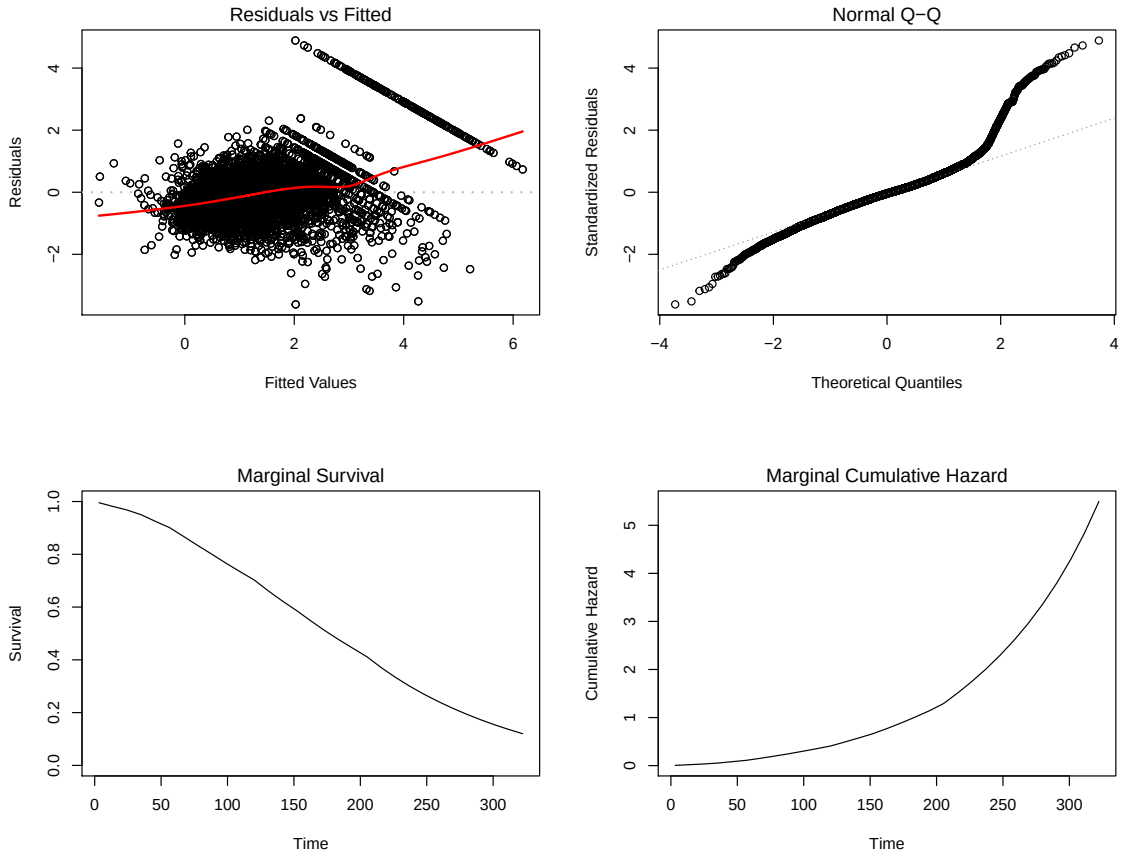

Figure 7.2: Diagnostic plots for the joint model fitted to the THR dataset. Included are the subject specific residuals versus the fitted values, the Q-Q plot of the subject specific residuals, and the marginal survival and cumulative risk functions for the event process (n = 1850).

The residuals for the survival sub-model also need to be explored. Martingale residuals are a standard type of residual used to address survival models. This residual is the difference between the observed number of events, in this case death, for the i-th patient until time t, and the expected number of events, using the fitted model. Therefore, it is similar to the residual definitions used previously, where the THR data is defined by the model and an error term.

Figure 7.4 shows the martingale residuals versus the subject specific fitted values of

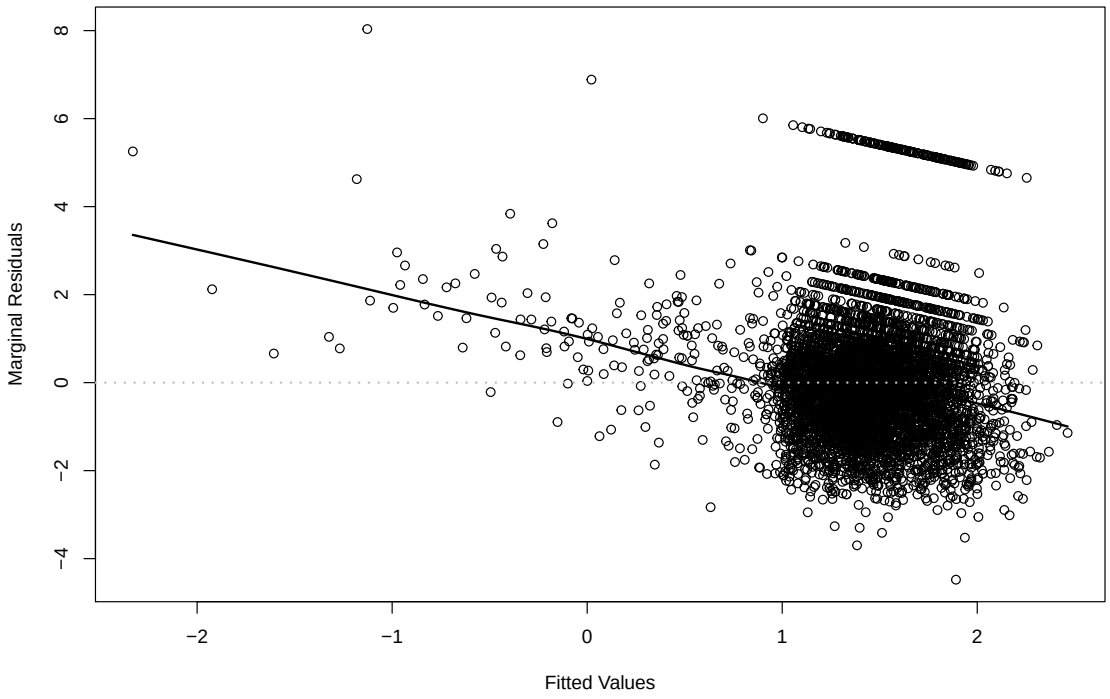

Figure 7.3: Marginal standardised residuals versus fitted values for the longitudinal outcome of the THR dataset (n = 1850).

the longitudinal outcome for the THR dataset. The grey solid line denotes the fit of the loess smoother. It is clear that for smaller fitted values of the longitudinal outcome the loess smoother is further from zero. To further investigate this potential systematic trend we condition on gender and then inspect the martingale residuals.

Figure 7.5 shows the martingale residuals versus the subject specific fitted values of the longitudinal outcome separated by gender for the THR dataset. The grey solid lines denotes the fit of the loess smoother. There still appears to be deviations in the loess smoother from zero but these appear to be consistent across both genders. Therefore, this deviation may be reflective of poor model behaviour for patient's with low fitted HHS values, potentially due to limited data in this area.

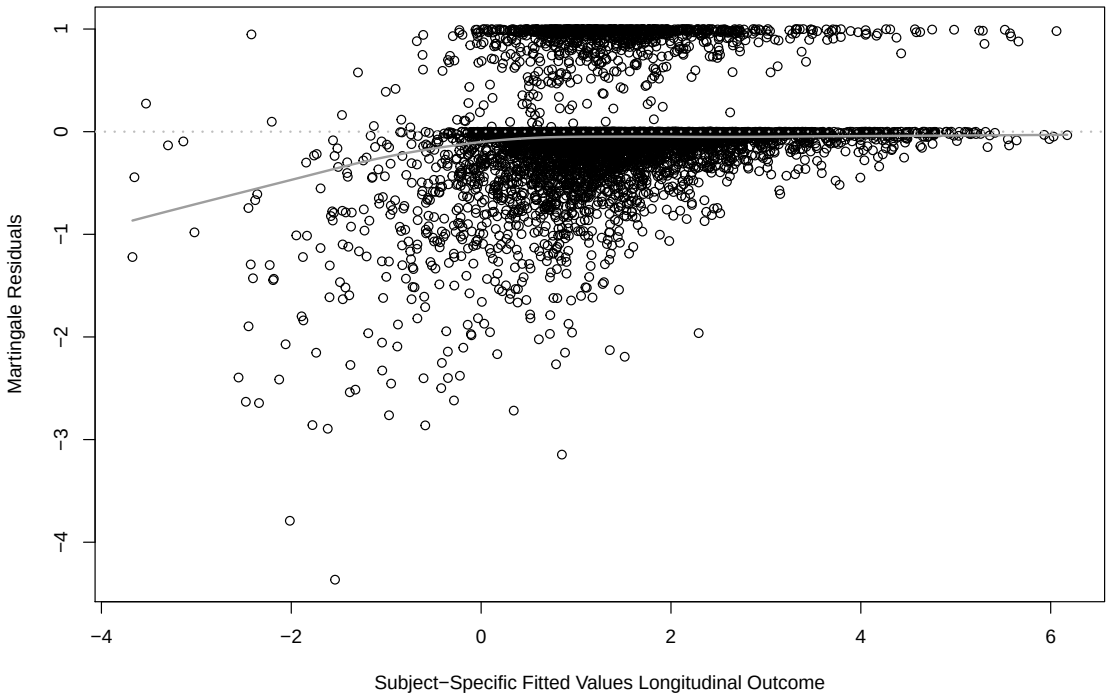

Figure 7.4: Martingale residuals versus the subject specific fitted values of the longitudinal outcome for the THR dataset. The grey solid line denotes the fit of the loess smoother (n = 1850).

Another type of residual that can be used to help inform us about the survival sub-model is the Cox-Snell residuals. These residuals are the estimated cumulative risk function evaluation for each patient at the observed event time T_i. Therefore, the Kaplan-Meier estimate of the survival function of the Cox-Snell residuals is plotted [28].

Figure 7.6 shows the Cox-Snell residuals for the survival sub-model. The black solid line denotes the Kaplan-Meier estimate of the survival function of the residuals (with the dashed lines corresponding to the 95% point-wise confidence intervals), and the grey solid line, the survival function of the unit exponential distribution. We compare the Kaplan-Meier estimate with the survival function of the unit exponential distribution as this is

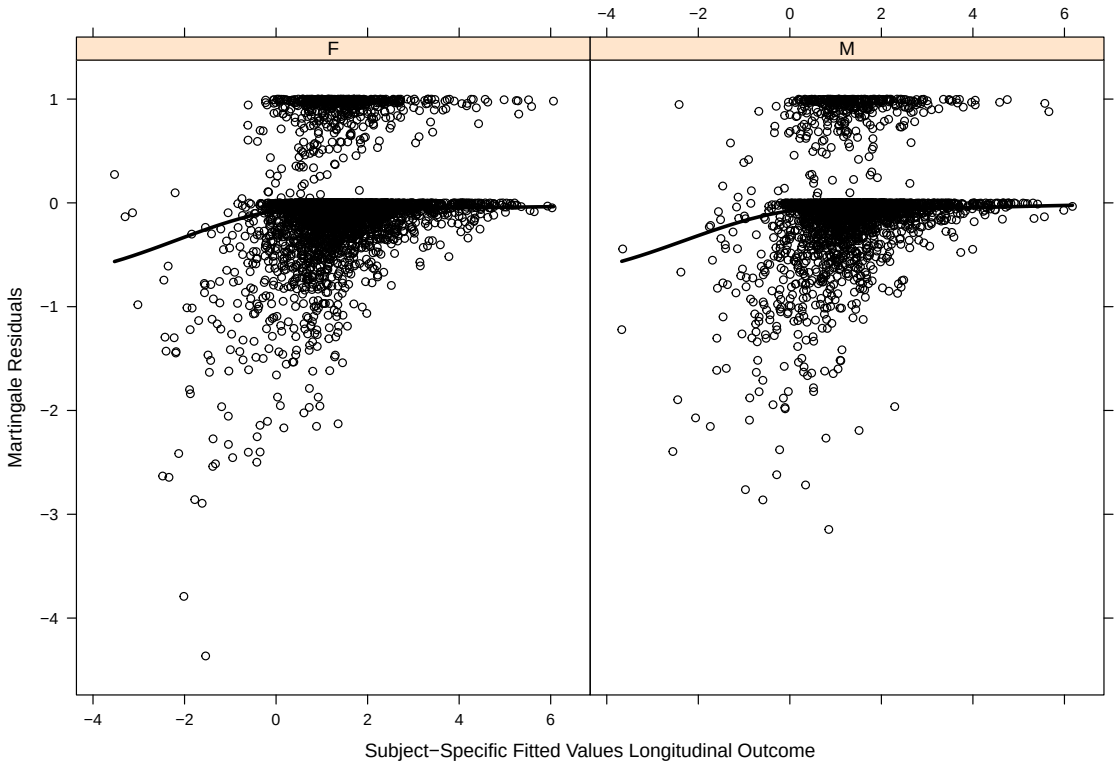

Figure 7.5: Martingale residuals versus the subject specific fitted values of the longitudinal outcome separated by gender for the THR dataset. The grey solid lines denotes the fit of the loess smoother (n = 1850).

the expected asymptotic distribution. It is clear that there are some strong differences between these two lines. This is particularly evident for smaller residuals. This is not ideal as we expect the survival function estimate to be close to the unit exponential distribution.

The diagnostics have shown that this model is potentially ill-specified due to systematic behaviour in the residuals. This may be because the joint model explored in this section only includes the predictors age, gender and follow-up time. In contrast, the optimal mixed-effects model in Chapter 4 includes more predictor variables. Therefore, we now consider including more variables to explore if this improves the model fit.

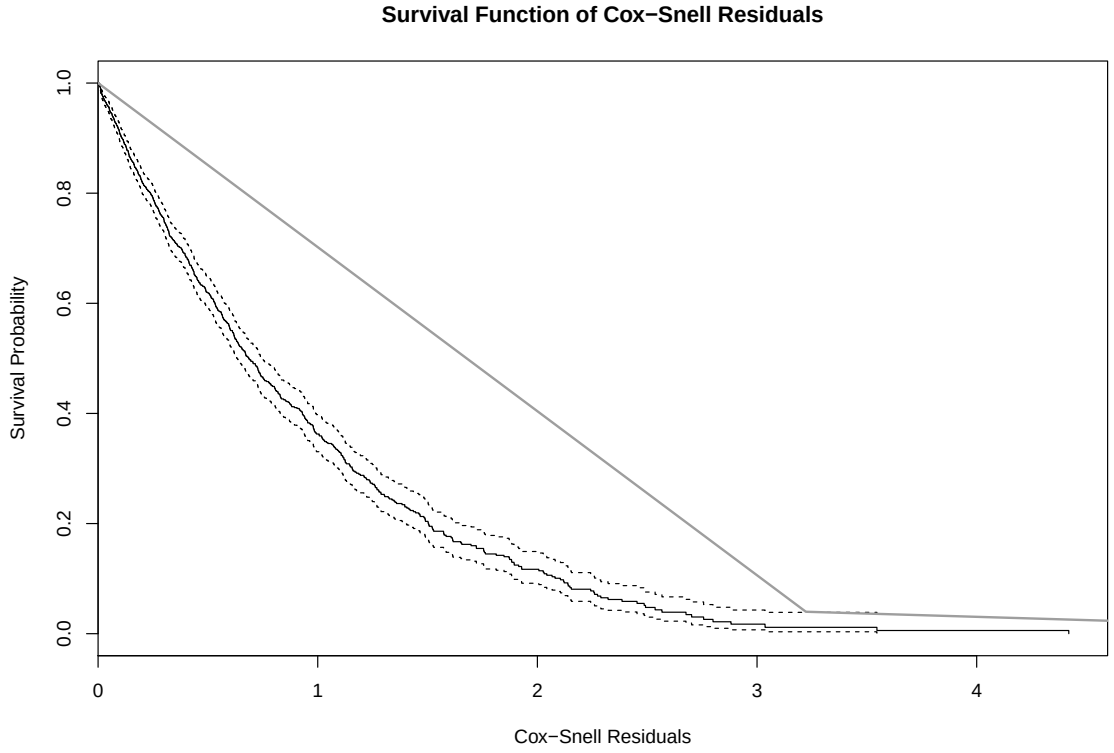

Figure 7.6: Cox-Snell residuals for the survival sub-model. The black solid line denotes the Kaplan-Meier estimate of the survival function of the residuals (with the dashed lines corresponding to the 95% point-wise confidence intervals), and the grey solid line, the survival function of the unit exponential distribution (n = 1850).

7.3 Joint Model with New Longitudinal Sub-model

The previous joint model showed some limitations when analysing the diagnostics. As a result, in this section, we introduce more predictor variables into the joint model (Equation 8.1). The variables that we include are those that were significant within the optimal mixed-effects model presented in Chapter 4. Therefore, we will look at including weight, AVN status, number of comorbidities and smoking status, along with the variables addressed in the previous analysis.

These variables were included and a joint model was formulated using the same process

outlined in the previous section. We note that the new predictor variables were only included in the longitudinal sub-model. The survival sub-model remained the same as the previous joint model.

Table 7.3 shows the results of this joint model, which includes more variables in the longitudinal sub-model. It is clear that the variables that have been added are all significant within the model. Similarly to previous analysis, this table indicates that having a heavier weight is associated with a decrease in HHS. Likewise, suffering from AVN is associated with a decrease in HHS when compared to not suffering from AVN. Interestingly, in this model suffering from more comorbidities is associated with a greater decrease in HHS. This was not the case for the separate mixed-effects model. Therefore, having four or more comorbidities is associated with a lower HHS than having three comorbidities. For the variable place of residence, patients who have external support are again associated with a decrease in HHS when compared to patients who live independently. Lastly, patients who smoke moderately are associated with a decrease in HHS when compared to patients who do not smoke at all.

Table 7.3 also shows the hazard ratio (HR), which is the ratio of the hazard rates relating to the two levels of the explanatory variables. For example, males appear to die at approximately twice the rate per unit time as females, as the HR is almost 2. Therefore, a higher HR indicates a higher hazard of death. We note that the HR is a relative measure of effect and does not inform us about absolute risk [22]. Interestingly age has a HR of approximately one. This indicates that an increase in age is associated with only a slightly higher hazard of death.

Table 7.3 also shows that the association between the transformed HHS and the survival outcome is significant and negative. This means that having a higher HHS is associated with a higher survival probability, which is as expected. Patients who experience

Variable	Estimate	SE	p-value	
Longitudinal sub-model				
Intercept	2.92	0.36	<0.001	
Baseline age (years)	-0.01	0.004	<0.001	
Follow-up time	0.02	0.002	<0.001	
(Follow-up time)^2	-0.0002	<0.001	<0.001	
Gender	0.43	0.09	<0.001	
Weight	-0.007	0.002	0.002	
AVN	-0.47	0.14	0.001	
Number of Comorbidities				
One	-0.02	0.11	0.83	
Two	-0.26	0.12	0.03	
Three	-0.41	0.14	0.003	
Four or more	-0.43	0.18	0.02	
Place of Residence				
Home alone & external support	-0.40	0.16	0.01	
Home with others & independent	-0.05	0.10	0.60	
Home with others & support	-0.54	0.12	<0.001	
Other	-0.54	0.29	0.06	
Smoker (non-smoker)				
Moderate	-0.48	0.13	0.0003	
Heavy	0.17	0.29	0.55	
Variable	Estimate	SE	p-value	HR
Survival sub-model				
logit(HHS)	-0.35	0.10	<0.001	0.71
Baseline age (years)	0.06	0.008	<0.001	1.07
Gender	0.66	0.14	<0.001	1.93

Table 7.3: Results of the joint model, which includes more variables in the longitudinal sub-model. For both sub-models, estimated parameters, standard errors (SE) and p-values are reported (n = 1850).

less pain and greater function are more likely to not only be younger, but also to suffer from less conditions that will increase their risk of death.

7.3.1 Diagnostics

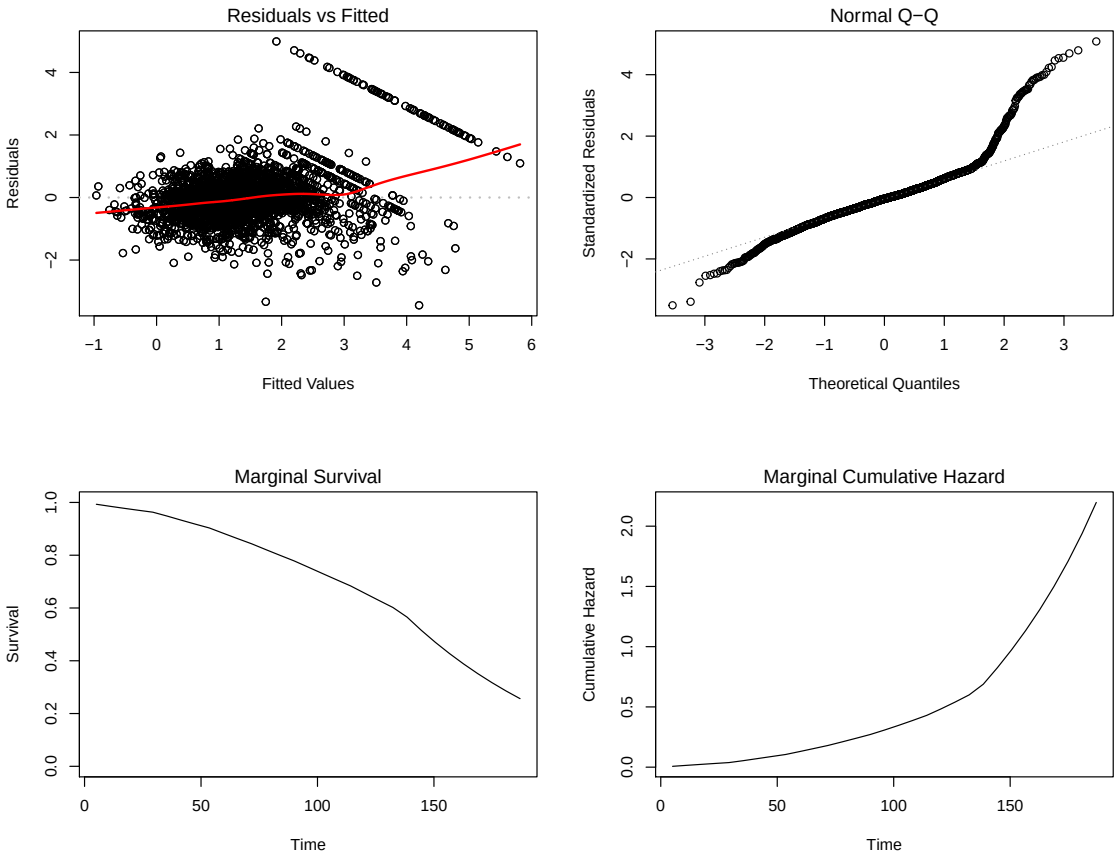

Figure 7.7: Diagnostic plots for the joint model fitted to the THR dataset. Included are the subject specific residuals versus the fitted values, the Q-Q plot of the subject specific residuals, and the marginal survival and cumulative risk functions for the event process (n = 1850).

It is again important to explore the diagnostics of this model. Particularly in comparison to the previous model to identify if the joint model with more predictor variables

within the longitudinal sub-model is more appropriate. Figure 7.7 shows the diagnostic plots for the joint model fitted to the THR dataset. The residuals versus fitted plot shows residuals that are slightly more centered around zero when compared to the residuals in the previous model (see Figure 7.2). The normal Q-Q plot shows no significant difference to the previous model.

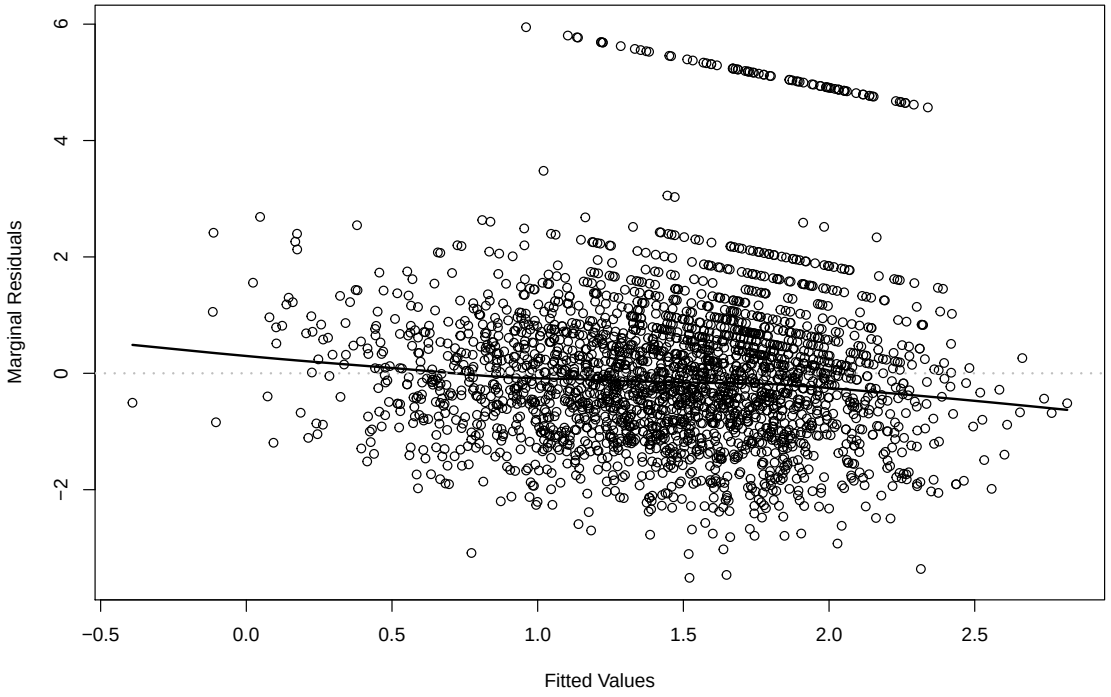

Figure 7.8: Marginal standardised residuals versus fitted values for the longitudinal outcome of the THR dataset (n = 1850).

Figure 7.8 shows the marginal standardised residuals versus fitted values for the longitudinal outcome of the THR dataset. There is a fitted loess curve that suggests that there are more positive residuals for smaller fitted values. Nevertheless, in comparison to the marginal residuals from the previous joint model (see Figure 7.3) this trend is less significant. It is also clear that the lower bound of the fitted values has increased

in comparison to the previous model. Therefore, the inclusion of more variables, which means more patients are ignored due to missingness, may have disregarded some patients with particularly low HHS measurements. Nevertheless, the systematic trend within the marginal residuals that is shown in Figure 7.8 is fairly small.

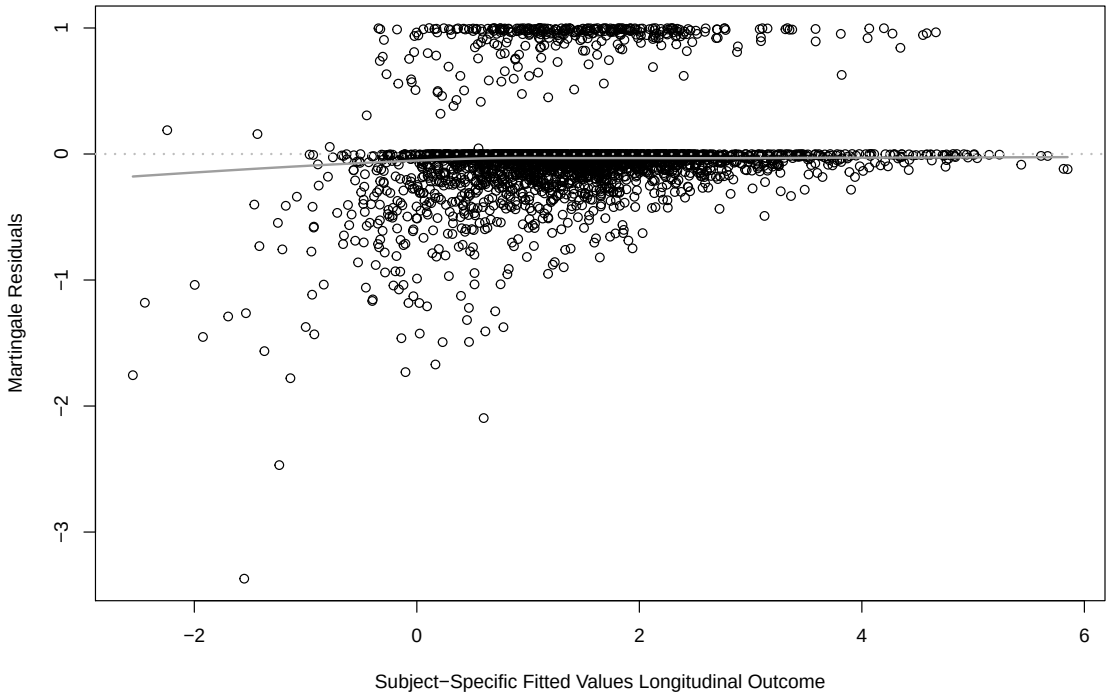

Figure 7.9: Martingale residuals versus the subject specific fitted values of the longitudinal outcome for the THR dataset. The grey solid line denotes the fit of the loess smoother (n = 1850).

Next we explore the diagnostics of the survival sub-model. Here we see that this joint model is more appropriate than the previous joint model. Figure 7.9 displays the martingale residuals versus the subject specific fitted values of the longitudinal outcome for the THR dataset. The grey solid line denotes the fit of the loess smoother. For smaller fitted values of the longitudinal outcome the loess smoother is only slightly further

from zero. In comparison to the previous model (Figure 7.4) the deviation from zero is significantly smaller.

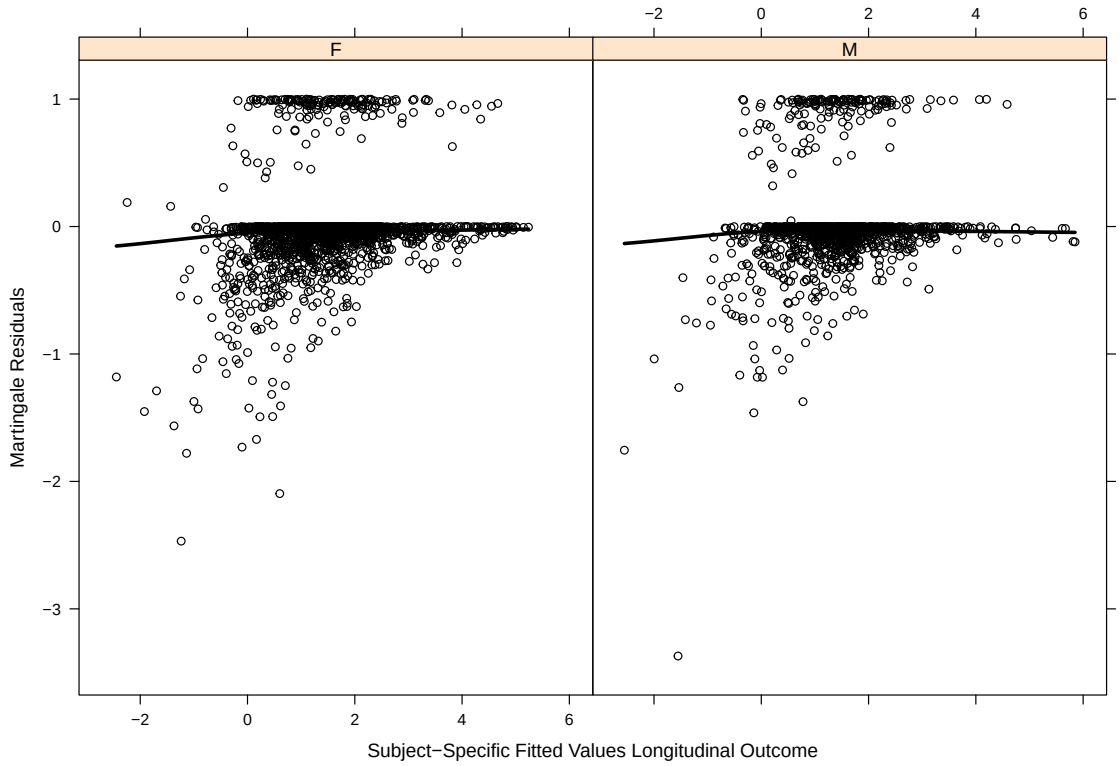

Figure 7.10: Martingale residuals versus the subject specific fitted values of the longitudinal outcome separated by gender for the THR dataset. The grey solid lines denotes the fit of the loess smoother (n = 1850).

Figure 7.10 displays the martingale residuals versus the subject specific fitted values of the longitudinal outcome separated by gender for the THR dataset. The grey solid lines denotes the fit of the loess smoother. There appears to be only minor deviations in the loess smoother from zero but these appear to be similar deviations across both genders. Therefore, the systematic trend that was more evident within the previous model has definitely diminished in this joint model.

The Cox-Snell residuals can also be explored to further analyse the survival sub-model

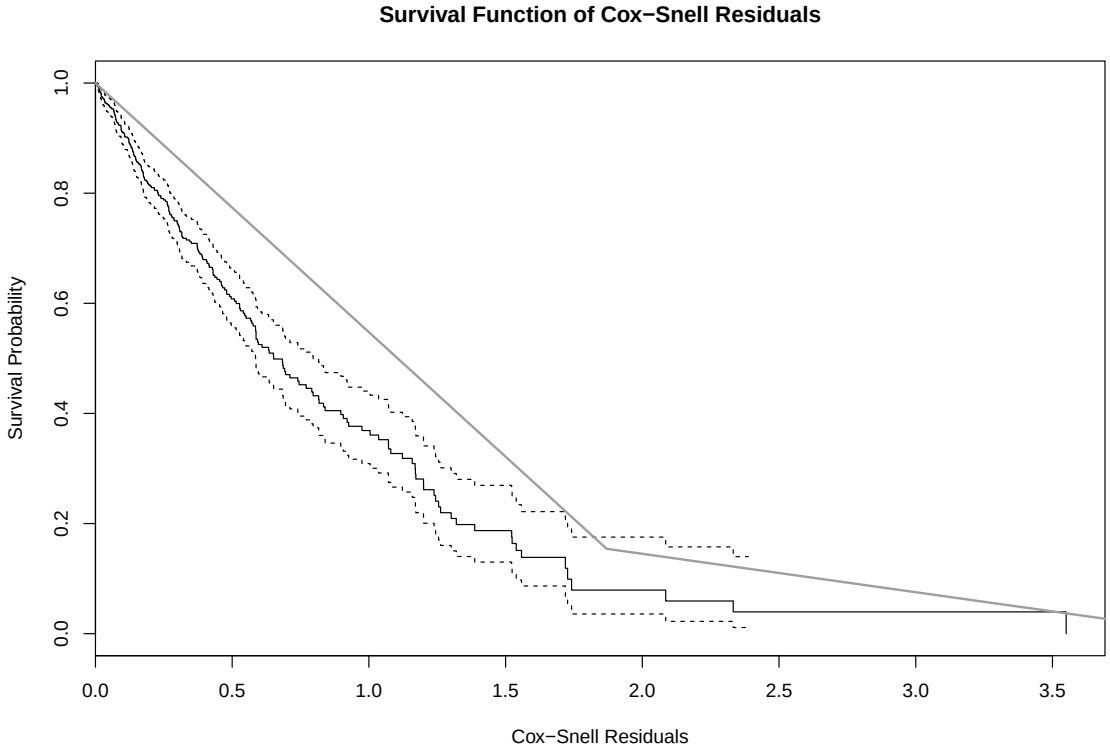

Figure 7.11: Cox-Snell residuals for the survival sub-model. The black solid line denotes the Kaplan-Meier estimate of the survival function of the residuals (with the dashed lines corresponding to the 95% pointwise confidence intervals), and the grey solid line, the survival function of the unit exponential distribution (n = 1850).

residuals. Figure 7.11 shows the Cox-Snell residuals for the survival sub-model. The black solid line denotes the Kaplan-Meier estimate of the survival function of the residuals (with the dashed lines corresponding to the 95% pointwise confidence intervals), and the grey solid line, the survival function of the unit exponential distribution.

When comparing the Kaplan-Meier estimate with the survival function of the unit exponential distribution it is clear that there are differences between these two lines, particularly for smaller residuals. Though we would expect the survival function estimate to be close to the unit exponential distribution, it is still a definite improvement on the

fit of the previous joint model. When comparing Figures 7.6 and 7.11 it is clear that the second joint model has a Kaplan-Meier estimate that sits closer to the survival function of the unit exponential distribution.

The diagnostics show that the addition of predictor variables, from the optimal mixed-effects model in Chapter 4, into the longitudinal sub-model of the joint model improves the model. Therefore, in future it would be beneficial to look further into determining an optimal survival sub-model that potentially includes more predictor variables. At this stage we do not explore this.

This section shows that our second joint model fits the data better than our initial joint model. Specifying differently the association structure in the joint model is not considered in this here but could also be a potential way of improving the model. Nevertheless, we can explore how to approach prediction for this joint model. We can discuss these predictions within the context of the predictions made using the mixed-effects model defined in Chapter 4. These predictions will be compared within the context of the original research question, which is how can we predict postoperative patient pain and function based on preoperative factors, such as age and gender.

7.3.2 Prediction

Joint models can be used to obtain individual patient level predictions across both longitudinal and survival outcomes. Particularly of interest to us is the dynamic prediction of the longitudinal outcomes, the HHS measurements. This is because our aim is to develop a model that can predict postoperative patient pain and function based on preoperative factors. Nevertheless, later in this section we will also address dynamic survival probability prediction alongside the longitudinal predictions.

In the case of longitudinal outcomes the expected HHS measurement at time $u > t$, given the observed HHS up until that point, $y_i(t)\{y_i(s), 0 \leq s < t\}$, can be written as

$$w_i(u|t) = E\{y_i(u)|T_i^* > t, Y_i(t), D_n; \theta^8\}, \quad u > t. \tag{7.2}$$

This is dynamic as with time more information is used to calculated the new prediction. Therefore, more HHS measurements are used to inform predictions made at later times after surgery. This estimation is completed using a Monte Carlo simulation scheme, which will be explored further in the context of survival probabilities later in this section [28].

This method of estimation will be used to determine predicted longitudinal outcomes in the results that follow. Initially we will compare prediction methods used in Chapter 4 that do not involve dynamic prediction. Then, later in this section, dynamic prediction will be explored.

Firstly, we use a similar process as used in Section 4.2.2 (Figure 4.11). This method selects a random patient and removes their last observation from the estimation model frame and then attempts to predict the HHS observation at that time. Therefore, this compares how well the model can predict the final HHS measurement given that it has a patient's previous measurements.

Figure 7.12 shows a plot of the observed against the predicted HHS for 500 patient's last HHS observation. We can see that for patients who have a low HHS observation the model appears to predict values that are higher than the observed values. In contrast, for patients who have a higher HHS observation the model appears to predict values that are lower than the observed value. Nevertheless, for patients whose HHS is within the area of highest density the predictions are fairly accurate. Overall, there is a systematic trend that exists within the prediction of the longitudinal outcomes. This trend was visible

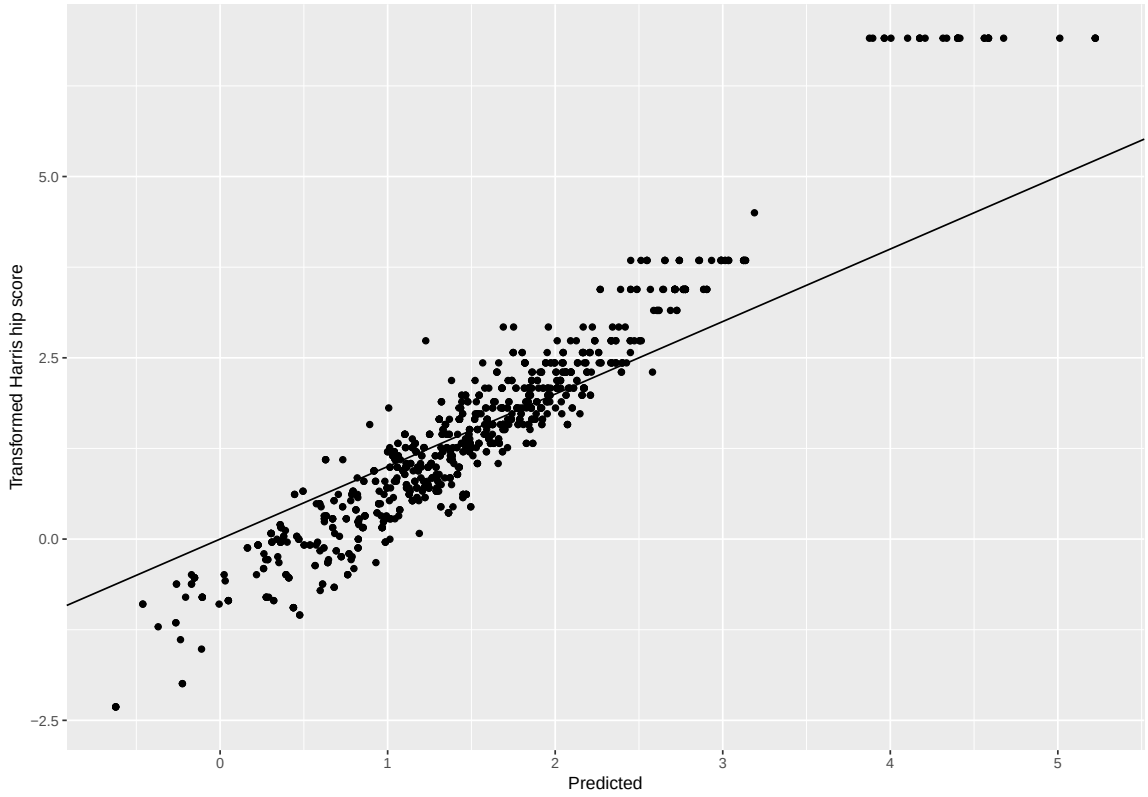

Figure 7.12: Plot of the observed against the predicted HHS for the last observation of 500 patient's last HHS observation.

within the longitudinal sub-model residuals and is further confirmed by this analysis.

This can be compared to a similar prediction plot relating to the optimal mixed-effects model in Chapter 4 (Figure 4.15). The predictions made by this joint model are more accurate as they sit closer to the linear line. Although, with this accuracy a systematic trend becomes clear.

We can also look closer at a subset of patients. Figure 7.13 shows a plot of the observed and predicted HHS measurements for 10 randomly selected patients last observation, given that we know the HHS measurements of earlier observations. The x-axis is the follow-up time and the y-axis is the transformed HHS. The dotted lines show the observed patient HHS trajectories. The smaller points are the observed patient HHS measurements. The

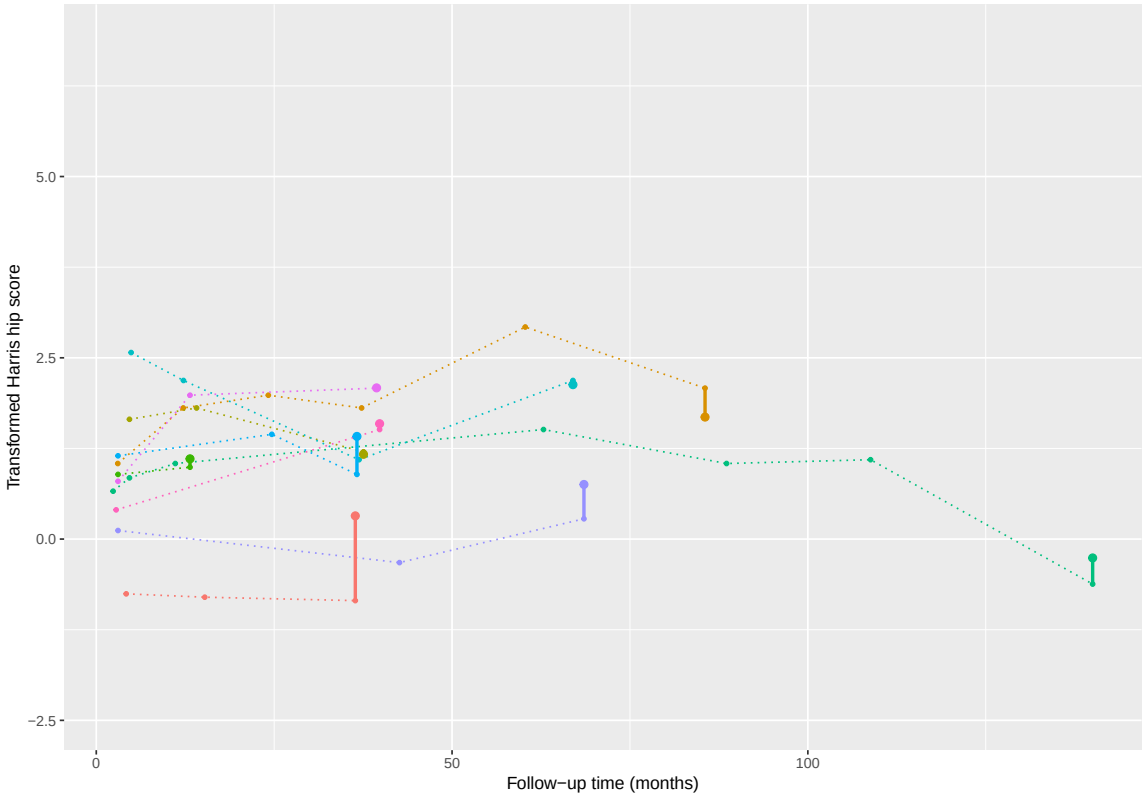

Figure 7.13: Plot of the observed and predicted HHS measurements for 10 randomly selected patients last observation, given that we know the HHS measurements of earlier observations. The x-axis is the follow-up time and the y-axis is the transformed HHS. The dotted lines show the observed patient HHS trajectories. The smaller points are the observed patient HHS measurements. The larger points are the predicted HHS measurements for the patient's last observation. The solid line shows the difference between the predicted and the observed HHS measurements for the patient's last observation.

larger points are the predicted HHS measurements for the patient's last observation. The solid line shows the difference between the predicted and the observed HHS measurements for the patient's last observation. This supports what is shown in Figure 7.12 that lower HHS measurements are consistently over predicted and higher HHS measurements appear to be under predicted. Those HHS measurements that lie in the middle are the most accurately predicted measurements.

We now illustrate the dynamic nature of longitudinal predictions. To do this we take

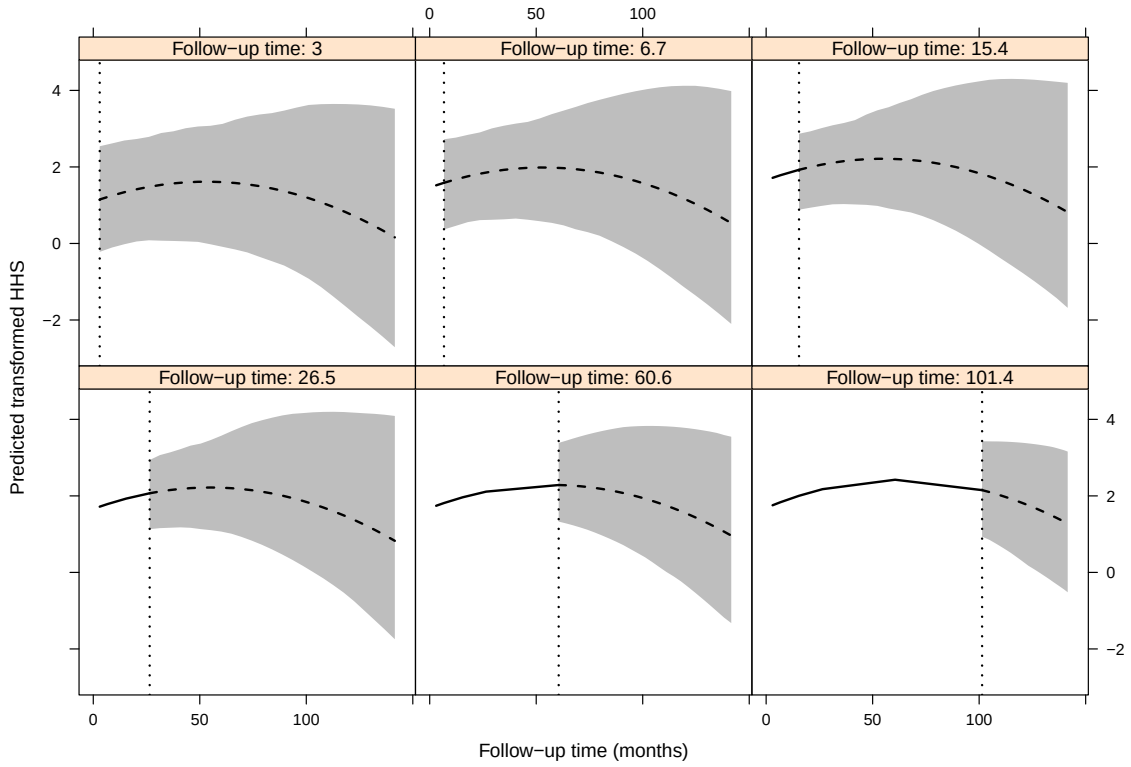

Figure 7.14: Dynamic predictions of the longitudinal responses for Patient 10056. In each panel the dotted vertical line indicates the time point of the last observed HHS measurement. The solid line denotes that fitted longitudinal trajectory prior to the last visit, and the dashed line represents the predicted longitudinal trajectory. The grey areas denote the 95% pointwise confidence intervals.

one patient and show how their HHS predictions update over time. Figure 7.14 shows dynamic predictions of the longitudinal responses for Patient 10056. In each panel the dotted vertical line indicates the time point of the last observed HHS measurement. The solid line denotes that fitted longitudinal trajectory prior to the last visit, and the dashed line represents the predicted longitudinal trajectory. The grey areas denote the 95% pointwise confidence intervals.

We see that at follow-up time three the predicted trajectory shows a sharper eventual descent than at later follow-up times. This is because the patient's HHS measurements

did not drop significantly over time, therefore the prediction updated accordingly. As we would expect the width of the confidence interval is larger at greater times as a prediction made near the time of the last measurement is much more reliable. These confidence intervals are not symmetric as they are not based on an asymptotic normality of $\hat{w}_i(u|t)$ [28]. This form of prediction is useful within personalised medicine as we are constantly receiving new information on a patient. Therefore, the best predictions occur when a patient, who is already in the system, is having their biomarkers measured regularly.

We can also examine the dynamic survival probabilities prediction alongside the longitudinal predictions. Therefore, we look at predicting a patient's survival probability at a certain time after surgery. To do this we let $D_n = \{T_i, \delta_i, y_i; i = 1, \ldots, n\}$ be a random sample that the joint model is fitted in. We look at predicting the survival probabilities of a new patient with HHS measurements $Y_i(t) = \{y_i(s); 0 \leq s < t\}$ and baseline covariates w_i. Therefore, the conditional probability of surviving time $u > t$ given survival up to t can be written as

$$\pi_i(u|t) = \Pr(T_i^* \geq u | T_i^* > t, Y_i(t), w_i, D_n; \theta^*), \qquad t > 0 \tag{7.3}$$

where θ^* are the true parameters. We note that this conditional probability again has a dynamic nature because as time increases the parameter t increases and the predictions are updated accordingly [28].

When estimating the individual level conditional survival probabilities we use the conditional independence assumption. The conditional independence assumption can be written as

$$p(T_i, \delta_i, y_i | b_i; \theta) = p(T_i, \delta_i | b_i; \theta) p(y_i | b_i; \theta),$$

as shown in Section 7.2. Applying the above assumption to Equation 7.3 we get

$$\begin{aligned}
&\Pr(T_i^* \geq u | T_i^* > t, Y_i(t), w_i, D_n; \theta) \\
&\qquad = \int \Pr(T_i^* \geq u | T_i^* > t, b_i; \theta) p(b_i | T_i^* > t, Y_i(t), w_i, D_n : \theta) db_i \\
&\qquad = \int \frac{S_i\{u | M_i(u, b_i, \theta); \theta\}}{S_i\{t | M_i(t, b_i, \theta); \theta\}} p(b_i | T_i^* > t, Y_i(t), w_i, D_n : \theta) db_i, \qquad (7.4)
\end{aligned}$$

where $S_i(.)$ is the survival function and $M_i(.)$ is the longitudinal history of the i-th patient. We note that this equation does have redundancy but it was used in this form to be consistent with the text.

Using Equation 7.4 and the posterior expectation of Equation 7.3 we can apply a Monte Carlo simulation scheme to estimate $\pi_i(u|t)$ [28]. After performing this simulation scheme the realisations of $\{\pi_i^l(u|t), l = 1, \ldots, L\}$ can be used to derive point estimates of $\pi_i(u|t)$, such as

$$\hat{\pi}_i(u|t) = \text{median}\{\pi_i^l(u|t), l = 1, \ldots, L\}.$$

This method of estimation will be used to determine predicted survival probabilities in the results that follow. We note that in this case we will again be looking at prediction for patients that already exist within the estimation model frame.

Figure 7.15 shows dynamic survival probabilities for Patient 10056 from the THR dataset during follow-up. The vertical dotted lines represent the time point of the last HHS measurement. Left of the vertical line, the fitted longitudinal trajectory is depicted. Right of the vertical line, the solid line represents the median estimator for $\pi_i(u|t)$, and the dashed lines the corresponding 95% confidence intervals. This prediction assumes that the last time point that the patient was still alive was when they last had a HHS measurement recorded. Therefore, for all previous time points $u < t$, $\pi_i(u|t) = 1$.

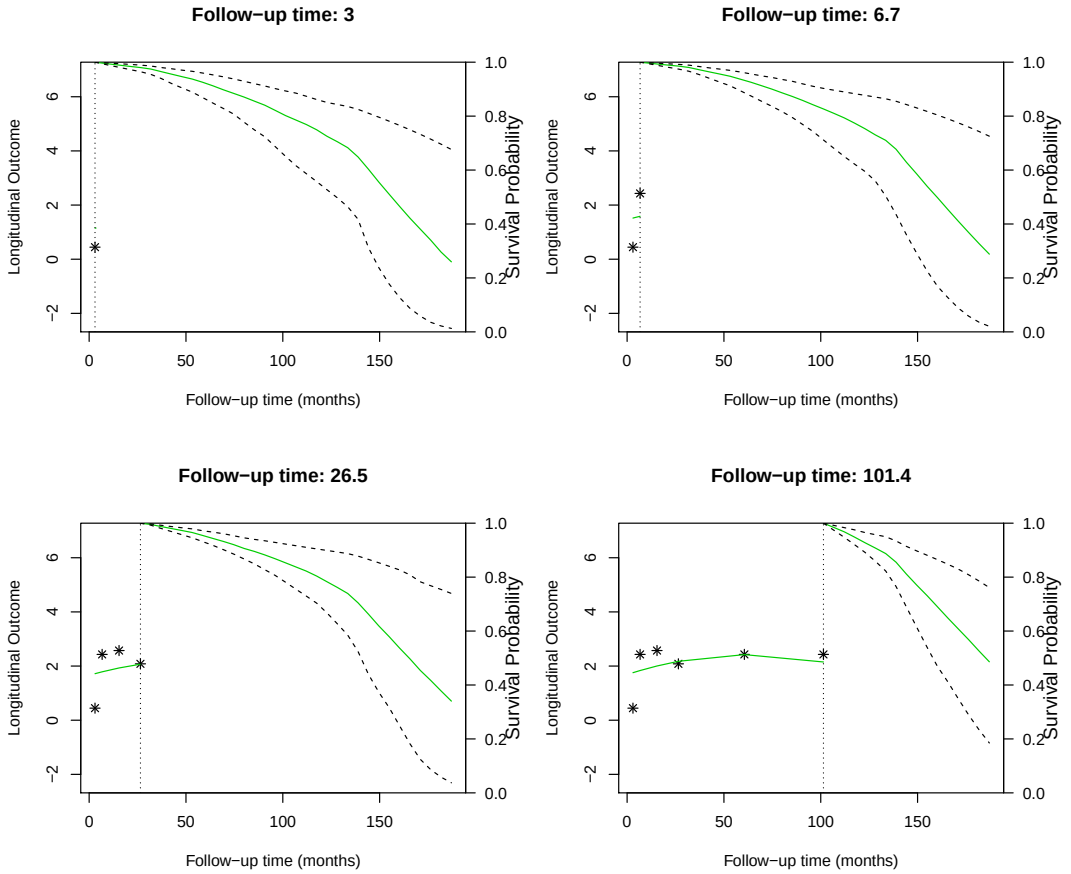

Figure 7.15: Dynamic survival probabilities for Patient 10056 from the THR dataset during follow-up. The vertical dotted lines represent the time point of the last HHS measurement. Left of the vertical line, the fitted longitudinal trajectory is depicted. Right of the vertical line, the solid line represents the median estimator for $\pi_i(u|t)$, and the dashed lines the corresponding 95% confidence intervals.

Figure 7.15 shows the dynamic predictions of the survival probability updated with each new HHS measurement. We observe that after the second HHS measurement, where there is a clear increase in HHS, the rate of decrease of the conditional survival function becomes less steep.

To further examine how changes in the observed HHS measurements effect the dynamic updates of the survival probabilities we can compare the predictions for two other patients.

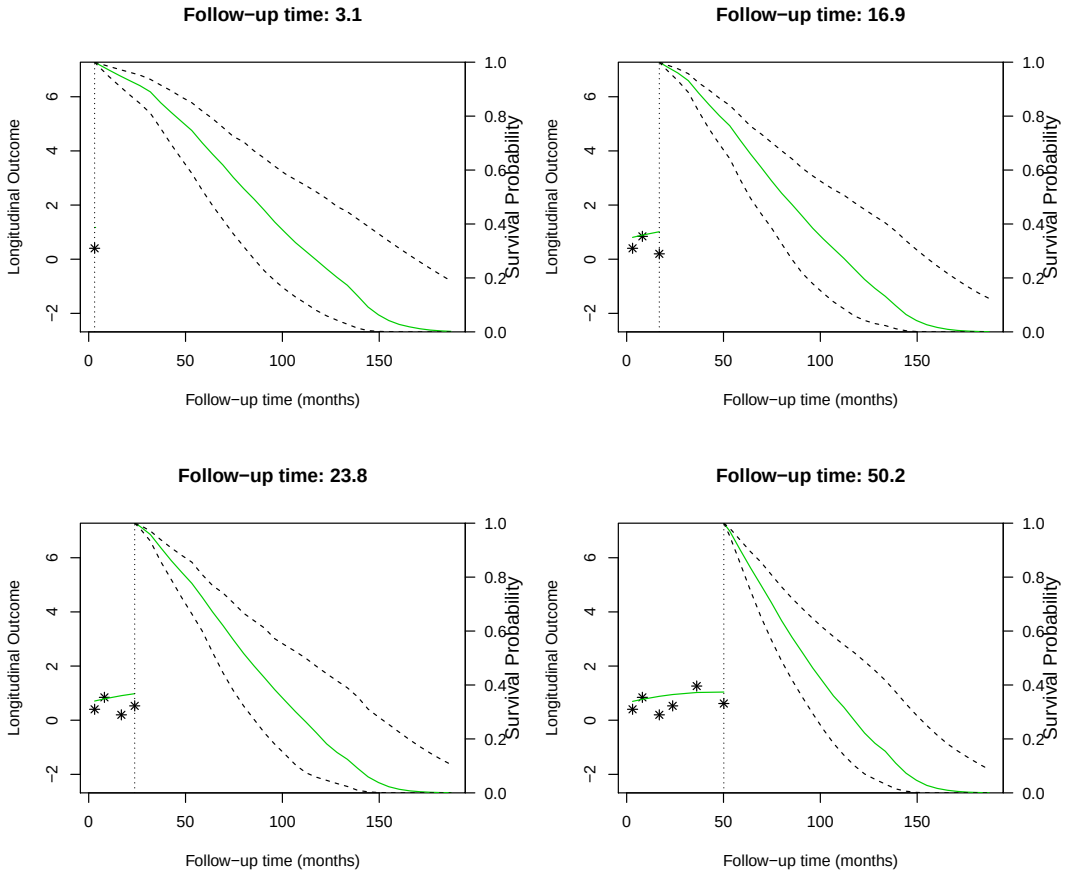

Figure 7.16: Dynamic survival probabilities for Patient 10298 from the THR dataset during follow-up. The vertical dotted lines represent the time point of the last HHS measurement. Left of the vertical line, the fitted longitudinal trajectory is depicted. Right of the vertical line, the solid line represents the median estimator for $\pi_i(u|t)$, and the dashed lines the corresponding 95% confidence intervals.

Figure 7.16 shows the dynamic survival probabilities for Patient 10298 from the THR dataset during follow-up. The vertical dotted lines represent the time point of the last HHS measurement. Left of the vertical line, the fitted longitudinal trajectory is depicted. Right of the vertical line, the solid line represents the median estimator for $\pi_i(u|t)$, and the dashed lines the corresponding 95% confidence intervals. It is clear that in comparison to Figure 7.15 this patient has overall a lower observed HHS. This is reflected in the steeper

conditional survival function over all time points. This patient is exactly 24 years older than Patient 10056 which also contributed to the lower survival probability that we see.

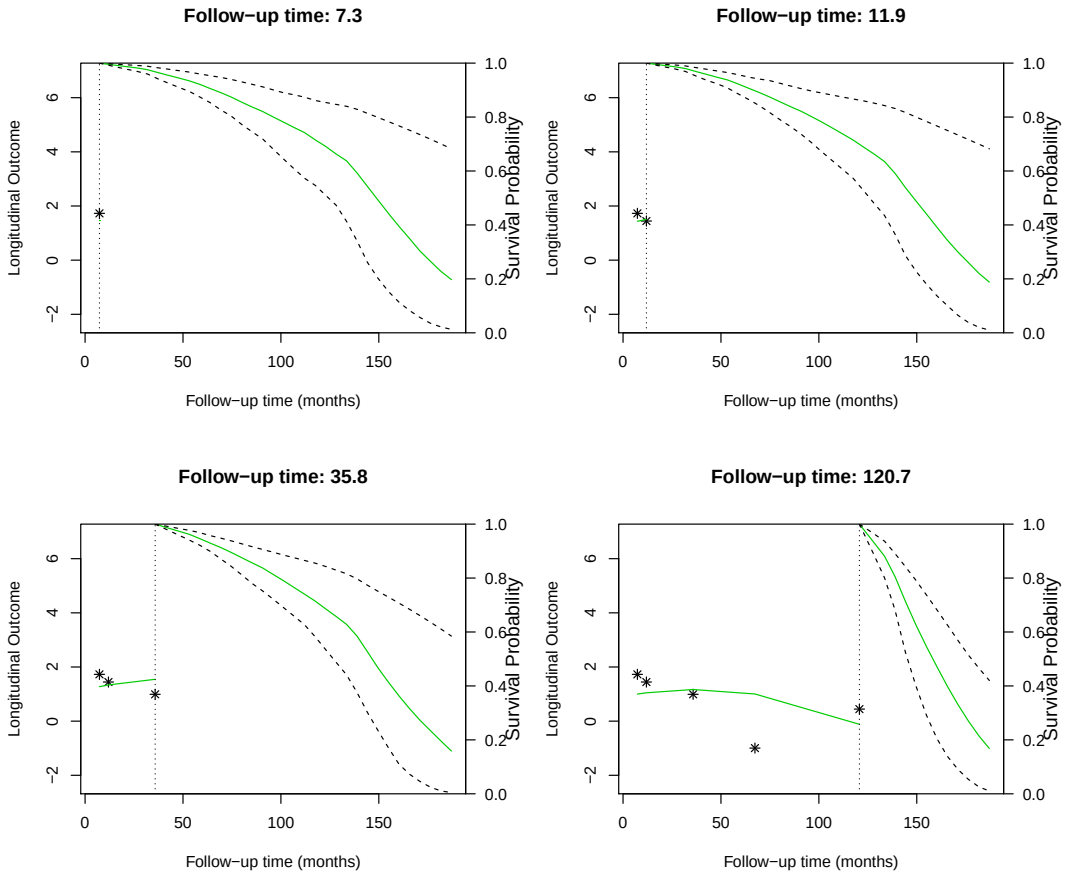

Figure 7.17: Dynamic survival probabilities for Patient 10268 from the THR dataset during follow-up. The vertical dotted lines represent the time point of the last HHS measurement. Left of the vertical line, the fitted longitudinal trajectory is depicted. Right of the vertical line, the solid line represents the median estimator for $\pi_i(u|t)$, and the dashed lines the corresponding 95% confidence intervals.

Figure 7.17 shows the dynamic survival probabilities for Patient 10268 from the THR dataset during follow-up. The vertical dotted lines represent the time point of the last HHS measurement. Left of the vertical line, the fitted longitudinal trajectory is depicted. Right of the vertical line, the solid line represents the median estimator for $\pi_i(u|t)$, and

the dashed lines the corresponding 95% confidence intervals. We can see that there is a decrease in the observed HHS over time for this patient. Therefore, the survival probability function becomes steeper over time.

These predictions show that it is possible to get accurate dynamic predictions of both the survival and longitudinal outcomes, that can be used in personalised medical treatment. These predictions are capable of incorporating not only the survival predictions, and the HHS predictions but also the relationship between these two outcomes.

Overall the longitudinal predictions made using this joint model are more accurate than those made using the mixed-effects model. Although, one has to be aware of the limitations of these predictions. In this case, a crucial limitation is that the low HHS observations are slightly over predicted. Likewise, the higher HHS observations are under predicted. Nevertheless, there are definite advantages to using this model. This joint model gives dynamic survival probability predictions that can help inform pain and function prediction. At a more general level, this model provides greater insight into a patient's health, as it relates to their THR.

Chapter 8

Extension of the Joint Model

In this chapter we explore two extensions of the joint model defined in the previous chapter. Firstly, we analyse a joint model where the survival outcome is the time until a patient undergoes their second surgery, instead of survival time. Secondly, we explore joint modelling with multivariate longitudinal outcomes. We are specifically interested in the effect of separating the HHS into separate pain and function scores, which in turn becomes two longitudinal outcomes. We see that combining two important measures, such as pain and function, into one measure, the HHS, has clear limitations.

8.1 Second Surgery Sub-Study

In the previous section we explored joint modelling of time-to-event and longitudinal outcomes for the total hip replacement (THR) dataset. Specifically analysing Harris hip score (HHS) as a longitudinal outcome and survival as the time-to-event outcome. Alternatively, we can consider the time until a patient undergoes their second surgery to be the time-to-event outcome. Therefore, we can create a joint model where the

longitudinal outcome is still the HHS but the time-to-event outcome is a patient's second surgery status.

Patient Demographics	**Second Surgery**	**One Surgery**
n of patients	269	1533
Mean (SD) age (yrs)	63.2 (10.6)	67.4 (12.3)
Mean (SD) weight (kgs)	81.3 (16.7)	78.2 (18.2)
Female (n,%)	168 (62.5)	891 (58.1)
Male (n,%)	101 (37.5)	642 (41.9)
Diagnosis (n,%)		
AVN	21 (7.8)	125 (8.2)
Number of Comorbidities (n,%)		
None	68 (25.3)	411 (26.8)
One	95 (35.3)	525 (34.2)
Two	67 (24.9)	347 (22.6)
Three	27 (10.0)	180 (11.7)
Four or more	11 (4.1)	70 (4.6)

Table 8.1: Summary of the distribution of demographic variables for second surgery and non-second surgery patients (n = 1802).

When we state that a patient is undergoing a second surgery we are referring to a patient having surgery on their other hip. For example, if a patient initially has a THR on their right side then their second surgery would be a left side THR. Within the THR dataset there are 450 (19.7%) patients who undergo a second surgery. Within this particular sub-study of 1802 patients there are 269 patients (14.9%) that have undergone a second surgery. Table 8.1 shows a comparison of the distribution of demographic variables across patients who have had a second surgery and those that only have one surgery. This shows that the mean age of patients who have had a second surgery (63 yrs) is lower than mean age of the whole dataset (67 yrs) and lower than the mean age of the subset of patients with one surgery (67 yrs). We also note that there is a higher percentage

of females who have undergone a second surgery when compared to the distribution of females that have only had one surgery. Nevertheless, the distribution of demographic variables is relatively consistent across patients who have or have not had a second surgery.

The same joint modelling frame work that was introduced in the Chapter 7 is again used in this chapter. It can be written as,

$$\begin{cases} y_i(t) = x_i(t)^T\beta + z_i(t)^T b_i + \epsilon_i(t), \\ h_i(t) = h_0(t)\exp\left[\gamma^T w_i + \alpha\{x_i(t)^T\beta + z_i(t)^T b_i\}\right], \quad t > 0, \end{cases}$$

where the hazard function, $h_i(t)$, refers to the rate of second surgery at a time point.

Initially a joint model was explored where the predictor variables for the longitudinal sub-model were simply baseline age, gender and follow-up time. For the survival sub-model baseline age and gender were also included. This joint model showed similar limitations to the initial joint model explored in Chapter 7. For example, there appeared to be strong systematic trends in the residuals. Therefore, we proceeded to include more predictor variables within the longitudinal sub-model. Specifically, the longitudinal sub-model was given the same structure as the optimal mixed-effects model identified in Chapter 4. The predictor variables specified within the time-to-event sub-model are still baseline age and gender. The joint model that contained a smaller set of predictor variables included 1802 patients, of which approximately 15% had experienced a second surgery. When more predictor variables were included in the longitudinal sub-model the number of patients within the estimation model frame became 801, of which approximately 17% have had a second surgery.

Table 8.2 shows the results of the joint model, with second surgery as the time-to-event outcome. For both sub-models, estimated parameters, standard errors (SE) and p-values

Variable	Estimate	SE	p-value	
Longitudinal sub-model				
Intercept	2.96	0.37	<0.001	
Baseline age (years)	-0.01	0.004	0.001	
Follow-up time	0.02	0.002	<0.001	
(Follow-up time)^2	-0.0002	0.000	<0.001	
Gender	0.41	0.09	<0.001	
Weight	-0.008	0.002	<0.001	
AVN	-0.46	0.15	0.002	
Number of Comorbidities				
One	-0.04	0.11	0.73	
Two	-0.31	0.12	0.01	
Three	-0.36	0.14	0.008	
Four or more	-0.39	0.18	0.03	
Place of Residence				
Home alone & external support	-0.45	0.17	0.007	
Home with others & independent	-0.04	0.10	0.65	
Home with others & support	-0.54	0.13	<0.001	
Other	-0.58	0.30	0.05	
Smoker (non-smoker)				
Moderate	-0.46	0.13	<0.001	
Heavy	0.25	0.29	0.40	
Variable	Estimate	SE	p-value	HR
Survival sub-model				
logit(HHS)	-0.42	0.12	<0.001	0.66
Baseline age (years)	-0.03	0.007	<0.001	0.97
Gender	-0.29	0.19	0.12	0.75

Table 8.2: Results of the joint model with second surgery as the time-to-event outcome. For both sub-models, estimated parameters, standard errors (SE) and p-values are reported (n = 1802).

are reported. The results for the longitudinal sub-model are similar to the results shown in Table 7.3. Patients with a heavier weight are associated with a decrease in HHS. Suffering from AVN is associated with a decrease in HHS when compared to not suffering from AVN. The more comorbidities you suffer from is associated with worse pain and less function. For the variable place of residence, patients who have external support are again associated with a decrease in HHS when compared to patients who live independently. Also patients who smoke moderately are associated with a decrease in HHS when compared to patients who do not smoke.

For the survival sub-model Table 8.2 shows the hazard ratio (HR) of each predictor variable. Since a higher HR indicates a higher hazard of having a second surgery we see that patients who are older are associated with a slightly lower hazard of having a second surgery. This may be due to it becoming less feasible for patients to undergo a THR on both the left and right side as they get older.

The association between the transformed HHS and the time-to-event outcome (Table 8.2) is significant and negative. This means that having a higher HHS is associated with a higher probability of not undergoing a second surgery, which is as expected. Patients who experience less pain and greater function are less likely to suffer from conditions that will increase their risk of second surgery.

The diagnostic plots for this model showed many similarities to the model explored in Section 7.3. While some assumptions of the model appear supported by the diagnostics there are still some limitations that become evident. Figure 8.1 shows the subject specific residuals versus the fitted values, the Q-Q plot of the subject specific residuals, and the marginal survival and cumulative risk functions for the event process. The first plot shows that those observations where the HHS is 100 behave poorly, otherwise the assumption of homoscedasticity is valid. The Q-Q plot shows that the assumption of normality is in

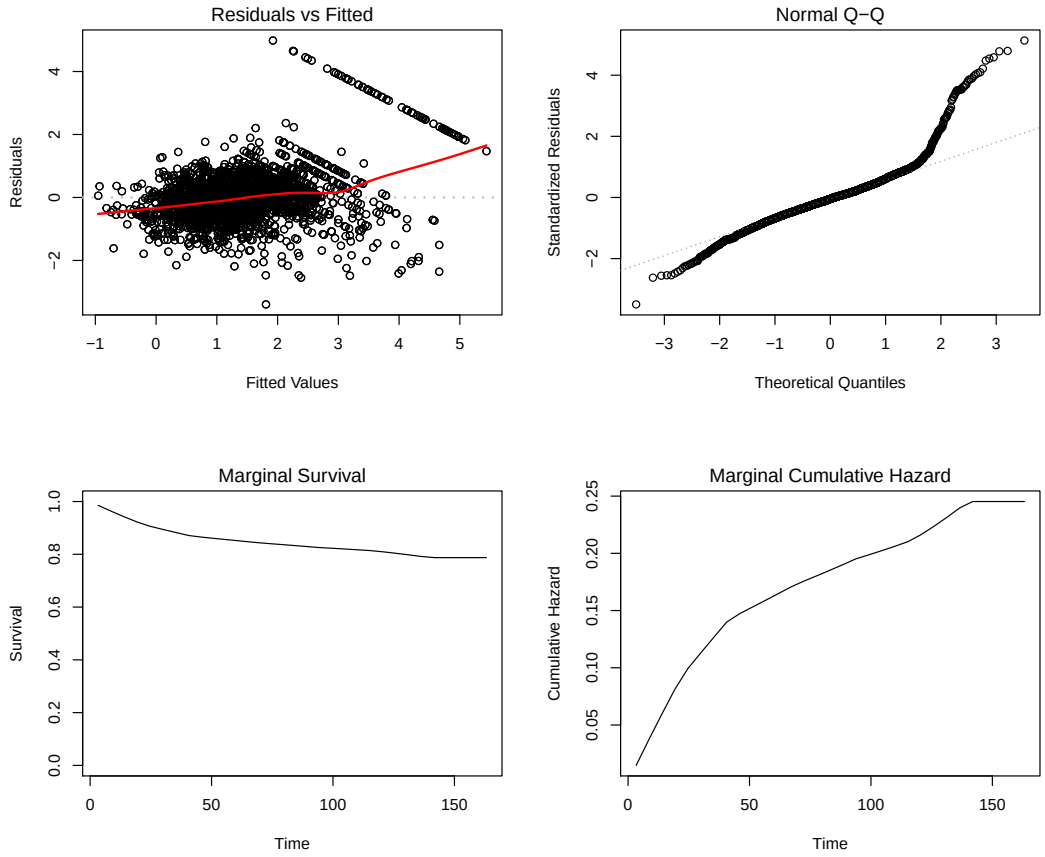

Figure 8.1: Diagnostic plots for the joint model with second surgery as the time-to-event outcome. Included are the subject specific residuals versus the fitted values, the Q-Q plot of the subject specific residuals, and the marginal survival and cumulative risk functions for the event process (n = 1802).

question due to the upper tail deviating from the reference line. Therefore, these results are similar to those found previously, showing that this model does not appropriately predict the upper bound of the HHS.

The survival sub-model diagnostics were also explored. For example, Figure 8.2 shows the Cox-Snell residuals for the survival sub-model. The black solid line denotes the Kaplan-Meier estimate of the survival function of the residuals (with the dashed lines corresponding to the 95% point-wise confidence intervals), and the grey solid line, the

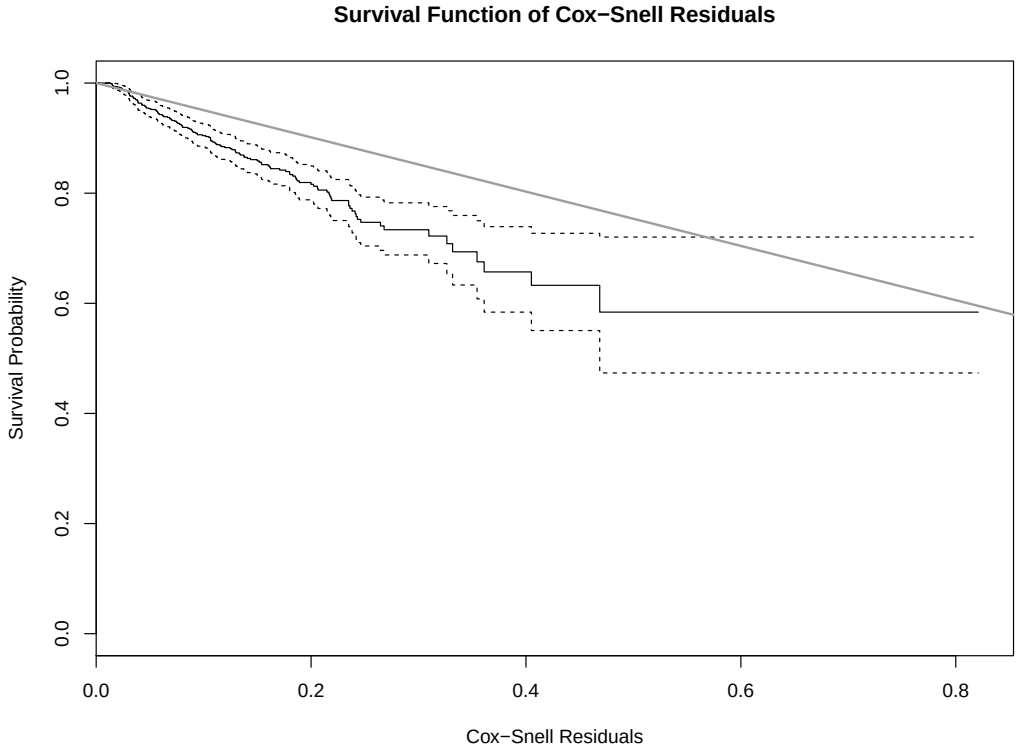

Figure 8.2: Cox-Snell residuals for the survival sub-model. The black solid line denotes the Kaplan-Meier estimate of the survival function of the residuals (with the dashed lines corresponding to the 95% point-wise confidence intervals), and the grey solid line, the survival function of the unit exponential distribution (n = 1802).

survival function of the unit exponential distribution. If this model was appropriate we would expect to see the Kaplan-Meier estimate close to the survival function of the unit exponential distribution, as this is the expected asymptotic distribution. Though it is clear that there are some differences in these lines the survival function of the unit exponential distribution lies close to the 95% point-wise confidence interval. We also note that because in this case our time-to-event outcome is undergoing a second surgery the survival probabilities are overall higher. This implies that there is a higher probability that a patient will not have a second surgery in comparison to the probability that a

patient will survive (see Figure 7.6).

Overall, this joint model shows that the variables age, gender, follow-up time, weight, suffering from AVN, number of comorbidities, place of residence and smoking status are all significant predictor variables for the HHS. Not only are they capable of predicting a patient's HHS but their HHS prediction is also capable of informing a patient's survival status or risk of second surgery.

Dynamic prediction using this joint model is possible but will not be explored here. Instead, we now consider separating the HHS into a pain score and function score. This is due to our interest in the separate prediction of pain and function. Therefore, we want to explore if it is more informative to predict pain and function separately instead of together within the HHS.

8.2 Joint Model with Multivariate Longitudinal Outcomes

Another extension of the joint model is joint models with multivariate longitudinal outcomes. This means considering pain and function as separate longitudinal outcomes. To begin this analysis each patient's pain and function scores had to be calculated. In this case, the pain score is defined as the answer to the first question in Table A.2 in Appendix A.2. The question is how much pain is the patient in. There are six possible answers with the highest scoring answer (score of 44) indicating that the patient is not in pain. We note that this variable, pain score, is more appropriately dealt with as a categorical variable but this is outside of the scope of this study. Therefore, we explore modelling with the assumption that the pain score is a continuous variable. The function score

will be defined as all other components of the HHS, including the four points that assess if there is an absence of deformity (see Tables A.2, A.3 and A.4 in Appendix A.2). To further understand these outcomes we explore their relationship to each other and to the predictor variables.

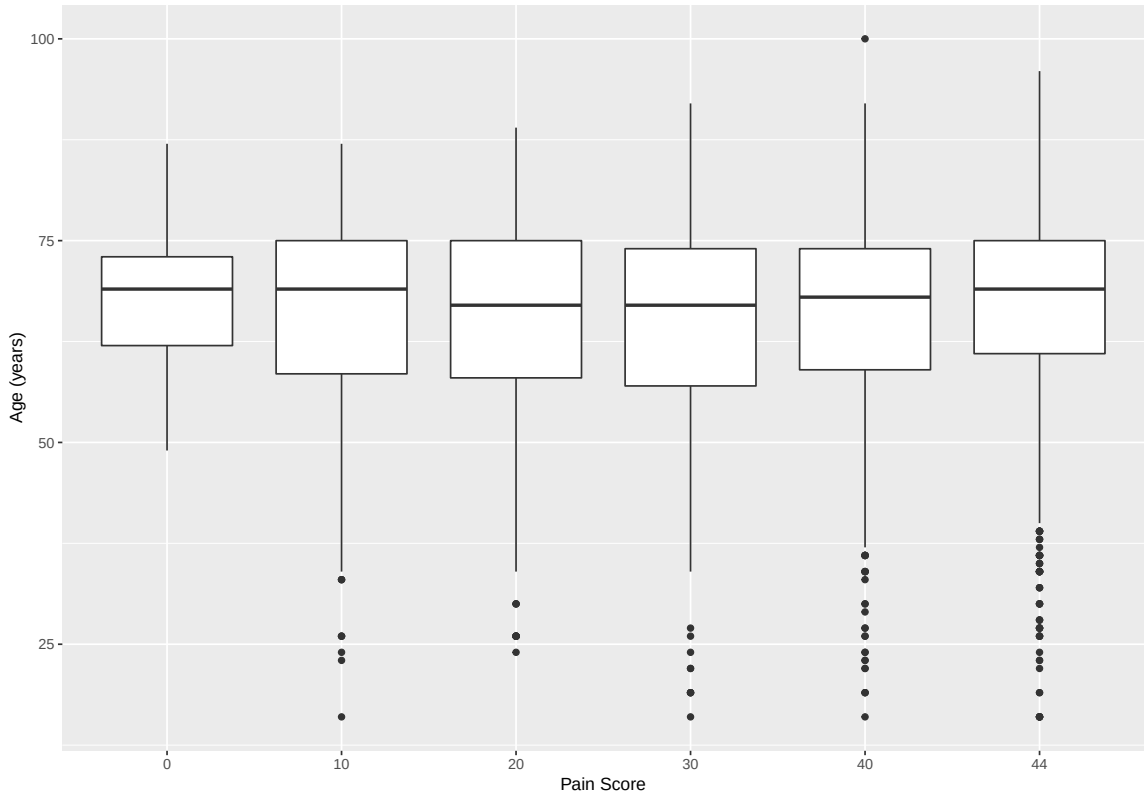

Figure 8.3: Side-by-side boxplot of the patient baseline age against pain score (n = 1796).

This joint modelling analysis with multivariate longitudinal outcomes will only address the predictor variables age, gender and follow-up time as this is merely foundational work. As a result, this analysis includes 1796 patients, as these variables contain little to no missingness. Figure 8.3 shows a side-by-side boxplot of the age of patients against the pain score. This shows that there is not significant change in age distribution as a patient's pain increases. There are also more young patients, evident in the outliers, as the pain

score increases. This implies that patient's who are younger, and are outliers in the age distribution, appear to have less pain.

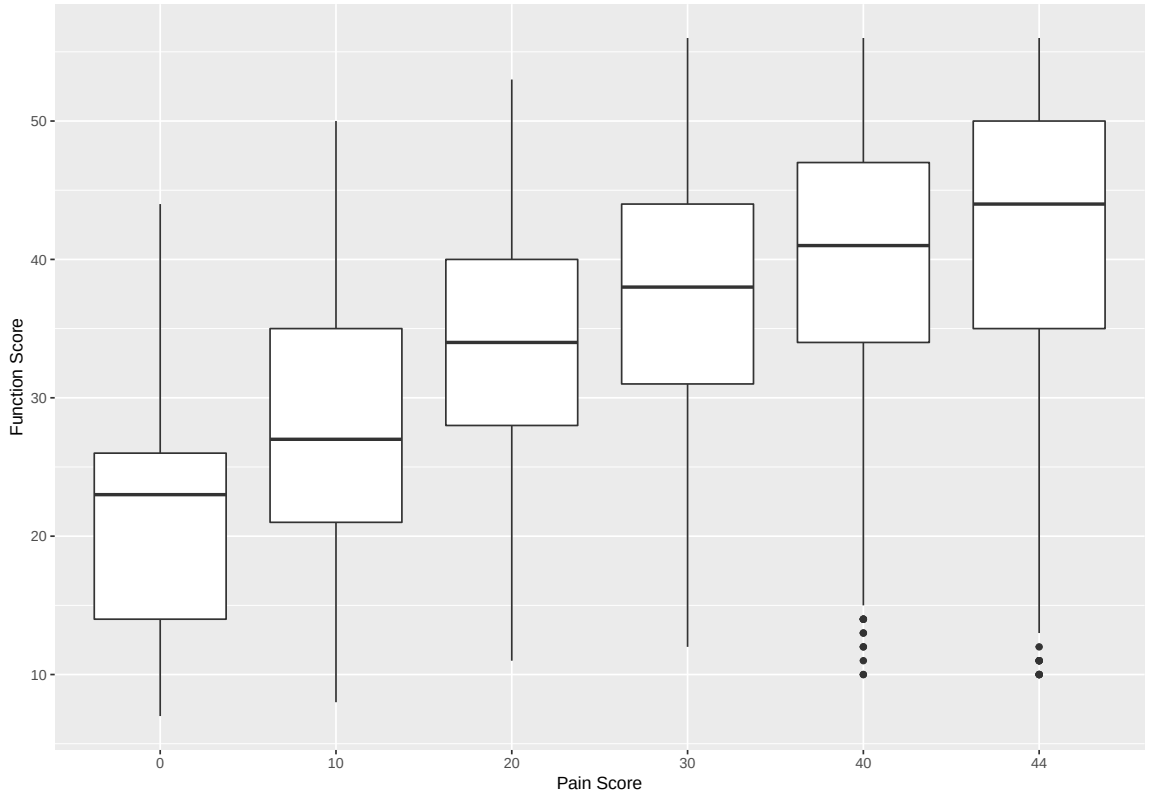

Figure 8.4: Side-by-side boxplot of the function score against the pain score for THR patients (n = 1796).

Figure 8.4 shows a side-by-side boxplot of the distribution of the function score against the pain score. It is clear from this plot that a patient who experiences greater pain is more likely to have poor function, and the converse is also true. Though there are some exceptions, as seen in the outliers, the medians show a clear increase in pain as function worsens. Therefore, there appears to be a linear relationship between these two longitudinal outcomes.

After completing univariate and bivariate analysis of both the pain and function outcomes we now proceed to modelling these outcomes separately, before exploring joint

modelling. A mixed-effects model frame work was used to model both the pain and function outcomes separately. The only predictor variables included were baseline age, gender and follow-up time. The optimal random effects structure was determined using the same model selection process shown in Chapter 4. Therefore, the random effects structure was determined first and the fixed effects structure was unchanged. This was done to keep the predictor variables across all sub-models the same as this is a preliminary model.

The two mixed-effects models, one with the pain score as the outcome and the other with the function score as the outcome, can be written as

$$\begin{aligned}\text{PainScore}_{ij} = \; & \alpha + \beta_1 \text{Sex}_i + \beta_2 \text{Age}_i + \beta_3 \text{FollowupTime}_{ij} + \\ & b_{i0} + b_{i1} \text{FollowupTime}_{ij} + \epsilon_{ij}. \\ \text{FunctionScore}_{ij} = \; & \alpha + \beta_4 \text{Sex}_i + \beta_5 \text{Age}_i + \beta_6 \text{FollowupTime}_{ij} + \\ & b_{i0} + b_{i1} \text{FollowupTime}_{ij} + \epsilon_{ij}.\end{aligned}$$

where the term b_{i0} is the random intercept and b_{i1} is the random slope for follow-up time. Therefore, the optimal random effects structure for both the pain and function score mixed-effects models included random slopes and intercepts. The estimated parameters for these models are calculated using REML and are shown in Table 8.3. We note that in this model, the correlation between the outcomes and survival endpoint are not incorporated into the random effects and so are difficult to interpret.

In relation to the pain score, Table 8.3 suggests that being male is associated with less pain. The variables follow-up time and age are not significant within this model. The variability of the intercept across patients has a standard deviation of 6.70, while 8.34 is the standard deviation of the variability within patients. This suggests that the mixed-effects model is potentially accounting for the correlations within patient repeated

		Pain Score		Function Score	
	Parameter	**Estimate**	**p-value**	**Estimate**	**p-value**
Fixed Effect Estimates	Intercept (score at 0 months)	34.01	<.0005	47.45	<.0005
	Slope (months)	-0.005	0.24	-0.009	0.02
	Sex	1.05	0.01	3.40	<.0005
	Age	0.03	0.08	-0.15	<.0005
			Standard Deviation		
Component of Variance Estimates	Within Patient Residual	8.34		6.22	
	Between Patient Residual (intercept)	6.70		8.68	
	Between Patient Residual (slope)	0.05		0.06	
	$\text{corr}(b_{i0}, b_{i1})$	-0.21		-0.34	

Table 8.3: Estimated model parameters for pain and function score mixed-effects models. For both models, estimated parameters, p-values, and components of variance estimates are reported (n = 1796).

measurements. There is a negative correlation between the intercept and slope. This implies that when a patient's intercept increases by one unit of standard deviation, that patient's slope decreases by 0.21 standard deviations. Though this sub-model is not optimal, for the purpose of exploring an extension of joint modelling, it is sufficient.

In relation to the function score, Table 8.3 suggests again that being male is associated with better function. This model suggests that follow-up time after surgery is associated with a decrease in function score. Patients who are older are also associated with worse function than younger patients. The within patient variability is smaller than the intercept variability. The correlation between the intercept and the slope is again negative.

This suggests that a patient that starts with better function is associated with a more negative slope or trajectory than a patient of the same characteristics who starts with worse function.

Up until this point we have only addressed univariate joint modelling. Nevertheless recent work in the field of joint modelling has provided new ways to address the application of multivariate joint models. Specifically the work done surrounding the joineRML package will be explored and applied in this section. This package implements a multivariate joint model that is an extension of the classical model for the multivariate longitudinal outcomes case and is described by Hickey *et al.* [14]. The implementation of this model relies on an algorithm proposed by Lin *et al* [18] with some changes to decrease computational time. These alterations include a quasi-Newton update approach, variance reduction method and dynamic Monte Carlo updates. What follows is a summary of the theory used to implement this joint model.

Let's suppose we have k longitudinal outcomes. For each subject $i = 1, \ldots, n$ the vector of longitudinal outcomes is $\boldsymbol{y}_i = (\boldsymbol{y}_{i1}^T, \ldots, \boldsymbol{y}_{iK}^T)$, where $\boldsymbol{y}_{ik}$ is the observed measurements for the k-th longitudinal outcome. Therefore, each measurement is recorded at time t_{ijk}, where j is the number of observations recorded for each patient. As with previous joint models, T_i^* is the time as which the event occurs. Due to potential censoring the observed event time is $T_i = \min(T_i^*, C_i)$, where C_i is the censoring time (in this case the end of 2013).

As with the joint model formulation shown in Chapter 7 these two separate submodels, longitudinal and survival, are linked using association parameters. This model relies on a latent process, specifically

$$W_i(t) = \left\{ W_{1i}^{(1)}(t), \ldots, W_{1i}^{(K)}(t), W_{2i}(t) \right\}$$

which is a $(K+1)$-variate Gaussian process with mean zero. This process is determined independently for each patient. In this case, the k-th longitudinal sub-model is written as

$$y_{ik}(t) = \boldsymbol{x}_{ik}^T(t)\boldsymbol{\beta}_k + W_{1i}^{(k)}(t) + \epsilon_{ik}(t)$$

where $\boldsymbol{x}_{ik}^T(t)$ is the vector of covariates and $\boldsymbol{\beta}_k$ are the fixed effects parameters for the covariates. $W_{1i}^{(k)}(t)$ is defined as

$$W_{1i}^{(k)}(t) = \boldsymbol{z}_{ik}^T(t)\boldsymbol{b}_{ik},$$

where $\boldsymbol{z}_{ik}^T(t)$ is the vector of covariates and $\boldsymbol{b}_{ik}$ is the corresponding random effects for the i-th subject relating to the k-th outcome. $\boldsymbol{b}_{ik}$ has a multivariate normal distribution with mean zero. The variance-covariance matrix for this term is $\boldsymbol{D}_{kk}$. To address the dependance between multiple longitudinal outcomes we specify that $\text{cov}(\boldsymbol{b}_{ik}, \boldsymbol{b}_{il}) = \boldsymbol{D}_{kl}$ for $k \neq l$. Similarly as before, we assume that the error terms, $\epsilon_{ik}(t)$, and the random effects terms, $\boldsymbol{b}_{ik}$, are uncorrelated.

It is also necessary to specify the survival sub-model. This can be written as

$$h_i(t) = h_0(t) \exp\left[\boldsymbol{v}_i^T(t)\boldsymbol{\gamma}_v + W_{2i}(t)\right],$$

where $\boldsymbol{v}_i^T(t)$ is the covariates associated with the hazard, $\boldsymbol{\gamma}_v$ is the vector of regression coefficients and $h_0(t)$ is the baseline hazard. This hazard model is conditional on both $W_i(t)$ and the inputed covariate data. Therefore, the longitudinal and time-to-event processes are conditionally independent. To create the desired latent association we define

$W_{2i}(t)$ as a linear combination of $\left\{W_{1i}^{(1)}(t), \ldots, W_{1i}^{(K)}(t)\right\}$, such that

$$W_{2i}(t) = \sum_{k=1}^{K} \gamma_{yK} W_{1i}^{(k)}(t)$$

where $\boldsymbol{\gamma}_y$ are the association parameters. In our case, k is two as there are two longitudinal outcomes, pain score and function score. The latent process that links our sub-models is

$$W_i(t) = \{W_{1i}^{(1)}(t), W_{1i}^{(2)}(t), W_{2i}(t)\}$$

where k can take the values one or two. Therefore, the joint model for multivariate longitudinal outcomes can be defined as

$$\begin{cases} y_{ik}(t) = \boldsymbol{x}_{ik}^T(t)\boldsymbol{\beta}_k + W_{1i}^{(k)}(t) + \epsilon_{ik}(t), \\ h_i(t) = h_0(t) \exp\left[\boldsymbol{v}_i^T(t)\boldsymbol{\gamma}_v + W_{2i}(t)\right], \end{cases} \tag{8.1}$$

where $W_{2i}(t) = \gamma_{y1} W_{1i}^{(1)}(t) + \gamma_{y2} W_{1i}^{(2)}(t)$ and the variables used are outlined previously in this section. This model is then estimated using likelihood estimation and the Monte Carlo Expectation Maximisation algorithm (MCEM). MCEM effectively determines the maximum likelihood estimates of the parameters through addressing the random effects as missing data [32].

Therefore, a joint model with multivariate longitudinal outcomes, namely pain score and function score, and survival outcome was analysed using the joineRML package [14]. The results of this joint model are shown in Table 8.4. Since this is a preliminary model the results shown are limited, nevertheless they are still of interest. It is clear that for both the pain and function scores, being male is associated with a higher score, meaning less pain

and greater function. Interestingly, the variable follow-up time is not significant within the pain score sub-model. However, follow-up time is significant within the function score sub-model, suggesting that greater time after surgery is associated with worse function. The variable age is not significant within the pain score sub-model but it is highly significant within the function score sub-model. The older a patient is the higher their association with a worse function score.

Longitudinal sub-model					
Outcome	Variable	Estimate	SE	p-value	
Pain Score	Intercept	34.14	1.26	<0.001	
	Baseline age (years)	0.03	0.02	0.11	
	Follow-up time	-0.0047	0.0057	0.40	
	Gender	1.10	0.45	0.01	
Function Score	Intercept	47.48	1.32	<0.001	
	Baseline age (years)	-0.15	0.02	<0.001	
	Follow-up time	-0.009	0.004	0.04	
	Gender	3.45	0.48	<0.001	
Survival sub-model					
	Variable	Estimate	SE	p-value	HR
	Pain Score	0.07	0.02	<0.001	1.08
	Function Score	-0.11	0.01	<0.001	0.90
	Baseline age (years)	-0.02	0.005	<0.001	0.98
	Gender	-0.18	0.13	0.17	0.84

Table 8.4: Results of the joint model, which includes multivariate longitudinal outcomes. For both sub-models, estimated parameters, standard errors (SE) and p-values are reported (n = 1796).

The results of the survival sub-model, Table 8.4, also show further insights into the

pain and function scores in relation to a patient's survival probability. In this joint model gender does not appear to be a significant predictor of survival. Interestingly, this survival sub-model suggests that older patients are slightly associated with a lower hazard of death than younger patients. We also see that patients who experience better function have a lower hazard of death, as the HR for function score is less than one. This may be because mobility and function is often a reflection of greater holistic health. In relation to the pain score, patients with worse pain are associated with a lower hazard of death. This implies that a better pain score does not increase survival probability the same way that an increased function score does. This result may be explained by the fact that regression associations often do not imply causal relationships [5]. Alternatively, patients with high levels of pain may have been a priority for treatment, decreasing their hazard of death.

Our aim in this section was to present an introductory analysis using the *joint modelling with multivariate longitudinal outcomes* approach. There are limitations to this implementation. For example, the joineRML package only allows an association structure of $W_{2i}(t) = \sum_{k=1}^{K} \gamma_{yk} W_{1i}^{(k)}(t)$, though this association may take different forms. This model also does not account for a potential interaction between the longitudinal outcomes, pain and function score, within the hazard sub-model. Nevertheless, there is ongoing work occurring in this area to address the gap that the package, and research around, joineRML has begun to fill.

Chapter 9

Conclusion

9.1 Summary

In this thesis we have investigated how to predict postoperative patient pain and function based on preoperative factors, such as age and gender, for total hip replacement (THR) patients. Previous analyses into patient recovery from THR are not in terms of subjective reports of symptomatic relief, such as the Harris hip score (HHS). Likewise, the increased risk of patient outcomes after THR due to comorbidities is not always addressed. Both of these concepts are addressed in a thorough exploration of the methods used to model longitudinal data. This exploration displayed several interesting results.

Using mixed-effects modelling, where the response variable was the HHS, we determined that suffering from comorbidities is associated with a lower HHS after THR surgery. This required a logit transformation of the HHS that behaved poorly at the upper bound, 100, throughout the analysis. The process of model selection showed that random intercepts, for each patient, and random slopes was the optimal random effects structure. Therefore, there is a relationship between HHS and follow-up time which is different for

each patient. In addition, patients who require external support before undergoing a THR are also associated with a lower HHS, in comparison to patients who do not require external support and patients who are older and heavier are also associated with a lower HHS after THR surgery.

The best prediction, for the optimal mixed-effects model, occurs for patients who exist within the estimation model frame making it optimal to have at least one HHS measurement to perform prediction. This was determined through an exploration of different methods of prediction for this model. For existing patients the best prediction utilised the best linear unbiased predictions (BLUPs) of the random effects of that patient. For new patients, those not within the estimation model frame, the best prediction required using the BLUPs of the random effects of patients with similar characteristics to the new patient. Nevertheless, if the sample size is not large enough to have a number of patients with similar characteristics, then using an average patient's BLUPs of the random effects is optimal.

Building on our previous results, we were able to predict whether a patient will experience any improvement in HHS at 24 months after surgery. This was done using a mixed-effect models with a new response variable, change in HHS, using baseline HHS measurements. For the majority of patients there is an overall improvement in HHS, for some this is not the case.

Within survival analysis, the number of comorbidities that a patient suffered from again proved to be significant. A patient with no comorbidities is associated with a significantly higher survival probability than a patient with any other number of comorbidities. Due to a large proportion of THR patients within this dataset who were deceased, survival status could also be applied as a survival outcome within joint modelling.

Joint modelling allowed us to perform dynamic prediction, a valuable tool for per-

sonalised medicine, using the foundation obtained from the separate longitudinal and survival analyses. We obtained the most appropriate joint model by including the previous optimal mixed-effects model predictors in the longitudinal sub-model. Joint modelling confirmed previous results, such as the association between comorbidities and HHS after THR surgery, but also allowed us to use dynamic prediction. Therefore, for specific patients, we can see how predictions relating to both the HHS and survival outcome are updated as new information is inputed.

Joint modelling was also applied where the survival outcome was second surgery. The results showed that patients who experience a lower HHS are more likely to undergo a second surgery. The longitudinal fixed effects results supported the results obtained in the previous joint model.

Despite our previous findings, analysis of pain and function separately provided more insight than when combining them into the HHS. This has important implications for the interpretation and application of results based on the HHS. Although pain and function were significantly correlated and had a joint predictor of sex when analysed separately, joint modelling revealed different predictors for each. In addition, with the survival sub-model, greater function lowered the hazard of death while less pain was associated with an increased hazard.

9.1.1 Future Work

An approach that was not explored within this study was greater scale imputation. Though some variables were imputed there were variables that were not imputed that contained missingness. This missingness in turn effected the sample size that could be used to create a certain model. Therefore, further work into appropriately imputing this

dataset would be valuable and help utilise all patient information.

The joint model with multivariate longitudinal outcomes relied on the package joineRML. For this method to be more widely used, it needs to have more flexible association structures and allow for interactions between longitudinal outcomes within the hazard sub-model. Therefore, further work in this area would be an invaluable place to start.

Future research could expand the scope of the survival analysis that was completed. Specifically, to improve the joint modelling performed more time could be spend understanding the appropriate predictor variables to predict the survival risk of a patient. In this research only a simple survival sub-model was used that contained age, gender and follow-up time. Futhermore, all results were obtained using ML estimation, therefore Bayesian inference could also be explored using the R package JMbayes.

Within this thesis we have assumed that both the pain score and function score are continuous quantitative variables. In both cases these variables could be considered ordinal categorical variables as they are the summation of discretely answered questions in a questionnaire. Therefore, more time could be spent exploring the results of modelling these outcomes as ordinal categorical rather than continuous quantitative variables to see if this increases predictive accuracy.

Finally, we consider the purpose of this thesis to be twofold. Firstly, this research identifies some key variables that are significantly and consistently associated with worse pain and function after THR surgery. Therefore, this information can inform clinicians to encourage potential changes in their patients before undergoing THR to improve pain and function outcomes. For example, the number of comorbidities that a patient suffers from is a significant predictor across all longitudinal models. This means that comorbidities are shown in this research to have a direct effect on the pain and function of a patient after

THR surgery. Therefore, if a patient presents for a THR and suffers from a large number of comorbidities it is important to ensure that these comorbidities are being treated and addressed before undergoing surgery.

Secondly, this thesis outlines a process that can be used to obtain predictions relating to joint pain and function measurements after THR surgery. It also clearly outlines the limitations of these predictions to ensure careful interpretation of the results. These models can be used to predict a patient's potential pain and function at two years after THR surgery, at which point there should be a resolution of symptoms. The dynamic prediction, used in conjunction with joint models, allows for updating predictions that can inform the clinician of any changing or negative pain and function trajectories so that intervention can occur. This can then inform clinicians of the best way to address the health of a specific THR patient.

Appendix A

Appendix

A.1 Summary of Variables Within the THR Dataset

	Dataset Name	Description (units)	Type
Demographic	age	age at operation (years)	integer
	gender	gender of patient	1=male, 2=female
	dob	date of birth (dd/mm/yyyy)	date
	status	survival status	categorical
	dod	date of death (dd/mm/yyyy)	date
Employment	employ	employment	catgorical
	emp_oth	specified other employment status	free text field
	occ_desc	occupation description	free text field
	profile	profile	free text field
Lifestyle	smoker	current cigarette/tobacco smoker	1=no, 2=yes, 999=not known
	cig_day	number of cigarettes per day	free text field
	alcohol	current alcohol consumption	1=no, 2=yes, 999=not known
	alc_day	number of standard drinks per day	free text field
	soc_drgs	social drugs	1=no, 2=yes, 999=not known
	drgs_sp	social drugs *specified*	free text field
Residence	plc_res	place of residence	categorical
	pr_oth	other place of residence *specified*	free text field
	hm_alone	specify if living at home alone	categorical
	hm_others	specify if living at home with others	categorical
	nrsh_dt	date entered nursing home	free text field
Past medical history	pmh	inactive past medical history	1=no, 2=yes, 999=not known
	pmh_spec	inactive past medical history *specified*	free text field

	oth_maj	other major orthopaedic/musculoskeletal conditions	1=no, 2=yes, 999=not known
	om_spec	other major orthopaedic/musculoskeletal conditions *specified*	free text field
Previous Operation	prev_op	previous operations/complications in this area/joint	1=no, 2=yes, 999=not known
	op_desc	previous operation description	free text field
	op_yr	previous operation year	numeric
	op_hosp	hospital of previous operation	categorical
	op_month	previous operation month	numeric
	op4prev	previous surgery	2=yes, 1=no
	op4num	number of previous surgeries	integer
	op4plast	last type of surgery	free text field
	op4lstyr	last year of surgery (yyyy)	integer
Illness	curr_med	current medications	1=no, 2=yes, 999=not known
	cmed_sp	current medications *specified*	free text field
Comorbidities	comorbs	comorbidities	1=no, 2=yes, 999=not known
	co_morb	cormorbidities	categorical
	com_spec	comorbidities *specified*	free text field
Surgery	pat_id	patient id	numeric
	adm_no	admission number	numeric
Diagnosis	op4jdoa	diagnosis is osteoarthritis	2=yes, 1=no
	op4jdra	diagnosis is rheumatoid arthritis	2=yes, 1=no
	op4jd1cs	osteoarthritis is secondary to	3=dysplasia, 4=trauma, 5=other diagnosis
	op4jd1oc	other diagnosis (in relation to op4jd1cs)	free text field
	op4AVN	diagnosis is avascular necrosis	2=yes, 1=no
	op4jdsot	diagnosis is another joint disease	free text field
	op4fra	diagnosis is neck of femur fracture	2=yes, 1=no

	op4fcoth	other fracture diagnosis	free text field
	op4othsp	other diagnosis specified	free text field
	adm_diag	admission diagnosis	free text field
	other_diag_all	other diagnosis *specified*	free text field
	hist_cmp	history of presenting complain	free text field
Operation	adm_date	admission date (dd/mm/yyyy)	date
	op1opdte	operation date (dd/mm/yyyy)	date
	op4side	joint side	1=left, 2=right, 3=bilateral
Implant	op5acnme	name of acetabulum implant	categorical
	op5acss	survival of acetabulum implant	categorical
	op5fnme	name of femoral implant	categorical
	op5fss	survival of femoral implant	categorical
Follow-up*	form_date	form date (dd/mm/yyyy)	date
	fuptime	follow-up time (months)	numeric
	fuptimeyrs	follow-up time (years)	numeric
	htrend	trendelenburg test	categorical
	hlurch	lurch	1=no, 2=yes, 999=not known
	hlength	leg length	1=equal, 2=short, 3=long, 999=not known
	hdist	leg length measurement (cms)	numeric
Power	hflexpw	flexion	categorical
	habdpw	abduction	categorical
Range of Motion	hextn	extension	only valid entry is 10
	hflexn1	flexion	categorical
	hflexn2	flexion	categorical
	habd1	abduction	categorical
	habd2	adduction	categorical

	hextint1	external rotation	categorical
	hextint2	internal rotation	categorical
Pain	pain	pain from hip	categorical
	hsrcbck	other sources of pain referred to the high/thigh region	1=no, 2=yes, 999=not known
	hsrcoth	other sources of pain referred to the high/thigh region *specified*	free text field
	hstress	pain on stressing the hip or extremes	1=no, 2=yes 999=not known
	htender	tenderness	categorical
	htenderother	tenderness *specified*	free text field
Mobility	mob_abhs	able to get around the house	categorical
	mob_outh	able to get out of the house	categorical
	mob_shop	able to go shopping	categorical
Scores	harris	harris hip score	numeric
	charnley	charnley score	numeric
Complications and	hcomp	complications since last hip surgery	1=no, 2=yes, 4=NA, 999=not known
Dislocations	hcompsp	complications since last hip surgery *specified*	free text field
	hdislcn	number of dislocations since surgery	0=Nil, 1=1, 2=2, 3=>2, 999=not known
	hdisyr	year of dislocations (yyyy)	numeric
Notes	alerts	alerts	free text field
	auditnotes	notes	free text field

Table A.1: Table of Variables contained within the Total Hip Replacement dataset; * repeated measurements.

A.2 Harris Hip Score Calculation

The equation used to calculate the Harris hips score is as follows:

Harris hip score = (pain value) + (limp value) + (support value) + (distance walked value) + (stairs value) + (shoes value) + (sitting value) + (public transportation value) + (absence of deformity value) + (range of motion value).

The overall score for the range of motion is calculated as follows:

Range of motion value = (SUM((value) $\times$ (index factor))) $\times$ 0.05.

This scoring system allows for a maximum of 100 points (pain 44 function 47 absence of deformity 4 range of motion 5). To calculate the value of (pain value) + (limp value) + (support value) + (distance walked value) + (stairs value) + (shoes value) + (sitting value) + (public transportation value) use Table A.2. To calculate the score for the absence of deformity use Table A.3. And lastly to calculate the score for the range of motion use Table A.4.

Variable (column name)	**Categories**	**Value in dataset**	**Equivalent value for HHS**
pain (parhpain)	none or ignore it	44	44
	slight, occasional pain	40	40
	mild plain	30	30
	moderate pain	20	20
	marked pain, severe at times	10	10
	totally disabled, crippled by pain	0	0
limp (palimp)	none	1	11
	slight	2	8
	moderate	3	5
	severe	4	0
	unable to walk without support	5	0
support (pasupp)	none	1	11
	walking stick for long walks	2	7
	walking stick most of the time	3	5
	one crutch	4	3
	two walking sticks	5	2
	two crutches	6	0
	frame	7	0
	not able to walk	8	0
walk (pawalk)	unrestricted distance	1	11
	greater than 10 blocks	2	8
	five to six blocks	3	8
	two to three blocks	4	5
	less than one block	5	5
	housebound	6	2
	wheelchair bound	7	0
	bedridden	8	0
stairs (pastair)	normally, without using railing	1	4
	normally, with railing or device	2	2
	two feet on each step	3	1
	any other method	4	1
	unable to	5	0
shoes (pashoe)	no difficulty	1	4
	some difficulty	2	2
	unable to	3	0
sitting (pasit)	in ordinary chair, for one hour	1	5
	in a high chair, for half an hour	2	3
	unable to sit comfortably	3	0
public transport (papublic)	yes	1	2
	no	2	1

Table A.2: Pain calculation for HHS table.

Variable (column name)	**All of the following must be present**	**4 points**
flexion (hflexn1&hflexn2)	$< 30^o$ fixed flexion contracture	
abduction (habd1)	$< 10^o$ fixed adduction	
internal rotation in extension (hextint2)	$< 10^o$ fixed internal rotation in extension	
limb-length discrepancy (hdist)	limb-length discrepancy $< 3.2cm$	

Table A.3: Absence of deformity calculation for HHS table.

Variable (column name)	**Range (degrees) in dataset**	**Values**	**Index factor**
flexion (hflexn1&hflexn2)	0-45	0-45	1.0
	45-90	0-45	0.6
	90-110	0-20	0.3
abduction (habd1)	0-15	0-15	0.8
	15-20	0-5	0.3
	> 20		0
external rotation in extension (hextint1)	0-15	0-15	0.4
	>15		0
internal rotation in extension (hextint2)	any		0
adduction(habd2)	0-15	0-15	*0.2*

Table A.4: Range of motion calculation for HHS table.

A.3 Summary of Variable Relabellings and Changes

Variable (column name)	**Regrouping**	**Levels**
Medications (cmed_sp)	**Paracetamol**	Paracetomol
	Paracetamol	Immune System Supplement
	Acetaminophen	NSAID's, non-selective cox
	Panadol	inhibitor
	Immune System Supplement	NSAIDs, cox 2 inhibitors
	Methotrexate	Opioids, weak
	NSAID's, non-selective cox	Opioid, strong
	inhibitor	
	Aspirin	
	Ibuprofen	
	Diclofenac, Voltaren	
	Indomethacin	
	Ketoprofen, Orudis	
	Naproxen, Naprosyn, Naproxyn	
	Piroxicam, Feldene	
	Sulindac, Clinoril	
	Dolobid	
	Meloxicam	
	NSAIDs, cox 2 inhibitors	
	Celecoxib, Celebrex	
	Opioids, weak	
	Codeine	
	Opioid, strong	
	Morphine	
	Oxycodone	
	Fentanyl	
	Tramadol	
	Methadone	

	Buprenorphine Endone Pethidine	
Employment Specified (emp_oth)	Free text field	Disability Pension No Disability Pension
Smoker (smoker & cig_day)	**Non-smoker** No (smoker) **Light Smoker** Yes (smoker, unspecified cig_day) < 25 a day (cig_day) **Heavy smoker** ≥25 a day (cig_day)	Non-smoker Light smoker Heavy smoker
Alcohol (alcohol & alc_day)	**Non-drinker** No (alcohol) **Light/Moderate Drinker** Yes (alcohol, unspecified alc_day) < 6 a day (alc_day) **Heavy Drinker** ≥6 a day (alc_day)	Non-drinker Light/Moderate Drinker Heavy Drinker
Social Drugs Specified (soc_drgs)	Free text field	No Marijuana Marijuana
Residence (plc_res & hm_alone & hm_others)	**Home Alone & Independent** Home Alone (plc_res, unspecified hm_alone) Independent (hm_alone) **Home Alone & External Support** External Support (hm_alone) **Home with others & Independent** Home with others (plc_res, unspecified hm_others) Independent (hm_others)	Home Alone & Independent Home Alone & External Support Home with others & Independent Home with others & Support Other

	Home with others & Support Dependent (hm_others) External Support (hm_others) **Other** Nursing Home (plc_res) Other (plc_res)	
Survival of Implant (op5acss & op5fss)	**Revised** Revised **Unrevised** Unrevised, re-operation only Unrevised, no re-operation Deceased, re-operation **Lost to follow up** Deceased, revision status unconfirmed Lost to follow-up	Revised Unrevised Lost to follow up
Implant Name (op5acnme & op5fnme)	**Trilogy & CPT** Trilogy (op5acnme) CPT (op5fnme) **Exeter & Exeter or Vitalock & Exeter** Exeter or Vitalock (op5acnme) Exeter (op5fnme) **Other** All other implants	Trilogy & CPT Exeter & Exeter or Vitalock & Exeter Other

Table A.5: Summary of variable relabellings and changes.

Bibliography

[1] http://medical-dictionary.thefreedictionary.com.

[2] American Academy of Orthopaedic Surgeons. Femoroacetabular impingement. http://orthoinfo.aaos.org.

[3] E. Andrinopoulou, D. Rizopoulos, M.L. Geleijnse, E. Lesaffre, A.J.J.C. Bogers, and J.J.M. Takkenberg. Dynamic prediction of outcome for patients with severe aortic stenosis: application of joint models for longitudinal and time-to-event data. *BMC Cardiovascular Disorders*, 15(1):28, 2015.

[4] H. Apold, H.E. Meyer, B. Espehaug, L. Nordsletten, L.I. Havelin, and G.B. Flugsrud. Weight gain and the risk of total hip replacement a population-based prospective cohort study of 265,725 individuals. *Osteoarthritis and cartilage*, 19(7):809–815, 2011.

[5] O. Asar, J. Ritchie, P.A. Kalra, and P.J. Diggle. Joint modelling of repeated measurement and time-to-event data: an introductory tutorial. *International Journal of Epidemiology*, 44(1):334–344, 2015.

[6] Australian Orthopaedic Association. *National joint replacement registry. Hip, knee & shoulder arthroplasty. Annual report 2016*, 2016.

[7] K. Bjorgul, W.M. Novicoff, and K.I. Saleh. Evaluating comorbidities in total hip and knee arthroplasty: available instruments. *Journal of Orthopaedics and Traumatology*, 11:203–9, 2010.

[8] R. Bourne, S. Mukhi, N. Zhu, M. Keresteci, and M. Marin. Role of obesity on the risk for total hip or knee arthroplasty. *Clinical Orthopaedics and Related Research*, 465:185–188, 2007.

[9] D.R. Cox and D. Oakes. *Analysis of Survival Data*. Chapman and Hall, 1984.

[10] P. Diggle, P. Heagerty, K.Y. Liang, and S. Zeger. *Analysis of longitudinal data*. Oxford University Press, 2002.

[11] A. Gałecki and T. Burzykowski. *Linear mixed-effects models using R: A step-by-step approach*. Springer Science & Business Media, 2013.

[12] W.H. Harris. Traumatic arthritis of the hip after dislocation and acetabular fractures: treatment by mold arthroplasty. *The Journal of bone and joint surgery. American volume.*, 51(4):737–55, 1969.

[13] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer-Verlag New York Inc., 2009. [Second Edition].

[14] G.L. Hickey, P. Philipson, A. Jorgensen, and R. Kolamunnage-Dona. joineRML: a joint model and software package for time-to-event and multivariate longitudinal outcomes. *BMC medical research methodology*, 18(1):50, 2018.

[15] E.L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.

[16] E.W. Karlson, L.A. Mandl, G.N. Aweh, O. Sangha, M.H. Liang, and F. Grodstein. Total hip replacement due to osteoarthritis: the importance of age, obesity, and other modifiable risk factors. *The American journal of medicine*, 114(2):93–98, 2003.

[17] J.E. Knowles, C. Frederick, and A. Whitworth. *merTools: Tools for Analyzing Mixed Effect Regression Models*, 2018. R package version 0.4.1. `https://CRAN.R-project.org/package=merTools`.

[18] H. Lin, C.E. McCulloch, and S.T. Mayne. Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables. *Statistics in Medicine*, 21(16):2369–2382, 2002.

[19] R.J.A Little and D.B Rubin. *Statistical analysis with missing data*, volume 333. John Wiley & Sons, 2014.

[20] D.F. Moore. *Applied survival analysis using R*. Springer-Verlag New York Inc., 2016.

[21] A.K. Nilsdotter, I.F. Petersson, E.M. Roos, and L.S Lohmander. Predictors of patient relevant outcome after total hip replacement for osteoarthritis, a prospective study. *Annals of the Rheumatic Diseases*, 62:923–930, 2003.

[22] M. Noordzij, M. Van Diepen, R.C. Caskey, and K.J. Jager. Relative risk versus absolute risk: one cannot be interpreted without the other. *Nephrology Dialysis Transplantation*, 32(suppl_2):ii13–ii18, 2017.

[23] Australian Bureau of Statistics. *National Health Survey: Summary of Results, 2007-2008*. Commonwealth of Australia, Canberra, 2009.

[24] J.C. Pinheiro and D.M Bates. *Mixed-effects Models in S and S-PLUS*. Springer-Verlag New York Inc., 2011.

[25] C. Proust-Lima and J.M.G. Taylor. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment psa: a joint modelling approach. *Biostatistics*, 10(3):535–549, 2009.

[26] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017.

[27] D. Rizopoulos. JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software*, 35(9):1–33, 2010.

[28] D. Rizopoulos. *Joint models for longitudinal and time-to-event data: With applications in R*. Taylor & Francis Ltd, 2012.

[29] D. Rizopoulos, G. Verbeke, and E. Lesaffre. Fully exponential laplace approximations for the joint modelling of survival and longitudinal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(3):637–654, 2009.

[30] G. Tarr, S. Müller, and A. Welsh. mplot: An R package for graphical model stability and variable selection procedures. *arXiv preprint arXiv:1509.07583*, 2015.

[31] W.N. Venables and B.D. Ripley. *Modern applied statistics with S-PLUS*. Springer Science & Business Media, 2013.

[32] G.C.G. Wei and M.A. Tanner. A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American statistical Association*, 85(411):699–704, 1990.

[33] A.F. Zuur, E.N. Ieno, N.J Walker, A.A. Saveliev, and G.M. Smith. *Mixed effects models and extensions in ecology with R*. Spring Science and Business Media, New York, NY, 2009.