Analysis of short- and long-term mortality after hip replacement based on the primary care data

Muhammad Ally
Hussein Wahedally

Doctor of philosophy
University of East Anglia
Faculty of Medicine and Health Sciences
December 2017

Declaration of Authorship

I declare that this thesis titled, *Analysis of short- and long-term mortality after hip replacement based on the primary care data* and the work presented in it, are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Muhammad Ally Hussein Wahedally
9^{th} October 2018
m.wahedally@uea.ac.uk

Abstract

The number of primary total hip replacement (THR) procedures is increasing worldwide. An increase in early post-operative mortality is associated with THR while in the long term, advancements in lifestyle factors and surgical techniques for THR suggest an improvement in survival following the procedure. Mortality after primary THR is affected by a large number of confounding variables each of which must be considered to enable valid interpretation. Routinely collected data by general practices can provide useful insights on variations in short and long term mortality after primary THR procedure.

The primary objectives of this research were to investigate how a history of various medical conditions before THR procedures affect the short and long term mortality risk after the surgery for patients in the United Kingdom.

Medical records from 1987 to 2011 from general practices contributing to The Health Improvement Network (THIN) database were used to develop two specific mortality models: to estimate odds ratio of death during the first 24 months after the procedure and to estimate the long term hazard ratio of all-cause mortality following THR. Both mortality models were multilevel and included preoperative comorbidities, lifestyle and socio-demographic factors. These models produced accurate estimates of mortality risk after THR procedures that could inform professional healthcare on future medical management of THR patients and financial planning for retirement by patients, the actuarial industry, and the government.

This research found that in the first 24 months after THR, an excess mortality risk is associated with THR cases compared to controls and preoperative overweight, obesity, smoking, myocardial infarction and male gender increased short term odds of death for all types of THR procedures, compared to controls without these conditions. In the long term, hazard of all-cause mortality for THR cases was lower than controls, and was higher for patients with preoperative hypercholesterolemia, myocardial infarction, osteoarthritis and smoking compared to those without these conditions. However overweight and obese THR cases had better survival prospects than controls and THR cases with normal BMI.

Contents

List of Tables

List of Figures

Expectancy

Acknowledgements

In the name of God, the Most Gracious, the Most Merciful.

I would like to thank my supervisors Prof Elena Kulinskaya and Prof Alexander Macgregor for their time, continuous support, motivation and enthusiasm. In particular, I would like to express my sincere gratitude to Elena for her guidance as a statistician and sharing her immense knowledge in a wide range of technical areas and her assistance in all forms of writing for conference abstracts, academic papers and this thesis. I would like to thank Alex for sharing his medical expertise and experience in epidemiological studies.

I thank my family: my parents Fateemah and Feroz Wahedally, my brothers Mehraj and Asghar Wahedally and my uncle Ayaz Sayad Hussein for their continuous motivation and financial support throughout this research. Their passionate commitment in supporting me to complete this thesis was unfaltering. I would also like to thank Tanwir Mohungeer for her endless inspirations, enjoyable discussions, and for proofreading parts of my work throughout this research.

Finally I would also like to thank my friends Nadir Soobratty and Juneida Khodabacus and research colleague, Dr Ilyas Backbergenuly, for their spiritual support and guidance during this research. Their inputs helped me in explaining my work comprehensively.

1 Introduction And Background Information

1.1 Introduction

Over the last two decades, there has been a rapid increase in the number of medical registries and other sources of electronically collected medical data. These medical databases permit researchers to carry out prospective, retrospective or cross-sectional studies. The availability of huge number of databases containing specialised information, has contributed vastly to inter-disciplinary research. One such example is the use of primary care records to extract conclusive and essential information for medical and actuarial application. In this research, the main objective is to develop survival models that will explain variations in mortality risks after total hip replacement, for adults in the United Kingdom using data from primary care records. The following topics are presented in this introductory chapter: description of the surgical procedure of total hip replacement, trends in number of total hip replacements, actuarial and medical interests in mortality analysis after total hip replacement, objectives and aims of this research and the outline of this thesis.

1.2 An Overview of Total Hip Replacement Procedure

Structure of Hip Joint in Adults

Figure 1 (McCarthy et al., (2016)) below illustrates the structure of hip joints in adults. An adult hip joint is scientifically known as the *acetabulo-femoral* joint. It is the joint connecting the femur and acetabulum of the pelvis. Its primary function is to support the weight of the body while being stationary (for example standing) and during motions (for example walking or running). Therefore, hip joints are the most important part in retaining balance and hence, any condition affecting the hip joint will cause painful distress to the individual's physical movements.

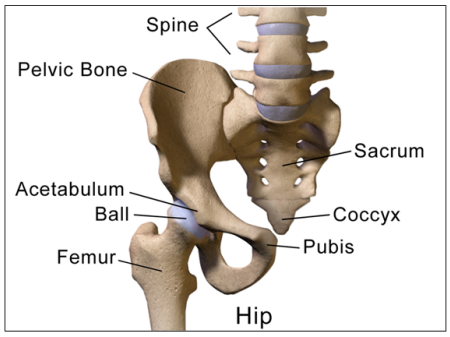


Figure 1: Frontal illustration of structure of hip in adults (McCarthy et al., (2016))

Hip Pain in Adults

Conditions affecting the hip joints are qualified as musculoskeletal. These are long term conditions that can cause a deterioration of an individual's quality of life due to pain in muscles, bones and joints. The most common problem with hip joint is hip pain. Such a condition is mainly caused by a degenerative disease known as osteoarthritis (commonly referred to as *coxarthrosis*) (McCarthy et al., (2016)). The term *osteoarthritis* (OA) is derived from three Greek words meaning bone, joint, and inflammation. OA is a progressive disorder of the joints caused by gradual loss of cartilage that acts as a protective cushion between the acetabulum ball and the pubis. As the cartilage is gradually worn away, the bone forms areas of abnormal hardening (commonly referred as spurs) and fluid-filled pockets (known as subchondral cysts) in the joint. As the disorder continues, pain results from deformation of the bones and fluid accumulation in the joints. Such a condition causes painful distress to the individual during physical movements. Figure 2 (Sanders, (2003)) below illustrates the gradual loss of cartilage between joints due to OA. In the advanced stage of OA, it can be observed that the cartilage is completely worn out causing the bone to erode as well causing uncomfortable pain to the individual.

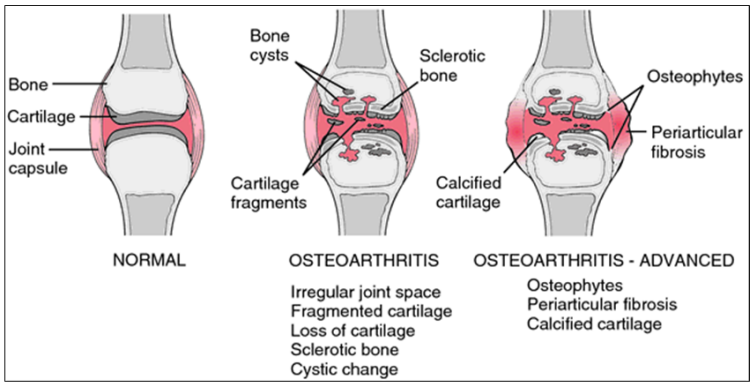


Figure 2: Loss of cartilage due to OA (Sanders, (2003))

History of Total Hip Replacement Procedures

Hstory of hip arthroplasty spans over more than 100 years. It has been seen as a landmark in twentieth-century surgery. The first known attempt to remodel a completely destroyed hip occurred in 1885 (Ollier,(1888)). Gluck (1891) was the first one to use foreign materials in the form of ivory components with nickel-plated steel screws for fixation. In the early twentieth century, the use of organic (fascia, fat) or inorganic (gold foil) materials in the form of membranes was common during resurfacing of hip. Mould arthroplasty (Smith-Petersen, (1948)) from 1923 and onwards, became famous while hemiprostheses of different designs and materials then prevailed until the early 1960's.

Since the early 1960, THR procedure has played an important role in alleviating pain and restoring mobility to millions of people suffering from arthritic joints. The success of THR was not the result of one single breakthrough. Its origins lay in the inter-positional and arthroplasty techniques developed between 1920 and 1950 in Europe and the United States. The evolution of procedures for putting materials between the articulating surfaces or for replacing one side of the hip joint led to materials, designs and surgical techniques that proved crucial to its success (Reynolds and Tansey, (2006)). However, THR was primarily a British innovation that started to take off in the late 1950s, under the National Health Service, but mainly in district general hospitals rather than in teaching hospitals. It was created

in hospital units in Norwich, Wrightington (near Wigan), Stanmore, Redhill and later Exeter (Parsons (1972), Duff-Barclay and Spillman (1966), Walker (1977), Klenerman (2002)).

The most important and highlighted breakthrough in the history of THR procedures occurred in the 1960s when Sir John Charnley introduced a new design of cemented hip prosthesis (Charnley, (1951)). Charnley hip prosthesis design is the standard reference to which new types of prosthesis are often compared to. The basic idea of Charnley was to use a cemented polyethylene acetabular cup and a small 22 mm femoral head that gave a low torque and low wear on a cemented femoral stem made of stainless steel. Further developments in the early 1970's concerned a new operation environment for THR procedures (Charnley (1972a, 1972b)).

Several designs with some attributes similar to the Charnley prosthesis but including new ideas were introduced to the market of hip prostheses in the 1970's. For example, the Norwegian orthopaedic surgeon Tor Christiansen designed a prosthesis (Christiansen (1969), Sudmann et al., (1983)) that was more anatomically correct and with a larger femoral head (37 mm wide) than the small Charnley prosthesis head (22 mm wide). Other prostheses from the 1970's and 1980's were the uncemented Bio-Fit femur prosthesis with a smooth surface and the double cup prostheses such as the Wagner prosthesis (Howie et al., 1990).

The history of hip replacement includes the use of several types of hip prostheses. However, there does not exist a general consensus on which type of prostheses are the best although cemented prostheses have been accepted as the best hip replacement technique for older patients. The debate on which type of hip replacement is most suitable for younger and active patients is still on although Lie (2002) showed a preference for newer uncemented designs with hydroxyapatite coating for younger and active patients.

Mechanism of Total Hip Replacement Procedure

The concept of total hip replacement consists of using a biochemical device, also known as a hip prosthesis, designed to completely replace the native joint (Figure 1) that has been painfully affected. The hip prosthesis consists

of a femoral head and an acetabular cup. The femoral stem is either of a modular or monoblock design as shown in Figure 3 (Hallan, (2007)). Under the monoblock design, the femoral head and stem come in one piece while under the modular design, the femoral head is free from the stem and is attached to the stem by using a taper locking mechanism. Similarly, the acetabular cup can either be monoblock where the cup comes together with the stem as one piece or modular where the cup consists of a shell that is attached to the pelvic bone and an insert (liner), which is fixed inside the shell. The bearing surface of the artificial joint is composed of a metal or ceramic femoral head while the inner surface of the cup is typically made of polyethylene (plastic), ceramics or metal.

The types of THR procedures are determined by the types of fixation technique used for total joint replacement. THR procedures involving artificial femoral head and acetabular cup of the joint fixed into the bone with acrylic cement, are referred as cemented THR (Morley et al., (2014)). Uncemented THR procedures involve fixing the artificial hip joint to the bone without using acrylic cement at all. The surface of the prosthetic joint is roughed up and often coated with bioactive materials such as hydroxyapatite and tricalcium phosphate to encourage the growth of bone onto the prosthetic joint so as to secure the prosthesis in place (Yamada et al., (2009)). THR procedures in which the femoral component is cemented into the bone while the cup is fixed without cement, is called hybrid THR procedures. Reverse hybrid THR procedures involve fixing the femoral component to the bone without cement while the acetabular cup is cemented into the bone (NICE, (2014)).

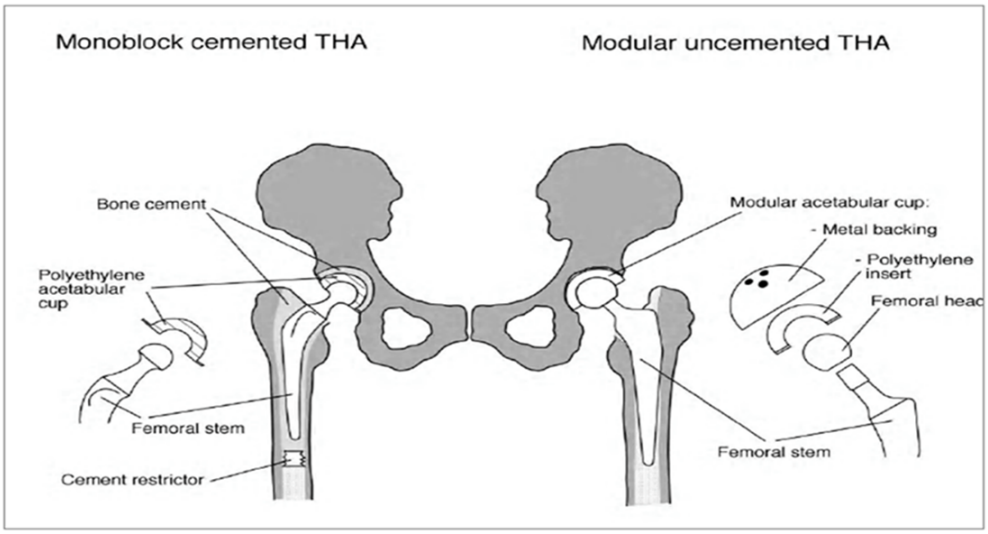


Figure 3: Illustration of hip prostheses (Hallan, (2007))

1.3 Increasing Number of Total Hip Replacement Procedures

There has been a rapid increase in the annual number of THR procedures performed worldwide. In 1980, the estimated number of THR procedures was between 300,000 and 400,000, per year. A decade later, the estimated number of THR cases jumped to approximately 800,000 annually (Levy et al., (1985)) and by 1991, over 1 million THR procedures per year were reported worldwide (Söderman, (2000)). In England and Wales only, the National Joint Registry (NJR) reported an increase of 80% in the number of THR procedures from 2003 to 2012 as shown by Figure 4 below (NJR, (2016)). This represents an average annual increase of 8% each year in England and Wales alone and therefore, the number of THR procedures is likely to keep on increasing with time. In this section, a summary of the impacts of the increasing trend in number of THR procedures on individual retirement planning, the actuarial and medical industry is provided.

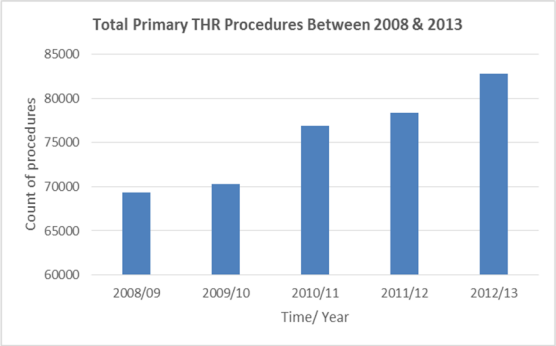


Figure 4: Total primary hip replacement procedures entered into the NJR between 2003 and 2013 (NJR, (2016))

1.3.1 Retirement Planning in UK

Retirement planning is the process of determining retirement income goals and the actions and decisions necessary to achieve those goals. It includes identifying income and expenses sources, implementing a savings program and managing assets for present and future access. In the UK, sources of income at retirement are (1) a pension provided by the state, (2) pension provided by employers, (3) defined benefits or defined contributions pension schemes and (4) personalised pension schemes (Office for National Statistics, (2013)). Access to the pension fund built up during the contribution years of the individual is allowed at the minimum retirement age, which is 55 years in the UK as at 2017 and this will be raised to 57 years in 2018 (Lain, (2016)).

Pension fund reforms enforced by the British government in April 2015 provided individuals with a greater flexibility in choosing retirement age and accessing their pension pots. From April 2015 onwards, people are allowed a 25% tax-free lump-sum from their total pension fund and then provided with the following options for the remaining 75% of the pension pot (Lain, (2016)):

(1) Withdrawal of the remaining 75% of the pension pot at the UK marginal tax rate

(2) Purchase of a life or term annuity from the government or private sector

(3) Purchase of flexible drawdown products, whereby individuals reinvest the remaining 75% of the pension pot in funds specifically designed and managed for providing an income after retirement. The income received, will vary depending on the fund's performance and is not guaranteed for life.

Changes in pension legislation in April 2015 were brought for two main reasons. Firstly, the life expectancy in the UK is increasing. The expected lifetime of a 65 years old in the UK was 16.1 years in 2001-2003 and 18.5 years in 2013-2015 for males and 19.1 years in 2001-2003 and 20.9 years in 2013-2015 for females, respectively (Office for National Statistics, (2016)). Secondly, the lifestyle of retired individuals have also changed. There are more active retired persons in the UK than there were 20 years ago (Lain, (2016)). Given that THR procedure is dominantly common among patients aged 55 or more (NJR (2017)), informing individuals who are planning for their retirement about mortality risk after THR, will assist them in making the appropriate choice for their pension fund, such that they have the optimum pension benefits at retirement.

1.3.2 Actuarial Impacts

Understanding and management of the longevity risks from various medical conditions and surgical interventions is an important part of actuarial research in the area of life insurance and pensions. The pricing of products sold to customers in the field of life insurance and pensions is highly dependent on the assumption or model used to estimate longevity risks for customers. An accurate estimation of longevity risks of customers is therefore highly desired in the insurance industry. In actuarial terms, longevity risk is defined as the uncertainty caused by future significant changes in mortality rates (Barrieu et al., (2012)). With the increasing number of THR procedures, the potential number of customers for life insurance companies undergoing THR procedures rises. Therefore it becomes important to know whether the longevity risks associated with this category of customers (those undergoing THR) are significantly different from those of the average customers. In this section, a summary of potential impacts caused by an increasing trend in number of THR procedures on the insurance and pensions industry, is provided.

The impact of THR procedures on life expectancy is still a controversial issue with conflicting results from epidemiological studies. If the life expectancy of patients undergoing THR procedures is improving, this presents challenges for life insurance and pension management companies in the way they adjust their existing assumptions on mortality during their pricing, reserving and profitability testing models.

Categorising THR patients as a group of individuals with different longevity risk allows insurers to build or adjust their pricing mechanism model of their products differently for customers with THR procedures and those without. For instance, when a group of individuals with a well defined medical condition, for example THR procedure, is identified to be associated with a higher risk of longevity, they are penalised in terms of the benefits they receive from pension and annuity products they bought, compared to insurance customers with a lower longevity risk. This benefit adjustment is referred as managing basis risk. In the insurance sector, basis risk is defined as the uncertainty associated with differences in assumptions that have been used for a particular model and the conditions observed in real life. Assuming a similar longevity risk for customers undergoing THR procedures as the average customers may represent a source of basis risk for the insurance and pension sector because of the existing uncertainty in the magnitude of the improvement in the life expectancy of individuals after THR procedures in the long run.

In addition, identifying sub-groups of individuals with different medical characteristic and classifying them according to their longevity risk, assist actuaries to design new types of enhanced annuities products that are tailored and more suitable to the customers' needs. In this way, life insurance and pension companies become more competitive in the insurance market. Managing longevity risks via categorisation of customers reduces the risk of overpricing or under-pricing their products. Therefore an educated management of longevity risk associated with THR procedures is of immense importance.

A common practice in the actuarial field is the categorisation of longevity risks according to their customers residential address or socio-economic factors such as Townsend and Mosaic scores. The use of postcodes in mortality investigations has become increasingly common in the UK, both for individual annuity and pension pricing. Therefore arises the need to investigate

how longevity risks associated with THR procedures, vary across residential areas with different socio-economic factors because such an investigation will aid the life insurance industry to identify areas of high or low longevity risks after THR procedures.

With increasing number of medical registries available today, insurers are looking to make better use of these huge databases. The appeal for the life insurance and pension industry is simple: "If you have already collected the data, then it is like leaving money on the table if it is not being exploited to the full. Worse, if your competitors make better use of their data, you can be selected against and lose money"(Richards, (2014)). As mentioned previously, use of residential addresses is increasingly common in the actuarial field. The biggest change has been in insurers' attitude towards the use of postcode related variables to analyse variability in longevity or mortality risk. Large medical databases are often linked to post-code related information and thus any extra value which can be squeezed out of these, represent a low-cost bonus information that insurance companies can use to edge out in the competitive insurance market.

Over the years, the insurance sector has employed gender as an underwriting factor when developing pension and annuity products. Women are statistically known to have higher life expectancy than men and thus the cost associated with insurance and pension policies sold to women is higher than males. This status quo is today challenged by the European Union with the introduction and implementation of the directive which requires all states of the European Union to ensure the respect of gender equality in all insurance contracts, as mentioned by the 5th Article of the European Council Directive (European Union Convention, (2004)). This challenging directive forces actuaries in the insurance sector to identify new sources of longevity risks. Understanding and learning from the longevity risk after medical procedures such as THR provides a tool for actuaries to identify new sources to explain variability in mortality or longevity risk.

In the light of the above arguments, there is a strong case to argue that the increasing trend in number of THR procedures in the United Kingdom and worldwide, will impact on the actuarial methods of managing mortality and longevity risk as it increases the likelihood that more customers of insurance and pension products will undergo THR procedures. Hence actu-

arial determination to price insurance and pension products, to reserve for future benefits and claims payment and to calculate the profitability of these products will be affected.

1.3.3 Medical and Healthcare Sector

Survival after THR or longevity risks associated with THR procedures is one of the main concerns of healthcare professional because of the growing demand for THR procedures. An increase in surgery age of patients undergoing the procedure, warrants an investigation of post-operative mortality rate after THR procedures. With an expected increase in number of THR procedures, medical professionals expect a larger distribution of ages at surgery time. Therefore, an analysis of mortality risks for different age groups at THR surgery will support medical professionals to identify age groups who are at high mortality risk.

Post-operative mortality risk associated with patients with a history of co-morbidities before their THR procedures, is still an area under development. Usually patients undergo a pre-selection assessment before their THR surgery. The medical history of patients is thoroughly checked for any cardiovascular diseases, kidney problems or any coronary heart disease before the surgery. With an expected increasing number of THR cases, surgeons face increasing difficulties in identifying patients at high mortality risk. This highlights the need to build a survival model that will determine the mortality risk of the patients given their medical history, prior to THR procedure.

Demand for THR prostheses is expected to increase in the future. As a result, new types of prostheses are expected to be offered into the market. Unfortunately, the impact of different types of prostheses on survival after THR is still a debatable area that needs to be addressed. The systematic reviews by Kynaston-Pearson et al. (2013) on primary hip replacement prostheses in 2013 found that 24% of all hip replacement implants available to surgeons in the UK have no evidence for their clinical effectiveness. This review also showed that 7.8% of the 136,593 components used in primary hip replacements in 2011 were implanted without readily identifiable evidence of clinical effectiveness. This points out that a considerable proportion of prostheses available to orthopaedic surgeons, have no readily available evidence of clinical effectiveness to support their use. A survival model estimating

post-THR mortality risk for different of types of THR prostheses will help to identify which types of prostheses are beneficial for survival after THR procedures.

With an ageing population in the United Kingdom, prevalence of chronic medical conditions such as chronic kidney diseases and cardiovascular diseases increases (Vos et al. (2015)). Higher prevalence of such medical conditions enhances the understanding in variability of mortality risks. With an increasing prevalence of chronic medical conditions, a higher degree of differentiation between patients is possible, including interaction of medical conditions with socio-demographic and lifestyle factors and treatments received. Hence a survival model assessing variations in mortality risk for patients with different chronic diseases prior to their THR surgery will provide an enhanced understanding of the role played by pre-existing medical conditions on individual mortality risk after THR procedures.

With increasing number of THR procedures in the UK, the prevalence of THR patients who receive drug therapy treatments for chronic conditions such as hypercholesterolemia, hypertension or type 2 diabetes, prior to their surgery, may also increase. The effects of drug therapy on mortality risk after THR procedures may vary across different socio-demographic, medical and lifestyle factors (Platt et al. (2008)). A survival model will help decision makers in the health sector to estimate the differences in effectiveness of drug therapy for THR patients, given that they were receiving treatments before their surgery.

The rising demand for THR procedures in UK also put pressure on professional healthcare providers such as clinicians, surgeons or policy makers. Allocation of resources may be more efficiently and strategically carried out to satisfy the increasing demand for THR procedures if a post-THR survival model is used to classify patients into categories of high or low mortality risk after surgery. Such an approach will aid professional healthcare providers in identifying group of THR patients with different risks of mortality and thus allocate their resources for the benefits of the patients, more efficiently.

In conclusion, there are strong reasons to argue that increasing number of THR procedures in the UK poses a challenge to healthcare professionals to provide their services more efficiently. Firstly, surgeons and general prac-

titioners need to assess the mortality risk for different age groups at time of THR procedures for patients with different medical conditions as demand for THR procedure at a younger age or for patients with different co-morbidities prior to surgery, may increase. Secondly, there is a strong need to understand the effects of different prostheses on mortality risk after THR procedures to assess their effectiveness. Thirdly, with increasing prevalence of THR procedures, effects of different drug therapy treatments on survival after THR procedures need a close examination. Therefore, the use of a survival model for mortality risk estimation after THR procedures will help to answer these questions and will provide a tool for health care service providers to work more efficiently and strategically.

1.4 Objectives and Aims of Research

The primary objectives of this research are to investigate mortality risks in the UK after THR procedure for individuals with different preoperative medical history and from different socio-economic groups. Using primary health care records from The Health Improvement Network (THIN) database, survival models to analyse variations in short and long term mortality risk after THR procedures, are developed, with the following goals:

(1) Estimate the mortality risk after primary THR procedures in the short and long term.

(2) Identify a list of risk factors or interactions between risk factors that cause variations in post-THR surgery mortality risk and estimate their effect size.

(3) Explain the medical and actuarial implications from estimated mortality risk after THR procedures.

The aims of the research are as follows:

(1) Select primary care records from the THIN database for identification of patients undergoing THR procedures.

(2) Determine the prevalence of co-morbidities such as angina, myocardial infarction, stroke and chronic conditions such as chronic kidney disease, type 2 diabetes, osteoarthritis and rheumatoid arthritis and modifiable conditions such as hypercholesterolemia, hypertension, body mass index and smoking status, prior to time of THR procedure of patients.

(3) Establish the prevalence of missing data among records extracted from the THIN database and determine the effects of missingness in the dataset on survival models after THR procedures.

(4) Estimate the effects of different types of THR procedures on mortality risk after THR procedures.

(5) Investigate whether the preoperative medical history (listed in aim (2)) and hormone replacement therapy (oestrogen, progesterone and testosterone) affect short and long term survival following THR.

(6) Investigate how mortality risks after THR procedures vary across demographic variables and socio-economic factors such as Townsend scores, Mosaic groups and index of multiple deprivation.

(7) Impact of research findings on clinical and actuarial management of patients undergoing THR procedures.

1.5 Thesis Outline

This section provides the outline of the following chapters of the thesis.

Chapter 2 is a review of mortality risk after primary total hip replacement. Two reviews are presented in this chapter; one for short term review of early post-operative mortality risk and another one for long term review of survival analysis following primary total hip replacement.

Chapter 3 is a review of statistical methods for survival analysis of censored hierarchical data. Firstly, two separate survival models, namely the Cox proportional hazards model with frailty and the logistic regression model with random effects, and their assumptions for model fitting, are, respectively, described in this chapter. Secondly, the process of model development with regard to the selection of covariates is described. Thirdly, the methodology in handling missing values in survival analysis is presented and fourthly, the assessment of the final survival models performance is explained.

Chapter 4 is a review of study designs used in the field of epidemiology, and primary care data in the UK and its use. Firstly, observational study designs are compared to experimental study designs. Secondly. an overview

of sources of primary care data and database in UK is discussed, in particular The Health Improvement Network (THIN) database, which is used in this research. Thirdly, a discussion on the generalisability of the THIN database is presented.

Chapter 5 describes the data set extracted from the THIN database for survival analysis after primary total hip replacement (THR). Firstly the criteria for identification of THR case and controls is described. Secondly, the distribution of patients across several demographic, life style and medical variables, is presented and compared to the UK population. Thirdly, a discussion on the proportion of missing data is provided.

Chapter 6 presents the short term survival model that estimated the all-cause odds of death during the first 24 months after THR procedure. Firstly, the model development strategy and the analysis procedure are described. Secondly, variations in short term odds of death are presented. Thirdly, an assessment and diagnostics of the survival model is provided. Finally the survival model and the variations in estimated short term odds of death are assessed and compared with the results of previous studies.

Chapter 7 presents the long term survival model that estimated the hazards of all-cause mortality after primary total hip replacement. The model development and analysis procedure is explained firstly. Secondly, the estimated variations in hazards of mortality are presented and compared with the results of previous studies. Thirdly, a discussion of the effect of THR procedures on effective age of THR cases with various preoperative medical conditions, is provided.

Chapter 8 is a discussion on the findings of this research. The main results of the short and long term models and their contributions to the existing evidence, are presented. Then the strengths and limitations of the research are discussed. Finally, the implications of this research findings in medical management and retirement planning for THR cases are presented by addressing the research aims.

2 Review Of Mortality Risk After Total Hip Replacement

2.1 Search strategy

This chapter is a literature review of mortality risk after primary unilateral total hip replacement (THR) procedures. The objectives of this review are to survey the existing mortality risk and survival models after THR procedures, report estimated risk of dying after THR procedures and produce a summary of the significant risk factors that contributed in explaining variations in mortality risks after THR procedures. A comprehensive review of all studies published between January 1990 and December 2017 inclusive in the English literature and containing mortality data for patients who had a primary THR procedure, was conducted electronically using the following online databases: *MEDLINE, Google Scholar,* the *Journal of Bone and Joint Surgery*, *ACTA Orthopedica* and the *Journal of Arthroplasty*, respectively. The search terms employed to identify relevant publications were: *total hip replacement* (THR), *total hip arthroplasty* (THA), *hip joint replacement, hip joint arthroplasty, short term THR* or *THA*, *long term THR* or *THA*, *survival analysis after THR* or *THA*, *multi-centre survival analysis after THR* or *THA*, *life expectancy after THR* or *THA* and *mortality risk after THR* or *THA*, respectively.

In total, 42 published journal articles were selected in this literature review. Mortality risk models after THR procedures were either developed to investigate survival in the short term or in the long term. It is essential to define the follow-up time of these 32 selected studies in order to classify them as either short or long term, respectively. According to the systematic review carried out by Berstock et al. (2014), an excess mortality risk in the short term (1, 3, 6 months and 2 years after THR procedure, respectively) is observed for individuals who underwent a THR procedure while in the long term, post-THR life expectancy is improved. Therefore mortality risk models presented in selected published studies in this review, were developed to investigate variations in survival after THR surgery either for the short or long term only. In this review, studies with a follow-up time smaller than 24 months after THR procedure are categorised as *short term* while studies with longer follow-up time are referred as *long term*. 17 studies (see Table

1) analysed mortality data in the short term while, 15 studies (see Table 4) reported estimated mortality risk in the long term after THR procedures. 11 studies were classified as controlled trial evidences of the effects of THR procedures on mortality risk after surgery (see Table)

The following items were derived from each published article and used as a basis of comparison, in this chapter, between different studies:

(1) *Follow-up time*: Length of study period over which individuals were observed after THR surgery.

(2) *Size*: Number of participants selected in the study.

(3) *Country*: Location of study.

(4) *Data source*: Database used as source of data to determine survival after THR procedures.

(5) *Study design*: Types of study carried out (Case-control, cohort matched, observational study).

(6) *Variables*: List of variables investigated to explain variations in mortality risk after THR procedures and identified risk factors.

(7) *Survival model*: Statistical model employed to estimate mortality risk.

(8) *Outcomes*: Reported mortality risk after THR procedures using different survival measures such as hazard of death (HR), odds of dying (OR), standardised mortality rate (SMR), crude mortality rate (CMR), Kaplan Meier survival estimate and percentage survivorship after THR procedures.

2.2 Short term review of mortality following THR procedure

17 studies out of 32 selected publications, listed in Table 1, reported mortality risk after THR procedures in the short term. Lie et al. (2000, 2002), Pedersen et al. (2007) and Boniello et al. (2017), respectively, investigated mortality risk after THR procedure exactly one month after the surgery while Lie et al. (2000) and Aynardi et al. (2009), respectively, had a follow-up time

of 2 months. The most common length of follow-up time in the short term is 3 months (9 studies out of 17: Lie et al. (2010), Blom et al. (2006), Williams et al. (2002), Fender et al. (1997), Hunt et al. (2013), Lovald et al. (2014), Barrett et al. (2005), Smith et al. (2015), Xu et al. (2017)). Only 2 studies investigated mortality risk in the short term with a follow-up time longer than 3 months; Jones et al. (2014) and Nunley and Lachiewicz (2003) analysed mortality risk 12 and 24 months after THR procedures, respectively.

The studies reporting short term mortality risk after THR procedures were carried out in various countries: England (3), England and Wales (3), Scotland (1), Denmark (1), Norway (3) and the United States (US) (6), respectively. The sample size of these 17 selected short term published articles on mortality risk after THR procedures vary between 835 patients for the smallest study in US (Nunley and Lachiewicz (2002)) and 409,096 patients for the largest study, carried out in England and Wales (Hunt et al. (2013)), respectively. In all 17 studies, the ratio of males to females in the sample size was always less than one (see Table 1), indicating that THR procedures are more prevalent among women than men, in various locations.

9 studies out of 17 used a register database as data source for survival analysis. A national registry database on patients who underwent THR procedures were used for studies carried out in Norway (Lie et al. (2002, 2000, 2010)), Denmark (Pedersen et al. (2011)), UK (Hunt et al. (2013) and Jones et al. (2014)), England alone via private surveys from various THR surgery centres located in East Anglia, Oxford, Trent, Northern and Yorkshire regions (Williams et al. (2002)) and US (Aynardi et al. (2009) and Boniello et al. (2017)), respectively. The source of data in all these 9 studies were therefore obtained from different surgery centres, hospitals or private clinics. None of these 9 studies accounted for the cluster effects due to grouping individuals by their surgery centre, hospital or private clinics in the database, to investigate mortality risk variations between each centre of data collection.

The remaining 8 short term studies investigate mortality after THR procedures using data from a single surgery centre, hospital or private clinic in their studies; Fender et al. (1997), Nunley and Lachiewicz (2003), Barrett et al. (2005), Ramiah et al. (2007), Blom et al. (2006), NHS Scotland (2002) and Lovald et al. (2014) used data from a single hospital or private

clinic while the US based study, Xu et al. (2017), used a private insurance company database in the US as a source of data to investigate short term mortality after THR procedures among insured customers. None of the 17 short term studies used primary care records as a source of data. Developing mortality risk models using primary care records can lead to different results in variations of mortality risk after THR procedures. Compared to registry or single centre database, primary care data has an extensive volume of socio-demographic, lifestyle and medical information, with a greater coverage or follow up of patients selected in a study. All these additional information can be used to explain post-operative mortality risk variations in the short term.

Nine studies out of seventeen studies reviewed were set up as a case-control study design using age at THR surgery, gender and year of birth as the main matching factors for selection of controls. Lie et al. (2000, 2002, 2010), Aynardi et al. (2009), Pedersen et al. (2011), Lovald et al. (2014) and Jones et al. (2014), respectively used age at time of THR procedure and gender as matching factors for selection of controls while Ramiah et al. (2007) used only age group at time of procedure. Barrett et al. (2005) used year of birth, age at time of surgery and race as matching factors for selection of controls for their study. The estimated effects of THR intervention on short term mortality risk in these case-control studies have a greater degree of generalisability than the cohort type studies (Blom et al. (2006), Williams et al. (2002), Fender et al. (1997), Hunt et al. (2013), Smith et al. (2015), Xu et al. (2017) and Boniello et al. (2017)) which investigated short term mortality risk after THR procedures by post-THR causes of death primarily, while Nunley and Lachiewicz (2003) estimated mortality risk after THR procedure across types and prognosis of THR procedure.

Among the variables used to explain variations in short term mortality risk in the seventeen studies reviewed, age at time of THR procedure, gender and osteoarthritis diagnosis prior to THR procedure (17) were the most common ones, followed by rheumatoid arthritis and myocardial infarction event (9), angina, stroke and osteoporosis (6), chronic kidney disease and revision surgery (5), respectively. Lie et al. (2000) and Nunley and Lachiewicz (2003) also compared variations in short term mortality for different types of THR procedure while Lie et al. (2002), Aynardi et al. (2009), Williams et al. (2002) and Xu et al. (2017) used revision surgery to compare short term mortality.

Hunt et al. (2013) was the only study to adjust for age, gender, body mass index, osteoarthritis, diabetes, chronic kidney disease, myocardial infarction, rheumatoid arthritis and osteoporosis, simultaneously in their analysis of short term mortality. Furthermore, variations in short term mortality were explained by ethnicity by Lovald et al. (2014) and Xu et al. (2017), while social deprivation index and hypertension were used additionally by Xu et al. (2017) to compare short term mortality.

Several statistical methods of mortality risk estimation were employed in the seventeen articles reviewed. Cox regression model and standardised mortality ratio were the most common methods used to estimate of mortality risks (5), followed by multivariate logistic regression (4), crude mortality rate and Kaplan Meier survival analysis (2), respectively. Furthermore, only one study reported proportion of missing values in their study; Hunt et al. (2013) reported that 10% of their whole study population had missing values and adjusted for this in their data analysis by employing multiple imputation technique assuming data were missing at random. The authors used imputation models that included all variables used in their Cox regression survival model to estimate mortality risk and also used the outcome variable (whether the patient is alive or dead by the end of the investigation) because the latter consist of information about missing values of the predictors of the imputed models.

In addition to the 17 publications listed in Table 1, the systematic review on short term mortality risk after THR procedures by Berstock et al. (2014) is also included. The authors critically estimated the 30- and 90-days overall mortality through meta-analysis of 32 published articles between 0.22%-0.38% and 0.50%-0.81%, respectively. Among the 17 studies reviewed in this section, Fender et al. (1997), Lie et al. (2002), Nunley and Lachiewicz (2003), Blom et al. (2006), Pedersen et al. (2011) and Jones et al. (2014) estimated the 30-, 60- and 90-days mortality rate after THR procedures between 0.27%-0.41%, 0.27%-0.75%, 0.45%-0.93%, respectively. Only Nunley and Lachiewicz (2003) reported 1-year mortality (1.66%) and 2-year mortality (4.0%) after THR procedure.

Among selected case-control studies, Lie et al. (2000), Williams et al. (2002), Ramiah et al. (2007), Aynardi et al. (2009) and NHS Scotland (2002), respectively, reported an estimated 30-, 60- and 90-days standardised mortal-

ity rate between 0.08%-0.10%, 0.06%-0.40% and 0.40%-0.90%, respectively while Lie et al. (2000), Hunt et al. (2013), NHS Scotland (2002) and Lovald et al. (2014) estimated the short term hazard ratio of death after THR procedures at 1.11 and 1.05-1.13, 60 and 90 days after THR procedures, respectively.

Based on these reported estimates of mortality rates or hazard ratio of death, it can be concluded that there is an excess mortality in the short term, at 30, 60, 90-days, 1 year and 2 years after THR procedure. There is strong evidence to suggest that mortality rates following THR procedures decrease with time; Hunt et al. (2013) reported a stable decrease of 90-days mortality rate from 0.56% in 2003, to 0.29% in 2011. These findings were based on a cohort of 409 096 patients with primary THR procedures in England and Wales and are also consistent with similar trends reported by Cram et al. (2011). Furthermore, Barrett et al. (2005) also investigated the increase in the short term mortality risk following THR surgery to estimate the duration of excess mortality among THR cases. The authors reported a crossing of survival curves of THR cases and age and gender matched controls at about 90 days, including adjustments for comorbidities. Barrett et al. (2005) also used a graph of smoothed Nelson–Aalen cumulative hazard estimates to demonstrate varying mortality risk over the first 90-days post-surgery. They found that the risk is highest in the first 30 days and plateaus at about 90 days, suggesting that short term mortality risk decreases with time after THR procedure until it has returned to its baseline level in patients undergoing THR. Similar trend for short term mortality is reported by Lie et al. (2002) with the duration of excess mortality continuing up to 24 months after THR procedures.

Ten studies out of 17 also reported short term survival after THR procedure by cause of death (see Table 1). Blom et al. (2006) reported that cardiovascular disease was the major cause of death (41.1%), followed by cerebrovascular diseases (23.1%) and pulmonary embolism (11.8%). Out of ten studies reporting post-THR survival by causes of death, six found myocardial infarction and pulmonary embolism, respectively, as the leading causes of death after the procedure (Blom et al. (2006), Aynardi et al. (2009), Pedersen et al. (2011), Hunt et al. (2013), NHS Scotland (2002) and Jones et al. (2014)), while in one small study, Nunley and Lachiewicz (2003) reported cerebrovascular events as the main cause of death. In study with

short term follow up over more than 2 months period post surgery, other causes of death such as malignancy (Pedersen et al. (2011)), rheumatic disease, mild liver disease, diabetes (Hunt et al. (2013)) and chronic kidney diseases (Boniello et al. (2017)) become significant cause of mortality. The pattern of mortality causes presented here is reflected by the conclusions of Singh et al. (2011) who estimated the short term incidence of adverse events after THR procedures. The authors reported that highest incidence was associated with cardiovascular complications (6.9%) followed by pulmonary embolism (4.0%), within 90 days post-THR procedure.

Reported risk factors for short term mortality in this review were either modifiable or non-modifiable. There is strong evidence which suggests that increasing age and male gender cause a significant increase in mortality risk following THR procedure. All studies but one (Aynardi et al. (2009)) found significant association between male gender and increased short term mortality risk. The 30- and 90-day mortality risk reported in the fourteenth (2017) Annual Report of the National Joint Registry for England, Wales and Northern Ireland ranges from 0.06% (both gender) and 0.16%, respectively, for men aged less than 55 (versus 0.06% and 0.21%, respectively, for women below 55) to 1.18% and 3.09%, respectively, for men aged over 80 years (versus 0.82% and 1.81%, respectively, in women aged above 80) (NJR (2017)). Furthermore, increasing age at time of surgery predisposes to a premature short term rise in mortality rate. In Hunt et al. (2013) study, Kaplan Meier estimates of mortality rate at 90 days were determined for different age group at time of THR procedure. The authors reported that 90 days mortality rate ranged from 0.08% in men aged less than 55 (versus 0.05% among women below 55) to 1.90% among men above 80 years (versus 1.13% among women above 80) and a strong association between increasing age at time of surgery and mortality rate.

Several studies investigated risk factors for variations in short term mortality rate after THR procedures by controlling for confounding variables using a multivariate regression analysis. In this review, a Charlson comorbidity index greater than three (Aynardi et al. (2009), Singh et al. (2011), Bozic et al. (2012) and Gaston et al. (2007)), use of general anaesthesia (Aynardi et al. (2009), Hunt et al. (2013)) and prior cardiovascular diseases (Mäkelä et al. (2014), Bozic et al. (2012), Memtsoudis et al. (2012)) were modifiable preoperative risk factors for early mortality following THR.

Comba et al. (2012) also reported that a history of cardiovascular diseases prior to THR procedure increased the risk of mortality by eight-fold. Bozic et al. (2012) reported a significant increase in 90-day hazard ratio of death among patients with metastatic cancer (HR=3.14), congestive heart failure (HR=2.11), dementia (HR=2.04), renal disease (HR=1.98), cerebrovascular disease (HR=1.40) and chronic pulmonary disease (HR=1.32) before their THR procedure. Hunt et al. (2013) estimated a ten-fold increase in mortality risk for patients with moderate to severe liver disease, a three-fold increase after myocardial infarction and two-fold increase following diabetes with complications and renal disease, with all conditions diagnosed prior to THR.

In addition to the articles listed in Table 1, Gaston et al. (2007) investigated the effects of pre-operative cardiovascular diseases and hypertension on short term mortality risk after THR procedures. The authors analysed prospectively collected data on 1744 patients who underwent primary elective THR between 1998 and 2004, inclusive and reported that 34% of THR patients were diagnosed with hypertension prior to their procedure. They found no statistically significant increased mortality rate at 3 months for both male and female patients with a history of cardiovascular disease and hypertension prior to their surgery, after adjusting for age and BMI. Similarly, the case-control study by Pedersen et al. (2011) analysed 90-days mortality rate after THR among patients with Type 2 diabetes, prior to their surgery. The authors concluded that mortality rate at 90 days was lower among THR cases with Type 2 diabetes than among the control population with Type 2 diabetes and hence, found no significant increase in 90 days mortality rate after THR for patients diagnosed with preoperative Type 2 diabetes.

Bozic et al. (2012) and Hunt et al. (2013) data analyses of the NJR database for England and Wales demonstrated that a lower 90-day hazard ratio of death was significantly associated with a body mass index (BMI) between 26 Kg/m^2 and 30 Kg/m^2 (HR=0.76), inclusive, relative to patients with a normal BMI between 19 Kg/m^2 and 25 Kg/m^2 (HR=1.00), inclusive. This suggests that there is a statistically significant protective effect of being overweight and obese, prior to THR, for overall short term hazard ratio of death in patients. This is known as the *obesity paradox* and is observed among other chronic conditions (Curtis et al. (2005), Bakaeen and Chu (2011), Stamou et al. (2011)). Obesity is in itself, strongly linked with

an increased risk of developing conditions such as cardiovascular disease or hypercholesterolemia. However once the condition is manifested, obesity protects against premature mortality when compared with non-obese patients.

6 studies reported variations in short term mortality risk after THR by types of fixation techniques used during the procedure. Lie et al. (2000), Nunley and Lachiewicz (2003), Ramiah et al. (2007), Hunt et al. (2013) and Bozic et al. (2012) estimated that 0 to 90 days hazard ratio of death following THR is between 1.07 and 1.13 for uncemented procedures and 0.85-1.08 for hybrid procedures, respectively, compared to cemented procedures (HR=1.00). Therefore, these studies suggest that uncemented and hybrid THR procedures are significantly associated with a higher short term hazard ratio of death than among cemented procedures. Aynardi et al. (2009) estimated short term mortality rates for 7478 THR procedures consisting of both primary and revision cases. The authors reported that overall 90-days mortality rate with revision surgery was 1.24%, compared to 0.41% for patients without revision surgery and also demonstrated that it decreases with increasing age group at THR surgery. A similar conclusion is reported by Lie et al. (2002) who estimated a significant association between increased short term mortality rate and revision surgery within 90 days after the procedure (odds of dying=1.20-1.90).

Among the diagnoses, osteoarthritis (OA) and rheumatoid arthritis (RA) were the main reasons patients chose to undergo THR procedures in all 17 studies while, Lie et al. (2002) also compared OA and RA to hip fracture. Estimated standardised mortality rate for THR cases with OA and RA ranged between 0.06%–1.60% and 0.20%–4.80%, across all age groups, respectively, in all reviewed studies. Lie et al. (2002, 2000) estimated short term odds ratio of death following THR between 1.40 and 3.90, respectively, among RA patients, compared to patients diagnosed with OA. Hence in the short term, RA is significantly associated with an elevated odds of death after THR procedures, compared to OA diagnosis prior to THR.

A number of studies have suggested that low surgeon experience and hospital volume is significantly associated with increased short term mortality risk after THR (Judge et al. (2006), SooHoo et al. (2010), Memtsoudis et al. (2012) and Clement et al. (2011)). They examined surgeon procedure volumes and observed a lower rate of adverse events and mortality rate after

THR in patients who received treatments by surgeons who performed on average more than 25 procedures per year (estimated mortality rate=0.57%), compared to surgeons performing less than 10 procedures, on average per year (estimated mortality rate=2.55%). The systematic review on short term mortality risk by Berstock et al. (2014) stated that it is difficult to make a firm conclusion regarding the effect of surgeon or surgery centre volume size on mortality because of the vast differences in surgeon's training and healthcare organisation across different locations. Similarly, Clement et al. (2011) investigated the effect of social deprivation index (using the Carstairs index) on 90-day mortality risk after THR procedures. The authors reported that odds of dying at 90 days after THR is about three times higher for patients from the most deprived areas (odds ratio=3.2), compared to patients from low deprivation index residential areas.

Table 1: Review of published studies on short term mortality risk after total hip replacement

Author (Year)	**Country**	**Follow up time (Months)**	**Data Source**	**Sample Size**	**Ratio (Male:Female)**	**Variables**	**Survival Model**	**Outcomes Reported**
Fender et al. (1997)	England	1	Private Institution	2036	0.6	Age group, cause of death	Logistic Regression	(1) 42-days post-THR surgery mortality rate ranges between 0.82%-1.05%.
Lie et al. (2000)	Norway	2	Norwegian arthroplasty registry (NAR)	39543	0.4	Gender, age at THR, age group, diagnosis, fixation technique	Cox Regression, Standardised Mortality Ratio	60-days SMR was higher among men with osteoarthritis than women with osteoarthritis across all age groups (0.2-1.6 for men versus 0.06-1.6 for women) and lower than rheumatoid arthritis and hip fracture prognosis combined (0.2-3.2 for women versus 0.3-4.8 for men).
Lie et al. (2002)	Norway	2	Norwegian arthroplasty registry (NAR)	67548	0.4	Gender, age at THR, year of surgery, diagnosis, surgical approach	Logistic Regression	(1) Mortality rates for cases were 0.41%, 0.75%, 0.93% and 0.19%, 0.56%, 0.83% for controls, for the first 20, 60, and 90 days, respectively. (2) Male (OR=1.7-2.4) cases had higher mortality than females across all age groups. (3) 2 month mortality was higher among patients diagnosed with rheumatoid arthritis (OR=1.4-3.9) than osteoarthritis. (4) Odds of dying was higher among patients with revision surgery (OR=1.2-1.9) than those without revision.

Table 1 continued from previous page

Author (Year)	Country	Follow up time (Months)	Data Source	Sample Size	Ratio (Male:Female)	Variables	Survival Model	Outcomes Reported
Williams et al. (2002)	England	3	NHS patients from East Anglia, Oxford, Trent, Northern and Yorkshire only	7151	0.6	Cause of deaths	Cox Regression, Standardised Mortality Ratio	All-cause 3-month standardised mortality rate was estimated to be 0.4% to 0.7%.
Nunley et al. (2003)	US	24	Private Institution	835	0.6	Age group, gender, diagnosis, procedure type, anesthesia types	Logistic Regression	(1) 30-, 60-, 90-days, 1- and 2-years crude mortality rates were 0.27%, 0.27%, 0.45%, 1.66% and 4.01%, respectively. (2) 1- and 2-years crude mortality rates were lower than the expected corresponding rates in the US general population. (3) No significant difference in early post-operative mortality risk across types of anaesthesia used.
Barrett et al. (2005)	US	3	Private Institution	28469	0.6	Gender, age group	Cox Regression	3-months hazard ratio of death was significantly higher among THR cases (HR=2.61) than matched controls (HR=1.00).

Table 1 continued from previous page

Author (Year)	Country	Follow up time (Months)	Data Source	Sample Size	Ratio (Male:Female)	Variables	Survival Model	Outcomes Reported
Blom et al. (2006)	England	3	Private Institution	1727	0.6	Age group, cause of death	Crude Mortality Rate	(1) 30-days mortality rates were 0%, 0.5%, 1.4% for ages <70, 70-79 and 80+, respectively. (2) 3-month mortality rates were 0.2%, 1.3% and 2.5% for ages <70 , 70-79 and >80, respectively. (3) During the first 90 days post-surgery, most prevalent cause of death was ischemic heart diseases, followed by cerebrovascular events and pulmonary embolism, respectively.
Ramiah et al. (2007)	England & Wales	3	NHS patients	5831	0.7	Type of THR procedure (cemented, uncemented), gender, age group	Kaplan Meier, Standardised Mortality Rate	30- and 90-days crude mortality rates were 0.4% and 0.9%, respectively.
Aynardi et al. (2008)	US	3	Institutional Database	7478	0.9	Age group, cause of death	Standardised Mortality Ratio	(1) Mortality rates free of revision surgery were 0.1% and 0.4% at 30 and 90 days, respectively. (2) Cases aged less than 65 years had lower mortality rate than older age group. (3) Mortality rates for patients with revision surgery were 0.8% and 1.2% at 30 and 90 days, respectively.
Lie et al. (2010)	Norway	1	Norwegian arthroplasty registry (NAR)	106254	0.7	Gender, age group	Aalen Survival Model	(1) THR cases have an excess mortality rate of 0.13% compared to controls after 30 days. (2) Males have higher mortality rate than females and male controls after 30 days. (3) Early mortality rate increases with age and is highest (0.3%) for those aged above 70.

Table 1 continued from previous page

Author (Year)	Country	Follow up time (Months)	Data Source	Sample Size	Ratio (Male:Female)	Variables	Survival Model	Outcomes Reported
Pedersen et al. (2011)	Denmark	3	Danish Hip Arthroplasty Registry (DAR)	44558	0.8	Age group, Comorbidities (Diabetes, Cardiovascular disease, any cancer)	Logistic Regression	(1) 90-days mortality rate was lower for THR cases (5.6% in women and 9.8% in men) than controls (9.8% in women and 10.8% in men) and across all age groups. (2) 90-days mortality rate for patients diagnosed with diabetes (18.7% for cases versus 25.1% for controls) and cardiovascular diseases (20.9% for cases versus 27.0% for controls) prior to THR was lower among THR cases than controls.
Hunt et al. (2013)	England & Wales	3	National Joint Registry (NJR)	409096	0.7	Surgical approach, implant type & fixation, anaesthetic, type, age, gender, post-surgery comorbidities	Kaplan Meier, Cox Regression	(1) Kaplan Meier estimate of death (KM) 90 days post-THR surgery was higher in men (KM=0.08%-1.90%) than women (KM=0.05%-1.13%), across all age groups. (2) Hazard of death (HR) after 90 days was higher among uncemented THR (HR=1.07-1.13) and hybrid THR (HR=0.85-1.08), respectively, than among cemented procedures (HR=1.00). (3) Diagnosis of myocardial infarction, congestive heart failure, cerebrovascular disease, dementia, chronic pulmonary disease, rheumatic disease, mild liver disease, diabetes and renal disease increases post-THR hazard ratio of death to 2.74, 2.62, 1.79, 1.54, 1.27, 1.32, 1.70, 1.13, and 2.18, respectively, compared to THR patient without these conditions after THR surgery (HR=1.00).

Table 1 continued from previous page

Author (Year)	Country	Follow up time (Months)	Data Source	Sample Size	Ratio (Male:Female)	Variables	Survival Model	Outcomes Reported
NHS Scotland (2002)	Scotland	1	Private Institution	5278	0.7	Age, gender, cause of death	Cox Regression, Standardised Mortality Rate	(1) Post-THR procedure 30-day mortality rate was estimated to be 0.08%. (2) The most common cause of death during the first 30 days post-THR was myocardial infarction.
Lovald et al. (2014)	US	12	Private Institution	19706	0.5	Age group, gender, cause of deaths	Cox Regression	Hazard ratio of death (HR) for THR cases was 0.26 times lower than controls, exactly one year after THR procedures.
Jones et al. (2014)	England & Wales	3	Joint Registry	3128	0.6	Age, gender, cause of death	Crude Mortality Rate, Logistic Regression	(1) Excess mortality rates 1- and 3-months post-THR were estimated at 0.26% and 0.03%, respectively, and do not differ significantly across gender. (2) Odds (OR) of dying during the first 3 months post-THR is higher for THR cases aged over 75 years (OR=0.61%) than THR cases under 75 years (OR=0.33%). (3) Myocardial infarction was the most common cause of death during the first 90 days post-THR procedure.

Table 1 continued from previous page

Author (Year)	Country	Follow up time (Months)	Data Source	Sample Size	Ratio (Male:Female)	Variables	Survival Model	Outcomes Reported
Xu et al. (2017)	US	3	Private Insurance Database	295572	0.9	Age, gender, race, year of surgery, state, income status, medical comorbidities	Logistic Regression	(1) Estimated 30- and 90-days odds ratio of dying (OR) were 1.63 and 1.58 for insured THR cases, respectively, versus 1.09 and 1.13, respectively, among uninsured THR cases. (2) Cardiovascular complications before THR cause a lower increase of OR in the insured population (OR=1.37) than in the non-insured population (OR=1.43). (3) Pulmonary, infectious and gastrointestinal complications post-THR surgery cause a higher increase in OR in insured THR cases than in non-insured THR cases. (4) Unit increase in length of hospital in-stay post-THR surgery, causes OR to increase by 18% in the insured THR cases and 9% among the non-insured ones.
Boniello et al. (2017)	US	1	National Surgeon Database	66839	0.5	Age, gender, medical comorbidities	Logistic Regression	(1) 11% of THR cases were more than 80 years old. (2) Estimated mortality rate was higher among 80+ THR cases (0.09%) than among younger THR cases (0.01%). (2) Adjusted odds ratio of dying (OR) for 80+ age group is twice the OR of those under 80. (3) Comorbidities post-THR cause an increase in OR, with chronic kidney diseases and malnutrition causing the biggest increase in OR after THR.

2.3 Long term review of mortality after THR procedure

15 studies out of 32 selected publications, listed in Table 4, reported mortality risk after THR procedures in the long term with follow-up time greater than 24 months. The shortest and longest mean follow-up time among studies reviewed, is 3 years (McMinn et al. (2012)) and 15 years (Mäkelä et al. (2014)), respectively. Mean follow-up of 5 years (Paavolainen et al. (2002), Barrett et al. (2005), Whitehouse et al. (2014)) and 10 years (Ritter et al. (1998), Visuri et al. (1994), Ramiah et al. (2007)) are the most common. The remaining publications had a mean follow-up time of 6 years (NHS Scotland (2002), Holmberg (1992)), 7 years (Lovald et al. (2014)), 8 years (Lie et al. (2000), Maradit-Kremers et al. (2016)), 12 years (Visuri et al. (1997)) and 13 years (Pedersen et al. (2011)), respectively.

The fifteen long term studies were carried out in various locations: US (4), England and Wales (3), Finland (3), Norway (1), Denmark (1), Sweden (1), Scotland (1) and Scandinavian region (1), respectively. The sample size for these fifteen studies vary between 646 patients for the smallest study (Holmberg 1992) and 438733 patients for the largest study (Mäkelä et al. (2014)). In all 15 studies, the ratio of men to women is always less than one (see Table 4), demonstrating that proportion of females undergoing THR procedures is always higher than males.

9 studies out of 15 used a register database as data source for survival analysis. A national registry database on patients who underwent THR procedures was used for studies carried out in the Scandinavian region (Mäkelä et al. (2014)), Norway (Lie et al. (2010)), Denmark (Pedersen et al. (2011)), Finland (Paavolainen et al. (2002)), England and Wales (McMinn et al. (2012), Ramiah et al. (2007), Whitehouse et al. (2014)), Scotland via private surveys from various THR surgery centres (NHS Scotland (2002)) and US (Maradit-Kremers et al. (2016)), respectively. The source of data in all these 9 studies were therefore obtained from different surgery centres, hospitals or private clinics. None of these 9 studies accounted for the cluster effects due to grouping individuals by their surgery centre, hospital or private clinics in the database, to investigate mortality risk variations between each centre of data collection. The following six studies investigated mortality after THR procedures using data from a single surgery centre, hospital or

private clinic in their studies and are all based in the US: Ritter et al. (1998), Visuri et al. (1994), Mäkelä et al. (2014), Lovald et al. (2014), Barrett et al. (2005), Holmberg (1992). None of the 15 long term studies used primary care records as a source of data.

Five studies out of fifteen were set up as a cohort study designs to estimate long term mortality risk after THR procedure (NHS Scotland (2002), Mäkelä et al. (2014), McMinn et al. (2012), Whitehouse et al. (2014), Maradit-Kremers et al. (2016)). The remaining 10 studies were set up under a case control study design. Age and gender were the most common matching factors employed in six studies for selection of controls (Visuri et al. (1997), Paavolainen et al. (2002), Holmberg (1992), Ritter et al. (1998), Lie et al. (2000), Lovald et al. (2014)). Visuri et al. (1994) and Ramiah et al. (2007) used only gender and age group, respectively, as matching factors while Pedersen et al. (2011) used gender and year of birth as matching factors for selection of controls. The US based study by Barrett et al. (2005) used year of birth, gender and race as matching factors.

Among variables used to explain long term mortality risk after THR procedures, gender and age at time of procedure (12) were the most common, followed by osteoarthritis preoperative diagnosis (8), procedure types, preoperative rheumatoid arthritis, preoperative osteoporosis (5), angina, myocardial infarction, stroke (3) and chronic kidney disease (2), respectively. Variables related to the THR surgery directly and investigated in the studies reviewed are prognosis of THR (6), types of procedure (3), revision surgery (1), respectively. Other long term studies in this review also used ethnic backgrounds and Charlson comorbidity index (Pedersen et al. (2011), Lovald et al. (2014)), body mass index and preoperative cardiovascular disease and Type 2 diabetes (Pedersen et al. (2011)) and social deprivation index (Whitehouse et al. (2014)) to investigate long term mortality risk following THR procedures. None of the studies reviewed reported the effects of preoperative smoking status, hypercholesterolemia and hypertension on long term mortality risk after THR.

Among the statistical methods of survival analysis employed, the most popular was Cox regression model (8) followed by standardised mortality rate and Kaplan Meier survival analysis (5), respectively. Only Pedersen et al. (2011) used crude mortality rate and McMinn et al. (2012) used a Royston-

Parmar survival model to estimate long term mortality risk following THR, respectively. Only one study (Mäkelä et al. 2014) reported that 0.5% of their data were missing because the records of these patients were incomplete and were therefore, excluded from the study.

Table 2 below summarises the reported long term mortality risks after THR that have been aggregated from several available published articles at each length of follow-up time. Between 3 to 15 years following THR procedure, mortality risks of THR cases are lower than matched controls or mortality of the general population who did not undergo THR in the reviewed articles. Across genders, reported long term mortality of male THR cases is higher than female counterparts 5, 10 and 12.7 years, respectively, post-surgery while no difference in mortality risk between males and females after THR procedures were found 3 and 12 years after the procedure (see Table 2). Only Mäkelä et al. (2014) reported a higher estimated mortality risk among women aged above 60 than men of the same age group, 8 years after the procedure.

Table 2: Summary of reported long term outcomes by follow-up time after THR procedure

Years after THR	Variations in mortality risk	References
3	Hazard ratio of death increases by 9% per unit increase in age at surgery and does not differ between men and women.	McMinn et al. (2012)
5	Standardised mortality rate (SMR) for THR patients is between 0.81%-0.87%. Estimated hazard ratio of death is 1.53 for male THR patients, compared to females (HR=1.00).	Paavolainen et al. (2002), Whitehouse et al.(2014), Barrett et al. (2005)
7	Estimated mortality rate among THR cases is 0.29% compared to 0.56% among controls.	Lovald et al. (2014)
8	(1) Estimated SMR is 0.80% and 0.83% among women and men, respectively, after THR. (2) Estimated SMR are 2.5% for patients below 49, 1.16% for ages 50-59, 0.86% for ages 60-69, 0.79% for ages 70-79, 0.76% for ages 80-89 and 0.87% for ages above 90, respectively.	Lie et al. (2000), Maradit-Kremerset et al. (2016)

Table 2 continued from previous page

Years after THR	Variations in mortality risk	References
10	(1) Estimated mortality rate among ages above 65, is 21% in male and 17% in female THR cases, and 26% in male and 12% in female controls, respectively. (2) Hazard ratio of death increases by 3% for every unit increase in age at primary procedure and does not differ significantly between men and women.	Visuri et al. (1994), Ritter et al. (1998), Ramiah et al. (2007)
12	Estimated relative mortality rate for THR cases is 10% higher than controls by the end of the 12^{th} post-operative year.	Visuri et al. (1997)
12.7	(1) Estimated mortality rate is 8.5% and 10.1% among female and male THR cases, respectively (versus 11.7% and 13.5% in female and male controls, respectively). (2) Mortality rates for ages 18-59, 60-69, 70-79 and 80+ are 2.3%, 5.0%, 11.8% and 22.8%, respectively (versus 2.4%, 6.8%, 16.1% and 35.9% for controls of same age group, respectively).	Pedersen et al. (2011)

Table 2 continued from previous page

Years after THR	Variations in mortality risk	References
15	Estimated post-THR mortality rates are 19.37% for males (versus 19.64% for females) aged below 60 and 15.00% for men (versus 10.77% for women) aged above 60.	Mäkelä et al. (2014)

All 12 studies suggested that mortality varies significantly for different age groups at surgery. 3 and 5 years after the procedure, hazard ratio of death increases by 9% (McMinn et al. (2012)) and 8% (Whitehouse et al. (2014)), respectively, per yearly increase in age at surgery for both gender. Lie et al. (2000) and Ramiah et al. (2007) concluded that 8 and 10 year mortality rate in their study also rises with increasing age at time of THR, respectively. For studies carried out in the Scandinavian region, 12.7 and 15 year mortality also increases with increasing age and for male THR cases (Pedersen et al. (2011), Mäkelä et al. (2014)). For England and Wales combined, the 14^{th} National Joint Registry annual review on joint replacement procedures (NJR (2017)) estimated long term mortality rate at 3, 5, 7, 10, 11 and 13 years after THR (see Table 3). This annual report also suggests that long term mortality rate following THR procedures, is higher among men than women for all age groups and increases for older ages at surgery time.

Age group		**Years after THR**					
		3	**5**	**7**	**10**	**11**	**13**
Males	**<55**	1.36%	2.21%	3.30%	4.93%	5.63%	6.88%
	55-59	1.86%	3.28%	5.05%	8.36%	9.96%	13.30%
	60-64	2.64%	4.81%	7.30%	12.56%	14.45%	18.60%
	65-69	3.61%	6.92%	11.07%	18.78%	21.67%	29.51%
	70-74	5.60%	10.67%	16.92%	29.39%	34.23%	44.83%
	75-79	8.63%	16.96%	27.75%	46.23%	52.96%	66.38%
	80-84	13.57%	27.08%	42.76%	66.42%	72.78%	83.81%
	85+	23.82%	44.00%	63.34%	85.70%	90.30%	95.72%
Females	**<55**	1.63%	2.49%	3.45%	4.93%	5.39%	6.37%
	55-59	1.71%	3.03%	4.47%	6.96%	7.90%	9.71%
	60-64	2.02%	3.76%	5.68%	9.44%	11.06%	14.87%
	65-69	2.55%	4.82%	7.70%	13.66%	15.98%	21.48%
	70-74	3.53%	7.19%	11.79%	21.63%	25.58%	34.74%
	75-79	5.61%	11.68%	19.34%	34.85%	40.73%	52.70%
	80-84	9.07%	18.69%	31.48%	53.57%	61.47%	74.63%
	85+	16.24%	32.12%	50.33%	74.08%	80.36%	90.19%

Table 3: Long term cumulative mortality rate reported by the 14^{th} National Joint Registry annual report (NJR (2017))

Three studies also reported variations in long term mortality risk by causes of death (Visuri et al. (1997), Paavolainen et al. (2002, Lovald et al. (2014)). For a mean follow-up time of 5 years, Paavolainen et al. (2002) reported that myocardial infarction (SMR=0.85), followed by hypertensive disease (SMR=0.85), cerebrovascular disease (SMR=0.86) and cancer (SMR=0.70), respectively, were the most common cause of death following THR. 7 years after THR procedure, Lovald et al. (2014) concluded that hazard of death

(HR) was only higher among THR cases who had an event of cardiovascular disease after surgery (HR=1.15), compared to those without these cardiovascular events after surgery and among THR cases diagnosed with diabetes after surgery (HR=1.02), relative to THR cases without diabetes, respectively. Similarly, Lovald et al. (2014) also reported cardiovascular diseases as the most common long term cause of death following THR. With a mean follow-up time of 12 years, Visuri et al. (1997) concluded that circulatory disease (relative mortality risk=1.50) was the most common cause of death, followed by cancer (relative mortality risk= 0.74) after THR procedure.

Four studies investigated difference in long term mortality risk for THR cases diagnosed with osteoarthritis (OA) or rheumatoid arthritis (RA). Lie et al. (2000) estimated that the relative mortality risk for THR cases diagnosed with RA is 2.57 times higher than OA while Whitehouse et al. (2014) concluded that RA increased hazard ratio of death to 1.66, compared to THR cases diagnosed with OA (HR=1.00). Maradit-Kremers et al. (2016) compared THR patients diagnosed with OA to those without OA and found that the standardised mortality rate among OA THR cases was 0.82%. A similar trend is reported by Holmberg (1992). All these studies provide evidence that estimated mortality risk is higher for THR cases diagnosed with RA, compared to OA THR cases.

Lie et al. (2000), McMinn et al. (2012) and Whitehouse et al. (2014) also reported the effect of different fixation techniques (cemented, uncemented and hybrid fixations) employed during the procedure on long term mortality risk. Lie et al. (2000) estimated lower relative mortality risk (RMR) for uncemented (RMR=0.78) and hybrid (RMR=0.94) fixation techniques, compared to cemented procedures while McMinn et al. (2012) and Whitehouse et al. (2014) demonstrated that the long term hazard ratio of death falls to 0.84-0.90 for uncemented procedures and 0.87 for hybrid procedures, compared to cemented fixation technique. Only McMinn et al. (2012) analysed the effect of revision surgery on long term mortality following THR procedure. The authors found that revision surgery after uncemented procedures increases hazard ratio of death to 1.60, compared to cemented procedures. In addition, Whitehouse et al. (2014) reported significant association between variations in long term mortality and types of surgical approach during THR interventions. This study concluded that, relative to lateral surgical approach, hazard ratio of death decreases to 0.93 and 0.82 for posterior and anterior

surgical approach, respectively.

Pedersen et al. (2011) compared mortality rates after mean follow-up of 12.7 years for patients with low, moderate and high Charlson's comorbidity index (C-index), prior to THR. The authors found that estimated mortality rates were 6.9%, 13.7% and 23.1% for THR cases with low, moderate and high C-index, respectively (versus 8.2%, 20.2% and 40.6% in the control population for low, moderate and high C-index, respectively). Hence mortality risk increases for THR cases with high number of preoperative comorbidities and is lower than control population. Furthermore, Pedersen et al. (2011) also reported that estimated mortality rates (MR) for THR cases with preoperative cardiovascular diseases (MR=17.8%), diabetes (MR=17.0%) and cancer (MR=16.1%), respectively, were lower among THR cases than controls (MR=28.5%, 27.0% and 24.7% for controls with a history of cardiovascular diseases, diabetes and cancer, respectively). This may be due to the fact that relatively healthy patients are allowed to undergo THR procedure and thus prevalence of preoperative comorbidities is lower among THR patients, compared to controls, as highlighted by Barrett et al. (2005), which reported that prevalence of comorbidities among THR cases, was on average 30% less than in controls, in their study.

McMinn et al. (2012) also compared the long term mortality risk for patients with different physical status prior to THR procedures. In this study, patients physical status was grouped into the following categories: category A for *healthy person*, category B for *mild systemic disease*, category C for *severe systemic disease*, category D for *severe systemic disease that is a constant threat to life* and category E for *moribund person who is not expected to survive without the operation*, respectively. The authors concluded that relative to category A, hazard ratio of death 3 years after THR increases to 1.17, 2.14 and 3.58 for categories B, C and D, respectively while hazard of death for category E does not differ significantly to category A. Additionally, Whitehouse et al. (2014) reported that long term hazard ratio of death varies significantly for different providers of the intervention. The authors concluded that hazard ratio of death decreases to 0.92, 0.74 and 0.59 in NHS private treatment centres, private hospitals and independent clinics, respectively, compared to public hospitals in the UK. Hence private surgery centres are significantly associated with a lower long term mortality risk after THR procedure, partly due to the fact that majority of surgeons in private surgery

centres have high level of experience in THR procedures and also because the waiting time to have the procedure is longer in public hospitals, than in private clinics or hospitals.

Table 4: Review of published studies on long term mortality risk after total hip replacement

Author (Year)	**Follow up time (Years)**	**Data Source**	**Sample Size**	**Ratio (Male:Female)**	**Country**	**Variables**	**Survival Model**	**Outcomes Found**
Ritter et al. (1998)	10	Private Institution	3807	0.7	US	Gender, Age at THR, Age group, unilateral or bilateral	Cox Regression	(1) Survival differs for age groups (61-70, 71-80 and >80). (2) Survival decreases with age. (3) Younger age cases (<60) survival did not differ from controls. (4) Females survival was better than males for age groups 51-60, 61-70 and 71-80, respectively.
Lie et al. (2000)	8	Norwegian Arthroplasty Registry (NAR)	39543	0.4	Norway	Gender, Age at THR, Age group, diagnosis, fixation technique	Cox Regression, Standardised Mortality Rate	(1) Standardised mortality rate (SMR) was 0.80 for women and 0.83 for men, respectively. (2) SMR was highest among younger age group (2.5 for <50 years) and decreases with age at surgery (0.9 for 90+ years). (3) SMR was 0.75, 0.92 and 0.84 for cemented, uncemented and hybrid fixation procedures, respectively.
McMinn et al. (2012)	3	National Joint Registry (NJR)	274473	0.6	England & Wales	Gender, fixation technique	Cox Regression	(1) Hazard ratio of death (HR) is higher among cemented (HR=1.1) than uncemented and increases at a rate of 9% per unit increase in age at surgery for patients free of revision surgery. (2) HR of men free of revision surgery (1.5-1.6) was higher than women free of revision surgery. (2) HR of cemented procedures among patients with revision surgery (HR=0.5-0.6) was lower than uncemented procedures of the same category.

Table 4 continued from previous page

Author (Year)	Follow up time (Years)	Data Source	Sample Size	Ratio (Male:Female)	Country	Variables	Survival Model	Outcomes Found
Visuri et al. (1994)	10	Private Institution	1018	0.4	Finland	Gender, THR prostheses type	Kaplan Meier	(1) 10-years survivorship reported were 84% among controls and 82%-85% among THR cases. (2) Among patients aged 65 or more, survivorship of THR cases (69% for men and 83% for women) was lower than controls (74% for men and 88% for women) and higher than the general Finnish population (62% for men and 80% for women).
NHS Scotland (2002)	6	National Health Services (NHS)	24983	0.6	Scotland	Age group, diagnosis, Consultant experience	Kaplan Meier	(1) Mortality rate after THR increases with age group. (2) Osteoarthritis THR patients mortality rate is better than rheumatoid arthritis. (3) Mortality rate was worst for consultants with less than 10 THR operations per year than those who have more than 10 THR operations per year.
Pedersen et al. (2011)	12.7	Danish Hip Arthroplasty Registry (DAR)	44558	0.8	Denmark	Age group, Comorbidities (Diabetes, Cardiovascular disease, any cancer), diagnosis	Crude Mortality Rate	(1) Mortality rate was lower for THR cases (8.5% in women and 10.1% in men) than controls (11.7% in women and 13.5% in men) and across all age groups. (2) Mortality rate for patients diagnosed with diabetes (17.0% for cases versus 27.0% for controls) and cardiovascular diseases (17.8% for cases versus 28.5% for controls) prior to THR was lower among THR cases than controls.

Table 4 continued from previous page

Author (Year)	Follow up time (Years)	Data Source	Sample Size	Ratio (Male:Female)	Country	Variables	Survival Model	Outcomes Found
Mäkelä et al. (2014)	15	Joint Registry	438733	0.6	Sweden, Denmark, Norway, Finland	Country, Gender, Age group, Diagnosis, Fixation technique	Kaplan Meier/ Cox Regression	(1) Survival after 15 years is 86.3%, 88.0%, 86.9% and 83.5 % in Denmark, Sweden, Norway and Finland, respectively. (2) 15-year survival was highest among cemented procedures (87.9%) than among uncemented (82.5%) and hybrid procedures (82.6%), across all countries, respectively. (3) 15-year survival for cemented procedures (79%) is lower for males aged less than 60 in Denmark, Sweden and Finland only, compared to uncemented procedures (82.4%) of the same age category.
Visuri et al. (1997)	12	Private Institution	1018	0.4	Finland	Gender, Cause of death	Standardised Mortality Rate	(1) Relative mortality risk of THR cases was 94% higher during the first 4 years post-surgery, than matched controls and decreases as follow-up period increases, becoming 4% lower than that of matched controls after 20 years of follow-up period. (2) Relative mortality risk of THR cases with cardiovascular diseases (CVD) was higher than those without CVD only for the first four years post-surgery.

Table 4 continued from previous page

Author (Year)	Follow up time (Years)	Data Source	Sample Size	Ratio (Male:Female)	Country	Variables	Survival Model	Outcomes Found
Paavolainen et al. (2002)	5	Finnish National Arthroplasty registry (FNAR)	24638	0.6	Finland	Cause of deaths, gender	Standardised Mortality Rate	(1) Standardised mortality rate (SMR) was 0.69%, with no difference between men and women THR patients and was lower than the Finnish population. (2) Estimated SMR for THR cases with neoplasm, diabetes, dementia, Alzheimer's disease, cardiovascular disease and accidents/violence as cause of death were 0.54%,0.36%, 0.51%, 0.70% and 0.74%, respectively.
Lovald et al. (2014)	7	Private Institution	19706	0.5	US	Events of heart failure, ischaemic heart disease, cardiovascular disease, diabetes, depression	Cox Regression	(1) Long term hazard ratio of death (HR) for THR cases was 0.52 times lower than controls.
Barrett et al. (2005)	5	Private Institution	28469	0.5	US	Gender, Age group, Comorbidities	Cox Regression	(1) Prevalence of serious comorbidities was on average, 30% less among THR cases than controls, prior to THR surgery. (2) From 3-months to 5 years post-surgery time, hazard of death (HR) of THR cases was 67% lower than matched controls. (4) HR 5 years post-THR surgery converge to that of matched controls.

Table 4 continued from previous page

Author (Year)	Follow up time (Years)	Data Source	Sample Size	Ratio (Male:Female)	Country	Variables	Survival Model	Outcomes Found
Ramiah et al. (2007)	10	NHS Patients	5831	0.7	England & Wales	Type of THR procedure (cemented, uncemented), gender, age group	Kaplan Meier, Standardised Mortality Rate	(1) 10-year survivorship after THR was higher among women than men and decreases with age at THR surgery. (2) Standardised mortality rate (SMR) was higher than unity (SMR=100%) among patients aged 64 or less, with a 15-fold increase in SMR of females and males under 45 years, respectively. (3) Having a THR above the age of 75 years is associated with better survival than other younger ages at surgery time.
Holmberg et al. (1992)	6	Private Institution	646	0.6	Sweden	Gender, Diagnosis, age group	Cox Regression	(1) Survival was highest among patients with osteoarthritis, followed by rheumatoid arthritis, respectively. (2) Mortality was lower among THR cases than that of the Swedish population.

Table 4 continued from previous page

Author (Year)	Follow up time (Years)	Data Source	Sample Size	Ratio (Male:Female)	Country	Variables	Survival Model	Outcomes Found
Whitehouse et al. (2014)	5	National Joint Registry (NJR)	382140	0.5	England & Wales	Age, gender, BMI, diagnosis, procedure types, surgical approach, surgeon experience, index of multiple deprivation	Cox Regression	(1) Hazard ratio of death (HR) is higher in men (HR=1.53) than in women after THR (HR=1.00). (2) Unit increase in age at surgery raises HR by 8%. (3) Compared to diagnosis of osteoarthritis (reference), rheumatoid arthritis (HR=1.51) and hip fracture (HR=1.51-2.65) cause an increase in HR. (3) HR in private independent hospitals/clinics (HR=0.59-0.74) was lower than NHS/public hospitals (reference). (4) Uncemented (HR=0.74) and hybrid (HR=0.87) procedures have a lower hazard ratio of death than cemented procedures. (5) Unit increase in index of multiple deprivation of patients' residential ward increases post-THR HR by 1%.
Maradit-Kremers et al. (2016)	12	US County Database	1645	0.7	Olmsted & County	Age, gender	Standardised Mortality Rate (SMR)	(1) SMR was estimated at 0.82% among THR patients and is lower than the US general population. (2) Men (SMR=0.81) and women (SMR=0.82%) long term mortality risk are close to each other.

2.4 Review of randomised controlled trial evidence of effects of THR procedures on mortality

In this section, published studies reporting evidences from randomised controlled trials (RCTs) evidences of the effect of THR procedures on mortality are reviewed. A comprehensive search from multiple databases (see section 2.1) was performed to identify 25 publications that investigated the clinical effects of THR procedures under a RCT study design. Reports on mortality after THR procedures in these studies were produced as the number of deaths during the follow-up time. This is mainly because these studies were interested in the survival of the prosthesis with revision surgery as end-point.

The 25 studies reviewed did not include any survival or mortality risk model for estimation of the effect of THR procedures on mortality risk. However the authors produced adequate information with respect to the number of deaths for mortality investigation. Table 5 below summarises the details of the included RCTs in this section. It can be observed that the number of deaths among patients undergoing metal-on-metal (MOM) THR procedures is lower than that of non-metal-on-metal (non-MOM) THR procedures in all studies. However these details do not provide conclusive evidence for the effect of THR procedures on mortality risk under RCT set up, as opposed to the observational studies reviewed in sections 2.2 and 2.3, respectively.

Findings from short term RCTs (follow-up$\leq 2 years$ - Pabinger (2003), Zagra (2013), Grubl (2006), Schouten (2012), Jensen (2011), Zijlstra (2011), Hanna (2012), Wessinger (2011), Tiusanen (2013), Penny (2012) and Gauthier (2013)) showed that the number of Non-MOM procedures was more prevalent than MOM procedures and that there was no death observed during the first year after the surgery for both procedures. However number of deaths during the second year of rehabilitation after surgery was one for MOM compared to three for non-MOM procedures. Similarly among long term RCTs reviewed (Brodner (2003), Macdonald (2005), Malviya (2011), Gustafson (2014), Wang (2012), Engh (2014), Bjorgul (2013), Zerahn (2011), Hailer (2011), Howie (2005), Desmarchelier (2013) and Ziljstra (2010)), there were 715 MOM procedures versus 997 non-MOM procedures, respectively while the number of deaths was higher among non-MOM (76) in comparison to 67 deaths among MOM procedures.

Table 5: Reported number of deaths in RCTs included in review of published RCT after THR procedures

Author	Publication	Metal-on-Metal (Mom)			Non-metal-on-metal (Non-Mom)					Follow-up time
	Year	N patients	N deaths	N revision	N patients	N deaths	N revision	Type of non-MOM	THR Types	Year
Zagra	2013	20	0	0	40	0	0	Polyethelene-Ceramic	Uncemented	0.33
Grubl	2006	15	0	0	13	0	0	Ceramic-Ceramic	Uncemented	1
Schouten	2012	39	0	0	42	0	0	Ceramic-metal	Cemented	1
Jensen	2011	21	0	0	22	0	0	Polyethelene-Ceramic	Uncemented	1
Zijlstra	2011	25	0	0	25	0	0	Polyethelene-metal	Uncemented	1
Hanna	2012	28	0	0	23	0	0	Polyethelene-metal	Cemented	2
Weissinger	2011	42	0	0	38	0	0	Ceramic-Ceramic	Uncemented	2
Tiusanen	2013	46	0	2	46	0	1	Polyethelene-metal	Cemented	2
Penny	2012	21	0	0	22	0	0	Polyethelene-metal	Cemented	2
Gauthier	2013	25	0	1	25	2	0	Ceramic-metal	Uncemented	2
Malviya	2011	50	1	2	50	2	2	Polyethelene-metal	Uncemented	4
Brodner	2003	50	1	1	50	2	1	Polyethelene-Ceramic	Cemented	5
Gustafson	2014	26	0	1	26	0	0	Polyethelene-metal	Cemented	5
Macdonald	2005	23	1	0	18	0	0	Polyethelene-metal	Uncemented	5
Wang	2012	37	0	0	40	0	0	Ceramic-Ceramic	Cemented	6
Engh	2014	68	2	1	37	2	1	Polyethelene-metal	Cemented	6.5
Bjorgul	2013	123	6	8	251	16	4	Mixed	Uncemented	7
Zerahn	2011	74	6	0	225	14	0	Mixed	Cemented	7.6
Hailer	2011	41	1	0	44	4	0	Polyethelene-metal	Uncemented	8
Howie	2005	11	0	8	13	0	2	Polyethelene-metal	Uncemented	10
Desmarchelier	2013	111	19	3	116	13	1	Ceramic-Ceramic	Cemented	10
Zijlstra	2010	101	30	4	97	23	2	Polyethelene-metal	Cemented	10.7
Pabinger	2003	31	1	1	28	1	0	Polyethelene-Ceramic	Cemented	2

The above findings showed higher number of non-MOM procedures than MOM procedures and higher number of deaths after non-MOM THR, respectively.

In addition, Pijls et al. [2016] carried out a systematic review of 25 RCT studies with mortality investigation as one of the outcomes and that included the first 12 publications listed in Table 5. The authors estimated through meta-analysis, that the overall mortality risk was higher for non-MOM THR procedures related to MOM ones in the short term, with follow-up less than 2 years. However for longer follow-up investigations, the same authors found no evidence of difference in overall mortality risk and that the death was independent of modifying factors such as procedure types.

Similarly, Abdulkarim et al. [2013] critically reviewed nine RCTs publications (not listed in Table 5) through meta-analysis of data on 778 patients, with mean age of 60.5 years and average follow-up period of 4.3 years (range: 2-8 years), comparing cemented to uncemented THR procedures. The authors concluded that uncemented THR provides an immediate postoperative benefit not only in terms of a reduction in mortality risk, but also improved other outcomes such as relief of joint pain after the procedure and risk of revision surgery. These conclusions are also in agreement with Morshed et al. [2007], a meta-analysis observational study comparing cemented procedures to uncemented ones with respect to mortality and revision surgery.

Although RCT studies benefits from strong experimental design to provide high level evidence, very few RCTs that estimated the effects of THR procedures on mortality were available in literature. Almost all published RCTs were primarily designed to investigate rate of revision surgery and patients' pain score after the procedure. Therefore statistics on mortality risk after THR were rarely reported in these papers with the few evidences provided by Abdulkarim et al. [2013] and Pijls et al. [2016], who respectively, reported an increase in mortality risk associated with MOM procedures in the short term and cemented THR procedures for long term follow-up studies, respectively.

2.5 Discussion of findings from review of short and long term mortality after THR procedure

The definition of follow-up time varies across different short and long term studies. In the short term studies, mortality risk was estimated at 1, 3, 6, 12 and 24 months following THR procedure while long term studies had a

mean follow-up of at least 3 years after the procedure. Hence in this research, mortality risk estimation is carried out for the first 24 months (short term mortality model) to investigate early post-operative excess mortality after THR and then in the long term for patients who survived more than 24 months after their THR procedure.

25 studies out of 32 reviewed in this chapter used data from multiple hospitals, clinics and private surgery centres or from secondary registry database. None of these studies accounted for the grouping effects that these hierarchical data could have on mortality risk estimation. The authors assumed that there was no significant differences in mortality risk between different hospitals. In case that this assumption is violated, failing to adjust for grouping effect can lead to false precision (Therneau and Grambsch 2013). In this research, for both short and long term mortality models, a random effect due to groupings of patients by different general practices, is added to investigate variations in mortality risk across different general practices.

None of the 32 studies reviewed, used primary care database as a source of information to investigate mortality risk after THR. Their source of data were either from secondary care database, single hospital/clinic, regional surveys or county based insurance database for US based studies. This research aims at developing mortality risk models using primary care records. It can lead to different results on variations of mortality risk after THR procedures. Volume of socio-demographic and medical information in primary care data is greater than registry or single centre database, and can thus be extensively used to explain long term variations in mortality risk following THR procedures.

Four short term studies (Aynardi et al. (2009), Singh et al. (2011), Bozic et al. (2012) and Gaston et al. (2007)) and two long term studies (Pedersen et al. (2011), Lovald et al. (2014)) assessed the physical and health status of patients prior to their THR procedures using Charlson's comorbidity (C-) index (Charlson et al. (1987)) or the American Society of Anaesthesiologists (ASA) classification (Daabiss 2011) of the physical status of patients prior to their THR procedure. These two classifications are not validated for orthopaedic patients such as THR cases. The ASA has been criticised for being inflexible in its definition of comorbidities used to classify patients' preoperative physical fitness for anaesthesia (Whitehouse et al. 2014). Additionally,

it is established that different anaesthetists assign different ASA grades to the same patient (Little, (1995), Mak et al. (2002)). Therefore the ASA cannot be used for a full assessment of a patient's comorbid status. The only validated gradings of the comorbid status for orthopaedic patients, is the POSSUM classification (Mohamed et al. (2002)). However, there is no evidence that it can be validated for surgical and orthopaedic interventions such as THR procedure. In this research, the extensive availability of primary care data provides important information on the preoperative comorbidities among THR patients and thus, no physical status classification is used.

For short term mortality models reviewed, only the following preoperative conditions were included in the data analysis among the 17 studies reviewed: body mass index (BMI), diabetes, chronic kidney diseases, cardiovascular diseases (angina, myocardial infarction, stroke), diagnosis for THR (osteoarthritis, rheumatoid arthritis, other arthritic diseases, osteoporosis) and hypertension. Similarly, the following preoperative comorbidities were assessed for long term mortality models reviewed: chronic kidney diseases, cardiovascular diseases (angina, myocardial infarction, stroke) and diagnosis for THR (osteoarthritis, rheumatoid arthritis, other arthritic diseases, osteoporosis). No studies investigated mortality risk after THR for the following preoperative conditions: BMI, diabetes and hypertension (long term only) and hypercholesterolemia (short and long terms), respectively. In this research, all reported risk factors for mortality risks following THR in this review are investigated together with BMI, diabetes, hypertension and hypercholesterolemia using primary care data, to investigate the effects of these preoperative comorbidities on mortality risk variations.

No studies distinguished between smokers and non-smokers in their data analysis. It is well established that mortality risk for smokers is significantly higher than non-smokers (Doll et al. (2004)). Hence failing to adjust for smoking status in the mortality model may lead to less precise estimates of mortality risk. All studies but one used secondary care database or single hospital/clinic data for their research and may therefore do not have access to lifestyle variables such as smoking status. In this research, it is important to adjust for smoking status in the survival model since smoking increases mortality risk. The Health Improvement Network (THIN) primary care database is used as source of data and consists of lifestyle information such as smoking that can be used for estimation of mortality risks after THR procedure.

2.6 Chapter summary

In this chapter, a review of short and long term mortality risk after total hip replacement procedures among adults for any reasons other than hip fracture and accident, is presented. For both short and long term, age at surgery, gender, types of THR procedures and fixations, and index of multiple deprivation were significantly associated with variations in mortality risk. Among preoperative risk factors for post-THR mortality risks discussed, cardiovascular diseases, diabetes, osteoarthritis, rheumatoid arthritis and osteoporosis were the most common. Overall, in the first 2 years after the procedure, THR cases have higher mortality risk than controls, with the excess mortality risk peaking at 1 and 3 months after the procedure and then decreasing until it converges to that of controls at about 24 months after the intervention. On the contrary, the literature reviewed in this chapter suggests that patients who survived the first 2 years after THR, have a lower mortality risk than control population that did not undergo THR. Therefore, it can be deduced that THR procedure improve survival in the long term only.

3 Review Of Statistical Methods For Survival Analysis

Introduction

Survival analysis is a set of statistical methods used to explain variations in data that measure the time taken for some event to occur. It is an essential analytical tool in many fields like medicine, biology, epidemiology, demography and engineering. It is used to model the time from the beginning of a follow-up of a subject until a pre-defined event occurs. Often the event is associated with failure. For example, in epidemiological studies, the event corresponds to death or the occurrence of a complication or disease among individuals who have undergone a particular treatment. In the field of economics, event may be an acceptance of a job offer for unemployed individuals while in the engineering sector, failure may be of interest for a component in an engine to stop working. The time for the event to occur is referred as the *survival* or *failure time*. This chapter details the different survival modelling strategies that exist in the analysis of epidemiological survival data and the techniques to fit these survival models. It covers a number of parametric and non-parametric survival models that are available in literature for the analysis of survival data.

3.1 Basic Ideas

According to Cox and Oakes (1984), to apply survival techniques to different situations, three components are needed: (1) determining a time origin that is well-defined at the start of the investigation period; (2) a scale to measure the change in time and (3) an exact definition of failure. Consider a set of homogeneous data where T is a positive random variable representing failure time, with distribution function, F(t) and unique probability density function, $f(t)$. For continuous time T, the *survival function* S(t) is defined as the probability that an individual survives beyond time t:

$$S(t) = P(T > t) = \int_t^{\infty} f(t)dt \ ; \ 0 < t < \infty \tag{3.1}$$

where $0<S(t)\leq 1$.

The hazard function or the force of mortality, $\lambda(t)$, at time t, is a fundamental quantity in survival analysis defined as

$$\lambda(t) = \frac{f(t)}{1 - F(t)} \tag{3.2}$$

$\lambda(t)$ can be interpreted as probability that failure occurs in the interval (t,t+δt) given that the individual survives past time t, i.e,

$$\lambda(t)\delta t \cong \ P(t < T < t + \delta t | T > t) \tag{3.3}$$

In terms of the survival function $S(t)$, the hazard function can be written as the probability density function divided by the probability that an event has not occurred before time t, as shown below.

$$\lambda(t) = \frac{f(t)}{S(t)} = \frac{\frac{-dS(t)}{dt}}{S(t)} = \frac{-d\log(S(t))}{d(t)} \tag{3.4}$$

Integrating λ(t) with respect to time t, the cumulative hazard function, $\Lambda(t)$, is obtained as

$$\Lambda(t) = \int_0^t \lambda(u)du = \int_0^t \frac{f(u)}{1 - F(u)}du = -\log(1 - F(t)) = -\log S(t) \tag{3.5}$$

Thus the survival function, $S(t)$, expressed in terms of the hazard function, is given by

$$S(t) = \exp(-\int_0^t \lambda(u)du) \tag{3.6}$$

3.2 Censoring

Censoring is a common characteristic of survival data where an observation does not contain complete survival information. In mortality data, some individuals cannot be followed-up completely during the investigation for practical and personal reasons. Therefore the time at which the event of interest occurs, is unknown to the researcher. *Right censoring* happens when a subject joins an experiment at the start of the study and leaves the investigation without experiencing failure (death). Right censoring is the most common form of censoring and usually occurs because either the study finishes before the subject experiences failure or because the individual is lost

to follow up or leaves the study before it actually ends. If the study started at time 0 and ended at time t, then the subject is said to be right-censored at time t and is assumed to have survived for a period of [0,t].

Left censoring is present when the subject experiences failure at a time before the investigation period occurs. It is a rare occurrence in real life data set. Suppose a subject who enrolled in an experiment at the start, experiences failure at a time that is unknown but before the starting time of the investigation period. The subject is said to be left censored and the subject's failure time is unknown. *Interval censoring* occurs when the event of interest happened within a time interval with the exact time of occurrence being unknown.

3.3 Parametric Modelling

Introduction

There exists a number of situations where survival data has a known distribution, or sometimes it is reasonable to assume that the data has a certain parametric specification. A selection of the distributions that are commonly used to fit such data, is detailed in this section. Fitting parametric models to survival data offers some advantages in comparison to non-parametric modelling. They have fully specified hazard functions that are dependent on the parameters that determine the overall distributional form. These parameters can be estimated to fit the model at any point in time and can be used to predict the hazard at future time points.

3.3.1 Selected survival distributions

Exponential distribution

If the hazard rate (equation (3.6)) is equal to a constant positive scale parameter, λ, the survival time is said to follow an exponential distribution, with parameter λ. The density function is given by

$$f(t) = \lambda \exp(-\lambda t), \text{ for } \lambda > 0 \tag{3.7}$$

and therefore, the survivor function is given by

$$S(t) = \exp(-\lambda t). \tag{3.8}$$

The Weibull distribution

The Weibull distribution is a generalisation of the exponential distribution where the hazard function takes the form of $h(t) = \alpha\lambda t^{\alpha^{-1}}$ $(\lambda, \alpha > 0)$. λ and α are referred as the scale and shape parameter, respectively. Under this distribution, survival time is assumed to follow a Weibull distribution with parameter λ and α, respectively, where the density function is given by

$$f(t) = \alpha\lambda t^{\alpha^{-1}} \exp(-\lambda t^{\alpha}), \text{ for } \lambda, \alpha > 0, \tag{3.9}$$

and the survivor function is written as

$$S(t) = \exp(-\lambda t^{\alpha}), \text{ for } \lambda, \alpha > 0. \tag{3.10}$$

The Log-Normal distribution

If survival time, T is assumed to be log-normally distributed with mean μ and variance σ^2, i.e. $\ln(T) \sim N(\mu, \sigma^2)$, then the survivor function is given by

$$\begin{aligned} S(t) &= 1 - P(T < t) = 1 - P(\ln(T) < \ln(t)) \\ &= 1 - P\left(\frac{\ln(T) - \mu}{\sigma} < \frac{\ln(t) - \mu}{\sigma}\right) \\ &= 1 - \Phi\left(\frac{\ln(t) - \mu}{\sigma}\right), \end{aligned} \tag{3.11}$$

where Φ is the cumulative probability distribution function of a standard normal distribution.

The generalised Gamma distribution

The generalised gamma distribution is given by

$$f(t) = \frac{\left(\frac{t-\mu}{\beta}\right)^{\gamma-1} \exp\left(-\frac{t-\mu}{\beta}\right)}{\beta\Gamma(\gamma)} \tag{3.12}$$

where γ is the shape parameter, μ is the location parameter, β is the scale parameter and Γ is the gamma function which has the formula

$$\Gamma(a) = \int_0^{\infty} s^{a-1} \exp(-s) ds, \text{ for } a > 0$$

The case where $\mu = 0$ and $\beta = 1$ is called the standard gamma distribution and the distribution is reduced to

$$f(t) = \frac{t^{\gamma-1}\exp(-t)}{\Gamma(\gamma)}, \text{ for } t \geq 0 \text{ and } \gamma > 0.$$

Since the general form of the probability functions for the gamma distribution can be expressed in terms of the standard distribution, all subsequent formulas in this section are given for the standard form of the function. The hazard function of the gamma distribution is

$$h(t) = \frac{t^{\gamma-1}\exp(-t)}{\Gamma(\gamma) - \Gamma_t(\gamma)}, \tag{3.13}$$

where $t \geq 0$, $\gamma > 0$ and $\Gamma_t(\gamma)$ is the gamma function evaluated at time t. The cumulative hazard function is given by

$$H(t) = -\log\left(1 - \frac{\Gamma_t(\gamma)}{\Gamma(\gamma)}\right), \text{ for } t \geq 0 \text{ and } \gamma > 0, \tag{3.14}$$

and therefore, the survival function is given by

$$S(t) = 1 - \frac{\Gamma_t(\gamma)}{\Gamma(\gamma)}, \text{ for } t \geq 0 \text{ and } \gamma > 0. \tag{3.15}$$

The Log-Logistic distribution

The survival time, T, is assumed to have a log-logistic distribution with location parameter, α, and scale parameter, β, if $\ln(T) = \alpha + \beta Z$, where Z

$$Z = \frac{\ln(T) - \alpha}{\beta},$$

follows a standard logistic distribution with density function

$$f(Z) = \frac{\exp(z)}{(1 + \exp(z))^2}$$

and cumulative distribution function

$$F(Z) = \frac{\exp(z)}{1 + \exp(z)}.$$

Therefore, the survivor function in terms of Z, is given by

$$S(z) = 1 - F(Z) = \frac{1}{1 + \exp(z)}.$$

Z can be transformed back to T using $\alpha = -\log(\lambda)$, where λ is the hazard rate, and letting $p = \beta^{-1}$. Hence the survival function, in terms of T, is re-written as

$$S(t) = \frac{1}{1 + (\lambda t)^p}, \tag{3.16}$$

while differentiating $S(t)$ with respect to t, the hazard function, $\lambda(t)$, is given by

$$\lambda(t) = \frac{\lambda p(\lambda t)^{p-1}}{1 + (\lambda t)^p}. \tag{3.17}$$

3.3.2 Parametric model fitting

Estimation of model parameters is commonly done either via *methods of moment estimation* or *Maximum Likelihood Method* (MLE). In this section, MLE method is described for estimation of parameters under a parametric modelling set up, since it is the most common approach in literature. Suppose there are n observations that are randomly censored. For a model consisting of p parameters, $\underline{\theta}=(\theta_1, ..., \theta_p)'$, the likelihood function is written as

$$L(\underline{\theta}) = \Pi_{i=1}^{n}[f(t_i|x_i, \theta_1, ..., \theta_p]^{\delta_i}[S(t_i|x_i, \theta_1, ..., \theta_p)]^{\delta_{1-i}}, \tag{3.18}$$

where the binary variable δ_i (for $i = 1, ..., n$) is equal to 1 if the individual experiences failure or 0 if they are censored. In this way, censored observations of individuals surviving the interval [0,t], contribute to the likelihood. To estimate a parameter, θ_j, equation (3.18) is differentiated with respect to θ_j, to obtain the score equation

$$\frac{\partial \log(L(\underline{\theta}))}{\partial \theta_j} = \sum_{i=1}^{n} \frac{\partial \log(L_{\underline{\theta}}(y_i, \theta_i))}{\partial \theta_j} = 0 \tag{3.19}$$

Solving equation (3.19) is not a straight-forward process as it requires the use of iterative methods. The most common methods to optimise equation (3.19) are the *Newton-Raphson Method* and the *Method of Scoring*. The likelihood functions for *j=1,...,p* are given by equation (3.18). Let

$\hat{\underline{\theta}}^0 = (\theta_1^0, ..., \theta_p^0)$ be the initial guess for the solution to equation (3.19). Usually the first guess is obtained by a simpler method, for instance method of moments. Expanding equation (3.19) about $\hat{\underline{\theta}}^0$ using the Taylor series, gives

$$\frac{\partial \log L(\hat{\underline{\theta}})}{\partial \underline{\theta}} = \frac{\partial \log L(\hat{\underline{\theta}^0})}{\partial \underline{\theta}} + \frac{\partial^2 \log L(\hat{\underline{\theta}^0})}{\partial \underline{\theta}^2} \left(\hat{\underline{\theta}} - \hat{\underline{\theta}^0} \right) + ... = 0 \tag{3.20}$$

Ignoring third order and higher terms, let $\underline{\theta}^1$ be the solution to (3.20). Then (3.20) can be written as

$$\hat{\underline{\theta}}^1 = \hat{\underline{\theta}}^0 + \left(-\frac{\partial^2 \log L(\hat{\underline{\theta}}^0)}{\partial \hat{\underline{\theta}}^2} \right)^{-1} \left(\frac{\partial \log L(\hat{\underline{\theta}}^0)}{\partial \hat{\underline{\theta}}} \right), \tag{3.21}$$

where

$$\frac{\partial \log L(\hat{\underline{\theta}})}{\partial \underline{\theta}}$$

is known as the score function at $\hat{\underline{\theta}}$ and

$$I(\hat{\underline{\theta}}) = -\frac{\partial^2 \log L(\hat{\underline{\theta}})}{\partial \underline{\theta}^2}$$

is referred as the sample information matrix at $\hat{\underline{\theta}}$. The Fisher information, $\underline{I}(\underline{\theta})$, is obtained by taking expectation of $I(\hat{\underline{\theta}})$, as shown below.

$$\mathbf{E}\left(I(\hat{\underline{\theta}}) \right) = -\mathbf{E}\left(\frac{\partial^2 \log L(\hat{\underline{\theta}})}{\partial \theta_k \partial \theta_j} \right) = \underline{I}(\underline{\theta}) = \sum_i^n \underline{I}_i(\underline{\theta}) = n\ \underline{I}_i(\underline{\theta})$$

Equation (3.21) is called the Newton-Raphson equation. Substituting in the the Fisher Information instead of the sample information in (3.21)

$$\hat{\underline{\theta}}^1 = \hat{\underline{\theta}}^0 + \underline{I}^{-1}(\hat{\underline{\theta}}^0) \frac{\partial \log L(\hat{\underline{\theta}}^0)}{\partial \underline{\theta}}. \tag{3.22}$$

This iterative method is known as the method of scoring. Once the estimate $\hat{\underline{\theta}}^1$ is obtained, (3.21) is expanded about $\hat{\underline{\theta}}^1$ again, using Taylor series and a new solution, $\hat{\underline{\theta}}^2$, that satisfies (3.19) is determined. This step is repeated until convergence is achieved. At convergence, the new estimate of θ is kept as the final result of the iteration and is referred as the MLE estimator, $\hat{\theta}$ (Cox and Oakes 1984).

3.4 Non-parametric Modelling

Introduction

Non-parametric models offer the possibility to explore survival data that are not restricted to any particular distributional form. In this section, the techniques of estimating and modelling the survivor function under a non-parametric framework are detailed. The methods summarised include the empirical survivor function, life table, actuarial method of survival estimation, the product limit (Kaplan Meier) survival estimator and the Nelson-Aalen survival function.

Empirical Survivor Function

In the absence of censored data, the empirical survivor function is used to estimate the survivor function at a specific time, t. This method estimates the probability that a subject survives beyond a time point, t, by determining the proportion of individuals who are still alive after time t. Thus the survival function, S(t) is given by

$$\hat{S}(t) = \frac{\text{Number of subjects with survival time} > t}{\text{Number of subjects in data set}} \tag{3.23}$$

However, survivor function (3.23) is invalid when censored data are present. Alternatively, one can divide the study period into a set of discrete time intervals and the survival function is estimated exactly at the end of each of these discrete time intervals. In this case the survival estimates are assumed to be proportional to the total number of individuals deemed at-risk in each time interval (Cox and Oakes (1984)).

Product Limit Estimator

The Product Limit (PL) estimation, also known as the Kaplan-Meier (KM) estimation method, offers an alternative way to estimate the survival function, S(t), under a non-parametric framework and produces better estimate of S(t) than the empirical survivor function since it does not exclude any information during the computation of S(t). Consider n individuals in an investigation where time period is divided into time intervals of variable length. Assuming that failure (or death) occurs at the start of each interval, then a series of intervals that contain only one failure at a time can be formed.

If there are $r \leq n$ failures, then we can have t_j (j=1,...,r), as the series of ordered failure times. Let n_j and d_j be the number of individuals at risk prior to t_j and the number of failures at t_j, respectively. Assuming independent failures, survival between t_j and t_{j+1} is given by $(n_j - d_j)/n_j$ and the KM estimate of S(t) for $t_j \leq t < t_{j+1}$, is given by

$$\hat{S}_{\mathrm{KM}}(t) = \prod_{j=1}^{r} \left(\frac{n_j - d_j}{n_j} \right), \tag{3.24}$$

It can be observed that $\hat{S}_{KM}(t)$ is a decreasing step-function where $\hat{S}_{KM}(0)$ is equal to 1 and $\hat{S}_{KM}(t)$ remains unchanged over each time interval $t_j \leq t < t_{j+1}$, j=1,...,r, where $t_{r+1}=\infty$. $\hat{S}_{KM}(t)$ allows the derivation of several quantities of interest such as mean, median, quartiles, associated standard errors, confidence intervals for $\hat{S}_{KM}(t)$, hazard and cumulative hazard functions. It also permits development of plots that provide useful inference about the form of the survival distribution. (Cox and Oakes (1984))

Nelson-Aalen Estimate

This method estimates the cumulative hazard of failure non-parametrically. The Nelson-Aalen cumulative hazard up to the k^{th} interval, is directly estimated as the cumulative sum of the ratio of the number of deaths (d_j) to the number of exposed subjects (n_j) for each time interval, t_j, i.e.,

$$\hat{\lambda}(t_k) = \sum_{j=1}^{k} \frac{d_j}{n_j} \tag{3.25}$$

To estimate the survival function from the Nelson-Aalen cumulative hazard function, Breslow (1972) suggested to use

$$\hat{S}(t) = \exp\left(-\hat{\lambda}(t)\right). \tag{3.26}$$

The Breslow estimator and the KM estimator for the survivor function, $\hat{S}(t)$ are asymptotically equivalent, and usually are quite close to each other, particularly when the number of deaths is small relative to the number exposed.

3.5 Cox Proportional Hazards Model

3.5.1 Model definition

The most important issue in survival modelling is to investigate the effect of covariates on survival time. Cox (1972) proposed the use of the proportional hazards (PH) model. He suggested that for an individual with a vector of covariates $\underline{x}$, the hazard at time t is made up of firstly, a baseline hazard function which does not depend on the covariates, $\underline{x}$, and secondly, a parametric function that represents the effects of the covariates on the failure time, over and above the baseline hazard. Under the PH model, no underlying distribution is assumed for the survival data. Cox (1972) defines the model as

$$h(t, \underline{x}) = h_0(t) \exp(\beta; \underline{x}) \tag{3.27}$$

where $\underline{x}$ is a vector of length p, containing explanatory variables, $h_0(t)$ is an unspecified baseline hazard function (when $\underline{x} = \underline{0}$) and β is a vector of p parameters. The Cox PH model produces constant hazard ratio over time, for two subjects with fixed covariate vectors, $\underline{X}_1$ and $\underline{X}_2$.

$$\frac{h_1(t)}{h_2(t)} = \frac{h_0(t) \exp(\beta; X_1)}{h_0(t) \exp(\beta; X_2)} = \frac{\exp(\beta; X_1)}{\exp(\beta; X_2)} \tag{3.28}$$

Hence the name of *proportional hazards* is given to the model. The PH model is popular because it allows to model the relationship of survival time, through its hazard function, to many covariates simultaneously. Similarly, Cox's survival model can easily accommodate censored data and the occurrence of multiple failures. Moreover, although the underlying survival distribution is unspecified, the model is easily fitted.

3.5.2 Model Fitting

Coefficients β's in the Cox's model are estimated through the maximisation of the partial log likelihood function of the Cox's model (Hosmer et al. (2011)). The likelihood function for the Cox's model is partial because it is based only on the number of individuals, k, who experienced failure during the investigation period instead of all N subjects who may or may not experience failure. Let the set R consists of individuals who experienced failure at the specified time or during the investigation period. Thus the partial likelihood function for the Cox's model is given by the product of the i^{th} individual

probability of experiencing failure at time t instead of other subjects j in the risk set, R. With distinct survival times arranged in ascending order, the partial likelihood function for the Cox's model is given by

$$PL(\beta) = \prod_{i=1}^{k} \left\{ \frac{\exp(\beta X_i)}{\sum_{j \in R(t_i)} \exp(\beta X_j)} \right\} \tag{3.29}$$

The log partial likelihood, $\ell(\beta)$, is written as

$$\ell(\beta) = \sum_{i=1}^{n} \left(\beta X_i - \log \left(\sum_{j \in R(t_i)} \exp(\beta X_j) \right) \right) \tag{3.30}$$

Differentiating (3.30) with respect to β produces the score vector, $U(\beta)$, given by

$$\begin{aligned} U(\beta) &= \sum_{i=1}^{n} \left(X_i - \frac{\sum_{j \in R(t_i)} X_j \exp(\beta X_j)}{\sum_{j \in R(t_i)} \exp(\beta X_j)} \right) \\ &= \sum_{i=1}^{n} \left(X_i - \sum_{j \in R(t_i)} w_{ij}(\beta) X_j \right) \\ &= \sum_{i=1}^{n} \left(X_i - \hat{X}_{w_i} \right), \end{aligned} \tag{3.31}$$

where $\sum_{j \in R(t_i)} w_{ij} X_j$ is the weighted mean of X, over those individuals still at risk at time t_i. The second derivative of the partial log-likelihood function with respect to β is called the Fisher's Information Matrix, and is given by

$$\begin{aligned} I(\beta) &= \sum_{i=1}^{n} \left(\frac{\left(\sum_{j \in R(t_i)} \exp(\beta X_j) \right) \left(\sum_{j \in R(t_i)} X_j^2 \exp(\beta X_j) \right) - \left(\sum_{j \in R(t_i)} X_j \exp(\beta X_j) \right)^2}{\left(\sum_{j \in R(t_i)} \exp(\beta X_j) \right)^2} \right) \\ &= \sum_{i=1}^{n} \sum_{j \in R(t_i)} w_{ij}(\beta) \left(X_j - \hat{X}_{w_{ij}} \right)^2 \end{aligned} \tag{3.32}$$

Partial likelihood estimator of β, denoted by $\hat{\beta}$, is obtained by solving the score equation $U(\hat{\beta}) = 0$ using the Newton-Raphson method presented in

Section 3.3.2. The variance of $\hat{\beta}$ is derived by taking the inverse of the Fisher's Information Matrix, i.e.,

$$\text{Var}(\hat{\beta}) = I(\hat{\beta})^{-1} \tag{3.33}$$

3.5.3 Stratified Cox Models

The Cox proportional hazard model allows the analysis of survival data that has been divided into different disjoint groups (strata). Each stratum is assigned a distinct baseline hazard function, $h_{0k}(t)$, but shares common values of the coefficients β (Therneau and Grambsch (2013). Let individuals $i = 1, ..., n_1$ be in group 1, individuals $n_1 + 1, ..., n_1 + n_2$ be in group 2, and so on, respectively. The hazard of a subject from the k^{th} group is given by $h_{0k}(t)exp(X_i\beta)$, where h_{0k} is the baseline hazard function associated with the k^{th} stratum of covariate X_i.

The log-likelihood function for stratified survival data with K groups, is given by

$$\ell(\beta) = \sum_{k=1}^{K} \ell_k(\beta) \tag{3.34}$$

where $\ell_k(\beta)$ is exactly similar to the log-likelihood function (3.30), but summed over only the individuals in the k^{th} group. Similarly, the score vector $U(\beta)$ and the information matrix $I(\beta)$ are given by $U(\beta) = \sum U_k(\beta)$ and $I(\beta) = \sum I_k(\beta)$, respectively.

3.5.4 Hypothesis Testing and Confidence Interval for Estimated Parameters

To test the statistical significance of covariates in the Cox's regression model, i.e., checking whether the model with the selected covariates explain an individual's survival more accurately than the model without the selected covariates (empty model equivalent to taking the average survival time of the whole population), the likelihood ratio test is used (Therneau and Grambsch (2013).

Let $\beta^{(0)}$ be the true theoretical value of the coefficients and $\hat{\beta}$ be the estimated coefficients. To test the global null hypothesis $H_0 : \hat{\beta} = \beta^{(0)}$, the

test statistic, denoted by LRT, for the likelihood ratio test is given by

$$\text{LRT} = 2(\ell(\hat{\beta}) - \ell(\beta^{(0)}), \tag{3.35}$$

i.e, twice the difference of the log partial likelihood function without any covariates given by $\ell(\beta^{(0)})$, and that of the Cox PH model with p parameters, given by $\ell(\hat{\beta})$. LRT follows a χ^2 distribution with p degrees of freedom, where p is equal to the difference in the number of estimated coefficients under the model with all covariates being tested and that under the model without the covariates being tested.

To check the statistical significance of a specific covariate x, i.e., to test whether x is significantly associated with survival time, the Wald test is used (Therneau and Grambsch (2013). The Wald test assesses the hypothesis that the estimated coefficient for covariate x, denoted by $\hat{\beta}_x$, from the Cox's model, is significantly different from zero. The test statistic, Z, is given by

$$Z = \frac{\hat{\beta}_x}{\widehat{\text{se}}(\hat{\beta}_x)}, \tag{3.36}$$

where $\widehat{\text{se}}(\hat{\beta}_x)$ is the standard error associated with $\hat{\beta}_x$. The test statistic Z follows a standard normal distribution and can also be used to formulate confidence interval for $\hat{\beta}_x$. The $(1-\alpha)\%$ confidence interval (CI) for $\hat{\beta}_x$ is given by

$$\text{CI}(\hat{\beta}_x) = \hat{\beta}_x \pm Z_{1-\alpha/2}\widehat{\text{se}}(\hat{\beta}_x), \tag{3.37}$$

where Z_α represents the critical value at $(1-\alpha)\%$ level from the standard normal distribution. When the $(1-\alpha)\%$ CI of an estimated coefficient contains zero, it means that the covariate associated with that estimated coefficient does not significantly affect hazard of failure.

3.5.5 Tied or Grouped Observations

Survival data may be tied mainly because of two reasons. Firstly, for continuous survival data, ties may exist because of rounding and thus they are in a form of incomplete data, where the true times are not tied. Secondly, tied survival data may arise when the survival times are discrete. Tied survival data affect the form of the partial likelihood for the Cox model.

As shown by the equation (3.30), the partial likelihood function for untied data is a product over event times of the likelihoods for different individuals. In the presence of ties, the exact form of the partial likelihood cannot be determined since the survival time data is not precise. To illustrate the problem of tied survival data, consider k individuals in time order, with the first two experiencing failures at the same recorded time. Let r_i be the risk score associated with the i^{th} individual. In the absence of ties, the first two terms of the log-likelihood function (i.e., for the first two subjects) are either

$$\left(\frac{r_1}{r_1 + r_2 + ... + r_k}\right)\left(\frac{r_2}{r_2 + r_3 + ... + r_k}\right) \tag{3.38}$$

or

$$\left(\frac{r_2}{r_1 + r_2 + ... + r_k}\right)\left(\frac{r_1}{r_1 + r_3 + ... + r_k}\right), \tag{3.39}$$

but one cannot precisely select which of (3.38) or (3.39) is the true likelihood function. To address this, Therneau and Grambsch (2013) presented three approximation methods; namely (1) Breslow approximation, (2) Effron approximation and (3) Discrete method.

The Breslow approximation is one of the simplest methods to estimate the likelihood function for tied survival data. It has been proposed independently by Breslow (1972) and Peto (1972), and uses the complete sum of $r_1 + r_2 + ... + r_k$ as the denominator of the partial log-likelihood function (Equation (3.40)) as shown below. Under this approximation, the fractions with the biggest risk pool for each time-tied event are used.

$$\left(\frac{r_1}{r_1 + r_2 + ... + r_k}\right)\left(\frac{r_2}{r_1 + r_2 + ... + r_k}\right) \tag{3.40}$$

Generalising (3.40) to the whole population for the example set up at the beginning of this section, Breslow's approximation is given by

$$\prod_{i=1}^{k} \frac{\exp(\beta X_{(i)+})}{\sum_{j \in R(t_i)} \left(\exp(\beta X_j)\right)^{d_i}} \tag{3.41}$$

where d_i is the number of failures at time t_i. $R(t_i)$ represents the set of individuals who experienced failure at time t_i and X_{i+} is the sum of the values of the covariates, i.e., $X_{i+} = \sum_{j \in D(t_i)} X_j$. According to Therneau and

Grambsch (2013), the Breslow approximation method is the least accurate, compared to the Effron and discrete method, because it accounts for failure more than once in the denominator, thereby producing biased estimate of β.

Under the Efron (1977) approximation method, a more accurate partial likelihood is derived by using the average denominator. The risk scores of the individuals with time-tied events in subsequent risk groups are multiplied by the probability that they will be in the subsequent risk group. Applying the Effron approximation method to the example set up at the start of this section, the risks of the two subjects are multiplied by a factor of 0.50 in the second risk set. The argument behind this adjustment is that subject 1 and 2 both have a 50% probability of being in the second risk group. The log-partial likelihood function (3.40) becomes

$$\left(\frac{r_1}{r_1 + r_2 + r_3}\right)\left(\frac{r_2}{0.5r_1 + 0.5r_2 + r_3}\right) \tag{3.42}$$

Generalising (3.42) to the whole population study, the partial log-likelihood function under the Effron approximation is given by

$$\prod_{i=1}^{k} \frac{\exp(\beta X_{(i)+})}{\prod_{k=1}^{d_i}\left(\sum_{j\in R(t_i)}\exp(\beta X_j) - \frac{k-1}{d_i}\sum_{j\in D(t_i)}\exp(\beta X_j)\right)}. \tag{3.43}$$

The discrete method does not assume any ordering of the tied events and treats time as a discrete variable. For any failure at time t, a Cox model with discrete event times is based on a proportional odds model where the risk score is multiplied by the odds ratio of the baseline hazard instead of the baseline itself (Therneau and Grambsch (2013)):

$$\frac{\hat{h}_i(t)}{1 + \hat{h}_i(t)} = \left(\frac{\hat{h}_0(t)}{1 + \hat{h}_0(t)}\right)\exp(\beta X_i) \tag{3.44}$$

Revisiting the scenario of two failures with tied times, under the discrete method, the partial log-likelihood function determines the probability that individual 1 and 2 belong to the group of individuals with time-tied failures instead of other individuals in the risk set and can be written as (Therneau and Grambsch (2013)):

$$\frac{r_1 r_2}{r_1 r_2 + r_1 r_3 + r_2 r_3}. \tag{3.45}$$

In general terms, for the whole population, the partial log-likelihood function under the discrete method is given by

$$\prod_{i=1}^{k} \frac{\exp(\beta X_{i+})}{\sum_{z \in Z(t_i)} \exp(\beta X_z)}, \tag{3.46}$$

where Z represents all combinations of selected individuals with time-tied events, z is one combination of time-tied individuals from the set Z, d_i is the number of failures at time t_i and X_z is the sum of the covariates values, $X_z = \sum_{j=1}^{d_i} X_j$.

In conclusion to this section on tied failure times, three different methods of approximation for the estimation of the partial likelihood function have been presented. The Breslow approximation is the simplest one but it produces biased estimates of β's. The discrete partial likelihood produces better approximation of the partial likelihood function than the Breslow method but it is time consuming and computationally ineffective. The Efron approximation method is the most reliable approximation among the three methods as it is computationally efficient and does not produce bias in the estimation of β's. In this research, for a dataset with time-tied events, the Efron approximation method is used for the formulation of partial likelihood function to estimate the model parameters.

3.5.6 Residuals

Introduction

In Cox PH model, the three common types of residuals of interest are the Score residuals, the Schoenfeld residuals and the Martingale residuals, respectively. In this section, a summary of the definitions and uses of these residuals in assessing the assumptions and adequacy of the Cox PH model is provided. To define these residuals, consider a set of n independent subjects such that the counting process $N_i \equiv \{N_i(t),\ t \geq 0\}$ for the i^{th} subject in the set, indicates the number of observed failures over time t. N_i is therefore analogous to a step function with increase of size $+1$ and $N_i(0) = 0$.

Martingale Residuals

The Martingale process for the i^{th} individual in a survival analysis is given by

$$M_i(t) = N_i(t) - E_i(t) = N_i(t) - \int_0^t Y_i(s)\hat{H}_0(s)\exp(X_i(s)\hat{\beta})ds, \qquad (3.47)$$

where $\hat{H}_0$ is the cumulative baseline hazard and $N_i(t)$ is the counting process (defined in the introduction of section 3.5.6) for the i^{th} individual in the set of subjects who experienced failure during the investigation (Therneau and Grambsch (2013)). In simpler terms, the Martingale residuals are defined as the difference between the observed number of failures and the conditionally expected number of failures associated with the i^{th} individual, given the fitted model, follow-up time and the complete course of time-varying covariates present in the fitted model. According to Therneau and Grambsch (2013), Martingale residuals are important, firstly, as a direct assessment of observations that are poorly fitted by the model and secondly, for the evaluation of the functional form of a covariate in the model.

Score residuals

For the i^{th} individual in a survival analysis study, the score process is given by

$$U_i(\beta, t) = \int_0^t [X_i(s) - \bar{x}(\beta, s)]dM_i(s), \qquad (3.48)$$

where $\bar{x}(\beta, s)$ is the weighted mean of the covariates over those at risk at time s and M_i is the martingale residual for the i^{th} individual (given by equation (3.47)). $U_i(\beta, t)$ is a row vector of length p, where p is equal to the number of covariates (Therneau and Grambsch (2013). The Score residuals are used for the assessment of influential individual observations in the model fit and to test for the robustness of the estimated variance of the Cox PH model. Furthermore, Lin and Wei (1989) proposed the use of the score process to assess the proportional hazards assumption of the fitted model globally.

Schoenfeld residuals

The Schoenfeld residual (Schoenfeld, (1980)), r_k, for the k^{th} covariate in a Cox model, are derived from the first derivative of the partial likelihood

function (3.30) and is given by

$$r_k = \sum_{i=1}^{m} \left(X_{ik} - \hat{X}_{q_{ik}} \right), \tag{3.49}$$

where m is the total number of failures in the study, X_i is the set of measured covariates for the i^{th} individual with ordered survival time t_i and $\widehat{X}_{q_{ik}}$ is the conditional means of the covariates for the i^{th} individual with ordered survival time t_i and at risk of encountering failure. Hence the Schoenfeld residuals for the k^{th} covariate associated with the i^{th} individual who experienced failure during the investigation period, can be written as

$$\hat{r}_{ik} = X_{ik} - \hat{X}_{q_{ik}} \tag{3.50}$$

The sum of the Schoenfeld residuals is expected to be zero since the coefficients, β, are estimated via the maximum likelihood method by equating the first derivative of the partial log-likelihood function to zero. In presence of censored data, the Shoenfeld residuals are not informative for the model fit because the partial likelihood function does not account for censored individuals. To address this issue, Therneau and Grambsch (2013) proposed the use of scaled Schoenfeld residuals. Under this approach, the Schoenfeld residuals are scaled by their estimated variance. The variance of the vector of Schoenfeld residuals associated with the k^{th} covariate, for the i^{th} individual, denoted by $\hat{r}_i' = (\hat{r}_{i1}, \hat{r}_{i2}, ..., \hat{r}_{ik})$, is estimated by

$$r_i^* = \frac{\hat{r}_i'}{m\widehat{\text{Var}}(\hat{\beta})}, \tag{3.51}$$

where m is the number of failures in the study and $\widehat{\text{Var}}(\hat{\beta})$ is the estimated variance-covariance matrix of the estimated coefficients, $\hat{\beta}$'s. Scaled Schoenfeld residuals are used in the assessment of proportional hazards assumption for the Cox model.

3.5.7 Assumption of proportional hazards

The key assumption in survival analysis of non-parametric data using the Cox proportional hazards model is that the relative hazard for any two individuals i and j, follows the time independent relationship

$$\frac{h_i(t)}{h_j(t)} = \frac{h_0(t)\exp(\beta; X_j)}{h_0(t)\exp(\beta; X_j)} = \frac{\exp(\beta; X_j)}{\exp(\beta; X_j)}.$$

In addition, the above relationship is valid for each single covariate in the model. It provides the basis to explain the multiplicative effect of each covariate in the hazards function on the fitted survival function, given the covariate is time independent. Therefore the assumption of proportional hazards is key and its assessment is crucial for model assessment and interpretation. Testing proportional hazards assumption for Cox survival model is carried out via (1) graphical test using the logarithmic form of the fitted survival curve and the residuals associated with the Cox PH model and (2) statistical tests using the scaled Schoenfeld residuals described in Section 3.5.6.

According to Therneau and Grambsch (2013), the simplest test to assess the assumption of proportional hazards is the graphical check of the fitted survival curve. Under the assumption of proportional hazards, the survival function for the i^{th} individual satisfies

$$S_i(t) = \exp(-h_0(t)\beta X_i),$$

hence rearranging $S_i(t)$ as

$$\log[-\log(S_i(t))] = \log[h_0(t)] - X_i\beta$$

shows that $\log[-\log(S_i(t))]$ is proportional to the measured covariate. If the fitted Cox model is correct, the survival curves for the different levels of the covariate, should be approximately parallel when plotted on the log-log scale, showing that the impact of the covariate of interest on the estimated hazard, is proportional. In the case when the covariate has many levels or is continuous, this method does not help in assessing the assumption of proportional hazards because the survival curves on the log-log scale become sparse and do not yield a good estimate of how close to parallel are the curves.

The Schoenfeld residuals can be used to test the assumption of constant proportional hazards for a fitted Cox model. Under the assumption of proportional hazards, a plot of the cumulative Schoenfeld residuals (Grambsch and Therneau (1994) and the cumulative score residuals (Lin et al. (1993)), respectively, versus the study time, scaled to $(0, 1)$, should yield a Brownian bridge pattern (identified as a random walk beginning and ending at 0) for each survival curve associated with each level. A major disadvantage with these plots is that they are often difficult to interpret and visualise the pattern required under the assumption of proportional hazards.

A different approach for testing proportional hazards for individual covariates in the Cox's model, firstly consists of ranking the survival time, t, of each individuals in the study, in the order of the occurrence of failure, to obtain the set of transformed survival times, denoted by $t^* = (1, 2, ..., m)$, where m is the number of failures during the study. Secondly, the correlation, ρ_k, between the scaled Schoenfeld residuals of the k^{th} covariate and t^* is estimated and its significance is tested. ρ_k asymptotically follows a χ^2-distribution with one degree of freedom. If ρ_k is not significantly different from zero, then the assumption of proportional hazards holds for the k^{th} covariate. In addition, a global test of proportional hazards over all covariates that assesses the statistical significance of the estimated correlation, $\hat{\rho}^*$, between the scaled Schoenfeld residuals of all covariates in the model together and the transformed survival time, t^*, is also useful. If $\hat{\rho}^*$ is not significantly different from zero, then assumption of proportional hazards is not violated (Grambsch and Therneau (1994)).

3.5.8 Time dependent covariates

A time-dependent covariate is an explanatory variable whose value changes over time and thus does not satisfy the assumption of proportional hazards, when incorporated in a Cox's model. Time dependent covariates arise (1) in investigation involving repeated measurements on a subject or (2) when the effect of the covariate of interest changes with time as the investigation period goes on. Under the presence of time-dependent covariates, two methods are proposed to account for time-varying effect of the covariates in the Cox's model. Firstly, the follow-up time is split in intervals over which the assumption of proportional hazards hold and a Cox PH model is fitted to the data for each interval. However such approach is highly dependent on the definition of the follow-up interval where assumption of proportional hazards are not violated.

Under the second method, an interaction between survival time, t and the time-dependent covariate, is added to the Cox model. Under this set up, the Cox's model is no longer a proportional hazards model and is referred as an extended Cox's model, written as

$$h(t, \underline{x}) = h_0(t) \exp(X\beta(t)), \tag{3.52}$$

where the estimated coefficient, $\hat{\beta}(t)$, is not constant and varies over time. A

plot of $\hat{\beta}_j(t)$, associated with the j^{th} covariate versus time therefore produces a non-horizontal line, showing the variation of the covariate effect with respect to time. An extended Cox model represented by (3.52), would compare the risk of failure between each level of the covariate at each event time, and would re-evaluate which risk group each person belonged in, based on the level of the covariate. The hazard of failure at time t depends on the value of the time dependent covariate at time t.

3.5.9 The Cox Model With Frailty

Introduction

The concept of frailty provides a convenient method to account for random effects, relationships and unobserved heterogeneity in the models for survival data. In its basic form, frailty is defined as an unknown random factor that affects the hazard of failure associated with an individual or related individuals. The concept of frailty was initiated by Greenwood and Yule (1920) and introduced by Vaupel et al. (1979) for survival models at the univariate level, while Clayton (1978) applied the concept of frailty to multivariate survival model. The frailty term allows the investigator to account for unknown and unmeasured covariates. In survival analysis, frailty model is a random effect time-to-event model whereby the frailty has a multiplicative effect on the baseline hazard function.

Univariate Cox model with frailty

The presumption of homogeneity of a study population in the field of epidemiology and medicine does not apply in many practical situations. The impact of a particular epidemiological or medical intervention or the effect of various explanatory variables vary greatly between different categories of patients. Vaupel et al. (1979) proposed the use of univariate frailty models in survival analysis to account for such unobserved heterogeneity among the study population.

> "The key idea is, that individuals possess different frailties, and that those patients who are most frail will die earlier than the others" (Vaupel et al. 1979).

There exist two major explanations as to why the inclusion of all important variables at the individual level into the analysis, is impossible. Firstly,

there may be too many covariates to be accounted for, in the model, and secondly, the investigator is unable to measure all the appropriate variables. Both arguments give rise to two sources of variability in survival data: (1) variation due to measurable risk factors, which can theoretically be formulated and (2) heterogeneity due to unknown covariates, which is therefore theoretically unpredictable.

Under the proportional hazards model set up, ignoring a set of the important covariates contributes to biased estimation of the regression coefficients and the hazard rates. For example, if there are two groups of patients in a study where one group experiences a higher risk of failure, then the remaining persons at risk, will to a large extent, have a lower risk of failure. Without accounting for unobserved frailty, estimated individual hazard rate would therefore be underestimated and the degree of underestimation would be higher as time progresses.

The univariate frailty model is an extension of the Cox model (3.27), where the hazard of failure is directly proportional to an unknown random variable, Z, as shown below (Therneau and Grambsch (2013).

$$h(t, \underline{x}) = Z h_0(t) \exp(\beta; \underline{x}) \tag{3.53}$$

The frailty Z is assumed to have an expected value of one and has a multiplicative effect on the individual hazard rate. It either increases ($Z > 1$) or decreases ($Z < 1$) the individual risk of failure. The survival function, S, which describes the proportion of individuals surviving, is given by

$$S(t|X, Z) = \exp(-Z \exp(\phi(\beta; \underline{x}) \int_0^t h_0(s) ds) \tag{3.54}$$

Under the setting of a shared frailty model, individuals in a study population are grouped into distinct homogeneous categories. The concept in this setting is that individuals who are at risk of failure within a defined and controlled environment, can be assumed to have the same level of frailty. For example, individuals living in the same accommodation or patients receiving treatments within the same ward, are exposed to the same level of unknown risks and can thus be assumed to share the same frailty. Consider a study population where individuals i ($i = 1, ..., n$), are grouped into j homogeneous categories ($j = 1, ..., q$). The shared frailty Cox model, for the i^{th} individual,

is written as (Therneau and Grambsch (2013):

$$h_i(t) = h_0(t) \exp(\beta X_i + \omega Z_i) \ , \tag{3.55}$$

where X_i and Z_i represent the i^{th} rows of the covariate matrices of size $n \times p$, for covariate set X and unknown frailty set Z, respectively. Under model (3.55), X and β estimate the fixed effects due to the measurable covariates, ω is the estimated vector of random effects (frailties) and Z is the design matrix where its Z_{ij} element is equal to 1 for subject belonging to the same frailty group j or 0 otherwise.

The shared frailties, Z_i's, are assumed to be identical and independently distributed random variables from a Gamma distribution with mean equal to 1 and variance Θ. Since the hazard function is non-negative, Z_i's are chosen to follow a Gamma distribution instead of a Gaussian distribution because the former distribution only takes positive values (Therneau and Grambsch (2013). The variance Θ determines the degree of variability between clusters and the level of correlation within each cluster. Higher value of Θ indicates that the cluster effects are more widespread between clusters and a stronger correlation within clusters. Under the assumption of Gamma distributed shared frailties, the correlation between individuals within the same cluster is determined by Kendall's τ, given by

$$\tau = \frac{\Theta}{2 + \Theta}, \tag{3.56}$$

where Θ is the estimated variance of the cluster effect.

Parameters of a shared frailty Cox proportional hazards model are estimated by maximising the partial log-likelihood function of the model, with respect to each of the parameters being estimated. Consider a population of n individuals, grouped into K clusters where the i^{th} group consists of n_i individuals. Let the number of failures in the i^{th} group be $F_i = \sum_{j=1}^{n_i} f_{ij}$ and $\Lambda_0(t)$ represent the cumulative baseline hazard. The partial log-likelihood for a shared frailty Cox proportional hazards model, is given by (Therneau

and Grambsch (2013):

$$\begin{aligned}\ell(\Theta,\beta) = &\sum_{i=1}^{K}\left(F_i\ln(\Theta) - \ln(\Gamma(\Theta^{-1})) + \ln(\Gamma(\Theta^{-1}+F_i))\right)\\ &- (\Theta^{-1}+F_i)\ln\left(1+\Theta\sum_{j=1}^{n_i}\Lambda_0(t)\exp(\beta X_{ij})\right)\\ &+\sum_{j=1}^{n_i} f_j\left(\beta X_{ij} + \ln(\Lambda_0(t))\right)\end{aligned} \tag{3.57}$$

The partial log-likelihood function, $\ell(\Theta,\beta)$, is maximised using the EM-algorithm described in Section 3.5.2 and can be summarised as follows, when applied to a shared frailty Cox's model:

(1) Coefficients $\hat{\beta}$ are estimated from the fitted model.

(2) The shared frailty terms, $\hat{Z}_i$, are estimated.

(3) Fitted Cox's model is updated with the estimated $\hat{Z}_i$ and new estimates of $\hat{\beta}$ are determined.

(4) Steps 2 and 3 are iterated until convergence of $\hat{\beta}$ and $\hat{Z}_i$, respectively, is achieved.

Statistical significance of estimated cluster effect, $\widehat{\Theta}$ is checked using the likelihood ratio test where the test statistics, LRT, is given by

$$\text{LRT} = 2\left(\ell(\widehat{\Theta},\hat{\beta}) - \ell(0,\hat{\beta})\right), \tag{3.58}$$

where $\ell(\widehat{\Theta},\hat{\beta})$ and $\ell(0,\hat{\beta})$ are the partial log-likelihood function evaluated at $\Theta = \widehat{\Theta}$ for the shared frailty Cox's model and the partial log-likelihood function evaluated at $\Theta = 0$. LRT is assumed to follow a χ^2-distribution with one degree of freedom (Hosmer et al. (2011)).

3.5.10 Model Diagnostics

A key aspect in evaluating the adequacy of a survival model is through the investigation of regression diagnostic statistics to identify subjects who (1)

have irregular configuration of covariates, (2) influence the estimates of the model parameters and (3) have a leverage on the model fit. In this section, the methods used to identify influential subjects with high leverage, are summarised.

Leverage is a statistical measure of how unusual are the values of a subject's covariates. Under the Cox proportional hazards set-up, Hosmer et al. (2011) used the score residuals, defined by equation (3.48) to assess leverage for continuous variables. Hosmer et al. (2011) showed that the score residuals follow a linear regression leverage property for continuous variables, i.e., the further away is the value of the covariate from the mean, the larger is the score residual value. A plot of the score residuals versus the continuous variable should yield a fan shape display with no observations having unexpectedly large values.

To assess the influence of a subject on the estimators of the fitted model, Cook's distance and jackknife residuals are employed. Cook's distance (Hosmer et al. (2011)) is defined as

$$\left(\hat{\beta} - \hat{\beta}_{-i}\right)' \left[\widehat{\text{var}}(\beta)\right]^{-1} \left(\hat{\beta} - \hat{\beta}_{-i}\right) \ , \tag{3.59}$$

while the jackknife residual (Therneau and Grambsch (2013), J_i, is defined as

$$J_i = \left(\hat{\beta} - \hat{\beta}_{-i}\right) \ , \tag{3.60}$$

where $\hat{\beta}$ and $\hat{\beta}_{-i}$ are the model estimators with all observations and without the observations for the i^{th} subject, respectively. A plot of the Cook's distance and the jackknife residuals versus the martingale residuals (equation (3.47)), respectively, yields a cup shaped figure where the largest values of the Cook's distance and the jackknife residuals, on either sides of the cup shape, correspond to subjects that have been poorly fitted.

3.5.11 Overall goodness of fit

Goodness-of-fit of a model is defined as the degree to which the observed data, used to fit a model, agrees with theoretical expected values. Several authors such as Schoenfeld, (1980), O'Quigley and Pessione (1989), Weissfeld (1990) and Lin et al. (1993) proposed different tests to assess the overall

goodness-of-fit of a proportional hazards model. However these tests were computationally difficult and time consuming to implement.

Grønnesby and Borgan (1996) proposed a test that was computationally feasible. The authors firstly partition the dataset into K groups, based on the ranked values of the estimated risk score. Secondly the test uses the sum of the martingale residuals (equation (3.47)) of each of the K groups and then compares the observed number of events to the model-based expected number of events. Grønnesby and Borgan (1996) showed that the test statistic for this test (the difference between observed and theoretical expected number of events) follows Chi-square distribution with $(K-1)$ degrees of freedom.

May and Hosmer (1998) showed that Grønnesby and Borgan (1996) method is equivalent to the score test, where May and Hosmer (1998) approximated the score test with the partial likelihood ratio test for computational feasibility and effectiveness. The expected and observed number of events are compared via the counting process theory. where the counting function is approximated by a Poisson random variable with mean equal to the variance of the cumulative hazard function. For large number of data points, the Poisson distribution can be approximated by the Normal distribution. Hence a simplified way to compare the observed and expected counts of events, is to determine the Z-score for each of the K groups by dividing the difference in observed and expected number of events by the square-root of the expected number of events. Then the two-tailed p-value of z-score is obtained from the standard Normal distribution.

In addition, the graphical test proposed by Arjas (1988) can also be used as an assessment technique for the goodness-of-fit of a Cox proportional hazards model. The author proposed to plot the cumulative observed number of events against the model-based cumulative estimated number of events for uncensored observations, within each partition of the data to graphically assess the model fit. Under a perfect model, the plotted points should follow a straight line, angled at 45° with both axes and starting at the origin.

A further assessment of the performance of a fitted model under the Cox proportional hazards set-up is the use of the R^2 statistics (Hosmer et al., 2011). The R^2, adjusted R^2 denoted by R^2_{adj}, or the predicted R^2 denoted by R^2_p, statistics of a fitted model can be used to assess the performance of the

model. R^2 is the proportion of variability among the observed data points that is explained by the fitted model. R^2_{adj} is a modified version of the R^2 and accounts for the number of covariates in the model, while R^2_p measures the degree to which the model predicts responses for new observations. R^2 and R^2_{adj} are defined as

$$R^2 = 1 - \exp\left(\frac{2}{n}(L_0 - L_p)\right) ,$$

$$R^2_{adj} = 1 - \left[\frac{(1-R^2)(n-1)}{(n-p-1)}\right] , \tag{3.61}$$

where L_0 is the partial log-likelihood function for the model with no covariates, L_p is the partial log-likelihood function for the model with p covariates, n is the number of observations in the data and p is the number of covariates in the model. As demonstrated by Schemper and Stare (1996), there is no straight-forward and easy method to determine the R^2 measure of a proportional hazards model. Furthermore, the measures are hugely dependent on the proportion of censored data. A perfectly adequate model may have a low R^2 measure because of the high amount of censoring in the data. Therefore assessment of model performance via the use of R^2 or similar adjusted measures should be carried out conservatively.

Another measure for the goodness of fit of a model is the Harrell's concordance, also known as as C-index (Harrell, (2001)). It is the percentage of all pairs of individuals whose survival times can be ordered such that the individual with the higher predicted survival, is the one who survived longer. Such a pair is known as a concordant pair (C). A discordant pair (D) is one where the individual with the higher predicted survival, is the one who has the shorter survival. A tied pair (T) is obtained when it is impossible to determine which of the two individuals experienced failures first. The C-index is a proportion of concordance ($0 \leq C$-index ≤ 1) between observed and predicted survival and is defined as

$$C\text{-index} = \frac{C + T/2}{C + D + T}, \tag{3.62}$$

where C represents number of concordant pairs, D is the number of discordant pairs and T is the number of tied pairs. A small value of C-index

(C-index $\leq 50\%$), indicates a model that is poorly predicting the outcomes, while a C-index close to one (C-index $\geq 70\%$) indicates a model with strong prediction of the outcomes. In the field of survival data analysis, Therneau and Grambsch (2013) suggested that an acceptable range for the C-index of a strong model is between 60% and 70%.

3.5.12 Interpretation of Hazard Rate

In this research, the event of interest is all-cause death after THR surgery. Adjusted hazard of death (HR) also referred to as risk of mortality, can be interpreted as the average change in the risk of death after THR procedure when comparing a group of subjects to the baseline category, with respect to one particular covariate, during the whole follow-up period of the study while adjusting for the remaining covariates (Therneau and Grambsch 2013). An HR smaller than one causes a decrease in mortality risk after THR procedure and thus survival after THR is significantly better than baseline group (controls). When HR is equal to zero, there is no significant change in mortality risk after THR and therefore cases and baseline group have similar survival. An estimated HR greater than one leads to an increase in mortality risk after THR and thus survival of cases is worse than baseline group.

Communicating statistically based risk information such as the HR is the essential part of any research. The HR can be explained in terms of the number of years gained or lost in effective age. Given the hazard rate associated with a particular subject, the effective age of this subject is defined as the age at which the baseline individual (or matched control) will have the same hazard rate as the former. In this research, effective age refers to the age at which controls have the same estimated mortality risk as the individuals who underwent THR surgery. The concept of effective age comes from the Gompertz model in which the annual one year increase in mortality risk associated with ageing is approximately unchanged for individuals aged between 30 and 95, respectively (Brenner et al., (1993), Vaupel, (2010)). In England and Wales, the estimated increase in mortality risk for individuals aged between 30 and 95, inclusive, is roughly equal to 1.1 for 2010-2012 (Spiegelhalter, (2016)). This means that the average risk of an individual in England and Wales dying before his next birthday is increased by approximately 10% per year. Assuming that the yearly increase in risk of mortality associated with ageing remains constant with age and secondly, proportional hazards hold

for the covariates over time (Brenner et al., (1993), the estimated number of years gained or lost in effective age, $\delta(\text{age})$, after THR procedure in England and Wales is given by

$$\delta(\text{age}) = \frac{\ln(\widehat{\text{HR}})}{\ln(1.1)} \backsimeq 10.49 \times \ln(\widehat{\text{HR}}) \tag{3.63}$$

3.6 Multiple Logistic Regression

3.6.1 Model Definition and parameter estimation

Logistic Regression

In many epidemiological studies, outcomes observed are often dichotomous in nature; i.e., the outcome of interest is defined as a binary discrete variable with two levels. For example, in survival analysis of longitudinal data, the outcome observed is either death or alive by the end of the study period, where the outcome variable is defined as one for death and zero for being alive. Logistic regression is a mathematical approach, based on the theory of generalised linear models of Nelder and Baker (1972), used to describe the relationship between a dichotomous variable and a list of variables measured during an experiment.

Consider a binary or dichotomous random variable Z that can take two possible outcomes. For a dataset of independent observations and of sample size N, Z can be described as a column vector of N Bernoulli random variables, Z_i. Let the values 1 and 0 represent the occurrence (success) or absence (failure) of the event of interest being investigated, respectively, for the dichotomous variable, Z. Let I be the total number of clusters, each with n_i observations, n be a column vector with elements n_i for $i = 1$ to I, where $\sum_{i=1}^{I} n_i = N$. Let Y denote a column vector of length N where its elements, Y_i, represent the number of occurrences for the event of interest (success) and the column vector $\underline{y}$ is defined such that its elements y_i represent the counts of the number of successes for each cluster in the dataset. Let π be a column vector, of length I, with its members, $\pi_i = \Pr(Z_i = 1|i)$, representing the probability of occurrence of the event of interest (success) for any given observation from the i^{th} cluster. For each cluster in the dataset, a set of K independent covariates are measured. Let X represent the design matrix composed of I rows and $K + 1$ columns, where the first element of each row

X, x_{i0}, is equal to one.

The general form of the logistic regression model for the above description is written as

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \sum_{k=0}^{K} x_{ik}\beta_k, \text{ for } i = 1, 2, ..., I. \tag{3.64}$$

The transformation $\ln[\pi_i/(1-\pi_i)]$ is also referred as the logit transform. The logistic regression (3.64) therefore, expressed the log-odds probability of an occurrence of the event of interest being investigated to a linear combination of the covariates X.

Parameter Estimation

The three common methods employed for estimation of parameters for logistic regression model are the maximum likelihood method (Hosmer Jr et al. (2013a)), non-iterative weighted least square method (Grizzle et al. (1969)) and discriminant function analysis method (Cornfield (1962)), respectively. In this research, estimation of logistic regression parameters is carried out via the maximum likelihood estimation (MLE) method and described in this section. The aim of this approach is to estimate the $K + 1$ unknown parameters β in equation (3.64) by maximising the log-likelihood function of equation (3.64) (Hosmer Jr et al. (2013a)).

The likelihood function for equation (3.64) is given by

$$\mathrm{L}(\beta|y) = \prod_{i=1}^{I} \frac{n_i!}{y_i!(n_i - y_i)!} \pi_i^{y_i} (1 - \pi_i)^{n_i - y_i}. \tag{3.65}$$

The factorial terms in equation (3.65) does not contain any of the π_i and are constants that can be ignored when equation (3.65) is maximised to estimate β. Re-arranging equation (3.65) and eliminating the factorial terms, the likelihood function of equation (3.64) can be expressed as

$$\mathrm{L}(\beta|y) \asymp \prod_{i=1}^{I} \left(\frac{\pi_i}{1-\pi_i}\right)^{y_i} (1 - \pi_i)^{n_i}. \tag{3.66}$$

Solving for π_i in Equation (3.64) gives

$$\pi_i = \left(\frac{\exp(\sum_{k=0}^{K} x_{ik}\beta_k)}{\exp(\sum_{k=0}^{K} x_{ik}\beta_k) + 1} \right). \tag{3.67}$$

Substituting the exponential of equation (3.64) and equation (3.67), for the first and second term of the likelihood function (3.66), yields in

$$\prod_{i=1}^{I} \left(\exp(y_i \sum_{k=0}^{K} x_{ik}\beta_k)(1 + \exp(\sum_{k=0}^{K} x_{ik}\beta_k))^{-n_i} \right). \tag{3.68}$$

Equation (3.68) is referred as the kernel of the likelihood function. The log-likelihood function of the logistic regression model, $\ell(\beta)$, is given

$$\ell(\beta) \asymp \sum_{i=1}^{I} y_i \left(\sum_{k=0}^{K} x_{ik}\beta_k \right) - n_i \ln(1 + \exp(\sum_{k=0}^{K} x_{ik}\beta_k)). \tag{3.69}$$

Hence, the first derivative of the log-likelihood function (3.69) with respect to each β_k is given by

$$\frac{\partial \ell(\beta)}{\partial \beta_k} = \sum_{i=1}^{N} y_i x_{ik} - n_i \pi_i x_{ik}, \tag{3.70}$$

The second derivative with respect to each β_k estimates the variance-covariance matrix for the estimates of βs and is written as

$$\frac{\partial^2 \ell(\beta)}{\partial \beta_k \partial \beta_{k'}} = -\sum_{i=1}^{N} n_i x_{ik} \frac{\partial}{\partial \beta_{k'}} \left(\frac{\exp(\sum_{k=0}^{K} x_{ik}\beta_k)}{1 + \exp(\sum_{k=0}^{K} x_{ik}\beta_k)} \right) \tag{3.71}$$

and can be simplified to the following form according to Czepiel (2002):

$$\frac{\partial^2 \ell(\beta)}{\partial \beta_k \partial \beta_{k'}} = -\sum_{i=1}^{N} n_i x_{ik} \pi_i (1 - \pi_i) x_{ik'} \tag{3.72}$$

To determine the MLE estimates of the β's, equation (3.70) is set to zero. This yields a system of $K+1$ non-linear equations, each with $K+1$ unknown parameters. The most common approach to find the solution to such a system of non-linear equations is the Newton-Raphson method (Hosmer Jr et al. (2013a)). Applied to the logistic regression model, the steps involved in the Newton-Raphson method to estimate βs are summarised below.

(1) Using equation (3.70), the first step is expressed in the form of

$$\beta^{(1)} = \beta^{(0)} + \left(X^T W X\right)^{-1} X^T (y - \mu), \tag{3.73}$$

where X is the design matrix defined in section 3.6.1, W is a square matrix of order N, with elements $n_i\pi_i(1-\pi_i)$ on the diagonal and zeros everywhere else and μ is a column vector of length N with elements $\mu_i = n_i\pi_i$.

(2) Substituting $\beta^{(0)}$ by the initial estimates of the β_k, new estimate, $\beta^{(1)}$ is obtained from step 1.

(3) $\beta^{(0)}$ in step 1 is again substituted by the newly estimated $\beta^{(1)}$ from step 2.

(4) Steps (1)–(3) are repeated until the estimated βs have converged, i.e., there is no difference in the estimated βs of two consecutive iterations.

Shared frailty multilevel logistic regression model

In public health, demography and sociology, large-scale surveys often follow a hierarchical data structure as the surveys are based on multi-stage stratified cluster sampling. The appropriate approach to analysing such data is therefore based on nested sources of variability which come from different levels of the hierarchy. When the variance of the residual errors is correlated between individual observations as a result of these nested structures, the effect of clustering should be accounted for during the data analysis stage. In this section, a multilevel logistic regression model with shared frailty effect is presented.

Consider the dataset and notation defined in section 3.6.1. The dataset is grouped into n_i clusters. To account for this clustering effect, a random effect, u_i, for $i = 1, ..., I$, is added to the logistic regression model (3.64) to create a 2-level hierarchical logistic regression model with shared frailty, as shown below:

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \sum_{k=0}^{K} x_{ik}\beta_k + u_i z_i, \tag{3.74}$$

for $i = 1, 2, ..., I$ and where u_i is the estimated vector of random effects (frailties) and Z is the design matrix where its Z_i element is equal to 1 for

the i^{th} subject belonging to the same frailty group i or 0 otherwise. u_i is assumed to follow a normal distribution with expected value equal to zero and variance that is estimated from the fitted data and denoted by σ_i^2. Under the shared frailty set up, all subjects within the same cluster are assumed to be at the same level of frailty due to clustering. In the field of epidemiology, grouping patients via their GP practice commonly assumes that effect of clustering is similar for all patients enrolled with the same GP practice as they are exposed to the same level of unknown risks. In this research, GP practice is used as a cluster and included as a random effect for mortality analysis.

Estimation of random effects

The most common methods employed for the estimation of shared random effects in a multi-level logistic regression model are the marginal quasi-likelihood (MQL) method (Goldstein and Rasbash (1996), Snijders (2011)) and the Penalized Quasi Likelihood (PQL) method (Laird (1978), Breslow and Clayton (1993)). The MQL method proceeds by linearising the model (3.74) via Taylor series expansion about the entire predicted value for the i^{th} cluster, such that the following penalised quasi-likelihood algorithm is derived from the first s terms of the Taylor series expansion:

$$Y_i^{(s)} = \beta_{ik}^T X_{ik}^T + Z_i^T u_i, \tag{3.75}$$

where Y_i is a column vector of length I with its elements, Y_i, denoting the number of occurrences of the event of interest, s denotes the order at which the Taylor series expansion has been restricted, X_{ik} is the set of k observed covariates for the i^{th} cluster, β_{ik}^T are the estimated effect due to the k^{th} covariate observed in cluster i, Z_i is the model matrix for the random effect due to cluster i, u_i is a vector of the random effect associated with the i^{th} cluster and Y_i is penalised using the following expression:

$$Y_i^{(s)} = \frac{Y_i - \pi_i^{(s)}}{\omega_i} + \ln\left(\frac{\pi_i}{1-\pi_i}\right), \tag{3.76}$$

where ω_i is the penalty term used to penalise the likelihood function of model (3.74) and is assumed to be equal to $\pi_i(1-\pi_i)$. Using the PQL algorithm defined by equation (3.75), estimated frailty terms, $\hat{u}_i$, are obtained by the following expression (Snijders (2011)

$$\hat{u}_i^{(s)} = (Z_i^T W_i^{(s)} Z_i)^{-1} Z_i^T W_i^{(s)} (Y_i^{(s)} - X_i \hat{\beta}_i^{(s)}), \tag{3.77}$$

where $W_i^{(s)} = \text{diag}\left\{\omega_i^{(s)}, ..., \omega_{n_i}^{(s)}\right\}$. $\hat{u}_i^{(s)}$ is interpreted as the combined effect of all omitted subject-specific covariates that causes some subjects in cluster i to be more frail to the event of interest than other subjects.

3.6.2 Hypothesis Testing and Confidence Intervals for Estimated Parameters

To test the statistical significance of covariates in the multiple logistic regression model, the likelihood ratio test is used (Hosmer Jr et al. (2013a)). Let $\beta^{(0)}$ be the hypothetical value of the coefficients and $\hat{\beta}$ be the estimated coefficients from the fitted model. To test the global null hypothesis $H_0 : \hat{\beta} = \beta^{(0)}$, the test statistic, denoted by LRT, for the likelihood ratio test is given by

$$\text{LRT} = 2(\ell(\hat{\beta}) - \ell(\beta^{(0)}), \tag{3.78}$$

i.e, twice the difference of the log partial likelihood function, evaluated at $\beta^{(0)}$ and denoted by $\ell(\beta^{(0)})$, and that of the multiple logistic regression model with K covariates, evaluated at $\hat{\beta}^{(0)}$ and denoted by $\ell(\hat{\beta})$. LRT follows a χ^2 distribution with K degrees of freedom.

To check the statistical significance of a specific covariate x_k, the Wald test is used (Hosmer Jr et al. (2013a)). The Wald test assess the hypothesis that the estimated coefficient for covariate x_k, denoted by $\hat{\beta}_k$, is significantly different from zero. The test statistic, Z, is given by

$$Z = \frac{\hat{\beta}_k}{\widehat{\text{se}}(\hat{\beta}_k)}, \tag{3.79}$$

where $\widehat{\text{se}}(\hat{\beta}_k)$ is the standard error associated with $\hat{\beta}_k$. The test statistic Z follows a standard normal distribution and can also be used to formulate confidence interval for $\hat{\beta}_k$. The $(1-\alpha)\%$ confidence interval (CI) for $\hat{\beta}_k$ is given by

$$\text{CI}(\hat{\beta}_k) = \hat{\beta}_k \pm Z_{1-\alpha/2}\widehat{se}(\hat{\beta}_k), \tag{3.80}$$

where Z_α represents the critical value at $(1-\alpha)\%$ level from the standard normal distribution. When the $(1-\alpha)\%$ CI of an estimated coefficient contains zero, it means that the covariate associated with that estimated coefficient does not significantly affect odds of occurrence of the event of interest being investigated.

3.6.3 Model diagnostics and assessing model fit

In this section, a review of several methods for assessing the fit of a logistic regression model is presented. To begin, the same description and notation defined in section 3.6.1 is used to explain the different methods of assessing the fit of a logistic regression model. Let the fitted value from the logistic regression model, shown by equation (3.74) be denoted by $\hat{y}_i = \hat{\pi}$. $\hat{\pi}$ is the estimated probability that $y_i = 1$ for the i^{th} observation (defined by equation(3.67)). Model fit assessment for logistic regression model relies on the analysis of two forms of the error component of a fitted logistic regression model, namely deviance (D) and Pearson chi-square statistic (Hosmer and Lemeshow (2000)).

The deviance, D, is given by

$$D = \sum_{i=1}^{N} d_i^2, \tag{3.81}$$

where D follows a chi-square distribution with $(N-k-1)$ degrees of freedom for N number of observations in the study population and k number of covariates in the fitted model, respectively. The individual components (d_i), known as the deviance residuals, are defined as:

$$\begin{aligned} d_i &= \left(2|\ln(\hat{\pi}_i)|\right)^{1/2}, \text{ if } y_i = 1, \\ d_i &= \left(2|\ln(1-\hat{\pi}_i)|\right)^{1/2}, \text{ if } y_i = 0, \end{aligned} \tag{3.82}$$

respectively. The Pearson chi-square statistic, χ^2, is defined as

$$\chi^2 = \sum_{i=0}^{n} r_i^2, \tag{3.83}$$

where χ^2 follows a chi-square distribution with $(N-k-1)$ degrees of freedom for N subjects in the study population and k covariates in the fitted model, respectively, and the individual components (r_i), known as the Pearson residuals, are defined as:

$$r_i = \frac{y_i - \hat{\pi}_i}{\left(\hat{\pi}_i(1-\hat{\pi}_i)\right)^{1/2}}. \tag{3.84}$$

Plots of the r_i and d_i, respectively, against the subject show which subject has relatively too low or too high values for r_i and d_i. A logistic model with subjects having too high value or too low value for r_i and d_i, respectively, indicates a poorly fitted model for these subjects. If the removal of these poorly fitted subjects affect the estimated model parameters, then these subjects are assumed to be influential observations. One drawback in using r_i and d_i to assess the goodness of fit of a logistic regression model is that they can only be analysed graphically and their aggregate statistics χ^2 and D, respectively, cannot be easily interpreted because their distribution under the hypothesis of the fitted logistic regression, cannot be approximated by a chi-square distribution (Hosmer et al. (1991)).

An alternative approach to assess the goodness of fit for the logistic regression model is the Hosmer and Lemeshow test (Hosmer Jr et al. (2013b)), which is carried out via the following steps described below:

(1) The observations are firstly split into j groups, based on deciles of the estimated probability of occurrence of the event of interest. The common cutting points for expected risks in logistic regression are the $10^{th}, 20^{th}, ..., 100^{th}$ percentiles.

(2) In each group, the sum of the probabilities of success is computed. This sum is equal to the expected number of events within each decile.

(3) The differences in observed and expected number of events within each decile are statistically compared using the Pearson goodness of fit statistic given by (Hosmer and Lemesbow (1980))

$$\chi^2_P = \sum_{k=0}^{1} \sum_{l=1}^{j} \frac{(O_{kl} - E_{kl})^2}{E_{kl}}, \tag{3.85}$$

where χ^2_P follows a chi-square distribution with $(j-2)$ degrees of freedom for a correctly fitted model, n is the number of subjects in the investigation, j is the number of groupings used in the test, O_{kl} represents the number of observed events of interest in the l^{th} group and E_{kl} represents the number of expected event of interest in the l^{th} group, with $k = 1$ for occurrence of the event of interest and $k = 0$ for non-occurrence. If the p-value of the χ^2_P statistic is significant, then the

null hypothesis that the data differ significantly from the fitted model, is accepted and thus, the model is a poor fit for the dataset.

One common criticism associated with the Hosmer and Lemeshow test is the choice of the number of clusters, j, to group the observations. Small values of j provides less opportunity to detect misspecification in the fitted model. Large values of j will group the observations into subsets of small size that makes determination of differences between observed and expected successes and failures due to chance or model misspecification, difficult. Hosmer Jr et al. (2013b) proposed to select the number of groupings based on using the rule $j > (K + 1)$, where K is the number of covariates in the fitted model.

3.7 Model Selection Procedures

Model selection in regression analysis is an essential step in obtaining the optimal (best) model. The best model from a regression analysis is one where its prediction error is minimised. Harrell (2001) defines the prediction error as a measure of the difference between the observed and predicted outcomes of a model. Estimating the prediction error takes into account both the bias between observed and predicted outcomes and the variance of the predicted outcomes. A model with minimised prediction error is regarded as the optimal regression model (M_{optimal}). A model that consists of more variables than M_{optimal}, over-fits the data as it includes covariates that do not contribute for prediction of outcomes. Therefore an over-fitted model will have low bias and an inflated variance of the predicted outcomes. A model with less variables than M_{optimal} under-fits the data and excludes covariates that are essential in predicting outcomes. Hence an under-fitted model does not capture the trend displayed by the data.

To find M_{optimal}, three methods of covariate selection for regression analysis are routinely used (Harrell (2001)). These methods are the forward elimination, backward elimination and stepwise selection. They rely on a mathematical information criteria that are used for estimation of the prediction error. For selection of covariates in regression analysis, a model with the lowest information criterion value, will be chosen over models with higher information criterion value. The two most common information criteria used for model selection are the Akaike information criterion (AIC) and

the Bayesian information criterion (BIC) (Harrell (2001)). AIC and BIC are respectively given by

$$\begin{aligned} AIC &= -2\ell - 2k, \\ BIC &= -2\ell + k\ln(n), \end{aligned} \tag{3.86}$$

where ℓ is the partial log-likelihood function of the regression model, k is the number of estimated parameters in the regression model and n is the number of events in the survival analysis. BIC penalizes larger models more heavily as it depends on the number of events, n and tends to perform better for smaller sample size model fitting in comparison to AIC.

Forward selection starts with an empty model. Covariates are added one at a time starting with the covariate with the highest correlation with the dependent variable to create model m_0. Variables of greater theoretical importance are usually entered first. Once selected, the covariate remains in the model and the variable with the second highest correlation with the dependent variable is entered into m_0 to obtain model m_1. If the estimated AIC of m_1 is smaller than that of m_0, then model m_1 is statistically more optimal than m_0. This process is repeated until the contributions of all remaining covariates are checked, one by one, using AIC values. Under the backward elimination method, all the covariates are entered into a full model (m_{all}) and deleted sequentially, one at a time, if they do not contribute to the regression model. If the removal of a covariate increases the AIC of model m_{all}, then it contributes significantly to the optimal model and should be kept in the model. Removal of covariates that decreases the AIC of m_{all} is continued until no decrease in AIC of m_{all} is observed when the next covariate is removed.

3.8 Dealing with Missing data

Introduction

Routinely collected data from clinical databases, such as primary care databases have long been recognised as rich data sources. However they include proportion of incomplete data (missing data). Incomplete dataset is common among observational studies with long follow-up time. Missing data cannot be ignored during data analysis. It is essential to account for the type and

degree of missingness in the dataset.

Datasets used in this research are from primary care records which consist of missing data. MacDonald and Morant (2008) and Hippisley-Cox and Coupland (2010a, 2010b) showed that there is a difference between observed and missing data among medical variables such as records of hypertension, hypercholesterol and chronic diseases. Marston et al. (2010) reported systematic difference between observed and unobserved data among lifestyle variables such as body mass index measurement, smoking and alcohol consumption. According to Shephard et al. (2011), there is a strong relationship between ill-health and primary care records. Subjects who are ill are more likely to have more complete records in primary care database, compared to individuals that are healthier because ill patients are more likely to visit their general practitioners more often than healthier patients. Furthermore, Marston et al. (2010) and Bartley (2016) demonstrated that proportion of missing data is lower among female patients, compared to male patients' primary care records.

There are three types of missing data, namely: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR), respectively. In this section, the methodology of dealing with these types of incomplete information is summarised. To explain these types of missing information, consider a set Y of observations with $k\%$ of missing data from a sample data set of X variables.

3.8.1 Types of missing data

Missing completely at random

Data are classified as missing completely at random (MCAR), when there is no systematic difference between recorded and missing information. Information is MCAR if the probability of missingness on Y is not related to other measured variables, nor to the values of Y itself (Enders, (2010)). Subjects with a full set of recorded information are therefore representative of those with missing information and the opposite is also true. Hence when complete case analysis is carried out on the data set, the estimates obtained from the analysis are not biased. However under such an approach, the analysis ignores subjects with missing information and thus the sample size is smaller

and therefore produces less precise estimates. Enders (2010) proposed the use of difference in mean value of complete case observations and that of missing data observations as a test for the data that are MCAR while Little (2010) suggests the use of likelihood ratio test to check for the MCAR assumption.

Missing at random

Data are missing at random (MAR) when the probability of missing data in Y, is related to some other measured variables in the analysis model but not to the values of Y itself. In other words, no relationship between the propensity for missing data on Y and the values of Y, after partialling out other variables, can be defined (Enders, (2010)). Although MAR means that data are missing in a haphazard way, it also means that a systematic relationship exists between one or more measured variables from the set X and the probability of missing data. One major drawback of the MAR mechanism is that there is no practical way to check if observations are MAR or if probability of missing data on Y is solely dependent on other measured variables from X. MAR mechanism is the most common assumption in the literature associated with missing information among survival data as it is the core rationale behind the use of maximum likelihood and multiple imputation methods to deal with data MAR (Enders, (2010)).

Missing not at random

Data are described as missing not at random (MNAR) if the probability of missing data in variable Y is dependent on Y itself even after adjusting for other variables from X (Enders, (2010)). MNAR data are very common in epidemiological and social sciences studies whereby the probability of missing data is related to the missing data itself. For instance, subjects who do not receive any drug prescription for hypertension, are less likely to visit their general practitioners for blood pressure measurements, compared to subjects who are on drug therapy for hypertension. Hence the propensity of missing values for blood pressure measurement can be higher among individuals who are suffering from hypertension and not on drug therapy.

3.8.2 Missing data handling methodology

Traditional methods

Several methods have been proposed over time for handling incomplete dataset. Presence of missing data affects the analysis of the data when proportion of missingness in the dataset exceeds 5%. Under such a scenario, Spratt et al. (2010) and van Buuren (2012) explained that the following methods of accounting for missing information in the analysis, do not produce unbiased and accurate analysis models:

- Analysis of complete records only (list-wise deletion)
- Excluding variables with missing data
- Grouping observations with missing data as one category
- Single imputation whereby missing data are replaced by an educated or reasonable guess (usually the mean value of the covariate being imputed)

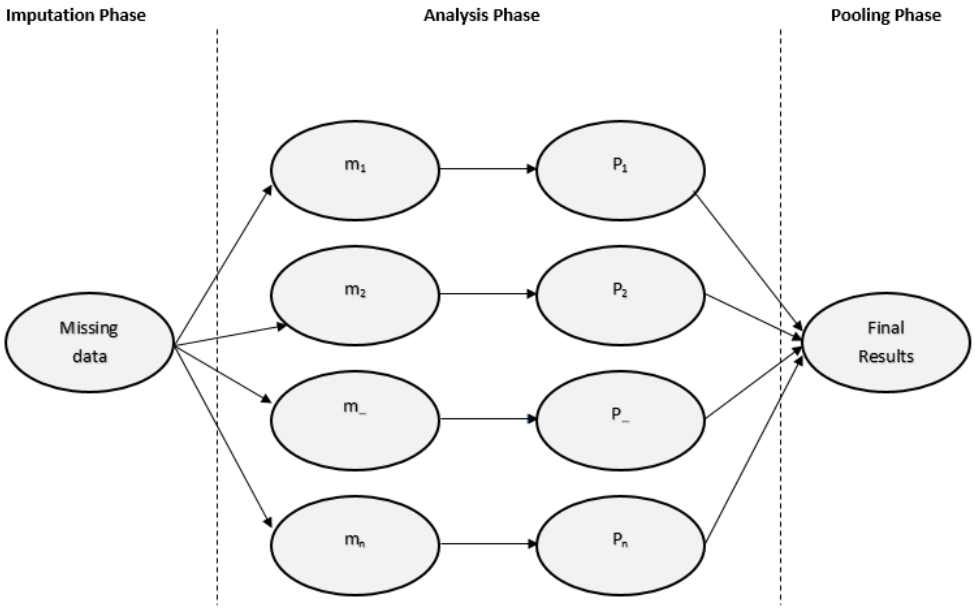
Under the list-wise deletion method, the analysis excludes subjects that have missing information. Hence the sample size of the analysis is reduced and the statistical power of tests involved in the analysis is lowered. This method will also yield biased estimates of parameters involved in the analysis if the records are not missing at random because the analysis is based towards the complete records sample set rather than the whole dataset. Similarly, removing covariates with missing data from the analysis results in estimating predicted outcomes that have not been adjusted for, by the excluded variables. Single imputation method and defining subjects with incomplete records as a separate category during analysis, produce biased estimates because they affect the correlation between covariates that has a level categorised as missing and variables without missing records (Spratt et al. (2010), van Buuren (2012)).

Multiple imputation method

An alternative and widely recommended method to handle missing data is the multiple imputation technique developed by Rubin (1987). Compared to the traditional imputation methods, multiple imputation technique is an analytical method that consists of three distinct stages: the imputation phase,

the analysis phase and the pooling phase, as shown in Figure 5. In this section, all these phases, applied to longitudinal data, are summarised.

Figure 5: Illustration of multiple imputation analysis (Enders (2010))



For n imputations, m_i's ($1 \leq i \leq n$) represent the imputed datasets and P_i's ($1 \leq i \leq n$) are the set of estimated parameters obtained from analysing the m_i^{th} imputed data set.

The imputation phase

The imputation stage is based on a data augmentation two-step procedure. The first step uses an estimate of the mean vector and the covariance matrix of variables with complete records to form a set of regression equations (labelled as the imputation model design (IMD)) that forecast missing data from set of observed variables (referred as the training set). Two key aspects should be considered at this stage: (1) defining the appropriate regression model as the IMD and (2) defining the measurement scale of covariates with incomplete records.

Firstly, van Buuren (2012) showed that the IMD used to impute the missing data, should include all covariates that are used for the data analysis and all variables that are associated with the missing values as well. However it may be impossible to account for all covariates in the imputation model design mainly because of strong correlation between covariates or because of limited computational capability to handle the complex nature of the model. van Buuren (2012) limits the optimal number of covariates in the IMD to 30, although he claimed that having more that 15 variables will hardly influence the explained variance in the imputed datasets.

Secondly, it is essential to specify the measurement scale for each covariates being imputed. According to van Buuren (2012) and Enders (2010), variables that are continuous in nature are imputed using a linear regression model as IMD. Binary variables are imputed via a logistic model, while incomplete categorical covariates with more than 2 levels are imputed using a multinomial regression model. In addition, variables that have been created or derived from incomplete covariates should also be imputed and compared to the imputed covariates used to derive the new variables to ensure consistency between the covariates used during the analysis stage. For a dataset with both continuous and categorical missing variables, joint modelling is used for imputation. Under this approach, the data is assumed to belong to a multivariate distribution (Gaussian distribution). Carrying out multiple imputations under such distributional assumption has been proved to be robust according to van Buuren (2012) and Enders (2010).

The second step of the imputation phase is a Bayesian iterative process using Monte Carlo simulation, as described below (Enders,2010).

(1) Using the regression coefficients of the IMD, values of the covariates with missing values are predicted. A random residual term which is normally distributed with mean zero and variance equal to the residual variance from the regression of the missing covariate value on the outcome variable, is added to the IMD. Adding random residual terms to the mean vector and the covariance matrix of the IMD produces parameter estimates that differ randomly to those that produced the coefficient estimates of the first IMD set up in the first step of the imputation phase. A new dataset (D_{new}) with observed and imputed values is obtained.

(2) Using D_{new}, the new sample means ($\hat{\mu}_{new}$) and the covariance matrix ($\hat{\mathrm{C}}_{new}$) are determined.

(3) Using $\hat{\mu}_{new}$ and $\hat{\mathrm{C}}_{new}$, a new posterior distribution is defined and used to obtain a new set of plausible estimates for the missing values.

(4) Steps (1)–(3) are iterated continuously until convergence of the estimated regression coefficients is achieved. This iteration process of convergence is referred as the burn-in-length.

Once the designated number of burn-in-length has been completed, the entire imputation process is repeated to generate multiple imputed datasets. The observed data stays the same across the imputed datasets; only the values that had originally been missing will differ. Authors like van Buuren (2012) and Enders (2010) indicated that between 5 to 10 imputed datasets is sufficient while other authors such as Graham et al. (2007) and Harel and Zhou (2007) suggests that, depending upon the amount of missing information in the data, increasing the number of imputations to as many as 40 imputed datasets can improve power.

Originally intended for analyzing multiple MCMC chains, $\hat{R}$ is calculated, in the context of multiple imputation, by discarding the burn-in iterations and dividing the single MCMC chain for each parameter into multiple segments (Asparouhov and Muthén (2010)). The $\hat{R}$ statistic then compares the variance within and between imputed datasets in order to detect a potential "drifting" of the chain, that is, regression chains used to simulate missing values that are more variable overall than one would expect, based on the variability within segments. Ideally, $\hat{R}$ should be close to one for all parameters (Gelman and Rubin, 1992). If larger values of $\hat{R}$ occur, a longer burn-in period is required (Enders (2010)). The second option to check convergence of parameters is through diagnostic plots. For each parameter in the imputation model, a trace plot for all iterations during and/or after burn-in is produced. The trace plot is a graphical representation of the MCMC chain for each parameter, and it shows the values of that parameter at each iteration.

The analysis phase

The imputation phase generated n imputed datasets, each of which contains different estimates of missing values. The purpose of the analysis phase is to analyse the imputed data set and involves n statistical analyses such as fitting a Cox regression model to each of the n imputed datasets using the same analysis procedure and software. This will result in obtaining n sets of estimated parameters for each of the imputed datasets.

The pooling phase

The analysis phase produces n sets of unbiased estimated parameters, $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2, ..., \hat{\alpha}_n)$, for MAR data. In the pooling phase, multiple imputation analysis combines the n estimated parameters into a single point estimate, $\bar{\alpha}$, using Rubin (1987) definition of multiple imputation point estimate (Enders, (2010))

$$\bar{\alpha} = \frac{1}{n} \sum_{i=1}^{n} \hat{\alpha}_i. \tag{3.87}$$

Rubin (1987) definition of multiple imputation point estimate is analogous to the formula for the sample mean, where the parameter estimates are used as data points to determine the sample mean. Although multiple imputation technique has been developed under a Bayesian framework, the pooled point estimate of the parameters is highly meaningful under a frequentist framework where $\bar{\alpha}$ is the point estimate of a fixed population parameter.

The analysis phase also yields in n estimates of standard errors associated with the parameters estimated for each of the n imputed datasets. Multiple imputation standard errors come from two sources of sampling fluctuation: (1) variance within imputation and (2) between imputation variance. Again Rubin (1987) definition of multiple imputation point estimate (equation 3.87) is applied to pool the n estimated standard errors into a point estimator. However, Rubin (1987) pooling formula (equation 3.87) works on a sampling variance metric. Hence the standard errors are scaled into their variances before applying equation 3.87. To illustrate this, consider n sets of imputed datasets. Let $\hat{\alpha}_i$ $(1 \leq i \leq n)$ be the estimated coefficients of the fitted model to the i^{th} imputed dataset and $\widehat{\text{se}}_i$ be the estimated standard errors associated with each $\hat{\alpha}_i$. Using Rubin (1987) pooling formula, the within imputation

variance of the parameter estimates, $\widehat{\omega}$, across the n imputed datasets, is therefore given by

$$\widehat{\omega} = \frac{1}{n}\sum_{i=1}^{n}\widehat{\text{se}}_i^2 \tag{3.88}$$

The between imputation variance, $\hat{V}_B$, measures the variability of the parameter estimates across the n imputed data sets and is given by

$$\hat{V}_B = \frac{1}{n-1}\sum_{i=1}^{n}(\hat{\alpha}_i - \bar{\alpha})^2 \tag{3.89}$$

Hence the variance of $\bar{\alpha}$, denoted by $\hat{V}_\alpha$, is estimated as (Rubin (1987))

$$\hat{V}_\alpha = \hat{\omega} + \left(1 + \frac{1}{n}\right)\hat{V_B} \tag{3.90}$$

Significance testing and confidence intervals for multiple imputation analysis

In the context of multiple imputation, Rubin (1987) proposed the use of an analogous t-statistic, T, defined below, to test whether the pooled estimate of the model coefficients is significantly different from a hypothesized value, α_0.

$$T = \frac{\bar{\alpha} - \alpha_0}{\sqrt{\hat{V}_\alpha}}, \tag{3.91}$$

where $\bar{\alpha}$ and $\hat{V}_\alpha$ are given by equations 3.87 and 3.90, respectively and T follows a student t-distribution with the number of degrees of freedom, DF, given by (Rubin, (1987))

$$\text{DF} = (n-1)\left(\frac{\hat{\omega} + \hat{V}_B(1+1\ n)}{\hat{V}_B(1+1\ n)}\right)^2, \tag{3.92}$$

where $\hat{\omega}$ and $\hat{V}_B$ are given by equations 3.88 and 3.89, respectively. The $(1-k)\%$ confidence interval for the pooled estimate of the model coefficient, $\bar{\alpha}$, is given by (Enders, (2010))

$$\bar{\alpha} \pm t_{p,1-k/2}\sqrt{V_\alpha}, \tag{3.93}$$

where $t_{p,1-k/2}$ is the critical value of a t distribution with p degrees of freedom and V_α is the pooled variance of $\bar{\alpha}$.

3.9 Chapter Summary

In this chapter, a review of the statistical methods used in the analysis of mortality data, is presented. Parametric models are employed for survival data assuming a particular distribution while non-parametric models are used for the analysis of data with no distributional assumption. In addition, the Cox proportional hazards model presented in this chapter can be extended to accommodate time-dependent, stratified variables and frailties. However when proportional hazards assumption fails, multiple logistic regression model with random effects can be used as an alternative approach for estimation of mortality risk.

4 Review Of Study Design And Primary Care Data In The United Kingdom

Introduction

This chapter is a review of the different study designs that are used in the field of epidemiology and sources of medical primary care data in the UK. Firstly a summary of the description of different study designs are provided, followed secondly by a discussion on the methods of data collection in these study designs. Thirdly a summarised description of existing primary care records database in the UK is presented. Fourthly an overview of the THIN database then follows.

4.1 Review of study designs in epidemiology

Different study designs provide information of distinct quality. Using the best possible study design is always desirable but not always practical or ethically acceptable. Therefore it is essential to evaluate the strengths and drawbacks of each type of study design, as applied to the research purpose. In the field of epidemiology, study designs are classified into experimental studies and observational studies as shown by Figure 6 below.

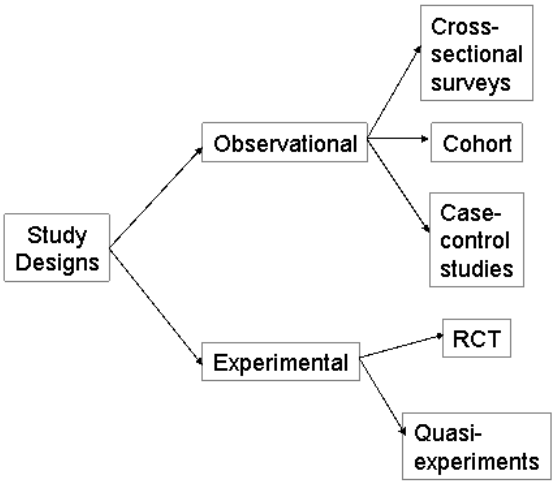


Figure 6: Types of observational study design (Song and Chung (2010))

Observational studies

Observational studies are a sub-class of analytic study design that identify and assess causes, risk factors or outcomes associated with health related events such as cardiovascular diseases or surgical interventions like total hip replacement. The researcher in an observational study only observes and systematically collects information, but does not intervene to alter the subjects.

Case-control study design was introduced by Lane-Claypon et al. (1926) and its mechanism was fully described and used by Doll and Hill (1950). Under the case-control study design, displayed by Figure 7, subjects are admitted into the study by their outcome status at the start of the investigation period. Outcomes of interest may include death of subjects who underwent a surgical intervention or experienced a medical complication or diagnosed with a disease. Upon the identification of subjects with the outcome of interest, they are grouped as cases. Subjects without the outcome of interest are selected from the same population source and are grouped as controls. Retrospective data about exposure to hypothetical risk factors are then collected for cases and controls, usually by interviewing the subjects or by extraction of data from medical database or via surveys.

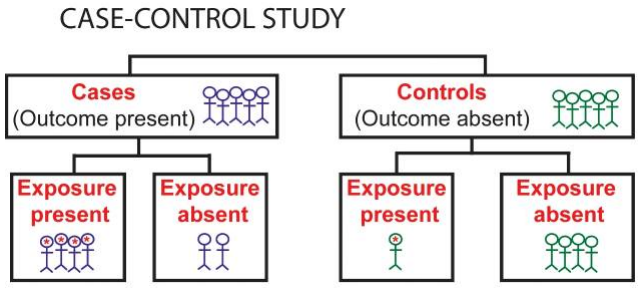


Figure 7: Case-control study design (Song and Chung (2010))

Case-control studies are efficient and economical because (1) such design permits the study of several risk factors simultaneously; (2) it permits the investigation of risk factors for rare conditions whereby there may be problems in generating a sufficiently large number of subjects with the rare conditions to produce accurate results and (3) it allows evaluation of confounding and interacting covariates between cases and controls due to the balanced na-

ture of the study design. However case-control studies are limited because of the following disadvantages: (1) case-control study does not involve a time-sequence and hence does not demonstrate causality; (2) such study design can only investigate one condition or surgical intervention at a time because selection of cases and controls is defined according to the condition or surgical intervention being investigated; (3) case-control study can only estimate relative risks between cases and controls and (5) case-control study design may be strongly biased if selection of controls is not carried out appropriately (Song and Chung,(2010)).

Based on the design of a case-control study, these four methodological aspects should be addressed carefully before the investigation (Song and Chung,(2010)):

(1) Selection of subjects as cases is dependent on the outcome of interest and hence it is important to explicitly define the inclusion and exclusion criteria for acceptance of subjects as cases. In addition, validity of the data source for cases' identification need to be carefully carried out, such that cases are representative of the target population.

(2) An essential principle during the selection of controls is that distribution of exposure should be similar among cases and controls. This can be ensured by selecting cases and controls from the same population source and adopting the same inclusion criteria for both.

(3) Matching cases to one or more controls based on their background variables is a key aspect of a case-control study design. This process establishes comparability between cases and controls and reduces variability and systematic differences that are insignificant to the investigator.

In a cohort study, an outcome-free study population is first identified by the exposure or event of interest and followed in time until the disease or outcome of interest occurs. For example, in survival analysis after a particular surgery, individuals who underwent the surgical procedure are followed from the time of surgery until the outcome of interest (death), if it occurs before the end of the investigation period. Since exposure is identified before the outcome, cohort studies have a temporal framework to assess causality and thus have the potential to provide the strongest scientific evidence (Everitt and Palmer, (2011)). Cohort study designs are beneficial because

firstly they permit the investigation of causality since the data being analysed take account of the sequence of outcome of interest and secondly they allow examination of multiple outcomes for a given exposure. However they are often criticised for requiring large sample size to produce accurate estimates.

Case-control and cohort studies can be prospective or retrospective. Prospective studies are carried out from the present time into the future. Since prospective studies are designed with specific data collection methods, they have the advantage of being tailored to collect specific exposure data and may be more complete. The disadvantage of prospective studies is the long follow-up period while waiting for outcome of interest to occur. Thus, this study design is inefficient for investigating diseases or conditions with long latency periods and is vulnerable to a high loss to follow-up rate.

Retrospective studies, also known as historical cohort studies, are carried out at the present time and look to the past to examine medical events or outcomes of interest. In other words, a cohort of subjects selected based on exposure status is chosen at the present time, and outcome data (i.e. disease status, event status), which was measured in the past, are reconstructed for analysis. The primary disadvantage of this study design is the limited control that the investigator has over data collection. The existing data may be incomplete, inaccurate, or inconsistently measured between subjects (Hulley et al. (2013)). However with the immediate availability of the data, retrospective study design is comparatively less costly and shorter than prospective studies.

Cross-sectional studies, also known as prevalence studies, analyse the data on disease and exposure at one particular time point only. Since the temporal relationship between disease occurrence and exposure cannot be established, cross-sectional studies cannot assess the cause and effect relationship. Cross-sectional study is one of the most efficient and economical designs because: (1) it is relatively quick and easy to conduct (no long periods of follow-up are required); (2) data on all variables are collected only once; (3) it measures prevalence for all factors under investigation; (4) it allows studies of multiple outcomes and exposures simultaneously. However cross-sectional studies are criticised because firstly it is difficult to determine whether the outcome followed exposure in time or exposure resulted from the outcome under this study design and secondly cross-sectional studies are not suitable

for studying rare outcomes or outcomes with a short duration.

Experimental studies

Under the experimental study set-up, the investigator intervenes during the experiment through a series of actions or decisions and then observes what happens to the subjects (Hulley et al. (2013)). Experimental studies are often performed in laboratories and in clinics to establish beneficial effects of drugs or procedures. Such a study design is considered to provide the most reliable evidence in epidemiological research. Experimental studies are subdivided into either preventative or therapeutic studies (Hulley et al. (2013)). Therapeutic study trials are conducted among individuals with a particular disease to assess the effectiveness of an agent or procedure to diminish symptoms, prevent recurrence, or reduce mortality from the disease. Preventative study trials are conducted to evaluate whether an agent or procedure reduces the risk of developing a particular disease among individuals free from that disease at the beginning of the trial.

A main characteristic of all experimental studies is that the therapeutic or preventative intervention being tested, is allocated by the investigator to a group of two or more study subjects which are then followed prospectively to compare the group of individuals with intervention to the control individuals who did not receive the intervention. There are two main types of experimental studies, namely cross-over trials and randomised controlled trials (Hulley et al. (2013)). A cross-over trial is one in which the subjects are first assigned to the treatment group and, after a brief interval for cessation of residual effect of the treatment intervention, are shifted into the control group. Thus, the subjects act as their own control at the end of the study. However, such studies are not feasible if there is loss of follow-up due to mortality, or if the disease is easily cured by one of the interventions.

The randomised controlled trial (RCT) is considered to be the most rigorous method of assessing whether a cause-effect relationship exists between an intervention and outcome. The strength of the RCT design lies in the process of randomisation that is unique to this type of epidemiological study design. Generally, study participants are randomly assigned to one of two groups: the experimental group receiving the intervention that is being tested and a comparison group (controls) which receives a conventional treatment. These

groups are then followed prospectively to assess the effectiveness of the intervention compared with the standard treatment. The random allocation of subjects is used to ensure that the intervention and control groups are similar in all respects (distribution of potential confounding factors are similar) with the exception of the therapeutic or preventative treatment being tested and to ascertain that any observed differences between the treatment groups are due to differences in the treatment alone.

4.2 Sources of medical data

In the UK, sources of medical data can be segregated into firstly, prospectively collected trial-cohort data and secondly, routinely collected data from primary and secondary healthcare, disease specific registers or mortality registers. In this section, the contrast in data collection between these two sources of medical data is discussed. They are compared in terms of determination of outcome of interest, precision of risk factors, proportion of missing values, all-cause mortality information, cost of data collection and the generalisability of the data.

Random trial controls (RCTs) and cohort studies are designed specifically to prospectively collect data whereby the outcome of interest is explicitly specified and recorded together with exposures and risk factors. With routinely collected data, patients' information is recorded only when they visit their primary or secondary healthcare services (MacDonald and Morant (2008), Wijlaars (2013)). Hence the amount of records for each patient in routinely collected data highly depends on the frequency of visits by the patients and also what information the healthcare professionals identify as relevant to be recorded. This may result in a proportion of patients with unknown records for different risk factors and outcomes for the periods they did not visit their primary or secondary healthcare. Thus in contrast to prospectively collected trial-cohort data, routine data collection may not contain complete records of outcomes or risk factors.

RCTs and cohort studies are carried out under rigid study protocols to ensure consistency and high precision during the data collection for outcomes and risk factors measurements. On the contrary, routinely collected data are recorded in coded form by clinicians. Thus there is a risk that different clinicians may use different methods of coding for patients' records or may even

employ the incorrect method to record information (MacDonald and Morant (2008), Hippisley-Cox and Coupland (2010a). In the UK, to ensure that primary and secondary care records are consistent and of high quality across all primary and secondary healthcare providers such as general practitioners, the Quality and Outcomes Framework (QOF) was introduced in 2004, as a scheme that pays primary and secondary healthcare to improve their services (Szatkowski et al. (2012)). In addition, published list of codes for classification of medical conditions and treatments in UK are also available as an online repository (See ClinicalCodes (2016)) where researchers can upload and download lists of clinical codes that enable researchers to better validate their studies, to build on previous code lists and to compare medical conditions' definitions across several published studies.

Information related to mortality is more consistent and readily available in routinely collected data, compared to RCT collected data. Primary and secondary care databases in UK have high precision and consistent all-cause mortality information because when an individual dies, his general practitioner is notified (HSCIC (2016)). Therefore, precise information on cause and time of death can be obtained from routinely collected data. Furthermore, RCT collected data and secondary healthcare data hold medical information that are specific to one particular medical condition only and cannot be used to test interaction of new risk factors with other medical conditions. On the other hand, routinely collected data such as primary healthcare records, consist of comprehensive medical history of individuals that are not specific to one condition only, thereby permits exploration and analysis of new factors that are not measured in RCT's or secondary healthcare databases.

Collecting data prospectively is expensive because of the ongoing costs involved during the investigation period and this may put financial limitation on the amount of data being recorded leading to the sample size not being optimal for analysis. On the other hand, routine data collection is associated with lower cost as it only involves an initial high cost for setting up a data collection system and training professional healthcare providers to use the system and a low cost of system maintenance to validate data collection and to provide a quality assessment of the data. Being less costly, routinely collected databases have large sample size and represent a rich source of long follow-up data with frequently updated information that can be used to pro-

duce the most recent statistics (Wijlaars (2013)).

Analysis of either prospectively collected trial-cohort or secondary care or disease specific data may not be generalisable to a population as a whole because such data are relatively small in size and are restricted to specific conditions that are specified by rigid study protocols. For example, analysis of secondary care data and disease specific registries can only be applied to individuals with the conditions being recorded in these databases. On the contrary, analysis of routinely collected data such as primary care data can be easily generalised to a population as a whole. In the UK, 99% of the population is registered to a general practitioner under the National Health Service (NHS (2013)) and all visits or admissions to secondary care providers are also recorded by the patients' general practitioners in primary care records (Hall (2009)). Therefore primary care data in the UK is representative of all patients with both mild and severe medical conditions, providing an almost complete medical history of the patients. However one cannot fully ascertain that primary care records reflect 100% of a patient's medical history because information such as self-medication of over the counter drugs may not be reported to the primary health care provider.

4.3 Overview of primary care databases in the United Kingdom

In UK, the National Health Service (NHS) is organised around primary care and, unless there is an accident or emergency, whenever citizens would like to use the NHS they have to go through their primary care physician, also referred as a general practitioner (GP) in the UK. From there, they can be referred to a specialist at a hospital if required. Secondary care clinicians can then feedback information to GPs. Since the vast majority of the population (99%) is registered with a general practice (NHS (2013)), GP's act not only as the main gatekeepers for the NHS but also as essential providers of a longitudinal electronic health records (Herrett et al. (2015)). There are now many ongoing primary care databases of anonymised patient records in UK that can be used for healthcare research. These population-based databases contain data originating from routine general practice. In this section an overview of the three largest primary care databases in the UK, namely the Clinical Practice Research Datalink (CPRD) (Walley and Mantgani (1997)),

The Health Improvement Network (THIN) (Bourke et al. (2004)) and QResearch (Hippisley-Cox et al. (2004)), respectively, is provided.

CPRD, THIN and QResearch databases are often used for cross-sectional surveys, case–control or cohort studies and for epidemiological, drug safety, clinical and healthcare usage research purposes. They rely heavily on individual general practices voluntarily contributing data via the propriety clinical systems they use to maintain these patient records. CPRD and THIN include medical records from approximately 600 practices, input using the Vision clinical system, while QResearch database consists of medical records from approximately 1000 practices that employ the EMIS clinical system for data collection. The records are usually anonymised at source by allocating a unique number to each patient to allow for the updates of the records and their linkage to other data sets, such as national mortality, national cancer registration and hospital records as well as with socio-economic, ethnicity and environmental data sets. Access to these data sets is usually granted after scientific and ethics review and can be tailored to customer requirements.

Table 6 summarises the main characteristics of the CPRD, THIN and QResearch databases and three other databases (Quality and Outcomes Framework, General Household Survey and Health Survey England) that contribute to CPRD , THIN and QResearch data collection and linkage. The CPRD (previously known as the General Practice Research Database) is a not-for-profit research service funded by the NHS National Institute for Health Research (NIHR) and the Medicines and Healthcare products Regulatory Agency (MHRA). It is owned by the UK Department of Health and consists of records for 11 million patients (4.4 million active) from 674 practices (Herrett et al. (2015)). The THIN database contains records of 12 million patients from approximately 600 general practices that use the Vision clinical system, set up by In Practice Systems (INPS) and Epidemiology and Pharmacology Information Core (EPIC) (IMS Health Incorporated (2017)). QResearch is a research service located at the University of Nottingham and its database consists of the health records of 18 million patients from 1000 general practices that use the EMIS clinical system (QResearch (2016)).

The key strengths of these databases lie in their size, representativeness of the UK population, long follow-up periods of patients and high data quality (Herrett et al. (2015)). They contain important and quality information on

morbidity and lifestyle, prescriptions, preventive care, current standards of care and inter-practice variations (Gnani and Majeed (2006)). Since they are continually updated, they are ideally set up for researchers to investigate and monitor healthcare trends and effectiveness of new interventions and treatments, with minimum cost. They are increasingly linked to secondary care and mortality data sets. However, their weaknesses include the fact that data are extracted from proprietary clinical systems developed for patient management and not for healthcare research. There are issues such as: (1) missing data (for example healthier patients are more likely to have incomplete records than less healthy patients who visit their GP more frequently), (2) variable definitions for diagnoses (although this is improved by published list of clinical diagnoses provided by ClinicalCodes (2016)), (3) incomplete secondary care data (for instance, incomplete records from hospital admissions) and (4) incomplete capture of wider health data such as treatment adherence or over the counter medication (Herrett et al. (2015)).

Nevertheless, CPRD, THIN and QResearch are highly regarded within the research community since they strongly support researchers obtaining definitive answers for various healthcare debates of considerable public interest. Validity of new or updated primary care records added to these databases in UK, is checked through external validation by comparing new records to existing ones. Although these three databases consist of different practices and patients, estimates of incidence, prevalence, morbidity and mortality rates are similar across all three databases when data are adjusted for gender, age and level of social deprivation of patients' residential areas (NHS (2013)). Therefore primary care records databases in the UK are good source of valid data that can be used to answer research questions under well-defined study designs.

Database	QResearch	Clinical Practice Research	The health Improvement Network	Quality and Outcome Framework	General Household Survey	Health Survey England
Abbreviation	QResearch	CPRD	THIN	QOF	GHS	HSE
Type	Primary Care Database	Primary Care				
Database	Primary Care					
Database	Incentive Payment					
Scheme	Repeated cross-sectional					
survey	Repeated cross-sectional					
survey						
Study Period	1993-Ongoing	1987-ongoing	1987-ongoing	2004-ongoing	1971-2011	1997-ongoing
Study population	Volunteering general practices using EMIS clinical system in United Kingdom	Volunteering general practices using EMIS clinical system United Kingdom	Volunteering general practices using EMIS clinical system in United Kingdom	Volunteering general practices in England	Probability stratified two-stage sample design in United Kingdom	Multi-stage stratified random sample in England
Sampe Size	Over 18 million patients from over 1000 general practices (in 2015)	Over 11 million patients from 674 general practice (in 2015)	Over 12 million patients from 587 general practices (in 2015)	Over 56 million patients from nearly 8000 general practices (in 2015)	18,367 individuals aged 16+ from 7,937 households (in 2015)	10,080 individuals (in 2014)
Data Collection	Electronic medical records	Electronic medical records	Electronic medical records	Electronic medical records	Annual telephone and face to face interview collecting data regarding education, employment and labour, health, housing, social indicators and quality of life, use and provision of specified social services	Face to face interview collecting data physical health, mental health and well-being, social care, lifestyle behaviours and physical measures
Reference	Hippisley-Cox et al. (2004)	Herett et al. (2015	Wijlaars (2013a)	Gillam et al. (2013)	ONS UK (2016)	HSCIC (2016)

Table 6: Primary care databases and national surveys in United Kingdom (Gitsels, 2017)

4.4 The Health Improvement Network database

The Health Improvement Network (THIN) database is used in this research for data extraction to investigate variations in mortality risks after THR procedures in the UK. In this section, the THIN database structure is described and a discussion of the key strengths of using THIN for the objectives of this research is provided.

4.4.1 THIN database structure

THIN is an electronic medical research database of anonymised patient records from approximately 600 different GP practices across the United Kingdom. The database represents almost 6% of the GP practices population in the UK. It is also equivalent to roughly 86 million patient-years of data. It was created by In Practice Systems Ltd (INPS) in collaboration with Epidemiology and Pharmacology Information Core (EPIC) in 2003 (IMS Health Incorporated, (2015)). The INPS data system collection is essentially similar to other existing operating clinical data recording systems such as EMIS for QResearch and CPRD databases (See section 4.3) in the UK (Department of Health, (2011)). EPIC is responsible for the data collection, for the quality of non-clinical information and for anonymising the records of each patient. It also connects medical data of patients to their post-code and environment related indicators (THIN Data Guide, (2011)).

The database is dynamic, in the sense that data is continuously collected and updated and patients may join and leave the database at different times. THIN data comes from routine health data that are collected directly and at regular intervals from the management software of GP practices using a modem arrangement which does not interrupt the programmes running and requires no human intervention. These data collections are then processed to provide coded longitudinal records of demographic details, lifestyle characteristics, medical events, treatment prescriptions, specialist referrals, and any diagnostic or laboratory results occurring at an individual patient level. These records also include information on various socio-economic markers and environmental variables (IMS Health Incorporated (2017)).

Raw data from GP practices are anonymised and restructured so that they can be manipulated in a simplified and flexible way. The structure of

THIN database is defined by seven American Standard Code for Information Interchange (ASCII) standardised files (Wijlaars, (2013)), as illustrated by Figure 8 below.

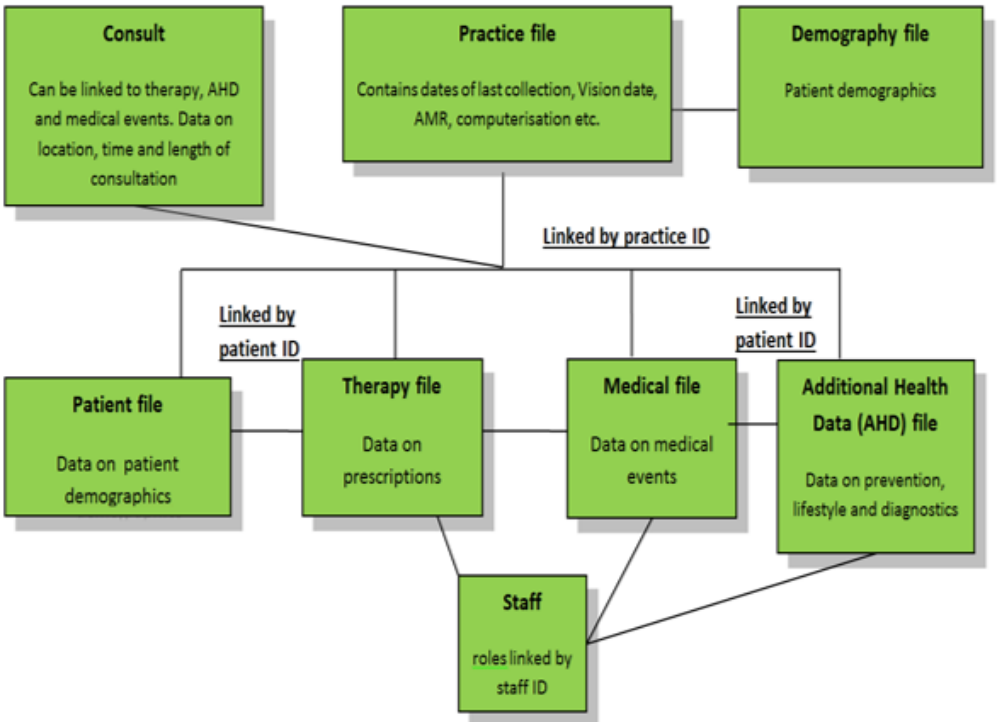


Figure 8: Structure of the THIN database (THIN Data Guide (2011))

(1) Patient file provides data on age, sex, registration date when entering the practice, and date when leaving the practice.

(2) Medical file consists of medical diagnoses, date of diagnosis, and location of the event and referrals to hospitals and specialists.

(3) Prescription file provides information on all prescriptions along with the date issued, formulation, strength, quantity, and dosing instructions, indication for treatment for all new prescriptions (inferred from cross reference to medical events on the same date), and events leading to withdrawal of a drug or treatment.

(4) Additional Health Data (AHD) file details information on vaccinations and prescription contraceptives; miscellaneous information such as smoking, height, weight, immunizations, pregnancy, birth, death, and laboratory results.

(5) Postcode variables indicator (PVI) file list out postcode linked area based socio-economic, ethnicity and environmental indices.

(6) Consultation file provides details on the date, time and duration of consultation.

(7) Staff file consists of the gender and roles of staff who entered the data.

In THIN, information is coded using hierarchical Read codes, which permits some standardisation of the method used to record information. In the context of a medical database, Read codes are defined as a clinical vocabulary used to store information on diagnostic, symptomatic and procedural data by codes, but translated into text when accessed. In the UK, Read codes were developed in 1982 by Dr James Read, a UK GP, and has become the de facto standard for coding diagnoses, operations, and procedures, signs and symptoms, and for all national minimum data sets and national statistics for the hospital and community health services (NHS Digital (2016)). THIN Read codes are interpreted using ancillary look up tables and dictionaries in which medical events are coded using the Read system and prescriptions of drugs are coded using multi-lexical code alongside a British National formulary code. A large part of the codes are allocated to patients' records by the consulting GP practices or healthcare providers themselves or by the administrative staff such as GP practice managers (THIN Data Guide, (2011)).

4.4.2 Generalisability of the THIN database

The degree of generalisability of patient databases such as THIN, to the general population, is important for interpreting primary care records based research. This section discusses the representativeness of THIN data to the UK population with regards to demographics, prevalence of medical conditions, and mortality rates by reporting published comparisons of THIN dataset to other medical databases in UK.

Blak et al. (2011) investigated the degree of generalisability of THIN patient database to the general population by comparing the THIN data to the

National Statistics data for UK and to that of Quality and Outcomes Framework (QOF) for 2006/2007. The authors concluded that THIN primary care database represents the UK demographically since the distribution of gender and age in THIN database is consistent with the UK population, although THIN consisted of a slightly smaller proportion of patients aged under 25, compared to other data sets used in the study. They also reported that population size of affluent areas is greater than deprived ones. Therefore adjusting for gender, age and level of patients' residential areas deprivation level, strengthen the representativeness and generalisability of estimates of THIN data to the UK population (Blak et al. (2011)).

Langley et al. (2011) assessed the validity of THIN data to monitor regional smoking prevalence in UK by comparing THIN dataset to the General Household Survey (GHS) between 2000 and 2008. The authors reported that THIN prevalence data on patients smoking status, were generally found to be highly comparable with GHS data from 2006 onwards. Loomis et al. (2016) compared prevalence of BMI and its association with prospective risk of nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH) from the THIN database to the Humedica EHR database in the UK. The authors reported that age, gender, smoking status and prevalence of diabetes were broadly similar between the two databases while the average BMI ($\pm$SD) was higher in Humedica EHR database (28.14$\pm$6.43 kg/m2) than in THIN (26.81$\pm$5.57 kg/m2).

Crude prevalence rates for selected medical conditions in THIN are slightly higher than those of QOF 2006/2007 dataset as reported by Blak et al. (2011). Crude prevalence for hypertension, ischaemic heart diseases, chronic kidney diseases and obesity was 0.1%, 0.2%, 0.2% and 0.8%, respectively, higher than prevalences reported in QOF 2006/2007 dataset, while that of diabetes is 0.2% higher than in QOF 2006/2007 dataset. Similarly, González et al. (2009) reported that prevalence of diabetes in THIN for 1995-2005, adjusted for gender and age, was 0.2% higher than in the Health Survey England (HSE) database. Hippisley-Cox and Coupland (2010a) found no difference in gender and age adjusted prevalence rate of chronic kidney diseases in THIN when compared to that of QResearch data between 2002 and 2008.

MacDonald and Morant (2008) compared prevalence of hypercholesterolemia and hypertension in THIN to national rates in 1998, 2003 and 2006, respec-

tively. The authors reported a lower prevalence of hypercholesterolemia and hypertension in THIN compared to the 1998, 2003 and 2006 national rates, respectively, and found that prevalence rates of hypercholesterolemia and hypertension from different databases, converge over time and with increasing age. Lewis et al. (2007) used THIN data from 1986–2003 to conduct case-control studies of associations between diseases and compare the results to that GPRD database for the same period. The authors found significant associations between stroke with hypertension and diabetes mellitus; between myocardial infarction with hypertension, hypercholesterolemia, obesity, and smoking, respectively, and also reported that similar associations were obtained with the GPRD database. Hence they concluded that THIN data that are collected outside of the GPRD, appear as valid as the data collected in GPRD.

Clegg et al. (2016) compared frailty index of ageing by taking account of hazards of death, unplanned hospitalisation and nursing home admissions among elderly people, aged 65–95 from the THIN dataset, to that of ResearchOne primary care database for the same age cohort. The authors reported that proportion of patients with mild frailty index of ageing in THIN is higher by 7% than in ResearchOne primary care database while no significant differences were observed in the frailty index of ageing for different levels of Townsend score of multiple deprivation, between THIN and ResearchOne primary care database, respectively. Blak et al. (2011) reported that mortality rates of THIN patients are of similar magnitude to the UK national death rates when adjusted for demographics and social deprivation indexes. Hall (2009) also investigated the validity of death data in THIN by comparing number of deaths due to suicide in THIN database to that of the General Practice Research Database (GPRD) and concluded that records of death information in THIN are reliably recorded.

In the light of the above discussion, although it is difficult for investigators to obtain complete medical information on each patients in THIN database, in general available information in THIN tends to be valid and precise enough to develop statistical models for THIN population estimates and to extend them to the general population in UK. Patients in THIN are representative of UK population with regards to demographics, prevalence of major medical conditions and mortality rates, adjusted for gender and level of deprivation.

4.5 Chapter Summary

In this chapter, study designs and sources of data used for epidemiological research, are presented and discussed with respect to their strengths and limitations. Additionally a review of primary care databases in UK is carried out and their eligibility for research purposes is discussed. In general, all the primary care databases in UK are comparable to each other when patients are matched on gender and age. For this research, a retrospective cohort matched study design is used to analyse longitudinal data from the THIN primary care database.

5 Data Description

5.1 Data extraction

This section details the step by step procedure involved in identifying relevant patient records from the THIN database, stored at the University of East Anglia data depository, for the purpose of this study. It covers definition of follow-up time for all patients, data manipulation, identification of cases who had a total hip replacement (THR) and the selection of controls to match with cases.

Description of follow-up time

Before leaving the GP practice computer system, patients data are completely anonymised. However, encrypted identifiers, that are unique to each patient and each GP practice are available. This allows to link patients to their GP practices and thus determines the duration of the period they are registered with their GP. In general, patients in THIN are followed from the latest of the following dates: (1) Registration date of the patient with the GP practice; (2) Date on which the practice started to fully use their computer system to record patients' diagnoses and treatment prescriptions (also referred as *Acceptable Computer Usage*, or ACU); and (3) Date by which computer generated records on patients' death become fully valid.

The importance of having ACU records arises from the fact that when a GP practice first starts to use a computerised system to record data instead of traditional paper records, there is a time delay for the computer recording system to be fully adopted by the practice and its staff. Therefore records from this initial period are likely to be incomplete, leading to biases and incorrect inferences from statistical analyses. To avoid such problem, the time point at which a GP practice fully utilises its computer recording system is determined based on empirical evaluation of the quantity of each type of record.

For a GP practice to have an acceptable ACU level, it needs to have all of the following records, for each patient, per patient year (McBride et al. (2010)); (1) an average of at least one medical record, (2) two prescription records, (3) one additional health data record. Dated records of

patients' deaths in the practice were assumed to be at an acceptable ACU level, based on the *"Acceptable Mortality Rate"* definition provided by the database provider (Maguire et al. (2009)). The definition classifies a practice records as ACU if the observed number of deaths within the practice is consistently within 30% of the total expected number of deaths for that period.

In this study, the complete investigation period for patients selected as cases starts on the date they underwent a total hip replacement procedure until July 2011 (date at which the latest ACU records in THIN dataset available at the University of East Anglia, are provided). Other ending points for the investigation period include (1) transfer out of patients to another GP practice and (2) death of patients. When patients are transferred out to either a GP practice within THIN or to another GP practice not registered with THIN, they are lost to follow-up. Thus their transfer out date is taken as the end of their investigation period. Death, which is the main event of interest in this study, also ends the follow-up period of a patient. The time-line in Figure 9 displays the investigation period for cases in this study.

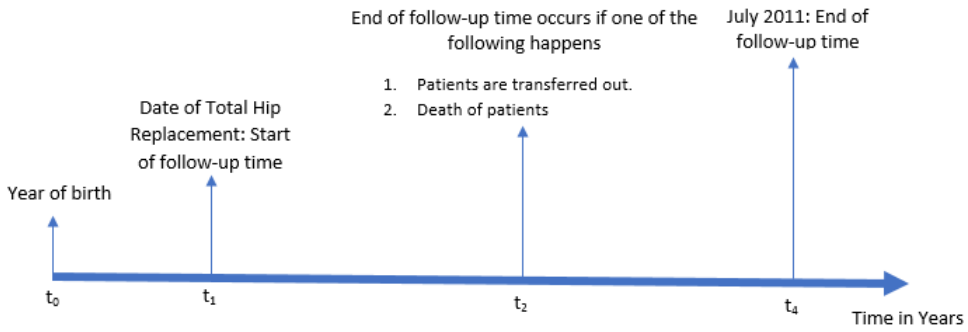


Figure 9: Time-line of follow-up of patients

5.1.1 Identification of Cases

A patient in THIN, born between 1920 and 1940 inclusive, is classified as a case for the purpose of this study if: (1) the patient underwent a *total hip replacement* surgery (THR) while the patient is registered with an active GP practice with ACU type records and valid time of death as described in the previous section; (2) the medical records of the patient have been accessed

at least once within the last ten years before their THR procedure date. Patients in THIN with these criteria are classified as *THR cases*; i.e. these represent individuals who underwent a THR procedure during their lifetime.

Medical events such as THR are coded in THIN by Read codes and thus getting the appropriate Read codes for THR procedure is essential to identify THR cases. From the literature review carried out in Chapter 2, a list of the types of THR procedures was built. It consists of *Cemented, Uncemented, Hybrid and Reverse Hybrid* THR procedures. In the THIN database, these THR procedure types are coded and can be identified by medical Read codes provided by the THIN database provider data dictionary (THIN Data Guide, (2011)). The Read codes for THR procedures selected to identify THR cases for this study, are listed in Table 27 in Appendix A. 40 different Read codes are identified in THIN and can be used to find patients who underwent a THR procedure, while they are registered with a GP practice in THIN. All of these Read codes represent specific types of THR procedures. 21 Read codes correspond to cemented THR procedures, 11 to uncemented THR procedures, 6 are classified as other types of THR procedures and 2 represent hybrid THR procedures, respectively (See Table 27 in Appendix A).

THR cases exclusions in the study

There is a list of patients who underwent THR procedures but were excluded from this study for the following reasons:

(1) This study investigates only unilateral THR procedure. Therefore patients with bilateral THR surgery are not included in this study. Survival analysis of these patients would be different due to the presence of multiple number of surgeries. Bilateral THR procedures are identified using the same Read codes from Table 27 in Appendix A. Patients with bilateral THR procedures can be identified in THIN by looking up the number of THR procedures they underwent. Patients with both hips replaced, either under one surgery or on two separate events, have two medical records of THR procedure and are not included into THR cases in this study.

(2) One of the objective of the study, as discussed in Chapter 1, is to analyse the impact of degenerative and chronic conditions such as arthritis

on survival of THR cases after their THR procedure. Patients who underwent a THR procedure as a consequence of hip fracture are excluded from this study. This is because the causal effect of THR procedure being investigated in this study is different for patients who had a hip fracture. In THIN, patients with hip fracture can be identified using the Read codes listed in Table 27 in Appendix A. These Read codes describe an event of hip fracture or a history of rehabilitation treatment given to the patient after THR surgery due to hip fracture.

The flowchart described in Figure 10 displays the total number of THR procedures identified in THIN database and patients excluded in this study.

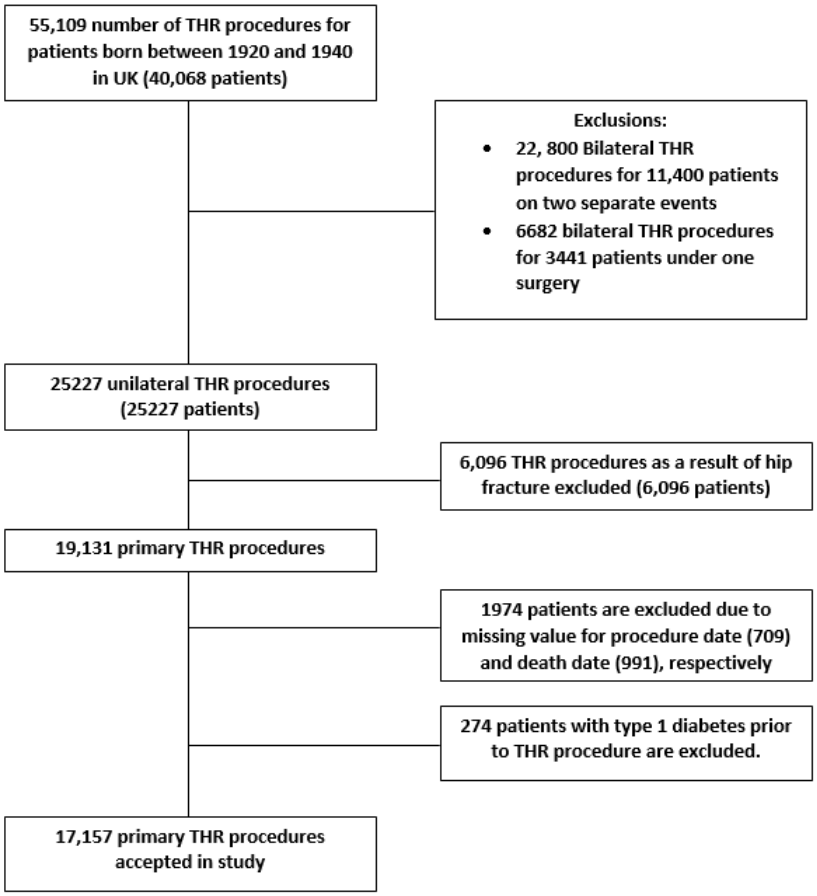


Figure 10: Identification and exclusions of THR cases in study

5.1.2 Identification of controls

Patients in THIN, born between 1920 and 1940 inclusive, are classified as controls in the context of this study if: (1) their medical history is free of any type of THR procedures while they are registered with an active GP practice with acceptable level of ACU patient records; (2) the medical records of the patient have been accessed at least once within the last ten years of registration with the GP practice and (3) the patient is alive at the time of THR surgery of the case with which the patient is matched as control. Initially all patients with these criteria in THIN are grouped in a set called A. Not all patients of set A will be eligible to be selected as controls for THR cases because of the matching criteria, defined below, in the context of this study. The next stage involves a matching process between THR cases and patients from *Set A*, based on three different factors, explained below. To ascribe patients from *Set A* as controls, THR cases are matched to controls on:

- *Sex*: Mortality rate is different for males and females (Shaw (2005)).
- *Year of birth category*: Taking the year of birth category of the patients into consideration will help to account for differences in medical advancements with time (Langholz and Clayton, (1994)). In this research, patients are grouped by their year of birth as follows: Category 1, 2 ,3 and 4, respectively for patients born between 1920–1924, 1925–1929, 1930–1934 and 1935–1940, respectively.
- *GP practice*: Individuals within the same GP practice are similar to each other in many ways; they reside in the same area and receive treatments from the same GP practice. Hence they are likely to be exposed to the same factors and risks and thus they all share similar level of frailty in health (Shaw, (2005)). This is an important aspect that needs to be accounted for during the matching process.

The last step in this matching process is finding the optimal matching ratio between exposed (cases) and unexposed (controls) patients that will maximise the efficiency of the statistical analysis of the data. According to Raboud and Breslow (1989), a matching ratio of 10 is the maximum optimal ratio. Nevertheless the efficiency of statistical analysis does not change significantly for ratios between 5 and 10 inclusive (Raboud and Breslow (1989)).

Therefore 5 controls are assumed to be the optimal number of matching controls per cases as reported by Hennessy et al. (1999) as well. A matching ratio of 5 between THR cases and controls is selected for this study.

The matching process of THR cases to controls in Set A is carried out using a programming language called *Python* which has built-in functions for random selection without replacement. For each patient in the set of THR cases, five active patients of similar gender, GP practice, year of birth category and who are alive at time of surgery of the THR case, are randomly selected without replacement as matched controls. Matching randomly without replacement ensures that each THR case is matched to 5 distinct controls; i.e. the same control cannot be matched to more than one THR case. As an illustrative example, consider a male THR case, with ID "aa01", born between 1920-1924 from GP practice "X11". Based on the above-mentioned matching process, "aa01" will be matched to five different male controls born between 1920-1924 and who are registered with GP practice "X11" at surgery time. Finally these five selected controls should be alive at the time of surgery of case with ID "aa01".

5.2 Data Set Description

This section describes the data set extracted from THIN database for the purpose of this study. The demographic, medical and prescription histories of THR cases and controls are detailed in the following format and order: for each variable being reported for the patients, (1) a definition of the variable in the context of this study, (2) definition and Read codes of the variable in THIN data dictionary and (3) descriptive statistics such as frequency of patients and prevalence of each condition among the study population.

5.2.1 Cases and controls

Number of THR cases and matched controls

$17,157$ patients from the THIN dataset were found to undergo a unilateral THR procedure and satisfy the definition of a THR case as presented in Section 5.1.1. The medical records of these patients include events that are coded by one of the Read codes from Table 27 in Appendix A. Similarly $85,785$ patients from Set A (described in Section 5.1.2) are selected based

on the matching process defined in Section 5.1.2 as controls. Their medical records in THIN dataset do not include Read codes from Table 27 in Appendix A. These 85,785 patients satisfy the definition of a control in the context of this study, as presented in Section 5.1.2. The selected 17,157 THR cases and 85,785 matched controls are all registered with active GP practices with ACU type records and that code valid time of death. In addition, the medical records of these selected THR cases and their matched controls have been accessed at least once within the last ten years before the relevant THR procedure date.

Comparing THR cases to controls

THR cases were matched to five different controls on the GP practice of the patient, their year of birth category and gender. In this section, a comparison between the number of patients across each of the matching factors for cases and controls is provided. The proportion of THR cases and matched controls classified as being male or female is given in Table 32 in Appendix B. The frequency distribution is 38% males and 62% females among the selected cases and controls, showing that THR procedures are more common among female patients, compared to males. The ratio of males to females is the same for cases and matched controls as gender is employed as a matching factor in this study.

THIN dataset at the University of East Anglia consists of patients born between 1920 and 1940 inclusive. However full dates of birth are not available in THIN dataset for adult patients to protect their anonymity. Only the year of birth (YOB) for each patient is allocated in THIN (THIN Data Guide, (2011)). To increase the efficiency of the matching process, i.e. to ensure that enough controls are available to match cases to controls, the YOB of the patient was categorised into 4 different categories, each of 5 years duration; namely *1920-24*, *1925-29*,*1930-34* and *1935-40*.

Table 32 in Appendix B shows the number of patients in each category of YOB for THR cases and controls. Number of male cases differs by a small margin across all categories of YOB with *Category 1930-34* providing the maximum number of THR male cases. Contrastingly, the frequency of female cases is highest in *1920-24* and decreases across each category of YOB. Exactly the same trends for both genders, are obtained for matched controls'

YOB category as the latter is utilised as a matching factor.

GP practices in THIN can be differentiated from each other via their unique identifier, referred to as *pracid*. Identified THR cases and matched controls are registered with 460 different active GP practices, spread across the UK. Male cases are spread among 445 distinct GP practices while female cases are distributed among 456 different GP practices. Eleven GP practices in THIN provide only female THR cases selected for this study.

GP practice is one of the matching factors to identify 5 different controls for each THR case (See Section 5.1.2). Therefore, a constant matching ratio of 5 (referred as M_R) is expected across GP practices when comparing the number of THR cases to matched controls. M_R's across the GP practices are computed and displayed in Table 31 in Appendix B. M_R's are constant and equal to 5 across all GP practices, showing a balanced ratio of population size between the exposed patients and the unexposed ones.

In addition, the ages at time of surgery for THR cases and controls are also compared to ensure that the exposed population is matched to selected controls that belong to the same age category at the time of the THR surgery. The patients are grouped by the GP practice and the mean age at surgery of THR cases and controls respectively, are determined for each of the 460 different GP practices. Table 31 in Appendix B provides the mean age at surgery time for THR cases and matched controls, respectively, across the GP practices. The minimum and maximum differences in age at THR surgery, between THR cases and matched controls, are -0.981 and 0.977 years respectively. This shows that THR cases are matched with controls whose ages at surgery time, are on average, within one calendar year or less.

5.2.2 Patient demographics

Table 32 in Appendix B describes the distribution of cases and controls across gender and their year of birth category. 38% of 17,157 THR cases are male and 62% are female. This shows that the prevalence of THR procedures is more common among female patients. The National Joint Registry (NJR) dataset (Set (2008)), which records data on joint procedures across England, Wales, Northern Ireland and the Isle of Man, also reported that among primary THR patients, 40% were male and 60% were female (NJR (2016)).

Therefore the extracted data from THIN dataset, reflects similar pattern to the NJR dataset, in respect to the ratio of male cases to female cases.

Figure 11 shows the distribution of the number of THR male and female cases by year of birth. The number of male THR cases increases for each year of birth between 1920 and 1932 inclusive and then decreases between 1933 and 1940 inclusive. The number of female cases shows a decreasing trend between 1920 and 1940 inclusive. THR cases are also categorised by year of birth category as explained in Section 5.2.1. Table 32 in Appendix B shows a trend similar to the one displayed by Figure 11 for number of male and female THR cases across each year of birth category.

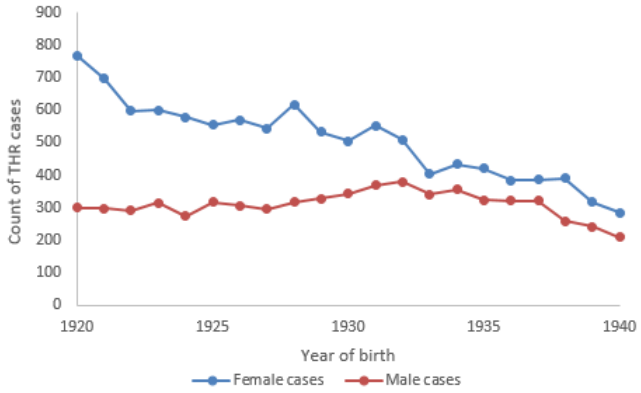


Figure 11: Number of patients across years of birth

5.2.3 Variables Related to Total Hip Replacement Procedure

In this section, the following variables are described in the context of this study: (1) age at surgery time (2) type of THR procedures and (3) revision surgery after THR procedure. The numbers and percentages of patients across the levels of these variables are reported in Table 33 in Appendix B. Age group *65-74* consists of the highest number of THR cases (45.6% for males and 42.6% for females) while only 2.7% of males and 2.6% of females were from the age group *18-54* at surgery, respectively. Proportion of patients aged above 85 at surgery time is also low (3.5% for males and 4.9%

for females).

According to the 2016 annual report from the National Joint Registry (NJR (2016)), 81.8% of male and 84.2% of female patients who underwent a THR procedure are aged between 54 and 84 years, inclusive at time of surgery (versus 93.7% for male cases and 92.3% for female cases in this research). The report also shows the proportions of males aged *55-64*, *65-74* and *75-84*, are 25.0%, 34.6% and 22.1%, respectively, versus 17.5%, 45.6% and 30.6%, respectively, for male THR cases. Similarly, the NJR 2016 report estimates that the percentage of females aged *55-64*, *65-74* and *75-84*, are 20.4%, 35.5% and 28.2%, respectively, versus 13.7%, 42.6% and 36.0%, respectively, for female THR cases in this study. Younger age groups in THIN have a lower number of THR cases than in NJR dataset while proportion of THR cases aged 55 years or more, is higher in the THIN dataset than in the NJR one. These differences show that more patients in THIN underwent their THR procedure at a higher age than patients from the NJR dataset.

The types of fixation techniques identified in THIN are listed in Table 27 in Appendix A and are further classified into *cemented*, *uncemented* and *other types*. Category *other types* include THR procedures described as either *Others* or *Unspecified type* or *Hybrid procedure* in THIN. Table 33 in Appendix B displays the number and percentages of THR cases across the different types of procedures. 45.1% of male and 44.4% of female cases underwent uncemented THR procedure while 36.3% of male and 37.5% of female cases underwent cemented THR surgery. According to the 2016 annual report from the UK National Joint Registry (NJR (2016)), by the end of 2015, 31.0% of THR procedures were cemented, 39.3% were uncemented and 29.7% were either hybrid or described as other types of THR. In this study, proportion of cases who underwent cemented, uncemented and hybrid or other type of THR procedure is equal to 37.1%, 44.6% and 18.3%, respectively. Hence a higher percentage of cases in this study underwent cemented and uncemented hip replacement.

Types of fixation technique employed depends on the age of the patients at surgery time. According to NJR (2016), uncemented procedures are more prevalent among younger patients aged 70 or less while cemented procedures were more common than uncemented ones among older patients, aged 70 year or more. This is because, uncemented procedures have a higher revision

surgery rates than other procedures and thus more older patients prefer to undergo cemented or hybrid THR procedures (NJR (2016)). Figure 12 below shows the variation in number of THR procedures for each type of fixation methods across gender and age groups. It displays a similar trend explained by the NJR (2016) report. Among both males and females, uncemented procedures are more prevalent than cemented and other types of THR for patients aged between 18 and 74, inclusive, while cemented procedures are more common among elderly female cases.

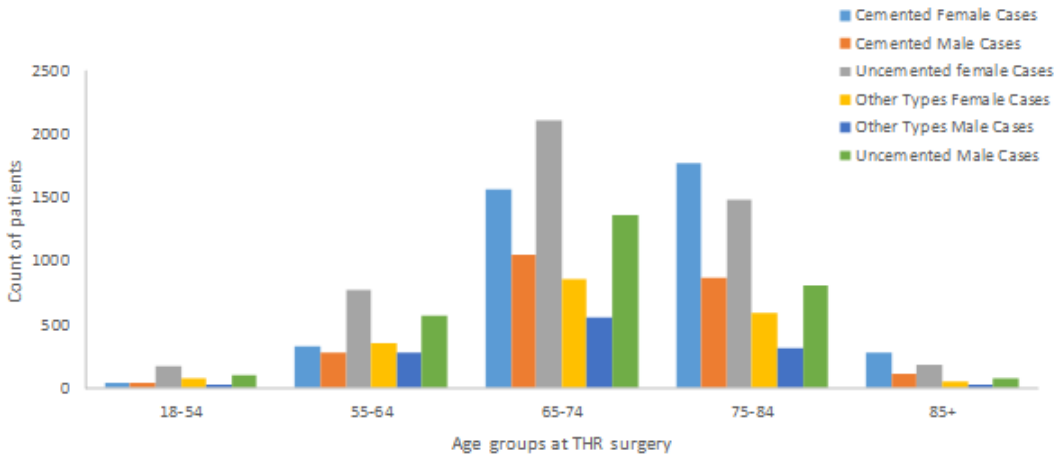


Figure 12: Proportion of THR procedures across each type of fixation methods and age groups

With an increasing number of THR procedures over the last two decades in UK, the mechanical devices which form the total hip prosthesis inevitably fail after surgery, in some patients. The latter therefore require a revision surgery to remove either the failed implant or any bone loss or soft tissue damage. Post-THR revision surgery contributes to an increased risk of death since it adds more surgical complications to the patients' medical history. Type of revision surgery varies at an individual level, depending on the cause of the implant failure (Berry et al. (2012)).

In THIN, Read codes for revision surgery can be identified in patients' MED file and are listed in Table 28 in Appendix A. Table 33 in Appendix B shows the proportion of THR cases who underwent a revision surgery after

their THR procedure. 1.9% of male and 1.6% of female cases underwent a revision surgery. According to the 2016 annual report from the UK National Joint Registry (NJR (2016)), by the end of 2015, 2.6% of patients in England, Wales, Northern Ireland and Isle of Wight who underwent a primary THR procedure, had to undergo a revision surgery. Hence the proportion of cases in this study who underwent a revision surgery (1.8% of cases) is marginally smaller than the reported proportion of revision surgery in the NJR dataset. This difference between the NJR dataset may be due to loss of follow-up of patients transferring out to new GP practices after undergoing their THR procedure and thus leading to under reporting of revision surgeries in THIN.

5.2.4 Post-Code Variable Indicators

Post-Code Variable Indicators (PVI) are Post-Code related indices derived using social, ethnicity and environmental factors of the patients' residential ward. PVI's have been added to the THIN database by UK CSD Medical Research team using the 2001 national survey carried out by the Office for National Statistics (ONS) in UK (THIN Data Guide (2011)). However PVI's are only available for patients who reside in England, Wales and Northern Ireland. THIN patients from Scotland, do not have records of PVI's, and thus their PVI variables are described as *unknown*. The PVI's for the residential ward of each patient from England, Wales and Northern Ireland are stored in the *Demographic* file in THIN and can be classified as follows:

- Ethnicity classification
- Urban-Rural classification
- Long-term illnesses classification
- Pollution classification

Each of the above PVI scores are provided in THIN database, split in the form of quintiles. The scores of each PVI variable were categorised into five groups of equal size, numbered 1 to 5, to indicate the level of ethnicity, pollution and deprivation in the patients' residential ward. Quintile 1 refers to the residential ward with minimum PVI score while quintile 5 refers to residential wards with the maximum PVI score.

Ethnicity in the patients' residential ward is reported by the variables described below. It is important to account for ethnicity in survival analysis because mortality rate differs across different ethnic groups. Wild and McKeigue (1997) showed that mortality across different ethnic backgrounds varies in England and Wales and also reported that prevalence of cardiovascular diseases and cancer differ for different ethnic groups. Gruer et al. (2016) compared life expectancy across different ethnic groups in Scotland and reported longer life expectancies for individuals belonging to such ethnic groups as Asian or Mixed ethnic background, compared to White Scottish population.

- *White ethnicity*: Proportion of population in the residential ward classified as White
- *Black ethnicity*: Proportion of population in the residential ward classified as Black
- *Asian ethnicity*: Proportion of population in the residential ward classified as Asian
- *Mixed ethnicity*: Proportion of population in the residential ward classified as Mixed
- *Other ethnicity*: Proportion of population in the residential ward classified as other ethnicity besides White, Black, Asian or Mixed

Table 34 in Appendix B describes the proportion of THR cases and matched controls across the quintiles of ethnicity. Among all types of ethnicity (White, Mixed, Asian, Black and Others), the distribution of THR cases is close to that of matched controls. Therefore it can be assumed that the population of THR cases belongs to a similar ethnic background as their matched controls in this study.

Patients' residential wards are also described in THIN according to their level of population. Residential wards with population size exceeding 10,000 are grouped as being urban while a ward with a population size less than 10,000 is classified as a rural area. Residential wards with sparse number of inhabitants are labelled as village. A full description of these classifications is provided by the *Office for National Statistics* 2004 national survey (Office for

National Statistics (2004)). Table 34 in Appendix B shows the number and percentage of THR cases and controls whose residential area is categorised as either *Urban*, *Rural* or *Village*. 63% of THR cases in this study were from an urban area while 14% and 9% of THR cases lived a rural area and village, respectively. No Urban-Rural classification were available for 13% and 16% of THR cases and controls, respectively, and their residential wards' Urban-Rural description were unknown in THIN database.

Long-term Illnesses (LLTI) is a term ascribed to individuals who suffer from long-term physical health conditions such as diabetes or coeliac disease and they are the most frequent users of health care services. With an increase in life expectancy, the number of people living with a long-term condition grows and therefore impacts on the individuals themselves, their families and their health care services. Hence it is important to account for the proportion of individuals suffering from LLTI during survival analysis (Lloyd and Heller, (2011)). In THIN, LLTI describes the level (in quintiles) of population with a limited long-term illness in residential wards. Table 34 in Appendix B shows the distribution of THR cases and controls living in a residential wards with different levels of LLTI. Across each quintile of LLTI, the distribution of THR cases is close to that of matched controls, showing that THR cases and matched controls come from similar residential areas. The residential wards of 13% of THR cases and 16% of matched controls were not available in the THIN database and are classified as *Unknown*.

THIN database also provides information on the level of air pollution by residential ward. Air pollution level describes the quintile estimates of the mean level of Nitrogen Dioxide (NO2), Nitrogen Oxides (NOX), Sulphur Dioxide (SO2) and Particulate Matter (PM) in the residential wards. Level 1 pollution represents the areas with the lowest level of pollution while level 5 corresponds to the maximum level of pollution in that residential area. Nevalainen and Pekkanen (1998) and Pope et al. (2009) demonstrate that level of pollution significantly affects life expectancy.

Table 34 in Appendix B shows the distribution of cases and matched controls living in residential wards with different levels of pollution. It can be observed that 39% of THR cases (versus 37% for controls) are from a residential area where the level of NO2 is 4 or 5, 40% of THR cases (versus 38% for controls) are from a residential area where the level of PM is 4 or

5, 36% of THR cases (versus 34% for controls) are from a residential area where the level of SO2 is 4 or 5, 39% of THR cases (versus 37% for controls) are from a residential area where the level of NOX is 4 or 5. Overall, 53% of THR cases (versus 51% for matched controls) are from a residential area where there is maximum pollution (Quintiles 4 or 5 for NO2, PM, SO2 or NOX).

5.2.5 Measures of Deprivation

An increasing availability of administrative data in the UK since the 1970's has led to the development of a number of definitions and measures of deprivation across the UK. Deprivation is defined, in general terms, as the ease of access of an individual to resources and distribution of wealth in the society as a whole (Cook (2000)). In THIN database, three different measures of deprivation are provided for all patients in the UK; namely *Townsend Score*, *index of Multiple Deprivation* and *Mosaic Score*, respectively. In this section, the distribution of patients by these three different deprivation indices is provided.

Townsend Score

Townsend score provides an index for the level of deprivation of the patients' residential ward by combining the following variables of each residential area, (THIN Data Guide (2011)):

- The proportion of residential accommodation having no access to a public transport.
- The amount of households that are rented and not occupied by the owners themselves.
- The percentage of residences that are overcrowded
- The unemployment rate of the active population, aged 16-74 years old inclusive

The above variables were used by the Office for National Statistics (ONS) in UK (THIN Data Guide (2011)) and combined into the Townsend score. The higher the score, the more deprived is the residential ward. Calculated Townsend scores are provided in quintiles in THIN dataset, with quintile 1

representing the least deprived area while quintile 5 is associated with the most deprived area. Data on Townsend Score are available in THIN for all patients residing in the UK.

Table 35 in Appendix B shows the proportion of THR cases and matched controls living in residential areas where the levels of deprivation (measured by Townsend Scores), are rated from level 1 (least deprived) to level 5 (most deprived). 28%, 22% and 17% of THR cases, respectively (versus 25%, 20% and 17% for controls, respectively) are from a residential area where the Townsend score is 3 or less. Table 35 also shows that the proportion of THR cases decreases as the level of deprivation increases. Only 13% and 7% of THR cases are from the highly deprived areas (level 4 and 5 Townsend Scores). This indicates that there are fewer individuals in the most deprived areas who underwent THR procedures, compared to the more affluent areas. 13% of THR cases and 16% of matched controls have unknown Townsend score in THIN dataset.

Index of Multiple Deprivation

Index of multiple deprivation (IMD) is a score attributed to a residential area to describe its level of socio-economic status. It is a combination of several index domains which measures different types or dimensions of deprivation within a residential area (English indices of deprivation (2015)). In UK, seven aspects of deprivation, listed below, are considered in order to estimate the IMD of a residential ward.

- Household income
- Employment
- Health deprivation and disability
- Education skills and training
- Barriers to housing and services
- Crime rate
- Living environment

The above variables were used by the Office for National Statistics (ONS) in UK (THIN Data Guide (2011) to calculate IMD scores that are provided in quintiles in THIN dataset, with quintile 1 representing the least deprived area while quintile 5 is associated with the most deprived area. Table 35 in Appendix B shows the distribution of THR cases and matched controls across the different quintiles of IMD score. For both THR cases and matched controls, the biggest proportion of patients are from an area where the IMD score is the lowest (Quintile 1). The percentages of THR cases and matched controls decreases as the residential areas' IMD score increases. 2% of THR cases have unspecified level of IMD in THIN.

Mosaic Score

Mosaic is a UK based geo-demographic classification system that was developed by a private company called *Experian* (Experian Ltd (2004)) as a tool for the consumer segmentation and marketing. It is a tool that enables businesses to get a better insight of their consumers' demographics and lifestyles, thereby allowing businesses to target the right customers at the right locations. Mosaic is a post code based classification system that categorises individuals' households into 15 major socio-economic groups, which can be further divided into 67 sub-groups as described in Table 29 in Appendix 27. These classifications occur at the level of the full UK postcode. Therefore all individuals living in the same accommodation, are assumed to be in the same Mosaic category (Experian Ltd (2004)).

Table 35 in Appendix B displays the proportions of THR cases and matched controls across the different Mosaic Groups described in Table 29 in Appendix 27. 35% of THR cases (versus 34% for matched controls) are from areas where the residents largely belong to the most affluent and most wealthy families who work in high status positions and who live in privately owned large detached accommodation (most affluent Mosaic group, namely A, B and C). 15% of THR cases (17% for controls) are elderly singles or belong to mature families who reside in rural areas where most residents own inexpensive homes and enjoying a comfortable retirement (Mosaic group D, E and F). 18% of cases (20% for controls) are from areas where most of the residents belong to relatively young families who have children and aspiring to become home makers through their own limited resources (Mosaic group G, H and I). 23% of THR cases (versus 21% for matched controls) are elderly

individuals relying on financial support from their local government to rent low cost accommodation in urban rental locations or elderly home owners of inexpensive homes whose mortgage is nearly paid off (least affluent Mosaic group, namely J, K, L, M, N and O). The remaining 9% of THR cases and 8% of matched controls had no Mosaic group description in THIN dataset.

5.2.6 Lifestyle variables

Smoking status

Information related to the smoking status of patients in THIN is recorded in either the *Demographic* (DEM) or *Medical* (MED) or *Addiditonal Health Data* (AHD) file. The DEM file consists of the recent smoking status record while the MED and AHD files contain smoking status records that have been collected, at several time points, during the period the patients were still active in THIN database. The smoking status of patients in the THIN database is defined as follows:

(1) *Current smoker* is ascribed to cases whose closest smoking status record is prior to their THR surgery date.

(2) *Ex-smoker* is ascribed to cases' whose closest smoking status record before their THR surgery date, is reported as non-smoker but also having previous records of current smoker or ex-smoker or whose last record is classified as ex-smoker.

(3) *Non-smoker* for cases who do not have any records classified as being current smoker or ex-smoker at any time and a smoking status record of non-smoker.

(4) For matched controls, their smoking status is determined in the same way as defined for THR cases in (1)—(3).

Table 36 in Appendix B displays the proportion of smokers, ex-smokers and non-smokers for cases and matched controls, prior to the relevant THR procedure data. Proportions of patients who were ex-smokers and smokers, respectively, prior to their THR procedure, are higher among males than females for cases and matched controls across all age groups. The latest records in the data extracted from THIN for this study are dated until July 2011. According to UK (2016), estimated average prevalence of smokers in

UK as at the end of 2012 was 22.3% for adults aged 64 years or less (versus 9.9% for cases and 7.1% for matched controls) and 10.1% for adults aged more than 64 years (versus 13.4% for cases and 11.9% for matched controls), respectively.

Body Mass Index

Body mass index (BMI) determines the healthy weight of an individual according to his/her body height. BMI is determined by the ratio of the weight of a person, in Kilograms, to the square of his/her body height, in metres. The resulting ratio is then grouped into different categories as described by the table below, to classify individuals as being *underweight, normal weight, overweight* and *obese*, according to the National institute for Health and Clinical Excellence (NICE) guidelines (Linsley et al., (2011)).

BMI Type	**BMI**
Underweight	<18.5
Normal	18.5 - 24.9
Overweight	25.0 - 29.9
Obese	>30

Classification of BMI in UK (Linsley et al. (2011))

In THIN database, the most recent records of weight (in Kilograms), height (in metres) and BMI measurements, for each patient, are provided in the DEM file, while historic records of weight, height and BMI measurements are available in the AHD files. The latest weight and height measurements of THR cases and matched controls before their THR procedure, are extracted from the AHD files. Their BMI are calculated and then classified using the categories described in the table above.

86% of THR cases and 75% of matched controls, respectively, have a measurement of weight and height, dated before their THR procedure date. 4% of 17,157 THR cases and 15% of 85,785 matched controls respectively, have no records of height measurements before THR surgery date, but had a recorded measurement of weight before the surgery. For these patients, their latest height measurements from the DEM file are used to estimate their BMI. Since all patients were adults at surgery time, their height measurement at THR surgery time, can safely be assumed to be close to that of the

most recent record in the DEM file in THIN. 10% of THR cases and 9.6% of matched controls, respectively, do not have any records of weight measurements prior to THR procedure date. Therefore the BMI of these patients before THR procedure cannot be determined and is classified as *unknown*.

Table 37 in Appendix B shows the distribution of THR cases and matched controls according to their BMI classification by age group prior to their THR surgery time. Distribution of overweight cases and controls is marginally different across each age groups for each gender. Proportion of male cases (33.2%) and matched controls (34.5%) who were overweight prior to their relevant THR surgery is higher across all age groups than female counterparts. Prevalence of obesity among female THR cases is higher than that of male THR cases and female controls prior to their THR surgery, across all age groups. Proportion of obesity among male THR cases is marginally higher to that of male controls across all age groups, except for THR cases aged 75 years or more (prevalence of obesity=17.7% versus 10.3% for male controls).

According to Health Survey for England (2015), prevalence of obesity by the end of 2012 was estimated at 24.7% in England, 27.1% in Scotland and 23% for both Wales and Northern Ireland, respectively. This amounts to an estimated mean prevalence of 24.5% for obesity in UK. In this dataset, proportions of THR cases and matched controls who are obese prior to their relevant THR surgery are 30% and 15%, respectively. Obesity is more prevalent among THR cases than in the UK on average while that of matched controls is significantly lower than the estimated UK average prevalence of obesity in 2012. Lower estimated prevalence of obesity among controls may be due to missing records of BMI that are dated prior to the relevant THR surgery time, while a higher prevalence among THR cases may be due to obesity itself creating the need for THR procedure.

According to Carl Baker (2017) statistics on obesity for England only, men are more likely to be overweight or obese than women. 68% of men were overweight or obese in 2015 (versus 50% for male cases and 43% for controls) compared with 58% of women (versus 68% for female cases and 50% for controls). These proportions vary by age groups; reported prevalence of overweight and obesity is between 71% and 75% among all age groups from 45 to 84 (versus 57% for cases and 42% for matched controls, aged between

45 and 84, respectively). Therefore prevalence of overweight and obesity in the dataset extracted is lower than the average prevalence of overweight and obese individuals in England alone.

5.2.7 Medical Variables

In this section, variables describing the medical profiles of THR cases and matched controls are reported. The list of medical variables include Type 2 diabetes mellitus (DM), cardiovascular diseases such as angina, myocardial infarction (MI) and stroke, osteoarthritis (OA), rheumatoid arthritis (RA), bone mass density (BMD), chronic kidney disease (CKD), hypertension (HP) and hypercholesterolaemia (HC). For each of these variables, the following items are provided in this section: (1) definition of each medical variable in the context of this study and (2) description of the prevalence of each variable among THR cases and matched controls.

As mentioned in Chapter 1, one of the objectives of this study is to identify potential medical conditions before THR surgery, that affects mortality risk after THR procedures. Therefore all the medical variables described in this section are dated before the THR surgery of THR cases while for controls, the medical variables are dated before the THR procedure of their matched THR cases. In THIN, medical conditions are coded and comparable to that of the International Classification of Disease 10th revision (ICD–10), provided by ICD (2015). The list of Read codes for the selected medical conditions in this study is based on literature review provided in Chapter 2 and are given in Table 28 in Appendix A.

Diabetes Mellitus

The human body includes of a number of systems and pathways that function in synchrony to create and maintain a healthy state of physiology. One of these systems is the ability of the human body to maintain a stable state (also referred as *homoeostasis*). Any form of homoeostasis instability can cause injury or pathological state in various organs. Diabetes (DM) is a condition where the human body loses its homoeostatic ability to regulate the glucose level in the blood stream because of a lack of insulin hormones in the body (Ahmad (2013)). DM can be classified into 4 major types (Ahmad (2013)):

- *Type 1 diabetes*: Patients who are diagnosed with an absolute insulin deficiency condition
- *Type 2 diabetes*: Patients whose body does not produce enough insulin or whose body's cells are inert to insulin
- *Gestational diabetes mellitus*: Patients diagnosed with DM in the second or third trimester of pregnancy
- *Other types*: Patients diagnosed with specific type of DM due to other causes such as monogenic diabetes syndromes, diseases of the exocrine pancreas, and drug- or chemical-induced diabetes such as in the treatment of HIV/AIDS

The risk of developing type 2 DM increases as individuals grow older, especially after the age of 45 probably because people tend to exercise less, lose muscle mass and gain weight as they age (Sharma et al., 2016). In this study, 99.7% of cases and matched controls were aged more than 45 years at the time of their relevant THR surgery. Hence only type 2 DM records are extracted for cases and matched controls. 158 male cases and 116 female cases were diagnosed, respectively, with type 1 DM before their THR surgery and were excluded from this study.

Table 38 in Appendix B displays the proportions of cases and matched controls diagnosed with type 2 DM prior to the relevant THR surgery. 6.3% of female cases (versus 6.8% for female controls) and 8.3% of male cases (versus 8.8% of male controls) had type 2 diabetes prior to their THR procedure. Hence prevalence between cases and controls differ by gender for all patients. Comparing age groups for cases and controls for each gender, prevalence of type 2 DM increases with age at time of THR surgery for all cases and controls.

Sharma et al. (2016) estimated prevalence of type 2 DM using primary care records in THIN database, dated between 2000 and 2013, inclusive, in the UK. The authors estimated the prevalence of type 2 DM in the UK as 5.32% (versus 7% for cases and 7.5% for controls in this study), with male type 2 DM being more prevalent than female. They also reported that prevalence of type 2 DM increases with age, with the age band 60–69 years having the highest crude percentage of 37.7% for type 2 DM whereas in this

study, prevalence of type 2 DM is more prevalent among patients aged 75 years or more, for both cases and matched controls.

Chronic Kidney Disease

The human kidney is a complex organ, made up of million of nephrons connected together by interstitial tissues. A nephron is made up of a glomerulus (blood vessel) and a renal tubule. Supply of blood to the kidney occurs via the glomerulus arteriole, which acts as a filter for the blood plasma. The rate at which the human kidney can filter the blood plasma entering via the glomerulus arterioles is known as the *glomerular filtration rate* (GFR) and it depends heavily on the adequacy of the blood supply to the kidney and the durability of the glomerular capillary blood vessels. GFR is used to estimate the quality of an individual's kidney health (Lewis, (2012)).

Kidney is a vascular organ and therefore it is vulnerable to diseases that affect blood vessels - for example, hypertension and diabetes. These conditions impact negatively on the kidney function by causing a process called *sclerosis*. During this process, the blood vessels are hardened and thickened, preventing normal blood flow to the kidney and thereby stopping filtration of the blood plasma. Sclerosed glomerulus can never be recovered. Hence an individual will exhibit a reduced GFR as more glomerular vessels are sclerosed. Figure 13 below illustrates the mechanism for the progression of Chronic Kidney Disease (CKD) as a result of self-damaging reduction in the number of functioning glomeruli (Lewis, (2012)).

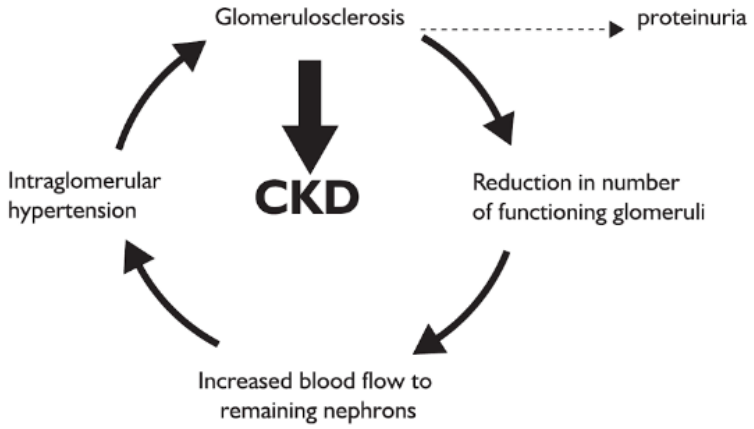


Figure 13: Mechanism for progression of CKD as a result of a reduction in number of functioning glomeruli due to sclerosis (Lewis, (2012))

CKD manifests differently depending on the amount of sclerosed glomerular blood vessels. The different stages of CKD can be classified as follows (Lewis, (2012)):

- CKD Stage 1 (GFR>90 mL/min)
- CKD Stage 2 (GFR = 60-89 mL/min)
- CKD Stage 3a and 3b (GFR = 30–45 mL/min)
- CKD Stage 4 (GFR = 15-29 mL/min)
- CKD Stage 5 (GFR<15 mL/min)

CKD stages 1-3 are classified as *early CKD* and patients at these stages have a reduced GFR of the kidney, therefore causing an increase in levels of urea and creatinine in blood, an increase in level of salt and water overload and hypertension. Stage 4 is recognised as an *advanced CKD* and patients at this stage, suffer from hypocalcaemia, hyperparathyroidism, bone disease, vascular calcification and anaemia. Stage 5 is referred as the *near end-stage CKD* and patients at this stage are, firstly no longer able to remove any waste and fluids from the body, causing an accumulation of toxins in the blood and secondly unable to regulate blood pressure. Therefore these patients require dialysis to remove excess water, solutes and toxins from the blood artificially

(Lewis (2012)).

THR cases and matched controls who have a Read code for CKD that is dated prior to their THR procedure date, are assumed to suffer from CKD. For patients who have multiple records of CKD in their MED file in THIN database, the closest record of CKD before the THR surgery date is assigned to the patients to determine the stage of their CKD. Table 38 in Appendix B shows the distribution of patients diagnosed with CKD stages 1–2 and 3–5, respectively, prior to their THR surgery. It can be noted that most cases and controls were aged 65 years or more when they were diagnosed with CKD. Only 2 female controls were in the age group *55-64* at the time they were diagnosed with CKD stage 3-5.

Among THR cases and matched controls, proportion of CKD stages 1–2 and 3–5 are respectively highest in age group *85+* for both males and females. Comparing patients aged 85 years or more, prevalence of CKD stage 3–5 is higher among females than males for cases and controls. Iwagami et al. (2017) estimated the prevalence of CKD (all stages altogether) in the UK by analysing primary care records from the CPRD database, dated between April 2004 and March 2014, inclusive. Out of 264,628 patients identified with CKD, the authors reported that CKD stages 1–5 were more prevalent among females (60.7%) than males and more prevalent among age groups 75–84 (42.3%) and *85+* (15.8%). In this study, out of 830 THR cases identified with CKD stages 1–5, prevalence of CKD stages 1–5 is more higher among female THR cases (68%) than male THR cases for all age groups at THR surgery. The highest proportion of CKD stages 1–5 is found in the age group *85+* for male and female THR cases, which follows closely the same trend as reported by Iwagami et al. (2017) for prevalence of CKD across different age groups in UK.

Cardiovascular diseases

The human heart is a muscle that is responsible for pumping blood (via blood vessels) to the whole human body to allow a constant flow of the blood stream so that essential nutrients such as Oxygen, and waste products such as urea, are transported to maintain the human body under homoeostatic conditions. Any conditions affecting the proper functioning of the human heart or its blood vessels are classified as cardiovascular disease (CVD) (Jevon, (2012)).

In this study, the types of CVD being investigated are coronary heart diseases (namely angina and myocardial infarction) and stroke.

Angina is the most common form of coronary heart disease and is defined as a condition that occurs when an individual's heart is insufficiently supplied with oxygenated blood by the coronary blood vessels. This happens because of atherosclerosis, i.e., the coronary blood vessels are narrowed or partially blocked by plaques made up of cholesterol and other cells, consequently resulting in a short-lived chest pain or tightness. Angina can be classified into either *stable* or *unstable* (Jevon (2012).

Stable angina occurs when the heart works harder than normal, during physical activities, for example. Chest pain or tightness caused by stable angina has a regular pattern that can be predicted and stopped usually within a few minutes, either by taking a rest or via medication. However *unstable angina* is a more serious chest pain or tightness that cannot be predicted or ascertained to a particular pattern. It can just as easily occur during physical movements as it can during resting position. The chest pain or tightness do not go away with rest and should be treated as an emergency cardiac condition that can cause further complication such as myocardial infarction (Jevon (2012).

Myocardial infarction (MI) causes severe chest pain and tightness that lasts longer than angina and can occur under any circumstances. When a coronary blood artery, is completely blocked by plaques made up of cholesterol and other cells, the heart is deprived of oxygenated blood, leading to the death of part of the heart muscle. Consequently the individual will feel chest pain and tightness that can spread to the upper part of the human body (Jevon (2012).

Stroke is clinically defined as a syndrome of rapidly developing symptoms or signs of loss of cerebral function as a result of vascular problems. It occurs when the blood supply to the brain is disrupted, causing that part of the brain to be deprived of oxygenated blood. Consequently, part of the brain stops working and thus causes loss of sensation in part of the body. The two main causes of stroke are due to either a *blockage* or *haemorrhage* of a blood vessel supplying blood to the brain.

Blockage of blood vessels (also referred as occlusion) is formed by blood clots made up of plaques of cholesterol and other cells. This blockage is either developed along a blood vessel or travels from a different source. Supply of blood to the brain is then disrupted, thereby leading to symptoms of stroke. Stroke occurring as a result of a brain haemorrhage is caused by the leaking or bursting of the cerebral artery, therefore disrupting the supply of blood to the brain. Brain cells die, consequently leading to loss of important brain functions (Lindley, (2008)).

Table 39 in Appendix B shows the distribution of THR cases and matched controls who had at least one event of angina, MI or stroke before the date of their relevant THR procedure. Prevalence of angina among male cases and controls, respectively, is higher than female counterparts across all age groups. For female cases and controls, angina was the most prevalent among the age group *85+* while among males, angina was the most prevalent among the age groups *75–84* for cases and *85+* for matched controls, respectively. Prevalence of angina is marginally higher among female cases than female controls (except for the age group *75–84*) and higher among male cases aged 84 years or less, compared to male controls of the same age categories.

6.2% of female cases (versus 5.9% of female controls) and 12.2% of male cases (versus 12.1% of male controls) had an event of MI prior to their THR surgery. Hence MI was more prevalent among males in this study. Proportion of MI increases with increasing age for both cases and controls across both genders, respectively. MI was the most prevalent among the age group *85+* for female cases and controls, respectively, whereas for male patients, proportion of MI was the highest in the age group *18–54* among cases and *75–84* among controls, respectively.

3.5% of female cases (versus 8.8% of female controls) and 4.4% of male cases (versus 9.8% of male controls) had an event of stroke prior to their THR surgery. Therefore prevalence of stroke among THR cases is marginally higher in men than in women and lower in THR cases than among matched controls. Among cases, the youngest age group *18–54* has the highest proportion of stroke events (4.6% for females and 6.2% for males) mainly because number of patients in the age group *18-54* is the smallest among all age groups. For controls, age groups *85+* and *75-84*, respectively, have the highest percentage of stroke events among female and male controls, respec-

tively. Prevalence of stroke prior to the relevant THR procedure increases with age group for controls whereas among cases, prevalence of stroke across age groups, for cases aged 55 years or more, is similar.

Using data from the CPRD database, Williams et al. (2014) estimated the prevalence of angina, MI and stroke in the UK for males and females, as summarised by the table below. Prevalence of angina and MI before THR surgery is higher among all age groups for all THR cases and matched controls, respectively, in this study (see Table 39 in Appendix B) than the reported prevalence of angina for males and females, respectively, in CPRD database by Williams et al. (2014), shown in the table below. Prevalence of stroke prior to THR surgery for age groups *18–54* (4.6% for female versus 6.2% for males), *55–64* (3.8% for females versus 4.6% for males) and *65–74* (3.3% for females versus 4.3% for males) is higher for cases than reported prevalences of MI by Williams et al. (2014), shown in table below, for the same age groups for both genders. However male and female THR cases aged 75 years or more in this study, have a lower prevalence of stroke prior to THR surgery than those provided by the table below for both genders. Among selected controls in this study, prevalence of stroke is higher than those presented in the table below across all age groups for both genders.

Reported prevalence of angina, myocardial infarction and stroke in the UK as of 2013 (Williams et al. (2014))

Gender		**Male**			**Female**	
Age group	**Angina**	**MI**	**Stroke**	**Angina**	**MI**	**Stroke**
0-44	0.05%	1.21%	0.11%	0.03%	0.02%	0.11%
45-54	0.92%	1.14%	0.89%	0.50%	0.29%	0.79%
55-64	3.60%	3.55%	2.69%	1.74%	0.89%	1.96%
65-74	8.83%	7.05%	6.40%	4.66%	2.06%	4.39%
75+	16.96%	12.08%	14.89%	11.15%	5.50%	12.43%
All ages	3.05%	2.46%	2.53%	1.79%	0.87%	1.99%

Osteoarthritis

The most common symptom of joint diseases is joint pain, which is mainly caused by a degenerative disease known as osteoarthritis (also commonly referred to as *coxarthrosis*). The term osteoarthritis (OA) is derived from three Greek words meaning bone, joint, and inflammation. OA is the most common form of arthritis and is a progressive disorder of the joints caused by

gradual loss of cartilage that acts as a protective cushion between the joint. As the cartilage is gradually worn away, the bone forms areas of abnormal hardening (commonly referred as spurs) and fluid-filled pockets (known as subchondral cysts) in the joint. As the disorder continues, pain results from deformation of the bones and fluid accumulation in the joints. Such a condition causes painful distress to the individual during movements. Figure 2 in section 1.2 illustrates the gradual loss of cartilage between joints due to OA where in the advanced stage of OA, it can be observed that the cartilage is completely worn out causing the bone to erode as well, leading to painful discomfort to the individual's joint.

OA is the major reason why patients undergo total joint replacement for hip, knee and shoulder. The 2016 National Joint Registry in UK reported that 92% of THR, 96% of total knee replacement (TKR) and 53% of total shoulder replacement procedures, respectively, were carried out because of degenerative joint OA (NJR (2016)). In this study, the impact of OA on mortality risk after THR procedure is investigated. THR cases and matched controls who have at least one record in MED file that is coded by the OA Read codes in Table 28 in Appendix A and dated before their THR surgery, are identified as patients diagnosed with OA prior to THR procedure.

Table 40 in Appendix B displays the proportions of THR cases and matched controls diagnosed with OA prior to their relevant THR surgery. Prevalence of OA is high across all age groups among THR cases, ranging from 80.3% among age group *85+* to 88.4% among age group *65–74* for males and ranging from 78.2% among age group *18-54* to 88.0% among age group *65-74* for females, respectively. On the contrary, very low prevalence of OA is observed among controls, with only 27.2% of women and 10.7% of men, diagnosed with OA, prior to their relevant THR surgery date. Compared to the 2016 report by the National Joint Registry in UK (NJR (2016)), OA was the predominant diagnosis in 92% of patients who underwent a THR and is higher than the prevalence of 87% for OA among THR cases in this study. This difference may be due to under-reporting of OA diagnosis in THIN medical records for identified THR cases in this study. Since the NJR dataset is a secondary care database specially for joint replacement patients in the UK, where records on degenerative conditions such OA, rheumatoid arthritis and osteoporosis are not under-reported.

Rheumatoid Arthritis

Rheumatoid Arthritis (RA) is a chronic long-lasting inflammatory joint disease that affects the lining of joints, causing a painful swelling that can eventually result in bone erosion and joint deformity. It is a disorder that occurs when an individual's immune system mistakenly attacks their own body's tissues mostly located between joints. When these tissues are completely worn out, the bones forming the joint erode, leading to limited joint movement because of severe joint swelling, pain, stiffness or fatigue (Newman and Matzko (2006)). The major cause of RA is the individual's abnormal immune system that destroys lining of joint bones. Therefore RA patients have a different state of immune system compared to patients with or without OA and thus have different life expectancies as demonstrated by Watson et al. (2003). The latter study reported that mortality risk is 60% to 70% higher in patients with RA compared to patients with OA and those with no arthritis. Hence it is also important to assess the role of RA as a risk factor for variations in mortality risk after THR procedures.

Table 40 in Appendix B shows the distribution of cases and matched controls, in this study, diagnosed with RA prior to their relevant THR surgery. Prevalences of RA across all age groups are higher among females than among males for both cases and controls. 2.6% of male THR cases (versus 1.5% for male controls) and 4.8% of female THR cases (versus 2.8% for female controls) suffered from RA before their THR surgery. Siebert et al. (2016) analysed the prevalence of RA among 502,649 UK patients who are registered with the UK Biobank organisation for chronic diseases and reported that 0.74% of males and 1.44% of females suffered from RA, across ages 37-73 years. Thus, the proportion of THR cases and controls in this study, who were diagnosed with RA prior to THR procedure, is higher.

Osteoporosis

Osteoporosis is a bone condition that is caused by a decrease in bone mass and a deterioration of the bone structure, thereby leading to an increasingly fragile bone. In clinical practice, osteoporosis is diagnosed by estimating the bone mineral density (BMD) of the patients' bones. BMD test is a surrogate marker for osteoporosis diagnosis because there exists a strong correlation between the bone mineral density of a bone and its weight tolerance ability

(Bhansali and Gogate (2015)). The human bone is made up of organic (30%-40%) and inorganic (60%-70%) components. BMD test estimates only the inorganic component of the bone and the results are produced in the form of T-scores. T-score is a conversion index that measures the deviation of BMD estimate from that of a reference young patient's BMD estimate (Bhansali and Gogate (2015).

The guideline to use T-score as a quantitative definition of osteoporosis was developed by the World Health Organisation (WHO). T-scores which are 2.5 standard deviations below that of the reference level (i.e., T-score of a young adult), can be classified as osteoporosis while a T-score that is between 1 and 2.5 standard deviation below that of a reference young adult, is classified as osteopenia, which represents the early stage of a decrease in bone mineral content (WHO (2017)).

In THIN database, patients suffering from osteoporosis at the hip can be identified in two ways, namely: (1) using medical Read codes for osteoporosis diagnosis in the MED file and (2) using BMD test results from the AHD file in THIN database. The Read codes for osteoporosis at the hip in the MED file and that for BMD test results in the AHD file are listed in Table 28 in Appendix A. Only osteoporosis diagnosis or BMD test results that are dated before the THR cases' and matched controls' THR surgery are extracted from the MED and AHD file respectively. Table 40 in Appendix B displays the distribution of cases and matched controls diagnosed with osteoporosis at the hip, prior to their relevant THR procedure.

The proportions of male and female THR cases who were diagnosed with osteoporosis before THR procedure are 1.8% and 7.3%, respectively, while prevalence of osteoporosis is 2.9% and 6.3% among male and female controls, respectively. For both cases and controls, osteoporosis is more prevalent among females. Comparing across age groups, osteoporosis becomes more prevalent as age increases among THR cases and matched controls, respectively. Proportion of osteoporosis is significantly lower among THR cases compared to matched controls. Osteoporosis is strongly associated with hip fracture (Johnell and Kanis (2005)) and also contributes to an increased mortality risk among males and females as shown by Johansson et al. (1998). In this study, THR procedures due to hip fracture are excluded. Hence the prevalence of osteoporosis among THR cases is significantly lower than

matched controls, although none of the latter had a hip fracture before their relevant THR procedure.

Hypercholesterolemia

Cholesterol is a fatty substance that is found in human body cells and it comes either from self-production by the human liver or from food consumption. It has essential functions in the human body and plays important role in the movement of substances through a cell. It is also important for the production of certain hormones such as progesterone and testosterone, vitamins such as vitamin D and bile acid. Cholesterol, produced by the liver, is moved around the human body in the form of *lipoproteins* while cholesterol from food consumption is stored in the form of *triglycerides*. Lipoproteins exist in two form; namely low density lipoprotein (LDL) and high density lipoprotein (HDL). LDL transports cholesterol from the liver to different tissues of the human body while HDL is responsible for the removal of excess cholesterol from human tissues and brings the excess back to the liver for removal from the body or for reconversion. Collectively, LDL, HDL and triglycerides are called *blood lipids* (Bull and Morrell, (2005)).

The amount of cholesterol in the human body can be determined by a blood test that estimates the amount of blood lipids, in millimoles per litre (mmol/L). In the UK, the general guideline for a healthy cholesterol level, recommends the levels displayed by the table below, for blood lipids in the human body. When the blood lipids measurement is above the minimum healthy level, an individual suffers from *hypercholesterolemia (HC)* or *hyperlipidemia*.

Blood Lipid Type	**Recommended Healthy Estimate (mmol/L)**
Total Cholesterol	<5
LDL	<3
HDL	>1

Recommended Cholesterol Level for Healthy Adults in the UK, according to NHS UK NHS (2015)

HC usually arises either from dietary factors such as consumption of highly saturated fats or via an overproduction of LDL in the human body or via genetic cause such as familial hypercholesterolemia (Bhatnagar et al.,

(2008)). An increased level of cholesterol in blood vessels affects the human blood circulatory system and thus leads to cardiovascular diseases as displayed by Figure 14 below.

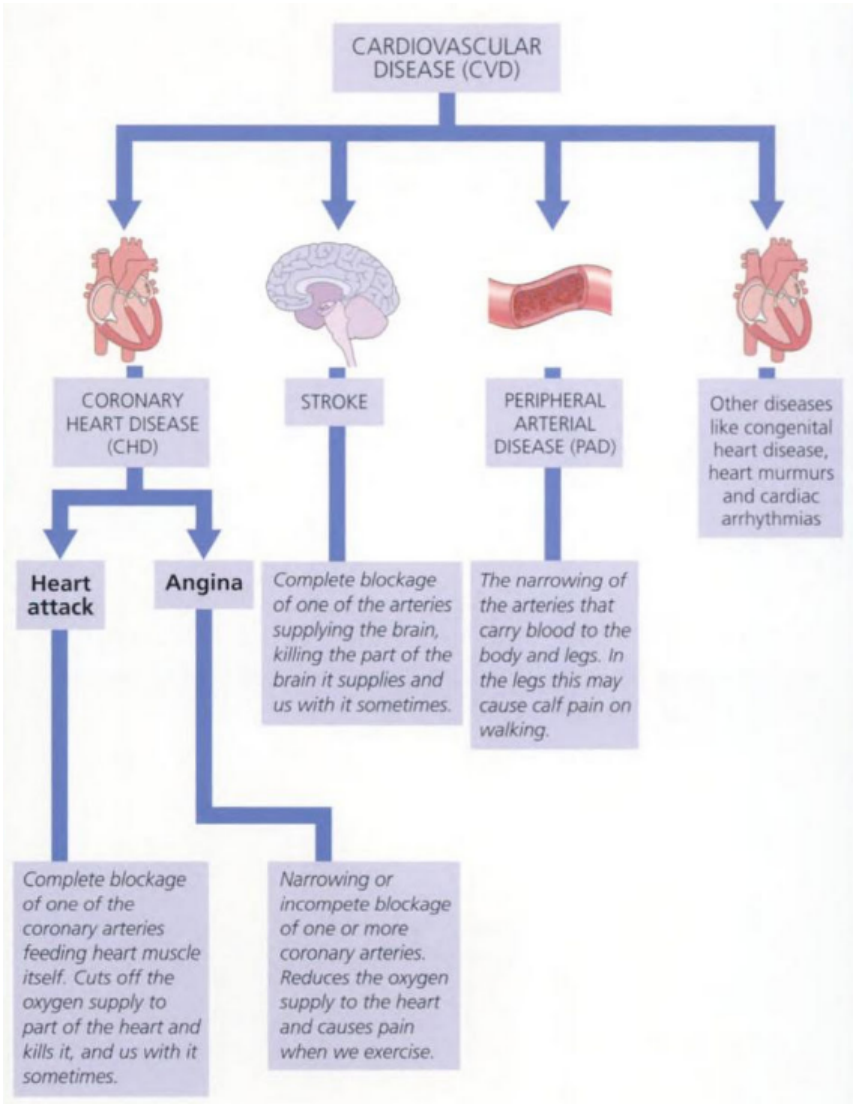


Figure 14: Types of cardiovascular diseases caused by hypercholesterolemia (Bull and Morrell (2005))

In the THIN database, HC for a patient can be identified by one or more of the following methods:

(1) A history of HC in the patients' records in the MED file, identified by Read codes, listed in Table 28 in Appendix A

(2) Recorded high cholesterol measurement (total cholesterol$\geq$ 5 mmol/L or LDL measurement$\geq$ 3 mmol/L or HDL$\leq$ 1 mmol/L in the patients' records in the AHD file), identified by Read codes (listed in Table 28 in Appendix A)

(3) A prescription of lipid-lowering drug treatment, namely statins (Atorvastatin, Fluvastatin, Pravastatin, Rosuvastatin, and Simvastatin), as recommended by the British National Formulary (BNF), Chapter 2.12 (British National Formulary 2.12, (2014).

Using the above methods to identify HC among THR cases and controls, four categories of HC were defined in this study, as shown below.

- *Category 1*: Patients with no records of HC and statins prescription prior to THR surgery
- *Category 2*: Patients with normal cholesterol level due to statins prescription prior to THR surgery
- *Category 3*: Patients with HC despite the statins prescription prior to THR surgery
- *Category 4*: Patients with HC but not with statins prescription prior to THR surgery

Table 41 in Appendix B displays the proportion of THR cases and matched controls in this study, across each of the above categories for HC. Proportion of THR cases and matched controls with normal cholesterol level due to prescription of statins prior to their THR surgery (Category 2 HC), increases with age at surgery peaking at 49.71% among women and 51.75% among men for THR cases aged 85 years or more. Similarly, prevalence of Category 2 HC is higher among male cases than female cases across all age groups. Proportion of controls with Category 2 HC is higher among male and female controls than among THR cases across all age groups, except for older age

groups *75–84* and *85+*, respectively.

Percentage of cases and matched controls with HC despite having statins prescription prior to their relevant THR surgery (Category 3 HC) increases as age of patient at surgery time increases, peaking in the age group *85+* among THR cases and among female controls, and in the age group *75–84* among male controls. Prevalence of Category 3 HC is marginally higher across all age groups for female cases than male cases and matched controls for both genders while that of male cases is similar to the prevalence of Category 3 HC among male and female controls, across all age groups. Proportion of patients with HC but not receiving statins prescription as treatment of HC prior to their relevant THR surgery (Category 4 HC) is highest among younger age groups and decreases with age at surgery for all THR cases and female controls, while proportion of Category 4 HC among male controls decreases with age at relevant surgery time. Prevalence of Category 4 HC is lower among women than men for both THR cases and matched controls.

MacDonald and Morant (2008) reported estimates of the prevalence of HC in the UK using data from the THIN database, from 1998 to 2006. For patients aged 55 years or older, the prevalence of HC in THIN, at the end of 2006, was 50.3% among men and 48.0% among women, respectively. For the same age group, prevalence of HC (Categories 1, 2 and 3, respectively) among THR cases in this study is 44.6% for females and 45.7% for males (versus 50.7% for females and 44.7% for males among controls, respectively). Hence proportion of HC (all categories inclusive) among male THR cases and controls, respectively, is lower than the reported prevalence of HC by MacDonald and Morant (2008) for male patients. Prevalence of female THR cases with any type of HC, is lower than the reported proportion of HC among female patients by MacDonald and Morant (2008) while that of female controls is marginally higher.

MacDonald and Morant (2008) also estimated the proportion of patients with HC and receiving treatment in the form of statins prescription (analogous to Category 2 and 3 HC, inclusive, in this study). They reported that 28.4% of men and 21.7% of women, aged 55 years or more by the end of 2006, respectively, received treatment in the form of statins prescription for HC. In this study, for the same age group, 38.7% of male THR cases (versus 33% for male controls) and 38.0% of female THR cases (versus 39% for

female controls) were on statins prescription for HC, prior to their surgery, by the end of July 2011. This shows that the prevalence of treated HC in this study is significantly greater among male and female THR cases and controls, respectively, than the reported estimates of treated HC for men and women, respectively, by MacDonald and Morant (2008). This may be because statins prescription for patients in THIN, was increasing over time and thus, a higher prevalence is treated HC is obtained in this study.

Hypertension

When blood travels from the heart to different parts of the human body through arteries and back to the heart through veins, pressure is applied against the arterial and venous walls. This pressure is referred as *blood pressure (BP)* and can be measured, in mmHg, using a medical tool called *sphygmomanometer*. The latter measures two types of BP: (1) Systolic BP which is the pressure in the human blood vessels when the heart beats and (2) Diastolic BP which is the pressure in the human blood vessels when the heart rests between beats (Lip and Hall (2007)). According to the guideline on BP measurement by the National Institute for Health and Care Excellence (NICE) (2016), BP is at a normal level if the systolic BP is lower than 140 mmHg and the diastolic BP is less than 90 mmHg. When an individual's systolic or diastolic BP (or both of them) exceeds the recommended level as described by the NICE guidelines, the individual is said to be suffering from *hypertension* (HT). HT is strongly associated with cardiovascular diseases, with patients suffering from HT having the greatest risk of developing cardiovascular diseases (Nadar and Lip (2015)).

In the THIN database, HT in patients can be identified by one or more of the following methods:

(1) A history of HT in the patients' records in the MED file, identified by Read codes listed in Table 28 in Appendix A.

(2) A recorded clinical measurement of BP readings (systolic BP $\geq$90 mmHg or diastolic BP $\geq$140 mmHg, averaged over two successive occasions).

(3) A record of prescription for anti-hypertensive drugs in the patient's AHD file, namely (1) α-adrenoceptor blockers (α-blocker) or β-adrenoceptor

blockers (β-blocker), unless prescribed for anxiety or with nitrates, (2) calcium channel blocker (C-blocker), unless prescribed with nitrates and (3) angiotensin-converting enzyme inhibitors (ACE-inhibitor), unless prescribed for congestive heart failure (BNF Chapter 2.12 (2014)).

Using the above methods to identify HT among THR cases and controls, four categories of HT were defined in this study, as shown below.

- *Category 1*: Patients with no records of HT in MED file or with normal systolic or diastolic BP based on measurements recorded prior to relevant THR surgery and not receiving any anti-hypertensive drug prescriptions prior to relevant THR surgery;
- *Category 2*: Patients with normal BP level due to prescription of anti-hypertensive drugs prior to THR surgery;
- *Category 3*: Patients with HT despite the prescription of anti-hypertensive drugs prior to THR surgery;
- *Category 4*: Patients with HT but without prescription of anti-hypertensive drugs prior to THR surgery.

Table 42 in Appendix B displays the distribution of cases and matched controls, across each category of HT, defined in this study. Comparing cases and controls with normal BP because of anti-hypertensive drug intake (Category 2 HT) prior to their relevant THR surgery, prevalence of Category 2 HT is higher among female cases, aged 74 years or less, than male cases of similar age groups, respectively, while among controls, Category 2 HT is more prevalent across all age groups among men than women. Prevalence of Category 2 HT increases with increasing age at THR surgery for all controls and for male THR cases only. Category 2 HT is more prevalent among the age group *65–74* for female THR cases. Proportions of patients with HT despite being prescribed with anti-hypertensive drugs prior to their relevant THR surgery (Category 3 HT) increases for older age groups, peaking in the age group *85+*, for both THR cases and matched controls, across all genders. Prevalence of Category 3 HT is higher across all age groups among women than men in this study.

Comparing percentage of patients with HT but not taking any anti-hypertensive drugs for treatment of HT (Category 4 HT), prior to their

relevant THR surgery, prevalence of Category 4 HT is higher among controls than THR cases aged 74 years or less while among older age groups, Category 4 HT was marginally higher among THR cases than matched controls, for both males and females, respectively. Among THR cases, age group *65–74* for men and *85+* for women, respectively, have the highest proportions of Category 4 HT whereas Category 4 HT is prevalent among the youngest age group, *18–54*, for male and female controls.

In their published paper, MacDonald and Morant (2008) provide estimates of prevalence of HT among THIN patients, aged 55 years or more, from 1998 to 2006. The authors reported that 55.0% of men and 57.0% of women, in THIN, suffered from HT by the end of 2006. In this study, prevalence of HT (Categories 2, 3 and 4, inclusive) among THR cases and matched controls across all age groups, is estimated at 54.4%, 60.0%, 56.7% and 61.5% among male cases, female cases, male controls and female controls, respectively, prior to their relevant surgery. Therefore reported prevalence of HT by MacDonald and Morant (2008), is close to that of male cases but marginally lower than prevalence of HT among female cases and all controls in this study. This small increase in estimated prevalence may be due to the fact that majority of patients in this study are aged 55 years or more and thus the study population is relatively older than the study population analysed by MacDonald and Morant (2008).

Similarly, MacDonald and Morant (2008) also estimated the proportion of patients, aged 55 years or more, with HT and receiving treatment in the form of anti-hypertensive drug prescriptions (defined as Category 2 and 3 HT, inclusive, in this study). They reported that 37.5% of men and 39.4% of women, aged 55 years or more, were receiving treatment for HT as at the end of 2006. In this study, 41.7% of male cases (versus 38.4% of male controls) and 51.0% of female cases (versus 41.8% of female controls), aged 55 years or more, were receiving treatment for HT in the form of anti-hypertensive drug prescriptions. Hence proportions Category 2 and 3 HT, inclusive, of male cases and controls (both genders), respectively, are close to the reported estimates of treated HT by MacDonald and Morant (2008) for males in UK. However prevalence of Category 2 and 3 HT, inclusive, is higher among female cases than reported prevalence of treated HT by MacDonald and Morant (2008) and this may be because of the larger proportion of women at older ages (55 years or more) in this study population than the one analysed by

MacDonald and Morant (2008).

5.2.8 Drug prescriptions

This section summarises the list of drug prescriptions that were given to THR cases and matched controls during the period they were registered with a GP practice in THIN for the treatment of hypercholesterolemia, hypertension and also prescriptions of hormone replacement therapy. To ensure that the information on drug prescriptions that are extracted from the *Therapy* file in THIN, consists of all the different types of drugs for the treatment of these conditions, publications and guidance on drug prescription from the *British National Formulary* (BNF Publications, (2016)) is used. Note that in this section, all drug prescriptions for all patients are dated before THR procedure for cases while for controls, prescriptions are dated before the surgery of their matched cases.

BNF is a UK based organisation that produces publications incorporating clinical evidence on medications from different sources such as consensus guidelines from the National Institute for Health and Care Excellence (NICE), the Scottish Medicines Consortium (SMC), the Scottish Intercollegiate Guidelines Network (SIGN) and systematic review databases, and approved by a panel of clinical experts with the consent of a *Joint Formulary Committee*. BNF publications reflect the latest best practice and provide legal and professional instructions in the management of medicines in the UK. They issue guidance on the management of drug prescriptions, monitoring, dispensing and administration of medicines and also provide details on the uses, cautions, contra-indications, side-effects, doses, and relative costs of medicines (BNF Publications, (2016)).

In THIN, records of drug prescriptions are available in the therapy file and can be identified using BNF Read codes. The full list of BNF Read codes used to identify drug prescriptions for treatments of hypercholesterolemia, hypertension and medication for hormone replacement therapy among cases and matched controls selected in this study, are presented in Table 30 in Appendix 27. These drug prescriptions have been cross-checked with the recommended NICE guidelines for UK population for the treatments of each condition.

Table 43 in Appendix B shows the proportion of THR cases and matched controls who received drug prescriptions for the treatment of the conditions listed, prior to their relevant THR surgery. 38.9% of male cases (versus 16.7% of male controls) and 38.0% of female cases (versus 14.2% of female controls) were on statins prescription prior to their THR procedure. 29.6% of male cases (versus 31.9% of male controls) and 33.0% of female cases (versus 34.5% of female controls) received an ACE inhibitor prescription for treatment of HT prior to surgery. 26.4% of male cases (versus 28.7% of male controls) and 26.5% of female cases (versus 34.3% of female controls) received an α- or β-blocker prescription for treatment of HT prior to surgery. Prevalence of Calcium-blocker prescription is 22.9% among male cases and 29.7% among female cases (versus 28.4% for male controls and 27.2% for female controls). Prescriptions of α- or β-blockers and C-blockers are lower than the percentage prescriptions of ACE inhibitors because the former two are given in conjunction with ACE inhibitor for patients in whom ACE inhibitor drugs are ineffective.

Hormone replacement therapy (HRT) is a treatment used to increase the level of natural hormones in the body, either in the form of estrogen-alone therapy, for women who have had a surgical menopause (hysterectomy) or both estrogen and progesterone therapy are given together to women who experience menopause naturally at mid-life. HRT helps the female body to process calcium (essential for the strengthening of bone), aids in controlling a healthy cholesterol levels, keeps the vagina healthy, relieves menopause symptoms and protects the female body against osteoporosis (Genazzani (2002)). A complete guidelines and instructions on the management and use of HRT for treatment of menopausal syndromes is provided by the NICE clinical guideline NG23. In men, testosterone is used as a HRT to treat unnaturally low levels of testosterone, caused by ageing. It relieves many symptoms such as reduced libido, lack of energy and concentration or endurance, loss of muscle mass, sexual dysfunction and depression (Kells and Ahlgrimm (2003)).

Among women, 7.6% of THR cases and 8.8% of matched controls received a prescription of estrogen prior to THR surgery, while 1.7% of cases (versus 0.9% of matched controls) were prescribed progesterone as HRT. Prevalence of estrogen prescription for HRT is higher than that of progesterone and marginally smaller among cases than among controls. Among men, 4.2% of cases had a prescription of testosterone for HRT prior to their THR procedure

(versus 4.6% for male controls). Thus proportions of drug prescription for HRT is smaller among males than among females. This is because HRT prescriptions is mainly given to female patients when they enter the stage of menopause, a condition that occurs in all women when they reach the age group *45-55* on average, whereas for men, testosterone prescription is only given as HRT when male patients suffer from a deficiency of androgens (a relatively rare condition referred as *hypogonadism*), which affects the proper functioning of the male reproductive system.

5.2.9 Missing data

Proportion of missingness

In the dataset extracted from the THIN database for survival analysis of primary THR procedures, the following variables have missing values for THR cases and matched controls in this study: smoking status, body mass index (BMI), Townsend score, index of multiple deprivation (IMD) of patients' residential area and the Mosaic category of the patients. Other post-code related variables describing ethnicity, pollution levels, proportion of long term illnesses and urban/rural classification of patients residential wards, have missing values only for patients in THIN that reside in Scotland. These variables are only used for descriptive analysis in this chapter and not included in survival analysis because missing values are systematic for these variables as missingness only applies to patients from Scotland. No missing values are present among medical and drug therapy related variables, because patients with no records of a particular medical diagnosis or drug therapy, are assumed to be free of the medical condition or the drug therapy before THR procedure.

Table 7 shows the percentage of missing values for cases and controls, across each gender for smoking status, BMI, Townsend score, IMD and Mosaic category. Proportion of missing data between male and female cases and between male and female controls, respectively, are close to each other across all variables. Smoking status of patients has the highest proportion of missing data (30.0% for THR cases versus 25.0% for matched controls), whereas the lowest proportion of missing data is observed for IMD (2.1% of missing data for both cases and controls, respectively).

Table 7: Proportion of missing data for cases and controls

Variables	Female cases	Male cases	Female controls	Male controls
Total number of patients	10646	6511	53230	32555
Body Mass Index	10.5%	9.1%	9.5%	9.8%
Smoking status	30.2%	29.5%	25.0%	25.1%
Townsend score	13.0%	13.7%	16.1%	16.9%
Index of Multiple Deprivation	2.0%	2.3%	2.0%	2.3%
Mosaic category	8.4%	8.4%	7.3%	7.7%

Proportion of patients with and without complete records at time of THR procedure, are compared to each other with respect to all variables with complete records, using a χ^2-test of independence (Kleinbaum and Klein, (2005)). The test statistic, χ^2, is given by

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}, \tag{5.1}$$

where O_{ij} and E_{ij} represents the observed and expected number of data points, respectively, in the i^{th} row and the j^{th} column, where r and c are the number of rows and columns in the contingency table. The test statistic distribution is approximated by a χ^2-distribution with $(r-1)(c-1)$ degrees of freedom. This approximation holds if the lowest number of expected observations is greater than five. The extracted dataset is sufficiently large, with sample size equal to $102,943$ subjects for this requirement to hold.

Figure 15 below illustrates the statistical differences between the proportion of patients with and without complete records, respectively, grouped by all the completely observed covariates. Unshaded cells in Figure 15 indicate a p-value smaller than 1% and hence there is a significant systematic difference between proportion of subjects with and without complete records in that category. For instance, there are significant differences in the proportion of patients with and without complete records, for all covariates, when grouped by year of birth. The shaded cells indicate no significant differences (p-values of χ^2-test $> 1\%$) in the proportion of subjects with and without complete records. Figure 15 shows that proportions of missing values differ significantly across the vast majority of covariates. Therefore, presence of missing values in the dataset should not be ignored during the analysis stage. Similarly, it also shows absence of a systematic trend in missingness in the dataset and thus, data is assumed to be missing at random (MAR) in the extracted data from THIN.

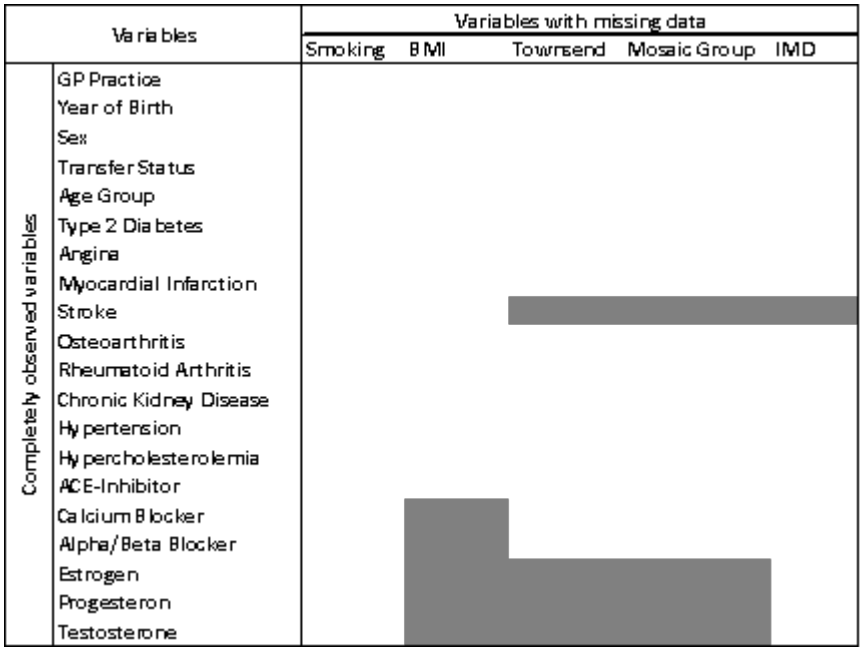


Figure 15: Comparing significance of systematic difference between complete and incomplete records for covariates with missing values

Imputing missing data

Missing data is treated using multiple imputation technique, described in section 3.8. The dataset extracted from the THIN database is classified as hierarchical (multilevel) since patients are clustered via their general practice (GP practice). The extracted dataset represents data at two levels: firstly at the individual level for each individual in each GP practice and secondly at the practice level, where a GP practice is a cluster. Patients within each cluster are assumed to be homogeneous. In the multiple imputation of incomplete multilevel data, it is essential that the imputation model takes the multilevel structure of the incomplete data into account to ensure valid statistical inferences in subsequent multilevel analyses (Black et al. (2011), Graham et al. (2007), Kleinbaum and Klein (2005)). To account for this, a random effect corresponding to a practice is added to the model used for imputing missing data.

In this research, the packages *mitml* (Grund et al. (2017)) and *jomo*

(Matteo and James (2017)) are used to impute missing values for incomplete records using a joint modelling approach. The function *jomoimpute* in the *mitml* package provides an interface to the *jomo* package, which uses the MCMC algorithms presented in section 3.8.2. Through this interface function, imputations for categorical and continuous variables for a multilevel dataset, can be generated at level 1 and level 2, by explicitly specifying the role of each variable in the imputation model in the *jomoimpute* function. The variables have to be described as:

- Target variables containing missing data, defined as type *1*.
- Predictors with fixed effect on all targets (completely observed), specified as type *2*.
- Predictors with random effect on all targets (completely observed), described as type *3*.
- Grouping variable within which the imputation is run separately, listed as type *-1*.
- Cluster indicator variable, defined as type *-2*.
- Variables not featured in the model are set as type *0*.

The list of target variables consists of the following variables: smoking status, BMI, Townsend score, index of multiple deprivation of residential area of patients and the Mosaic category of each patients. These target variables are specified as type *1* in the imputation model. The list of predictors with fixed effect on all target variables consists of all the variables with complete records, as listed in Figure 15 above. These completely observed variables are defined as type *2*. GP practice is used to cluster the patients and is thus defined as type *-2* in the imputation model to account for the multilevel structure of the dataset. Multiple imputation on the incomplete THIN dataset is carried out at the level 1 (individual) only, but accounting for random effects due to clustering by GP practice. For the THIN dataset, the burn-in-length (see section 3.8.2) is set to 500 iterations; the number of iterations until convergence of estimated model parameters is set to 500 and the number of imputed datasets to be simulated is set to 10, respectively. Analysis and description of imputed dataset for the short and long term survival model developed in this thesis, are presented in Chapter 6 and 7, respectively.

5.3 Contribution of chapter

In this chapter, the THIN database is presented as the source of data to complete aims 1–3, described in section 1.4. The information presented in this chapter helps to address the following aims as follows:

Aim (1) : Firstly, primary care records in THIN can be used to identify primary THR procedures in UK, although the description of procedure types can only be categorised as cemented, uncemented, hybrid or as other types. No information describing the material of the hip prosthesis (metal or non-metal) used during the procedure, is available in THIN.

Aim (2) : Secondly, the prevalence of co-morbidities described in this chapter is compared with published studies using similar dataset or other primary care database such as CPRD and QResearch. Conditions in this study that have a marginally lower prevalence among THR cases than relevant published results are proportion of smokers, prevalence of angina, myocardial infarction, osteoarthritis, rheumatoid arthritis, osteoporosis and hypercholesterolemia. Obesity, type 2 diabetes, stroke among cases aged 74 or less, treated hypercholesterolemia (Category 2 and 3 HC, inclusive), treated hypertension (Category 2 and 3 HT, inclusive) and hypertension (among female cases only) were more prevalent among THR cases than relevant published studies. Chronic kidney diseases, hypertension among male THR cases and proportion of stroke among THR cases aged 75 or more have a prevalence level close to the relevant published studies.

Aim (3) : Unknown records in body mass index (9.7%), smoking status (25.8%), Townsend scores (15.9%), Mosaic group (7.6%) and index of multiple deprivation (2.1%) are missing at random in the extracted dataset and are treated using multiple imputation in this study.

6 Short Term Survival Analysis After Primary Total Hip Replacement

Chapter outline

This chapter details the results of investigating mortality risk within the first 24 months (short term) after the THR procedure. Firstly, Kaplan-Meier survival analysis is carried out as preliminary analysis to compare mortality between THR cases and matched controls, 12- and 24-months after the procedure, for different gender and THR procedure types. Secondly, the full procedure to carry out the multiple imputation of missing data is described and assessed. Thirdly, a multilevel Cox regression model is fitted to the imputed dataset to compare all-cause hazard ratio of death of THR cases relative to matched controls. A discussion on the validity of the Cox regression model is then provided. Fourthly, a multivariate logistic regression model with frailty is fitted to the imputed data set to estimate all-cause odds ratio of death within the first 24 months post-surgery. The results of all the survival and mortality models described in this chapter, are presented using forest plots and discussed in details.

6.1 Preliminary analysis

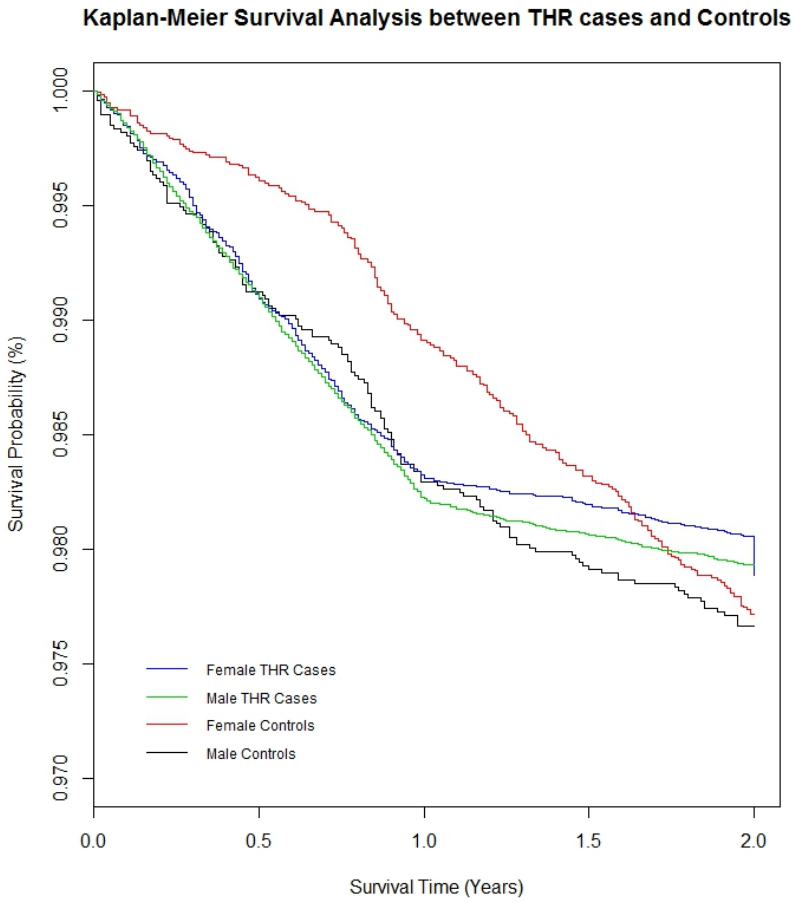
Table 8 displays the distribution of deaths and transferred out patients among THR cases and matched controls within the first two years after THR procedures, across each age group at surgery time and each gender. As a preliminary analysis for the first 24 months after THR procedures, Kaplan-Meier survival analysis is carried out to estimate survivorship for each age group and types of THR procedures. Patients who were transferred out to a new GP practice were assumed to be alive during the 24 months follow-up. The strength of this assumption is assessed in section 6.4.4.To compare the statistical difference between survivorship in each category, the log-rank test (Grambsch and Therneau [1994]), presented in section 3.4, is used.

Table 8: Distribution of transfers out and deaths after THR procedures during the first two years after surgery

Gender		Age Group	No. of Patients	No. Transferred Out	%	No. of deaths after 2 years	%
THR	**Male**	**18-54**	177	2	1.13%	0	0.00%
Cases		**55-64**	1141	3	0.26%	20	1.75%
		65-74	2972	7	0.24%	66	2.22%
		75-84	1993	6	0.30%	39	1.96%
		85+	228	2	0.88%	5	2.19%
		All ages	6511	20	0.31%	130	2.00%
	Female	**18-54**	281	3	1.07%	0	0.00%
		55-64	1457	2	0.14%	32	2.20%
		65-74	4538	6	0.13%	85	1.87%
		75-84	3847	5	0.13%	58	1.51%
		85+	523	1	0.19%	11	2.10%
		All ages	10646	17	0.16%	186	1.75%
Controls	**Male**	**18-54**	1002	5	0.50%	0	0.00%
		55-64	5859	4	0.07%	70	1.19%
		65-74	14606	11	0.08%	304	2.08%
		75-84	9851	3	0.03%	229	2.32%
		85+	1237	2	0.16%	20	1.62%
		All ages	32555	25	0.08%	623	1.91%
	Female	**18-54**	1480	8	0.54%	0	0.00%
		55-64	7594	6	0.08%	0	0.00%
		65-74	22460	13	0.06%	487	2.17%
		75-84	18978	17	0.09%	415	2.19%
		85+	2718	12	0.44%	47	1.73%
		All ages	53230	56	0.11%	949	1.78%

Figure 16 shows the differences in survival over two years between male and female THR cases and their matched controls, respectively. It can be observed that the survivorship of female THR cases is better than male cases. Survival of male THR cases was lower than male controls. A similar trend is observed when comparing female cases to female controls. However by the end of the two year period, the THR cases survivorship improves significantly until they are better than matched controls for both genders.

Figure 16: Survival plot comparing THR cases to matched controls

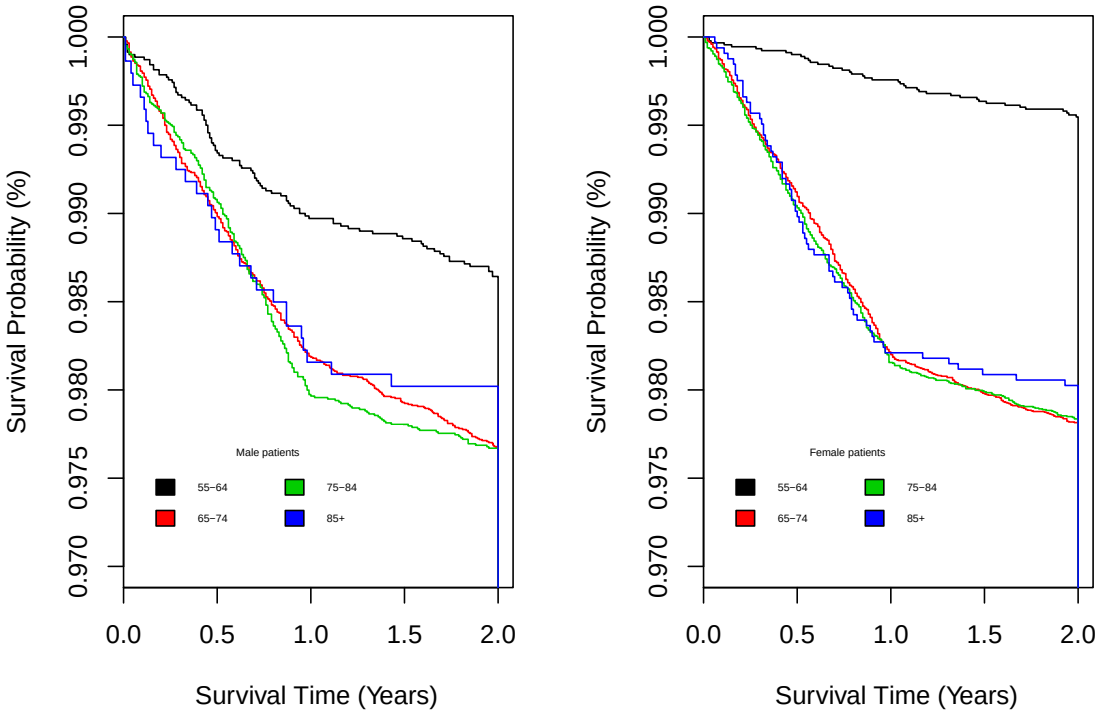


Comparing the Kaplan-Meier survival curves between THR cases and matched controls for both gender

Figure 17 shows the differences in survival two years after THR procedures between different age groups among men and women. There were no death for the youngest age group (18-54) during the first two years after the procedure (Table 8) although five cases from this age group category were transferred out to a new GP practice. Figure 17 shows that survival becomes poorer for older age groups. There are significant differences in survival of older age groups, relative to the youngest category (55-64), for both genders. The

survival of the oldest age category (85+) is the worst among men while the *75-84* age category has the poorest survival among women, one year after the procedure, respectively.

Figure 17: Survival plot comparing THR cases of different age groups

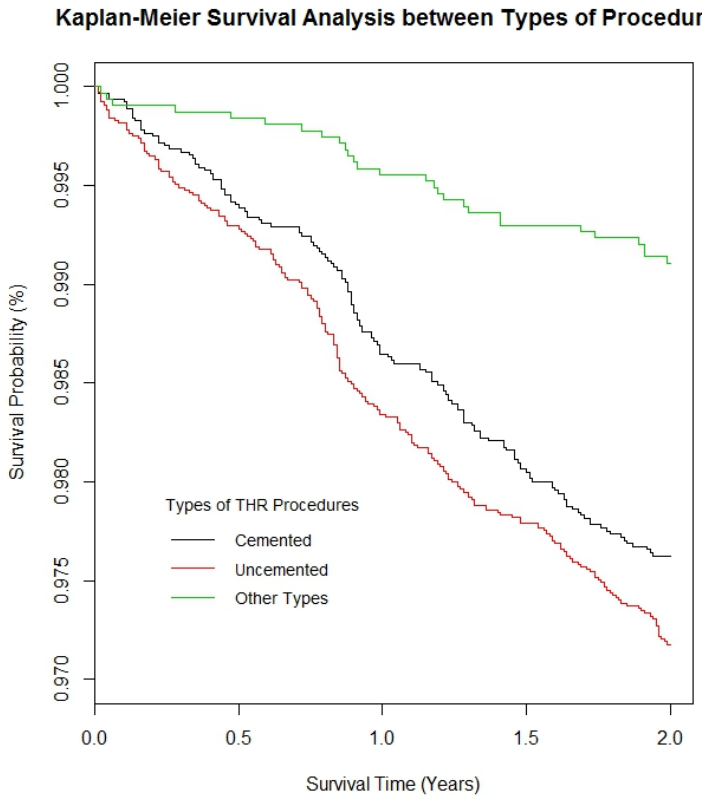


Comparing the Kaplan-Meier survival curves of patients across age groups at surgery time

Figure 18 compares survival of THR cases over two years, across each types of procedures they underwent. The percentage survivorship for cases who underwent *hybrid or other types* of procedures is higher than *cemented* and *uncemented* procedures, respectively. In addition, survival of cases with cemented THR procedures was better than uncemented procedures during

the first two years after surgery. However the preliminary variations in survival of cases depicted in Figure 18 do not account for the fact that the types of THR procedures carried depend strongly on the age group and health status of the patients. Thus further analysis, whereby the age group and medical comorbidities of the patients are factored in, is required to understand and estimate the variations in survival of THR cases with different types of procedures.

Figure 18: Survival plots comparing types of THR procedures



Comparing the Kaplan-Meier survival curves between cemented, uncemented and other types of THR procedures

6.2 Multiple imputation of missing values

6.2.1 The imputation phase

Table 7 in section 5.2.9 displays the proportions of missing data for BMI, smoking status, Townsend score, Index of multiple deprivation and Mosaic category among the 17,157 THR cases and 85,785 matched controls, used in this research. Missing values within these variables are assumed to be missing at random (See section 5.2.9). In the first stage of multiple imputation, there are two key aspects that should be considered:

- Defining the measurement scale of the covariates with incomplete records, and
- Defining the appropriate regression model as the imputation model (IM).

In the dataset analysed in this chapter, BMI, smoking status, Townsend scores, Index of multiple deprivation and Mosaic category are categorical in nature with more than two levels. Therefore the appropriate IM model to impute missing records for these variables, is defined as a multivariate linear mixed model (Schafer and Yucel [2002]) and is of the form shown below.

$$y_i = X_i\beta + Z_i b_i + \varepsilon_i,\ i = 1, ..., m,$$

where y_i is the matrix of incomplete multivariate data for the i^{th} cluster, β is the matrix of coefficients common to the population (fixed effects), b_i is the matrix of coefficients specific to the i^{th} cluster (random effects), X_i is the matrix of covariates for fixed effects, Z_i is the matrix of covariates for random effects b_i and ε_i is the matrix of residual errors.

The covariates with complete and incomplete records, and the cluster variables in this study are categorised as:

- *Fixed effect predictors*: Case-Control Indicator, Sex, Year of Birth Category, Transfer Status of patients, Age group at surgery, Survival time, Type 2 Diabetes, Heart Attack, Angina, Stroke, Chronic Kidney Disease, Rheumatoid arthritis, Osteoarthritis, Oestrogen prescription, Progesterone prescription, Testosterone prescription, Hypercholesterol, Hypertension, Beta Blocker prescription, Statin prescription, Calcium Blocker prescription, ACE Inhibitor prescription

- *Target variables*: Townsend Score, BMI Category, Smoking status, Mosaic Category, Index of Multiple Deprivation
- *Cluster variable*: General Practice

According to van Buuren [2012], the IM used to impute the missing records in the data set should include all the variables that are used for data analysis and showed that having over 25 covariates in the IM hardly influence the explained variance in the imputed datasets. The IM used in this chapter consists of 28 covariates in total and therefore adding any interaction between covariates in the IM will hardly have an impact on the generated datasets.

To carry out the above procedure, the package *jomo* (Matteo and James [2017]) in R, is used because it permits to impute data that are hierarchical in nature. The *jomo* package imputes the missing records based on a Bayesian iterative process involving Monte Carlo simulation as described below.

(1) Using the regression coefficients of the IM, missing values for Townsend Score, BMI Category, Smoker, Mosaic Category and Index of Multiple Deprivation, respectively, are predicted. A random residual term which is normally distributed with mean zero and variance equal to the residual variance from the regression of the missing covariate value on the outcome variable, is added to IM. Adding random residual terms to the mean vector and the covariance matrix of IM produces parameter estimates that differ randomly to those that produced the coefficient estimates of the first IM set up in the first step of the imputation phase. A new dataset (D_{new}) with observed and imputed values is obtained.

(2) Using D_{new}, the new sample means ($\hat{\mu}_{new}$) and the covariance matrix ($\hat{\mathrm{C}}_{new}$) are determined.

(3) Using $\hat{\mu}_{new}$ and $\hat{\mathrm{C}}_{new}$, a new posterior distribution is defined and used to obtain a new set of plausible estimates for the missing values.

(4) Steps (1)–(3) are iterated continuously until convergence of the estimated regression coefficients is achieved. This iteration process of convergence is referred as the burn-in-length and is set to 500 iterations for the imputation of missing records in this dataset.

Once the designated number of burn-in-length has been completed in this dataset, then the entire imputation process is repeated ten times to generate ten imputed datasets. The observed data stays the same across the ten imputed datasets. Only the values that had originally been missing will differ. Table 9 below compares the average distribution of patients in this study across the ten imputed datasets generated to the dataset with full case records only. All imputed variables have a distribution close to that of the incomplete dataset.

Table 9: Comparing average distribution of imputed variables across ten imputed datasets to full case dataset for short term mortality analysis

Variables	Levels	Full Case Dataset*	Average Distribution (SD) in 10 Imputed Datasets**
Townsend Scores	**1**	30.5%	29.3% (5%)
	2	24.9%	22.8% (1%)
	3	19.8%	20.7% (1%)
	4	15.9%	15.6% (2%)
	5	9.0%	12.6% (3%)
	Missing**	16.3%	
Body Mass Index	**Normal**	46.6%	46.6% (2%)
	Overweight	34.0%	32.5% (5%)
	Obese	19.4%	21.6% (1%)
	Missing**	9.9%	
Smoking	**Non-smoker**	69.6%	70.2% (3%)
	Ex-smoker	13.1%	13.3% (2%)
	Smoker	16.4%	16.5% (1%)
	Missing**	26.3%	
Mosaic Categeory	**Cat A-C**	37.5%	37.6% (3%)
	Cat D-F	17.5%	17.6% (5%)
	Cat G-I	20.6%	20.3% (1%)
	Cat J-O	24.4%	24.5% (4%)
	Missing**	7.8%	
Index of Multiple Deprivation	**1**	23.5%	23.5% (4%)
	2	21.9%	22.0% (1%)
	3	19.9%	19.9% (1%)
	4	18.1%	18.1% (1%)
	5	16.6%	16.7% (2%)
	Missing**	2.0%	

**Proportions determined out of a total of 74,787 patients with complete records only.*

***Average proportions calculated out of a total of 102,942 patients with complete and imputed data, across 10 imputed datasets.*

Average distribution of Townsend scores, smoking status, BMI, Mosaic category and index of multiple deprivation across the 10 imputed datasets, is close to that of the full cases only dataset.

6.2.2 Checking for convergence of imputed regression coefficients and comparing distribution of complete and imputed datasets

For the analysis of imputed datasets to yield reliable results, it must be ensured that the iterative algorithm, described above, has converged and that the imputed datasets are approximately independent draws from the predictive distribution (Gill (2014)). The first option to check for convergence is to examine the potential scale reduction factor, also called $\hat{R}$ (Gelman and Rubin (1992)) for the parameters of the imputation model (See section 3.8 for a full theoretical description of $\hat{R}$).

The summary output for the imputation carried out using the *jomo* package in R is shown in Figure 19. In this analysis, $\hat{R}$ is estimated at $\hat{R} = 1.007$, represented by *Psi* in Figure 19, and is very close to one, showing that estimated variance within and between imputed datasets is low and close to one. In addition, the summary also shows how the estimated coefficients of imputation model (represented by *Beta* in Figure 19) used to predict the missing values evolve against the percentage of missing data imputed. It is observed that the estimated *Beta* has a minimum and maximum value of 1.00, indicating that estimated coefficients of the imputation model did not change each time a missing variable is imputed. The estimated variance of random effects between the ten imputed datasets have a maximum estimated value of 1.001 (represented by *Sigma* in Figure 19). This means that the estimated variance of random effects between two randomly imputed datasets in this imputation differ by a maximum of 0.001 only.

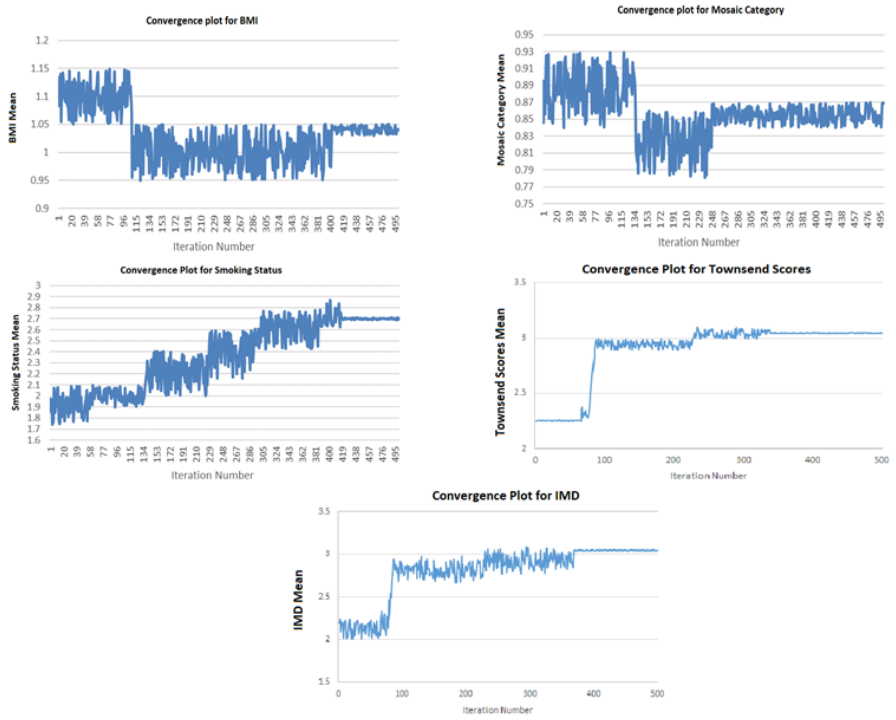
Figure 19: Summary output from multiple imputation of dataset for short term mortality

```
Call:

jomoImpute(data=df.training, type=variable.type, n.burn=500, n.iter=10, m=10)

Cluster variable:         pracid
Target variables:         townsend BMI smoker mosaic_cat IMD
Fixed effect predictors:  (Intercept) cc sex yob_cat xfer status age_group slong diabetes ha
                          angina stroke ckd ra oa osteo oestrogen progesterone testosterone
                          hypercholesterol hypertension beta.blocker statin calcium.blocker
                          ace.inhibitor
Random effect predictors: (Intercept)

Performed 500 burn-in iterations, and generated 10 imputed data sets, each 10 iterations apart.

Potential scale reduction (Rhat, imputation phase):

       Min   25%   Mean  Median  75%  Max
Beta:  1.000 1.000 1.000 1.000  1.000 1.000
Psi:   1.000 1.000 1.001 1.000  1.001 1.007
Sigma: 1.000 1.000 1.000 1.000  1.000 1.001
Largest potential scale reduction:
Beta: [1,4], Psi: [1,1], Sigma: [1,1]

Missing data per variable:
    pracid townsend BMI smoker mosaic_cat IMD cc sex yob_cat xfer status age_group slong diabetes ha angina stroke
MD% 0      16.3     9.9 26.3   7.8        2.2 0  0   0       0    0      0         0     0        0  0      0
    ckd ra oa osteo oestrogen progesterone testosterone hypercholesterol longcat hypertension beta.blocker statin
MD% 0   0  0  0     0         0            0            0                0       0            0             0
    calcium.blocker ace.inhibitor
MD% 0               0
```

10 datasets were generated after carrying out multiple imputation for missing values in BMI, Townsend scores, Index of multiple deprivation, smoking status and Mosaic group, respectively. The potential scale reduction value (Psi) is estimated at $\hat{R} = 1.007$ and is very close to one, showing that estimated variance within and between imputed datasets is close to one. Estimated coefficients of the imputation model stays unchanged with the number of missing data imputed (Minimum and maximum value of Beta = 1.00). Estimated variance of random effects between the 10 imputed data sets (Sigma) is very low and differs by a maximum of 0.001 only across the 10 data sets.

The second option to check convergence of the IM parameters is through diagnostic trace plots that graphically represent the MCMC chain for each imputed covariate at each iteration. Figure 20 below displays the associated convergence plots for each of the imputed variables. They appear to converge to a common value after about 400 iterations.

Figure 20: Convergence plots for multiple imputation of dataset for short term mortality



The convergence plots show that the imputation model parameters used to generate substitute values for missing records in BMI, Townsend scores, Index of multiple deprivation, smoking status and Mosaic group, respectively, have all converged to a common estimate. Hence a burn-in-length of 500 iterations is sufficient to carry out the multiple imputation.

6.3 Multilevel Cox's Regression Analysis

6.3.1 Model development strategy

In this section, two separate multilevel Cox's regression models are fitted to the imputed THIN dataset to estimate all-cause hazard ratio of death after THR procedures at one year and two years, respectively, after the surgery. To address the objectives set out in Chapter 1, a retrospective case control study where THR patients are matched up to five controls with the same gender, year of birth category and GP practice, is used. This allows the estimation of the relative hazard ratio of death between THR cases and controls, taking into account the preoperative history of comorbidities and related treatments, lifestyle factors and residential demographics of the patients.

The two level Cox's regression model (see section 3.5.9) fitted to the data, is of the form

$$h(t, \underline{x}) = Z h_0(t) \exp(\beta; \underline{x}),$$

where $h_0(t)$ is the baseline hazard rate, $\underline{x}$ is the set of covariates fitted and Z is the random effect (frailty) added to the Cox's regression model to account for the clustering effect of patients by their GP practice. The frailty Z is assumed to have an expected value of one and has a multiplicative effect on the individual hazard rate. It either increases ($\hat{Z} > 1$) or decreases ($\hat{Z} < 1$) the individual all-cause hazard ratio of death, where $\hat{Z}$ is the realisation of Z for a GP practice in question.

The set of variables used to fit the above Cox's regression model is provided in Table 44 in Appendix C. The outcome of interest in this analysis is death after THR procedure before the end of target study periods. Patients who left their GP practice before the end of the investigation period, were censored. For each of the one and two years survival models presented in this section, the model selection is carried out using the backward elimination method for both the full case analysis and for the imputed dataset analysis. Initially a full model with all variables listed in Table 44 in Appendix C and without interaction terms, is set up and backward elimination is carried out using a significance level of 5% until the best model is achieved. The short term follow-up time for THR cases and controls is described in Table 10 which defines the start and follow-up time to the target time. The

chosen target times are one and two years after the THR procedure.

Table 10: Follow-up time for THR cases and matched controls for the one and two years survival models

Follow-up	THR Cases	Controls	Target end of study period
Start of follow-up	THR procedure date	Date at which relevant matched THR case underwent the procedure	1 and 2 years exactly after surgery (for cases) and after procedure date of relevant THR cases (for controls).
End of follow-up	The earliest of either transfer out date or death date or the end of study period	The earliest of either transfer out date or death date or the end of the study period	

6.3.2 Multilevel Cox's regression analysis one and two year after THR procedure

Table 11 displays the results of the one and two years multilevel Cox's regression models fitted to the full case and imputed datasets to compare the all-cause hazard ratio of death after THR procedures between THR cases and matched controls. The final one year survival models (full case analysis and imputed dataset analysis) consist of the following variables which have significant effect on all-cause hazard ratio of death: type of THR procedure, year of birth category, gender, smoking status, Mosaic category, body mass index (BMI) category, osteoarthritis, type 2 diabetes, angina, myocardial infarction, stroke and hypercholesterolemia (HC), respectively. Estimated hazard ratios from the full-case analysis are marginally higher than estimated hazard ratios from the imputed data sets analysis, showing that removal of patients with incomplete data caused a marginal but not significant overestimation of hazard ratios. Similarly, the two-year survival model fitted using full case records includes the same list of variables as the one-year survival model in addition to rheumatoid arthritis and hypertension, respectively, but excludes Mosaic category (Table 11).

Table 11: Results of Cox's regression analysis one and two years after THR procedures and estimated all-cause hazard ratios (95% confidence interval) of death

Variables	Levels	One-year Survival Model				Two-year Survival Model			
		Full Case Analysis		Imputed Data Set Analysis		Full Case Analysis		Imputed Data Set Analysis	
		HR (95% CI)	P-value	HR (95% CI)	P-value	HR (95% CI)	P-value	HR (95% CI)	P-value
Procedures Types	**Controls**	1.00		1.00		1.00		1.00	1.00
	Cemented	1.27 (1.17,1.38)	1.30E-08	1.25 (1.15,1.36)	3.80E-08	1.02 (0.96,1.09)	2.00E-16	1.03 (0.95,1.12)	2.00E-16
	Uncemented	1.19 (1.1,1.28)	5.90E-06	1.18 (1.09,1.27)	9.10E-04	1.06 (1.01,1.12)	5.42E-02	1.09 (1.01,1.17)	5.14E-11
	Others	1.30 (1.16,1.45)	5.40E-06	1.27 (1.14,1.43)	2.60E-03	1.09 (1.01,1.17)	1.44E-02	1.07 (0.96,1.18)	1.95E-07
Year of Birth Category	**1920-1924**	1.00		1.00		1.00		1.00	1.00
	1925-1929	0.71 (0.69,0.74)	<2.2e-16	0.70 (0.67,0.72)	<2.2e-16	0.72 (0.70,0.74)	1.77E-02	0.70 (0.67,0.73)	4.82E-06
	1930-1934	0.44 (0.42,0.46)	<2.2e-16	0.44 (0.42,0.46)	<2.2e-16	0.45 (0.44,0.47)	2.00E-16	0.43 (0.41,0.45)	2.00E-16
	1935-1940	0.27 (0.25,0.28)	<2.2e-16	0.26 (0.25,0.28)	<2.2e-16	0.28 (0.27,0.29)	2.00E-16	0.26 (0.24,0.27)	2.00E-16
Gender	**Female**	1.00		1.00		1.00		1.00	1.00
	Male	1.28 (1.23,1.32)	<2.2e-16	1.24 (1.2,1.29)	<2.2e-16	1.29 (1.25,1.32)	2.00E-16	1.30 (1.26,1.34)	2.00E-16
Smoking status	**Non-smoker**	1.00		1.00		1.00		1.00	1.00
	Ex-Smoker	2.33 (2.21,2.46)	<2.2e-16	2.30 (2.18,2.43)	<2.2e-16	1.64 (1.58,1.71)	2.00E-16	1.90 (1.81,1.99)	2.00E-16
	Smoker	2.22 (2.11,2.34)	<2.2e-16	2.21 (2.1,2.33)	<2.2e-16	1.72 (1.66,1.78)	2.00E-16	2.04 (1.94,2.14)	2.00E-16
Mosaic Category	**Group A-C**	1.00		1.00					
	Group D-F	1.08 (1.02,1.14)	8.20E-03	1.05 (0.99,1.11)	8.00E-09			NA	
	Group F-I	1.14 (1.07,1.21)	4.40E-05	1.11 (1.04,1.18)	7.50E-03				
	Group J-O	1.03 (0.98,1.09)	2.30E-01	1.01 (0.96,1.06)	4.90E-06				
BMI	**Normal**	1.00		1.00		1.00		1.00	1.00
	Overweight	0.93 (0.91,0.96)	9.00E-09	0.92 (0.9,0.94)	1.20E-03	0.85 (0.82,0.88)	2.00E-16	0.86 (0.83,0.89)	2.00E-16
	Obese	0.94 (0.91,0.96)	6.15E-03	0.91 (0.89,0.94)	6.80E-10	0.93 (0.89,0.97)	2.00E-16	0.91 (0.87,0.95)	2.00E-16
Osteoarthritis	**No**	1.00		1.00		1.00		1.00	1.00
	Yes	1.18 (1.14,1.22)	<2.2e-16	1.15 (1.11,1.20)	<2.2e-16	1.15 (1.11,1.18)	2.00E-16	1.17 (1.13,1.21)	1.09E-03
Rheumatoid Arthritis	**No**	NA	NA	NA	NA	1.00		1.00	1.00
	Yes					1.41 (1.32,1.51)	2E-16	1.40 (1.29,1.52)	2.00E-16
Type 2 Diabetes	**No**	1.00		1.00		1.00		1.00	1.00
	Yes	1.65 (1.57,1.74)	<2.2e-16	1.63 (1.55,1.72)	<2.2e-16	1.65 (1.59,1.73)	2.00E-16	1.63 (1.55,1.71)	1.85E-15
Angina	**No**	1.00		1.00		1.00		1.00	1.00
	Yes	1.19 (1.14,1.25)	2.30E-13	1.19 (1.14,1.25)	<2.2e-16	1.18 (1.14,1.23)	2.00E-16	1.24 (1.19,1.30)	2.00E-16
Myocardial Infarction	**No**	1.00		1.00		1.00		1.00	1.00
	Yes	1.51 (1.43,1.59)	<2.2e-16	1.47 (1.4,1.55)	4.40E-16	1.53 (1.47,1.59)	2.00E-16	1.53 (1.45,1.60)	2.00E-16
Stroke	**No**	1.00		1.00		1.00		1.00	1.00
	Yes	1.27 (1.21,1.34)	<2.2e-16	1.24 (1.18,1.31)	<2.2e-16	1.31 (1.25,1.36)	2.00E-16	1.29 (1.23,1.36)	2.00E-16
Hypercholesterol	**HC Category 1**	1.00		1.00		1.00		1.00	1.00
	HC Category 2	1.14 (1.13,1.16)	<2.2e-16	1.13 (1.11,1.14)	<2.2e-16	1.46 (1.38,1.55)	2.00E-16	1.41 (1.32,1.50)	2.00E-16
	HC Category 3	1.23 (1.1,1.38)	<2.2e-16	1.21 (1.09,1.35)	<2.2e-16	1.46 (1.39,1.53)	2.00E-16	1.47 (1.47,1.64)	2.96E-11
	HC Category 4	1.31 (1.19,1.45)	<2.2e-16	1.30 (1.18,1.44)	1.10E-12	1.68 (1.62,1.73)	2.00E-16	1.55 (1.41,1.52)	2.00E-16
Hypertension	**HT Category 1**							1.00	1.00
	HT Category 2	NA						1.20 (1.15,1.28)	2.00E-16
	HT Category 3							1.46 (1.40,1.52)	2.00E-16
	HT Category 4							1.56 (1.50,1.67)	2.00E-16

HT Category 1 refers to patients with no records of HT in MED file or with normal systolic or diastolic BP based on measurements recorded prior to relevant THR surgery and not receiving any anti-hypertensive drug prescription prior to relevant THR surgery. HT Category 2 represents patients with normal BP level due and on anti-hypertensive drugs prescription prior to THR surgery. HT Category 3 are patients with HT despite on anti-hypertensive drugs prescription prior to THR surgery. HT Category 4 represent patients with HT and without anti-hypertensive drugs prescription prior to THR surgery. HC Category 1 refers to patients with no records of hypercholesterolemia and statins prescription prior to THR surgery. HC Category 2 represents patients with normal cholesterol level due to statins prescription prior to THR surgery. HC Category 3 are patients with HC despite the statins prescription prior to THR surgery. HC Category 4 are patients with HC but not on statins prescription prior to THR surgery

The estimates of all-cause hazard ratio of death from the full-case analysis are marginal but not significantly higher than the estimates obtained using the multiple imputation method, showing that the exclusion of patients with incomplete records from the analysis, lead to a slight underestimation of the hazard ratios obtained from the survival model using patients with full records only.

The proportionality of the hazards is a critical assumption of Cox proportional hazards analysis. Specifically, the model assumes that each covariate has a multiplicative effect in the hazards function that is constant over time. Violation of the PH assumption can raise questions regarding the validity of the model, and possibly lead to misleading and erroneous scientific findings. To test for the validity of this assumption in this case-control study, a correlation test of Schoenfeld residuals and event time (Grambsch and Therneau [1994]) for each variable contributing significantly to the fitted Cox's regression model, is carried out. Table 12 shows the estimated correlation between Schoenfeld residuals and the event time, and their associated statistical significance for each variable kept in the fitted one-year and two-year survival models (complete case and imputed dataset analysis).

For the one-year survival model (complete case and imputed dataset analysis, respectively), there is no significant correlation between the Schoenfeld residuals and the event time for only Mosaic category, pre-operative osteoarthritis, angina and myocardial infarction, respectively. These variables satisfy the assumption of proportional hazards. However the remaining covariates in the model have strong correlation between the Schoenfeld residuals and the event time. In addition, the global test of proportional hazards for the multivariate Cox regression model is also strongly significant (Table 12), indicating that the one-year model does not satisfy the assumption of proportional hazards. Similarly, for the two-year survival model (complete case and imputed dataset analysis, respectively), there are significant correlations between the Schoenfeld residuals and the event time for all covariates kept in the model (Table 12). In addition, the global test of proportional hazards for the two-year multivariate Cox's regression model is also strongly significant, indicating that the model does not satisfy the assumption of proportional hazards.

Table 12: Testing assumption of proportional hazards for one and two year all-cause mortality Cox regression model

Variables	Levels	One-year Survival Model						Two-year Survival Model					
		Complete Case Analysis			Imputed Data Set Analysis			Complete Case Analysis			Imputed Data Set Analysis		
		Rho	Chi	P-value	Rho	Chi	P-value	Rho	Chi	P-value	Rho	Chi	P-value
Procedures Types	Controls												
	Cemented	0.084	114.01	2.00E-16	0.062	82.40	2.00E-16	0.093	152.77	2.00E-16	0.074	141.14	2.00E-16
	Uncemented	0.135	293.30	2.00E-16	0.106	238.00	2.00E-16	0.138	334.48	2.00E-16	0.107	294.05	2.00E-16
	Others	0.096	144.79	2.00E-16	0.077	125.00	2.00E-16	0.105	194.96	2.00E-16	0.087	199.30	2.00E-16
Year of Birth Category	1920-1924												
	1925-1929	-0.049	36.39	8.27E-10	-0.068	94.30	2.00E-16	-0.051	43.32	2.38E-11	-0.062	97.53	2.00E-16
	1930-1934	-0.091	124.91	2.00E-16	-0.111	250.00	2.00E-16	-0.086	124.99	2.00E-16	-0.094	229.34	2.00E-16
	1935-1940	-0.110	181.61	2.00E-16	-0.129	335.00	2.00E-16	-0.106	189.26	2.00E-16	-0.117	351.32	2.00E-16
Gender	Female												
	Male	0.042	26.55	1.33E-07	0.034	23.10	8.00E-07	0.046	35.44	1.35E-09	0.023	13.83	1.07E-04
Smoking status	Non-smoker												
	Ex-Smoker	-0.305	1461.10	2.00E-16	-0.243	1230.00	2.00E-16	-0.322	1836.78	2.00E-16	-0.233	1417.07	2.00E-16
	Smoker	-0.310	1484.88	2.00E-16	-0.221	1040.00	2.00E-16	-0.308	1672.91	2.00E-16	-0.213	1249.29	2.00E-16
Mosaic Category	Group A-C							NA			NA		
	Group D-F	0.003	0.15	9.48E-01	-0.002	0.11	1.13E+00						
	Group F-I	-0.005	0.37	5.48E-01	-0.008	1.44	1.62E-01						
	Group J-O	0.006	0.73	3.25E-01	0.001	0.01	3.58E+00						
BMI	Normal												
	Overweight	-0.025	10.03	8.38E-04	-0.027	15.10	5.40E-05	-0.010	1.80	1.21E-01	0.013	4.32	2.21E-02
	Obese	0.082	10.15	7.84E-04	-0.022	15.10	5.40E-05	-0.016	4.43	2.07E-02	0.033	3.63	3.41E-02
Osteoarthritis	No												
	Yes	0.011	1.94	1.09E-01	0.008	1.26	1.89E-01	0.021	8.04	2.53E-03	0.011	3.27	4.30E-02
Rheumatoid Arthritis	No	NA			NA								
	Yes							0.021	7.44	3.54E-03	0.017	7.31	3.82E-03
Type 2 Diabetes	No												
	Yes	-0.052	40.57	9.70E-11	0.014	3.82	3.02E-02	-0.040	26.42	1.42E-07	-0.034	29.83	2.43E-08
Angina	No												
	Yes	-0.009	1.32	1.80E-01	-0.066	86.10	2.00E-16	0.012	2.41	7.70E-02	0.033	27.97	6.37E-08
Myocardial Infarction	No												
	Yes	0.009	1.18	2.03E-01	-0.012	2.95	5.31E-02	0.017	4.77	1.68E-02	0.025	16.07	3.22E-05
Stroke	No												
	Yes	0.027	11.23	4.34E-04	0.009	1.80	1.21E-01	0.030	16.13	3.12E-05	0.031	24.67	3.53E-07
Hypercholesterol	HC Category 1												
	HC Category 2	0.055	46.63	4.38E-12	0.024	12.40	2.30E-04	-0.054	49.19	1.18E-12	-0.026	17.49	1.52E-05
	HC Category 3	0.066	13.25	1.45E-04	0.031	20.20	3.65E-06	-0.056	54.38	8.41E-14	-0.032	26.13	1.65E-07
	HC Category 4	-0.065	2.59	6.79E-02	0.294	30.10	2.12E-08	0.075	97.67	2.00E-16	0.048	61.37	2.40E-15
Hypertension	HT Category 1	NA			NA			NA					
	HT Category 2										-0.139	483.19	2.00E-16
	HT Category 3										-0.129	440.26	2.00E-16
	HT Category 4										0.011	3.16	4.62E-02
GLOBAL TEST OF PROPORTIONAL HAZARDS		NA	3260.27	2.00E-16	NA	2980.00	2.00E-16	NA	4245.37	2.00E-16	NA	5019.90	2.00E-16

HT Category 1 refers to patients with no records of HT in MED file or with normal systolic or diastolic BP based on measurements recorded prior to relevant THR surgery and not receiving any anti-hypertensive drug prescription prior to relevant THR surgery. HT Category 2 represents patients with normal BP level due and on anti-hypertensive drugs prescription prior to THR surgery. HT Category 3 are patients with HT despite on anti-hypertensive drugs prescription prior to THR surgery. HT Category 4 represent patients with HT and without anti-hypertensive drugs prescription prior to THR surgery. HC Category 1 refers to patients with no records of hypercholesterolemia and statins prescription prior to THR surgery. HC Category 2 represents patients with normal cholesterol level due to statins prescription prior to THR surgery. HC Category 3 are patients with HC despite the statins prescription prior to THR surgery. HC Category 4 are patients with HC but not on statins prescription prior to THR surgery

Figures 32, 33, 34 and 35 in Appendix C show the plot of Schoenfeld residuals versus time of death for the one- and two-year Cox regression models fitted to the multiple imputed dataset. Graphical inspection for the one-year model shows violation of the proportional hazards assumption for all covariates, except for Mosaic category, pre-operative osteoarthritis, angina and myocardial infarction. This shows that the estimated hazard ratio of death is not constant over time for the first year model. Similarly, the graphical check of the Schoenfeld residuals plots for the second year model shows violation of the proportional hazards assumption for all covariates kept in the two-year Cox regression model, indicating that the estimated hazards of death is time dependent during the first 2 years after surgery.

Additionally, the assumption of proportional hazards was also checked for Cox's survival models fitted using the imputed data set at 1-, 3- and 6-months after the procedures. The aim was to determine whether hazard ratios estimated using the 1-, 3- and 6-months survival models were time independent. The global test of proportional hazards were strongly significant for the 1-month (p-value$< 2.00 \times 10^{-16}$), 3-months (p-value$= 2.04 \times 10^{-08}$) and 6-months (p-value$= 1.19 \times 10^{-11}$) survival models, showing that violation of the assumption of proportional hazards. Therefore estimated hazard ratios were time dependent at 1-, 3-, 6-, 12- and 24-months after the procedure.

As a conclusion, the estimated all-cause hazards of death for the first two years after THR procedures cannot be validated and used to explain variation in short term mortality risk. Consequently, to estimate the short term mortality risk, a multilevel logistic regression model, explained in detail in the following section, is fitted to estimate the all-cause odds of death after THR procedures.

6.4 Multilvel Logistic Regression Analysis

6.4.1 Study design, list of variables and model development strategy

To address the objectives described in Chapter 1, a retrospective matched cohort study design is used. The strengths and limitations of such study design are discussed in Chapter 4. Using a matched cohort study allows estimation of mortality risk at 24 months, associated with THR cases, compared

to controls. In this study, THR cases were matched to 5 controls of the same gender, year of birth category and general practice. The list of variables used to fit a multivariate logistic regression model with frailty to the data is presented in Table 44 in Appendix C. Under a multilevel logistic modelling set-up, the length of time survived by cases and controls during the two-year investigation period is not used. Instead, the number of deaths (represented by death status in Table 44 in Appendix C) within the study population within 24 months, is used to estimate the odds ratio of death between cases and matched controls.

A multivariate logistic regression model with frailty term for GP practice, of the form given by equation (3.74) is used to estimate the effect of THR procedure on all-cause odds of death within the first 24 months after the procedure. The follow-up period for THR cases started from the time they underwent their surgery, while matched controls were followed up from the time their matched case underwent their THR procedure. The end point of the follow-up period is similar for both cases and controls and is defined as the earlier of the following: (1) until patients die, (2) until patients transfer out to a new GP practice, or (3) until the end of the investigation, at exactly 24 months after the procedure. The following models were fitted to the imputed and full case datasets to address the different aspects of the data.

(1) Two main effects only models for full cases analysis (model m_{full}) and multiple imputation analysis (model m_{imp}),

(2) Two models with second order interaction effects for full cases analysis (model $model_{full}$) and imputed dataset analysis (model $model_{imp}$), and respectively.

(3) One model with second order interaction effects for imputed dataset analysis (model $model_{trans}$) that exclude transferred out patients, respectively.

For the main effects only model, the initial models include all covariates presented in Table 44 in Appendix C (excluding survival time variable). For models with interaction effects, second-order interactions between all variables in Table 44 in Appendix C (excluding survival time) with the matching factors gender and year of birth category, THR procedure type and medical variables, respectively, are set up. A similar procedure is set up for the

model that excludes censored patients who are transferred out before the end of the 2 year investigation period. Backward elimination is then carried out on each of the above mentioned models to achieve the most efficient and economical model, with the alpha level set to 5% for fixed effects and 1% for interaction terms, respectively. Finally each of the models also includes a random effect due to clustering effect of general practice (frailty term) to adjust for the in-between patients' correlation within the same GP practice. The models are compared to each other with respect to covariates that contribute significantly to the model. The significance of the contribution of each variable kept in the final model, after backward elimination, is tested by decomposing the multivariate logistic regression model using analysis of variance (ANOVA).

For the remaining sections in this chapter, the outline is as follows: a comparison of m_{full} and m_{imp} for the fixed effects only logistic regression models, (2) a comparison of $model_{full}$ and $model_{imp}$ for logistic regression models including interaction effects and (3) a comparison of $model_{trans}$ model that excludes censored patients versus the $model_{imp}$ model including censored observations. Each comparison comprises of the estimated effect size of the odds ratio of death (95 % confidence interval) and their significance, and are graphically displayed using forest plots.

6.4.2 Logistic regression model with main effects only

Results

The following variables had significant effect (p-value $< 5\%$) on all-cause odds of death after THR procedures for the full case and multiple imputation analysis models: sex, age at surgery, year of birth category, smoking status, BMI, type 2 diabetes, angina, myocardial infarction, stroke, osteoarthritis, hypercholesterolemia and types of THR procedure. Table 13 displays the estimated odds ratio (OR-95% confidence interval) for each covariate kept in both models and their associated p-value for statistical significance. There are small differences in the estimated odds ratio of death from the full case analysis and imputed dataset analysis. This shows the importance of accounting for missing data during the analysis. Completely ignoring the effect of missing data in the analysis would have led to an underestimation or overestimation of the odds ratio of death.

Table 13: Main effects only logistic regression model fitted to full case records and imputed datasets

Variables	Levels	Full Case Analysis		Multiple Imputation Analysis		
		OR (95% CI)	P-value	OR (95% CI)	P-value	RIV**
Sex	**Female**	1.00		1.00		
	Male	2.09 (1.99,2.20)	<2e-16	1.93 (1.45,2.57)	4.20E-06	2.94E-03
Age at surgery	**Unit increase**	1.05 (1.02,1.10)	<2e-16	1.06 (1.01,1.10)	1.51E-02	3.46E-03
Year of Birth Category	**1920-24**	1.00		1.00		
	1925-29	0.62 (0.59,0.66)	<2e-16	0.67 (0.65,0.68)	2.84E-04	2.21E-03
	1930-34	0.65 (0.61,0.69)	<2e-16	0.64 (0.55,0.74)	8.12E-03	3.28E-03
	1935-40	0.65 (0.61,0.70)	<2e-16	0.59 (0.51,0.69)	1.21E-02	3.40E-03
Smoking status	**Non-smoker**	1.00		1.00		
	Ex-smoker	1.85 (1.66,2.05)	<2e-16	1.66 (1.40,1.97)	8.94E-03	3.24E-03
	Smoker	1.92 (1.76,2.10)	<2e-16	2.04 (1.78,2.35)	3.11E-03	4.24E-03
BMI	**Normal**	1.00		1.00		
	Overweight	0.80 (0.76,0.84)	<2e-16	0.84 (0.79,0.88)	7.58E-03	7.01E-03
	Obese	0.95 (0.88,1.01)	5.21E-06	0.85 (0.80,0.89)	8.49E-03	3.89E-03
Type 2 Diabetes	**No**	1.00		1.00		
	Yes	2.31 (2.10,2.53)	<2e-16	2.09 (1.73,2.51)	5.23E-03	2.77E-03
Angina	**No**	1.00		1.00		
	Yes	1.56 (1.47,1.66)	<2e-16	1.34 (1.19,1.51)	1.31E-02	4.73E-03
Myocardial Infarction (MI)	**No**	1.00		1.00		
	Yes	1.95 (1.81,2.09)	<2e-16	2.05 (1.75,2.40)	3.98E-03	2.27E-03
Stroke	**No**	1.00		1.00		
	Yes	1.45 (1.36,1.55)	<2e-16	1.57 (1.51,1.62)	4.61E-04	3.24E-03
Osteoarthritis	**No**	1.00		1.00		
	Yes	1.44 (1.37,1.51)	2.38E-13	1.27 (1.22,1.31)	2.47E-06	3.68E-03
Hypercholesterolemia	**No Hypercholesterolemia**	1.00		1.00		
	Hypercholesterolemia M1*	1.40 (1.29,1.53)	2.65E-10	1.43 (1.38,1.49)	8.85E-04	4.68E-03
	Hypercholesterolemia M2*	1.82 (1.69,1.95)	<2e-16	1.76 (1.64,1.88)	1.73E-03	2.28E-03
	Hypercholesterolemia M3*	1.84 (1.75,1.94)	<2e-16	1.83 (1.44,2.33)	1.28E-02	4.97E-03
Procedure types	**No surgery**	1.00		1.00		
	Cemented procedures	1.28 (1.04,1.58)	1.64E-02	1.41 (1.19,1.66)	1.85E-02	4.41E-04
	Uncemented procedures	1.32 (1.09,1.60)	6.86E-03	1.39 (1.19,1.61)	1.67E-02	3.05E-03
	Others procedure types	1.24 (0.93,1.64)	3.30E-03	1.25 (1.19,1.31)	3.83E-03	4.54E-03

Estimated odds ratio of death during the first 24 months after THR procedures, adjusted for sex, age at surgery, year of birth category, smoking status, BMI, Type 2 diabetes, angina, myocardial infarction, stroke, osteoarthritis, hypercholesterolemia and types of procedures.
**M1 represents patients with normal cholesterol level due to statins prescription prior to THR surgery. M2 are patients with HC despite the statins prescription prior to THR surgery. M3 are patients with HC but not on statins prescription prior to THR surgery.*
*** RIV indicates the reduction in value of estimated model coefficients if the results of the 10 imputed dataset were not combined into a single effect size using Rubin's rule.*

The random effect (frailty term) due to grouping of patients within GP practice, was significant (p-value $< 2 \times 10^{-16}$) in the model fitted to the imputed dataset. This showed that odds ratio of death (OR) in this study varied strongly between GP practices. OR of death within 24 months varied between 0.87 (0.76,0.96) and 6.14 (4.51,7.61) across GP practices. OR of death of THR cases were higher than matched controls (OR=1.00) for all types of THR procedures. OR was the highest for THR cases who underwent cemented procedures [OR=1.41 (1.19,1.66)] followed by uncemented procedures [1.39 (1.19,1.61)] and other types of procedures [1.25 (1.19,1.31)], respectively. Male patients had a higher OR [1.93 (1.45,2.45)] compared to females. The youngest patient in the study was 18 years old at the time of surgery and yearly increase in age at the time of surgery caused the OR to increase by 1.06 (1.01,1.10). Relative to patients born in the *1920-1924* (OR=1.00), patients from the *1925-1929*, *1930-1934* and *1935-1940* year of birth categories had a lower OR of death estimated at 0.67 (0.59,0.66), 0.65 (0.61,0.69) and 0.59 (0.51,0.69), respectively.

Lifestyle factors such as smoking status and BMI prior to the surgery also had significant effect on the short term OR of death after THR procedures. Relative to non-smokers (OR=1.00), patients who were ex-smokers and smokers prior to the surgery, had an OR of death estimated at 1.66 (1.40,1.97) and 2.04 (1.78,2.35). Compared to patients with normal BMI (OR=1.00) prior to the surgery, patients who were overweight and obese before the procedure had a lower OR of death estimated at 0.84 (0.79,0.88) and 0.85 (0.80,0.89), respectively.

Among preoperative medical conditions that had significant effect on OR of death, patients with type 2 diabetes had an estimated OR of 2.09 (1.73,2.51) compared to those without the condition (OR=1.00). Similarly having an event of angina before the procedure increased OR of death to 1.34 (1.19,1.51), relative to patients without angina (OR=1.00). Myocardial infarction (MI) also doubled OR to 2.05 (1.75,2.40) relative to those without MI (OR=1.00). Additionally, having stroke increased OR to 1.57 (1.51%,1.62) , compared to patients without the condition prior to the surgery (OR=1.00). Osteoarthritis (OA) caused OR of death to increase to 1.27 (1.22,1.31) relative to patients not diagnosed with this degenerative condition (OR=1.00).

Preoperative hypercholesterolemia (HC) also increased OR of death fol-

lowing THR procedures for all patients. Relative to participants without hypercholesterolemia, patients with normal cholesterol level due to statins prescription (HC category M1 in Table 13) had an estimated OR of death of 1.43 (1.38,1.49), while those with HC despite being on statins prescription (HC category M2) had an estimated OR of 1.76 (1.64,1.88). Similarly, the OR of death for patients with HC but not on statins prescription (HC category M3) was estimated at 1.83 (1.44,2.33) and was therefore higher than HC category M1 [1.43 (1.38,1.49)] and M2 [1.76 (1.64,1.88)], respectively.

Model comparisons and performance

Two models with main effects only are presented in this section; model m_{full} for the full case analysis and model m_{imp} for the model fitted using imputed datasets. Both models include the same predictors, indicating that missing data did not have any influence on the selection of significant covariates in the optimal model. Relative to the estimated adjusted OR of death of m_{full}, short term OR of death from m_{imp} is higher by 9% for uncemented procedures, 5% for uncemented procedures and 0.80% for other types of procedures. These differences in the estimated OR of death during the first 24 months after the surgery arise due to the of exclusion of patients with incomplete records from the analysis. Incorporating missing data into the model using multiple imputation technique showed the underestimation of the OR of death under the full case analysis method.

Furthermore, the McFadden's pseudo R^2 value (Gordon [2012]) for model m_{full} is estimated between 10.80% and 13.15% (versus 16.40% to 20.56% for model m_{imp}) and thus explains only between 10.80% and 13.15% of the variation in short terms adjusted OR of death following THR procedure. Therefore model m_{imp} explains a higher degree of variability than m_{full}. In addition, for 10 separate portions of the dataset used to fit the model, the test statistics for the Hosmer-Lemeshow test is estimated at 9.65 (p-value=18.71%) for m_{full} and 12.41% (p-value=26.21%) for m_{imp}, respectively. Both m_{full} and m_{imp} are good fit for the data with model m_{imp} performing better than model m_{full}.

Discussion of estimated short term OR of death

In this short term mortality study, adjusted OR of death following THR procedure were estimated using a multilevel logistic regression model with main effects only. The results showed higher OR of death associated with the male gender, increasing age at surgery, preoperative smoking status and medical conditions such as type 2 diabetes, angina, MI, stroke, osteoarthritis and hypercholesterolemia, respectively. These results are in agreement with a number of studies which reported similar findings when comparing THR cases to a control population: Fender et al. (1997), Lie et al. (2002), Nunley and Lachiewicz (2003), Blom et al. (2006), Pedersen et al. (2011), Jones et al. (2014), Lie et al. (2000), Williams et al. (2002), Ramiah et al. (2007), Aynardi et al. (2009), NHS Scotland (2002), Lie et al. (2000), Hunt et al. (2013), NHS Scotland (2002) and Lovald et al. (2014). However the model estimated a lower OR among patients who were overweight or obese prior to their relevant THR procedure. This result is similar to the reported findings by Bozic et al. (2012) and Memtsoudis et al. (2012) who both reported a lower 90-day mortality for THR cases who were either overweight or obese pre-operatively, compared to patients with normal BMI.

However several studies, reviewed in chapter 2 of this thesis, demonstrated that mortality risk during the first 24 months after the procedure varied significantly for patients with various combinations of preoperative characteristics or medical conditions. Table 14 summarises the different publications who investigated the effect of interactions between various medical conditions and gender, age group and procedure types, respectively, on short term mortality risk after the surgery.

Table 14: Studies investigating effect of interactions between covariates on short term mortality after THR

Significant interaction term reported	Studies
Gender and Procedure types	Barett et al. (2005), Lie et al. (2010), Pedersen et al. (2011), Lovald et al. (2014), Xu et al. (2017)
Age group and Procedure types	Lie et al. (2010), Pedersen et al. (2011), Lovald et al. (2014), Xu et al. (2017)
Osteoarthritis and Procedure types	Lie et al. (2000, 2002)
Osteoarthritis and Gender	Lie et al. (2000, 2002)
Rheumatoid Arthritis and Procedure types	Lie et al. (2002)
Diabetes and Gender	Pedersen et al. (2011)
Diabetes and Procedure types	Pedersen et al. (2011)
Cardiovascular diseases and Procedure types	Pedersen et al. (2011), Hunt et al. (2013), Jones et al. (2014)

The studies in this table tested the signficance of the mentioned interaction terms using either a Cox's regression or Logistic regression model. All authors reported a significant p-value (p-value $< 1\%$) for these interaction terms and used the estimated hazard ratios of death or odds ratio of death to further explain the variations in short term mortality risk for each sub-category of THR cases, relative to controls.

Based on the reported findings in Table 14, it is highly desirable to have a mortality risk model that includes the effects of interactions between covariates on short term mortality risk. In this section, the two level logistic regression model for the full case analysis and imputed datasets analysis, respectively, only tested the main effects of the covariates included in the model. To have a deeper insight on the variability of adjusted OR of death at 24 months after the procedure, second order interaction terms are added to the models presented in this section and their effects on the OR of death within 24 months are reported in section 6.4.3.

6.4.3 Logistic regression model with interaction terms

Model selection strategy

In order to obtain a deeper insight on the variability of OR of death after THR procedures, several interaction terms of second order are added to the main effect models developed in section 6.4.2. Initially a full model with all variables listed in Table 44 from Appendix C and that includes the following interaction terms, listed below, is set up.

1. THR procedure types and demographic variables (age at surgery, gender)

2. THR procedure types and lifestyle factors (smoking status, BMI)

3. THR procedure types and social deprivation indices (Townsend score, Mosaic score)

4. THR procedure types and preoperative medical conditions (angina, myocardial infarction, stroke, osteoarthritis, rheumatoid arthritis, osteoporosis, hypertension, hypercholesterolemia, chronic kidney diseases)

5. THR procedure types and drug prescriptions (Ace-inhibitor, oestrogen, progesterone, testosterone)

6. Gender and lifestyle factors (smoking status, BMI)

7. Gender and preoperative medical conditions (angina, myocardial infarction, stroke, osteoarthritis, rheumatoid arthritis, osteoporosis, hypertension, hypercholesterolemia, chronic kidney diseases)

8. Age at surgery and lifestyle factors (smoking status, BMI)

9. Age at surgery and preoperative medical conditions (angina, myocardial infarction, stroke, osteoarthritis, rheumatoid arthritis, osteoporosis, hypertension, hypercholesterolemia, chronic kidney diseases)

Backward elimination is carried out firstly by excluding, one at a time, interaction terms having no significant effect (p-value $> 1\%$) on the OR of death from the model and secondly, by removing main effect covariates, one at a time, that do not statistically explain variation in the OR of death (p-value $> 5\%$) in order to achieve the most efficient and economical model. Finally, the model with interaction terms also includes a random effect term due to grouping of patients via their GP practice to adjust for between patients' correlation within the same GP practice. Two separate multivariate logistic regression models are developed and compared; one for a full case analysis for patients with complete records only ($model_{full}$) and a second model ($model_{imp}$) using the 10 imputed datasets. $model_{imp}$ is fitted using Rubin's rules (See section 3.87). $model_{full}$ and $model_{imp}$ are compared to each other with respect to covariates that contribute significantly to the model. The significance of the contribution of each variable kept in the final model, after backward elimination, is tested by decomposing the multivariate logistic regression model using ANOVA (analysis of variance).

Results

After backward elimination, $model_{imp}$ and $model_{full}$ include the same predictors and interaction terms, as shown in Table 15. $model_{imp}$ and $model_{full}$ include the following main effects: (1) gender, (2) year of birth category, (3) age at surgery, (4) smoking status, (5) body mass index (BMI), (6) type 2 diabetes, (7) angina, (8) myocardial infarction (MI), (9) stroke, (10) osteoarthritis, (11) hypercholesterolemia, respectively, and significant interaction terms between (1) procedure types and gender, (2) procedure types and smoking status, (3) procedure types and BMI, (4) procedure types and MI, (5) gender and type 2 diabetes and (6) smoking status and BMI. The results from $model_{imp}$ presented in Table 15, are converted into OR of death (95% confidence interval) and discussed in this section. To explain the effect of interaction terms on short term OR of death, the results are presented using forest plots to compare the variations in OR between several categories of THR cases and controls.

$model_{imp}$ does not include interaction between types of THR procedures and (1) age at surgery, (2) year of birth category, (3) angina, (4) stroke, (5) osteoarthritis and (6) hypercholesterolemia, indicating that there was no difference in short term OR of death between THR cases and matched controls in these categories or with these preoperative comorbidities. For all patients in this study, yearly unit increase in age at surgery time raises the adjusted OR of death by 30% (18%,44%) while OR of death for patients born between 1925-29, 1930-34 and 1935-40, are equal to 0.65 (0.53,0.80), 0.64 (0.47,0.88) and 0.63 (0.44,0.91), respectively, relative to patients born in the 1920-24 year of birth category (OR=1.00). This shows that younger patients are at a lower odds of death than older ones within the first 24 months after THR and may also reflect the advancement in modern technology in surgical interventions such as THR procedure with time.

Table 15: Estimated coefficients of $model_{imp}$ and $model_{full}$ and their significance level

Variables	Levels	Full Case Analysis				Multiple Imputation Analysis					
		Estimate[1]	Std.Err	z	Pr(>|z|)	Estimate[1]	Std.Er	t.value	df	P(>|t|)	RIV
(Intercept)		-0.6787	0.039	-17.59	<2e-16	-0.6783	0.026	-2.60	1	4.1E-02	3.8E-03
Sex	Male	0.6301	0.024	26.34	<2e-16	0.6244	0.059	8.18	1	4.7E-03	2.9E-03
Age at surgery	Unit increase	0.0551	0.008	9.09	<2e-16	0.0542	0.005	5.52	1	1.0E-02	3.5E-03
Year of Birth Category	1925-29	-0.4801	0.026	-18.71	<2e-16	-0.4299	0.106	-4.05	1	1.8E-02	2.2E-03
	1930-34	-0.4631	0.029	-37.35	<2e-16	-0.4432	0.162	-3.24	1	2.8E-02	3.3E-03
	1935-40	-0.4362	0.033	-49.24	<2e-16	-0.4561	0.182	-2.50	1	4.4E-02	3.4E-03
Smoking status	Ex-smoker	0.2422	0.052	12.26	<2e-16	0.2420	0.143	3.08	1	3.0E-02	3.2E-03
	Smoker	0.3304	0.046	13.74	<2e-16	0.3430	0.105	6.00	1	8.6E-03	4.2E-03
BMI	Overweight	0.2084	0.027	-11.14	<2e-16	0.2073	0.089	-3.34	1	2.6E-02	7.0E-03
	Obese	0.2581	0.034	-5.62	2E-08	0.2590	0.074	-2.57	1	4.2E-02	3.9E-03
Type 2 Diabetes	Yes	0.7124	0.046	15.39	<2e-16	0.7114	0.197	3.61	1	2.3E-02	2.8E-03
Angina	Yes	0.2851	0.031	9.35	<2e-16	0.2843	0.071	4.01	1	1.9E-02	4.7E-03
Myocardial Infarction (MI)	Yes	0.6320	0.036	17.54	<2e-16	0.6319	0.043	4.41	1	1.6E-02	2.3E-03
Stroke	Yes	0.3359	0.033	10.06	<2e-16	0.3354	0.029	3.09	1	3.0E-02	3.2E-03
Osteoarthritis	Yes	0.1782	0.023	-7.75	9E-15	0.1779	0.065	2.72	1	3.8E-02	3.7E-03
Hypercholesterolemia	Hypercholesterolemia M1*	0.3258	0.043	-7.59	3E-14	0.3252	0.042	2.66	1	3.9E-02	4.7E-03
	Hypercholesterolemia M2*	0.4350	0.035	-16.94	<2e-16	0.4344	0.099	4.38	1	1.6E-02	2.3E-03
	Hypercholesterolemia M3*	0.5990	0.025	-17.71	<2e-16	0.5981	0.058	10.26	1	3.0E-03	5.0E-03
Procedure types	Cemented procedures	0.2154	0.105	-2.05	4E-02	0.2145	0.081	2.64	1	4.0E-02	4.4E-04
	Uncemented procedures	0.1864	0.097	-1.92	5E-02	0.1859	0.057	3.28	1	2.7E-02	3.0E-03
	Others procedure types	0.0294	0.144	-2.67	1E-02	0.0294	0.005	6.53	1	7.3E-03	4.5E-03
Procedure types*Sex	Male Cemented procedures	0.2576	0.100	2.57	1E-03	0.2572	0.075	3.41	1	2.5E-02	2.9E-03
	Male Uncemented procedures	0.1706	0.089	1.92	5E-03	0.1698	0.057	2.95	1	3.3E-02	3.4E-03
	Male Other procedures	0.0504	0.138	-0.37	7E-03	0.0499	0.017	3.02	1	3.2E-02	3.3E-03
Procedure types*Smoking status	Ex-smoker Cemented procedures	0.2713	0.132	-5.00	6E-07	0.2708	0.050	5.37	1	1.1E-02	1.9E-03
	Smoker Cemented procedures	0.4613	0.132	-2.06	4E-02	0.4606	0.013	2.19	1	5.5E-02	8.6E-03
	Ex-smoker Uncemented procedures	0.2622	0.125	-5.51	4E-08	0.2615	0.061	4.25	1	1.7E-02	2.4E-03
	Smoker Uncemented procedures	0.4904	0.112	-2.35	2E-02	0.4895	0.138	3.54	1	2.4E-02	4.8E-03
	Ex-smoker Other procedures	0.4156	0.184	-3.75	2E-04	0.4149	0.154	2.70	1	3.8E-02	3.7E-03
	Smoker Other procedures	0.3879	0.171	-2.43	2E-02	0.3687	0.059	2.65	1	4.0E-02	3.8E-03
Procedure types*BMI	Overweight Cemented procedures	0.1412	0.115	3.22	2E-03	0.1402	0.015	9.06	1	3.8E-03	1.1E-03
	Obese Cemented procedures	0.0755	0.122	2.62	1E-02	0.0747	0.028	2.64	1	4.0E-02	4.8E-03
	Overweight Uncemented procedures	0.1305	0.103	2.97	3E-03	0.1309	0.089	3.44	1	2.5E-02	2.9E-03
	Obese Uncemented procedures	0.1272	0.112	2.43	1E-02	0.1279	0.067	4.08	1	1.8E-02	2.5E-03
	Overweight Other procedures	0.1261	0.160	2.49	2E-02	0.1260	0.079	3.32	1	2.6E-02	4.0E-03
	Obese Other procedures	0.1691	0.167	2.78	8E-03	0.1687	0.049	3.45	1	2.5E-02	2.9E-03
Procedure types*MI	Cemented procedures with MI	0.2366	0.153	-2.01	5E-02	0.2365	0.030	7.93	1	5.0E-03	1.3E-03
	Uncemented procedures with MI	0.1500	0.138	-1.91	6E-02	0.1490	0.042	3.54	1	2.4E-02	2.8E-03
	Others procedures types with MI	0.8066	0.232	-3.48	5E-04	0.1058	0.184	4.38	1	1.6E-02	2.3E-03
Sex*Diabetes	Male Type 2 Diabetes	0.2527	0.068	-3.72	2E-04	0.2520	0.074	3.40	1	2.5E-02	2.9E-03
BMI*Smoking status	Overweight Ex-smoker	0.0273	0.009	3.47	5E-04	0.0272	0.106	2.57	1	4.2E-02	3.9E-03
	Overweight Smoker	0.0342	0.010	4.46	8E-06	0.0341	0.176	1.94	1	6.7E-02	9.1E-03
	Obese Ex-smoker	0.0341	0.012	4.37	1E-05	0.0344	0.070	4.88	1	1.3E-02	2.1E-03
	Obese Smoker	0.0419	0.013	3.19	1E-03	0.0426	0.131	3.20	1	2.8E-02	3.1E-03

[1] *Estimated coefficients for multilevel logistic regression model fitted to full case records and imputed datasets*

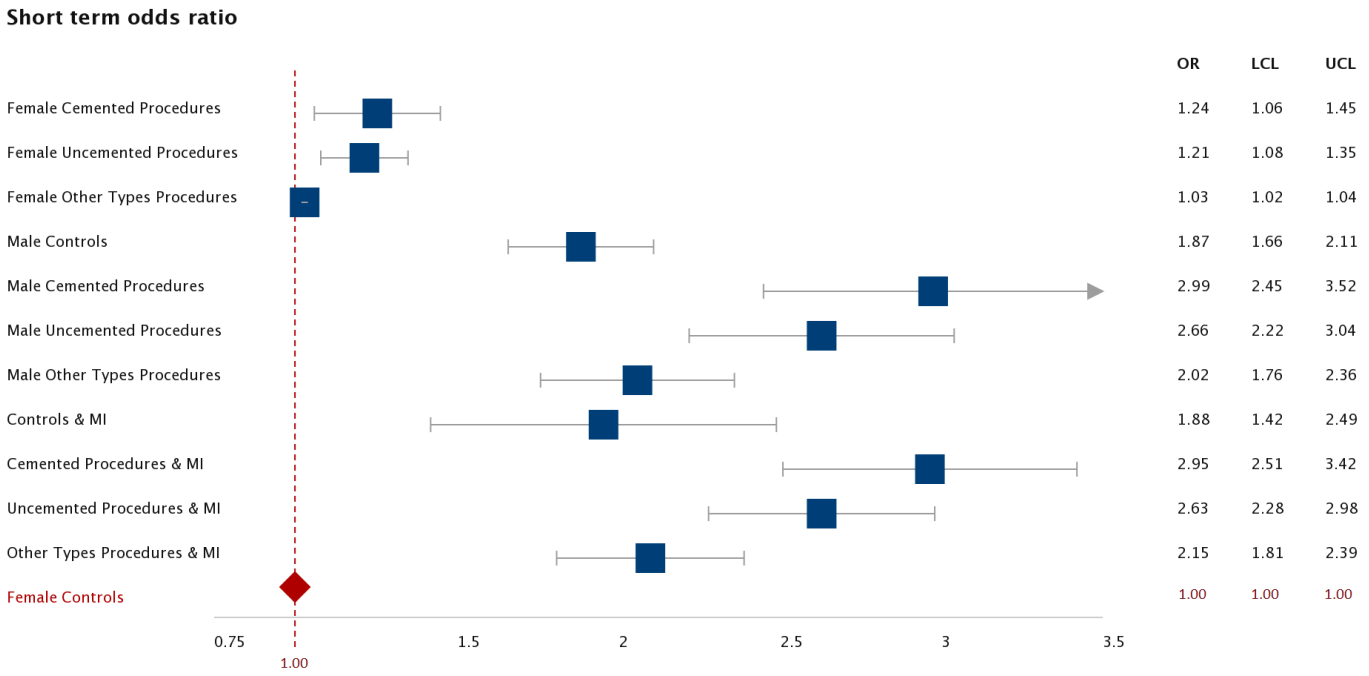
**M1 represents patients with normal cholesterol level due to statins prescription prior to THR surgery. M2 are patients with HC despite the statins prescription prior to THR surgery. M3 are patients with HC but not on statins prescription prior to THR surgery.*

*** RIV indicates the reduction in value of estimated model coefficients if the results of the 10 imputed dataset were not combined into a single effect size using Rubin's rule.*

Preoperative angina, stroke and osteoarthritis increased short term adjusted OR of death to 1.33 (1.16,1.53), 1.40 (1.13,1.73) and 1.19 (1.05,1.36), respectively, compared to patients without these medical conditions in this study (OR=1.00). Preoperative hypercholesterolemia did not cause any significant difference in OR of death between cases and controls. Relative to patients with no HC (OR=1.00), OR of death increased to: (1) 1.38 (1.09,1.76) for patients with normal cholesterol level due to statins prescription prior to THR surgery (Category M1), (2) 1.54 (1.27,1.88) for patients with HC despite the statins prescription prior to THR surgery (Category M2) and (3) 1.82 (1.62,2.04) for patients with HC but not on statins prescription prior to THR surgery, respectively. Furthermore, OR of death for patients with preoperative type 2 diabetes did not vary significantly between THR cases and matched controls, but differed between female (OR=1.00) and male [OR=2.40 (1.95,2.72)] patients in this study.

$model_{imp}$ shows significant interaction between types of THR procedures and (1) gender and (2) MI, respectively (Table 15). This indicates that OR of death varies between male and female THR cases and matched controls. Figure 21 shows the variation in adjusted OR of death for cases and controls for both genders. Relative to female controls [OR=1.00], female THR cases have a higher OR of death for cemented procedures [OR=1.24 (1.06,1.45)], uncemented procedures [OR=1.21 (1.08,1.35)] and other types of procedures [OR=1.03 (1.02,1.04)], respectively. Furthermore, male controls [OR=1.87 (1.66,2.11)] and male THR cases [OR=2.99 (2.45,3.52) for cemented procedures, OR=2.66 (2.22,3.04) for uncemented procedures and OR=2.02 (1.76,2.36) for other types of procedure, respectively] have a higher OR of death than female controls [OR=1.00]. Thus, being male increases the short term OR of death for controls and cases across all types of THR procedures, with cemented procedures having the highest risk of mortality, followed by uncemented and other types of THR procedures. Relative to controls with no MI (OR=1.00), OR of death increases for controls with MI [OR=1.88 (1.42,2.49)] and THR cases with MI [OR=2.95 (2.51,3.42) for cemented procedures, OR=2.63 (2.28,2.98) for uncemented procedures and OR=2.15 (1.81,2.39) for other types of procedure, respectively]. Hence having an event of MI before the surgery, increases OR of death for all types of THR procedures.

Figure 21: Effects of procedure types, gender and myocardial infarction on short term OR of death after THR procedure

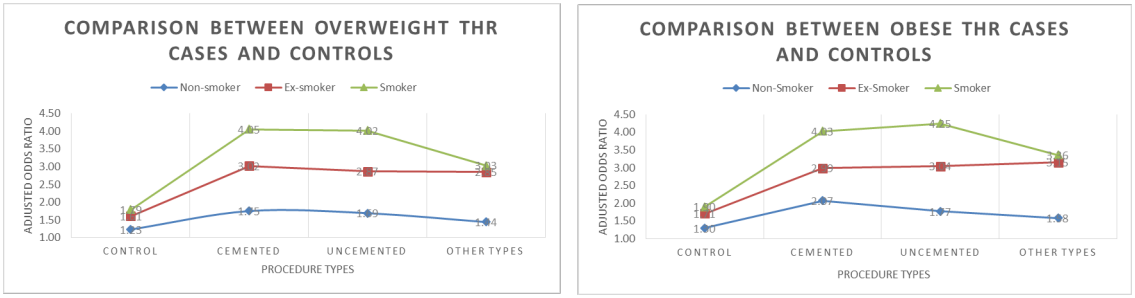


There was significant difference in OR of THR cases and matched controls for both genders and with pre-operative myocardial infarction. Being a male and having a history of preoperative MI before the surgery increased short term OR of death for controls, cemented, uncemented and other types of procedures. All OR of death presented in this Forest plot are estimated relative to a female controls with no preoperative history of MI, adjusted for age at surgery and preoperative comorbidities. The reference level (OR=1.00) represents female controls without MI.

Variations in short term OR are furthermore explained by significant interactions between types of THR procedures and BMI, and also between smoking status and BMI, respectively. This shows that preoperative BMI and smoking status has strong effects on the estimated OR of death of different THR procedures. Relative to controls who are non-smokers with normal BMI [OR=1.00], a higher OR of death is estimated among obese and overweight controls and THR cases for all types of procedures. Figure 23 below shows the difference in adjusted OR of death for the first 24 months after the surgery for cases and controls with different level of preoperative BMI and smoking status.

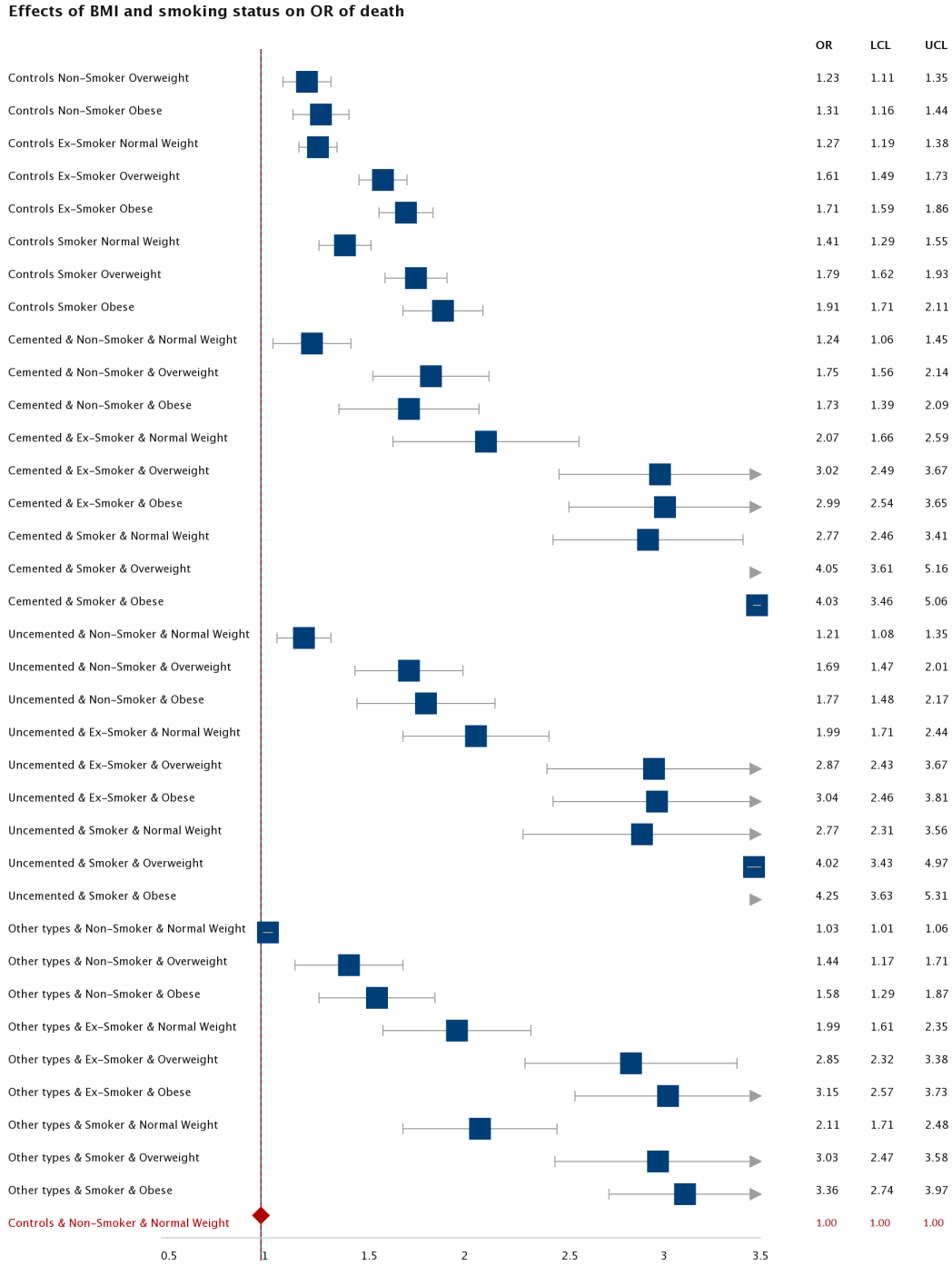
Figure 22 below displays the movement of OR of death for THR cases and matched controls across different types of BMI and smoking status. Among overweight patients, cemented procedures are associated with the highest OR of death, followed by uncemented procedures, other types and matched controls for both ex-smokers and smokers, respectively. However among obese patients, uncemented procedures have the highest estimated OR of death, followed by cemented procedures, other types of procedures and matched controls among ex-smokers and smokers, respectively. Hence obesity causes the highest increase in short term OR for uncemented procedures, while being overweight pre-operatively yields the highest increase in OR of death for cemented procedures, for both ex-smokers and smokers, respectively.

Figure 22: Comparing OR of death for different types of THR procedures



Movement of OR for cases and controls across BMI types and smoking status. For overweight patients, cemented procedures are associated with the highest OR of death, for all types of smoking status, while for obese patients, uncemented procedures had the highest OR of death for smokers and ex-smokers.

Figure 23: Variations in short terms OR of death due to second order interaction between procedures types, BMI and smoking status



Significant interactions between types of procedures and BMI, and between BMI and smoking status showed additional source of variability in short term OR of death. Being overweight and obese raised OR of death for all cases and controls, compared to normal BMI patients. Relative to non-smokers, OR were higher for ex-smokers and smokers.

OR of death within the first 24 months varied considerably between GP practices. Variance of random effects due to GP practice is estimated at 0.021 and the 95% confidence interval of the short term OR of death associated with GP practice is 0.88 to 6.91. For some GP practices, estimated OR of death shows better mortality risk (OR$<$ 1.00) and worse for different GP practices (OR$>$ 1.00). Adjusting for the age groups, gender, lifestyle factors and comorbidities of the patients provides a more precise estimation of OR of death among patients in this study. This shows the importance of accounting for all possible sources of variations in the data. In addition, a general practice can serve different types of patients with regards to the deprivation level of their residential areas (Townsend scores) and socio-economic group (Mosaic group). However, in this analysis, these factors do not significantly explained variations in short term odds of death following THR procedure. Thus, short term mortality after THR does not significantly depend on the deprivation level and socio-economic group of the patients. Perhaps this effect was confounded with the effect of GP practice on short term OR of death in this analysis.

Multiple imputation model diagnostics and performance assessment

In this section, the multivariate logistic regression model with interaction terms fitted to estimate short term mortality after THR is assessed using diagnostic plots firstly to check non-linearity among predictors using Pearson residual plots and secondly to check for the presence of outliers that can be influential. Figures 36 and 37 in Appendix C, show the plots of the Pearson residuals for each predictor kept in the short term multivariate logistic regression model in this study. Visual inspection of these plots demonstrate no trend associated with each predictor. Figure 38 in Appendix C displays the diagnostic plots combining Cook's distance, studentized residuals and hat-values. Observations indexed as 17154, 22000 and 45756 are most likely to be outliers since they have the largest studentized residuals. To assess whether these three observations are influential, a plot of studentized residuals against hat-values is used as shown in Figure 39 in Appendix C, in which the size of the circle around each observation point is proportional to Cook's distance. Observations indexed as 17154, 22000 and 45756, respectively, have the largest Cook's distance, but moderate hat values.

According to Martin and Pardo (2009), 0.2 is suggested as the threshold for acceptable leverage, and Cook's distance should not be greater than $4/(N-p-1)$, where N is the number of observations and p is the number of predictors in the model, respectively. In this study, the maximum value for leverage and Cook's distance are estimated at 0.067 and 3.15×10^{-4}, respectively, and are therefore, smaller than the cut off values suggested by Martin and Pardo (2009). Estimated Cook's distance of 3.15×10^{-4} is associated with observation indexed at 45756. As a sensitivity check, this observation was removed and the new estimates of the short term logistic regression model were obtained and compared to the results of the model that included observation indexed as 45756. There are very small differences of order 1×10^{-4} between the estimates of the two models. Removal of the suspected influential observation from the model does not significantly affect the results of the model and thus, it can be concluded that no outliers are influential in the model fitted to estimate short term OR of death within the first 24 months after THR procedures.

The McFadden's pseudo R^2 value (Gordon (2012)) is estimated between 18.5% and 22.1% and thus, the fitted model explains 18.5% to 22.1% of the variations in short term odds of death following THR in this study. Similarly, the Hosmer-Lemeshow test is used to assess the goodness of fit of the short term logistic model. For 10 separate segments of the dataset used to fit the model, the test statistic for the Hosmer-Lemeshow test is estimated at 19.54 with p-value equal to 26.52%. Therefore, there is no significant difference between the predicted OR of death among observed proportions of deaths and that of the data set, indicating that the multivariate logistic regression model is a good fit for the data.

Comparing performance of full case analysis to imputed data set analysis and to the model with main effects only

Table 15 shows the summary of the full case analysis model denoted by $model_{full}$ and fitted to observations with complete records only (N= 74,787 patients). Both models, $model_{imp}$ and $model_{full}$, respectively, include the same predictors and interaction terms. Relative to estimated OR of death of $model_{imp}$, estimated OR of death from $model_{full}$, at the baseline level, are lower by 0.05% for controls and 0.04% for other types of procedures, respectively, and higher for cemented and uncemented procedures by 0.04% and

0.01%, respectively. There are marginal differences between the estimated OR of death of $model_{full}$ and $model_{imp}$, respectively, because the full case analysis may include patients that are on average more sick than those excluded from the analysis due to missing records. Completeness of primary care records depends heavily on the ill-health status of a patient and thus, sicker patients are more likely to have more complete primary care data records as they visit their GP practice more often.

In addition, the McFadden's pseudo R^2 value (Gordon (2012)) for $model_{full}$ is estimated between 11.5% and 14.1% (versus 18.5% to 22.1% for $model_{imp}$) and thus, the $model_{full}$ explains 11.5% to 14.1% of the variations in short term odds of death following THR. Hence $model_{imp}$ explains a higher degree of variations in short term OR following THR, compared to $model_{full}$. Moreover, for 10 separate portions of the dataset used to fit the model, the test statistic for the Hosmer-Lemeshow test is estimated at 11.36 with p-value equal to 16.08%. Therefore, $model_{full}$ is also a good fit for the data. However the Hosmer-Lemeshow test statistic of $model_{imp}$ (test statistic=19.54%, p-value=26.52%) is higher and more significant than that of $model_{full}$, suggesting that $model_{imp}$ performs better than $model_{full}$.

Comparing the model with main effects only, denoted by m_{imp} (section 6.4.2) to the model including second order interaction terms, denoted by $model_{imp}$ (section 6.4.3), this section outlines the main differences between these models and provides a discussion on why $model_{imp}$ represents the best approach in estimating variations in adjusted OR of death within 24 months for THR cases and matched controls. m_{imp} only provides an overall estimate of the OR of death for cases and controls, but does not provide a deeper insight about which preoperative characteristics or comorbidities cause significant variation in OR of death for different types of THR procedures, relative to matched controls. $model_{imp}$ shows that gender, smoking status, BMI, type 2 diabetes and MI cause significant variations among sub-categories of patients in this study. This improves the proportion of explained variability in the estimation of OR of death.

6.4.4 Assessing the impact of patients transferred out

Table 8 in section 6.1 shows the proportion of patients who are transferred out to new GP practice during the first 24 months. In total, 37 cases and 81

controls transferred out during the short term investigation period. These patients are censored and information on their death status are thus, unknown. Models m_{imp} and $model_{imp}$, respectively, assume that patients transferred out, are still alive at the end of the investigation period. To assess the validity of this assumption, a sensitivity analysis whereby the patients transferred out are excluded from the analysis, is carried out and presented in this section. This will aid in assessing the impact of transfer out patients on the estimated OR of death.

Firstly the 118 transferred out patients are excluded from the dataset and multiple imputation method (similar to section 6.2) is carried out again to generate ten imputed datasets. A similar imputation model design to section 6.2, is used to impute missing values for this sensitivity analysis. A multivariate logistic regression model with frailty effect to account for the grouping effect of GP practice, given by equation (3.74), is fitted using the 10 imputed datasets to estimate the effect of THR procedure on all-cause OR of death within the first 24 months after the procedure. A full model with all variables listed in Table 44 from Appendix C fixed as main effects and that also includes the interaction terms listed in section 6.4.3, is set up. Backward elimination is then carried out firstly by excluding, one at a time, interaction terms having no significant effect (p-value $> 1\%$) on the OR of death from the model and secondly, by removing main effect covariates, one at a time, that does not statistically explain variation in the OR of death (p-value $> 5\%$) in order to achieve the most efficient and economical model.

The final model obtained, is labelled as $model_{trans}$ and includes the same main effect and interaction terms as $model_{imp}$. Table 16 provides a comparison of the estimated coefficients of $model_{trans}$ and $model_{imp}$, respectively. Δ_{trans}, the percentage difference between estimated coefficients of $model_{trans}$ and $model_{imp}$, respectively, is determined using equation (6.1).

$$\Delta_{trans} = \frac{\hat{C}_{imp} - \hat{C}_{trans}}{\hat{C}_{imp}} \times 100\% \ , \tag{6.1}$$

where $\hat{C}_{imp}$ and $\hat{C}_{trans}$ are the estimated coefficients of $model_{trans}$ and $model_{imp}$, respectively. Estimated Δ_{trans} varies between -0.99% and 1.97% (Table 16). Therefore, relative to $\hat{C}_{imp}$, excluding patients who are transferred out before the end of the investigation period overestimates $\hat{C}_{trans}$ by

a maximum value of 0.99% for the following covariates in the model: (1) year of birth category, (2) smoking status, (3) hypercholesterolemia status and (4) interaction between procedure types and smoking status, respectively, while, the remaining $\hat{C}_{trans}$ are smaller than $\hat{C}_{imp}$ by a maximum of 1.97%.

In addition, the McFadden's pseudo R^2 value (Gordon (2012)) for $model_{trans}$ is estimated between 18.2% and 21.8%, versus 18.5% to 22.1% for $model_{imp}$. Therefore $model_{trans}$ only causes a marginal decrease in explained degree of variations in short term OR following THR, compared to $model_{full}$. This reduction can mainly be associated to the exclusion of 118 observations from the analysis. Moreover, for 10 separate portions of the dataset used to fit $model_{trans}$, the test statistic for the Hosmer-Lemeshow test is estimated at 17.76 with p-value equal to 28.08%. Therefore, $model_{trans}$ is also a good fit for the data.

As a check for non-linearity among predictors in $model_{trans}$, visual inspection of Pearson residual plots demonstrates no trend associated with each predictor kept in $model_{trans}$. Similarly, graphical check of the plot of studentized residuals against Cook's distance reveals the presence of six outliers in $model_{trans}$. These outliers are not influential because they have low Cook's distance. The maximum value of Cook's distance is estimated at 2.1×10^{-02} and is associated with observation 65152. Exclusion of this observation from the analysis as a sensitivity check, does not impact on the estimated parameter of $model_{trans}$.

As a conclusion, this sensitivity analysis shows that exclusion of patients who are transferred out before the end of the investigation period causes marginal differences in the estimated coefficients of $model_{trans}$, when compared to the estimated coefficients of $model_{imp}$. There are also marginal differences in the performance and model fit of $model_{trans}$ and $model_{imp}$. Therefore keeping transferred out patients in the study and assuming that they stay alive until the end of the study, is a reasonable assumption that impacts marginally on the estimation of short term OR of death.

Table 16: Comparing model coefficients of imputed datasets analysis with and without patients transferred out before end of study

Variables	Levels	Model including transferred out patients						Model excluding transferred out patients						% Difference
		Est, $\hat{C}_{imp}$	Std.Er	t.value	df	P-value	RIV	Est, $\hat{C}_{trans}$	Std.Er	t.value	df	P-value	RIV	in Est, Δ_{trans}
(Intercept)		-0.6783	0.261	-2.60	1	4.1E-02	3.8E-03	-0.6768	0.034	-2.49	1	4.4E-02	1.0E-02	0.21%
Sex	Male	0.6244	0.059	8.18	1	4.7E-03	2.9E-03	0.6166	0.032	8.26	1	4.6E-03	6.6E-03	1.25%
Age at surgery	Unit increase	0.0542	0.003	5.52	1	1.0E-02	3.5E-03	0.0548	0.004	5.61	1	9.8E-03	6.7E-03	1.20%
Year of Birth Category	1925-29	-0.4299	0.106	-4.05	1	1.8E-02	2.2E-03	-0.4266	0.021	-3.95	1	1.9E-02	9.6E-03	0.78%
	1930-34	-0.4432	0.162	-3.24	1	2.8E-02	3.3E-03	-0.4451	0.022	-3.17	1	2.9E-02	3.2E-03	-0.44%
	1935-40	-0.4561	0.182	-2.50	1	4.4E-02	3.4E-03	-0.4521	0.023	-2.45	1	4.5E-02	7.2E-03	0.89%
Smoking status	Ex-smoker	0.2420	0.143	3.08	1	3.0E-02	3.2E-03	0.2444	0.012	3.09	1	3.0E-02	3.0E-03	-0.99%
	Smoker	0.3430	0.105	6.00	1	8.6E-03	4.2E-03	0.3426	0.017	6.09	1	8.4E-03	7.6E-03	0.12%
BMI	Overweight	0.2073	0.089	-3.34	1	2.6E-02	7.0E-03	0.2047	0.010	-3.24	1	2.8E-02	5.0E-03	1.26%
	Obese	0.2590	0.074	-2.57	1	4.2E-02	3.9E-03	0.2551	0.013	-2.54	1	4.3E-02	2.3E-03	1.51%
Type 2 Diabetes	Yes	0.7114	0.197	3.61	1	2.3E-02	2.8E-03	0.7035	0.035	3.62	1	2.3E-02	4.6E-03	1.11%
Angina	Yes	0.2843	0.071	4.01	1	1.9E-02	4.7E-03	0.2804	0.014	4.06	1	1.8E-02	7.1E-03	1.37%
Myocardial Infarction (MI)	Yes	0.6319	0.143	4.41	1	1.6E-02	2.3E-03	0.6239	0.031	4.51	1	1.5E-02	2.4E-03	1.27%
Stroke	Yes	0.3354	0.109	3.09	1	3.0E-02	3.2E-03	0.3339	0.017	3.11	1	3.0E-02	9.7E-03	0.44%
Osteoarthritis	Yes	0.1779	0.065	2.72	1	3.8E-02	3.7E-03	0.1747	0.010	2.76	1	3.7E-02	7.0E-03	1.79%
Hypercholesterolemia	Hypercholesterolemia M1*	0.3252	0.122	2.66	1	3.9E-02	4.7E-03	0.3259	0.016	2.72	1	3.8E-02	1.5E-03	-0.24%
	Hypercholesterolemia M2*	0.4344	0.099	4.38	1	1.6E-02	2.3E-03	0.4354	0.022	4.47	1	1.5E-02	1.0E-02	-0.23%
	Hypercholesterolemia M3*	0.5981	0.058	10.26	1	3.0E-03	5.0E-03	0.6000	0.030	10.32	1	3.0E-03	1.6E-03	-0.31%
Procedure types	Cemented procedures	0.2145	0.081	2.64	1	4.0E-02	4.4E-04	0.2116	0.011	2.73	1	3.8E-02	9.4E-03	1.37%
	Uncemented procedures	0.1859	0.057	3.28	1	2.7E-02	3.0E-03	0.1855	0.010	3.36	1	2.6E-02	1.2E-03	0.22%
	Others procedure types	0.0294	0.005	6.53	1	7.3E-03	4.5E-03	0.0290	0.002	6.58	1	7.2E-03	9.0E-03	1.34%
Procedure types*Sex	Male Cemented procedures	0.2572	0.075	3.41	1	2.5E-02	2.9E-03	0.2547	0.013	3.49	1	2.4E-02	7.3E-03	0.97%
	Male Uncemented procedures	0.1698	0.057	2.95	1	3.3E-02	3.4E-03	0.1672	0.009	3.03	1	3.1E-02	2.0E-03	1.55%
	Male Other procedures	0.0499	0.017	3.02	1	3.2E-02	3.3E-03	0.0493	0.003	3.03	1	3.1E-02	4.5E-03	1.12%
Procedure types*Smoking status	Ex-smoker Cemented procedures	0.2708	0.050	5.37	1	1.1E-02	1.9E-03	0.2711	0.013	5.43	1	1.0E-02	9.1E-03	-0.11%
	Smoker Cemented procedures	0.4606	0.301	2.19	1	5.5E-02	8.6E-03	0.4615	0.023	2.27	1	5.2E-02	9.9E-03	-0.21%
	Ex-smoker Uncemented procedures	0.2615	0.061	4.25	1	1.7E-02	2.4E-03	0.2617	0.013	4.26	1	1.7E-02	4.2E-03	-0.07%
	Smoker Uncemented procedures	0.4895	0.138	3.54	1	2.4E-02	4.8E-03	0.4900	0.025	3.59	1	2.3E-02	2.6E-03	-0.09%
	Ex-smoker Other procedures	0.4149	0.154	2.70	1	3.8E-02	3.7E-03	0.4153	0.021	2.79	1	3.6E-02	9.6E-03	-0.10%
	Smoker Other procedures	0.3687	0.259	2.65	1	4.0E-02	3.8E-03	0.3693	0.018	2.76	1	3.7E-02	1.0E-02	-0.15%
Procedure types*BMI	Overweight Cemented procedures	0.1402	0.015	9.06	1	3.8E-03	1.1E-03	0.1393	0.008	9.15	1	3.8E-03	9.6E-03	0.59%
	Obese Cemented procedures	0.0747	0.028	2.64	1	4.0E-02	4.8E-03	0.0742	0.004	2.71	1	3.8E-02	1.0E-02	0.75%
	Overweight Uncemented procedures	0.1309	0.089	3.44	1	2.5E-02	2.9E-03	0.1300	0.007	3.52	1	2.4E-02	3.1E-03	0.68%
	Obese Uncemented procedures	0.1279	0.067	4.08	1	1.8E-02	2.5E-03	0.1270	0.006	4.08	1	1.8E-02	8.8E-03	0.70%
	Overweight Other procedures	0.1260	0.079	3.32	1	2.6E-02	4.0E-03	0.1249	0.007	3.38	1	2.6E-02	4.9E-03	0.91%
	Obese Other procedures	0.1687	0.049	3.45	1	2.5E-02	2.9E-03	0.1658	0.009	3.50	1	2.4E-02	6.3E-04	1.70%
Procedure types*MI	Cemented procedures with MI	0.2365	0.030	7.93	1	5.0E-03	1.3E-03	0.2330	0.012	7.97	1	4.9E-03	9.0E-03	1.47%
	Uncemented procedures with MI	0.1490	0.042	3.54	1	2.4E-02	2.8E-03	0.1476	0.008	3.54	1	2.4E-02	7.0E-03	0.98%
	Others procedures types with MI	0.1058	0.184	4.38	1	1.6E-02	2.3E-03	0.1055	0.005	4.40	1	1.6E-02	8.5E-03	0.32%
Sex*Diabetes	Male Type 2 Diabetes	0.2525	0.074	3.40	1	2.5E-02	2.9E-03	0.2507	0.013	3.47	1	2.4E-02	7.6E-03	0.69%
BMI*Smoking status	Overweight Ex-smoker	0.0272	0.106	2.57	1	4.2E-02	3.9E-03	0.0269	0.002	2.62	1	4.0E-02	1.8E-03	1.23%
	Overweight Smoker	0.0341	0.176	1.94	1	6.7E-02	9.1E-03	0.0341	0.002	1.99	1	6.4E-02	9.0E-03	0.06%
	Obese Ex-smoker	0.0341	0.070	4.88	1	1.3E-02	2.1E-03	0.0341	0.002	4.92	1	1.3E-02	9.9E-03	0.05%
	Obese Smoker	0.0419	0.131	3.20	1	2.8E-02	3.1E-03	0.0414	0.002	3.29	1	2.7E-02	1.0E-04	1.05%

**M1 represents patients with normal cholesterol level due to statins prescription prior to THR surgery. M2 are patients with HC despite the statins prescription prior to THR surgery. M3 are patients with HC but not on statins prescription prior to THR surgery.*
*** RIV indicates the reduction in value of estimated model coefficients if the results of the 10 imputed dataset were not combined into a single effect size using Rubin's rule.*
Δ_{trans} is the percentage difference between estimated $\hat{C}_{imp}$ from $model_{imp}$ and $\hat{C}_{trans}$ from $modelC_{trans}$, respectively.

6.5 Discussion and validation of results

In this chapter, adjusted OR of death for the first 24 months after THR procedures for UK patients born between 1920 and 1940, inclusive, are estimated using a multilevel logistic regression model. Estimated OR of death for cases with no preoperative medical comorbidities undergoing cemented, uncemented and other types of THR procedures are higher than estimated OR for matched controls, higher for male cases compared to female ones and yearly increase in age of patients at surgery, increases short term OR, respectively (Table 15 and Figure 21). These results are in agreement with the following publications which reported that for the first 24 months after the surgery, survival of male THR cases is worse than that of female THR cases, matched controls and increases for patients undergoing the procedure at older ages: Fender et al. (1997), Lie et al. (2002), Nunley and Lachiewicz (2003), Blom et al. (2006), Pedersen et al. (2011), Jones et al. (2014), Lie et al. (2000), Williams et al. (2002), Ramiah et al. (2007), Aynardi et al. (2009), NHS Scotland (2002), Lie et al. (2000), Hunt et al. (2013), NHS Scotland (2002) and Lovald et al. (2014).

In addition, the Kaplan Meier survival analysis revealed that the excess mortality for THR cases peaks during the first 3 months and then decreases and converges to that of matched controls between 12 to 24 months after the procedure, for both genders(Figure 16). Barrett et al. (2005) found that duration of excess mortality associated with THR cases, compared to age and gender matched controls, peaks at about 30 to 60 days after the surgery and then decreases slowly to zero until about 90 days after the surgery. Furthermore Lie et al. (2002) reported a similar trend with the duration of excess mortality in the short term continuing up to 24 months after the procedure. Hence estimated duration of excess mortality following THR in this study is consistent with published articles.

Compared to controls with normal BMI, this study concluded that being overweight prior to THR increases short term OR of death for all types of THR procedures. This result is in accordance to the studies by Bozic et al. (2012) and Memtsoudis et al. (2012), Bozic et al. (2012) and Memtsoudis et al. (2012), respectively. Preoperative obesity on its own increases OR of death for all types of THR procedures, compared to controls with normal BMI in this study. Furthermore being overweight and obese pre-operatively

increased short term OR of death for all types of THR procedures among ex-smokers or smokers prior to their THR procedure, respectively, compared to non-smoking controls with normal BMI (Figure 23). The results of this short term study also revealed significant effects on short term OR of death due to the interaction between BMI and smoking status among cases and controls. No study in literature reviewed in Chapter 2 reported effects of preoperative smoking status and BMI, either as main effects or as interaction effects on short term OR of death after THR.

Pre-surgery myocardial infarction (MI) significantly increases short term OR of death for all types of THR procedures compared to controls with no history of MI. This result is in agreement with Hunt et al. (2013), which estimated a three-fold increase for mortality after THR for cases with preoperative events of MI, relative to controls without a history of MI. Similarly, Comba et al. (2012) reported that a history of cardiovascular diseases prior to THR procedure increased the risk of mortality of THR cases eight-fold, relative to controls without cardiovascular diseases. However the authors did not distinguish between the types of cardiovascular disease as carried out in this study for different types of THR procedures.

Finally this study found no significant difference in short term OR of death of cases and matched controls with either preoperative angina, stroke, hypercholesterolemia (HC) or type 2 diabetes. However these comorbidities increase odds of death of patients given these preoperative medical conditions, compared to those without. Osteoarthritis is the main reason for THR in this study but for cases with OA, OR of death for cemented, uncemented and other types of procedures do not differ significantly. Thus, OA increases OR of death to the same level for all patients in this study. Similar finding is shown in studies by Fender et al. (1997), Lie et al. (2000), Williams et al. (2002), Barrett et al. (2005), Pedersen et al. (2011), Xu et al. (2017) and Boniello et al. (2017), respectively.

Among the preoperative medical conditions that do not vary significantly for different type of procedures in this study, only the effect of preoperative type 2 diabetes on short term mortality at 3 months after the procedure, was investigated by Pedersen et al. (2011), who also reported no significant difference in mortality between cases and controls diagnosed with type 2 diabetes before the surgery.

6.6 Strengths and limitations

In this short term investigation, primary care data that are routinely collected in the THIN database, are used to estimate odds of death at 24 months after THR. THIN data is representative of the UK population for various demographic, lifestyle factors and medical conditions (Blak et al. (2011), Langley et al. (2011), González et al. (2009), Hippisley-Cox and Coupland (2010a), MacDonald and Morant (2008)). Therefore primary care database such as THIN provides good coverage of THR cases during the study period and also allows to generalise the findings of this study to the general UK population.

This study is designed as a matched cohort study. This allows to directly estimate the effect of different types of THR procedures on short term odds ratio of death of cases compared to controls, matched on gender, year of birth category and GP practice, while adjusting for a number of confounders such as preoperative lifestyle factors, demographic covariates, comorbidities and interactions between each of these confounders. In literature reviewed in Chapter 2, no previous study has adjusted for so many confounders. Furthermore, this study found several preoperative lifestyle factors (smoking status) and medical conditions (obesity and MI), that significantly cause a difference in odds ratio of death for cases undergoing different types of THR procedures. These findings may be used by healthcare professionals for medical and risk management for THR cases and patients waiting to undergo THR, respectively.

However, no information describing the type of prosthesis, surgical approach and surgeon experience is available in THIN. These variables have significant effect on survival after THR procedure (Whitehouse et al. (2014)). Estimated odds of death in this analysis are not adjusted for these variables and survival after THR could not be differentiated for various types of prosthesis, surgical approach and surgeon experience. Additionally, patients who are transferred out before the end of the two years investigation period are lost to follow-up and their death status is unknown for the remaining period of the investigation. However these patients are assumed to be alive until the end of the two year period. The strength of this assumption has been tested using a sensitivity analysis whereby it is shown that exclusion of transfer out patients from the analysis does not impact significantly on the estimated odds ratio of death.

Variables related to pollution levels and ethnicity are not available for patients from Scotland and Northern Ireland. Pollution levels and ethnicity were thus not included in the analysis since these variables had values not missing at random. However missing values for variables that were missing at random for all English, Welsh, Northern Irish and Scottish patients (BMI, Townsend scores, IMD, smoking status and Mosaic groups) were handled by employing the multiple imputation method. Finally, estimated short term OR was adjusted for drug prescriptions given prior to THR procedure. One limitation with this approach is that patient's adherence to the drug therapy is unknown, thereby not precisely reflecting the effects of drug therapy on short term mortality following THR.

6.7 Conclusions

The study described in this chapter estimated the adjusted all-cause odds of death at 24 months after total hip replacement procedures. Odds ratio of death for THR cases were higher than matched controls for both genders and increased for with age at surgery. Short term odds ratio of death was the highest for cemented procedures, followed by uncemented and other types of procedures. Among preoperative risk factors that significantly impacted on short term odds ratio of death, obesity, smoking and myocardial infarction increased odds of death for all types of THR procedures, compared to cases without these preoperative conditions and lifestyle factors.

7 Long Term Survival Analysis After Primary Total Hip Replacement

Chapter outline

This chapter presents the long term survival analysis after had a total hip replacement procedures. The dataset used for this analysis is described in Chapter 5 and represents patients who survived their first two years after the surgery. Firstly, Kaplan-Meier survival analysis is carried out as preliminary analysis to compare the survival between THR cases and matched controls for different gender, age groups and THR procedure types. Secondly, the full procedure to carry out the multiple imputation of missing data is described and assessed. Thirdly, multilevel Cox's regression analysis is used to fit survival models for full case data, using complete records only and for survival analysis of the multiple imputed datasets, to compare the results of these two approaches. A discussion on the validity of each Cox's regression models, evaluated in terms of performance, both internally and externally, is then provided. These survival models explain variations in the estimated hazard ratios between cases and controls and also between various general practices in the UK. Finally, the results of these survival models are presented using forest plots and their validity is discussed in details.

7.1 Model development strategy

7.1.1 Patients exclusion

According to the review of 15 published studies carried out in Chapter 2, section 2.3, long term survival analysis after total hip replacement (THR) procedures included patients who only survived at least one year after the procedure. This is because in the early post-operative period after surgery, there is an excess mortality risk associated with THR cases, compared to controls. This excess mortality disappears between 12 and 24 months after the procedure. Therefore to allow for this change in mortality among THR cases, several authors (see section 2.3) developed their long term survival models conditional on the assumption that the patients survived the short term period where their risk of mortality was highest.

In this research, the follow-up period for the short term study (Chapter

6) was defined as the first 24 months after the surgery. Thus, patients who died or transferred out within that follow-up period, were excluded from the dataset used in this long term study. Table 17 below shows the proportion of THR cases and controls who were transferred out or died within the first 24 months after the procedure and in those surviving more than 24 months. In total, 2006 patients died or transferred out in the first 24 months of follow-up and they are excluded from the dataset used in this chapter. Therefore, all survival models developed in this long term analysis are conditional on the patients staying in the study for at least 24 months after the surgery.

Table 17: Distribution of transfers out and deaths after THR procedures in the first 24 months and after 24 months

Gender		Age group	0-2 years follow-up					2+ years follow-up*				
			No. of Patients	Transfers out	%	No. of Deaths	%	No. of Patients	Transfers out	%	No. of Deaths	%
THR	**Male**	**18-54**	177	2	1.13%	0	0.00%	175	100	57.1%	39	22.3%
Cases		**55-64**	1141	3	0.26%	20	1.75%	1118	605	54.1%	298	26.7%
		65-74	2972	7	0.24%	66	2.22%	2899	1144	39.5%	728	25.1%
		74-85	1993	6	0.30%	39	1.96%	1948	730	37.5%	582	29.9%
		85+	228	2	0.88%	5	2.19%	221	60	27.1%	82	37.1%
		All ages	6511	20	0.31%	130	2.00%	6361	2639	41.5%	1729	27.2%
	Female	**18-54**	281	3	1.07%	0	0.00%	278	150	54.0%	62	22.3%
		55-64	1457	2	0.14%	32	2.20%	1423	729	51.2%	309	21.7%
		65-74	4538	6	0.13%	85	1.87%	4447	1794	40.3%	926	20.8%
		74-85	3847	5	0.13%	58	1.51%	3784	1462	38.6%	879	23.2%
		85+	523	1	0.19%	11	2.10%	511	131	25.6%	107	20.9%
		All ages	10646	17	0.16%	186	1.75%	10443	4266	40.9%	2283	21.9%
Controls	**Male**	**18-54**	1002	5	0.50%	0	0.00%	997	497	49.8%	195	19.6%
		55-64	5859	4	0.07%	70	1.19%	5785	2697	46.6%	1252	21.6%
		65-74	14606	11	0.08%	304	2.08%	14291	4884	34.2%	3482	24.4%
		74-85	9851	3	0.03%	229	2.32%	9619	2649	27.5%	3353	34.9%
		85+	1237	2	0.16%	20	1.62%	1215	135	11.1%	528	43.5%
		All ages	32555	25	0.08%	623	1.91%	31907	10862	34.0%	8810	27.6%
	Female	**18-54**	1480	8	0.54%	0	0.00%	1472	689	46.8%	237	16.1%
		55-64	7594	6	0.08%	0	0.00%	7588	3168	41.8%	1232	16.2%
		65-74	22460	13	0.06%	487	2.17%	21960	7467	34.0%	4226	19.2%
		74-85	18978	17	0.09%	415	2.19%	18546	5370	29.0%	5009	27.0%
		85+	2718	12	0.44%	47	1.73%	2659	318	12.0%	948	35.7%
		All ages	53230	56	0.11%	949	1.78%	52225	17012	32.6%	11652	22.3%

**Proportion of transfers out and deaths for various age groups determined after excluding patients who dropped out or died within the first 24 months after the procedure*

7.1.2 Follow-up time, outcome and censoring

Cases join the study period from the time they undergo their THR procedure while controls follow up period starts on the date when their matched THR case undergo their surgery. Cases and controls are both followed up until the earliest of the following outcomes: (1) the end of the study period on 1^{st} July 2011, (2) the date of transfer out to new GP practice and, (3) the date of death. The main outcome of interest in this long term study is time to death after THR procedures. For patients with death dates before

the end of the investigation period (1^{st} July 2011), their follow-up time is completely observed. For patients who are alive at 1^{st} July 2011, the outcome of interest is unknown (right censoring) and their follow-up time is estimated as the difference between the time they underwent their surgery until 1^{st} July 2011. Patients who are transferred out to new GP practice before 1^{st} July 2011, are also right censored as their death information is also unknown and their follow-up time is defined as the difference between the time they joined the study and their transfer out date. Male and female cases have a mean follow-up time of 9.63 years (range: 2.01-52.15 years) and 9.53 years (range: 2.01-58.07 years) respectively. Male and female controls have a mean follow-up time of 9.93 years (range: 2.01-56.08 years) and 9.98 years (range: 2.01-59.50 years) respectively. The overall mean follow-up time for cases and controls combined, is 9.89 years (range: 2.01-59.50 years).

7.1.3 Multiple imputation of missing data

The imputation phase

Table 7 in section 5.2.9 displays the proportions of missing data for BMI, smoking status, Townsend score, Index of multiple deprivation and Mosaic category among THR cases and matched controls, included in this study. Missing values within these variables are assumed to be missing at random (see section 5.2.9). In the first stage of multiple imputation, there are two key aspects that should be considered:

- Defining the measurement scale of the covariates with incomplete records, and
- Defining the appropriate regression model as the imputation model (IM).

In the dataset analysed in this chapter, BMI, smoking status, Townsend scores, Index of multiple deprivation and Mosaic category are categorical in nature with more than two levels. Therefore the appropriate IM model to impute missing records for these variables, is defined as a multivariate linear mixed model (Schafer and Yucel [2002]) and is of the form shown below.

$$y_i = X_i\beta + Z_i b_i + \varepsilon_i,\ i = 1, ..., m,$$

where y_i is the matrix of incomplete multivariate data for the i^{th} cluster, β is the matrix of coefficients common to the population (fixed effects), b_i is

the matrix of coefficients specific to the i^{th} cluster (random effects), X_i is the matrix of covariates for fixed effects, Z_i is the matrix of covariates for random effects b_i and ε_i is the matrix of residual errors.

The covariates with complete and incomplete records, and the cluster variables in this study are categorised as:

- *Fixed effect predictors*: Case-Control Indicator, Sex, Year of Birth Category, Transfer Status of patients, Age group at surgery, Survival time, Type 2 Diabetes, Heart Attack, Angina, Stroke, Chronic Kidney Disease, Rheumatoid arthritis, Osteoarthritis, Oestrogen prescription, Progesterone prescription, Testosterone prescription, Hypercholesterolemia, Hypertension, Beta Blocker prescription, Statin prescription, Calcium Blocker prescription, ACE Inhibitor prescription
- *Target variables*: Townsend Score, BMI Category, Smoking status, Mosaic Category, Index of Multiple Deprivation
- *Cluster variable*: General Practice

According to van Buuren [2012], the IM used to impute the missing records in the data set should include all the variables that are used for data analysis, and having over 25 covariates in the IM hardly influences the explained variance in the imputed datasets. The IM used in this chapter consists of 28 covariates in total and therefore adding any interactions between covariates in the IM will hardly have an impact on the generated datasets.

To carry out the multiple imputation, the package *jomo* (Matteo and James [2017]), is used because it permits to impute data that are hierarchical in nature. The *jomo* package imputes the missing records based on a Bayesian iterative process involving Monte Carlo simulation as described below.

(1) Using the regression coefficients of the IM, missing values for Townsend Score, BMI Category, Smoker, Mosaic Category and Index of Multiple Deprivation, respectively, are predicted. A random residual term which is normally distributed with mean zero and variance equal to the residual variance from the regression of the missing covariate value on the outcome variable, is added to the IM. Adding random residual terms to the mean vector and the covariance matrix of the IM produces parameter estimates that differ randomly to those that produced the

coefficient estimates of the first IM initiated in the first step of the imputation phase. A new dataset (D_{new}) with observed and imputed values is obtained.

(2) Using D_{new}, the new sample means ($\hat{\mu}_{new}$) and the covariance matrix ($\hat{\mathrm{C}}_{new}$) are determined.

(3) Using $\hat{\mu}_{new}$ and $\hat{\mathrm{C}}_{new}$, a new posterior distribution is defined and used to obtain a new set of plausible estimates for the missing values.

(4) Steps (1)–(3) are iterated continuously until convergence of the estimated regression coefficients is achieved. This iteration process of convergence is referred as the burn-in-length and is set to 500 iterations for the imputation of missing records in this dataset.

Once the designated number of burn-in-length has been completed in this dataset, the entire imputation process is repeated ten times to generate ten imputed datasets. The observed data stay the same across the ten imputed datasets. Only the values that had originally been missing will differ. Table 18 below compares the average distribution of patients across the ten imputed datasets generated to the dataset with full case records only. All imputed variables have a distribution close to that of the incomplete dataset.

Table 18: Comparing average distribution of imputed variables across ten imputed datasets to incomplete full cases only, dataset for long term survival analysis

Variables	Levels	Incomplete Dataset*	Average Distribution (SD) in imputed dataset**
Townsend Score			
	1	29.5%	29.3% (5%)
	2	23.8%	22.8% (1%)
	3	20.9%	20.7% (1%)
	4	14.3%	15.6% (2%)
	5	11.5%	12.6% (3%)
	Missing**	16.3%	0%
Body Mass Index			
	Normal	45.6%	46.6% (2%)
	Overweight	33.6%	32.5% (5%)
	Obese	20.8%	21.6% (1%)
	Missing**	9.9%	0%
Smoking			
	Non-smoker	62.6%	68.8% (3%)
	Ex-smoker	13.8%	14.1% (2%)
	Smoker	23.6%	24.3% (1%)
	Missing**	26.3%	0%
Mosaic Categeory			
	Cat A-C	36.5%	36.6% (3%)
	Cat D-F	17.4%	17.4% (5%)
	Cat G-I	23.2%	23.2% (1%)
	Cat J-O	22.9%	23.0% (4%)
	Missing**	7.8%	0%
Index of Multiple Deprivation			
	1	23.3%	23.4% (4%)
	2	21.9%	22.0% (1%)
	3	20.1%	20.1% (1%)
	4	18.1%	18.1% (1%)
	5	16.6%	16.6% (2%)
	Missing**	2.2%	0%

**Proportions determined out of a total of 74,762 patients with complete records only.*

***Average proportions calculated out of a total of 100,936 patients with complete and imputed data, across 10 imputed datasets.*

Average distribution of Townsend scores, BMI, smoking status, Mosaic category and index of multiple deprivation across the 10 imputed datasets, is close to that of the full cases only dataset.

Checking for convergence of imputed regression coefficients and comparing distribution of complete and imputed datasets

For the analysis of imputed datasets to yield reliable results, it must be checked that the iterative algorithm, described above, has converged and that the imputed datasets are approximately independent draws from the predictive distribution (Gill (2014)). The first option to check for convergence is to examine the potential scale reduction factor, also called $\hat{R}$ (Gelman and Rubin (1992)) for the parameters of the imputation model (See section 3.8 for a full theoretical description of $\hat{R}$).

Figure 19 shows the summary output for the multiple imputation carried out using the *jomo* package in R. $\hat{R}$, represented by *Psi* in Figure 19, is estimated at 1.011 and is very close to one, showing that estimated variance within and between imputed datasets is low and close to one. Furthermore, the summary also shows how the estimated coefficients of imputation model, represented by *Beta* in Figure 19, used to predict the missing values, change against the percentage of missing data imputed. It is observed that the estimated *Beta* has a minimum and maximum value of 1.00, indicating that estimated coefficients of the imputation model did not evolve each time a missing variable is imputed during the iterative process. The estimated variance of random effects between the ten imputed datasets have a maximum estimated value of 1.001 (represented by *Sigma* in Figure 24). This means that the estimated variance of random effects between two randomly imputed datasets in this imputation differ by a maximum of 0.001 only.

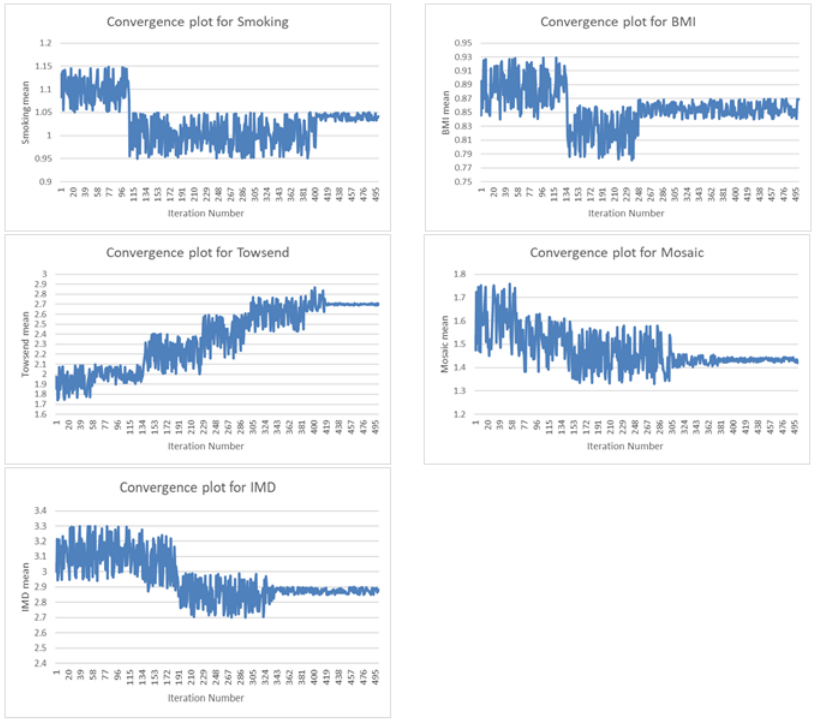
Figure 24: Summary output for multiple imputation of dataset for long term mortality

```
Call:

jomoImpute(data = df.training, type=variable.type, n.burn=500, n.iter=10, m=10)

Cluster variable:         pracid
Target variables:         townsend BMI smoker mosaic_cat IMD
Fixed effect predictors:  (Intercept) cc sex yob_cat xfer status age_group
                          sshort diabetes ha angina stroke ckd ra oa osteo
                          oestrogen progesterone testosterone hypercholesterol
                          hypertension beta.blocker statin calcium.blocker
                          ace.inhibitor
Random effect predictors: (Intercept)

Performed 500 burn-in iterations, and generated 10 imputed data sets,
each 500 iterations apart.

Potential scale reduction (Rhat, imputation phase):

       Min   25%   Mean  Median  75%  Max
Beta:  1.000 1.000 1.000 1.000  1.000 1.000
Psi:   1.000 1.000 1.001 1.000  1.001 1.011
Sigma: 1.000 1.000 1.000 1.000  1.000 1.001
Largest potential scale reduction:
Beta: [1,6], Psi: [1,1], Sigma: [1,1]

Missing data per variable:
    pracid townsend BMI smoker mosaic_cat IMD cc sex yob_cat xfer status age_group sshort diabetes ha angina
MD% 0      15.9     9.7 25.8   7.6        2.0 0  0   0       0    0      0         0      0         0  0
    stroke ckd ra oa osteo oestrogen progesterone testosterone hypercholesterol hypertension beta.blocker
MD% 0      0   0  0  0     0         0            0            0                0            0
     statin calcium.blocker ace.inhibitor
MD%  0      0               0
```

10 datasets were generated after carrying out multiple imputation for missing values in BMI, Townsend scores, Index of multiple deprivation, smoking status and Mosaic group, respectively. The potential scale reduction value (Psi) is estimated at 1.011 and is very close to one, showing that estimated variance within and between imputed datasets is close to one. Estimated coefficients of the imputation model stay unchanged with the number of missing data imputed (Minimum and maximum value of Beta = 1.00). Estimated variance of random effects between the 10 imputed data sets (Sigma) is very low and differs by a maximum of 0.001 across the 10 data sets.

The second option to check convergence of the IM parameters is through visual inspection of diagnostic trace plots that graphically represent the MCMC chain for each imputed covariate at each iteration. Figure 25 below displays the associated convergence plots for each of the imputed variables. The parameters of the IM model have converged to a common value after 500 iterations.

Figure 25: Convergence plots for multiple imputation of dataset for long term mortality



The above convergence plots show that the imputation model parameters used to generate substitute values for missing records in BMI, Townsend scores, Index of multiple deprivation, smoking status and Mosaic group, respectively, have all converged to a common estimate. Hence a burn-in-length of 500 iterations is enough to carry out the multiple imputation for the dataset used to estimate long term hazards of death.

7.2 Kaplan-Meier survival analysis

As a preliminary analysis, Kaplan-Meier survival analysis is carried out to compare the long term survival of patients for each gender, age group and types of THR procedures, respectively. Patients who were transferred out to a new GP practice were censored. To compare the statistical difference between survivorship in each category, the log-rank test (Grambsch and Therneau [1994]), presented in section 3.4, is used.

Figure 26 shows the differences in survival between male and female THR cases and their matched controls who stayed in the study for more than two years after the procedure. It can be observed that the survivorship of female THR cases is better than male cases (p-value of log-rank test=0.0003). Survival of male THR cases is better than male controls (p-value of log-rank test=0.0011) up to about 30 years after the surgery while survival of female cases stays higher than that of female controls (p-value of log-rank test=0.0086) and male controls (p-value of log-rank test=0.0009), respectively, until the end of the study period.

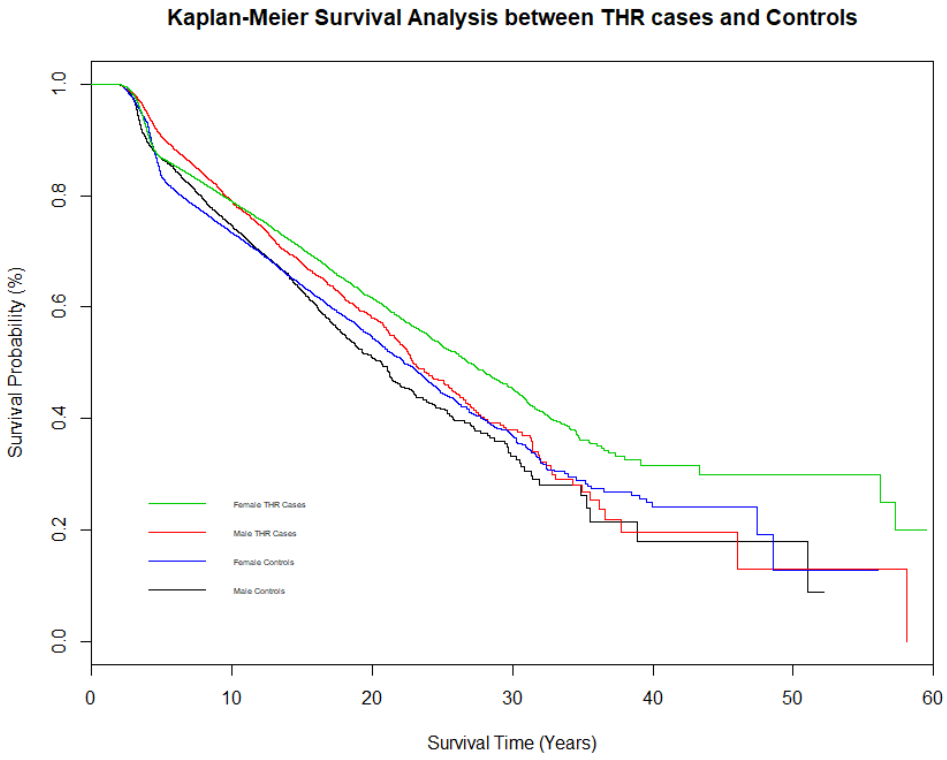
Figure 27 compares the differences in survival between different age groups for THR cases surviving the first two years after surgery. It can be observed that survival of younger patients are better than older age groups. There are significant differences in survival of older age groups, relative to the youngest category (18-54). The p-values of the log-rank test comparing each category to the *18-54* age group are all significant (p-value<0.05).

Figure 28 compares the long term survival of THR cases across procedure types. The percentage survivorship for male cases who underwent *hybrid or other types* of procedures is higher than *cemented* procedures until about 18 years after the surgery (p-value of log-rank test=0.0004) and *uncemented* procedures until about 23 years after the surgery (p-value of log-rank test=0.0002), respectively. Similarly, survival of female cases who underwent *hybrid or other types* of procedures is higher than *cemented* (p-value of log-rank test=0.0010) and *uncemented* procedures (p-value of log-rank test=0.0004), respectively, until about 25 years after the surgery.

In addition, survival of male cases with uncemented THR procedures was better than cemented procedures until about 10 years after surgery (p-value

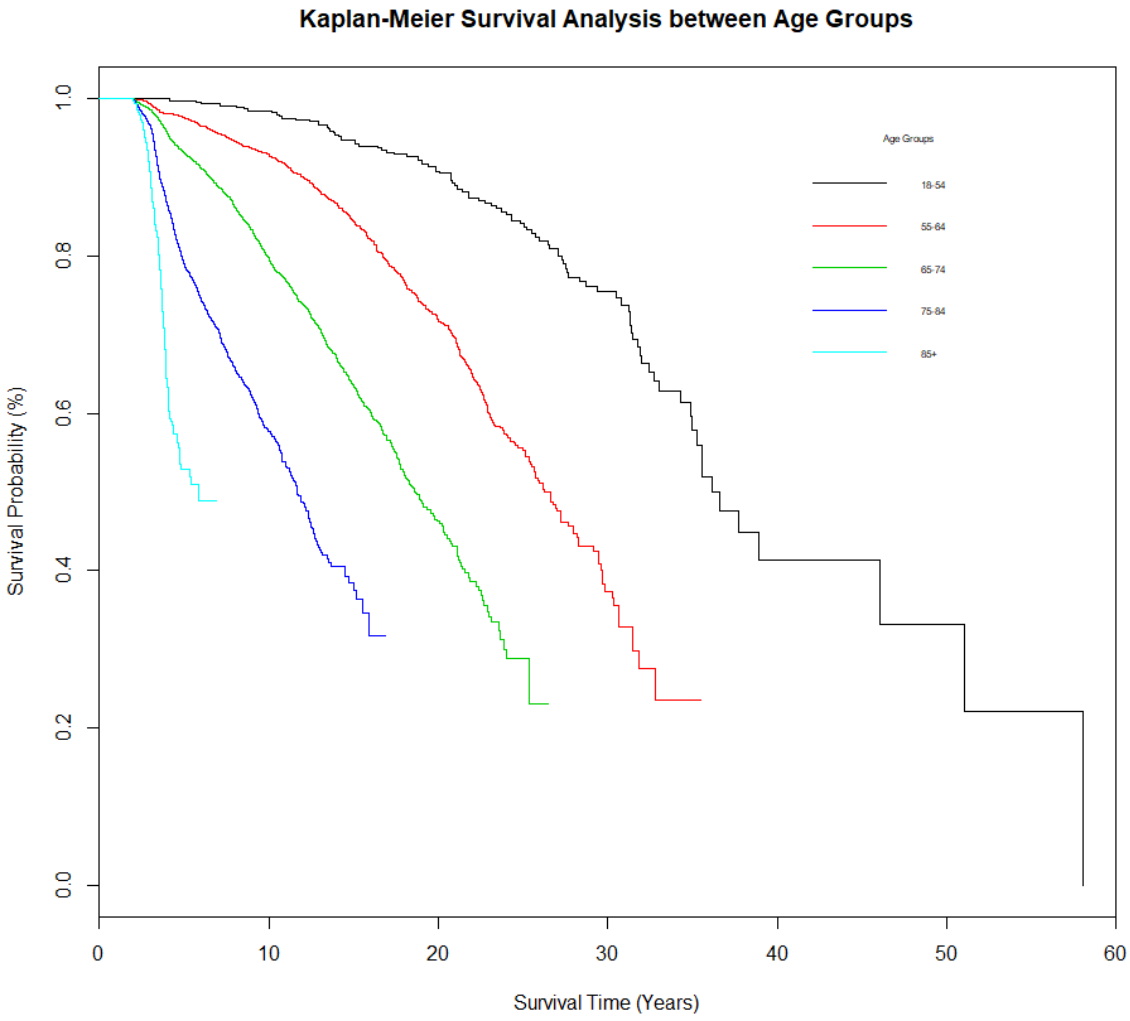
of log-rank test=0.0018). Among female cases, survival of uncemented THR procedures was better than cemented THR procedures until about 25 years after the surgery (p-value of log-rank test=0.0002). However the preliminary variations in survival of male and female cases, depicted in Figure 28, do not account for the fact that the types of THR procedures carried out, depend strongly on the age group and health status of the patients. Thus further analysis, whereby the age group and medical comorbidities of the patients are factored in, is required to understand and estimate the variations in survival of THR cases with different types of procedures.

Figure 26: Survival plot comparing THR cases to matched controls



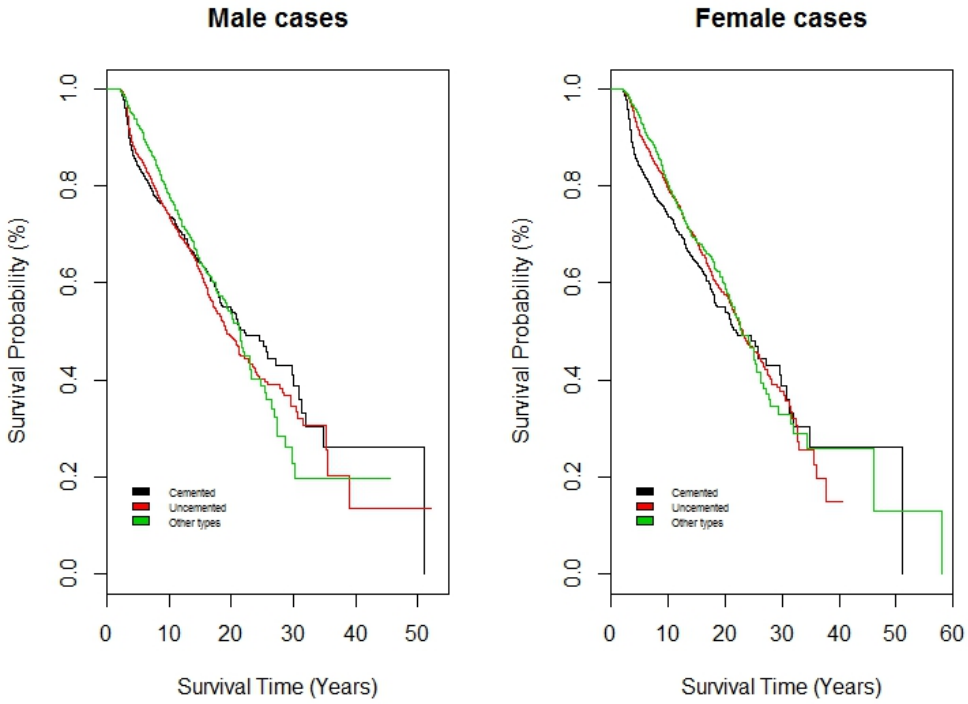
Comparing the Kaplan-Meier survival curves between THR cases and matched controls for both genders

Figure 27: Survival plot comparing THR cases of different age groups



Comparing the Kaplan-Meier survival curves for THR cases across age groups at surgery time

Figure 28: Survival plots comparing types of THR procedures



Comparing the Kaplan-Meier survival curves between cemented, uncemented and other types of THR procedures

7.3 Multilevel Cox's regression analysis

7.3.1 Model development strategy

To answer the research objectives set out in chapter one, a retrospective case control study where THR cases are matched to five controls with the same gender, year of birth category and GP practice, is used. This permits the estimation of the relative hazard ratio (HR) of death between THR cases and controls, taking into account the preoperative history of comorbidities and related treatments, lifestyle factors and residential demographics of the patients. The strengths of a case control study design, applied to this long term survival analysis, are discussed in section 4.1.

The two level Cox's regression model (see section 3.5.9) fitted to the data, is of the form

$$h(t, \underline{x}) = Z h_0(t) \exp(\beta; \underline{x}),$$

where $h_0(t)$ is the baseline hazard rate, $\underline{x}$ is the set of covariates fitted and Z is the random effect (frailty) added to the Cox regression model to account for the clustering effect of patients by their GP practice. The frailty Z is assumed to have an expected value of one and has a multiplicative effect on the individual HR of death. It either increases ($Z > 1$) or decreases ($Z < 1$) the individual all-cause hazard ratio of death. Including a second level random effect, Z, in the regression model allows the estimation of the variation in HR of death between clusters (GP practices) of patients.

The set of variables used to fit the above Cox's regression model is provided in Table 44 in Appendix C. The outcome of interest in this analysis is death after THR procedures. Patients who left their GP practice before the end of the investigation period, were censored. Section 7.1.2 for details.

For the main effects only model, a full model with all variables listed in Table 44 in Appendix C as main effects, is set up and backward elimination is carried out using a significance level of 5% until the best model is achieved. For survival models including second-order interaction terms, an initial full model with all variables listed in Table 44 in Appendix C as main effects and second-order interaction effects, listed in section 7.3.3, is set up. Backward elimination is then carried out using a significance level of 5% for main effects and 1% for second-order interaction terms, respectively, until

the optimal model is achieved. Four survival models are presented in this section, namely the main effect model (1) for full case analysis (L_{full}) and (2) for multiple imputation analysis (L_{imp}), respectively, and the survival model with interaction effects for (3) full case analysis ($Long_{full}$) and (4) multiple imputation analysis ($Long_{imp}$), respectively.

7.3.2 Cox's regression models with main effects only

Results

Table 19 shows the analysis of variance table for the final models with main effects only, fitted using the full case and imputed datasets. The full case analysis results show that the final model, denoted by L_{full}, consists of the following main effects: (1) GP practice, (2) Sex, (3) Townsend score, (4) Mosaic category, (5) Type of procedure, (6) Smoking status (7) Stroke, (8) BMI, (9) Osteoarthritis, (10) Osteoporosis, (11) Hypercholesterolemia and (12) Type 2 diabetes, respectively. In addition to these variables, the Cox's model fitted using ten imputed datasets, labelled as L_{imp}, showed that pre-operative myocardial infarction (MI) and angina also had significant effect on the variations of estimated hazard ratios of death. This shows the importance of comparing the results of full case analysis to multiple imputation analysis since the latter also reveals that MI and angina are risk factors for long term mortality after THR procedures. Under the full case analysis, majority of the patients who had an event of MI or angina before surgery were excluded from the survival analysis because they had missing records for other covariates. Therefore the full case Cox's model failed to detect MI and angina as risk factors.

Table 19: Analysis of variance results comparing full case analysis to multiple imputation analysis for main effects only model

Variables	Full Case Analysis				Multiple Imputation Analysis			
	Loglik	**Chisq**	**Df**	**P-value**	**Loglik**	**Chisq**	**Df**	**P-value**
Empty Model	-171429				-171426			
GP Practice	-171313	233.25	301	<2.0E-16	-171288	211.53	401	<2.0E-16
Stroke	-171250	124.32	1	<2.0E-16	-171235	102.98	1	<2.0E-16
Townsend Scores	-171097	307.62	4	<2.0E-16	-171096	284.04	4	<2.0E-16
Mosaic Category	-171066	61.68	3	1.26E-13	-171060	48.10	3	9.9E-11
Procedures Types	-171064	0.43	3	2.11E-01	-171036	15.82	3	5.8E-04
Myocardial Infarction		NA			-170923	21.25	1	2.1E-06
Angina		NA			-170889	72.95	1	<2.0E-16
BMI	-170935	258.19	2	<2.0E-16	-170863	240.12	2	<2.0E-16
Osteoarthritis	-170891	89.20	1	<2.0E-16	-169894	78.76	1	<2.0E-16
Osteoporosis	-170891	5.17	1	1.32E-02	-169557	15.98	1	3.4E-05
Hypercholesterolaemia	-169915	1951.33	3	<2.0E-16	-169497	1940.36	3	<2.0E-16
Type 2 Diabetes	-169574	681.99	1	<2.0E-16	-167223	664.01	1	<2.0E-16
Sex	-169520	107.56	1	<2.0E-16	-167223	89.92	1	<2.0E-16
Smoking Status	-167242	4555.86	2	<2.0E-16	-167208	4551.33	2	<2.0E-16

The above results show that the full case model is a subset of the multiple imputation final model fitted using 10 imputed datasets. In addition to the covariates kept in the full case model, myocardial infarction and angina were also significant risk factors causing variations in long term hazard ratio of death estimated from the Cox's model fitted using multiple imputation analysis.

Assessing assumption of proportional hazards

The proportionality of the hazards is a critical assumption of Cox proportional hazards analysis. Specifically, the model assumes that each covariate has a multiplicative effect in the hazard function that is constant over time. Violation of the proportional hazards (PH) assumption can raise questions regarding the validity of the model, and possibly lead to misleading and erroneous scientific findings, when interpreting the estimated parameters of the fitted Cox's regression model. To test for the validity of the PH assumption in this analysis, a correlation test between Schoenfeld residuals and event time (Grambsch and Therneau [1994]) for each variable contributing significantly to the fitted Cox regression model, is carried out.

Table 20 shows the estimated correlation between Schoenfeld residuals and the event time, and their associated statistical significance for each variable kept in L_{full} and L_{imp}, respectively. For both models, there is no significant correlation between the Schoenfeld residuals and the event time for all covariates kept in the model (p-value of Chi-square test $> 5\%$). Therefore these variables satisfy the assumption of proportional hazards. Furthermore,

the global test of proportional hazards test for both L_{full} and L_{imp}, respectively, is also insignificant (Table 20), indicating that estimated HR of death for both models are independent of time.

Table 20: Assessing assumption of proportional hazards for main effects model

Variables	Levels	Full Case Analysis			Multiple Imputation Analysis		
		rho	chisq	P-value	rho	chisq	P-value
Stroke	**No**						
	Yes	0.0170	0.547	4.1E-01	0.0230	0.669	3.49E-01
Townsend Scores	**1**						
	2	-0.0028	1.067	2.3E-01	0.0062	1.143	2.11E-01
	3	0.0049	0.221	7.6E-01	0.0091	0.292	6.39E-01
	4	0.0142	0.953	2.5E-01	0.0241	1.108	2.18E-01
	5	0.0054	0.292	6.4E-01	0.0071	0.405	5.12E-01
Mosaic Category	**Cat A-C**						
	Cat D-F	-0.0046	0.198	8.1E-01	0.0007	0.351	5.66E-01
	Cat G-I	0.0025	0.059	1.6E-01	0.0066	0.272	6.67E-01
	Cat J-O	-0.0080	0.675	3.5E-01	-0.0026	0.702	3.35E-01
Procedures Types	**Controls**						
	Cemented	0.0236	1.311	1.8E-01	0.0311	1.319	1.80E-01
	Uncemented	0.0376	1.071	2.3E-01	0.0456	1.203	1.99E-01
	Others	0.0756	1.050	2.3E-01	0.0850	1.179	2.04E-01
Angina	**No**						
	Yes	NA			-0.0082	1.038	2.33E-01
Myocardial Infarction	**No**						
	Yes	NA			-0.0258	0.875	2.75E-01
BMI	**Normal**						
	Overweight	-0.0121	1.235	1.9E-01	0.0093	1.405	1.67E-01
	Obese	-0.0268	0.052	1.7E-01	0.0673	0.152	9.48E-01
Ostearthritis	**No**						
	Yes	0.0054	0.254	7.0E-01	0.0077	0.264	6.80E-01
Osteoporosis	**No**						
	Yes	0.0644	0.965	2.5E-01	0.0151	1.135	2.12E-01
Hypercholesterolaemia	**HC Category 0**						
	HC Category 1	-0.0008	1.006	2.4E-01	0.0030	1.048	2.31E-01
	HC Category 2	0.0140	1.500	1.5E-01	0.0136	1.579	1.44E-01
	HC Category 3	-0.0028	0.068	1.9E-01	0.0349	0.782	3.05E-01
Type 2 Diabetes	**No**						
	Yes	0.0043	0.159	9.2E-01	-0.0541	0.263	6.82E-01
Sex	**Female**						
	Male	0.0342	1.276	1.9E-01	-0.0628	1.497	1.54E-01
Smoking	**Non-Smoker**						
	Ex-smoker	-0.0578	0.797	3.0E-01	0.0448	0.960	2.52E-01
	Smoker	-0.0692	1.114	2.2E-01	0.0929	1.302	1.82E-01
Global Test of Proportional hazards		NA	15.050	1.4E-01	NA	20.942	3.58E-01

All variables kept in the full case and multiple imputation analysis models, respectively, have a p-value greater than 5%, showing that the estimated hazard ratio of death associated with these covariates are independent of survival time in this study. A similar results are obtained for the global test of proportional hazards assumption for both models.
HC Category 0 refers to patients with no records of hypercholesterolemia and statins prescription prior to THR surgery. HC Category 1 represents patients with normal cholesterol level due to statins prescription prior to THR surgery. HC Category 2 are patients with HC despite the statins prescription prior to THR surgery. HC Category 3 are patients with HC but not on statins prescription prior to THR surgery

Appendix D shows the plots of Schoenfeld residuals versus time of death for L_{full} (Figures 40 and 41) and L_{imp} (Figures 42 and 43), respectively. Graphical inspection of these plots shows no observed trend against event time among the Schoenfeld residuals associated with each covariate kept in both models. These residuals do not vary with time and thus there are no violation of the proportional hazards assumption for all covariates kept in L_{full} and L_{imp}, respectively, in this long term survival analysis.

Estimated hazard ratios of death

This section details out the variations in the long term hazard ratios (HR [95% CI]) of death, estimated from L_{imp}. Table 21 displays the estimated HR of death for patients with demographic characteristics, lifestyle factors and preoperative comorbidities that were detected as risk factors for mortality in this analysis.

Relative to controls (HR=1.00), cases undergoing cemented, uncemented and other types of THR procedure have a lower HR of death estimated at 0.86 [0.84,0.89], 0.75 [0.71,0.79] and 0.88 [0.86,0.90], respectively. Therefore THR cases who survived their first two years after surgery, have a lower HR of death than matched controls, indicating that THR procedure improves mortality in the long term. Similarly, being overweight or obese prior to the procedure is beneficial for long term mortality. Estimated HRs of death for obese (0.83 [0.80,0.86]) and overweight (0.80 [0.78,0.83]) patients are lower than for those with normal weight (HR=1.00). Higher BMI appears to have a long term protective effect on survival of patients who survived the first two years of the study.

The rest of the findings apply equally to cases and controls. The frailty term, GP practice was the most significant variable in the model. This shows that HR of death varies significantly between GP practices, with estimated HR of death ranging from 0.68 [0.58,0.79] to 2.64 [2.16,2.97], respectively (variance of random effects due to frailty terms estimated at 0.0325). Male participants in this study have a higher HR of death (HR=1.16 [1.14,1.19]) than female patients (HR=1.00).

Relative to non-smokers (HR=1.00), ex-smokers (1.81 [1.63,2.00]) and smokers (2.58 [2.19,3.03]), respectively, have a higher HR of death. HR of

death increases for patients who are from more deprived residential areas (measured by Townsend scores). Compared to the least deprived residential ward (Townsend score=1 and HR=1.00), the estimated HR of death increases to 1.09 [1.07,1.10], 1.12 [1.09,1.14], 1.16 [1.13,1.19] and 1.19 [1.16,1.23] for residential areas with Townsend scores of 2, 3, 4 and 5 (most deprived), respectively.

In comparison to patients belonging to the socio-economic group where majority of individuals belong to the most affluent and wealthy families with high status jobs and private detached accommodation (Mosaic groups A, B and C, respectively, HR=1.00), patients who are elderly singles or belong to mature families who reside in rural areas and owning inexpensive accommodation (Mosaic groups D, E and F, respectively) have HR of death estimated at 1.07 [1.03,1.11]. Similarly, patients from socio-economic group where most individuals belong to relatively young families with children and aspiring to become home owners through their own limited resources (Mosaic groups G, H and I, respectively), also have a higher HR of death estimated at 1.13 [1.09,1.16], than patients from Mosaic groups A, B and C. Patients from Mosaic groups *J–O*, inclusive, represent the least affluent and least wealthy socio-economic group where majority of the individuals are elders who rely on financial support from their local government to rent low cost accommodation in urban areas or elderly home owners of inexpensive accommodation whose mortgage is nearly paid off. Estimated hazard of death for patients from Mosaic groups *J–O* is estimated at 1.06 [1.02,1.09] relative to Mosaic groups *A–C*.

Preoperative angina, myocardial infarction and stroke increase HR of death to 1.62 [1.49,1.75], 1.43 [1.35,1.52] and 1.37 [1.30,1.44], respectively, compared to patients without these comorbidities. Degenerative conditions, namely, osteoarthritis and osteoporosis also raise estimated HRs of death to 1.08 [1.04,1.13] and 1.49 [1.40,1.60], respectively, relative to patients without these conditions. Patients with type 2 diabetes have a higher HR of death, estimated at 2.14 [1.88,2.44], compared to those without this co-morbidity. Relative to patients with no hypercholesterolemia (HC) and no statins prescription (Category 0, HR=1.00), patients with normal cholesterol due to statins prescription (Category 1) and those with HC despite the intake of statins prescription (Category 2), have a higher HRs of death estimated at 1.14 [1.09,1.18] and 1.62 [1.49,1.76], respectively. Similarly, compared to Cat-

egory 0, patients with HC but not on statins prescription (HC Category 3) have a HR of death estimated at 1.70 [1.55,1.87]. Thus patients with pre-operative HC have higher HRs of death than those without, while intake of statins improve HR of death for some patients.

Table 21: Long term Cox's regression analysis and estimated all-cause hazard ratios (95% confidence interval) of death after THR procedures

Variables	Levels	Full Case Analysis, L_{full}		Multiple Imputation Analysis, L_{imp}		
		HR (95% CI)	P-value	HR (95% CI)	P-value	RIV*
Sex	**Female**	Ref (1.00)		Ref (1.00)		
	Male	1.15 (1.11,1.18)	<2.0E-16	1.16 (1.14,1.19)	2.3E-03	2.9E-03
Procedure Type	**Controls**	Ref (1.00)		Ref (1.00)		
	Cemented	0.85 (0.78,0.92)	7.8E-03	0.86 (0.84,0.89)	2.9E-03	5.9E-03
	Uncemented	0.73 (0.68,0.78)	3.3E-05	0.75 (0.71,0.79)	1.5E-03	4.9E-03
	Others	0.86 (0.78,0.96)	<2.0E-16	0.88 (0.86,0.90)	1.2E-02	4.4E-03
Townsend Scores	**1**	Ref (1.00)		Ref (1.00)		
	2	1.08 (1.04,1.13)	<2.0E-16	1.09 (1.07,1.10)	1.8E-02	7.8E-03
	3	1.12 (1.06,1.18)	3.2E-04	1.12 (1.09,1.14)	2.5E-03	6.3E-04
	4	1.19 (1.12,1.27)	9.4E-06	1.16 (1.13,1.19)	7.7E-03	1.1E-03
	5	1.24 (1.15,1.33)	8.5E-09	1.19 (1.16,1.23)	3.2E-03	2.4E-03
Mosaic Category	**Cat A-C**	Ref (1.00)		Ref (1.00)		
	Cat D-F	1.06 (1.02,1.11)	2.5E-09	1.07 (1.03,1.11)	2.3E-02	3.1E-03
	Cat G-I	1.14 (1.07,1.17)	2.4E-02	1.13 (1.09,1.16)	3.8E-03	7.5E-03
	Cat J-O	1.06 (1.01,1.09)	8.1E-06	1.06 (1.02,1.09)	3.5E-02	6.8E-03
Smoking	**Non-Smoker**	Ref (1.00)		Ref (1.00)		
	Ex-smoker	1.67 (1.59,1.76)	<2.0E-16	1.81 (1.63,2.00)	3.4E-03	1.0E-03
	Smoker	2.52 (2.40,2.65)	<2.0E-16	2.58 (2.19,3.03)	8.6E-03	5.4E-03
BMI	**Normal**	Ref (1.00)		Ref (1.00)		
	Overweight	0.79 (0.76,0.82)	3.4E-03	0.80 (0.78,0.83)	2.7E-02	1.1E-03
	Obese	0.84 (0.80,0.87)	2.0E-15	0.83 (0.80,0.86)	8.8E-03	1.3E-03
Angina	**No**	NA		Ref (1.00)		
	Yes			1.62 (1.49,1.75)	2.3E-03	1.4E-03
Myocardial Infarction	**No**	NA		Ref (1.00)		
	Yes			1.43 (1.35,1.52)	2.8E-03	6.7E-03
Stroke	**No**	Ref (1.00)		Ref (1.00)		
	Yes	1.34 (1.28,1.41)	<2.0E-16	1.37 (1.30,1.44)	3.8E-03	7.7E-03
Ostearthritis	**No**	Ref (1.00)		Ref (1.00)		
	Yes	1.09 (1.05,1.13)	<2.0E-16	1.08 (1.04,1.13)	1.3E-02	2.3E-03
Osteoporosis	**No**	Ref (1.00)		Ref (1.00)		
	Yes	1.39 (1.29,1.5)	4.4E-06	1.49 (1.40,1.60)	7.0E-03	1.1E-03
Hypercholesterolaemia	**Category 0**	Ref (1.00)		Ref (1.00)		
	Category 1	1.14 (1.07,1.22)	<2.0E-16	1.14 (1.09,1.18)	3.0E-03	2.7E-05
	Category 2	1.57 (1.51,1.63)	<2.0E-16	1.62 (1.49,1.76)	4.8E-03	3.6E-03
	Category 3	1.69 (1.6,1.79)	<2.0E-16	1.70 (1.55,1.87)	1.7E-02	5.0E-03
Type 2 Diabetes	**No**	Ref (1.00)		Ref (1.00)		
	Yes	2.08 (1.97,2.19)	3.4E-06	2.14 (1.88,2.44)	7.2E-03	6.7E-03

HC Category 0 refers to patients with no records of hypercholesterolemia and statins prescription prior to THR surgery. HC Category 1 represents patients with normal cholesterol level due to statins prescription prior to THR surgery. HC Category 2 are patients with HC despite the statins prescription prior to THR surgery. HC Category 3 are patients with HC but not on statins prescription prior to THR surgery.

** RIV indicates the reduction in value of estimated model coefficients if the results of the 10 imputed dataset were not combined into a single effect size using Rubin's rule.*

Performance and comparison of the models

This section assesses the performance of L_{full} and L_{imp}, respectively, by comparing the estimated Royston's R^2, Harrell's concordance and shrinkage of the models (Table 45 in Appendix D). L_{full} explains 23.4% to 28.5% of the differences in survival of patients, compared to between 22.6% and 27.4% for L_{imp}. Harrell's concordance for L_{full} is estimated at 70.3% (95% CI: 67.4%-73.1%) and hence, there exists 67.4% to 73.1% agreement between the estimated HR of death and the observed survival time, compared to between 67.9% and 74.8% for L_{imp}. This is a reasonably good performance for both L_{full} and L_{imp}, respectively, as it lies in the range of 60%–70%, as recommended by Therneau and Grambsch (2013) for survival analysis of medical data. Estimated shrinkage slope indicates that the adjusted HRs of death were overestimated by 3.2% for L_{full} and by 2.4% for L_{imp}, respectively. These small shrinkage slope values suggest that the estimated results of these survival models are robust.

Furthermore, Δ_L, the percentage difference between estimated coefficients of L_{full} and L_{imp}, described by equation 6.1 in section 6.4.4, is computed as a sensitivity check, see Table 46 in Appendix D. Estimated Δ_L ranged from -18% to a maximum of 21%, showing that the full case analysis underestimates the model coefficients between 1%-18% for Townsend Score 3, 4, and 5, Mosaic categories G-I and J-O, types of THR procedures, overweight, OA and HC category 1, respectively, and overestimates those of stroke, Townsend Score 2, Mosaic Category D-F, obese, osteoporosis, HC categories 2 and 3, type 2 Diabetes, sex and smoking status, respectively, between 1% and 18%.

Furthermore, the multiple imputation analysis also reveals that preoperative angina and myocardial infarction are risk factors for mortality in this long term study, unlike the full case analysis. This is mainly because the majority of the low number of patients with preoperative angina and myocardial infarction in this study, were excluded during the full case analysis and thus their effects on HR of death were insignificant in the analysis. These arguments show the importance of including all possible patients, using multiple imputed datasets because such analysis improves the estimation of the model parameters and provide precise and unbiased estimates of variability in the data.

Validation and discussion of results

The results of model L_{imp} show that THR procedures improve survival of THR cases who survived their first 24 months after surgery, for all types of THR procedures, with the highest improvement in mortality associated with other types of procedures, followed by uncemented and cemented. Long term survival of female patients were better than male patients. These results are in agreement with the following studies, which reported similar findings at 3 to 15 years after the procedure: McMinn et al. (2012), Paavolainen et al. (2002), Whitehouse et al. (2014), Barrett et al. (2005), Lovald et al. (2014), Lie et al. (2000), Maradit-Kremers et al. (2016), Visuri et al. (1994), Ritter et al. (1998), Ramiah et al. (2007), Visuri et al. (1997), Pedersen et al. (2011) and Mäkelä et al. (2014).

However no studies on survival after THR in literature investigated preoperative treated or untreated hypercholesterolemia. This may be because the studies reviewed in literature (See section 2.3) used secondary care databases in which measurements of hypercholesterol and treatments of hypercholesterolemia were not available. It is well established that mortality risk for smokers is significantly higher than non-smokers (Doll et al. (2004)). The results are adjusted for smoking status in the analysis, thus producing more precise estimates. Among all studies reviewed in section 2.3, none of the authors adjusted for smoking status.

In addition, preoperative cardiovascular diseases (angina, myocardial infarction and stroke) increase long term HRs of death, compared to those without these conditions. Pedersen et al. (2011) found that estimated mortality rates per 100,000 person-years (MR) for THR cases with preoperative cardiovascular diseases (MR=28.5%) was lower than controls with similar conditions (MR=17.8%), implying that an interaction between procedure type and cardiovascular diseases, which is investigated and presented in section 7.3.3). However the authors did not distinguish between the types of cardiovascular diseases as carried out in this study.

Survival model L_{imp} also indicates that there is a statistically significant protective effect of being overweight or obese at the baseline for long term HR of death. This finding is also observed among other interventions for treatments of chronic conditions (Curtis et al. (2005), Bakaeen and Chu (2011),

Stamou et al. (2011)) and is referred as the *obesity paradox*. No studies in the literature reviewed investigated effects of being overweight and obese prior to surgery, on long term survival after THR. Whitehouse et al. (2014) excluded BMI from their Cox's regression model due to large number of missing data for BMI.

HR of death increased for patients living in residential areas with high deprivation level. This result is in agreement with Whitehouse et al. (2014), which reported that unit increase in index of multiple deprivation score of an individual residential ward, causes a 1% increase in post-THR risk of death. Similarly, less wealthy and less affluent socio-economic groups (Mosaic groups D-F, G-I and J-O, respectively) have a higher estimated HRs of death than the most wealthy and affluent groups. No previous studies adjusted for socio-economic factors in their long term survival analysis.

Preoperative osteoporosis, osteoarthritis, hypercholesterolemia, type 2 diabetes and smoking status of patients increase long term HR of death. Long term survival analyses by Holmberg (1992), Pedersen et al. (2011), Lie et al. (2000), Barrett et al. (2005) and Whitehouse et al. (2014) demonstrate that mortality among patients diagnosed with osteoarthritis and osteoarthritis, is higher than those without these degenerative conditions. Similarly Pedersen et al. (2011) and Barrett et al. (2005) also found type 2 diabetes as risk factors for mortality in their results comparing THR cases to controls.

Several studies (See section 2.3) demonstrated that long term mortality risk after THR procedure varies significantly for patients with various preoperative characteristics or medical conditions. Table 22 summarises the different publications that investigated the effect of second order interactions between procedure types and gender, age group and various medical conditions, respectively, on mortality risk after the surgery. Having a survival model that tests for such second order interactions between various covariates assist in explaining further sources of variations in survival among THR cases. In this section, the two-level Cox's regression models (L_{full} and L_{imp}) only tested the main effects of the covariates. To have a deeper insight of explained variability in adjusted HR of death, second order interaction terms are added to the models presented in this section and their effects on the HRs of death are reported in section 7.3.3.

Table 22: Studies investigating effects of interactions between covariates on long term survival analysis after THR

Interaction with types of THR tested	Studies
Age at surgery	McMinn et al. (2012), Whitehouse et al. (2014)
Gender	McMinn et al. (2012), Makela et al. (2014), Whitehouse et al. (2014)
Ethnicity	Lovald et al. (2014)
Deprivation of residential ward	Whitehouse et al. (2014)
Type 2 diabetes, Cardiovascular diseases	Barett et al. (2005)
Osteoarthritis/Osteoporosis/Rheumatoid Arthritis	Holmberg et al. (1992), Lie et al. (2000), Barrett et al. (2005), McMinn et al. (2012), Lovald et al. (2014)

The above authors investigated and reported the significant effects of interactions between types of THR procedures and the above variables on the long term survival of THR cases. Such analysis permitted the authors to further investigate which categories of THR cases were associated with the highest improvement in long term survival/mortality.

7.3.3 Cox's regression models with interaction effects

Second order interaction terms investigated

To obtain a deeper insight of variability in long term HRs of death after THR procedures, several second order interaction terms, listed below, are added to the main effect models presented in section 6.4.2.

1. THR procedure types and demographics variables (age at surgery, gender)
2. THR procedure types and lifestyle factors (smoking status, BMI)
3. THR procedure types and social deprivation indices (Townsend score) and socio-economic group (Mosaic score)
4. THR procedure types and preoperative medical conditions (angina, myocardial infarction, stroke, osteoarthritis, rheumatoid arthritis, osteoporosis, hypertension, hypercholesterolemia, chronic kidney diseases)
5. THR procedure types and drug prescriptions (Ace-inhibitor, oestrogen, progesterone, testosterone)
6. Gender and lifestyle factors (smoking status, BMI)
7. Gender and preoperative medical conditions (angina, myocardial infarction, stroke, osteoarthritis, rheumatoid arthritis, osteoporosis, hypertension, hypercholesterolemia, chronic kidney diseases)
8. Age at surgery and lifestyle factors (smoking status, BMI)

9. Age at surgery and preoperative medical conditions (angina, myocardial infarction, stroke, osteoarthritis, rheumatoid arthritis, osteoporosis, hypertension, hypercholesterolemia, chronic kidney diseases)

Initially a full model with all variables listed in Table 44 from Appendix C as main effects and that also includes the above second order interaction terms is created. Backward elimination is then carried out using a significance level of 5%. Two separate multivariate Cox's regression models are developed and compared in this section; one for full case analysis ($long_{full}$) and a second model ($long_{imp}$), using the 10 imputed datasets.

Abbreviations

The following set of abbreviations are used for discussions of all results that follow in this section.

• MI: Myocardial Infarction
• BMI: Body Mass Index
• OA: Osteoarthritis
• HC Category 0: Patients with no records of hypercholesterolemia and statins prescription prior to THR surgery
• HC Category 1: Patients with normal cholesterol level due to statins prescription prior to THR surgery
• HC Category 2: Patients with HC despite being on statins prescription prior to THR surgery
• HC Category 3: Patients with HC but not on statins prescription prior to THR surgery

Results

Table 23 displays the main effects and interaction terms that were kept in the full case analysis model ($long_{full}$) and multiple imputation analysis model ($long_{imp}$), respectively. Model $long_{full}$ consists of the following main effects: (1) GP practice, (2) Townsend scores, (3) Mosaic category, (4) angina, (5) osteoporosis, (6) sex, (7) type 2 diabetes, (8) procedure type, (9) BMI, (10) smoking status, (11) MI, (12) OA and (13) HC, respectively, in addition to the following interaction terms:

(1) types of procedures and MI,

(2) types of procedures and BMI,
(3) types of procedures and OA
(4) types of procedures and HC
(5) sex and type 2 diabetes
(6) smoking status and BMI.

Similarly model $long_{imp}$ includes the same list of main effects and interaction effects as in $long_{full}$ in addition to preoperative angina (Table 23). Therefore the multiple imputation analysis showed angina as an additional risk factor to survival for all patients. This is because under the full case analysis, majority of the low number of patients diagnosed with angina are excluded from the analysis since they have missing records for other variables. Thus the analysis fails to pick angina as a significant risk factor to long term survival.

Table 23: Analysis of variance results comparing full case analysis to multiple imputation analysis for model including interaction terms

Variables	**Full Case Analysis,** $long_{full}$				**Multiple Imputation Analysis,** $long_{imp}$			
	Loglik	**Chisq**	**Df**	**P-value**	**Loglik**	**Chisq**	**Df**	**P-value**
Empty Model	-171429				-171420			
GP Practice	-171313	1233.25	301	<2.0E-16	-171309	217.05	401	8.1E-16
Stroke	-171250	24.32	1	4.2E-07	-171228	12.15	1	2.6E-04
Townsend Scores	-171097	37.62	4	6.4E-08	-171073	28.67	4	4.3E-06
Mosaic Category	-171066	41.68	3	2.3E-09	-171056	53.31	3	7.7E-12
Procedures Types	-171064	12.94	3	2.2E-03	-171062	23.11	3	1.8E-05
MI	-170843	24.18	1	4.6E-07	-170930	24.62	2	2.3E-06
Angina	NA				-170880	8.09	1	2.4E-03
BMI	-170707	21.46	2	1.1E-05	-170873	16.35	1	2.8E-05
OA	-170660	9.28	1	1.3E-03	-169894	19.89	3	8.5E-05
Osteoporosis	-169998	9.48	1	1.1E-03	-169554	6.48	1	6.1E-03
HC	-169738	18.23	3	1.9E-04	-169496	29.25	1	3.3E-08
Type 2 Diabetes	-169413	6.26	1	7.0E-03	-167239	9.41	2	4.5E-03
Sex	-169379	66.75	1	<2.0E-16	-167203	74.69	1	<2.0E-16
Smoking Status	-168139	28.59	2	3.1E-07	-167207	22.91	1	8.8E-07
Procedures Types*MI	-168134	11.15	3	5.1E-03	-167203	0.04	4	8.6E-03
Procedures Types*BMI	-168116	35.79	6	1.4E-06	-167198	0.09	8	7.5E-06
Procedures Types*OA	-168115	17.82	3	2.3E-04	-167185	0.03	3	6.3E-02
Procedures Types*HC	-168027	22.15	9	3.0E-03	-167170	22.96	9	2.3E-03
Type 2 Diabetes*Sex	-168017	18.37	1	9.6E-06	-167156	24.90	2	2.0E-06
BMI*Smoke	-167026	1.14	8	8.8E-03	-167136	39.73	4	2.3E-08

The above results show that the full case analysis is a subset of the multiple imputation analysis final model fitted using 10 imputed datasets. In addition to the covariates kept in the full case analysis model, angina was also a significant source of variation hazard ratio of death estimated from the Cox's model fitted using multiple imputation analysis. MI, BMI, OA, HC and smoking status cause significant difference in HR of death for each types of procedures.

Assessing assumption of proportional hazards and model comparison

To test for the validity of the assumption of Cox's proportional hazards, a correlation test between Schoenfeld residuals and event time (Grambsch and Therneau [1994]) for each variable and second order interactions contributing significantly to the fitted Cox's regression models, namely, $long_{full}$ and $long_{imp}$, respectively, is carried out. Table 24 shows the associated statistical significance of these correlation tests for $long_{full}$ and $long_{imp}$, respectively. There is no significant correlation between the Schoenfeld residuals and the event time for all covariates kept in the model (p-value of Chi-square test $> 5\%$). Therefore these variables satisfy the assumptions of proportional hazards. Furthermore, the global test of proportional hazards test for both $long_{full}$ and $long_{imp}$, respectively, is insignificant (Table 20), indicating estimated HRs of death for both models are independent of time.

Appendix D shows the plots of Schoenfeld residuals versus time of death for $long_{full}$ (Figures 44, 45 and 46) and $long_{imp}$ (Figures 47, 48 and 43), respectively. Graphical inspection of these plots shows no trends over event time among the Schoenfeld residuals associated with covariates and interaction terms kept in both models. These residuals do not vary with time and thus there are no violation of the proportional hazards assumption for all covariates kept in $long_{full}$ and $long_{imp}$, respectively.

Δ_{long}, the percentage difference between estimated coefficients of $long_{full}$ and $long_{imp}$, described by equation 6.1 in section 6.4.4, is computed as a sensitivity check, see Table 25. Estimated Δ_{long} shows the full case analysis underestimates the model coefficients between 0.5%-4.0% and overestimates some coefficients between $0.1\% - 3.9\%$, respectively. Compared to the models of main effects only (L_{full} and L_{imp}), the models with interaction terms (Table 25) have significantly lower values for Δ_{long}.

Table 24: Assessing assumption of proportional hazards for models with interaction effects

Variables	Levels	Full Case Analysis			Multiple Imputation Analysis		
		rho	chisq	P-value	rho	chisq	P-value
Sex	**Female**						
	Male	0.0202	1.0328	6.9E-01	0.0355	0.3772	4.6E-01
Procedures Types	**Controls**						
	Cemented	-0.0047	0.1836	3.3E-01	0.0264	0.7432	6.1E-01
	Uncemented	0.0075	0.4627	5.0E-01	0.0545	1.3707	7.6E-01
	Others	0.0133	1.4693	7.7E-01	0.0774	1.3046	7.5E-01
Townsend Scores	**1**						
	2	-0.0031	0.0838	2.3E-01	0.0002	0.7727	6.2E-01
	3	0.0043	0.1706	3.2E-01	0.0278	1.1512	7.2E-01
	4	0.0134	0.7474	6.1E-01	0.0283	0.3221	4.3E-01
	5	0.0035	0.1210	2.7E-01	0.0300	1.1358	7.1E-01
Mosaic Category	**Cat A-C**						
	Cat D-F	-0.0029	0.0774	2.2E-01	0.0085	0.4518	5.0E-01
	Cat G-I	0.0042	0.1701	3.2E-01	0.0123	0.3863	4.7E-01
	Cat J-O	-0.0067	0.4634	5.0E-01	-0.0077	0.3252	4.3E-01
BMI	**Normal**						
	Overweight	-0.0239	1.1450	7.2E-01	-0.0059	1.1812	7.2E-01
	Obese	-0.0256	1.3325	7.5E-01	-0.0029	1.4715	7.7E-01
Smoking	**Non-Smoker**						
	Ex-smoker	-0.0495	1.2574	7.4E-01	-0.0478	1.5795	7.9E-01
	Smoker	-0.0579	1.2665	7.4E-01	-0.0623	1.0306	6.9E-01
Stroke	**No**						
	Yes	0.0154	1.0916	7.0E-01	0.0211	0.6895	5.9E-01
Angina	**No**						
	Yes				-0.0623	1.0306	6.9E-01
MI*	**No**						
	Yes	0.0236	1.4738	7.8E-01	0.0965	1.0782	7.0E-01
OA**	**No**						
	Yes	0.0010	0.0083	7.2E-02	0.0251	0.1747	3.2E-01
Osteoporosis	**No**						
	Yes	-0.0102	0.092	2.4E-01	0.0681	0.3689	4.6E-01
HC***	**Category 0**						
	Category 1	0.0111	1.0619	7.0E-01	0.0149	1.1750	7.2E-01
	Category 2	0.0320	0.9744	6.8E-01	0.0217	1.0875	7.0E-01
	Category 3	0.0176	0.6973	6.0E-01	-0.0006	1.6007	7.9E-01
Type 2 Diabetes	**No**						
	Yes	-0.0019	0.0324	1.4E-01	0.0085	0.8365	6.4E-01
Procedures Types*MI	**Controls & No MI**						
	Cemented & MI	0.0079	0.547	5.4E-01	0.0180	0.9395	6.7E-01
	Uncemented & MI	0.0078	0.508	5.2E-01	0.0158	0.6482	5.8E-01
	Others & MI	-0.0025	0.054	1.8E-01	0.0039	0.5192	5.3E-01
Procedures Types*BMI	**Controls & Normal BMI**						
	Cemented & Obese	0.0218	0.050	1.8E-01	0.0246	0.1270	2.8E-01
	Uncmented & Obese	0.0146	0.556	5.4E-01	0.0186	0.1585	3.1E-01
	Others & Obese	-0.0022	0.040	1.6E-01	0.0208	0.5832	5.5E-01
	Cemented & Overweight	0.0196	0.745	6.1E-01	0.0443	0.1314	2.8E-01
	Uncemented & Overweight	0.0119	1.200	7.3E-01	0.0254	0.7458	6.1E-01
	Others & Overweight	-0.0077	0.505	5.2E-01	0.0155	1.2356	7.3E-01
Procedures Types*OA	**Controls & OA**						
	Cemented & OA	0.0076	0.494	5.2E-01	0.0087	0.5494	5.4E-01
	Uncmented & OA	0.0105	0.945	6.7E-01	0.0232	0.6065	5.6E-01
	Others & OA	0.0074	0.464	5.0E-01	0.0240	1.0009	6.8E-01
Procedures Types*HC	**Controls & No HC**						
	Cemented & HC Category 1	-0.0161	0.751	6.1E-01	0.0056	0.5150	5.3E-01
	Uncmented & HC Category 1	-0.0168	1.394	7.6E-01	-0.0029	0.8069	6.3E-01
	Others & HC Category 1	-0.0164	1.321	7.5E-01	-0.0072	1.4497	7.7E-01
	Cemented & HC Category 2	-0.0179	1.765	8.2E-01	-0.0115	1.4214	7.7E-01
	Uncmented & HC Category 2	-0.0182	0.790	6.3E-01	0.0047	1.8760	8.3E-01
	Others & HC Category 2	-0.0129	1.389	7.6E-01	-0.0021	0.8503	6.4E-01
	Cemented & HC Category 3	-0.0123	1.272	7.4E-01	0.0065	1.4104	7.7E-01
	Uncmented & HC Category 3	0.0055	0.258	3.9E-01	0.0083	1.3461	7.5E-01
	Others & HC Category 3	-0.0125	1.310	7.5E-01	0.0095	0.3508	4.5E-01
Type 2 Diabetes*Sex	**Controls & No Type 2 diabetes**						
	Male & Type 2 diabetes	0.0038	0.120	2.7E-01	0.0085	1.3957	7.6E-01
BMI*Smoke	**Normal Weight & Non-smoker**						
	Obese & Ex-smoker	0.0164	1.030	6.9E-01	0.0260	0.1229	2.7E-01
	Overweight & Ex-smoker	0.0096	1.798	8.2E-01	0.0345	1.0747	7.0E-01
	Obese & Smoker	0.0079	0.547	5.4E-01	0.0230	1.8698	8.3E-01
	Overweight & Smoker	0.0114	1.118	7.1E-01	0.0214	0.6043	5.6E-01
Global Test OF PROPORTIONAL HAZARDS		NA	36.2913	1.1E-01	NA	44.3622	3.0E-01

Table 25: Comparison of estimated coefficients of $long_{full}$ and $long_{imp}$

Variables	Levels	Full Case Analysis			Multiple Imputation Analysis				Estimated
		Estimate	S.Error	P-value	Estimate	Std.Er	P-value	RIV	Delta, Δ_{long}
Sex	**Female**								
	Male	0.1319	0.0173	1.4E-14	0.1366	0.0320	1.23E-04	3.5E-03	3.4%
Procedures Types	**Controls**								
	Cemented	-0.3304	0.0946	2.2E-03	-0.3370	0.0967	2.66E-02	6.2E-03	1.9%
	Uncemented	-0.5102	0.0860	1.5E-09	-0.5078	0.1060	4.24E-04	2.9E-03	-0.5%
	Others	-0.3008	0.0902	4.4E-04	-0.3030	0.0956	2.08E-03	7.7E-03	0.7%
Townsend Scores	**1**								
	2	0.0769	0.0234	5.4E-04	0.0783	0.0135	3.40E-02	5.6E-03	1.7%
	3	0.1136	0.0267	1.1E-05	0.1171	0.0188	1.07E-03	6.4E-03	2.9%
	4	0.1732	0.0314	1.8E-08	0.1769	0.0195	7.60E-04	7.4E-03	2.1%
	5	0.2029	0.0363	1.2E-08	0.1955	0.0178	6.43E-04	7.7E-03	-3.8%
Mosaic Category	**Cat A-C**								
	Cat D-F	0.0544	0.0270	2.6E-02	0.0545	0.0093	5.33E-03	6.5E-03	0.2%
	Cat G-I	0.1192	0.0299	3.5E-05	0.1212	0.0149	1.13E-02	2.1E-04	1.6%
	Cat J-O	0.0567	0.0244	1.2E-02	0.0589	0.0159	1.28E-02	2.0E-03	3.9%
BMI	**Normal**								
	Overweight	-0.2565	0.0217	<2.0E-16	-0.2643	0.0903	1.62E-05	6.9E-04	3.0%
	Obese	-0.2731	0.0277	<2.0E-16	-0.2741	0.0626	3.46E-05	4.9E-04	0.4%
Smoking	**Non-Smoker**								
	Ex-smoker	0.8539	0.0351	<2.0E-16	0.8713	0.0445	4.71E-07	6.1E-03	2.0%
	Smoker	1.0110	0.0307	<2.0E-16	1.0183	0.1837	5.38E-07	4.6E-03	0.7%
Stroke	**No**								
	Yes	0.2780	0.0244	<2.0E-16	0.2883	0.0268	1.96E-05	4.6E-03	3.6%
Angina	**No**								
	Yes	NA			0.3816	0.0915	7.12E-04	1.3E-04	NA
MI*	**No**								
	Yes	0.4191	0.0252	<2.0E-16	0.4219	0.1007	4.40E-06	2.6E-03	0.7%
OA**	**No**								
	Yes	0.0824	0.0196	1.4E-05	0.0849	0.0105	2.36E-03	6.0E-03	3.0%
Osteoporosis	**No**								
	Yes	0.0712	0.0103	8.1E-05	0.0728	0.0199	3.76E-03	4.1E-04	2.2%
HC***	**Category 0**								
	Category 1	0.1618	0.0481	<2.0E-16	0.1631	0.0632	9.22E-05	7.3E-03	0.8%
	Category 2	0.3938	0.0203	<2.0E-16	0.4010	0.0984	4.96E-07	6.6E-03	1.8%
	Category 3	0.5779	0.0317	1.7E-07	0.5796	0.1044	9.35E-04	7.8E-03	0.3%
Type 2 Diabetes	**No**								
	Yes	0.7977	0.0359	<2.0E-16	0.8077	0.1025	1.37E-06	4.8E-03	1.2%
Procedures Types*MI	**Controls & No MI**								
	Cemented & MI	0.2518	0.0644	5.0E-03	0.2557	0.0799	2.78E-02	5.5E-03	1.5%
	Uncemented & MI	0.1552	0.0846	4.0E-02	0.1555	0.0968	5.23E-03	5.4E-03	0.2%
	Others & MI	0.3043	0.0979	1.2E-02	0.3077	0.0386	1.14E-02	7.2E-03	1.1%
Procedures Types*BMI	**Controls & Normal BMI**								
	Cemented & Obese	0.1523	0.0934	6.5E-02	0.1543	0.0450	6.72E-03	7.1E-03	1.3%
	Uncmented & Obese	0.3127	0.0743	1.3E-05	0.3208	0.0908	2.35E-03	1.9E-03	2.5%
	Others & Obese	0.2530	0.0749	1.6E-02	0.2587	0.0144	5.40E-03	4.9E-03	2.2%
	Cemented & Overweight	0.2300	0.0892	5.6E-03	0.2349	0.0825	3.21E-02	3.8E-03	2.1%
	Uncemented & Overweight	0.2418	0.0704	3.2E-04	0.2484	0.0769	7.81E-03	3.4E-03	2.6%
	Others & Overweight	0.1136	0.0337	3.5E-02	0.1092	0.0305	3.73E-02	5.5E-03	-4.0%
Procedures Types*OA	**Controls & OA**								
	Cemented & OA	0.1374	0.0560	1.9E-03	0.1381	0.0317	5.16E-02	5.4E-03	0.5%
	Uncmented & OA	0.4177	0.1012	2.3E-02	0.4327	0.0887	1.81E-02	4.9E-04	3.5%
	Others & OA	0.3789	0.0987	2.6E-03	0.3821	0.0118	6.75E-03	4.7E-03	0.8%
Procedures Types*HC	**Controls & No HC**								
	Cemented & HC Category 1	0.0385	0.0951	1.3E-03	0.0398	0.0092	4.59E-02	6.2E-03	3.2%
	Uncmented & HC Category 1	0.1421	0.0825	2.4E-03	0.1444	0.0459	2.42E-02	3.7E-03	1.6%
	Others & HC Category 1	0.1236	0.0271	2.7E-03	0.1265	0.0104	4.59E-02	4.9E-03	2.3%
	Cemented & HC Category 2	0.6780	0.1005	1.9E-09	0.6790	0.0324	5.15E-04	2.8E-03	0.1%
	Uncmented & HC Category 2	0.9909	0.0952	<2.0E-16	1.0222	0.2808	3.01E-05	4.9E-03	3.1%
	Others & HC Category 2	0.3547	0.0988	1.0E-02	0.3650	0.0593	9.30E-02	5.4E-03	2.8%
	Cemented & HC Category 3	0.5409	0.1007	9.9E-04	0.5202	0.1050	9.54E-02	5.0E-03	-4.0%
	Uncmented & HC Category 3	0.4674	0.1019	4.6E-05	0.4766	0.0634	3.95E-03	5.9E-03	1.9%
	Others & HC Category 3	0.4932	0.0972	2.3E-03	0.5116	0.1048	5.32E-03	2.1E-03	3.6%
Type 2 Diabetes*Sex	**Controls & No Type 2 diabetes**								
	Male & Type 2 diabetes	0.2416	0.0515	1.4E-06	0.2480	0.0564	7.21E-04	4.8E-04	2.6%
BMI*Smoke	**Normal Weight & Non-smoker**								
	Obese & Ex-smoker	0.1833	0.0388	4.3E-03	0.1845	0.0730	2.73E-02	6.5E-03	0.6%
	Overweight & Ex-smoker	0.2112	0.0553	7.0E-05	0.2117	0.0714	6.02E-03	4.1E-03	0.3%
	Obese & Smoker	0.1292	0.0373	1.5E-02	0.1277	0.0437	2.53E-02	5.3E-03	-1.2%
	Overweight & Smoker	0.1921	0.0534	1.7E-04	0.1976	0.0840	5.41E-03	4.7E-03	2.7%

Estimated Δ_{long}'s between full case and multiple imputation analyses are low and therefore indicate a marginal difference between estimated coefficients.

Estimated hazard ratios of death

This section details the estimated in long term hazard ratios (HR [95% CI]) of death in the model $long_{imp}$. The frailty term, GP practice was the most significant variable in the model, indicating that HR of death varies significantly between GP practices, with estimated HRs of death ranging from 0.60 [0.57,0.64] to 3.64 [3.21,4.06], respectively (variance of random effects due to frailty terms estimated at 0.0218).

Types of THR procedures is a significant risk factor for long term survival in this analysis. Therefore there is significant difference in HRs of death between controls and different types of cases. Figure 29 shows that, compared to controls without any preoperative medical condition (HR=1.00), estimated HRs of cases who undergo cemented, uncemented and other types of THR are equal to 0.71 (0.52,0.88), 0.60 (0.53,0.68) and 0.74 (0.65,0.83), respectively. Therefore for cases who survived the first 24 months after the surgery, their survival is better than matched controls.

Figure plot 29 is a forest plot showing the variation in long term HRs of death between cases and controls with preoperative MI, OA and HC. Model $long_{imp}$ includes significant interaction between types of THR procedures with MI, OA and HC. Therefore, for patients with these comorbidities, survival varies between cases and controls, and also between types of THR procedures. Relative to controls without MI (HR=1.00), preoperative MI increases HR of death to 1.52 (1.35,1.68), 1.41 (1.38,1.43), 1.07 (0.91,1.23) and 1.53 (1.38,1.68) for controls, cemented, uncemented and other types of THR procedures with MI, respectively. Degenerative condition OA causes HRs of death to increase to 1.33 (1.07,1.56), 1.09 (1.03,1.16), 1.23 (1.11,1.36) and 1.44 (1.26,1.65) for controls, cemented, uncemented and other types of THR procedures with the condition, respectively, compared to controls without OA.

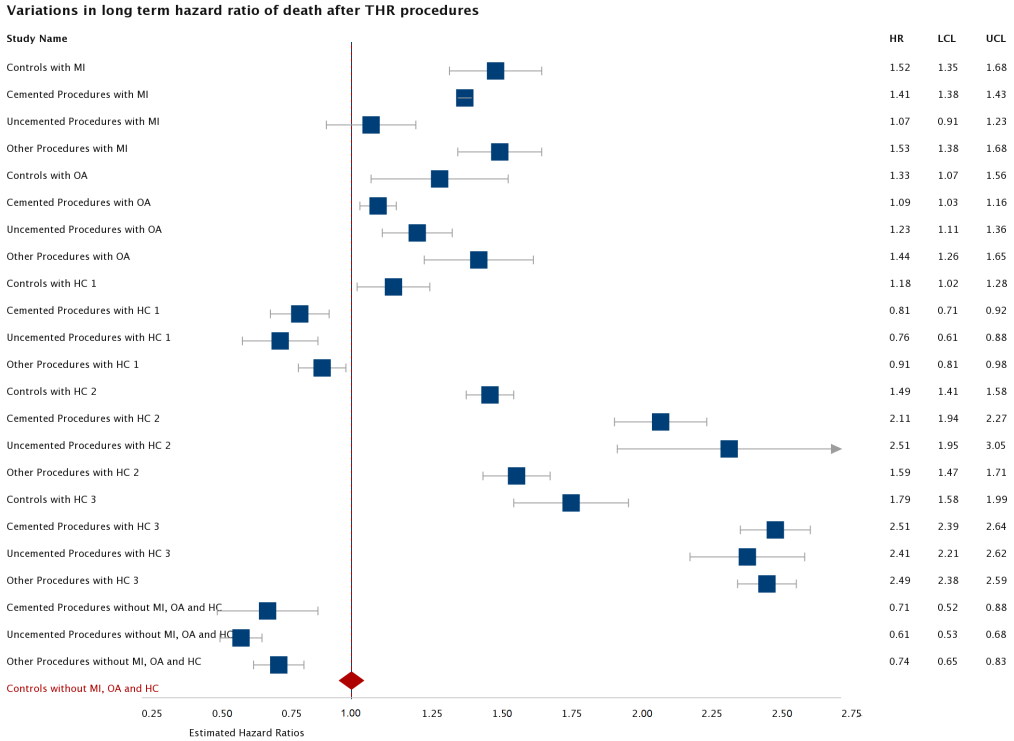
Compared to controls without HC (HR=1.00), controls with normal cholesterol due to statins prescriptions (HC 1 in Figure 29) have a higher HR of death (1.18 [1.02,1.28]), while estimated HR of death is lower for cemented procedures (0.81 [0.70,0.92]), uncemented procedures (0.76 [0.61,0.88]) and other types of procedures (0.91 [0.81,0.98]) with HC category 1, respectively. This shows that THR cases with HC category 1 still have better survival

than matched controls with and without HC category 1, but have higher estimated HR of death than cases without HC category 1. However for cases and controls who either have HC despite being on statins prescriptions (HC category 2) or not on statins prescriptions (HC category 3), HR of death is higher than in controls without HC (HR=1.00). HR of death increases to 1.49 (1.41,1.58), 2.10 (1.94,2.27), 2.50 (1.95,3.05) and 1.59 (1.47,1.71) for controls, cemented, uncemented and other types of procedures with HC category 2, respectively, and rises to 1.79 (1.58,1.99), 2.51 (2.39,2.64), 2.41 (2.20,2.62) and 2.49 (2.38,2.59) with HC category 3, respectively. These results show that HC causes HR of death to increase for all patients but statins prescriptions are beneficial for the long term survival of THR cases, i.e., for those who take statins prescription (HC category 1).

Variations in HR of death are furthermore explained by significant interactions between types of THR procedures and BMI, and also between smoking status and BMI, respectively. Figure 30 shows the difference in adjusted HRs of death for cases and controls with different levels of preoperative BMI and smoking status. Relative to controls who are non-smokers and with normal BMI (HR=1.00), HRs of death among overweight cases are estimated at 0.94 (0.88,0.99) for cemented procedures, 0.80 (0.76,0.84) for uncemented procedures, 0.85 (0.81,0.90) for other types of procedures, respectively. Similarly, compared to non-smoking controls with normal BMI (HR=1.00), obese cases have a lower HR of death estimated at 0.90 (0.85,0.95) for cemented and 0.89 (0.84,0.95) for uncemented procedures, respectively.

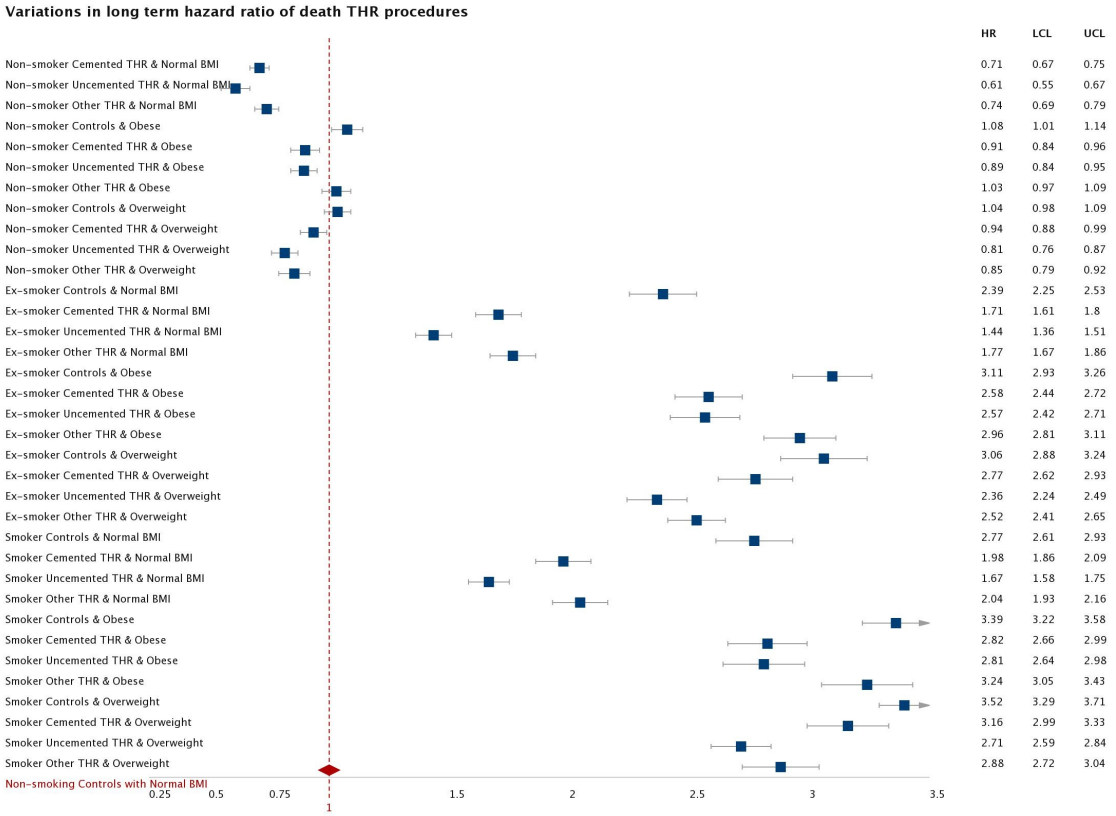
Relative to non-smokers with normal BMI (HR=1.00), HR of death increases significantly for ex-smokers and smokers present in this study population. Figure 30 below displays the variations of HRs of death for THR cases and matched controls across different types of BMI and smoking status. For all overweight and obese patients, HR of death associated with ex-smokers and smokers are higher than for non-smokers. Ex-smokers and smokers' HRs are highest for controls followed by other types of procedures, uncemented and cemented procedures, respectively, for the overweight category, while among obese patients, HR is highest for controls, followed by cemented, other types of procedures and uncemented procedures, respectively.

Figure 29: Effect of preoperative MI, OA and HC on survival after THR



Forest plot showing the effect of preoperative MI, OA and HC on long term survival of THR cases and matched controls, respectively. HC 1 refers to patients with normal cholesterol level due to statins prescriptions. HC 2 refers to patients with hypercholesterolemia despite being on statins prescription and HC 3 refers to patients with hypercholesterolemia but not on statins prescriptions, respectively.

Figure 30: Effect of preoperative BMI and smoking status on survival after THR



Forest plot showing the effect of being overweight and obese on long term survival after THR among non-smokers, ex-smokers and smokers, respectively.

The remaining findings reported applied to both cases and controls equally. Patients who were diagnosed with preoperative angina and stroke have a higher HRs of death estimated at 1.46 (1.22,1.75) and 1.33 (1.27,1.41), respectively, compared to patients without these conditions. Patients with osteoporosis have an estimated HR of death of 1.08 (1.03,1.12) relative to those without this degenerative condition. Similarly, patients diagnosed with type 2 diabetes have a higher HRs of death estimated at 3.29 (2.59,4.04) and 2.24 (1.83,2.74) among males and females, respectively, compared to male patients (HR=1.15 [1.08,1.22]) and female patients (HR=1.00) without this condition.

Relative to patients living in residential areas with the lowest deprivation index (Townsend score=1, HR=1.00), HRs of death increase in more deprived areas. Estimated HR for patients from residential areas with Townsend score of 2, 3, 4 and 5 are 1.08 (1.05,1.11), 1.12 (1.08,1.17), 1.19 (1.15,1.24), and 1.22 (1.17,1.26), respectively. In comparison to patients belonging to the top socio-economic groups (Mosaic groups A, B and C, HR=1.00), patients who are elderly singles or belong to mature families who reside in rural areas and owning inexpensive accommodation (Mosaic groups D, E and F) have HRs of death estimated at 1.06 (1.04,1.08). Similarly, patients from socio-economic group where most individuals belong to relatively young families with children and aspiring to become home owners through their own limited resources (Mosaic groups G, H and I), also have a higher HR of death estimated at 1.13 (1.10,1.16). Patients from Mosaic groups *J–O*, inclusive, represent the least affluent and least wealthy socio-economic group where majority of the individuals are elders who rely on financial support from their local government to rent low cost accommodation in urban areas or elderly home owners of inexpensive accommodation whose mortgage is nearly paid off. Estimated hazard ratio of death for patients from Mosaic group *J–O* is estimated at 1.06 (1.03,1.09) relative to Mosaic group *A–C*.

Performance statistics

This section assesses the performance of the models $long_{full}$ and $long_{imp}$ by comparing the estimated Royston's R^2, Harrell's concordance and shrinkage of the models (Table 47 in Appendix D). Model $long_{full}$ explains 22.7% to 28.1% of the differences in survival of patients, compared to between 22.1% and 27.6% for $long_{imp}$. Harrell's concordance for model $long_{full}$ is estimated at 71.6% (95% CI: 68.4%-74.2%) and hence, there exists 68.4% to 74.2%

agreement between the estimated HR of death and the observed survival time, compared to between 68.9% and 75.4% for model $long_{imp}$. This is a reasonably good concordance for both models $long_{full}$ and $long_{imp}$, as it lies in the range of 60%–70%, as recommended by Therneau and Grambsch (2013) for survival analysis of medical data. Estimated shrinkage slopes indicate that the adjusted HR of death were overestimated by 3.9% for models $long_{full}$ and by 3.1% for $long_{imp}$, respectively. These small shrinkage slope values suggest that the estimated results of these survival models are robust.

Validation and discussion of results

The results of model $long_{imp}$ showed that THR procedure improved survival of THR cases who survived their first 24 months after surgery, for all types of THR procedures, while female patients had lower HR of death than male individuals. The following studies also reported similar findings, 3 to 15 years after the procedure: McMinn et al. (2012), Paavolainen et al. (2002), Whitehouse et al. (2014), Barrett et al. (2005), Lovald et al. (2014), Lie et al. (2000), Maradit-Kremers et al. (2016), Visuri et al. (1994), Ritter et al. (1998), Ramiah et al. (2007), Visuri et al. (1997), Pedersen et al. (2011) and Mäkelä et al. (2014). The highest improvement in survival was associated with uncemented procedures, followed by cemented and other types of procedures, respectively. These findings are also reported by Visuri et al. (1997), Kendal et al. (2013) and Mäkelä et al. (2014).

Among cardiovascular diseases included in this analysis, only MI caused significant increase in HR of death for different procedure types, unlike stroke and angina, which did not cause a difference in survival of different types of cases and controls. This finding is supported by Visuri et al. (1997), Barrett et al. (2005) and Pedersen et al. (2011) who found that THR cases with a history of cardiovascular diseases before surgery, had higher HR of death compared to controls and THR cases without these medical conditions. However these studies did not differentiate between the types of cardiovascular diseases in their study, whereas this analysis compared survival of patients with preoperative stroke, angina and MI.

Both osteoporosis and OA increased HR of death, a finding also reported by Holmberg (1992), Lie et al. (2000), Barrett et al. (2005), Pedersen et al. (2011) and Whitehouse et al. (2014, while Pedersen et al. (2011) also reported

that higher HR of death is associated with osteoporosis, compared to OA. Furthermore, only OA was found as a risk factor for variation in survival of various kinds of THR in this current study (Figure 29). Whitehouse et al. (2014 also reported higher survival among uncemented THR cases with OA compared to cemented ones.

This study also found that preoperative HC increased HR of death for all patients, compared to those without HC, except for THR cases who had normal cholesterol due to intake of statins (HC category 1), see Figure 29. Patients with HC not controlled (HC categories 2 and 3) had poorer survival compared to those without HC. HR was lower than controls without HC for cases with HC category 1. This may be due to the long term benefits of statins on survival (Gitsels et al. (2016)), coupled with the improvement in survival due to the THR procedure itself. No studies in literature were found to investigate preoperative treated or untreated hypercholesterolemia. This may be because the studies reviewed in literature (see section 2.3) used secondary care databases in which measurements of hypercholesterol and treatments of hypercholesterolemia were not available.

Type 2 diabetes increased long term mortality risk for cases and controls in this study. However it did not cause significant difference in survival of various types of THR procedures. This is in agreement with the results of Pedersen et al. (2011) and Barrett et al. (2005), respectively, who reported type 2 diabetes as a risk factor for mortality when comparing THR cases to controls. It is well established that mortality risk for smokers is significantly higher than non-smokers (Doll et al. (2004)). The results are adjusted for smoking status in this analysis, and showed that HR of death is highest for smokers and ex-smokers, respectively. Among all studies reviewed in section 2.3, none of the authors adjusted for smoking status.

Being overweight or obese had a protective effect on long term HR of death in this study and improved longevity for THR cases without any pre-operative comorbidities. This is referred as the *obesity paradox* and several studies reported this relationship between being overweight or obesity and longevity in men and women of all ages, races, and ethnicities (Curtis et al. (2005), Adams et al. (2006), Bakaeen and Chu (2011), Stamou et al. (2011)). No studies in the literature reviewed investigated effects of being overweight or obese prior to surgery, on long term survival after THR while Whitehouse

et al. (2014) excluded BMI from their Cox's regression model due to large number of missing data for BMI. Further research is needed to investigate why higher levels of BMI before the surgery is beneficial for long term survival after THR procedures whereas being overweight or obese is associated with a higher overall mortality risk in the general population (Banack and Kaufman (2014)). A comparison of BMI readings pre and post-THR surgery may aid in understanding the improvement in survival associated with higher BMI levels.

HR of death increases for patients living in residential areas with high deprivation level. This result is in agreement with Whitehouse et al. (2014) who reported that unit increase in index of multiple deprivation score of an individual residential ward, causes a 1% increase in post-THR risk of death. Similarly, patients from less wealthy and less affluent socio-economic groups (Mosaic groups D-F, G-I and J-O, respectively) have a higher estimated HR of death than the most wealthy and affluent categories. No previous studies adjusted for socio-economic factors in their long term survival analysis.

7.4 Strengths and limitations

In this research, routinely collected primary care data from the THIN database is used as the source of data to investigate long term survival after THR. These data are representative of the UK population as explained in section 4.4.2 for various demographic, lifestyle factors and medical conditions (Blak et al. (2011), Langley et al. (2011), González et al. (2009), Hippisley-Cox and Coupland (2010a), MacDonald and Morant (2008)). Primary care records consist of an extensive availability of socio-demographic, lifestyle factors and medical information, thereby improving coverage of THR cases during the study period and also allowing to validate the results of this study to the general UK population.

Furthermore, this study is designed as a case control study, allowing estimation of effect of THR procedures on long term survival between THR cases and controls matched on gender, year of birth category and GP practice, while adjusting for several confounders such as preoperative lifestyle factors and demographic variables. No previous study has adjusted estimated mortality risk by so many confounders in literature. Additionally, this study estimated HR of death after THR given the preoperative medical

history of the patients. The results of this study can therefore be used to inform healthcare professionals on medical and risk management for patients waiting to undergo THR and for THR cases after the procedure.

In THIN database, data on the type of prosthesis, surgical approach and surgeon experience are not available. These variables were reported to have significant effect on long term survival following THR (Whitehouse et al. (2014)). Therefore estimates of long term mortality risk in this study were not adjusted by these variables and survival after THR could not be distinguished for different types of prosthesis, surgical approach and surgeon experience, respectively. Another limitation in this study is the presence of missing values for lifestyle factors. Variables related to pollution levels and ethnicity were not available for patients from Scotland and Northern Ireland and were thus not included in the survival model. However, missing values for variables that were available for the whole UK population (BMI, Townsend scores, IMD, smoking status and Mosaic groups) were dealt with using multiple imputation analysis, which is the widely accepted method to deal with bias and lack of precision in estimates when missing values are present (Enders (2010)). Additionally, drug therapy was included as a confounder in this study. However one limitation with this approach is that adherence of patients to the drug therapy is unknown, thereby, not precisely reflecting the effects of drug therapy on long term mortality risk following THR.

7.5 Conclusions

The findings of this study suggest that adjusted HR of death of THR cases with no preoperative medical condition, are lower than matched controls, for cemented, uncemented and other types of THR procedures. HR of death is higher for males for all types of procedures. Preoperative stroke and angina, respectively, increase HR of death for cases and controls, but did not vary between different types of THR procedures, compared to patients without these cardiovascular diseases. Preoperative type 2 diabetes, hypercholesterolemia and osteoporosis, respectively, increase hazards of death for all types of THR procedure compared to those without these conditions. Furthermore THR cases who were on statins prescriptions for treatment of hypercholesterolemia have better long term survival than those without the same prescriptions. Being overweight and obese prior to the procedure improved long term survival

of THR cases for all types of THR procedures, compared to controls with normal BMI. However survival of ex-smokers and smokers across all types of THR procedures is worse than non-smoking controls and THR cases, respectively.

8 Discussion

Introduction

This thesis is relevant to the development of mortality risk models using primary care data to estimate all-cause short term odds ratio (OR) and long term hazards ratio (HR) of death, following THR procedures in the UK. This chapter firstly, summarises the results and contributions of this research findings to the existing evidence available in literature. Secondly, the objectives of this research are addressed and an evaluation of its clinical and actuarial validity and implications are presented, and thirdly, a review of the strengths and limitations of this research follows.

8.1 Main findings

8.1.1 Short term mortality risk model

In this large study, age at surgery, year of birth category, types of THR procedure, lifestyle factors (smoking status and BMI) and preoperative medical conditions (Type 2 diabetes, angina, myocardial infarction, stroke, osteoarthritis and hypercholesterolemia) were the most important risk factors for OR of death during the first 24 months after the procedure. Estimated OR of death after primary THR procedures were higher among THR cases than matched controls, given they had no preoperative medical conditions. Estimated OR of death increased for older THR cases and were higher for male THR cases compared to female cases. These findings are similar to published studies by Fender et al. (1997), Lie et al. (2002), Nunley and Lachiewicz (2003), Blom et al. (2006), Pedersen et al. (2011), Jones et al. (2014), Lie et al. (2000), Williams et al. (2002), Ramiah et al. (2007), Aynardi et al. (2009), NHS Scotland (2002), Lie et al. (2000), Hunt et al. (2013), NHS Scotland (2002) and Lovald et al. (2014).

The findings of this research demonstrated that OR of death within the first 24 months for patients with preoperative medical conditions such as Type 2 diabetes, angina, stroke, osteoarthritis and hypercholesterolemia increased estimated OR of death in the first 2 years equally for cases and controls, respectively, compared to patients without these conditions. These results are in accordance with previous studies by Gaston et al. (2007) and

Pedersen et al. (2011). Although a proportion of patients in this research were from deprived areas, variations in short term OR of death were not different by deprivation levels (Townsend scores, Index of multiple deprivation) and socio-economic categories (Mosaic group).

Across types of THR procedures, OR of death was the highest for cemented procedures, followed by uncemented and other types of THR procedures. This result is in agreement with previous studies by Lie et al. (2000), Nunley and Lachiewicz (2003), Ramiah et al. (2007), Hunt et al. (2013) and Bozic et al. (2012). Furthermore, preoperative myocardial infarction increased OR of death for all types of THR procedures compared to patients without the condition - a result in accordance with Hunt et al. (2013) and Comba et al. (2012).

BMI and smoking status prior to THR surgery contributed to variations in short term OR of death for different type of procedures. Being overweight and obese before the surgery increased short term OR of death for all types of THR procedures. This result is also reported by Bozic et al. (2012) and Memtsoudis et al. (2012), respectively. Ex-smokers and smokers also have an increased OR of death for all types of procedures, compared to non-smokers. No studies in literature adjusted for the effect of smoking status prior to THR procedure while investigating short term mortality following THR, improving precision of estimated OR of death.

8.1.2 Long term mortality risk model

Among patients who survived longer than 2 years after the procedure, variations in estimated all-cause HR of death were significantly contributed to by gender, types of THR procedure, deprivation level (Townsend scores), socio-economic status (Mosaic groups), preoperative lifestyle factors (BMI and smoking status) and medical conditions (stroke, angina, myocardial infarction, hypercholesterolemia, osteoarthritis, osteoporosis and type 2 diabetes).

Estimated HR of death for THR cases without any preoperative medical conditions was lower than matched controls without any medical condition and was lower for female patients than male individuals. These results are in accordance with McMinn et al. (2012), Paavolainen et al. (2002), Whitehouse et al. (2014), Barrett et al. (2005), Lovald et al. (2014), Lie et al.

(2000), Maradit-Kremers et al. (2016), Visuri et al. (1994), Ritter et al. (1998), Ramiah et al. (2007), Visuri et al. (1997), Pedersen et al. (2011) and Mäkelä et al. (2014), who also found significant improvement in long term survival of THR cases, compared to controls. Preoperative stroke, angina, osteoporosis and Type 2 diabetes increased long term HR of death equally for both cases and controls compared to patients without these comorbidities. These results agree with Pedersen et al. (2011).

Deprivation and socio-economic factors are considered to be main indicators of health and health related behaviours (Macintyre et al. (2002), Diez Roux (2001), Pickett and Pearl (2001)) and (Pickett and Pearl (2001)). This is reflected in the higher HR of death associated firstly, with residential areas with high level of deprivation (Townsend scores ≥ 3) —a finding similar to the results provided by Whitehouse et al. (2014), and secondly, with socio-economic factor *Mosaic category*, where HR of death was higher among patients from less affluent categories.

Compared to matched controls, estimated long term HR of death was the lowest for uncemented procedures, followed by cemented and other types of THR procedures, respectively. This result is in agreement with Lie et al. (2000) and McMinn et al. (2012). Preoperative diagnosis of osteoarthritis also increased long term hazard of death for various types of procedures; a result also reported by Holmberg (1992), Pedersen et al. (2011), Lie et al. (2000), Barrett et al. (2005) and Whitehouse et al. (2014).

HR of death of THR cases with preoperative hypercholesterolemia was the highest among those not on statins prescriptions, followed by those on statin prescription. However for those with normal cholesterol level due to statins prescription, their HR of death was lower than controls without hypercholesterolemia, showing the survival benefits of statins on long term survival as shown also by Gitsels et al. (2016). No previous studies considered for effects of preoperative hypercholesterolemia on survival of THR cases.

Being overweight or obese before the surgery decreased HR of death for all types of THR procedures among non-smokers, indicating a protective effect on long term survival of cases, compared to controls with normal BMI. This obesity paradox is reported in men and women of all ages, races, and ethnicities for various surgeries (Curtis et al. (2005), Adams et al. (2006),

Bakaeen and Chu (2011), Stamou et al. (2011)). However this protective effect disappears among THR cases who were either smokers or ex-smokers prior to their THR surgery. Estimated HR of death increased for smokers and ex-smokers who are either overweight or obese, compared to controls with normal BMI.

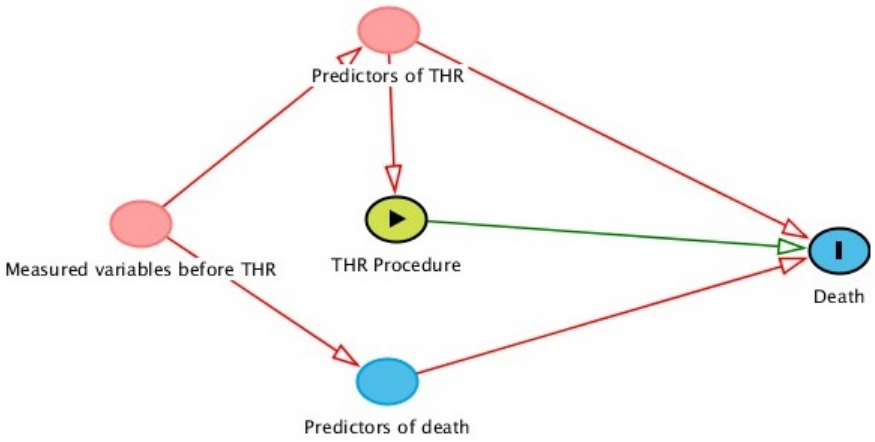
8.2 Is association between THR and mortality risk a result of study bias?

Introduction

Almost all studies are prone to error because they use samples drawn from a population to estimate what is occurring or what might occur in the whole population. These errors can broadly be divided into random error and systematic error, respectively. Random error is the play of chance and results in an estimate of effect being equally likely to be above or below the true value and is assessed with statistical measures such as p-values and confidence intervals. Systematic error is called bias and also leads to the estimate being above or below the true value. This section addresses how various sources of bias are addressed and discusses whether the research findings can be a result of study bias. The presence of study bias commonly causes the association between exposure and outcome among those selected for the analysis to differ from the association among those eligible for the experiment (Szklo and Nieto (2000)). Study biases can be classified as causal effect, confounding result or selection bias (Hernán et al. (2004)).

Figure 31 shows a directed acylic graph (Greenland (2003)) that depicts the causal association between THR procedures and death. The green arrow shows the direct causal path between THR and death while, the pink arrows shows the ancestors that are also predictors of THR and death, respectively. In this research, measured covariates can be grouped into two categories: (1) predictors of THR procedure such as age, gender, degenerative inflammatory conditions such as OA and BMI (Schäfer et al. (2010), Smith et al. (2012)), and (2) predictors of death itself, such as demographic variables and history of medical comorbidities before the procedure. Because of the relationships between the measured covariates (Figure 31) and death, not adjusting the selection of patients and analysis for these links will result in biased conclusions.

Figure 31: Directed acylic graph depicting the causal association between THR procedures and death



1. Cause and effect

Cause and effect relationship can give rise to bias as a consequence of reverse causation. If the outcome of interest precedes the measured covariates, then the association of the outcome with the measured covariates can partially represent bias attributable to the outcome's effect on measured exposure (Hernán et al. (2002)). In both case-control studies carried out in this research, all measured covariates were measured before the THR procedure and before death, the outcome of interest. Therefore the risk of incorporating bias as a result of reverse causation has been minimised.

2. Confounders

Generally when the exposure and outcome of interest have a common cause, the estimated relationship measure is different from the effect measure between exposure and outcome (Hernán et al. (2004)), giving rise to confounding results. Figure 31 shows the structure of the link between measured covariates that are predictors of THR procedure and death, respectively. In this research, predictors of THR procedure are age, gender, diagnosis of inflammatory conditions such as OA and BMI (Schäfer et al. (2010), Smith

et al. (2012)), and measured covariates that are also predictors of death are age, gender, BMI, lifestyle factors and history of medical comorbidities before the procedure. To minimise the bias due to these confounders on the data analysis, all mortality risk models developed, are adjusted for these covariates so that estimated odds and hazard ratios of death are unbiased.

3. Selection bias

Selection bias is present when the exposure status of cases or controls affects the likelihood that they are accepted into the experiment. All cases should be equally likely to be included in the study population. If cases or controls are kept in, or excluded from, a study on criteria related to exposure to the risk factor under examination, estimated effect will be biased. Selection bias arises due to several sources during the study design and data collection process.

Firstly, selection bias may be present in a case-control study if there is inappropriate selection of controls. In this research, patients identified as controls were not selected due to their medical comorbidities. The only condition to be accepted as controls in both studies carried out was that they did not undergo a THR procedure. Moreover, controls were randomly matched to cases using gender, year of birth category and GP practice to ensure that both cases and controls might be assumed were exposed to the same level of measured and unmeasured covariates. Selecting controls in this way avoided introducing selection bias into the study design.

Secondly, loss to follow-up (censoring) can cause selection bias in estimated effects if the censoring is informative. This is because conditioning on censored variable to estimate the effect of interest will yield biased results (Hernán et al. (2002)). In the short term study, a sensitivity analysis comparing the model that included censoring due to transfer outs and another model that excluded transferred out patients, revealed marginal differences in estimated OR of death, indicating that selection bias due to patients lost for follow-up after transferring to new GP practice had insignificant impact on the data analysis. A reason for this ma be the low volume of transfer outs during the first 24 months after surgery. In the long term study, transferred out patients were also censored.However, the analysis revealed no significant effect of transfer outs on HR of death, when the patients' transfer to new

GP practice was treated as a factor. Thus, the long term results were not affected by selection bias due to loss to follow-up.

Thirdly, data analysis that excludes measured covariates with missing data (full case analysis) yields biased results (Spratt et al. (2010), van Buuren (2012)). In this research, missing data for Townsend score, index of multiple deprivation, Mosaic category, body mass index and smoking status were imputed using multiple imputation technique (Enders (2010)). Sensitivity analyses carried out to compare the difference in results between full case analysis and multiple imputation analysis showed differences in results that would have been undetected under full case analysis, leading to inappropriate conclusions from the analysis. Therefore data analysis in both studies carried out was not affected by selection bias arising due to missing data because missing data were included in the analysis by employing multiple imputation techniques.

Conclusion

In the light of the arguments presented in this section, potential sources of bias due to confounders and selection bias were controlled during the design and analysis stage of both studies carried out in this research. This helped to minimise systematic errors and to produce appropriate unbiased results. Therefore, it can be concluded that the reported association between THR and mortality risk in this research is due to the surgical procedure itself, although perfect causal effect of THR on mortality cannot be attributed because of the presence of random errors. Nonetheless, the use of confidence intervals for all estimated results helped to account for errors occurring randomly.

8.3 Research implications

8.3.1 An evaluation of research objectives

The first aim of this research was to identify THR procedures from routinely collected data in primary care records using the THIN database. Table 27 in Appendix A displays the Read codes for identification of primary THR procedures in THIN. These Read codes have been validated by comparing them to the International Classification of Disease (ICD-10) for surgical in-

tervention. Therefore primary care records in THIN can be used to identify primary THR procedures in UK. However the description of procedure types can only be categorised as cemented, uncemented, hybrid or as other types. No information describing the material of the hip prosthesis (metal or non-metal) used during the procedure, is available in THIN.

The second aim was to estimate the preoperative prevalence of various medical conditions among patients identified as THR cases in this research and compare with the relevant literature for the UK population. Prevalence of preoperative smoking, angina, myocardial infarction, osteoarthritis, rheumatoid arthritis, osteoporosis and hypercholesterolemia was marginally lower in this study population, compared to relevant publications using primary care records. However obesity, type 2 diabetes, stroke among cases aged 74 or less, treated hypercholesterolemia and hypertension (among female cases), respectively, were more prevalent among THR cases than relevant published studies on these comorbidities. Prevalence of chronic kidney diseases, hypertension among male THR cases and stroke among THR cases aged 75 or more, is similar to relevant published studies. These lower estimated prevalences may be due to under reporting of records in THIN because patients with no records of a particular preoperative medical condition, were assumed to be free of that co-morbidity prior to their procedure.

The third aim was to account for missing data in the study population when developing mortality/survival models. Smoking status (27.5%), Townsend scores (14.9%), BMI (9.7%), Mosaic group (8.4%) and index of multiple deprivation (2.1%), respectively had unknown records in THIN for THR cases and matched controls. These unknown observations were assumed to be missing at random and handled using multiple imputation technique in this research. For both short and long term mortality models, results from the full case analysis for patients with complete records were compared to results for the analysis from the imputed datasets in terms of estimated model parameters, performance and evaluation.

The fourth aim was to estimate the short and long term mortality risk after THR procedures. Short term odds ratio of death were higher among THR cases than matched controls during the first 2 years after the procedure. Estimated all-cause odds ratio of death was the highest for cemented procedures, followed by uncemented and other types of procedures. How-

ever, for patients surviving the first two years after surgery, estimated long term all-cause hazards ratio of death among THR cases were lower than matched controls, and was the lowest for uncemented procedures, followed by cemented and other types of procedures. These findings suggest that THR procedures improve mortality risk in the long term, given the patients survived the first 24 months after surgery.

The fifth aim was to investigate how the presence of preoperative medical conditions affects the short and long term mortality risk after THR. Risk factors for variations in short term odds ratio of death following THR procedures include gender, myocardial infarction, BMI and smoking status. For the long term model, risk factors for THR cases for all-cause hazards ratio of death include myocardial infarction, smoking status, BMI, osteoarthritis and hypercholesterolemia. Hazards of death associated with the comorbidities remained constant during the follow-up after the procedure. This shows that irrespective of the number of years the patients had already survived with the co-morbidity, they were still at a higher hazard of death than those without the condition.

The sixth aim was to identify socio-economic factors that explained variations in short and long term risk of death following THR procedures. During the first 24 months after the procedure, estimated odds ratio of death do not vary significantly between different levels of deprivation of patients residential wards and between different socio-economic groups. However estimated long term hazards ratio of death is higher for cases from more deprived residential areas and from low income and less affluent categories (Mosaic group G-O in Table 29 in Appendix A). Explanation for variations in long term hazards of death for different levels of deprivation and Mosaic groups can be explained by poor ecology, lifestyle and access to transport and medical facilities in highly deprived areas is poor, and thus, patients from these residential areas have poorer health status than those from less deprived areas.

8.3.2 Clinical relevance of findings

Age and gender

Short term study concluded that OR of death was higher among elderly patients after the surgery indicating that the longer elderly patients are made to

wait for their THR, the higher is the immediate post-operative mortality risk. The researched study population comprised of a majority of elderly individuals who were aged 60 years or more and therefore included riskier age groups with higher prevalence of medical comorbidities (obesity, stroke, angina, myocardial infarction, hypertension and hypercholesterolemia). However in the long run, this research found that age at surgery did not significantly affect all-cause HR of death. This indicates that for patients who survived the early post-operative months, THR procedure is beneficial to each age group.

Gender was also an important factor for short and long term mortality after THR procedure. Male cases had higher mortality risk than female cases and controls. Interaction between gender and types of procedures was only significant in the short term with the highest OR of death attributed to cemented procedures, followed by uncemented and other types of procedure among male cases. This interaction disappeared in the long term. Long term HR of death associated with different types of procedures did not differ by gender. These short and long term findings classified both age and gender as risk factors for mortality after THR and similar conclusions were reported by Visuri et al. (1997), Lie et al. (2000), NHS Scotland (2002), NHS Scotland (2002), Ramiah et al. (2007), Aynardi et al. (2009), Pedersen et al. (2011), Mäkelä et al. (2014) and Lovald et al. (2014), respectively.

Cemented versus cementless procedures

For both short and long term studies, mortality risk of THR cases with no preoperative medical comorbidities was the highest for cemented procedures, followed by uncemented and other types of procedures, respectively, compared to matched controls. Cemented procedures, which are also described as a biologic fixation, have been strongly linked with a progressive time dependent loss of fixation function (Amstutz et al. (1988), Hozack et al. (1990)) occurring due to loosening of the artificial joint (Harris and McGann (1986)). This is a common cause of complications after surgery according to systematic review by Abdulkarim et al. (2013) and is a determinant factor in causing a higher mortality risk for cemented procedures, in contrast to uncemented or hybrid THR. On the contrary, uncemented procedures have been associated with greater joint fixation stability (Engh et al. (2006)) because loosening of the joint in cementless procedures were less common and not found to cause further complications (Röder et al. (2010)).

Nevertheless, despite the proven benefits of cementless THR, such procedure still remains the second most used type of fixation technique for THR, after cemented ones, firstly, due to the high cost of the uncemented components (Unnanuntana et al. (2009)) and, secondly, because of the posible formation of blood clots in deep veins around the joint after surgery which eventually travel up the leg to the lungs, causing pulmonary embolism. Such complication is more prevalent among uncemented procedures, compared to cemented ones (Blom et al. (2006), Aynardi et al. (2009), Pedersen et al. (2011), Hunt et al. (2013) and Jones et al. (2014)). These findings suggest that while clinicians should recommend cementless procedures, an assessment of the risk of developing pulmonary embolism should also be carried out such that the benefits of cementless procedures on mortality risk are not overshadowed by that of pulmonary embolism.

Cardiovascular and inflammatory diseases

Mortality risk of patients with inflammatory diseases such as osteoarthritis (OA) and osteoporosis was found to be higher than individuals without these conditions. Unlike the short term study, in the long run OA was a risk factor for mortality after THR, with HR of death being the lowest for cemented procedures, followed by uncemented and other types of procedures. This finding strengthens the previous comment on why cemented procedures are more prevalent than cementless ones despite being associated with lower mortality risk. This is mainly attributed to the fact that THR is a treatment allocated mostly to patients diagnosed with OA and thus, cemented procedures represent the least risky procedure compared to cementless procedures. OA is the main cause of the degenerative wear and tear condition that causes painful hip joint inflammation and dysfunction (Berenbaum (2013)). The most common treatment modality for symptomatic osteoarthritis is joint replacement, a procedure that itself brings about degenerative inflammation like OA (Haynes et al. (2004)), thereby, explaining the perceived increase in mortality risk after surgery for patients with OA and osteoporosis. Diagnosis of joint OA and osteoporosis before the procedure should be carried out and followed by clinicians since such information could help in selecting the type of THR procedure to carry out.

Cardiovascular conditions such as stroke, angina and myocardial infarc-

tion (MI) increased mortality risk among the study population and these conform to results from previous studies by Gaston et al. (2007) and Pedersen et al. (2011). Firstly, no difference in mortality risk between cases and controls, both with either stroke or angina was observed. Thus selection of fixation type (cemented versus cementless) can be independent of their medical history of stroke and angina. However for cases with a preoperative history of MI, mortality risk varied for different types of procedures with HR of death being lowest among uncemented procedures, followed by other types of THR and cemented procedures. This finding showed that a history of preoperative MI increased mortality risk among patients who underwent cemented procedures and can therefore aid clinicians in advising patients with a history of MI towards cementless procedures such as hybrid or uncemented hip joint replacement. Inflammatory diseases such as OA is the most common diagnosis for THR procedures. Hansson and Hermansson (2011) and Libby and Hansson (2012, 2015) showed that the pathological process of cardiovascular disease is causally related to inflammatory process so that conditions like OA increased the risk of cardiovascular events (Kaplan and McCune (2003), Mason and Libby (2014)).

Hypercholesterolemia

Hypercholesterolemia (HC) was associated with an increase in short term OR of death for all patients but did not vary for different types of procedures. In the long run however, HC was a risk factor for mortality varying across different types of procedures. Table 26 summarises the effect of hypercholesterolemia on the long term survival after THR.

Table 26: Effects of hypercholesterolemia on long term survival for different procedure types

Increasing HR of death	**No HC**	**No HC due to statins prescription**	**HC despite statins prescription**	**HC but not on statins prescription**
Lowest HR	Uncemented	Uncemented	Controls	Controls
	Cemented	Cemented	Cemented	Uncemented
	Other types	Other types	Uncemented	Other types
Highest HR	Controls	Controls	Other types	Cemented

Patients with no HC due to statins prescription may be directed to uncemented procedures as this will minimise the risk of mortality after the procedure, in comparison to cemented ones. Additionally, cases with HC

and statin prescription have reduced mortality risk compared to those with HC but not on statins; a valid finding also reported by Gitsels et al. (2016). Therefore, patients with HC but not on statins should be referred to their general practitioner so that they can be prescribed statins before they undergo their THR procedure since this will reduce their potential mortality risk after the procedure.

Body mass index

In the short term, OR of death for overweight (26-30 kg/m^2) and obese ($> 30kg/m^2$) patients were higher than individuals with normal BMI (19-25 kg/m^2). Significant interaction between BMI and procedure types showed OR of death was higher for cemented procedures compared to cementless procedures, for both obese and overweight patients; a result also reported bu Hunt et al. (2013) and Jämsen et al. (2013), respectively. However this excess mortality risk disappears in the long term as being overweight and obese before the procedure, resulted in a protective effect on long term HR of death, with uncemented procedures resulting in the highest improvement in HR, followed by cemented and other types, respectively, for both overweight and obesity; a result in accordance with Adams et al. (2006), Bakaeen and Chu (2011) and Stamou et al. (2011).

In this research, approximately two-thirds of the THR cases were either overweight or obese before the procedure and had higher prevalence for diabetes, hypertension, hypercholesterolemia and osteoarthritis, respectively, across all age groups, compared to THR cases with normal BMI. However obese and overweight patients had their surgical procedure at a younger age compared with normal BMI patients, suggesting that obese and overweight patients should most likely be referred for THR procedure at younger age to experience the protective effect on mortality risk in the long term.

The protective effect of overweight and obesity on long term survival after THR procedures is a paradoxical finding because high BMI is associated with an increased risk of developing conditions such as coronary heart diseases, hypercholesterolemia or hypertension. However, once the condition is manifested, overweight or obesity protects against increasing mortality risk when compared to patients with normal BMI. This result is also observed among obese patients for other chronic conditions and in heart surgeries (Curtis

et al. (2005), Bakaeen and Chu (2011), Stamou et al. (2011) and Haass et al. (2011).

Type 2 Diabetes and smoking status

Type 2 diabetes increased short and long term mortality risk for all patients in this research, but did not vary for different types of THR procedures. Furthermore, there was significant interaction between gender and diabetes showing that diagnosis of type 2 diabetes and being a male patient, increased mortality risk. This is in agreement with the results of Pedersen et al. (2011) and Barrett et al. (2005) who reported type 2 diabetes as a risk factor for mortality when comparing THR cases to controls. It is well established that mortality risk for smokers is significantly higher than non-smokers (Doll et al. (2004)) due to the various health complications it causes. Mortality risk estimated in this research was adjusted for smoking status and findings showed significant increase in mortality among smokers and ex-smokers compared to non-smokers, after the procedure.

Deprivation index and socio-economic classification

Short term OR of death did not vary significantly across residential areas with various levels of deprivation and across socio-economic classifications. However, variations in long term survival were significantly different between residential areas with low and high deprivation level, measured by Townsend score. Long term HR of death increased for patients residing in areas classified as *highly deprived* (Townsend score > 3). Similarly patients from less wealthy socio-economic groups, indexed using post code based Mosaic classification, had a higher estimated HR of death than patients from the most wealthy and affluent areas.

Deprivation level was also reported to significantly increase mortality risk after THR procedures among patients from highly deprived residential areas by Clement et al. (2011), Whitehouse et al. (2014) and Xu et al. (2017). However, no previous studies adjusted for socio-economic factors in their long term survival analysis. One reason for this may be that data and census tools needed to classify patients by their socio-economic factors were not readily available, in contrast to this research. Patients from deprived residential areas (Townsend score > 3) and poor socio-economic groups (Mosaic category

G-O) have limited access to a public transport, consist of families who are not house owners and are mostly unemployed, leading to less healthy lifestyle in comparison to less deprived areas and more affluent socio-economic groups.

Blak et al. (2011) demonstrated significant association between poor health outcomes such as death and social deprivation using primary care data; a finding which validates the results of this research where mortality risk after THR increases in highly deprived residential areas and for patients from poor socio-economic groups. Furthermore, mortality risk varied significantly between general practices in this research. This result is supported by the conclusion of Blak et al. (2011) which helps to explain why some GP practices are associated with poor survival after THR procedure, both in the short and long term. GP practices in highly deprived residential areas are more likely to provide care for patients with poor health and thus associated with higher risk of mortality. This research demonstrated the impact of residential deprivation and patients' socio-economic classification on survival after THR procedures and this points that post code related variables can be used as an informative tool for clinicians to have a preliminary assessment of the THR case health status before the procedure.

8.3.3 Actuarial relevance of findings

Understanding and management of the mortality risks from various medical interventions and treatments is an important part of actuarial research in the area of life insurance and pensions. The pricing of life assurance and pension products sold to customers is highly dependent on the assumption or model used to estimate mortality risks after the medical intervention. In actuarial terms, mortality risk is defined as the uncertainty caused by future significant changes in mortality rates (Barrieu et al., (2012)). With the increasing number of THR procedures, the potential number of customers for life insurance companies undergoing THR procedures rises. An accurate estimation and interpretation of mortality risk of customers who underwent THR procedure is therefore highly informative for the insurance industry. This research demonstrated a temporary higher risk of death among individuals who had the procedure in comparison to those who did not have THR. In the long term however, given that the individuals survived the first 24 months, THR procedure is found to improve survival compared to those who did not have the intervention.

This variation affects the mortality risk assumptions and pricing for life assurance and pension products. Traditionally for life assurance products, individuals are required to go through a series of medical check-ups that help to estimate the mortality risk of the individuals so that their contract can be underwritten and priced accordingly. THR procedure is excluded from the medical history requested by underwriters. The findings of this research demonstrated that THR procedures significantly impact on mortality risk. Failing to account for the impact of THR procedure on mortality will classify individuals who had THR in the same cohort as those without the procedure and will thus give rise to basis risk. Under such a scenario, the overall profitability of the portfolio of life assurance products will consequently be affected since appropriate mortality risk assumptions was not tuned in the pricing of the products. Therefore there is an incentive for actuaries and underwriters to request information about the timing of THR procedure in order to price their insurance products using the appropriate mortality risk assumptions.

THR procedure is a surgical intervention mostly performed on people aged more than 55 years. Thus it is a procedure that affects the cohort of individuals either close to retirement age or pensioners receiving their benefits. The findings of the long term survival analysis in this research showed improvement in mortality risk, suggesting that life expectancy of THR cases to exceed that of controls. This information can aid actuaries and pension fund managers firstly, to adjust the pricing of their pension products such that the fund collected is enough to cover for the increase in life expectancy among those who have THR before retirement, and secondly, to adjust the benefits paid out to pensioners who have THR during retirement period so that the pension fund is not outlived by the individuals life expectancy. Not adjusting variation in mortality risk assumption due to THR procedure, may lead to an increasing liability for actuaries if the pension pot is not big enough to pay the financial benefits to individuals who had THR.

8.4 Strengths

The study population in this research was obtained using primary care records that were representative of the UK population after adjusting for gender, age and deprivation (Blak et al. (2011), Langley et al. (2011), González

et al. (2009), Hippisley-Cox and Coupland (2010a), MacDonald and Morant (2008)). These primary care records consist of medical histories that provide extensive insight on the clinical practice of medical conditions, diagnoses, interventions and treatments among the general population. Similarly, primary care data provide extensive socio-demographic, lifestyle factors and medical information. Therefore, selected THR cases and matched controls in this research were more representative of THR patients in the UK than previous UK based studies which selected patients either through hospital admissions or using secondary care databases. Furthermore, this research selected THR cases and matched controls from the same population source and thus, comparisons between THR cases and controls are valid because using cases and controls from the same source eliminates selection bias (Raboud and Breslow 1989).

The mean follow-up period for the long term study in this research was 9.8 years. The long follow-up allowed monitoring of changes in medical conditions, interventions, treatments and lifestyle factors of patients over time, thereby providing detailed insights on the past and present health status of the study population. Hence potential risk factors for variations in long term hazards of death could be identified by testing whether the effects of medical conditions, interventions, treatments and lifestyle factors on long term survival, vary over time. Similarly, the long follow-up period meant that the event of interest (death) could be observed over a long period of time, thereby estimating hazards of death more accurately. With an increasing life expectancy in UK population, estimating hazards of death more accurately becomes more important for medical management and resource allocation as well as for retirement planning.

The short and long term mortality risk models developed in this research are designed as matched cohort studies in which THR cases are compared to matched controls from the same general population. Such a design permits the estimation of the effect of THR procedure on short and long term survival between patients who underwent the procedures (cases) and those who did not (controls). Furthermore, the estimated short term odds and long term hazards of death were both adjusted for a range of risk factors that are known to cause variations in survival following THR procedures, thereby improving precision of estimated effects on mortality risk after the procedure. Interactions between all risk factors were tested and not limited to

testing only interactions between the main exposure, gender and age group. This allowed the investigation of variations in short and long term mortality risk in more comprehensive details, thereby permitting the formulation of recommendations for patients with different medical histories, interventions, treatments and retirement plans.

Patients in the short and long term studies were grouped by their GP practice and were assumed to be exposed to the same level of unmeasured frailty within the GP practice. This permits to include correlations of patients within each GP practice. Therefore, inferences from the results of this research can not only be extended to the GP practices in the short and long term study, but also to the whole UK population (Brown and Prescott (2014)).

Finally, the methodology used to develop short and long term mortality risk models, permits the use of the research findings for medical management of patients by healthcare professionals and for planning of retirement benefits. Presented short and long term survival models include risk factors that are routinely collected by primary care providers in the UK and are available patients. Therefore estimates of mortality risk in the short and long term can be determined for patients with various medical conditions and lifestyle factors. Similarly, the findings of this research are important and instructive for medical management of patients with respect to the administration of their health status, ongoing treatments and medical interventions. With respect to planning retirement period and benefits, the findings of this research are useful for individuals who want to estimate their life expectancy after THR procedures for financial planning during retirement, for actuaries to determine the price of annuity products for the insurance industry and for assisting the government or other regulatory bodies in the UK to bring changes to the UK pension system.

8.5 Limitations

National death records for UK population are not linked with the THIN database. Moving to a new GP practice may also indicate a change in patients residential area. According to Uren and Goldring (2007), national trends for the UK population show that patients from the age group 60-69 moves from more affluent areas in good health whereas the age group 70 or

older move from deprived areas in worse health conditions. The mortality risk models developed in this research, assumed that patients who transferred to a new GP practice, had similar mortality rate to patients who did not move to a new GP practice. This assumption was tested via sensitivity analysis and it was concluded that transfer outs were not informative of changes in short and long term mortality risks after THR procedures.

This research focussed on analysing the impact of degenerative and chronic conditions such as arthritis on survival of THR cases after unilateral THR procedure. Exclusion of bilateral THR procedures from the analysis may limit the generalisability of the findings to the UK population in the sense that the results may only apply to individuals having unilateral procedure. A further research using time dependent survival model that accounts for the time span between the two THR procedures may be employed to estimate the effect of mortality risk after the bilateral procedure.

Gossec et al. (2005), Schäfer et al. (2010) and Smith et al. (2012) reported that age, gender, body mass index, duration of symptoms affecting the normal functioning of the hip joint, measurement level of hip pain and movement using indices such as the Oxford hip score (Wylde et al. (2005)), the WOMAC (Wolfe (1999)) and the Harris hip score (Nilsdotter and Bremander (2011)), respectively, are predictors of THR procedure. Among these reported predictors for THR, age, gender and body mass index were also found to be predictors for mortality risk after the procedure. THIN data has no information with respect to measurement indices such as the WOMAC, Oxford and Harris hip score and thus, the analysis did not adjust for these predictors of THR.

The long follow-up period for the long term study meant that over time, the reasons for, and methods of collecting data by general practitioners, have changed (Marston et al. (2010)). For instance, pre-1990 data collection in primary care databases required no clinical audits but this process became contractual in the early 1990 while, as from 2004, the Quality and Outcomes Framework (QOF) was set up to provide financial incentives for general practitioners to collect data. Consequently, quality of primary care records and management of common chronic diseases, preventative and therapeutic measures, medical interventions and lifestyle choices improved (Campbell et al. (2007), Langley et al. (2011), Sharma et al. (2010)). In this research, no

significant interactions between year of birth category of patients with other risk factors, were observed for the short and long term mortality models. Therefore, short term odds and long term hazards of death following THR procedures did not vary over time.

Medical records extracted from the THIN database are incomplete for lifestyle factors (BMI and smoking status) and post-code related variables (Townsend scores, Index of multiple deprivation and Mosaic group). Accounting for missingness during the data analysis phase for the short and long term mortality risk models required additional analyses and model assumptions, thereby reducing the precision of the estimates from both models. Missingness in this was dealt with by employing the method of multiple imputation. The distributions of complete records are close to that of the imputed ones, while the survival model based on cases with complete records only, estimated similar hazard ratios and performance statistics as the model based on the imputed dataset, for both short and long term model.

In THIN database, data on the type of prosthesis, surgical approach and surgeon experience are not available. Therefore estimates of short and long term mortality risk in this study were not adjusted for these variables, and survival after THR could not be distinguished for different types of prosthesis, surgical approach and surgeon experience. Additionally, drug therapy was included as a confounder for both short and long term models. However one limitation with this approach is that adherence of patients to the drug therapy is unknown, thereby, not precisely reflecting the effects of drug therapy on mortality risks following THR.

A THIN Read Codes Summary Tables

Table 27: Read codes in THIN database to identify types of THR procedures (THIN Data Guide (2011))

Read-codes	Description in THIN database	Procedure Type
7K20.00	Total prosthetic replacement of hip joint using cement	Cemented THR
7K20.11	Arthroplasty of hip joint using cement	Cemented THR
7K20.12	Arthroplasty of hip joint using cement	Cemented THR
7K20.13	Arthroplasty of hip joint using cement	Cemented THR
7K20.14	Arthroplasty of hip joint using cement	Cemented THR
7K20.15	Arthroplasty of hip joint using cement	Cemented THR
7K20.16	Arthroplasty of hip joint using cement	Cemented THR
7K20.17	Arthroplasty of hip joint using cement	Cemented THR
7K20.18	Arthroplasty of hip joint using cement	Cemented THR
7K20.19	Arthroplasty of hip joint using cement	Cemented THR
7K20.1A	Arthroplasty of hip joint using cement	Cemented THR
7K20.1B	Arthroplasty of hip joint using cement	Cemented THR
7K20.1C	Arthroplasty of hip joint using cement	Cemented THR
7K20.1D	Arthroplasty of hip joint using cement	Cemented THR
7K20.1E	Arthroplasty of hip joint using cement	Cemented THR
7K20.1F	Arthroplasty of hip joint using cement	Cemented THR
7K20.1G	Arthroplasty of hip joint using cement	Cemented THR
7K20000	Primary cemented total hip replacement	Cemented THR
7K20011	Charnley cemented total hip replacement	Cemented THR
7K20300	Primary hybrid total replacement of hip joint NEC	Other types
7K20y00	Total prosthetic replacement of hip joint using cement OS	Cemented THR
7K20z00	Total prosthetic replacement of hip joint using cement NOS	Cemented THR
7K21.00	Total prosthetic replacement of hip joint not using cement	Uncemented THR
7K21.11	Freeman total replacement of hip joint not using cement	Uncemented THR
7K21.12	Freeman total replacement of hip joint not using cement	Uncemented THR
7K21.13	Freeman total replacement of hip joint not using cement	Uncemented THR
7K21.14	Freeman total replacement of hip joint not using cement	Uncemented THR
7K21.15	Freeman total replacement of hip joint not using cement	Uncemented THR

7K21.16	Freeman total replacement of hip joint not using cement	Uncemented THR
7K21.17	Freeman total replacement of hip joint not using cement	Uncemented THR
7K21000	Primary uncemented total hip replacement	Uncemented THR
7K21y00	Total prosthetic replacement hip joint not using cement OS	Uncemented THR
7K21z00	Total prosthetic replacement hip joint not using cement NOS	Uncemented THR
7K22.00	Other total prosthetic replacement of hip joint	Other types
7K22.11	Other arthroplasty of hip joint	Other types
7K22.12	Other arthroplasty of hip joint	Other types
7K22000	Primary total prosthetic replacement of hip joint NEC	Other types
7K22011	Primary hybrid total replacement of hip joint NEC	Other types
7K22y00	Other specified total prosthetic replacement of hip joint	Other types
7K22z00	Total prosthetic replacement of hip joint NOS	Other types

Table 28: Read codes in THIN database to identify medical conditions (THIN Data Guide (2011))

Medical Condition	**Read-code**	**Description in THIN Database**
Diabetes Mellitus (DM)	1434	H/O: diabetes mellitus
	C10..00	Diabetes mellitus
	C100011	Insulin dependent diabetes mellitus
	C100112	Non-insulin dependent diabetes mellitus
	C108.00	Insulin dependent diabetes mellitus
	C108.12	Type 1 diabetes mellitus
	C108.13	Type I diabetes mellitus
	C108400	Unstable insulin dependent diabetes mellitus
	C108411	Unstable type I diabetes mellitus
	C108412	Unstable type 1 diabetes mellitus
	C109.00	Non-insulin dependent diabetes mellitus
	C109.12	Type 2 diabetes mellitus
	C109.13	Type II diabetes mellitus
	C10E.00	Type 1 diabetes mellitus
	C10E.11	Type I diabetes mellitus
	C10E.12	Insulin dependent diabetes mellitus
	C10E400	Unstable type 1 diabetes mellitus
	C10E411	Unstable type I diabetes mellitus
	C10E412	Unstable insulin dependent diabetes mellitus
	C10F.00	Type 2 diabetes mellitus
	C10F.11	Type II diabetes mellitus
	Cyu2.00	Diabetes mellitus
	Cyu2000	Other specified diabetes mellitus
Angina	G311500	Acute coronary syndrome
	14A5.00	H/O: angina pectoris
	G33..00	Angina pectoris
	G331.11	Variant angina pectoris
	G33z.00	Angina pectoris NOS
	G33zz00	Angina pectoris NOS
	Gyu3000	Other forms of angina pectoris
	G33z400	Ischaemic chest pain

	G332.00	Coronary artery spasm
	G331.00	Prinzmetal's angina
	G311.13	Unstable angina
	G311100	Unstable angina
	G33z700	Stable angina
	G331.11	Variant angina pectoris
Heart Attack (HA)	G30..14	Heart attack
	14AH.00	H/O: Myocardial infarction in last year
	14AT.00	History of myocardial infarction
	G30..00	Acute myocardial infarction
	G30..13	Cardiac rupture following myocardial infarction (MI)
	G30..15	MI - acute myocardial infarction
	G30..17	Silent myocardial infarction
	G301.00	Other specified anterior myocardial infarction
	G301z00	Anterior myocardial infarction NOS
	G304.00	Posterior myocardial infarction NOS
	G305.00	Lateral myocardial infarction NOS
	G306.00	True posterior myocardial infarction
	G307100	Acute non-ST segment elevation myocardial infarction
	G308.00	Inferior myocardial infarction NOS
	G30B.00	Acute posterolateral myocardial infarction
	G30X.00	Acute transmural myocardial infarction of unspecif site
	G30X000	Acute ST segment elevation myocardial infarction
	G30y.00	Other acute myocardial infarction
	G30yz00	Other acute myocardial infarction NOS
	G30z.00	Acute myocardial infarction NOS
	G310.00	Postmyocardial infarction syndrome
	G311000	Myocardial infarction aborted
	G311011	MI - myocardial infarction aborted
	G312.00	Coronary thrombosis not resulting in myocardial infarction
	G32..00	Old myocardial infarction
	G32..12	Personal history of myocardial infarction
	G35..00	Subsequent myocardial infarction
	G350.00	Subsequent myocardial infarction of anterior wall
	G351.00	Subsequent myocardial infarction of inferior wall

	G353.00	Subsequent myocardial infarction of other sites
	G35X.00	Subsequent myocardial infarction of unspecified site
	G38..00	Postoperative myocardial infarction
	G380.00	Postoperative transmural myocardial infarction anterior wall
	G381.00	Postoperative transmural myocardial infarction inferior wall
	G382.00	Postoperative transmural myocardial infarction other sites
	G383.00	Postoperative transmural myocardial infarction unspec site
	G384.00	Postoperative subendocardial myocardial infarction
	G38z.00	Postoperative myocardial infarction, unspecified
	Gyu3400	Acute transmural myocardial infarction of unspecif site
	G311500	Acute coronary syndrome
	G30..12	Coronary thrombosis
Stroke	G61..12	Stroke due to intracerebral haemorrhage
	G64..13	Stroke due to cerebral arterial occlusion
	G66..00	Stroke and cerebrovascular accident unspecified
	G66..12	Stroke unspecified
	ZV12511	Personal history of stroke
	L440.11	CVA - cerebrovascular accident in the puerperium
	ZV12512	Personal history of cerebrovascular accident (CVA)
	1477	H/O: cerebrovascular disease
	1JA1000	Suspected cerebrovascular accident
Osteoarthritis (OA)	N05..00	Osteoarthritis and allied disorders
	N05..11	Osteoarthritis
	N050.00	Generalised osteoarthritis - OA
	N050000	Generalised osteoarthritis of unspecified site
	N050100	Generalised osteoarthritis of the hand
	N050200	Generalised osteoarthritis of multiple sites
	N050z00	Generalised osteoarthritis NOS
	N051.00	Localised, primary osteoarthritis
	N051000	Localised, primary osteoarthritis of unspecified site
	N051100	Localised, primary osteoarthritis of the shoulder region
	N051200	Localised, primary osteoarthritis of the upper arm
	N051300	Localised, primary osteoarthritis of the forearm
	N051400	Localised, primary osteoarthritis of the hand
	N051500	Localised, primary osteoarthritis of the pelvic region/thigh

N051600	Localised, primary osteoarthritis of the lower leg
N051700	Localised, primary osteoarthritis of the ankle and foot
N051800	Localised, primary osteoarthritis of other specified site
N051D00	Localised, primary osteoarthritis of the wrist
N051E00	Localised, primary osteoarthritis of toe
N051F00	Localised, primary osteoarthritis of elbow
N051z00	Localised, primary osteoarthritis NOS
N052.00	Localised, secondary osteoarthritis
N052000	Localised, secondary osteoarthritis of unspecified site
N052100	Localised, secondary osteoarthritis of the shoulder region
N052200	Localised, secondary osteoarthritis of the upper arm
N052300	Localised, secondary osteoarthritis of the forearm
N052400	Localised, secondary osteoarthritis of the hand
N052500	Localised, secondary osteoarthritis of pelvic region/thigh
N052600	Localised, secondary osteoarthritis of the lower leg
N052700	Localised, secondary osteoarthritis of the ankle and foot
N052800	Localised, secondary osteoarthritis of other specified site
N052z00	Localised, secondary osteoarthritis NOS
N053.00	Localised osteoarthritis, unspecified
N053000	Localised osteoarthritis, unspecified, of unspecified site
N053100	Localised osteoarthritis, unspecified, of shoulder region
N053200	Localised osteoarthritis, unspecified, of the upper arm
N053300	Localised osteoarthritis, unspecified, of the forearm
N053400	Localised osteoarthritis, unspecified, of the hand
N053500	Localised osteoarthritis, unspecified, pelvic region/thigh
N053600	Localised osteoarthritis, unspecified, of the lower leg
N053611	Patellofemoral osteoarthritis
N053700	Localised osteoarthritis, unspecified, of the ankle and foot
N053800	Localised osteoarthritis, unspecified, of other spec site
N053z00	Localised osteoarthritis, unspecified, NOS
N054.00	Oligoarticular osteoarthritis, unspecified
N054000	Oligoarticular osteoarthritis, unspec, of unspecified sites
N054100	Oligoarticular osteoarthritis, unspecified, of shoulder
N054200	Oligoarticular osteoarthritis, unspecified, of upper arm
N054300	Oligoarticular osteoarthritis, unspecified, of forearm

N054400	Oligoarticular osteoarthritis, unspecified, of hand
N054500	Oligoarticular osteoarthritis, unspecified, of pelvis/thigh
N054600	Oligoarticular osteoarthritis, unspecified, of lower leg
N054700	Oligoarticular osteoarthritis, unspecified, of ankle/foot
N054800	Oligoarticular osteoarthritis, unspecified, other spec sites
N054900	Oligoarticular osteoarthritis, unspecified, multiple sites
N054z00	Osteoarthritis of more than one site, unspecified, NOS
N05z.00	Osteoarthritis NOS
N05z000	Osteoarthritis NOS, of unspecified site
N05z100	Osteoarthritis NOS, of shoulder region
N05z200	Osteoarthritis NOS, of the upper arm
N05z211	Elbow osteoarthritis NOS
N05z300	Osteoarthritis NOS, of the forearm
N05z311	Wrist osteoarthritis NOS
N05z400	Osteoarthritis NOS, of the hand
N05z411	Finger osteoarthritis NOS
N05z412	Thumb osteoarthritis NOS
N05z500	Osteoarthritis NOS, pelvic region/thigh
N05z511	Hip osteoarthritis NOS
N05z600	Osteoarthritis NOS, of the lower leg
N05z611	Knee osteoarthritis NOS
N05z700	Osteoarthritis NOS, of ankle and foot
N05z711	Ankle osteoarthritis NOS
N05z712	Foot osteoarthritis NOS
N05z713	Toe osteoarthritis NOS
N05z800	Osteoarthritis NOS, other specified site
N05z900	Osteoarthritis NOS, of shoulder
N05zA00	Osteoarthritis NOS, of sternoclavicular joint
N05zB00	Osteoarthritis NOS, of acromioclavicular joint
N05zC00	Osteoarthritis NOS, of elbow
N05zD00	Osteoarthritis NOS, of distal radio-ulnar joint
N05zE00	Osteoarthritis NOS, of wrist
N05zF00	Osteoarthritis NOS, of MCP joint
N05zG00	Osteoarthritis NOS, of PIP joint of finger
N05zH00	Osteoarthritis NOS, of DIP joint of finger

	N05zJ00	Osteoarthritis NOS, of hip
	N05zK00	Osteoarthritis NOS, of sacro-iliac joint
	N05zL00	Osteoarthritis NOS, of knee
	N05zM00	Osteoarthritis NOS, of tibio-fibular joint
	N05zN00	Osteoarthritis NOS, of ankle
	N05zP00	Osteoarthritis NOS, of subtalar joint
	N05zQ00	Osteoarthritis NOS, of talonavicular joint
	N05zR00	Osteoarthritis NOS, of other tarsal joint
	N05zS00	Osteoarthritis NOS, of 1st MTP joint
	N05zT00	Osteoarthritis NOS, of lesser MTP joint
	N05zU00	Osteoarthritis NOS, of IP joint of toe
	N05zz00	Osteoarthritis NOS
	N11..12	Osteoarthritis of spine
	N110.12	Osteoarthritis cervical spine
	N11D.00	Osteoarthritis of spine
	N11D000	Osteoarthritis of cervical spine
	N11D100	Osteoarthritis of thoracic spine
	N11D200	Osteoarthritis of lumbar spine
	N11D300	Osteoarthritis of spine NOS
	N11z.11	Osteoarthritis spine
Rheumatoid Arthritis (RA)	Nyu1000	[X]Rheumatoid arthritis+involvement/other organs or systems
	Nyu1100	[X]Other seropositive rheumatoid arthritis
	Nyu1200	[X]Other specified rheumatoid arthritis
	Nyu1G00	[X]Seropositive rheumatoid arthritis, unspecified
	N04..00	Rheumatoid arthritis and other inflammatory polyarthropathy
	N040.00	Rheumatoid arthritis
	N040000	Rheumatoid arthritis of cervical spine
	N040100	Other rheumatoid arthritis of spine
	N040200	Rheumatoid arthritis of shoulder
	N040300	Rheumatoid arthritis of sternoclavicular joint
	N040400	Rheumatoid arthritis of acromioclavicular joint
	N040500	Rheumatoid arthritis of elbow
	N040600	Rheumatoid arthritis of distal radio-ulnar joint
	N040700	Rheumatoid arthritis of wrist
	N040800	Rheumatoid arthritis of MCP joint

	N040900	Rheumatoid arthritis of PIP joint of finger
	N040A00	Rheumatoid arthritis of DIP joint of finger
	N040B00	Rheumatoid arthritis of hip
	N040C00	Rheumatoid arthritis of sacro-iliac joint
	N040D00	Rheumatoid arthritis of knee
	N040E00	Rheumatoid arthritis of tibio-fibular joint
	N040F00	Rheumatoid arthritis of ankle
	N040G00	Rheumatoid arthritis of subtalar joint
	N040H00	Rheumatoid arthritis of talonavicular joint
	N040J00	Rheumatoid arthritis of other tarsal joint
	N040K00	Rheumatoid arthritis of 1st MTP joint
	N040L00	Rheumatoid arthritis of lesser MTP joint
	N040M00	Rheumatoid arthritis of IP joint of toe
	N040P00	Seronegative rheumatoid arthritis
	N040S00	Rheumatoid arthritis - multiple joint
	N040T00	Flare of rheumatoid arthritis
	N047.00	Seropositive errosive rheumatoid arthritis
	N04X.00	Seropositive rheumatoid arthritis, unspecified
Osteoporosis	N330.00	Osteoporosis
	N330300	Idiopathic osteoporosis
	N330800	Localized osteoporosis - Lequesne
	N330z00	Osteoporosis NOS
	NyuBC00	[X]Osteopenia
Chronic Kidney Disease (CKD)	1Z10.00	Chronic kidney disease stage 1
	1Z11.00	Chronic kidney disease stage 2
	1Z12.00	Chronic kidney disease stage 3
	1Z13.00	Chronic kidney disease stage 4
	1Z14.00	Chronic kidney disease stage 5
	1Z15.00	Chronic kidney disease stage 3A
	1Z16.00	Chronic kidney disease stage 3B
	1Z17.00	Chronic kidney disease stage 1 with proteinuria
	1Z18.00	Chronic kidney disease stage 1 without proteinuria
	1Z19.00	Chronic kidney disease stage 2 with proteinuria
	1Z1A.00	Chronic kidney disease stage 2 without proteinuria
	1Z1B.00	Chronic kidney disease stage 3 with proteinuria

	1Z1C.00	Chronic kidney disease stage 3 without proteinuria
	1Z1D.00	Chronic kidney disease stage 3A with proteinuria
	1Z1E.00	Chronic kidney disease stage 3A without proteinuria
	1Z1F.00	Chronic kidney disease stage 3B with proteinuria
	1Z1G.00	Chronic kidney disease stage 3B without proteinuria
	1Z1H.00	Chronic kidney disease stage 4 with proteinuria
	1Z1J.00	Chronic kidney disease stage 4 without proteinuria
	1Z1K.00	Chronic kidney disease stage 5 with proteinuria
	1Z1L.00	Chronic kidney disease stage 5 without proteinuria
Hypercholesterolemia	C320.00	Pure hypercholesterolaemia
	C320y00	Other specified pure hypercholesterolaemia
	C320z00	Pure hypercholesterolaemia NOS
	C320400	Fredrickson's hyperlipoproteinaemia, type IIa
	C328.00	Dyslipidaemia
Hypertension	G20..00	Essential hypertension
	G203.00	Diastolic hypertension
	G20z.00	Essential hypertension NOS
	G20z.11	Hypertension NOS
	14A2.00	H/O: hypertension
Blood Pressure	1005010500	Blood pressure measurement
Cholesterol Measurement	1001400017	Serum cholesterol
	1001400182	Lipoprotein electrophesis
	1001400054	Very low density lipoprotein
	1001400035	Low Density Lipoprotein
	1001400031	High Density Lipoprotein
	1001400239	Blood lipid ratios
	1001400069	Blood lipids
	1001400045	Triglycerides
Hip Fracture (HF)	122E.00	No family history of hip fracture in first degree relative
	12I4.00	FH: Maternal hip fracture
	12I5.00	FH: Hip fracture in first degree relative
	12I8.00	FH: maternal hip fracture before age 75
	14G7.00	H/O: hip fracture
	38DB.00	WHO FRAX 10 year hip fracture probability score
	7P20100	Delivery of rehabilitation for hip fracture

S30..11	Hip fracture
S30y.11	Hip fracture NOS

Table 29: Description of Mosaic Groups (Experian Ltd (2004))

Mosaic Group	Sub-Group	Description
City Prosperity - A	World-Class Wealth	Global high flyers and families of privilege living luxurious lifestyles in London's most exclusive boroughs
	Uptown Elite	accessible inner suburbs where they enjoy city High status households owning elegant homes in
	Penthouse Chic	City suits renting premium-priced flats in prestige central locations where they work hard and play hard
	Metro High-Flyers	Ambitious 20 and 30-somethings renting expensive apartments in highly commutable areas of major cities
Prestige Positions - B	Premium Fortunes	Influential families with substantial income established in distinctive, expansive homes in wealthy enclaves
	Diamons Days	Retired residents in sizeable homes whose finances are secured by significant assets and generous pensions
	Alpha Families	High-achieving families living fast-track lives, advancing careers, finances and their school-age kids' development
	Bank of Mum and Dad	Well-off families in upmarket suburban homes where grown-up children benefit from continued financial support
	Empty-Nest Adventure	Mature couples in comfortable detached houses who have the means to enjoy their empty-nest status
Country Living - C	Wealthy Landowners	Prosperous owners of country houses including the rural upper class, successful farmers and second- home owners
	Rural Vogue	Country-loving families pursuing a rural idyll in comfortable village homes while commuting some distance to work
	Scattered Homesteads	Older households appreciating rural calm in stand- alone houses within agricultural landscapes
	Village Retirement	Retirees enjoying pleasant village locations with amenities to service their social and practical needs
Domestic Success - D	Satellite Settlers	Mature households living in expanding developments around larger villages with good transport links
	Local Focus	Rural families in affordable village homes who are reliant on the local economy for jobs
	Outlying Seniors	Pensioners living in inexpensive housing in out of the way locations
	Far-Flung Outposts	Mid-income families in remote areas with long travel times to larger towns
Senior Security - E	Legacy Elders	Time-honoured elders now mostly living alone in comfortable suburban homes on final salary pensions
	Bungalow Haven	Peace-seeking seniors appreciating the calm of bungalow estates designed for the elderly
	Classic Grandparents	Lifelong couples in standard suburban homes enjoying retirement through grandchildren and gardening
	Solo Retirees	Senior singles whose reduced incomes are satisfactory in their affordable but pleasant owned homes
Surburban Stability - F	Boomerang Boarders	Long-term couples with mid-range incomes whose adult children have returned to the shelter of the family home
	Family Ties	Active families with teens and adult children whose prolonged support is eating up household resources
	Fledginling Free	Pre-retirement couples with respectable incomes enjoying greater space and spare cash since children left home
	Dependable Me	Single mature owners settled in traditional suburban semis working in intermediate occupations
Domestic Success - G	Cafes and Catchments	Affluent families with growing children living in upmarket housing in city environs
	Thriving Independence	Well-qualified older singles with incomes from successful professional careers in good quality housing
	Modern Parents	Busy couples in modern detached homes juggling the demands of school-age children and careers
	Mid-Career Convention	Professional families with children in traditional mid-range suburbs where neighbours are often older
Aspiring Homemakers - H	Primary Ambitions	Forward-thinking younger families who sought affordable homes in good suburbs which they may now be out-growing
	Affordable Fringe	Settled families with children owning modest, 3-bed semis in areas where there's more house for less money
	First-Rung Futures	Pre-family newcomers who have bought value homes with space to grow in affordable but pleasant areas

	Contemporary Starts	Fashion-conscious young singles and partners setting up home in developments attractive to their peers
	New Foundations	Occupants of brand new homes who are often younger singles or couples with children
	Flying Solo	Bright young singles on starter salaries choosing to rent homes in family suburbs
Family Basics - I	Solid Economy	Stable families with children renting better quality homes from social landlords
	BudgetGenerations	Families supporting both adult and younger children where expenditure can exceed income
	ChildcareSqueeze	Younger families with children who own a budget home and are striving to cover all expenses
	Families withNeeds	Families with many children living in areas of high deprivation and who need support
Transient Renters - J	Make Do& Move On	Yet to settle younger singles and couples making interim homes in low cost properties
	DisconnectedYouth	Young people endeavouring to gain employment footholds while renting cheap flats and terraces
	StreetwiseStrain Younger	Younger hard-pressed singles in social housing with financial challenges
	Renting a Room	Transient renters of low cost accommodation often within subdivided older properties
Municipal Challenge - K	Inner CityStalwarts	Long-term renters of inner city social flats who have witnessed many changes
	CrowdedKaleidoscope	Multi-cultural households with children renting social flats in over-crowded conditions
	High RiseResidents	Renters of social flats in high rise blocks where levels of need are significant
	StreetwiseSingles	Hard-pressed singles in low cost social flats searching for opportunities
	Low IncomeWorkers	Older social renters settled in low value homes in communities where employment is harder to find
Vintage Value - L	DependentGreys	Ageing social renters with high levels of need in centrally located developments of small units
	Pocket Pensions	Penny-wise elderly singles renting in developments of compact social homes
	Aided Elderly	Including retirement homes and complexes of Supported elders in specialised accommodation in small homes
	Estate Veterans	Long-standing elderly renters of social homes who have seen neighbours change to a mix of owners and renters
	SeasonedSurvivors	Deep-rooted single elderly owners of low value properties whose modest home equity provides some security
Modest Traditions - M	Down-to-EarthOwners	Ageing couples who have owned their inexpensive home for many years while working in routine jobs
	Offspring Overspill	Lower income owners whose adult children are still striving to gain independence meaning space is limited
	Self Supporters	Hard-working mature singles who own budget terraces manageable within their modest wage
Urban Cohesion - N	CommunityElders	Established older households owning city homes in diverse neighbourhoods
	CulturalComfort	Thriving families with good incomes in multi-cultural urban communities
	Asian Heritage	Large extended families in neighbourhoods with a strong South Asian tradition
	Ageing Access	Older residents owning small inner suburban properties with good access to amenities
Rental Hubs - O	Career Builders	Motivated singles and couples in their 20s and 30s progressing in their field of work from commutable properties
	Central Pulse	Entertainment-seeking youngsters renting city centre flats in vibrant locations close to jobs and night life
	FlexibleWorkforce	Self-starting young renters ready to move to follow worthwhile incomes from service sector jobs
	Bus-RouteRenters	Singles renting affordable private flats away from central amenities and often on main roads
	Learners & Earners	Inhabitants of the university fringe where students and older residents mix in cosmopolitan locations
	Student Scene	Students living in high density accommodation close to universities and educational centres

Table 30: BNF Read codes for drug prescriptions in THIN database (THIN Data Guide (2011))

Conditions	**Drug Name**	**Sub-types**	**BNF Read codes in THIN**	**[1]NICE Guideline Code**
Hypercholesterolemia	Statins	Astorvastatin Rosuvastatin Simvastatin Fluvastatin Pravastatin	02120000	NICE CG181
Hypertension	Channel blockers	α-blocker	02040000	
		β-blocker	02020100	NICE GC127
		Calcium channel blocker	01070400	
	ACE-Inhibitor		02050500, 02050501, 02050502, 02060200	
Hormone Replacement Therapy	Estrogen		06040101, 07030100, 08030100, 07030200, 07030000, 06040200, 07020100, 020000000	NICE NG23
	Progesterone		06040102, 08030200, 07030202, 08010500, 07030500	NICE NG23
	Testosterone		06040200, 08030402, 03040100, 013060200, 07040100	NICE NG23

[1] *NICE stands for National Institute for Healthcare and Excellence*

B Data Description Tables

Table 31: Mean age (THR_{age}) at THR procedure and ratio of frequency of THR cases to controls, (M_R), across GP practices

Practice ID	THR Cases	Controls	M_R	Cases THR_{age}	Controls THR_{age}
a6706	58	290	5	72.88	72.71
a6732	38	190	5	70.74	70.62
a6744	63	315	5	72.38	72.58
a6775	100	500	5	72.29	72.50
a6802	53	265	5	73.05	72.87
a6810	123	615	5	70.26	70.22
a6811	163	815	5	69.92	69.81
a6873	5	25	5	67.48	68.03
a6875	50	250	5	72.50	72.42
a6883	37	185	5	71.09	70.95
a6897	123	615	5	71.10	70.98
a7714	38	190	5	72.54	72.53
a7778	66	330	5	71.95	71.69
a7832	99	495	5	70.85	70.77
a7840	95	475	5	71.31	71.51
a7916	93	465	5	71.17	71.19
a7939	48	240	5	73.69	73.76
a9000	90	450	5	70.98	70.88
a9799	3	15	5	73.05	73.12
a9800	22	110	5	68.11	68.11
a9849	20	100	5	71.19	71.09
a9854	102	510	5	69.09	69.10
a9856	13	65	5	72.82	72.70
a9870	78	390	5	70.88	70.68
a9873	90	450	5	70.71	70.36
a9874	26	130	5	70.44	70.19
a9884	86	430	5	70.84	70.48
a9887	59	295	5	70.32	70.39
a9889	39	195	5	73.90	74.17
a9898	32	160	5	71.95	71.61
a9899	46	230	5	71.62	71.98
a9910	95	475	5	70.54	70.41
a9914	112	560	5	71.08	70.99
a9915	36	180	5	69.40	69.61
a9916	2	10	5	68.12	67.62
a9918	48	240	5	69.06	69.13
a9919	76	380	5	70.92	70.55
a9923	100	500	5	73.01	73.02
a9925	37	185	5	72.27	72.05
a9928	45	225	5	71.00	70.75
a9937	66	330	5	70.96	70.90
a9939	54	270	5	70.46	70.76
a9945	48	240	5	69.81	69.78
a9957	22	110	5	69.78	69.62
a9968	53	265	5	71.82	71.48
a9975	14	70	5	70.62	70.95
a9977	40	200	5	71.54	71.08
a9984	7	35	5	72.22	71.90
a9990	5	25	5	59.96	59.68
a9991	81	405	5	70.59	70.27
a9992	33	165	5	70.04	69.58
b6702	12	60	5	71.19	71.62
b6732	24	120	5	72.66	72.61
b6797	48	240	5	74.21	73.84
b6834	2	10	5	73.15	74.05
b6837	31	155	5	73.15	72.72
b6862	11	55	5	77.15	76.97
b6868	4	20	5	70.74	70.89
b6876	7	35	5	76.42	76.05
b6878	20	100	5	69.81	69.90
b6882	97	485	5	72.86	72.82
b6883	38	190	5	72.14	72.43
b6884	65	325	5	72.29	72.26
b6887	44	220	5	68.23	68.01
b6888	27	135	5	69.40	69.21
b6892	43	215	5	71.35	71.08
b6894	22	110	5	70.07	69.94

b7691	91	455	5	72.41	72.42
b7858	114	570	5	70.78	70.64
b7915	3	15	5	76.60	76.54
b7938	46	230	5	71.36	71.61
b7942	5	25	5	76.61	77.37
b9000	63	315	5	68.94	68.86
b9774	21	105	5	73.30	73.58
b9814	25	125	5	70.32	70.91
b9818	53	265	5	70.80	69.98
b9846	14	70	5	74.14	73.21
b9849	8	40	5	69.72	70.27
b9856	70	350	5	70.41	70.29
b9857	62	310	5	71.85	71.52
b9867	82	410	5	71.34	71.29
b9870	33	165	5	70.11	69.39
b9872	40	200	5	72.39	72.41
b9878	58	290	5	69.58	69.66
b9880	6	30	5	73.29	72.62
b9884	80	400	5	73.30	73.43
b9885	50	250	5	70.59	70.37
b9889	29	145	5	72.51	72.35
b9892	123	615	5	70.55	70.35
b9898	32	160	5	70.69	70.71
b9908	67	335	5	72.97	72.68
b9909	28	140	5	71.73	71.50
b9916	23	115	5	70.30	70.52
b9925	97	485	5	70.89	70.72
b9927	25	125	5	71.15	71.09
b9928	10	50	5	70.18	70.36
b9931	11	55	5	74.45	75.23
b9960	8	40	5	72.49	72.04
b9964	62	310	5	71.15	71.22
b9989	26	130	5	70.18	70.60
b9990	51	255	5	69.81	69.65
b9997	68	340	5	71.45	71.04
b9998	36	180	5	69.96	69.99
c6698	4	20	5	71.34	72.19
c6705	7	35	5	67.39	67.02
c6709	18	90	5	69.78	69.70
c6724	11	55	5	73.61	72.69
c6746	6	30	5	73.00	72.39
c6816	16	80	5	72.24	72.03
c6830	64	320	5	70.34	70.27
c6850	12	60	5	75.56	76.06
c6854	6	30	5	65.00	64.94
c6871	23	115	5	71.97	71.89
c6876	2	10	5	76.12	76.19
c6878	18	90	5	74.22	74.54
c6880	16	80	5	69.41	69.78
c6889	18	90	5	74.38	74.69
c6899	75	375	5	72.58	72.24
c7699	2	10	5	70.83	70.05
c7781	18	90	5	71.76	71.54
c7805	3	15	5	74.10	74.77
c9774	11	55	5	71.81	71.45
c9782	23	115	5	69.47	69.96
c9785	54	270	5	71.37	71.43
c9800	29	145	5	70.21	70.42
c9855	73	365	5	70.28	69.93
c9857	68	340	5	72.00	71.92
c9885	24	120	5	72.36	72.01
c9894	72	360	5	67.78	67.57
c9896	12	60	5	66.85	66.94
c9897	43	215	5	68.81	68.49
c9898	169	845	5	71.78	71.72
c9905	32	160	5	68.74	68.41
c9906	86	430	5	71.86	72.01
c9910	118	590	5	70.09	70.00
c9913	52	260	5	71.63	71.53
c9953	85	425	5	69.39	69.29
c9972	26	130	5	71.19	71.08
c9975	110	550	5	71.67	71.66
c9981	12	60	5	69.33	68.59
c9991	17	85	5	73.94	73.59
c9993	17	85	5	71.44	71.03
c9996	73	365	5	70.31	70.54
c9998	60	300	5	69.99	69.84

d6710	27	135	5	71.64	72.07
d6727	44	220	5	72.49	72.49
d6753	20	100	5	73.76	73.45
d6764	23	115	5	69.48	69.65
d6804	2	10	5	78.01	77.21
d6808	10	50	5	69.32	69.60
d6838	35	175	5	72.73	73.04
d6857	7	35	5	63.79	63.91
d6869	4	20	5	61.65	60.85
d6878	4	20	5	65.17	65.47
d6879	3	15	5	70.99	71.72
d6883	28	140	5	72.50	72.01
d7700	3	15	5	73.79	74.13
d7793	44	220	5	71.17	71.33
d7813	9	45	5	70.27	69.59
d7907	10	50	5	71.05	71.53
d7941	4	20	5	73.03	73.39
d9779	6	30	5	75.18	74.71
d9783	28	140	5	71.70	71.67
d9795	9	45	5	73.93	72.95
d9832	17	85	5	71.47	70.81
d9834	28	140	5	68.00	67.25
d9864	66	330	5	71.11	70.96
d9866	7	35	5	70.79	71.25
d9905	101	505	5	71.27	71.03
d9908	23	115	5	70.84	70.62
d9914	64	320	5	71.32	70.76
d9918	70	350	5	68.76	68.93
d9928	49	245	5	72.11	72.04
d9935	10	50	5	73.36	74.34
d9937	33	165	5	72.22	72.32
d9953	68	340	5	70.26	70.09
d9957	36	180	5	71.71	71.70
d9961	54	270	5	70.04	70.03
d9965	152	760	5	69.86	69.75
d9970	14	70	5	70.76	70.77
d9971	164	820	5	69.47	69.41
d9974	140	700	5	70.90	70.64
d9978	7	35	5	67.73	66.76
d9986	32	160	5	71.25	71.44
d9997	15	75	5	70.78	71.15
e6765	47	235	5	73.94	73.90
e6776	5	25	5	67.47	67.55
e6799	24	120	5	72.52	72.53
e6802	112	560	5	70.76	70.71
e6806	69	345	5	70.92	70.69
e6817	95	475	5	72.09	72.09
e6857	33	165	5	70.09	70.60
e6866	10	50	5	74.19	73.25
e6867	8	40	5	70.33	69.63
e6877	48	240	5	70.85	70.76
e6881	2	10	5	74.11	73.60
e6888	5	25	5	69.42	70.30
e7686	22	110	5	71.55	71.32
e7692	40	200	5	71.43	71.50
e7712	11	55	5	70.23	70.58
e7734	6	30	5	66.73	67.29
e7805	9	45	5	71.59	70.99
e7916	21	105	5	74.58	74.37
e7923	2	10	5	72.51	72.41
e9003	6	30	5	64.41	64.01
e9776	34	170	5	70.15	70.07
e9820	26	130	5	70.09	70.06
e9825	49	245	5	72.09	71.93
e9846	58	290	5	70.26	69.98
e9857	10	50	5	75.86	75.64
e9881	25	125	5	67.17	67.15
e9888	96	480	5	72.71	72.71
e9908	38	190	5	69.39	68.92
e9909	28	140	5	72.42	72.00
e9912	33	165	5	67.10	67.17
e9914	31	155	5	71.27	71.13
e9924	4	20	5	64.34	63.48
e9931	48	240	5	70.87	70.49
e9935	27	135	5	71.60	71.38
e9957	21	105	5	69.91	69.67
e9972	26	130	5	70.64	70.66

e9973	108	540	5	71.07	70.81
e9981	17	85	5	70.76	70.39
e9986	6	30	5	63.44	64.24
e9990	19	95	5	68.49	67.74
e9994	65	325	5	71.61	71.67
f6704	2	10	5	73.11	72.24
f6718	11	55	5	70.42	71.07
f6749	16	80	5	72.82	72.53
f6753	44	220	5	72.59	71.92
f6772	59	295	5	70.58	70.49
f6797	80	400	5	71.61	71.39
f6803	47	235	5	71.37	71.39
f6808	100	500	5	72.96	72.90
f6809	11	55	5	71.65	70.90
f6811	9	45	5	67.97	67.61
f6818	59	295	5	72.08	72.01
f6860	5	25	5	75.75	75.35
f6871	5	25	5	72.73	72.69
f6890	35	175	5	71.37	71.01
f7687	6	30	5	72.31	72.58
f7713	77	385	5	70.66	70.70
f7812	17	85	5	71.31	71.95
f7815	7	35	5	70.80	71.15
f7849	68	340	5	73.00	72.86
f9100	5	25	5	68.19	67.91
f9772	27	135	5	70.68	70.42
f9776	5	25	5	74.42	74.90
f9790	4	20	5	67.73	67.95
f9793	13	65	5	72.25	72.51
f9794	16	80	5	69.82	69.60
f9814	18	90	5	71.62	71.62
f9823	7	35	5	71.02	70.30
f9837	27	135	5	69.65	69.49
f9846	33	165	5	71.02	71.24
f9866	35	175	5	74.78	74.87
f9878	142	710	5	71.24	71.19
f9882	72	360	5	69.62	69.30
f9885	15	75	5	70.14	70.46
f9892	59	295	5	71.95	71.95
f9897	3	15	5	68.24	69.11
f9913	22	110	5	75.66	75.73
f9919	73	365	5	72.29	72.24
f9921	34	170	5	70.11	69.53
f9923	81	405	5	72.30	72.21
f9931	3	15	5	68.72	68.59
f9932	74	370	5	71.86	71.92
f9934	81	405	5	71.70	71.64
f9936	51	255	5	70.80	70.46
f9948	20	100	5	66.52	66.22
f9954	23	115	5	69.36	69.92
f9955	48	240	5	72.75	72.25
f9956	60	300	5	72.52	72.33
f9959	40	200	5	71.34	71.05
f9961	50	250	5	70.36	70.31
f9962	66	330	5	72.68	72.61
f9964	95	475	5	72.07	71.96
f9967	25	125	5	73.12	73.24
f9971	33	165	5	68.80	68.26
f9983	3	15	5	59.09	59.89
f9991	73	365	5	69.35	68.95
f9994	85	425	5	72.07	71.81
f9995	135	675	5	69.38	69.08
g6714	24	120	5	69.63	69.80
g6736	30	150	5	70.14	69.69
g6857	8	40	5	77.34	78.22
g6868	2	10	5	75.23	74.93
g6878	2	10	5	73.18	74.01
g6884	80	400	5	71.99	72.08
g6895	3	15	5	68.28	68.95
g6896	4	20	5	78.11	78.11
g6897	7	35	5	70.57	70.05
g7688	3	15	5	69.79	70.39
g7779	31	155	5	71.96	71.95
g7788	46	230	5	73.43	73.11
g7807	7	35	5	71.54	71.11
g7856	63	315	5	71.73	71.68
g7858	1	5	5	77.13	77.13

g9781	34	170	5	69.55	69.88
g9807	10	50	5	68.84	68.16
g9835	20	100	5	71.92	72.21
g9840	8	40	5	70.14	69.88
g9846	12	60	5	71.77	71.52
g9851	42	210	5	72.00	71.42
g9853	13	65	5	68.45	68.56
g9859	29	145	5	69.41	69.28
g9861	12	60	5	71.54	70.59
g9884	67	335	5	71.97	71.97
g9885	19	95	5	70.01	70.35
g9895	27	135	5	67.75	67.68
g9910	14	70	5	69.96	69.71
g9911	78	390	5	70.71	70.85
g9912	29	145	5	70.12	69.47
g9917	107	535	5	69.22	68.98
g9920	138	690	5	71.56	71.52
g9936	13	65	5	68.65	68.46
g9938	64	320	5	69.14	68.97
g9956	24	120	5	69.72	69.83
g9959	41	205	5	70.54	70.45
g9965	74	370	5	69.99	70.01
g9987	42	210	5	69.35	68.70
g9993	4	20	5	73.43	72.55
h6720	36	180	5	72.32	72.45
h6744	6	30	5	69.28	69.88
h6748	17	85	5	73.21	72.92
h6755	18	90	5	65.60	65.43
h6800	5	25	5	64.59	65.11
h6812	13	65	5	68.84	69.16
h6813	5	25	5	69.01	69.73
h6871	19	95	5	72.46	72.28
h6872	28	140	5	72.51	72.74
h6873	8	40	5	72.06	72.06
h6879	12	60	5	72.66	72.86
h7691	67	335	5	70.78	70.23
h7694	168	840	5	73.26	73.15
h7706	1	5	5	73.86	74.76
h7711	121	605	5	74.34	74.21
h7753	12	60	5	76.66	76.66
h7822	16	80	5	72.81	73.10
h7954	7	35	5	75.29	75.23
h9004	31	155	5	70.80	70.64
h9781	12	60	5	66.25	66.03
h9784	48	240	5	72.33	71.99
h9799	29	145	5	70.83	70.98
h9819	29	145	5	71.24	71.04
h9832	14	70	5	75.49	75.40
h9845	23	115	5	71.79	71.64
h9860	42	210	5	71.05	71.01
h9864	12	60	5	71.34	72.29
h9882	59	295	5	68.90	68.66
h9888	77	385	5	70.74	70.62
h9899	69	345	5	71.02	70.62
h9900	85	425	5	71.67	71.57
h9901	24	120	5	68.93	68.79
h9921	12	60	5	74.91	73.98
h9936	2	10	5	71.34	71.53
h9938	100	500	5	71.50	71.45
h9946	21	105	5	67.07	66.71
h9950	10	50	5	72.12	72.40
h9957	45	225	5	69.62	69.59
h9959	30	150	5	72.22	72.14
h9973	11	55	5	73.32	72.52
h9978	83	415	5	72.38	72.32
h9981	12	60	5	71.73	70.98
h9986	4	20	5	74.18	73.43
h9987	36	180	5	65.16	64.90
h9988	37	185	5	69.72	69.13
h9990	56	280	5	70.63	70.44
h9995	81	405	5	73.48	73.20
h9998	53	265	5	71.91	71.54
i6714	26	130	5	70.33	70.21
i6727	12	60	5	70.64	70.01
i6731	2	10	5	62.43	62.93
i6732	12	60	5	69.31	70.14
i6744	9	45	5	71.09	71.61

i6749	21	105	5	73.52	73.07
i6750	21	105	5	71.28	71.29
i6783	40	200	5	72.69	72.77
i6805	13	65	5	69.76	69.87
i6859	37	185	5	71.27	70.82
i6861	34	170	5	74.62	74.68
i6880	12	60	5	70.74	70.52
i6894	3	15	5	75.81	75.87
i6895	8	40	5	72.71	72.51
i6897	7	35	5	65.26	65.63
i7689	20	100	5	71.73	71.18
i7698	54	270	5	72.46	72.78
i7816	27	135	5	73.37	73.76
i7876	30	150	5	75.01	75.17
i7927	13	65	5	73.85	72.88
i9777	21	105	5	72.81	72.21
i9820	8	40	5	68.59	68.02
i9823	88	440	5	70.12	69.97
i9829	15	75	5	74.93	75.45
i9840	8	40	5	73.24	73.28
i9859	22	110	5	72.84	72.43
i9863	17	85	5	72.28	72.12
i9869	18	90	5	69.11	68.65
i9881	16	80	5	72.79	72.36
i9898	9	45	5	70.06	69.33
i9907	55	275	5	71.63	71.35
i9912	6	30	5	71.27	70.37
i9916	35	175	5	71.94	71.69
i9922	6	30	5	69.87	68.93
i9934	32	160	5	69.42	69.67
i9949	112	560	5	72.04	71.92
i9952	10	50	5	70.19	69.93
i9955	60	300	5	69.36	68.73
i9966	14	70	5	69.47	69.02
i9983	10	50	5	70.72	70.66
i9984	30	150	5	69.34	69.47
i9986	4	20	5	72.03	72.63
i9990	105	525	5	71.16	71.37
i9993	41	205	5	71.51	71.60
i9994	55	275	5	70.00	69.81
j6725	11	55	5	73.60	74.29
j6737	24	120	5	72.42	72.10
j6739	17	85	5	72.43	73.23
j6744	7	35	5	79.07	80.01
j6811	6	30	5	71.16	70.33
j6865	1	5	5	69.61	68.69
j6871	32	160	5	74.29	73.52
j6887	1	5	5	71.99	71.05
j6890	32	160	5	70.08	69.67
j7696	4	20	5	72.96	72.96
j7710	100	500	5	72.76	72.45
j7712	3	15	5	78.77	78.37
j7733	8	40	5	68.85	68.80
j7785	7	35	5	74.34	74.70
j7877	49	245	5	71.04	71.59
j7915	3	15	5	77.05	77.32
j7922	17	85	5	72.94	73.19
j7924	8	40	5	75.33	75.63
j7927	27	135	5	73.08	73.12
j7934	16	80	5	66.79	67.10
j7951	2	10	5	66.56	66.46
j9100	2	10	5	68.10	67.60
j9855	45	225	5	73.51	73.23
j9864	16	80	5	69.65	70.17
j9870	19	95	5	73.86	73.87
j9881	47	235	5	70.18	69.96
j9884	19	95	5	71.47	71.67
j9885	10	50	5	66.29	65.63
j9904	24	120	5	75.54	75.41
j9916	35	175	5	70.15	70.01
j9919	9	45	5	73.18	74.03
j9926	22	110	5	72.93	72.70
j9930	7	35	5	75.71	76.36
j9945	43	215	5	73.60	73.88
j9953	4	20	5	74.32	73.37
j9956	131	655	5	71.96	71.89
j9961	5	25	5	70.90	71.34

j9963	2	10	5	72.18	72.98
j9969	5	25	5	69.15	69.07
j9973	213	1065	5	72.83	72.70
j9977	54	270	5	74.16	73.97
j9981	31	155	5	66.92	66.33
j9987	13	65	5	71.89	72.29
j9989	103	515	5	71.21	71.02
j9998	27	135	5	68.90	68.74

Each THR cases are matched to 5 controls of the same gender, year of birth category and GP practice. Mean age at THR of cases is within plus-minus one year to that of matched controls across each GP practices. Hence THR cases are either younger by a maximum of one year or older by a maximum of one year than matched controls.

Table 32: Distribution of cases and controls by age, gender and year of birth category

Gender	Age Group at THR	Number of Patients	%	Year of Birth Category							
				1920-24	%	1925-29	%	1930-34	%	1935-1940	%
Female Cases	18-54	281	2.6%	30	10.7%	50	17.8%	74	26.3%	127	45.2%
	55-64	1457	13.7%	198	13.6%	265	18.2%	365	25.1%	629	43.2%
	65-74	4538	42.6%	890	19.6%	1037	22.9%	1247	27.5%	1364	30.1%
	75-84	3847	36.1%	1658	43.1%	1407	36.6%	718	18.7%	64	1.7%
	85+	523	4.9%	464	88.7%	59	11.3%	0	0.0%	0	0.0%
	All ages	10646	100.0%	3240	30.4%	2818	26.5%	2404	22.6%	2184	20.5%
Male Cases	18-54	177	2.7%	15	8.5%	26	14.7%	40	22.6%	96	54.2%
	55-64	1141	17.5%	123	10.8%	183	16.0%	342	30.0%	493	43.2%
	65-74	2972	45.6%	412	13.9%	602	20.3%	921	31.0%	1037	34.9%
	75-84	1993	30.6%	727	36.5%	729	36.6%	487	24.4%	50	2.5%
	85+	228	3.5%	203	89.0%	25	11.0%	0	0.0%	0	0.0%
	All ages	6511	100.0%	1480	22.7%	1565	24.0%	1790	27.5%	1676	25.7%
Female Controls	18-54	1480	2.8%	183	12.4%	230	15.5%	415	28.0%	652	44.1%
	55-64	7594	14.3%	1056	13.9%	1405	18.5%	1850	24.4%	3283	43.2%
	65-74	22460	42.2%	4487	20.0%	5139	22.9%	6185	27.5%	6649	29.6%
	75-84	18978	35.7%	8072	42.5%	7000	36.9%	3570	18.8%	336	1.8%
	85+	2718	5.1%	2402	88.4%	316	11.6%	0	0.0%	0	0.0%
	All ages	53230	100.0%	16200	30.4%	14090	26.5%	12020	22.6%	10920	20.5%
Male Controls	18-54	1002	3.1%	91	9.1%	152	15.2%	210	21.0%	549	54.8%
	55-64	5859	18.0%	611	10.4%	912	15.6%	1762	30.1%	2574	43.9%
	65-74	14606	44.9%	2135	14.6%	2940	20.1%	4491	30.7%	5040	34.5%
	75-84	9851	30.3%	3481	35.3%	3666	37.2%	2487	25.2%	217	2.2%
	85+	1237	3.8%	1082	87.5%	155	12.5%	0	0.0%	0	0.0%
	All ages	32555	100.0%	7400	22.7%	7825	24.0%	8950	27.5%	8380	25.7%

Table 33: Distribution of cases by types of THR procedures and revision surgery

Gender	Age Group	Number of Patients		Types of THR procedure						Revision Surgery			
		Count	%	Cemented	%	Uncemented	%	Other types	%	Yes	%	No	%
Female Cases	**18-54**	281	2.6%	41	14.6%	168	59.8%	72	25.6%	20	7.1%	261	92.9%
	55-64	1457	13.7%	332	22.8%	776	53.3%	349	24.0%	56	3.8%	1401	96.2%
	65-74	4538	42.6%	1572	34.6%	2110	46.5%	856	18.9%	72	1.6%	4466	98.4%
	75-84	3847	36.1%	1767	45.9%	1482	38.5%	598	15.5%	24	0.6%	3823	99.4%
	85+	523	4.9%	285	54.5%	187	35.8%	51	9.8%	3	0.6%	520	99.4%
	All ages	10646	100.0%	3997	37.5%	4723	44.4%	1926	18.1%	175	1.6%	10471	98.4%
Male Cases	**18-54**	177	2.7%	43	24.3%	99	55.9%	35	19.8%	24	13.6%	153	86.4%
	55-64	1141	17.5%	285	25.0%	575	50.4%	281	24.6%	52	4.6%	1089	95.4%
	65-74	2972	45.6%	1048	35.3%	1367	46.0%	557	18.7%	38	1.3%	2934	98.7%
	75-84	1993	30.6%	871	43.7%	810	40.6%	312	15.7%	11	0.6%	1982	99.4%
	85+	228	3.5%	117	51.3%	82	36.0%	29	12.7%	1	0.4%	227	99.6%
	All ages	6511	100.0%	2364	36.3%	2933	45.0%	1214	18.6%	126	1.9%	6385	98.1%

Table 34: Number and percentage of patients by PVI variables, measured in quintiles level

Variable Description	Levels	THR Cases	%	Controls	%
White ethnicity	1	3151	18	15457	18
(Quintile)	2	3385	20	16398	19
	3	3091	18	15072	18
	4	2884	17	13342	16
	5	2392	14	11417	13
	Unknown	2254	13	14099	16
Mixed ethnicity	1	2223	13	10531	12
(Quintile)	2	3012	18	14275	17
	3	3215	19	15341	18
	4	3418	20	16556	19
	5	3035	18	14983	17
	Unknown	2254	13	14099	16
Asian ethnicity	1	2700	16	13044	15
(Quintile)	2	2726	16	13148	15
	3	3003	18	13965	16
	4	3325	19	16062	19
	5	3149	18	15467	18
	Unknown	2254	13	14099	16
Black ethnicity	1	1888	11	8649	10
(Quintile)	2	3448	20	16554	19
	3	2828	16	13769	16
	4	3809	22	17997	21
	5	2950	17	14717	17
	Unknown	2234	13	14099	16
Other ethnicity	1	1967	11	9086	11
(Quintile)	2	2786	16	13209	15
	3	3692	22	17548	20
	4	3571	21	17637	21
	5	2887	17	14206	17
	Unknown	2254	13	14099	16
Urban/Rural Classification	Urban	10867	63	52678	61
(Quintile)	Rural	2418	14	11830	14
	Village	1618	9	7178	8
	Unknown	2254	13	14099	16
Long term illnesses	1 (Lowest)	2858	17	13748	16
(Quintile)	2	3463	20	16155	19
	3	2879	17	13902	16
	4	2648	15	13219	15
	5 (Highest)	3055	18	14662	17
	Unknown	2254	13	14099	16
Mean level of Nitrogen Dioxide	1 (Lowest)	2381	14	11354	13
	2	2986	17	14272	17
	3	2844	17	13617	16
	4	3588	21	17520	20
	5 (Highest)	3104	18	14923	17
	Unknown	2254	13	14099	16
Mean level of particulate Matter	1 (Lowest)	1871	11	8989	10
	2	2880	17	13506	16
	3	3277	19	15844	18
	4	3566	21	17536	20
	5 (Highest)	3309	19	15811	18
	Unknown	2254	13	14099	16
Mean level of Sulphur dioxide	1 (Lowest)	2599	15	12320	14
	2	3019	18	14466	17
	3	3149	18	15301	18
	4	2518	15	12038	14
	5 (Highest)	3618	21	17561	20
	Unknown	2254	13	14099	16
Mean level of Nitrogen oxides	1 (Lowest)	2372	14	11325	13
	2	2967	17	14127	16
	3	2863	17	13741	16
	4	3598	21	17567	20
	5 (Highest)	3103	18	14926	17
	Unknown	2254	13	14099	16

Table 35: Distribution of patients by socio-economic status categories

Variable Description	Levels	Cases	%	Controls	%
Residential ward Townsend Score	**1 (Lowest)**	4727	28	21336	25
	2	3749	22	17704	20
	3	2919	17	14302	17
	4	2272	13	11698	14
	5 (Highest)	1213	7	6646	8
	Unknown	2277	13	14099	16
Index of Multiple Deprivation	**1 (Lowest)**	3972	23	19773	23
(IMD)	**2**	3675	22	18271	21
	3	3304	19	16760	20
	4	3050	18	15241	18
	5 (Highest)	2790	16	13910	16
	Unknown	366	2	1830	2
Mosaic Group	**Group A, B and C**	6039	35	29273	34
	Group D, E and F	2629	15	14128	17
	Group G, H and I	3063	18	16902	20
	Group J—O	3987	23	19109	21
	Unknown	1439	9	6373	8

Table 36: Distribution of cases and matched controls by smoking status

Gender	Age Group	Number of Patients	%	Smoking Status							
				Ex-smoker	%	Non-smoker	%	Smoker	%	Unknown	%
Female Cases	18-54	281	2.6%	4	1.4%	18	6.4%	9	3.2%	250	89.0%
	55-64	1457	13.7%	66	4.5%	345	23.7%	146	10.0%	900	61.8%
	65-74	4538	42.6%	380	8.4%	2155	47.5%	603	13.3%	1400	30.9%
	75-84	3847	36.1%	462	12.0%	2250	58.5%	491	12.8%	644	16.7%
	85+	523	4.9%	84	16.1%	377	72.1%	39	7.5%	23	4.4%
	All ages	10646	100.0%	996	9.4%	5145	48.3%	1288	12.1%	3217	30.2%
Male Cases	18-54	177	2.7%	6	3.4%	5	2.8%	13	7.3%	153	86.4%
	55-64	1141	17.5%	75	6.6%	244	21.4%	134	11.7%	688	60.3%
	65-74	2972	45.6%	440	14.8%	1276	42.9%	457	15.4%	799	26.9%
	75-84	1993	30.6%	368	18.5%	1070	53.7%	279	14.0%	276	13.8%
	85+	228	3.5%	60	26.3%	136	59.6%	24	10.5%	8	3.5%
	All ages	6511	100.0%	949	14.6%	2731	41.9%	907	13.9%	1924	29.5%
Female Controls	18-54	1480	2.8%	12	0.8%	1118	75.5%	15	1.0%	335	22.6%
	55-64	7594	14.3%	88	1.2%	5180	68.2%	208	2.7%	2118	27.9%
	65-74	22460	42.2%	838	3.7%	14517	64.6%	1319	5.9%	5786	25.8%
	75-84	18978	35.7%	1422	7.5%	11035	58.1%	1899	10.0%	4622	24.4%
	85+	2718	5.1%	440	16.2%	1388	51.1%	445	16.4%	445	16.4%
	All ages	53230	100.0%	2800	5.3%	33238	62.4%	3886	7.3%	13306	25.0%
Male Controls	18-54	1002	3.1%	6	0.6%	701	70.0%	21	2.1%	274	27.3%
	55-64	5859	18.0%	136	2.3%	3762	64.2%	215	3.7%	1746	29.8%
	65-74	14606	44.9%	1013	6.9%	8647	59.2%	1185	8.1%	3761	25.7%
	75-84	9851	30.3%	1435	14.6%	4997	50.7%	1212	12.3%	2207	22.4%
	85+	1237	3.8%	358	28.9%	503	40.7%	205	16.6%	171	13.8%
	All ages	32555	100.0%	2948	9.1%	18610	57.2%	2838	8.7%	8159	25.1%

Table 37: Distribution of cases and matched controls by BMI categories

Gender	Age Group	Number of Patients	%	Body Mass Index Categories							
				Normal	%	Overweight	%	Obese	%	Unknown	%
Female Cases	18-54	281	2.6%	61	21.7%	86	30.6%	102	36.3%	32	11.4%
	55-64	1457	13.7%	307	21.1%	429	29.4%	531	36.4%	190	13.0%
	65-74	4538	42.6%	1000	22.0%	1345	29.6%	1748	38.5%	445	9.8%
	75-84	3847	36.1%	828	21.5%	1141	29.7%	1472	38.3%	406	10.6%
	85+	523	4.9%	132	25.2%	135	25.8%	216	41.3%	40	7.6%
	All ages	10646	100.0%	2328	21.9%	3136	29.5%	4069	38.2%	1113	10.5%
Male Cases	18-54	177	2.7%	79	44.6%	53	29.9%	26	14.7%	19	10.7%
	55-64	1141	17.5%	458	40.1%	370	32.4%	187	16.4%	126	11.0%
	65-74	2972	45.6%	1241	41.8%	1006	33.8%	478	16.1%	247	8.3%
	75-84	1993	30.6%	808	40.5%	654	32.8%	346	17.4%	185	9.3%
	85+	228	3.5%	82	36.0%	80	35.1%	48	21.1%	18	7.9%
	All ages	6511	100.0%	2668	41.0%	2163	33.2%	1085	16.7%	595	9.1%
Female Controls	18-54	1480	2.8%	648	43.8%	439	29.7%	301	20.3%	92	6.2%
	55-64	7594	14.3%	3345	44.0%	2162	28.5%	1380	18.2%	707	9.3%
	65-74	22460	42.2%	10259	45.7%	6306	28.1%	3830	17.1%	2065	9.2%
	75-84	18978	35.7%	9460	49.8%	5048	26.6%	2549	13.4%	1921	10.1%
	85+	2718	5.1%	1471	54.1%	682	25.1%	291	10.7%	274	10.1%
	All ages	53230	100.0%	25183	47.3%	14637	27.5%	8351	15.7%	5059	9.5%
Male Controls	18-54	1002	3.1%	398	39.7%	377	37.6%	155	15.5%	72	7.2%
	55-64	5859	18.0%	2372	40.5%	2047	34.9%	879	15.0%	561	9.6%
	65-74	14606	44.9%	5997	41.1%	5191	35.5%	2025	13.9%	1393	9.5%
	75-84	9851	30.3%	4552	46.2%	3223	32.7%	1032	10.5%	1044	10.6%
	85+	1237	3.8%	619	50.0%	379	30.6%	108	8.7%	131	10.6%
	All ages	32555	100.0%	13938	42.8%	11217	34.5%	4199	12.9%	3201	9.8%

Table 38: Type 2 diabetes and chronic kidney disease in THR cases and controls

Gender	Age Group	Number of Patients	%	Type 2 Diabetes				Chronic Kidney Disease (CKD)					
				Yes	%	No	%	No CKD	%	Stage 1-2	%	Stage 3-5	%
Female Cases	**18-54**	281	2.6%	4	1.4%	277	98.6%	281	100.0%	0	0.0%	0	0.0%
	55-64	1457	13.7%	29	2.0%	1428	98.0%	1457	100.0%	0	0.0%	0	0.0%
	65-74	4538	42.6%	274	6.0%	4264	94.0%	4424	97.5%	11	0.2%	103	2.3%
	75-84	3847	36.1%	311	8.1%	3536	91.9%	3541	92.0%	30	0.8%	276	7.2%
	85+	523	4.9%	53	10.1%	470	89.9%	381	72.8%	9	1.7%	133	25.4%
	All ages	10646	100.0%	671	6.3%	9975	93.7%	10084	94.7%	50	0.5%	512	4.8%
Male Cases	**18-54**	177	2.7%	4	2.3%	173	97.7%	177	100.0%	0	0.0%	0	0.0%
	55-64	1141	17.5%	40	3.5%	1101	96.5%	1141	100.0%	0	0.0%	0	0.0%
	65-74	2972	45.6%	270	9.1%	2702	90.9%	2904	97.7%	9	0.3%	59	2.0%
	75-84	1993	30.6%	201	10.1%	1792	89.9%	1845	92.6%	14	0.7%	134	6.7%
	85+	228	3.5%	26	11.4%	202	88.6%	176	77.2%	6	2.6%	46	20.2%
	All ages	6511	100.0%	541	8.3%	5970	91.7%	6243	95.9%	29	0.4%	239	3.7%
Female Controls	**18-54**	1480	2.8%	13	0.9%	1467	99.1%	1480	100.0%	0	0.0%	0	0.0%
	55-64	7594	14.3%	198	2.6%	7396	97.4%	7592	100.0%	0	0.0%	2	0.0%
	65-74	22460	42.2%	1380	6.1%	21080	93.9%	22114	98.5%	40	0.2%	306	1.4%
	75-84	18978	35.7%	1710	9.0%	17268	91.0%	18068	95.2%	63	0.3%	847	4.5%
	85+	2718	5.1%	293	10.8%	2425	89.2%	2378	87.5%	10	0.4%	330	12.1%
	All ages	53230	100.0%	3594	6.8%	49636	93.2%	51632	97.0%	113	0.2%	1485	2.8%
Male Controls	**18-54**	1002	3.1%	17	1.7%	985	98.3%	1002	100.0%	0	0.0%	0	0.0%
	55-64	5859	18.0%	215	3.7%	5644	96.3%	5859	100.0%	0	0.0%	0	0.0%
	65-74	14606	44.9%	1257	8.6%	13349	91.4%	14397	98.6%	23	0.2%	186	1.3%
	75-84	9851	30.3%	1208	12.3%	8643	87.7%	9379	95.2%	48	0.5%	424	4.3%
	85+	1237	3.8%	165	13.3%	1072	86.7%	1110	89.7%	8	0.6%	119	9.6%
	All ages	32555	100.0%	2862	8.8%	29693	91.2%	31747	97.5%	79	0.2%	729	2.2%

Table 39: Angina, myocardial infarction and stroke among cases and matched controls

Gender	Age Group	Number of Patients	%	Angina				Myocardial Infarction				Stroke			
				Yes	%	No	%	Yes	%	No	%	Yes	%	No	%
Female Cases	**18-54**	281	2.6%	33	11.7%	248	88.3%	11	3.9%	270	96.1%	13	4.6%	268	95.4%
	55-64	1457	13.7%	143	9.8%	1314	90.2%	87	6.0%	1370	94.0%	56	3.8%	1401	96.2%
	65-74	4538	42.6%	496	10.9%	4042	89.1%	275	6.1%	4263	93.9%	150	3.3%	4388	96.7%
	75-84	3847	36.1%	425	11.0%	3422	89.0%	240	6.2%	3607	93.8%	137	3.6%	3710	96.4%
	85+	523	4.9%	78	14.9%	445	85.1%	50	9.6%	473	90.4%	18	3.4%	505	96.6%
	All ages	10646	100.0%	1175	11.0%	9471	89.0%	663	6.2%	9983	93.8%	374	3.5%	10272	96.5%
Male Cases	**18-54**	177	2.7%	29	16.4%	148	83.6%	27	15.3%	150	84.7%	11	6.2%	166	93.8%
	55-64	1141	17.5%	172	15.1%	969	84.9%	135	11.8%	1006	88.2%	52	4.6%	1089	95.4%
	65-74	2972	45.6%	473	15.9%	2499	84.1%	369	12.4%	2603	87.6%	127	4.3%	2845	95.7%
	75-84	1993	30.6%	329	16.5%	1664	83.5%	239	12.0%	1754	88.0%	87	4.4%	1906	95.6%
	85+	228	3.5%	33	14.5%	195	85.5%	22	9.6%	206	90.4%	11	4.8%	217	95.2%
	All ages	6511	100.0%	1036	15.9%	5475	84.1%	792	12.2%	5719	87.8%	288	4.4%	6223	95.6%
Female Controls	**18-54**	1480	2.8%	134	9.1%	1346	90.9%	60	4.1%	1420	95.9%	100	6.8%	1380	93.2%
	55-64	7594	14.3%	683	9.0%	6911	91.0%	372	4.9%	7222	95.1%	545	7.2%	7049	92.8%
	65-74	22460	42.2%	2298	10.2%	20162	89.8%	1208	5.4%	21252	94.6%	1796	8.0%	20664	92.0%
	75-84	18978	35.7%	2192	11.6%	16786	88.4%	1264	6.7%	17714	93.3%	1907	10.0%	17071	90.0%
	85+	2718	5.1%	329	12.1%	2389	87.9%	213	7.8%	2505	92.2%	320	11.8%	2398	88.2%
	All ages	53230	100.0%	5636	10.6%	47594	89.4%	3117	5.9%	50113	94.1%	4668	8.8%	48562	91.2%
Male Controls	**18-54**	1002	3.1%	139	13.9%	863	86.1%	104	10.4%	898	89.6%	84	8.4%	918	91.6%
	55-64	5859	18.0%	800	13.7%	5059	86.3%	638	10.9%	5221	89.1%	455	7.8%	5404	92.2%
	65-74	14606	44.9%	2215	15.2%	12391	84.8%	1713	11.7%	12893	88.3%	1369	9.4%	13237	90.6%
	75-84	9851	30.3%	1666	16.9%	8185	83.1%	1325	13.5%	8526	86.5%	1143	11.6%	8708	88.4%
	85+	1237	3.8%	211	17.1%	1026	82.9%	164	13.3%	1073	86.7%	124	10.0%	1113	90.0%
	All ages	32555	100.0%	5031	15.5%	27524	84.5%	3944	12.1%	28611	87.9%	3175	9.8%	29380	90.2%

Table 40: Prevalence of osteoarthritis, rheumatoid arthritis and osteoporosis among cases and matched controls

Gender	Age Group	Number of Patients	%	Osteoarthritis				Rheumatoid Arthritis				Osteoporosis			
				Yes	%	No	%	Yes	%	No	%	Yes	%	No	%
Female Cases	**18-54**	281	2.6%	220	78.2%	61	21.8%	21	7.5%	260	92.5%	0	0.0%	281	100.0%
	55-64	1457	13.7%	1259	86.4%	198	13.6%	98	6.7%	1359	93.3%	21	1.4%	1436	98.6%
	65-74	4538	42.6%	3995	88.0%	543	12.0%	215	4.7%	4323	95.3%	263	5.8%	4275	94.2%
	75-84	3847	36.1%	3360	87.3%	487	12.7%	162	4.2%	3685	95.8%	425	11.0%	3422	89.0%
	85+	523	4.9%	444	84.9%	79	15.1%	18	3.4%	505	96.6%	65	12.4%	458	87.6%
	All ages	10646	100.0%	9278	87.2%	1368	12.8%	514	4.8%	10132	95.2%	774	7.3%	9872	92.7%
Male Cases	**18-54**	177	2.7%	144	81.3%	33	18.7%	5	2.8%	172	97.2%	1	0.6%	176	99.4%
	55-64	1141	17.5%	980	85.9%	161	14.1%	33	2.9%	1108	97.1%	6	0.5%	1135	99.5%
	65-74	2972	45.6%	2626	88.4%	346	11.6%	78	2.6%	2894	97.4%	40	1.3%	2932	98.7%
	75-84	1993	30.6%	1733	87.0%	260	13.0%	45	2.3%	1948	97.7%	56	2.8%	1937	97.2%
	85+	228	3.5%	183	80.3%	45	19.7%	8	3.5%	220	96.5%	13	5.7%	215	94.3%
	All ages	6511	100.0%	5666	87.0%	845	13.0%	169	2.6%	6342	97.4%	116	1.8%	6395	98.2%
Female Controls	**18-54**	1480	2.8%	270	18.3%	1210	81.7%	47	3.2%	1433	96.8%	13	0.9%	1466.68	99.1%
	55-64	7594	14.3%	2223	29.3%	5371	70.7%	204	2.7%	7390	97.3%	244	3.2%	7349.8	96.8%
	65-74	22460	42.2%	6226	27.7%	16234	72.3%	651	2.9%	21809	97.1%	1244	5.5%	21216	94.5%
	75-84	18978	35.7%	5078	26.8%	13900	73.2%	519	2.7%	18459	97.3%	1666	8.8%	17311.6	91.2%
	85+	2718	5.1%	707	26.0%	2011	74.0%	78	2.9%	2640	97.1%	207	7.6%	2510.6	92.4%
	All ages	53230	100.0%	14504	27.2%	38726	72.8%	1499	2.8%	51731	97.2%	3375	6.3%	49855	93.7%
Male Controls	**18-54**	1002	3.1%	70	7.0%	932	93.0%	17	1.7%	985	98.3%	3	0.3%	998.994	99.7%
	55-64	5859	18.0%	469	8.0%	5390	92.0%	85	1.5%	5774	98.5%	161	2.8%	5697.8	97.2%
	65-74	14606	44.9%	1753	12.0%	12853	88.0%	213	1.5%	14393	98.5%	498	3.4%	14108.2	96.6%
	75-84	9851	30.3%	1084	11.0%	8767	89.0%	164	1.7%	9687	98.3%	218	2.2%	9633.2	97.8%
	85+	1237	3.8%	124	10.0%	1113	90.0%	15	1.2%	1222	98.8%	50	4.0%	1187.4	96.0%
	All ages	32555	100.0%	3499	10.7%	29056	89.3%	494	1.5%	32061	98.5%	929	2.9%	31626	97.1%

Table 41: Hypercholesterolemia in cases and matched controls

Gender	Age Group	Number of Patients	%	Category 1	%	Category 2	%	Category 3	%	Category 4	%
Female Cases	**18-54**	281	2.64%	254	90.39%	3	1.07%	0	0.00%	24	8.54%
	55-64	1457	13.69%	1189	81.61%	90	6.18%	56	3.84%	122	8.37%
	65-74	4538	42.63%	2720	59.94%	918	20.23%	565	12.45%	335	7.38%
	75-84	3847	36.14%	1590	41.33%	1301	33.82%	738	19.18%	218	5.67%
	85+	523	4.91%	127	24.28%	260	49.71%	114	21.80%	22	4.21%
	All ages	10646	100.00%	5880	55.23%	2572	24.16%	1473	13.84%	721	6.77%
Male Cases	**18-54**	177	2.72%	156	88.14%	4	2.26%	0	0.00%	17	9.60%
	55-64	1141	17.52%	907	79.49%	85	7.45%	45	3.94%	104	9.11%
	65-74	2972	45.65%	1588	53.43%	785	26.41%	376	12.65%	223	7.50%
	75-84	1993	30.61%	812	40.74%	770	38.64%	311	15.60%	100	5.02%
	85+	228	3.50%	60	26.32%	118	51.75%	41	17.98%	9	3.95%
	All ages	6511	100.00%	3523	54.11%	1762	27.06%	773	11.87%	453	6.96%
Female Controls	**18-54**	1480	2.78%	727	49.12%	0	0.00%	2	0.14%	751	50.74%
	55-64	7594	14.27%	3714	48.91%	27	0.36%	146	1.92%	3707	48.81%
	65-74	22460	42.19%	10660	47.46%	570	2.54%	2355	10.49%	8875	39.51%
	75-84	18978	35.65%	9606	50.62%	833	4.39%	2983	15.72%	5556	29.28%
	85+	2718	5.11%	1534	56.44%	163	6.00%	492	18.10%	529	19.46%
	All ages	53230	100.00%	26241	49.30%	1593	2.99%	5978	11.23%	19418	36.48%
Male Controls	**18-54**	1002	3.08%	571	56.99%	0	0.00%	1	0.10%	430	42.91%
	55-64	5859	18.00%	3367	57.47%	50	0.85%	150	2.56%	2292	39.12%
	65-74	14606	44.87%	7761	53.14%	756	5.18%	1833	12.55%	4256	29.14%
	75-84	9851	30.26%	5516	55.99%	751	7.62%	1593	16.17%	1991	20.21%
	85+	1237	3.80%	782	63.22%	113	9.14%	183	14.79%	159	12.85%
	All ages	32555	100.00%	17997	55.28%	1670	5.13%	3760	11.55%	9128	28.04%

Category 1 refers to patients with no records of hypercholesterolemia (HC) and statins prescription prior to THR surgery.

Category 2 represents patients with normal cholesterol level due to statins prescription prior to THR surgery.

Category 3 are patients with HC despite the statins prescription prior to THR surgery.

Category 4 are patients with HC but not on statins prescription prior to THR surgery

Table 42: Hypertension in cases and matched controls

Gender	Age Group	Number of Patients	%	Category 1	%	Category 2	%	Category 3	%	Category 4	%
Female Cases	**18-54**	281	2.64%	251	89.3%	8	2.8%	4	1.4%	18	6.4%
	55-64	1457	13.69%	952	65.3%	210	14.4%	161	11.1%	134	9.2%
	65-74	4538	42.63%	1887	41.6%	808	17.8%	1364	30.1%	479	10.6%
	75-84	3847	36.14%	1071	27.8%	602	15.6%	1779	46.2%	395	10.3%
	85+	523	4.91%	99	18.9%	60	11.5%	306	58.5%	58	11.1%
	All ages	10646	100.00%	4260	40.0%	1688	15.9%	3614	33.9%	1084	10.2%
Male Cases	**18-54**	177	2.72%	165	93.2%	1	0.6%	5	2.8%	6	3.4%
	55-64	1141	17.52%	843	73.9%	73	6.4%	102	8.9%	123	10.8%
	65-74	2972	45.65%	1312	44.1%	336	11.3%	875	29.4%	449	15.1%
	75-84	1993	30.61%	611	30.7%	308	15.5%	784	39.3%	290	14.6%
	85+	228	3.50%	39	17.1%	36	15.8%	127	55.7%	26	11.4%
	All ages	6511	100.00%	2970	45.6%	754	11.6%	1893	29.1%	894	13.7%
Female Controls	**18-54**	1480	2.78%	791	53.4%	10	0.7%	11	0.7%	668	45.1%
	55-64	7594	14.27%	3664	48.2%	343	4.5%	778	10.2%	2809	37.0%
	65-74	22460	42.19%	9107	40.5%	2377	10.6%	5910	26.3%	5066	22.6%
	75-84	18978	35.65%	6142	32.4%	3076	16.2%	7418	39.1%	2342	12.3%
	85+	2718	5.11%	791	29.1%	564	20.8%	1178	43.3%	185	6.8%
	All ages	53230	100.00%	20495	38.5%	6370	12.0%	15295	28.7%	11070	20.8%
Male Controls	**18-54**	1002	3.08%	562	56.1%	9	0.9%	13	1.3%	418	41.7%
	55-64	5859	18.00%	3099	52.9%	257	4.4%	528	9.0%	1975	33.7%
	65-74	14606	44.87%	6431	44.0%	1759	12.0%	3613	24.7%	2803	19.2%
	75-84	9851	30.26%	3625	36.8%	1852	18.8%	3354	34.0%	1020	10.4%
	85+	1237	3.80%	391	31.6%	299	24.2%	461	37.3%	86	7.0%
	All ages	32555	100.00%	14108	43.3%	4176	12.8%	7969	24.5%	6302	19.4%

Category 1 refers to patients with no records of HT in MED file or with normal systolic or diastolic BP based on measurements recorded prior to relevant THR surgery and not receiving any anti-hypertensive drug prescription prior to relevant THR surgery.

Category 2 represents patients with normal BP level due and on anti-hypertensive drugs prescription prior to THR surgery.

Category 3 are patients with HT despite on anti-hypertensive drugs prescription prior to THR surgery.

Category 4 represent patients with HT and without anti-hypertensive drugs prescription prior to THR surgery.

Table 43: Drug prescriptions prior to THR surgery by gender and case-control status

Drug Name	*Prescription*[1]	Female Cases	%	Male Cases	%	Female Controls	%	Male Controls	%
Number of patients		10646	100.0%	6511	100.0%	53230	100.0%	32555	100.0%
Statins[2]	No	6601	62.0%	3976	61.1%	45659	85.8%	27125	83.3%
	Yes	4045	38.0%	2535	38.9%	7571	14.2%	5430	16.7%
Alpha/Beta Blocker	No	7830	73.5%	4791	73.6%	34973	65.7%	23222	71.3%
(α-/β-Blocker)	Yes	2816	26.5%	1720	26.4%	18257	34.3%	9333	28.7%
Calcium Channel Blocker	No	7489	70.3%	5023	77.1%	38777	72.8%	23311	71.6%
	Yes	3157	29.7%	1488	22.9%	14453	27.2%	9244	28.4%
Angiotensin converting enzyme inhibitor	No	7131	67.0%	4581	70.4%	34849	65.5%	22165	68.1%
(ACE-Inhibitor)	Yes	3515	33.0%	1930	29.6%	18381	34.5%	10390	31.9%
Estrogen	No	9841	92.4%	6511	100.0%	48551	91.2%	32555	100.0%
	Yes	805	7.6%	0	0.0%	4679	8.8%	0	0.0%
Progesterone	No	10470	98.3%	6511	100.0%	52731	99.1%	32555	100.0%
	Yes	176	1.7%	0	0.0%	499	0.9%	0	0.0%
Testosterone	No	10646	100.0%	6239	95.8%	53230	100.0%	31045	95.4%
	Yes	0	0.0%	272	4.2%	0	0.0%	1510	4.6%

[1] *All prescriptions are dated before relevant THR surgery time for cases and matched controls.*

[2] *Statins prescription is not used as a variable in data analysis as it is used to define categories of hypercholesterolemia (See Table 41 in Appendix B)*

C Short Term Survival Model

Table 44: List of variables for model selection

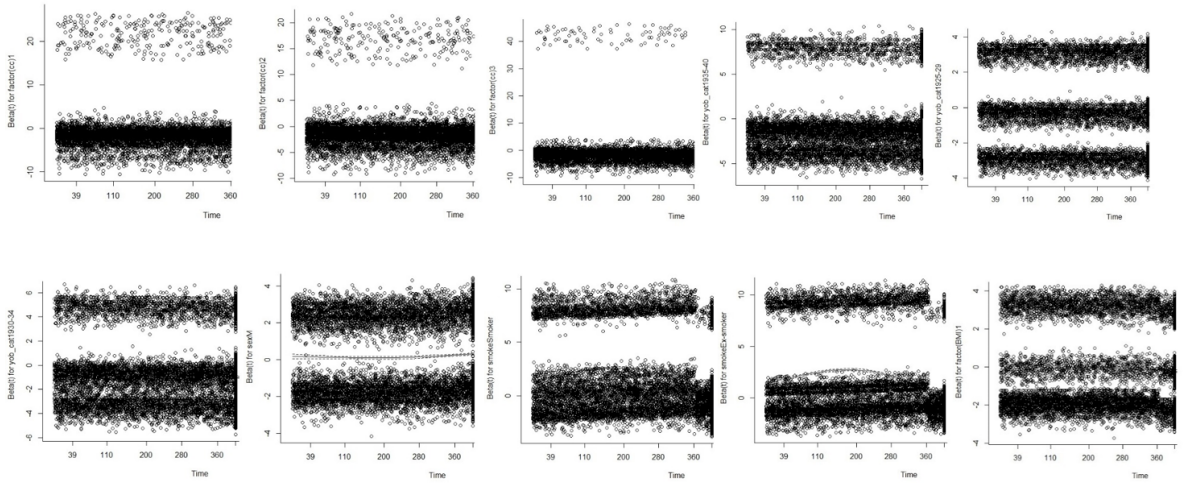
Category	Code	Variable Description	Levels
Demographic	cc	Procedure types	0=Controls, 1=Cemented 2=Uncemented, 3=Other types
	pracid	Identity of GP practice	Frailty random variable
	yob_cat	Year of birth category	1920-24, 1925-29, 1930-34,1935-40
	sex	Gender	0=Female, 1=Male
	deathstatus	Death indicator	0=Alive, 1=Dead
	smoke	Smoking status	0=Non-smoker, 1=Ex-smoker, 2=Smoker
	xfer	Transfer status	0=Not transferred, 1=Transferred
	surage	Age at surgery	Continuous variable
	survtime	Time (years) survived after THR	Continuous variable
Post-code Related Variables	town	Townsend score	Level 1 (Least deprived), 2, 3, 4 and 5 (Most deprived)
	IMD	Index of Multiple Deprivation	Level 1 (Least deprived), 2, 3, 4 and 5 (Most deprived)
	mosaic	Mosaic Category	0=Category A-C, 1=Category D-F, 2=Category G-I, 3=Category J-O
Medical	BMI	Body Mass Index	0=Normal, 1=Overweight, 2=Obese
	diab	Type 2 Diabetes	0=No, 1=Yes
	angina	Preoperative event of angina	0=No, 1=Yes
	stroke	Preoperative event of stroke	0=No, 1=Yes
	ha	Preoperative event of myocardial infarction	0=No, 1=Yes
	oa	Preoperative diagnosis of osteoarthrtitis	0=No, 1=Yes
	ra	Preoperative diagnosis of rheumatoid arthrtitis	0=No, 1=Yes

Table 44 continued from previous page

Category	Code	Variable Description	Levels
	ckd	Preoperative chronic kidney disease	0=No CKD, 1=CKD stage 1-2, 2=CKD stage 3-5
	HT	Hypertension	0=No HT and no drug prescription, 1=No HT with drug prescription, 2=HT with drug prescription, 3=HT without drug prescription
	HC	Hypercholesterolaemia	0=No HC and no drug prescription, 1=No HC with drug prescription, 2=HC with drug prescription, 3=HC without drug prescription
Drug Prescription	ace	ACE-inhibitor prescription	0=No, 1=Yes
	calcium	Calcium-blocker prescription	0=No, 1=Yes
	beta	Alpha/Beta-blocker prescription	0=No, 1=Yes
	estro	Estrogen for HRT	0=No, 1=Yes
	proges	Progesteron for HRT	0=No, 1=Yes
	testos	Testosterone prescription	0=No, 1=Yes

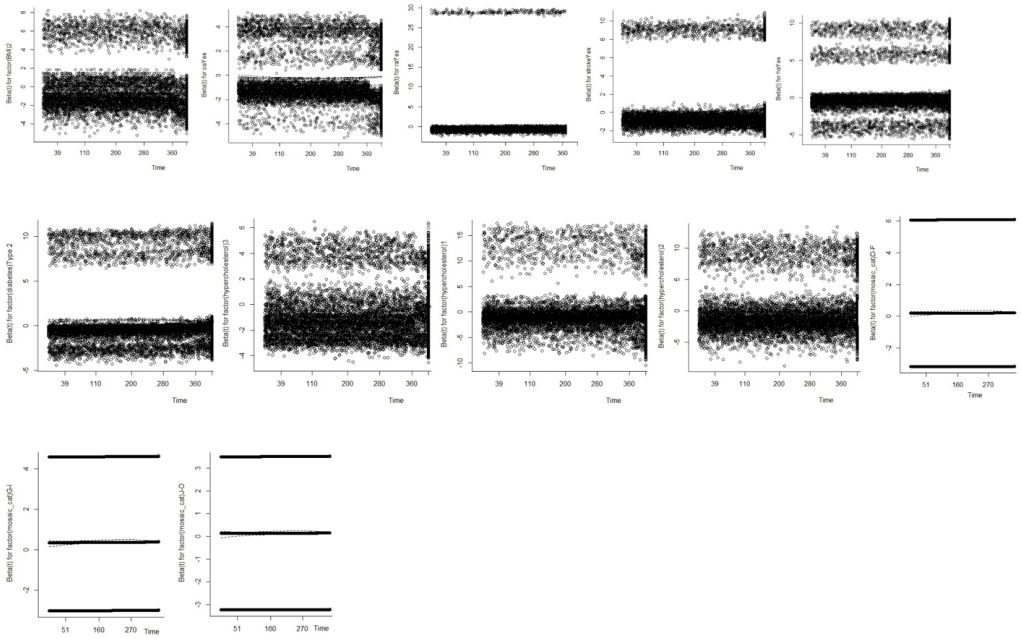
The variables listed here are used for model selection for short term logistic regression model and for the long term Cox proportional hazards model, both with shared frailty.

Figure 32: Schoenfeld residuals distribution for one-year survival model - Part 1



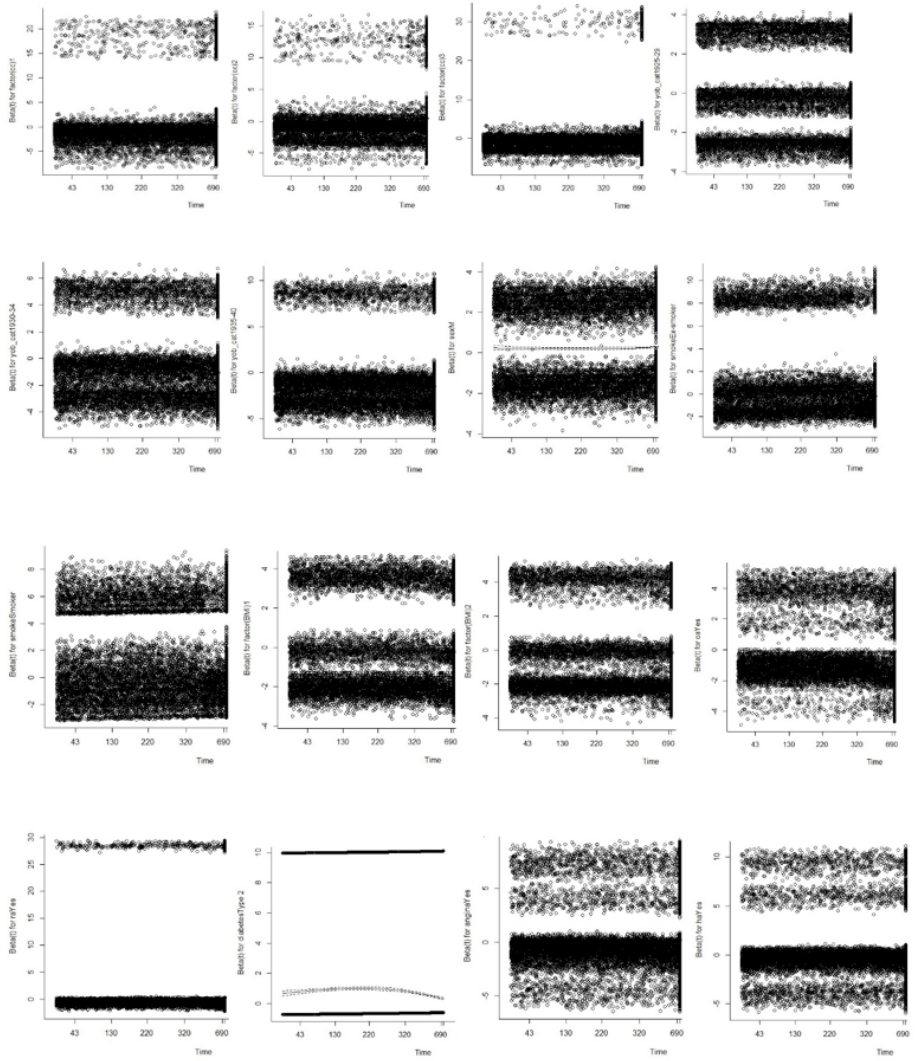
Plots of Schoenfeld residuals versus event time for the two-year survival models show that the assumption of proportional hazards is violated. The effects of the variables kept in the fitted model, on hazard rate, are time dependent.

Figure 33: Schoenfeld residuals distribution for one-year survival model - Part 1



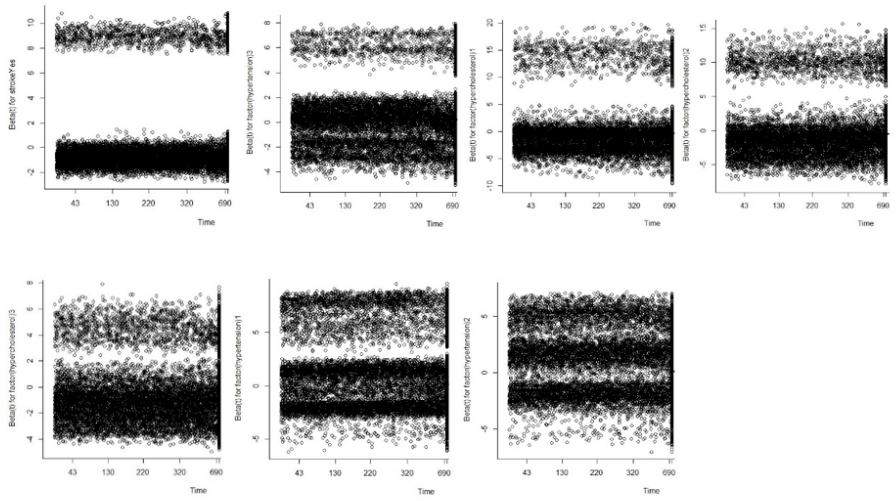
Plots of Schoenfeld residuals versus event time for the two-year survival models show that the assumption of proportional hazards is violated. The effects of the variables kept in the fitted model, on hazard rate, are time dependent.

Figure 34: Schoenfeld residuals distribution for two-year survival model - Part 2



Plots of Schoenfeld residuals versus event time for the two-year survival models show that the assumption of proportional hazards is violated. The effects of the variables kept in the fitted model, on hazard rate, are time dependent.

Figure 35: Schoenfeld residuals distribution for two-year survival model - Part 2



Plots of Schoenfeld residuals versus event time for the two-year survival models show that the assumption of proportional hazards is violated. The effects of the variables kept in the fitted model, on hazard rate, are time dependent.

Figure 36: Pearson residuals plots for covariates selected in short term logistic regression model with interaction terms fitted using 10 imputed datasets

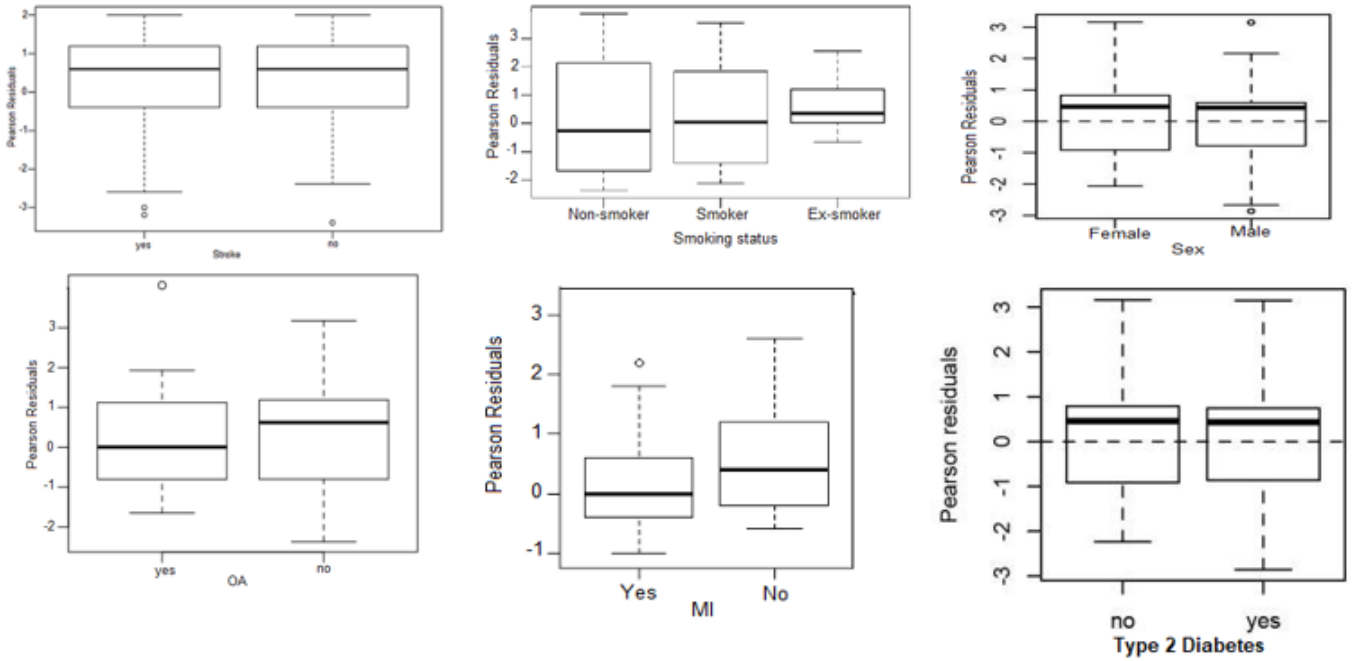


Figure 37: Pearson residuals plots for covariates selected in short term logistic regression model with interaction terms fitted using 10 imputed datasets

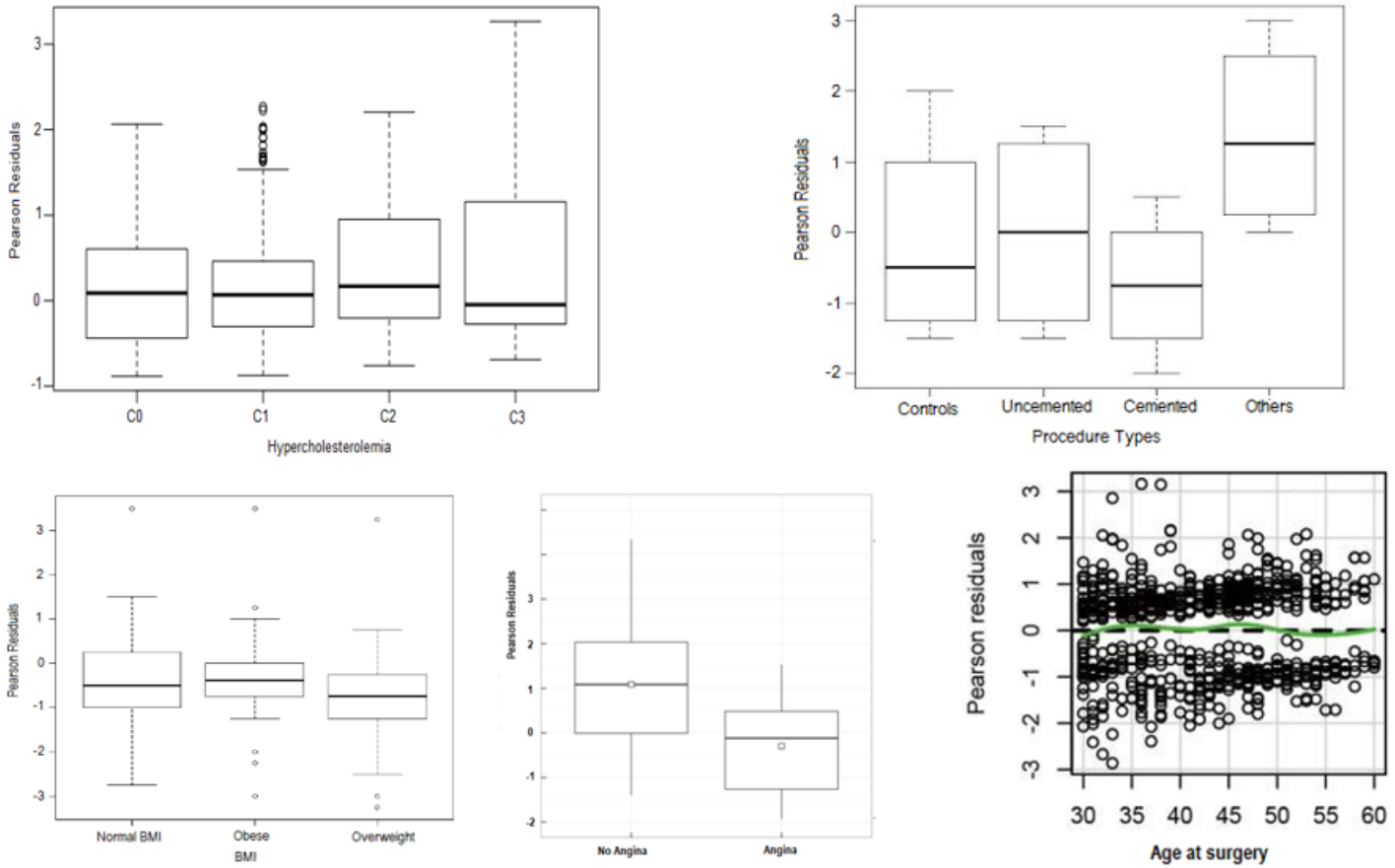
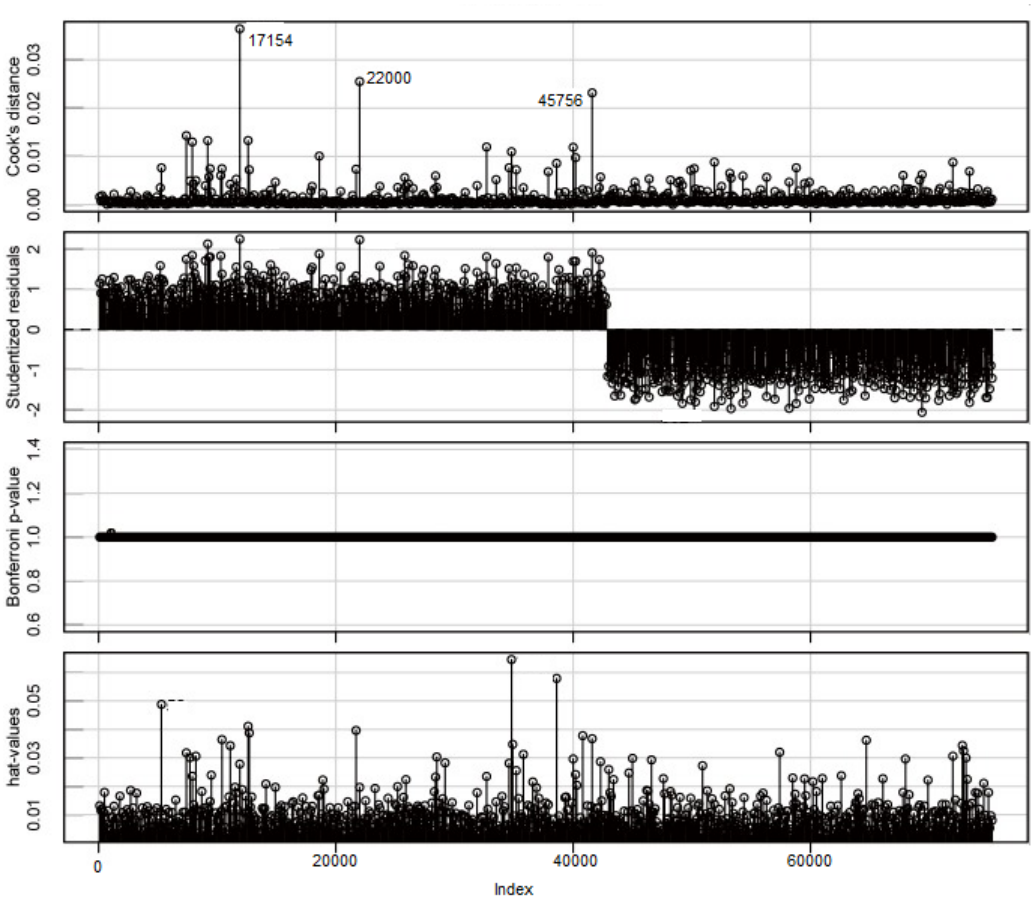
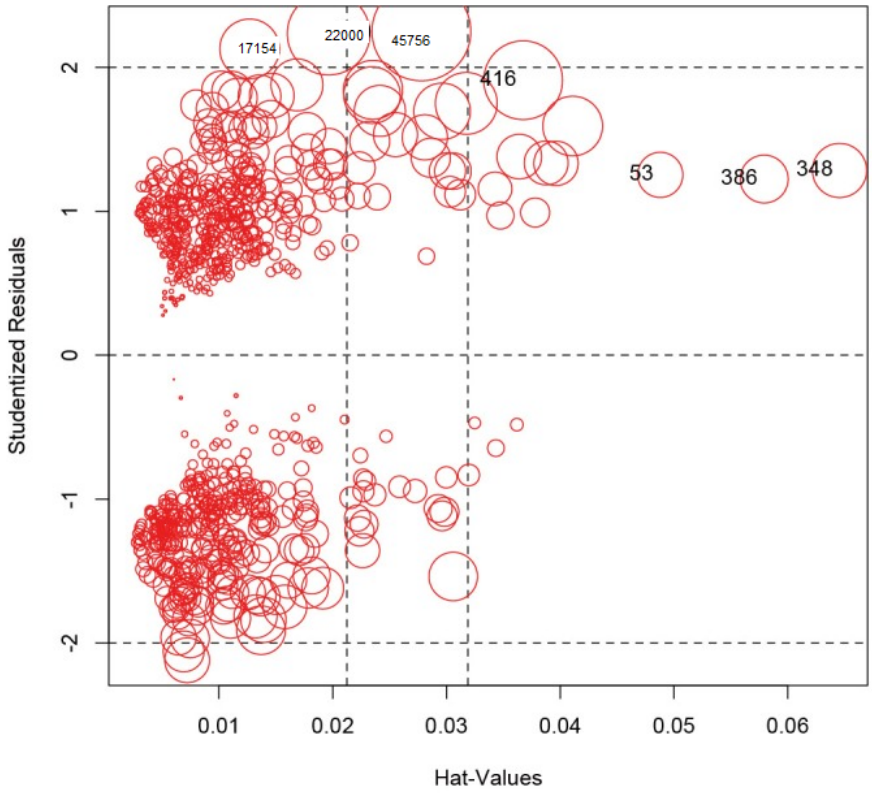


Figure 38: Diagnostic plots for short term logistic regression model with interaction terms fitted using 10 imputed datasets



The above figure displays the Cook's distance, studentized residuals and hat values associated with each observations in $model_{imp}$. Observations indexed at 17154, 22000 and 45756 are mostly likely to be outliers because they have the three largest Cook's distance.

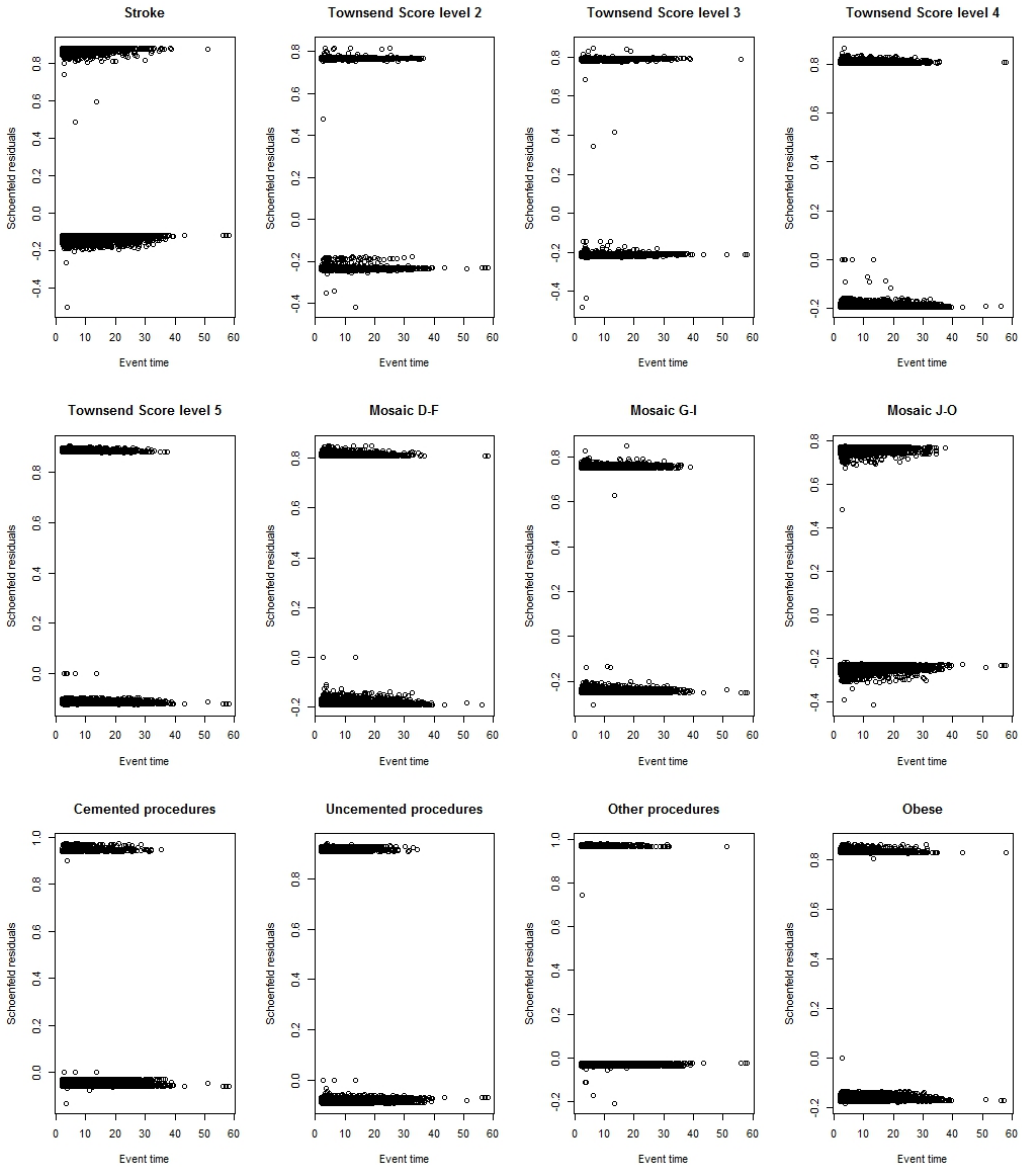
Figure 39: Assessment of influential outliers in short term logistic regression model with interaction terms fitted using 10 imputed datasets



In the above figure, studentized residuals are plotted against hat-values. The diameter of the circle around each observation is proportional to Cook's distance. Observation indexed at 45756 has the largest circle (Cook's distance=0.315) and is thus suspected to be an influential outlier. However, as a sensitivity check, removal of observation 45756 did not significantly impact on the estimated parameters of $model_{imp}$.

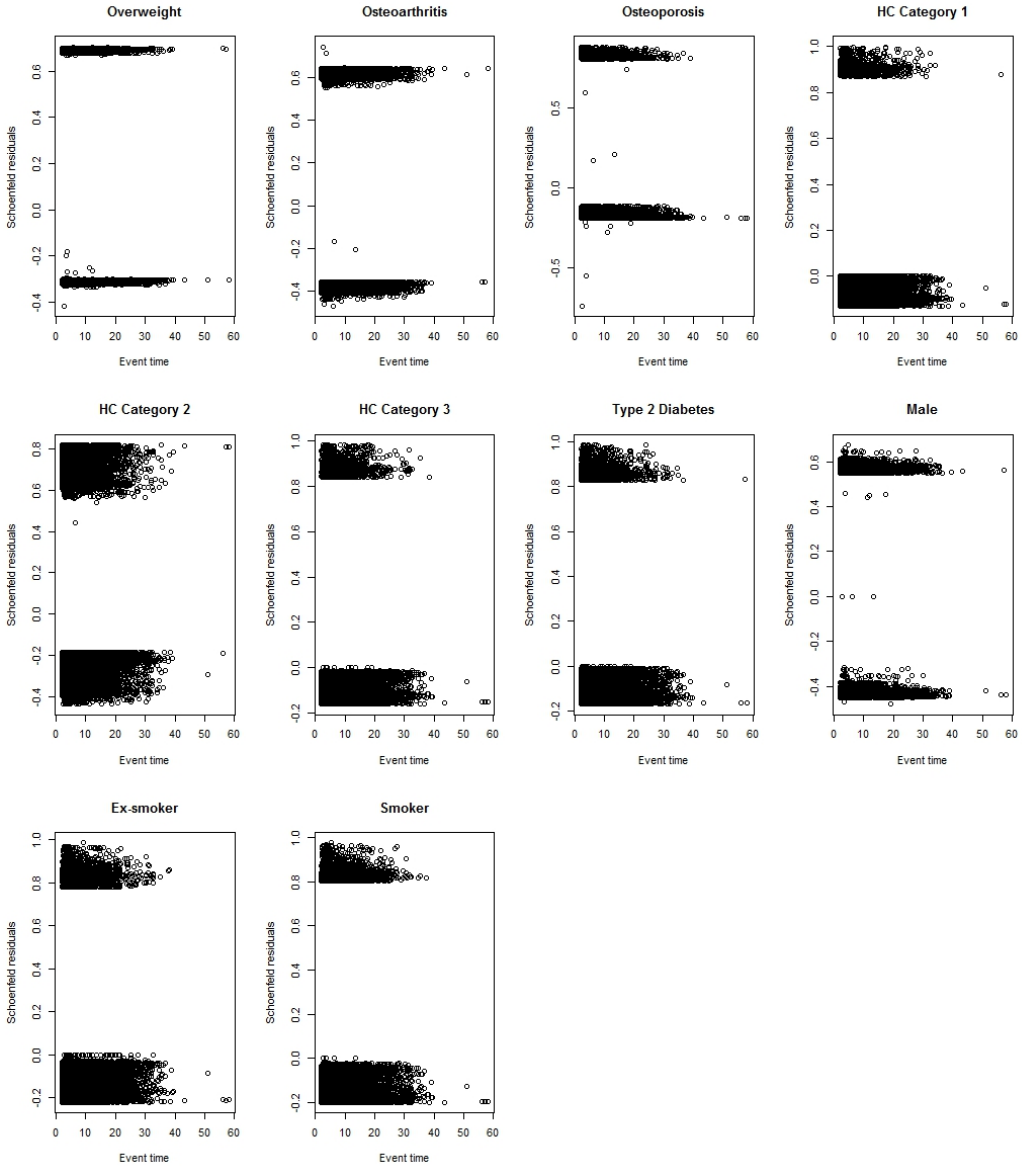
D Long Term Survival Model

Figure 40: Schoenfeld residuals distribution for model L_{full}- Part 1



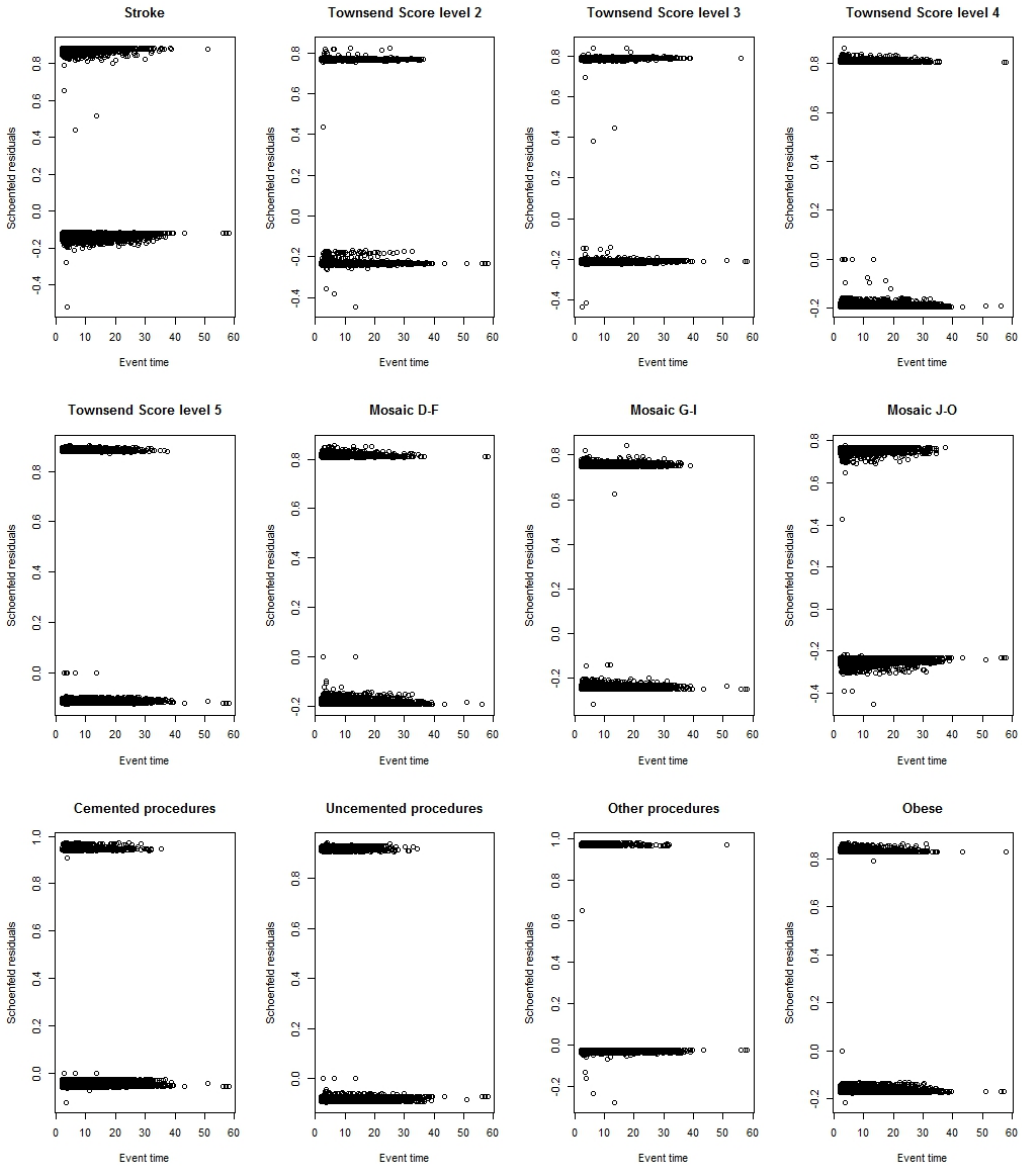
Plots of Schoenfeld residuals versus event time for the full case analysis model show that the assumption of proportional hazards is not violated. Variations in hazard ratio of death, due to stroke, Townsend scores, Mosaic categories, procedure types and obesity, respectively, are independent of the event time.

Figure 41: Schoenfeld residuals distribution for model L_{full}- Part 2



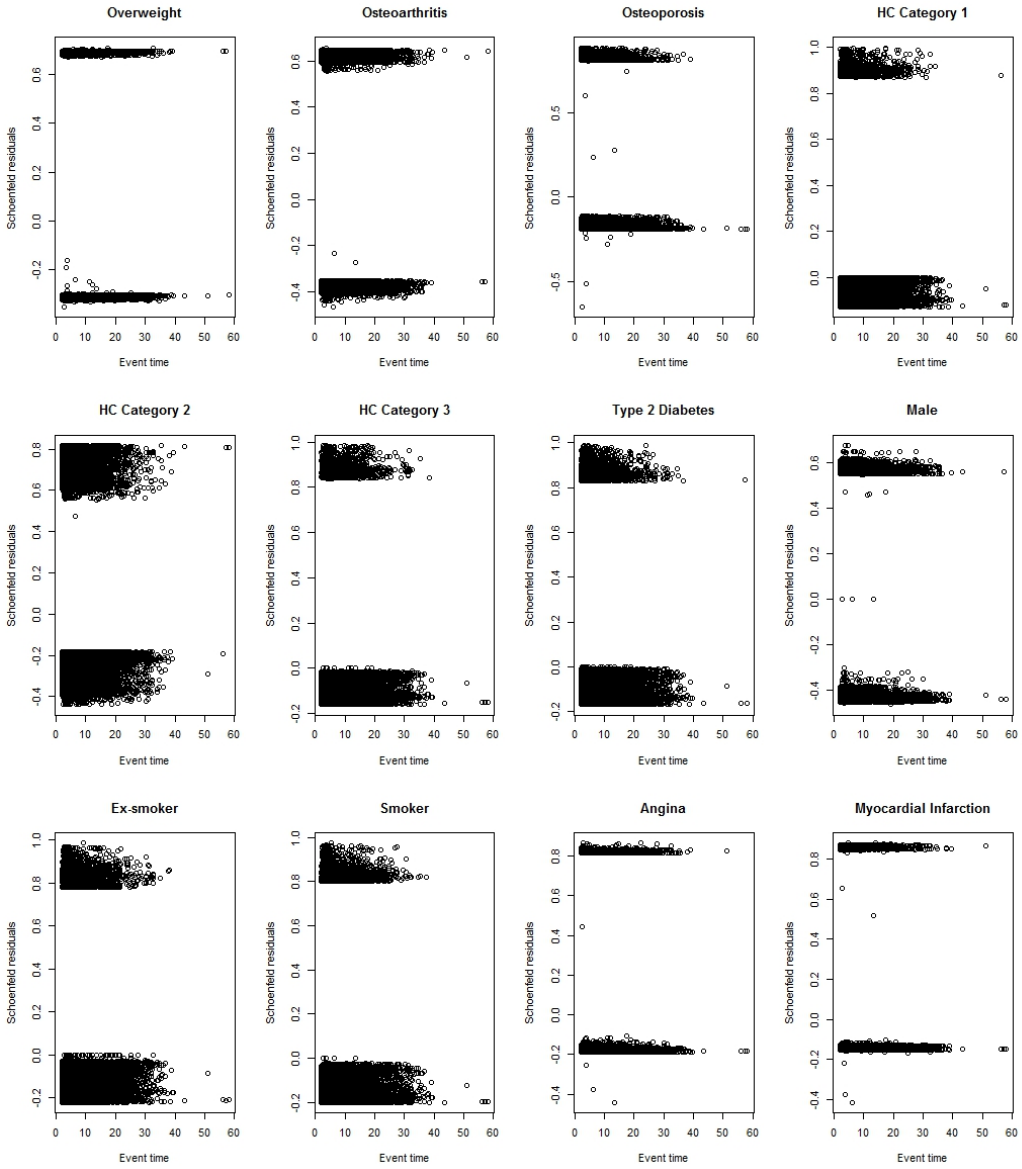
Plots of Schoenfeld residuals versus event time for the full case analysis model show that the assumption of proportional hazards is not violated. Variations in hazard ratio of death, due to overweight, osteoarthritis, osteoporosis, hypercholesterolemia, type 2 diabetes, sex and smoking status are independent of the event time.

Figure 42: Schoenfeld residuals distribution for model L_{imp}- Part 1



Plots of Schoenfeld residuals versus event time for the full case analysis model show that the assumption of proportional hazards is not violated. Variations in hazard ratio of death, due to stroke, Townsend scores, Mosaic categories, procedure types and obesity, respectively, are independent of the event time.

Figure 43: Schoenfeld residuals distribution for model L_{imp}- Part 2



Plots of Schoenfeld residuals versus event time for the full case analysis model show that the assumption of proportional hazards is not violated. Variations in hazard ratio of death, due to overweight, osteoarthritis, osteoporosis, hypercholesterolemia, type 2 diabetes, sex, smoking status, angina and myocardial infarction are independent of the event time.

Table 45: Performance statistics comparison for long term survival models of main effects only

Statistics	*Fullcaseanalysis*	*Multipleimputationanalysis*
Royston's R^2	26.1% [23.4%-28.5%]	25.2% [22.6%-27.4%]
Harrell's Concordance, C	70.3% [67.4%-73.1%]	71.5% [67.9%-74.8%]
Shrinkage	0.968	0.976

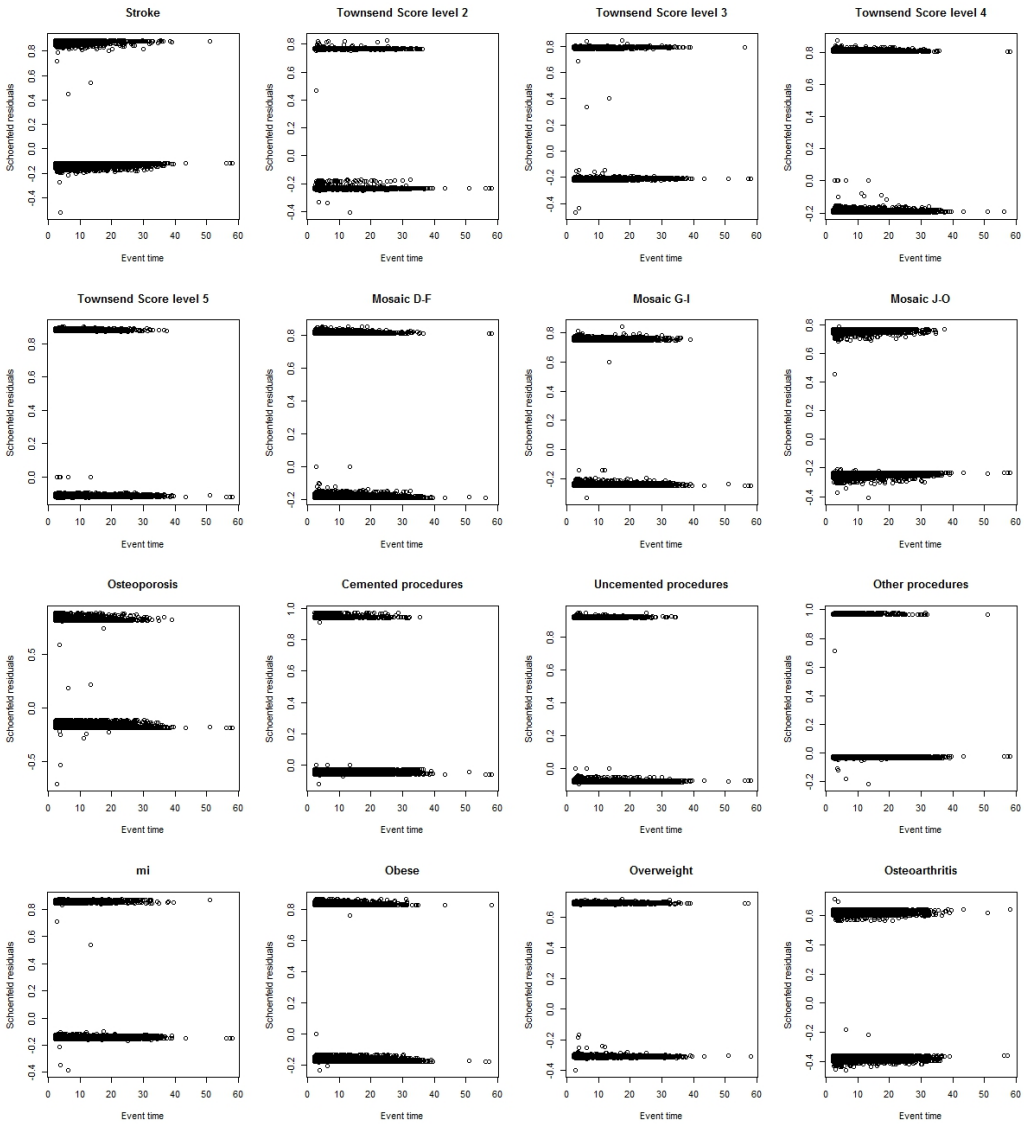
The estimated performance statistics (Royston's R^2, Harrell's concordance and shrinkage) of models fitted using full case and multiple imputation analysis are close to each other. R^2 is marginally higher for the full case record model. Estimated shrinkage is marginally less for the model fitted using multiple imputation.

Table 46: Comparing coefficients of full case and multiple imputation analyses

Coefficients	Full Case Analysis	Multiple Imputation Analysis	Δ_L
Stroke	0.2960	0.3122	5%
Townsend Score 2	0.0803	0.0830	3%
Townsend Score 3	0.1146	0.1092	-5%
Townsend Score 4	0.1773	0.1519	-17%
Townsend Score 5	0.2129	0.1763	-21%
Mosaic Category D-F	0.0550	0.0655	16%
Mosaic Category G-I	0.1292	0.1246	-4%
Mosaic Category J-O	0.0602	0.0594	-1%
Uncemented THR	-0.1669	-0.1487	-12%
Cemented THR	-0.3167	-0.2875	-10%
Other THR	-0.1478	-0.1278	-16%
Obese	-0.1787	-0.1878	5%
Overweight	-0.2364	-0.2211	-7%
OA*	0.0831	0.0760	-9%
Osteoporosis	0.3278	0.4021	18%
HC** Category 1	0.1317	0.1303	-1%
HC** Category 2	0.4531	0.4814	6%
HC** Category 3	0.5265	0.5327	1%
Type 2 Diabetes	0.7313	0.7626	4%
Sex	0.1363	0.1512	10%
Ex-smoker	0.5126	0.5933	14%
Smoker	0.9248	0.9475	2%
Angina	NA	0.4795	NA
MI	NA	0.3578	NA

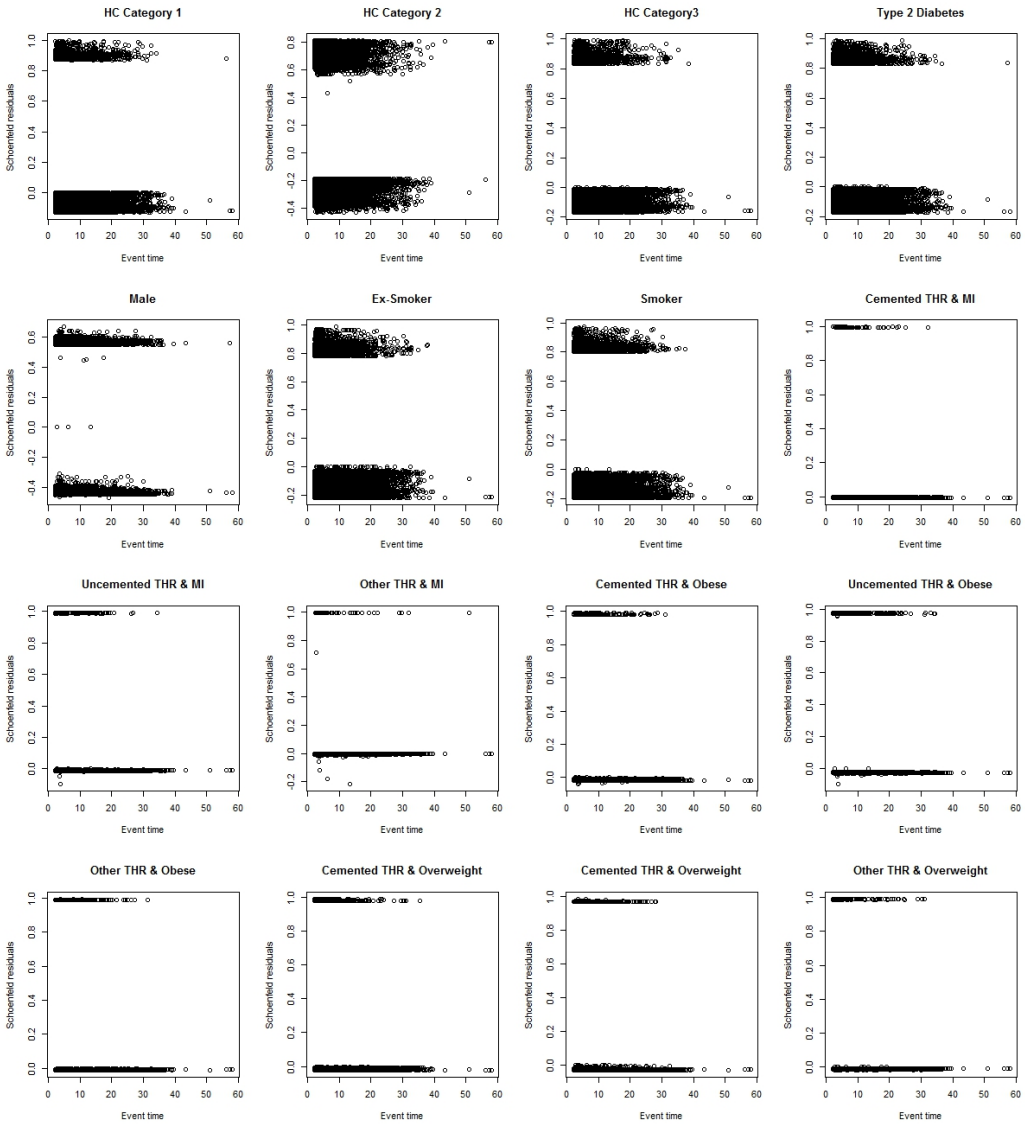
**OA=Osteoarthritis, **HC=Hypercholesterolemia*
Estimated Δ_L's between full case and multiple imputation analyses indicate that full case analysis underestimate ($-\Delta_L$) and overestimate ($+\Delta_L$) for most model coefficients.

Figure 44: Schoenfeld residuals distribution for model $long_{full}$- Part 1



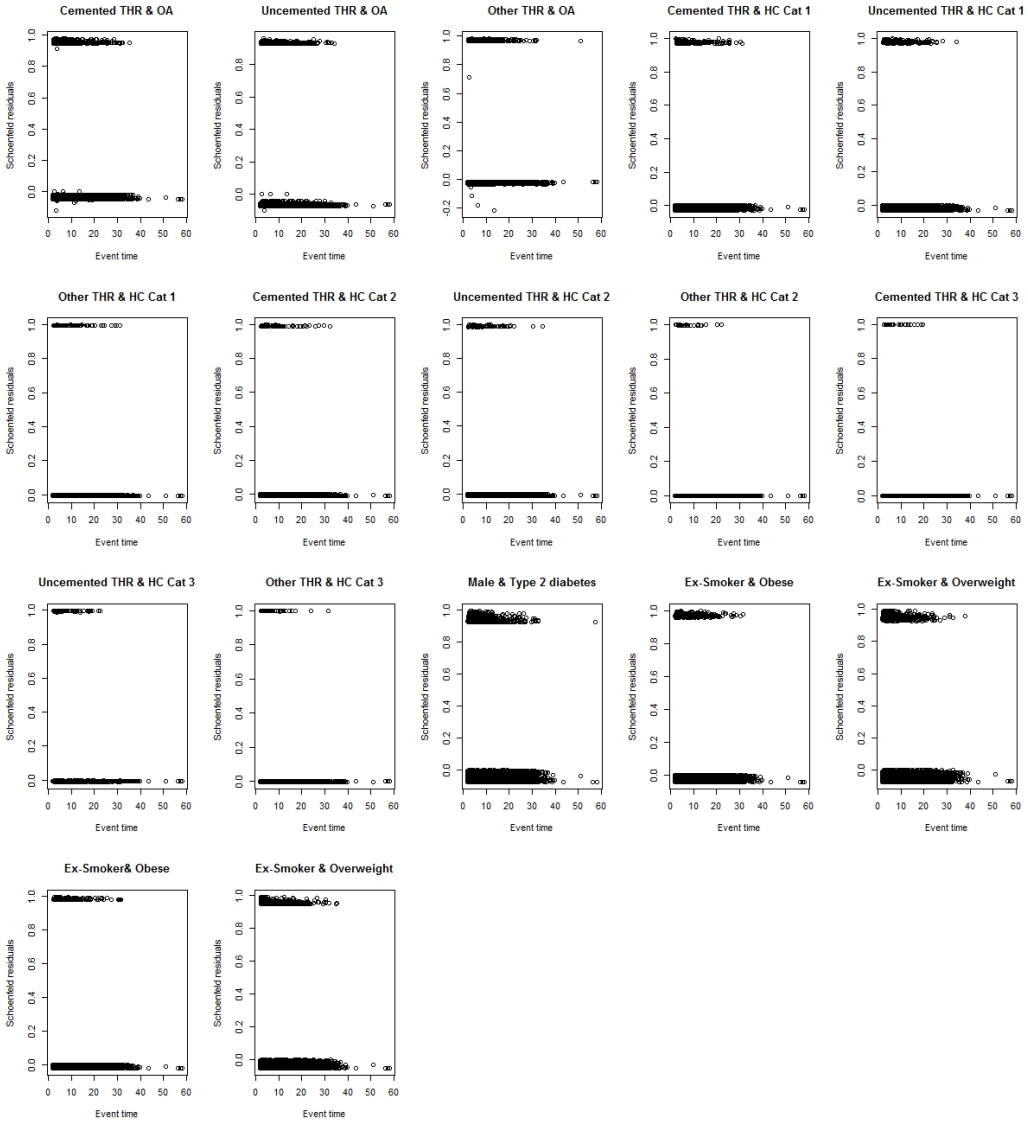
Plots of Schoenfeld residuals versus event time for the full case analysis model show that the assumption of proportional hazards is not violated. Variations in hazard ratio of death are independent of time.

Figure 45: Schoenfeld residuals distribution for model $long_{full}$- Part 2



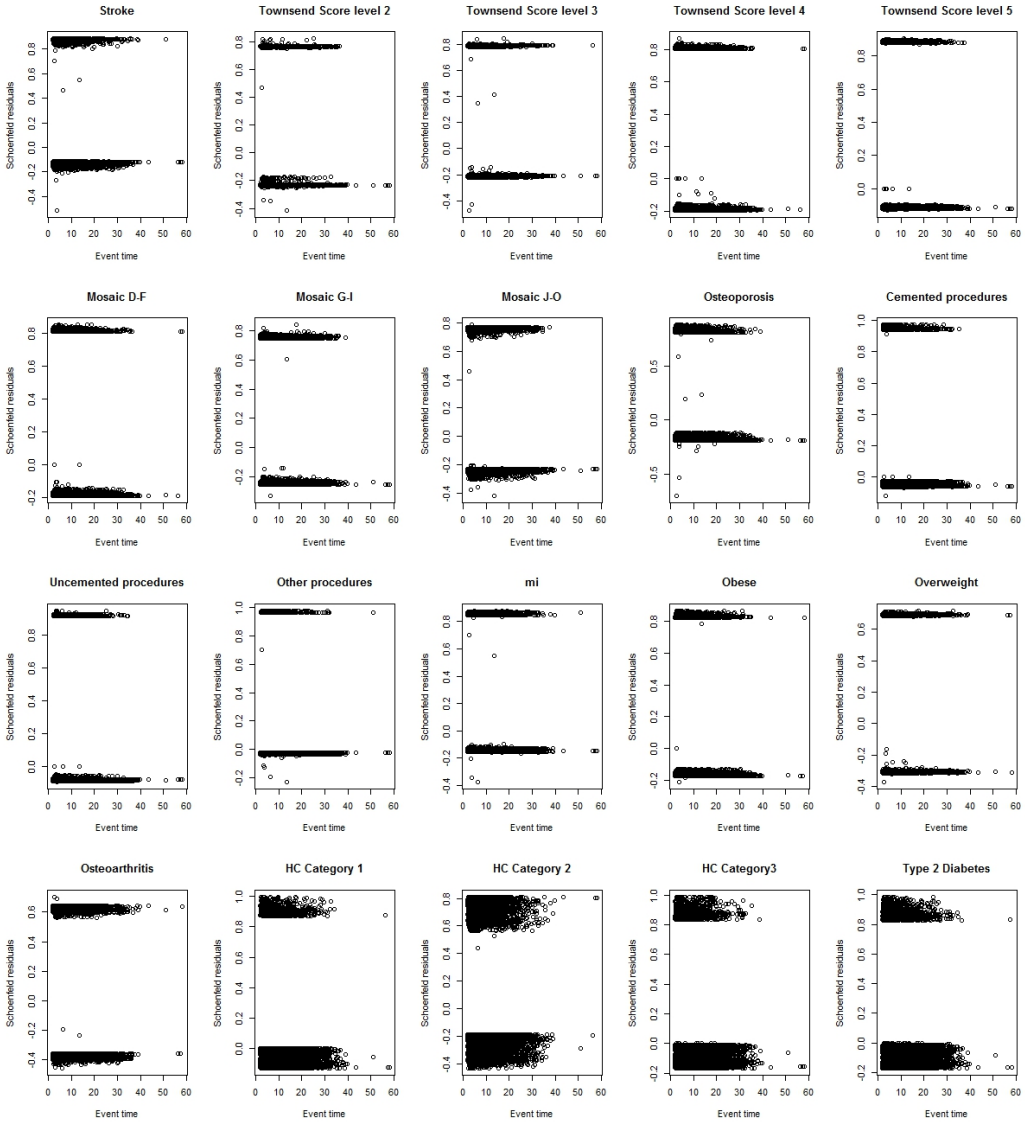
Plots of Schoenfeld residuals versus event time for the full case analysis model show that the assumption of proportional hazards is not violated. Variations in hazard ratio of death are independent of time.

Figure 46: Schoenfeld residuals distribution for model $long_{full}$- Part 3



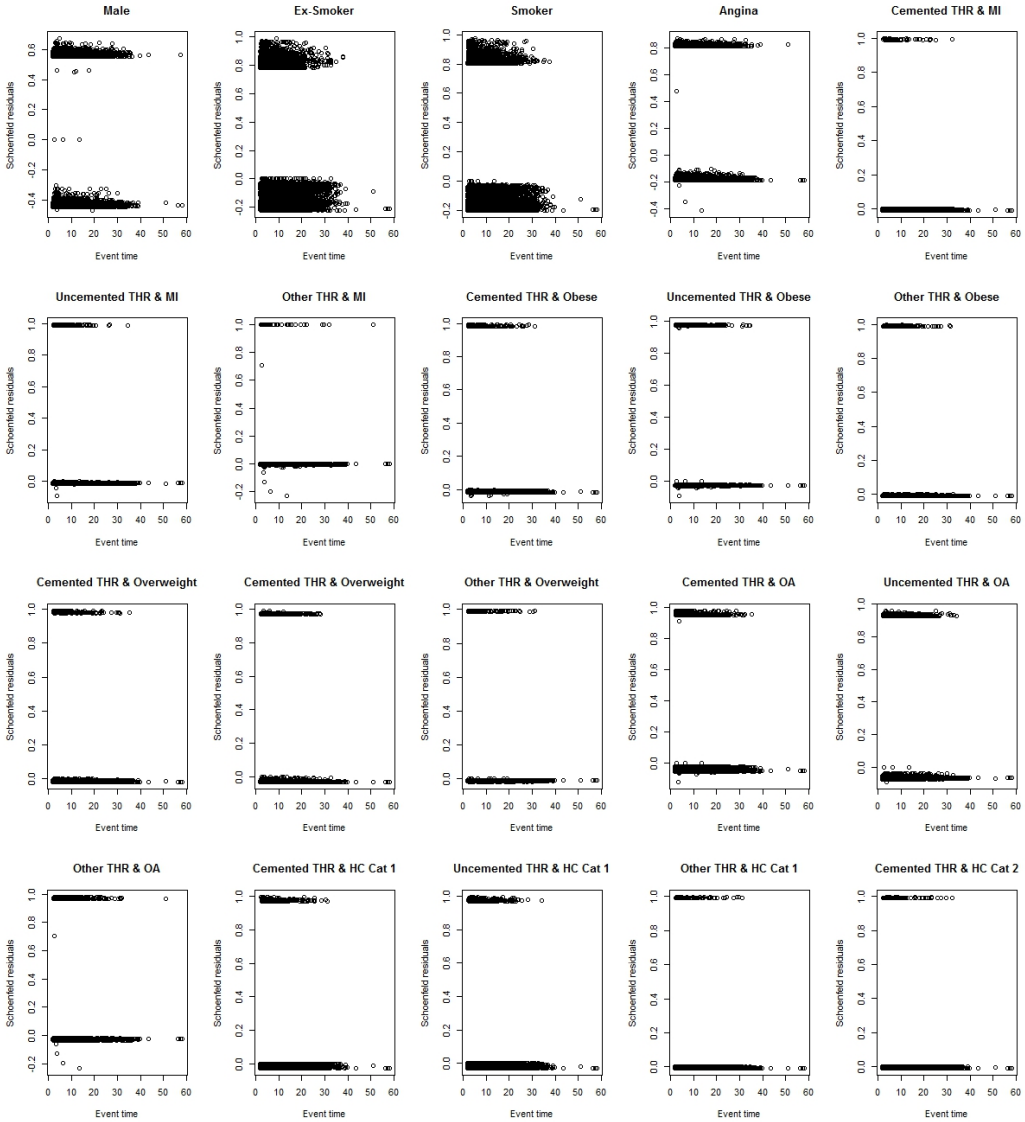
Plots of Schoenfeld residuals versus event time for the full case analysis model show that the assumption of proportional hazards is not violated. Variations in hazard ratio of death are independent of time.

Figure 47: Schoenfeld residuals distribution for model $long_{imp}$- Part 1



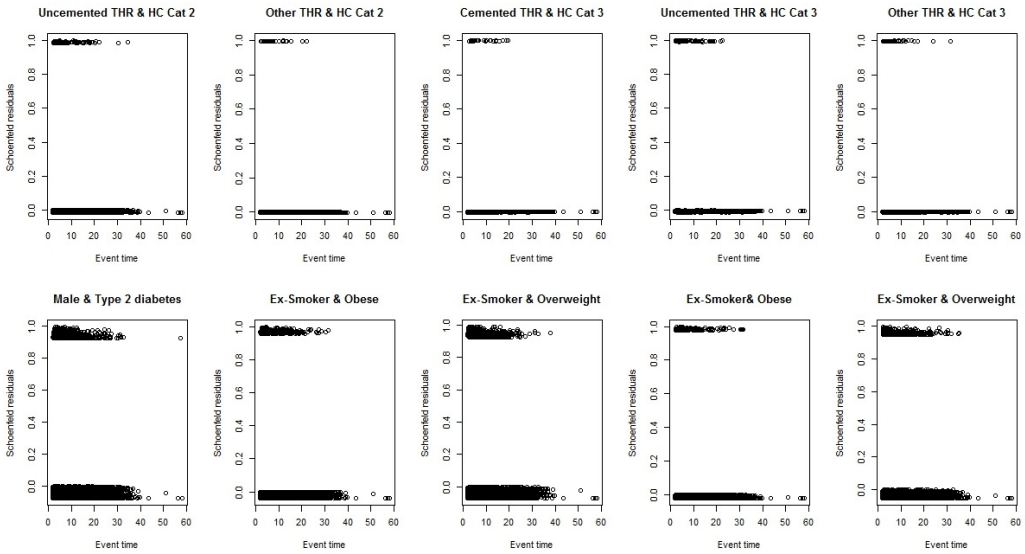
Plots of Schoenfeld residuals versus event time for the full case analysis model show that the assumption of proportional hazards is not violated. Variations in hazard ratio of death are independent of time.

Figure 48: Schoenfeld residuals distribution for model $long_{imp}$- Part 2



Plots of Schoenfeld residuals versus event time for the full case analysis model show that the assumption of proportional hazards is not violated. Variations in hazard ratio of death are independent of time.

Figure 49: Schoenfeld residuals distribution for model $long_{imp}$- Part 3



Plots of Schoenfeld residuals versus event time for the full case analysis model show that the assumption of proportional hazards is not violated. Variations in hazard ratio of death are independent of time.

Table 47: Performance statistics comparison for long term survival models including interaction effects

Statistics	*Fullcaseanalysis*	*Multipleimputationanalysis*
Royston's R^2	25.7% [22.7%-28.1%]	25.4% [22.1%-27.6%]
Harrell's Concordance, C	71.6% [68.4%-74.2%]	72.3% [68.9%-75.4%]
Shrinkage	0.961	0.969

The estimated performance statistics (Royston's R^2, Harrell's concordance and shrinkage) of models fitted using full case and multiple imputation analysis are close to each other. R^2 is marginally higher for the full case record model. Estimated shrinkage is marginally less for the model fitted using multiple imputation.

References

[1] Abdulkarim, A., Ellanti, P., Motterlini, N., Fahey, T., and O'Byrne, J. M. (2013). Cemented versus uncemented fixation in total hip replacement: a systematic review and meta-analysis of randomized controlled trials. *Orthopedic reviews*, 5(1).

[2] Adams, K. F., Schatzkin, A., Harris, T. B., Kipnis, V., Mouw, T., Ballard-Barbash, R., Hollenbeck, A., and Leitzmann, M. F. (2006). Overweight, obesity, and mortality in a large prospective cohort of persons 50 to 71 years old. *New England Journal of Medicine*, 355(8):763–778.

[3] Ahmad, S. (2013). *Diabetes: An Old Disease, a New Insight*. Springer Science & Business Media.

[4] Amstutz, H., Yao, J., Dorey, F., and Nugent, J. (1988). Survival analysis of t-28 hip arthroplasty with clinical implications. *The Orthopedic clinics of North America*, 19(3):491–503.

[5] Arjas, E. (1988). A graphical method for assessing goodness of fit in cox's proportional hazards model. *Journal of the American Statistical Association*, 83(401):204–212.

[6] Asparouhov, T. and Muthén, B. (2010). Weighted least squares estimation with missing data. *Mplus Technical Appendix*, 2010:1–10.

[7] Aynardi, M., Pulido, L., Parvizi, J., Sharkey, P. F., and Rothman, R. H. (2009). Early mortality after modern total hip arthroplasty. *Clinical orthopaedics and related research*, 467(1):213–218.

[8] Bakaeen, F. G. and Chu, D. (2011). The obesity paradox and cardiac surgery: are we sending the wrong message? *The Annals of thoracic surgery*, 92(3):1153.

[9] Banack, H. R. and Kaufman, J. S. (2014). The obesity paradox: understanding the effect of obesity on mortality among individuals with cardiovascular disease. *Preventive medicine*, 62:96–102.

[10] Barrett, J., Losina, E., Baron, J. A., Mahomed, N. N., Wright, J., and Katz, J. N. (2005). Survival following total hip replacement. *JBJS*, 87(9):1965–1971.

[11] Barrieu, P. et al. (2012). *2012*. Understanding, Modeling and Managing Longevity Risk: Key Issues and Main Challenges. Scand. Actuar.

[12] Bartley, M. (2016). *Health inequality: an introduction to concepts, theories and methods*. John Wiley & Sons.

[13] Berenbaum, F. (2013). Osteoarthritis as an inflammatory disease (osteoarthritis is not osteoarthrosis!). *Osteoarthritis and Cartilage*, 21(1):16–21.

[14] Berry, D., Trousdale, R., Dennis, D., and Paprosky, W. (2012). *Revision Total Hip and Knee Arthroplasty*. Wolters Kluwer Health.

[15] Berry, D. J., Harmsen, W. S., Cabanela, M. E., and Morrey, B. F. (2002). Twenty-five-year survivorship of two thousand consecutive primary charnley total hip replacements: factors affecting survivorship of acetabular and femoral components. *JBJS*, 84(2):171–177.

[16] Berstock, J., Beswick, A., Lenguerrand, E., Whitehouse, M., and Blom, A. (2014). Mortality after total hip replacement surgery. *Bone and Joint Research*, 3(6):175–182.

[17] Bhansali, A. and Gogate, Y. (2015). *Clinical Rounds in Endocrinology: Volume I - Adult Endocrinology*. Springer India.

[18] Bhatnagar, D., Soran, H., and Durrington, P. N. (2008). Hypercholesterolaemia and its management. *BMJ*, 337.

[19] Black, A. C., Harel, O., and Betsy McCoach, D. (2011). Missing data techniques for multilevel data: Implications of model misspecification. *Journal of Applied Statistics*, 38(9):1845–1865.

[20] Blak, B., Thompson, M., Dattani, H., and Bourke, A. (2011). Generalisability of the health improvement network (thin) database: demographics, chronic disease prevalence and mortality rates. *Informatics in Primary Care*, 19(4):251–255.

[21] Blom, A., Pattison, G., Whitehouse, S., Taylor, A., and Bannister, G. (2006). Early death following primary total hip arthroplasty: 1,727 procedures with mechanical thrombo-prophylaxis. *Acta orthopaedica*, 77(3):347–350.

[22] BNF Publications (2016). British National Formulary, General Information.

[23] Boniello, A. J., Simon, M. S., Emenari, C. C., and Courtney, P. M. (2017). Complications and mortality following total hip arthroplasty in the octogenarians: An analysis of a national database. *The Journal of Arthroplasty*.

[24] Bourke, A., Dattani, H., and Robinson, M. (2004). Feasibility study and methodology to create a quality-evaluated database of primary care data. *Journal of Innovation in Health Informatics*, 12(3):171–177.

[25] Bozic, K. J., Lau, E., Kurtz, S., Ong, K., Rubash, H., Vail, T. P., and Berry, D. J. (2012). Patient-related risk factors for periprosthetic joint infection and postoperative mortality following total hip arthroplasty in medicare patients. *JBJS*, 94(9):794–800.

[26] Brenner, H., Gefeller, O., and Greenland, S. (1993). Risk and rate advancement periods as measures of exposure impact on the occurrence of chronic diseases. *Epidemiology*, 4(3):229–236.

[27] Breslow, N. E. (1972). Discussion of professor cox's paper. *J Royal Stat Soc B*, 34:216–217.

[28] Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American statistical Association*, 88(421):9–25.

[29] British National Formulary 2.12 (2014). Lipid-regulating drug, Chapter 2.12. `https://www.evidence.nhs.uk/formulary/bnf/current/2-cardiovascular-system/212-lipid-regulating-drugs/statins`. [Online; accessed 25-January-2017].

[30] Brown, H. and Prescott, R. (2014). *Applied mixed models in medicine*. John Wiley & Sons.

[31] Bull, E. and Morrell, J. (2005). *A Simple Guide to Cholesterol*. Bestmedicine health handbooks. CSF Medical Communications Limited.

[32] Campbell, S., Reeves, D., Kontopantelis, E., Middleton, E., Sibbald, B., and Roland, M. (2007). Quality of primary care in england with the

introduction of pay for performance. *New England Journal of Medicine*, 357(2):181–190.

[33] Carl Baker (2017). Obesity Statistics. `http://www.parliament.uk/commons-library`. [Online; accessed 17-January-2017].

[34] Charlson, M. E., Pompei, P., Ales, K. L., and MacKenzie, C. R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of chronic diseases*, 40(5):373–383.

[35] Charnley, J. (1951). Arthroplasty of the hip. *A new operation.. Lancet*, 1(7187):1129–1132.

[36] Charnley, J. (1972a). The long-term results of low-friction arthroplasty of the hip performed as a primary intervention. *J Bone Joint Surg [Br]*, 54:61–76.

[37] Charnley, J. (1972b). Postoperative infection after total hip replacement with special reference to air contamination in the operating room. *Clin Orthop 87*, 187:167–87.

[38] Christiansen, T. (1969). A new hip prosthesis with trunnion-bearing. *Acta Chir Scand*, 135:43–6.

[39] Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, pages 141–151.

[40] Clegg, A., Bates, C., Young, J., Ryan, R., Nichols, L., Ann Teale, E., Mohammed, M. A., Parry, J., and Marshall, T. (2016). Development and validation of an electronic frailty index using routine primary care electronic health record data. *Age and ageing*, 45(3):353–360.

[41] Clement, N., Muzammil, A., MacDonald, D., Howie, C., and Biant, L. (2011). Socioeconomic status affects the early outcome of total hip replacement. *J Bone Joint Surg Br*, 93(4):464–469.

[42] ClinicalCodes (2016). ClinicalCodes Organisation. `https://clinicalcodes.rss.mhs.man.ac.uk`.

[43] Comba, F., Hidalgo, I. A., Buttaro, M., and Piccaluga, F. (2012). Risk factor analysis for 30-day mortality after primary tha in a single institution. *HSS Journal®*, 8(2):111–115.

[44] Cook, S. (2000). *Comparative Spatial Deprivation in Ireland: A Cross Border Analysis*. Oak Tree Press.

[45] Cornfield, J. (1962). Joint dependence of risk of coronary heart disease on serum cholesterol and systolic blood pressure: a discriminant function analysis. In *Fed Proc*, volume 21, pages 58–61.

[46] Cox, D. and Oakes, D. (1984). *Analysis of Survival Data*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

[47] Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220.

[48] Cram, P., Lu, X., Kaboli, P. J., Vaughan-Sarrazin, M. S., Cai, X., Wolf, B. R., and Li, Y. (2011). Clinical characteristics and outcomes of medicare patients undergoing total hip arthroplasty, 1991-2008. *Jama*, 305(15):1560–1567.

[49] Curtis, J. P., Selter, J. G., Wang, Y., Rathore, S. S., Jovin, I. S., Jadbabaie, F., Kosiborod, M., Portnay, E. L., Sokol, S. I., Bader, F., et al. (2005). The obesity paradox: body mass index and outcomes in patients with heart failure. *Archives of internal medicine*, 165(1):55–61.

[50] Czepiel, S. A. (2002). Maximum likelihood estimation of logistic regression models: theory and implementation. *Available at czep. net/stat/mlelr. pdf*.

[51] Daabiss, M. (2011). American society of anaesthesiologists physical status classification. *Indian journal of anaesthesia*, 55(2):111.

[52] Department of Health (2011). The Good Practice Guidelines for GP electronic patient records. `https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/215680/dh_125350.pdf`.

[53] Diez Roux, A. V. (2001). Investigating neighborhood and area effects on health. *American journal of public health*, 91(11):1783–1789.

[54] Doll, R. and Hill, A. B. (1950). Smoking and carcinoma of the lung. *British medical journal*, 2(4682):739.

[55] Doll, R., Peto, R., Boreham, J., and Sutherland, I. (2004). Mortality in relation to smoking: 50 years' observations on male british doctors. *Bmj*, 328(7455):1519.

[56] Duff-Barclay, I. and Spillman, D. (1966). Paper 10: Total human hip joint prostheses—a laboratory study of friction and wear. In *Proceedings of the Institution of Mechanical Engineers, Conference Proceedings*, volume 181, pages 90–103. SAGE Publications Sage UK: London, England.

[57] Efron, B. (1977). The efficiency of cox's likelihood function for censored data. *Journal of the American statistical Association*, 72(359):557–565.

[58] Enders, C. (2010). *Applied Missing Data Analysis*. Methodology in the social sciences. Guilford Press.

[59] Engh, C. A., Bobyn, J., and Glassman, A. H. (2006). Porous-coated hip replacement. the factors governing bone ingrowth, stress shielding, and clinical results. *The Journal of bone and joint surgery. British volume*, 69(1):45–55.

[60] English indices of deprivation (2015). Department for Communities and Local Government.

[61] European Union Convention (2004). European union convention on human rights. *Article 5 of Council Directive. OJ L, Volume 37, p. 373.*, 37:373.

[62] Everitt, B. S. and Palmer, C. (2011). *Encyclopaedic companion to medical statistics*. John Wiley & Sons.

[63] Experian Ltd (2004). Mosaic uk: The consumer classification for the uk.

[64] Fender, D., Harper, W., Thompson, J., and Gregg, P. (1997). Mortality and fatal pulmonary embolism after primary total hip replacement. *J Bone Joint Surg Br*, 79(6):896–899.

[65] Gaston, M., Amin, A., Clayton, R., and Brenkel, I. (2007). Does a history of cardiac disease or hypertension increase mortality following primary elective total hip arthroplasty? *The Surgeon*, 5(5):260–265.

[66] Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, pages 457–472.

[67] Genazzani, A. (2002). *Hormone Replacement Therapy and Cancer: The Current Status of Research and Practice*. Controversial issues in climacteric medicine series. Taylor & Francis.

[68] Gill, J. (2014). *Bayesian methods: A social and behavioral sciences approach*, volume 20. CRC press.

[69] Gitsels, L. A. (2017). *Cardiovascular disease and its impact on longevity and longevity improvement*. PhD thesis, University of East Anglia, Norwich.

[70] Gitsels, L. A., Kulinskaya, E., and Steel, N. (2016). Survival benefits of statins for primary prevention: a cohort study. *PloS one*, 11(11):e0166847.

[71] Gluck, T. (1891). Erklärung. *Arch Klin Chir*, 41:747.

[72] Gnani, S. and Majeed, A. (2006). *A user's guide to data collected in primary care in England*. ERPHO.

[73] Goldstein, H. and Rasbash, J. (1996). Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, pages 505–513.

[74] González, E. M., Johansson, S., Wallander, M.-A., and Rodríguez, L. G. (2009). Trends in the prevalence and incidence of diabetes in the uk: 1996–2005. *Journal of Epidemiology & Community Health*, 63(4):332–336.

[75] Gordon, R. (2012). *Applied Statistics for the Social and Health Sciences*. Taylor & Francis.

[76] Gossec, L., Tubach, F., Baron, G., Ravaud, P., Logeart, I., and Dougados, M. (2005). Predictive factors of total hip replacement due to primary osteoarthritis: a prospective 2 year study of 505 patients. *Annals of the Rheumatic Diseases*, 64(7):1028–1032.

[77] Graham, J. W., Olchowski, A. E., and Gilreath, T. D. (2007). How many imputations are really needed? some practical clarifications of multiple imputation theory. *Prevention science*, 8(3):206–213.

[78] Grambsch, P. M. and Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, pages 515–526.

[79] Greenland, S. (2003). Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*, pages 300–306.

[80] Greenwood, M. and Yule, G. U. (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *Journal of the Royal statistical society*, 83(2):255–279.

[81] Grizzle, J. E., Starmer, C. F., and Koch, G. G. (1969). Analysis of categorical data by linear models. *Biometrics*, pages 489–504.

[82] Grønnesby, J. K. and Borgan, Ø. (1996). A method for checking regression models in survival analysis based on the risk score. *Lifetime data analysis*, 2(4):315–328.

[83] Gruer, L., Cézard, G., Clark, E., Douglas, A., Steiner, M., Millard, A., Buchanan, D., Katikireddi, S. V., Sheikh, A., and Bhopal, R. (2016). Life expectancy of different ethnic groups using death records linked to population census data for 4.62 million people in scotland. *Journal of Epidemiology & Community Health*.

[84] Grund, S., Robitzsch, A., and Luedtke, O. (2017). Package 'mitml'.

[85] Haass, M., Kitzman, D. W., Anand, I. S., Miller, A., Zile, M. R., Massie, B. M., and Carson, P. E. (2011). Body mass index and adverse cardiovascular outcomes in heart failure patients with preserved ejection fraction: results from the irbesartan in heart failure with preserved ejection fraction (i-preserve) trial. *Circulation: Heart Failure*, 4(3):324–331.

[86] Hall, G. (2009). Validation of death and suicide recording on the thin uk primary care database. *Pharmacoepidemiology and Drug Safety*, 18(2):120–131.

[87] Hallan, G. (2007). Wear fixation and revision of total hip prosthesis'. *Faculty of Medicine, University of Bergen, Norway*.

[88] Hansson, G. K. and Hermansson, A. (2011). The immune system in atherosclerosis. *Nature immunology*, 12(3):204.

[89] Harel, O. and Zhou, X.-H. (2007). Multiple imputation: review of theory, implementation and software. *Statistics in medicine*, 26(16):3057–3077.

[90] Harrell, F. (2001). *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Graduate Texts in Mathematics. Springer.

[91] Harris, W. H. and McGann, W. (1986). Loosening of the femoral component after use of the medullary-plug cementing technique. follow-up note with a minimum five-year follow-up. *The Journal of bone and joint surgery. American volume*, 68(7):1064–1066.

[92] Haynes, D., Crotti, T., and Zreiqat, H. (2004). Regulation of osteoclast activity in peri-implant tissues. *Biomaterials*, 25(20):4877–4885.

[93] Health Survey for England (2015). Prevalence of obesity in UK and Ireland.

[94] Hennessy, S., Bilker, W. B., Berlin, J. A., and Strom, B. L. (1999). Factors influencing the optimal control case ratio in matched case control studies. *American Journal of Epidemiology, 1999*, 149(2):195–197.

[95] Hernán, M. A., Hernández-Díaz, S., and Robins, J. M. (2004). A structural approach to selection bias. *Epidemiology*, pages 615–625.

[96] Hernán, M. A., Hernández-Díaz, S., Werler, M. M., and Mitchell, A. A. (2002). Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *American journal of epidemiology*, 155(2):176–184.

[97] Herrett, E., Gallagher, A. M., Bhaskaran, K., Forbes, H., Mathur, R., van Staa, T., and Smeeth, L. (2015). Data resource profile: clinical practice research datalink (cprd). *International journal of epidemiology*, 44(3):827–836.

[98] Hippisley-Cox, J. and Coupland, C. (2010a). Predicting the risk of chronic kidney disease in men and women in england and wales: prospective derivation and external validation of the qkidney® scores. *BMC family practice*, 11(1):49.

[99] Hippisley-Cox, J. and Coupland, C. (2010b). Unintended effects of statins in men and women in england and wales: population based cohort study using the qresearch database. *Bmj*, 340:c2197.

[100] Hippisley-Cox, J., Stables, D., and Pringle, M. (2004). Qresearch: a new general practice database for research. *Journal of Innovation in Health Informatics*, 12(1):49–50.

[101] Holmberg, S. (1992). Life expectancy after total hip arthroplasty. *The Journal of arthroplasty*, 7(2):183–186.

[102] Hosmer, D., Lemeshow, S., and May, S. (2011). *Applied Survival Analysis: Regression Modeling of Time to Event Data*. Wiley Series in Probability and Statistics. Wiley.

[103] Hosmer, D. W. and Lemesbow, S. (1980). Goodness of fit tests for the multiple logistic regression model. *Communications in statistics-Theory and Methods*, 9(10):1043–1069.

[104] Hosmer, D. W. and Lemeshow, S. (2000). *Special topics*. Wiley Online Library.

[105] Hosmer, D. W., Taber, S., and Lemeshow, S. (1991). The importance of assessing the fit of logistic regression models: a case study. *American journal of public health*, 81(12):1630–1635.

[106] Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X. (2013a). *Applied logistic regression*, volume 398. John Wiley & Sons.

[107] Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X. (2013b). *Applied logistic regression*, volume 398. John Wiley & Sons.

[108] Howie, D., Campbell, D., McGee, M., and Cornish, B. (1990). Wagner resurfacing hip arthroplasty. *The results of one hundred consecutive arthroplasties after eight to ten years.*, 72:708–14.

[109] Hozack, W. J., Rothman, R. H., Booth, R. E., Balderston, R. A., Cohn, J. C., and Pickens, G. T. (1990). Survivorship analysis of 1,041 charnley total hip arthroplasties. *The Journal of arthroplasty*, 5(1):41–47.

[110] HSCIC (2016). Health and Social Care Information Centre. `https://www.gov.uk/government/organisations/health-and-social-care-information-centre`.

[111] Hulley, S. B., Cummings, S. R., Browner, W. S., Grady, D. G., and Newman, T. B. (2013). *Designing clinical research*. Lippincott Williams & Wilkins.

[112] Hunt, L. P., Ben-Shlomo, Y., Clark, E. M., Dieppe, P., Judge, A., MacGregor, A. J., Tobias, J. H., Vernon, K., Blom, A. W., et al. (2013). 90-day mortality after 409 096 total hip replacements for osteoarthritis, from the national joint registry for england and wales: a retrospective analysis. *The Lancet*, 382(9898):1097–1104.

[113] ICD (2015). International classification of diseases, tenth revision, clinical modification.

[114] IMS Health Incorporated (2015). THIN Data Statistics. `http://www.csdmruk.imshealth.com/our-data/statistics.shtml`.

[115] IMS Health Incorporated (2017). THIN Database Homepage. `http://csdmruk.cegedim.com/our-data/our-data.shtml`. [Online; accessed 10-January-2017].

[116] Iwagami, M., Tomlinson, L. A., Mansfield, K. E., McDonald, H. I., Smeeth, L., and Nitsch, D. (2017). Prevalence, incidence, indication, and choice of antidepressants in patients with and without chronic kidney disease: a matched cohort study in uk clinical practice research datalink. *Pharmacoepidemiology and drug safety*.

[117] Jämsen, E., Puolakka, T., Eskelinen, A., Jäntti, P., Kalliovalkama, J., Nieminen, J., and Valvanne, J. (2013). Predictors of mortality following primary hip and knee replacement in the aged: a single-center analysis of 1,998 primary hip and knee replacements for primary osteoarthritis. *Acta orthopaedica*, 84(1):44–53.

[118] Jevon, P. (2012). *Angina and Heart Attack*. The Facts. OUP Oxford.

[119] Johansson, C., Black, D., Johnell, O., Odén, A., and Mellström, D. (1998). Bone mineral density is a predictor of survival. *Calcified Tissue International*, 63(3):190–196.

[120] Johnell, O. and Kanis, J. (2005). Epidemiology of osteoporotic fractures. *Osteoporosis International*, 16(2):S3–S7.

[121] Jones, M. D., Parry, M. C., Whitehouse, M. R., and Blom, A. W. (2014). Early death following primary total hip arthroplasty. *The Journal of arthroplasty*, 29(8):1625–1628.

[122] Judge, A., Chard, J., Learmonth, I., and Dieppe, P. (2006). The effects of surgical volumes and training centre status on outcomes following total joint replacement: analysis of the hospital episode statistics for england. *Journal of public health*, 28(2):116–124.

[123] Kaplan, M. J. and McCune, W. J. (2003). New evidence for vascular disease in patients with early rheumatoid arthritis. *The Lancet*, 361(9363):1068–1069.

[124] Kells, J. and Ahlgrimm, M. (2003). *HRT Solution (rev. edition): Optimizing Your Hormonal Potential*. Penguin Publishing Group.

[125] Kendal, A. R., Prieto-Alhambra, D., Arden, N. K., Carr, A., and Judge, A. (2013). Mortality rates at 10 years after metal-on-metal hip resurfacing compared with total hip replacement in england: retrospective cohort analysis of hospital episode statistics. *Bmj*, 347:f6549.

[126] Kleinbaum, D. and Klein, M. (2005). *Survival Analysis: A Self-Learning Text*. Statistics for biology and health. Springer.

[127] Klenerman, L. (2002). Arthroplasty of the hip. In L., K., editor, *The Evolution of Orthopaedic Surgery*, pages 13–23. The Royal Society of Medicine Press, London.

[128] Kynaston-Pearson, F., Ashmore, A., Malak, T., Rombach, I., Taylor, A., Beard, D., Arden, N., Price, A., Prieto-Alhambra, D., Judge, A., et al. (2013). Primary hip replacement prostheses and their evidence base: systematic review of literature. *BMJ*, 347:f6956.

[129] Lain, D. (2016). *Reconstructing Retirement: Work and Welfare in the UK and USA*. Policy Press.

[130] Laird, N. (1978). Empirical bayes methods for two-way contingency tables. *Biometrika*, 65(3):581–590.

[131] Lane-Claypon, J. E. et al. (1926). A further report on cancer of the breast with special reference to its associated antecedent conditions. *A Further Report on Cancer of the Breast with Special Reference to its Associated Antecedent Conditions.*, (32).

[132] Langholz, B. and Clayton, D. (1994). Sampling strategies in nested case-control studies. *Environmental health perspectives, 102(Suppl)*.

[133] Langley, T. E., Szatkowski, L. C., Wythe, S., and Lewis, S. A. (2011). Can primary care data be used to monitor regional smoking prevalence? an analysis of the health improvement network primary care data. *BMC Public Health*, 11(1):773.

[134] Levy, R. N., Volz, R. G., Kaufer, H., Matthews, L. S., Capozzi, J., Sturm, P., and Sherry, H. (1985). Progress in arthritis surgery: With special reference to the current status of total joint arthroplasty. *Clinical orthopaedics and related research*, 200:299–321.

[135] Lewis, J. D., Schinnar, R., Bilker, W. B., Wang, X., and Strom, B. L. (2007). Validation studies of the health improvement network (thin) database for pharmacoepidemiology research. *Pharmacoepidemiology and drug safety*, 16(4):393–401.

[136] Lewis, R. (2012). *Understanding Chronic Kidney Disease: A Guide for the Non-Specialist*. M & K Update Limited.

[137] Libby, P. (2012). Inflammation in atherosclerosis. *Arteriosclerosis, thrombosis, and vascular biology*, 32(9):2045–2051.

[138] Libby, P. and Hansson, G. K. (2015). Inflammation and immunity in diseases of the arterial tree: players and layers. *Circulation research*, 116(2):307–311.

[139] Lie, S. A. (2002). Survival studies of total hip replacements and postoperative mortality. *Department of Public Health and Primary Health Care and Institute of Surgical Sciences, Section 3. 2*, 2.(2).

[140] Lie, S. A., Engesæter, L. B., Havelin, L. I., Furnes, O., and Vollset, S. E. (2002). Early postoperative mortality after 67,548 total hip replacements. *Acta Orthopaedica Scandinavica*, 73(4):392–399.

[141] Lie, S. A., Engesæter, L. B., Havelin, L. I., Gjessing, H. K., and Vollset, S. E. (2000). Mortality after total hip replacement: 0-10-year follow-up of 39,543 patients in the norwegian arthroplasty register. *Acta Orthopaedica Scandinavica*, 71(1):19–27.

[142] Lie, S. A., Pratt, N., Ryan, P., Engesæter, L. B., Havelin, L. I., Furnes, O., and Graves, S. (2010). Duration of the increase in early postoperative mortality after elective hip and knee replacement. *JBJS*, 92(1):58–63.

[143] Lin, D. Y. and Wei, L.-J. (1989). The robust inference for the cox proportional hazards model. *Journal of the American statistical Association*, 84(408):1074–1078.

[144] Lin, D. Y., Wei, L.-J., and Ying, Z. (1993). Checking the cox model with cumulative sums of martingale-based residuals. *Biometrika*, 80(3):557–572.

[145] Lindley, R. (2008). *Stroke*. The Facts. OUP Oxford.

[146] Linsley, P., Kane, R., and Owen, S. (2011). *Nursing for Public Health: Promotion, Principles and Practice*. OUP Oxford.

[147] Lip, G. and Hall, J. (2007). *Comprehensive Hypertension E-Book*. Elsevier Health Sciences.

[148] Little, J. (1995). Consistency of asa grading. *Anaesthesia*, 50(7):658.

[149] Little, R. (1987). A test of missing completely at random for multivariate data with missing values. *Journal of American Statistical Association*, 83(404):1198–1202.

[150] Lloyd, C. and Heller, T. (2011). *Long-Term Conditions: Challenges in Health & Social Care*. SAGE Publications.

[151] Loomis, A. K., Kabadi, S., Preiss, D., Hyde, C., Bonato, V., St. Louis, M., Desai, J., Gill, J. M., Welsh, P., Waterworth, D., et al. (2016). Body mass index and risk of nonalcoholic fatty liver disease: two electronic

health record prospective studies. *The Journal of Clinical Endocrinology & Metabolism*, 101(3):945–952.

[152] Lovald, S. T., Ong, K. L., Lau, E. C., Schmier, J. K., Bozic, K. J., and Kurtz, S. M. (2014). Mortality, cost, and downstream disease of total hip arthroplasty patients in the medicare population. *The Journal of arthroplasty*, 29(1):242–246.

[153] MacDonald, T. M. and Morant, S. V. (2008). Prevalence and treatment of isolated and concurrent hypertension and hypercholesterolaemia in the united kingdom. *British Journal of Clinical Pharmacology*, 65(5):775–786.

[154] Macintyre, S., Ellaway, A., and Cummins, S. (2002). Place effects on health: how can we conceptualise, operationalise and measure them? *Social science & medicine*, 55(1):125–139.

[155] Maguire, Blak, and Thompson (2009). The importance of defining periods of complete mortality reporting for research using automated data from primary care. *Pharmacoepidemiol Drug Saf*, 18(1):76–83.

[156] Mak, P., Campbell, R., and Irwin, M. (2002). The asa physical status classification: inter-observer consistency. *Anaesthesia and intensive care*, 30(5):633.

[157] Mäkelä, K. T., Matilainen, M., Pulkkinen, P., Fenstad, A. M., Havelin, L. I., Engesaeter, L., Furnes, O., Overgaard, S., Pedersen, A. B., Kärrholm, J., et al. (2014). Countrywise results of total hip replacement: an analysis of 438,733 hips based on the nordic arthroplasty register association database. *Acta orthopaedica*, 85(2):107–116.

[158] Maradit-Kremers, H., Larson, D. R., Noureldin, M., Schleck, C. D., Jiranek, W. A., and Berry, D. J. (2016). Long-term mortality trends following total hip and knee replacement: A population-based study. *The Journal of arthroplasty*, 31(6):1163.

[159] Marston, L., Carpenter, J. R., Walters, K. R., Morris, R. W., Nazareth, I., and Petersen, I. (2010). Issues in multiple imputation of missing data for large general practice clinical databases. *Pharmacoepidemiology and drug safety*, 19(6):618–626.

[160] Martin, N. and Pardo, L. (2009). On the asymptotic distribution of cook's distance in logistic regression models. *Journal of Applied Statistics*, 36(10):1119–1146.

[161] Mason, J. C. and Libby, P. (2014). Cardiovascular disease in patients with chronic inflammation: mechanisms underlying premature cardiovascular events in rheumatologic conditions. *European heart journal*, 36(8):482–489.

[162] Matteo, Q. and James, C. (2017). Multilevel joint modelling multiple imputation using package 'mitml'.

[163] May, S. and Hosmer, D. W. (1998). A simplified method of calculating an overall goodness-of-fit test for the cox proportional hazards model. *Lifetime data analysis*, 4(2):109–120.

[164] McBride, D., Hardoon, S., Walters, K., Gilmour, S., and Raine, R. (2010). Explaining variation in referral from primary to secondary care: cohort study. *BMJ*, 341.

[165] McCarthy, J., Noble, P., and Villar, R. (2016). *Hip Joint Restoration: Worldwide Advances in Arthroscopy, Arthroplasty, Osteotomy and Joint Preservation Surgery*. Springer New York.

[166] McMinn, D., Snell, K., Daniel, J., Treacy, R., Pynsent, P., and Riley, R. (2012). Mortality and implant revision rates of hip arthroplasty in patients with osteoarthritis: registry based cohort study. *Bmj*, 344:e3319.

[167] Memtsoudis, S. G., Pumberger, M., Ma, Y., Chiu, Y.-L., Fritsch, G., Gerner, P., Poultsides, L., and Valle, A. G. D. (2012). Epidemiology and risk factors for perioperative mortality after total hip and knee arthroplasty. *Journal of Orthopaedic Research*, 30(11):1811–1821.

[168] Mohamed, K., Copeland, G., Boot, D., Casserley, H., Shackleford, I., Sherry, P., and Stewart, G. (2002). An assessment of the possum system in orthopaedic surgery. *Bone & Joint Journal*, 84(5):735–739.

[169] Morley, J., Barker, R., and Howell, J. R. (2014). Cemented total hip replacement. In *European Surgical Orthopaedics and Traumatology*, pages 2397–2422. Springer.

[170] Morshed, S., Bozic, K. J., Ries, M. D., Malchau, H., and Colford Jr, J. M. (2007). Comparison of cemented and uncemented fixation in total hip replacement: a meta-analysis. *Acta Orthopaedica*, 78(3):315–326.

[171] Nadar, S. and Lip, G. (2015). *Hypertension*. Oxford Cardiology Library. Oxford University Press.

[172] Nelder, J. A. and Baker, R. J. (1972). *Generalized linear models*. Wiley Online Library.

[173] Nevalainen, J. and Pekkanen, J. (1998). The effect of particulate air pollution on life expectancy. *Science of The Total Environment*, 217(1–2):137 – 141.

[174] Newman, E. and Matzko, C. (2006). *Rheumatoid Arthritis FAQs*. FAQs (Frequently Asked Questions) Series. Decker Publishing.

[175] NHS (2013). Improving General Practice – A Call to Action. `https://www.england.nhs.uk/2014/03/cta-emerging-findings/`.

[176] NHS (2015). High cholesterol.

[177] NHS Digital (2016). Read Codes. `https://digital.nhs.uk/article/1104/Read-Codes`.

[178] NHS Scotland (2002). Primary hip and knee replacements in scotland. *Analysis of 6 years of operations on NHS patients April 1996 - March 2002*.

[179] NICE, G. (2014). Total hip replacement and resurfacing arthroplasty for end-stage arthritis of the hip.

[180] NICE Guidance (2016). Hypertension in adults: diagnosis and management. `https://www.nice.org.uk/guidance/cg127/chapter/1-Guidance#measuring-blood-pressure`. [Online; accessed 15-January-2017].

[181] Nilsdotter, A. and Bremander, A. (2011). Measures of hip function and symptoms: Harris hip score (hhs), hip disability and osteoarthritis outcome score (hoos), oxford hip score (ohs), lequesne index of severity for osteoarthritis of the hip (lisoh), and american academy of orthopedic

surgeons (aaos) hip and knee questionnaire. *Arthritis care & research*, 63(S11):S200–S207.

[182] NJR (2016). 13th annual report for national joint registry for england, wales, northern ireland and the isle of man. NJR Team.

[183] NJR (2017). 14th annual report for national joint registry for england, wales, northern ireland and the isle of man. NJR Team.

[184] Nunley, R. M. and Lachiewicz, P. F. (2003). Mortality after total hip and knee arthroplasty in a medium-volume university practice. *The Journal of arthroplasty*, 18(3):278–285.

[185] Office for National Statistics, U. K. (2004). Rural and urban area classification for output areas.

[186] Office for National Statistics, U. K. (2013). Pension trends chapter 6: Private pensions.

[187] Office for National Statistics, U. K. (2016). Life expectancy at birth and at age 65 by local areas, uk.

[188] Ollier, L. (1888). *Traité des résections et des opérations conservatrices qu'on peut pratiquer sur le système osseux*, volume 1. Masson.

[189] O'Quigley, J. and Pessione, F. (1989). Score tests for homogeneity of regression effect in the proportional hazards model. *Biometrics*, pages 135–144.

[190] Paavolainen, P., Pukkala, E., Pulkkinen, P., and Visuri, T. (2002). Causes of death after total hip arthroplasty: a nationwide cohort study with 24,638 patients. *The Journal of arthroplasty*, 17(3):274–281.

[191] Parsons, D. (1972). *Total hip replacement- Modern Trends in Orthopaedics*. London: Apley A G.

[192] Pedersen, A. B., Baron, J., Overgaard, S., and Johnsen, S. (2011). Short-and long-term mortality following primary total hip replacement for osteoarthritis. *Bone and Joint Journal*, 93(2):172–177.

[193] Peto, R. (1972). Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society. Series A (General)*, pages 185–207.

[194] Pickett, K. E. and Pearl, M. (2001). Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. *Journal of Epidemiology & Community Health*, 55(2):111–122.

[195] Pijls, B., Meessen, J., Schoones, J., Fiocco, M., van der Heide, H., Sedrakyan, A., and Nelissen, R. (2016). Increased mortality in metal-on-metal versus non-metal-on-metal primary total hip arthroplasty at 10 years and longer follow-up: a systematic review and meta-analysis. *PloS one*, 11(6):e0156051.

[196] Platt, R., Madre, L., Reynolds, R., and Tilson, H. (2008). Active drug safety surveillance: a tool to improve public health. *Pharmacoepidemiology and drug safety*, 17(12):1175–1182.

[197] Pope, C. A. I., Ezzati, M., and Dockery, D. W. (2009). Fine-particulate air pollution and life expectancy in the united states. *New England Journal of Medicine*, 360(4):376–386. PMID: 19164188.

[198] QResearch (2016). QResearch General Information. `http://www.qresearch.org/SitePages/Home.aspx`.

[199] Raboud, J. and Breslow, N. E. (1989). Efficiency gains from the addition of controls to matched sets in cohort studies. *Statistics in Medicine*, 8(8):977–985.

[200] Ramiah, R., Ashmore, A., Whitley, E., and Bannister, G. (2007). Ten-year life expectancy after primary total hip replacement. *Bone & Joint Journal*, 89(10):1299–1302.

[201] Reynolds, L. and Tansey, E. (2006). *EARLY DEVELOPMENT OF TOTAL HIP REPLACEMENT*. London: Wellcome Trust Centre.

[202] Richards, S. (2014). *Sweating your data set*. [Online] Available at: https://goo.gl/UR4SnA [Accessed 31.

[203] Ritter, M. A., Albohm, M. J., Keating, E. M., Faris, P. M., and Meding, J. B. (1998). Life expectancy after total hip arthroplasty. *The Journal of arthroplasty*, 13(8):874–875.

[204] Röder, C., Bach, B., Berry, D. J., Eggli, S., Langenhahn, R., and Busato, A. (2010). Obesity, age, sex, diagnosis, and fixation mode differently affect early cup failure in total hip arthroplasty: a matched case-control study of 4420 patients. *JBJS*, 92(10):1954–1963.

[205] Rubin, D. B. (1987). Comment. *Journal of the American Statistical Association*, 82(398):543–546.

[206] Sanders (2003). *Miller-Keane Encyclopedia and Dictionary of Medicine, Nursing, and Allied Health*. Elsevier.

[207] Schafer, J. L. and Yucel, R. M. (2002). Computational strategies for multivariate linear mixed-effects models with missing values. *Journal of computational and Graphical Statistics*, 11(2):437–457.

[208] Schäfer, T., Krummenauer, F., Mettelsiefen, J., Kirschner, S., and Günther, K. (2010). Social, educational, and occupational predictors of total hip replacement outcome. *Osteoarthritis and cartilage*, 18(8):1036–1042.

[209] Schemper, M. and Stare, J. (1996). Explained variation in survival analysis. *Statistics in medicine*, 15(19):1999–2012.

[210] Schoenfeld, D. (1980). Chi-squared goodness-of-fit tests for the proportional hazards regression model. *Biometrika*, 67(1):145–153.

[211] Set, N. D. (2008). *National Joint Registry Annual Report, 2008*. NJR [Accessed online].

[212] Sharma, A., Sarah, L., and Lisa, S. (2010). Insights into social disparities in smoking prevalence using mosaic, a novel measure of socioeconomic status: An analysis using a large primary care dataset. *BMC Public Health*, 10(755).

[213] Sharma, M., Nazareth, I., and Petersen, I. (2016). Trends in incidence, prevalence and prescribing in type 2 diabetes mellitus between 2000 and 2013 in primary care: a retrospective cohort study. *BMJ Open*, 6(1).

[214] Shaw, M. (2005). Health inequality: An introduction to theories, concepts and methods. mel bartley. cambridge: Polity press, 2004, pp. 224, £16.99 (pb), isbn: 0745627803. *International Journal of Epidemiology*, 34(2):500–502.

[215] Shephard, E., Stapley, S., and Hamilton, W. (2011). The use of electronic databases in primary care research.

[216] Siebert, S., Lyall, D. M., Mackay, D. F., Porter, D., McInnes, I. B., Sattar, N., and Pell, J. P. (2016). Characteristics of rheumatoid arthritis and its association with major comorbid conditions: cross-sectional study of 502 649 uk biobank participants. *RMD Open*, 2(1).

[217] Singh, J. A., Jensen, M. R., Harmsen, W. S., Gabriel, S. E., and Lewallen, D. G. (2011). Cardiac and thromboembolic complications and mortality in patients undergoing total hip and total knee arthroplasty. *Annals of the rheumatic diseases*, 70(12):2082–2088.

[218] Smith, E., Maru, M., and Siegmeth, A. (2015). Thirty-day mortality after elective hip and knee arthroplasty. *the surgeon*, 13(1):5–8.

[219] Smith, G. H., Johnson, S., Ballantyne, J. A., Dunstan, E., and Brenkel, I. J. (2012). Predictors of excellent early outcome after total hip arthroplasty. *Journal of orthopaedic surgery and research*, 7(1):13.

[220] Smith-Petersen, M. (1948). The classic: Evolution of mould arthroplasty of the hip joint by M. *N. Smith-Petersen. J*, 134:5–11.

[221] Snijders, T. A. (2011). Multilevel analysis. In *International Encyclopedia of Statistical Science*, pages 879–882. Springer.

[222] Söderman, P. (2000). On the validity of the results from the swedish national total hip arthroplasty register. *Acta Orthopaedica Scandinavica*, 71(sup296):i–33.

[223] Song, J. W. and Chung, K. C. (2010). Observational studies: cohort and case-control studies. *Plastic and reconstructive surgery*, 126(6):2234.

[224] SooHoo, N. F., Farng, E., Lieberman, J. R., Chambers, L., and Zingmond, D. S. (2010). Factors that predict short-term complication rates after total hip arthroplasty. *Clinical Orthopaedics and Related Research®*, 468(9):2363–2371.

[225] Spiegelhalter, D. (2016). How old are you, really? communicating chronic risk through 'effective age'of your body and organs. *BMC medical informatics and decision making*, 16(1):104.

[226] Spratt, M., Carpenter, J., Sterne, J. A., Carlin, J. B., Heron, J., Henderson, J., and Tilling, K. (2010). Strategies for multiple imputation in longitudinal studies. *American journal of epidemiology*, 172(4):478–487.

[227] Stamou, S. C., Nussbaum, M., Stiegel, R. M., Reames, M. K., Skipper, E. R., Robicsek, F., and Lobdell, K. W. (2011). Effect of body mass index on outcomes after cardiac surgery: is there an obesity paradox? *The Annals of thoracic surgery*, 91(1):42–47.

[228] Sudmann, E., Havelin, L., Lunde, O., and Rait (1983). The Charnley versus the Christiansen total hip arthroplasty. *A comparative clinical study.*, 54:545–52.

[229] Szatkowski, L., Lewis, S., McNeill, A., Huang, Y., and Coleman, T. (2012). Can data from primary care medical records be used to monitor national smoking prevalence? *J Epidemiol Community Health*, 66(9):791–795.

[230] Szklo, M. and Nieto, F. (2000). Understanding lack of validity: bias. *Epidemiology Beyond the basics. Maryland: Aspen Publishers Inc*, pages 125–176.

[231] Therneau, T. and Grambsch, P. (2013). *Modeling Survival Data: Extending the Cox Model*. Statistics for Biology and Health. Springer New York.

[232] THIN Data Guide (2011). *THIN Data Guide*. Cegedim Company, UK.

[233] UK, O. (2016). Statistical bulletin:adult smoking habits in the uk: 2016.

[234] Unnanuntana, A., Dimitroulias, A., Bolognesi, M. P., Hwang, K. L., Goodman, S. B., and Marcus, R. E. (2009). Cementless femoral prostheses cost more to implant than cemented femoral prostheses. *Clinical Orthopaedics and Related Research®*, 467(6):1546.

[235] Uren, Z. and Goldring, S. (2007). Migration trends at older ages in england and wales. *Population Trends*, (130):31.

[236] van Buuren, S. (2012). *Flexible Imputation of Missing Data*. Chapman & Hall/CRC Interdisciplinary Statistics. CRC Press.

[237] Vaupel, J. W. (2010). Biodemography of human ageing. *Nature*, 464(7288):536–542.

[238] Vaupel, J. W., Manton, K. G., and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3):439–454.

[239] Visuri, T., Pulkkinen, P., Paavolainen, P., Koskenvuo, M., and Turula, K. B. (1997). Causes of death after hip arthroplasty in primary arthrosis. *The Journal of arthroplasty*, 12(4):397–402.

[240] Visuri, T., Pulkkinen, P., Turula, K. B., Paavolainen, P., and Koskenvuo, M. (1994). Life expectancy after hip arthroplasty: case-control study of 1018 cases of primary arthrosis. *Acta Orthopaedica Scandinavica*, 65(1):9–11.

[241] Vos, T., Barber, R. M., Bell, B., Bertozzi-Villa, A., Biryukov, S., Bolliger, I., Charlson, F., Davis, A., Degenhardt, L., Dicker, D., et al. (2015). Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the global burden of disease study 2013. *The Lancet*, 386(9995):743.

[242] Walker, P. (1977). *Human Joints and Their Artificial Replacements*. s. l, :Springfield, IL.

[243] Walley, T. and Mantgani, A. (1997). The uk general practice research database. *The Lancet*, 350(9084):1097–1099.

[244] Watson, D. J., Rhodes, T., and Guess, H. A. (2003). All-cause mortality and vascular events among patients with rheumatoid arthritis, osteoarthritis, or no arthritis in the uk general practice research database. *The Journal of Rheumatology*, 30(6):1196–1202.

[245] Weissfeld, L. A. (1990). Influence diagnostics for the proportional hazards model. *Statistics & probability letters*, 10(5):411–417.

[246] Whitehouse, S. L., Bolland, B. J., Howell, J. R., Crawford, R. W., and Timperley, A. J. (2014). Mortality following hip arthroplasty—inappropriate use of national joint registry (njr) data. *The Journal of arthroplasty*, 29(9):1827–1834.

[247] WHO (2017). Who criteria for diagnosis of osteoporosis.

[248] Wijlaars, L. (2013). Thin database. `http://www.ucl.ac.uk/pcph/research-groups-themes/thin-pub/database`.

[249] Wild, S. and McKeigue, P. (1997). Cross sectional analysis of mortality by country of birth in england and wales, 1970-92. *BMJ : British Medical Journal*.

[250] Williams, N., Bhatnagar, J., Wickramasinghe, P., and Rayner, K. (2014). Cardiovascular disease statistics. *British Heart Foundation*.

[251] Williams, O., Fitzpatrick, R., Hajat, S., Reeves, B. C., Stimpson, A., Morris, R. W., Murray, D. W., Rigge, M., and Gregg, P. J. (2002). Mortality, morbidity, and 1-year outcomes of primary elective total hip arthroplasty. *The Journal of arthroplasty*, 17(2):165–171.

[252] Wolfe, F. (1999). Determinants of womac function, pain and stiffness scores: evidence for the role of low back pain, symptom counts, fatigue and depression in osteoarthritis, rheumatoid arthritis and fibromyalgia. *Rheumatology (Oxford, England)*, 38(4):355–361.

[253] Wylde, V., Learmonth, I. D., and Cavendish, V. J. (2005). The oxford hip score: the patient's perspective. *Health and quality of life outcomes*, 3(1):66.

[254] Xu, H. F., White, R. S., Sastow, D. L., Andreae, M. H., Gaber-Baylis, L. K., and Turnbull, Z. A. (2017). Medicaid insurance as primary payer predicts increased mortality after total hip replacement in the state inpatient databases of california, florida and new york. *Journal of Clinical Anesthesia*, 43:24–32.

[255] Yamada, H., Yoshihara, Y., Henmi, O., Morita, M., Shiromoto, Y., Kawano, T., Kanaji, A., Ando, K., Nakagawa, M., Kosaki, N., et al. (2009). Cementless total hip replacement: past, present, and future. *Journal of Orthopaedic Science*, 14(2):228–241.