Hormone Replacement Therapy and its Long-term Impact on the Survival of Women in the United Kingdom

Nurunnahar Akter

Doctor of Philosophy

University of East Anglia

School of Computing Sciences

December 2022

Abstract

Hormone replacement therapy (HRT) is an effective treatment for relieving symptoms of menopause. However, because of some adverse health consequences, including an increased risk of breast cancer, many symptomatic women are cautious about using HRT. Untreated menopausal symptoms may deteriorate quality of life, increase the risk of developing other health problems, and place additional pressure on the healthcare system. Past studies were mostly based on survey or register data, whereas data from routine primary care may provide greater insights about the effects of HRT in the general population. While previous studies mainly investigated the impact of HRT on morbidities, all-cause mortality may summarise the net benefits and risks.

This study investigated the long-term hazards of all-cause mortality associated with estrogen-only and combined (combination of estrogen and progesterone) HRT using a large electronic primary care records from The Health Improvement Network database. 105,199 HRT users who started the treatment at ages 46 to 65 and 224,643 matched non-users were selected for survival modelling. The hazards of all-cause mortality associated with HRT were estimated by a newly developed Weibull-Double-Cox model adjusting for important medical, lifestyle, and socio-demographic factors. Multilevel multiple imputation techniques were used to deal with missing data.

The length of study follow-up was up to 32 years (1984−2017), with an average follow-up per participants was almost 14 years. During study follow-up, a total of 21,751 women died, of whom 6,329 were HRT users, and 15,422 non-users. This research found that estrogen only HRT has no long-term impact on mortality at any age, but combined HRT reduces the hazards of death from all-causes. Furthermore, starting combined HRT between the ages of 51 to 55 reduces the hazards of mortality the most. The findings of this study may help women in making an informed choice, and further educating the clinicians and resource planners.

Contents

List of Tables

List of Figures

Research Dissemination

Publication

Akter, N., Kulinskaya, E., Steel, N. and Bakbergenuly, I. (2022): The effect of hormone replacement therapy on the survival of UK women: a retrospective cohort study 1984−2017, *BJOG: An International Journal of Obstetrics and Gynaecology*, 129(6), 994-1003.

Conference presentations

- ICCB 2021: 15th International Conference on Clinical Biostatistics, 18-19 November 2021, London, UK (This research was recognised as outstanding by the conference committee and has been awarded the best presentation certificate).

- Actuarial Teachers and Researchers Conference, 14-15 June 2021, University College Dublin, Ireland (Virtual).

- ARC Webinar Series on Use of Primary Health Care Records Data in Actuarial Research, 9 March 2021, London (Virtual).

- Online International Conference (OICA) in Actuarial Science, Data Science, and Finance, 28-29 April 2020, University of Lyon, France.

- Technical Workshop on Statistical methods for assessment of the impact of medical advances and health interventions on longevity and population projections. Actuarial Research Centre Event, 30 October 2019, Staple Inn Hall, Holborn, London.

- 42nd Research Students Conference (RSC) in Probability and Statistics (Sponsored by Alan Turing Institute), 18-21 June 2019, University of Exeter, UK.

- Protection, Health, and Care Conference, Institute and Faculty of Actuaries, 6 June 2019, UK.

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisor, Professor Elena Kulinskaya for her continuous support, guidance, and assistance throughout this PhD. It would not have been possible to complete this thesis without her constant engagement, encouragement, and valuable time. I would also like to thank Professor Nicholas Steel for his guidance and advice on presenting and explaining my research to medical audiences in a clear and concise manner.

I would like to thank my husband Dr Zillur Rahman Shabuz for sharing his statistical knowledge and proofreading statistical methodology chapter of my thesis. I would like to thank Dr Ilyas Bakbergenuly for assisting me with data extraction, and Dr Lisanne Andra Gitsels for sharing her knowledge in various topics. I would also like to thank my project colleagues Padma and Njabulo for their time and sharing knowledge during PhD.

I am grateful to my parents for their continuous mental support and inspiration to help me to continue my PhD. I would also like to express my gratitude to my son Nehan Sadit for his understanding and patience throughout my PhD.

Finally, I would like to thank the Institute and Faculty of Actuaries (IFoA) for funding this research. I am thankful to the Charles Wallace Bangladesh Trust and the UEA Student Services Fund for awarding a small grant during the thesis writing period of my PhD.

Chapter 1

Introduction

This chapter first presents the background of hormone replacement therapy (HRT) and then it explains the rationale for developing survival models to estimate the impact of HRT on longevity and morbidity in women population at postmenopausal age. After that, the aims and objectives of this study are listed. Next, the contributions of this study to HRT research are discussed. Finally, an outline of this thesis is provided.

1.1   Background

Most women experience troubling menopausal symptoms as they approach menopause due to a sharp fall in female sex hormones estrogen/oestrogen and progesterone. Postmenopausal women are more likely to develop life-threatening

conditions such as osteoporosis, cardiovascular disease, and neurodegenerative disorders as a result of these hormone deficiencies (Christiansen, 1996; Kawecka-Jaszcz et al., 2002; Zárate et al., 2017). Hormone replacement therapy has been used as an effective treatment in ameliorating menopausal symptoms since its initiation in 1960s (National Institute for Health and Care Excellence, 2019; Cagnacci and Venier, 2019). Other known benefits of HRT include slowed bone loss, reduced cardiovascular disease, and improved quality of life after menopause (Grady et al., 1992; Grodstein et al., 1994; Folsom et al., 1995; Persson et al., 1996; Pentti et al., 2006). However, the actual risks and benefits of HRT have long been disputed, owing to inconsistencies in results between observational studies and randomised controlled trials, and more specifically, to the possible increased risks of breast cancer. Although numerous studies have been conducted in the past to investigate the effects of HRT, the majority of them focused on either morbidity (Grady et al., 1995; Hodis et al., 2003; Margolis et al., 2004; Beral et al., 2007) or cause-specific mortality (Persson et al., 1996; Schuetz et al., 2007; Jang et al., 2019). The impact of HRT on the hazards of all-cause mortality summarises the net measurement of the risks and benefits, but this has received far less attention in previous studies.

The main aims of this study were to estimate the effect of estrogen-only and combined HRT on the hazards of overall and age-specific all-cause mortality in women using routinely collected primary care data in the United Kingdom (UK).

1.2 Rationale

Each year, approximately 1.5 million women in the UK suffer from moderate to severe menopausal symptoms. Currently, there are around 13 million women who are either in peri-menopausal or menopausal transition age, accounting for one third of the entire UK female population (Menopause support UK, 2021). Although HRT is an effective treatment for soothing menopausal symptoms and is also protective against some other chronic medical conditions, due to the uncertainty of its impact, many symptomatic women are left untreated. Untreated menopausal symptoms increase the risks of developing other health conditions and eventually put an additional burden on the resources of the health care sector, budget, and time. Postmenopausal women are not only at high risk of developing osteoporosis, cardiovascular disease, and neurological disease, these conditions are also the leading cause of death in women in the UK (Climént-Palmer and Spiegelhalter, 2019). According to the International Longevity Centre UK (ILCUK, 2010), the total cost for hospital stays to the National Health Services (NHS) for osteoporotic women is more than £400 million per year. Deaths from osteoporotic fractures in women is approximately 6,000 each year in the UK (ILCUK, 2010). The burden of neurodegenerative and cardiovascular disease is also considerable and it is increasing over time (ONS UK, 2019).

Longevity prospects in a population can be projected based on their co-morbidities, treatments histories, lifestyle choices, and socio-economic status.

Survival models are useful for estimating the impact of medical conditions, treatments or other risk factors on mortality and morbidity. A well developed survival model, which ideally consists of a wide range of risk factors that influence the variation in survival, may accurately forecast longevity. A more precise estimates of survival and a better understanding of its variation can help to allocate resources in a strategic way among patients, hospitals, and medical practices. Health professionals also benefit from the survival modelling because it can help to tailor the treatments according to the individual patient's needs, and individuals can gain knowledge about how certain medical conditions, treatments, and lifestyle choices may affect their long-term survival prospects.

Accurate estimates of longevity prospects and understanding their variations are also important for many other organisations, such as insurance companies, government, and pension fund providers. Actuaries are interested in survival models because they provide insights on survival variations and hence can be used for calculating the present value of the annuity of a pension fund, or pricing of the insurance products. The life insurance companies have to continue providing a payment to their clients until their death, and thereby it is of great importance to them to estimate the life expectancy as accurately as possible. If life expectancy is overestimated, clients pay lower insurance premiums for life insurance, causing the insurers to lose money. On the other hand, if life expectancy is underestimated, clients pay more for life insurance, and the insurers gain profits. A well-developed survival model can provide more insight about the

factors affecting longevity, and these factors should be considered when pricing annuities and other insurance products in order to set better annuity prices.

Survival models are also important for the government because they can assist in making informed decisions regarding taxation, state pensions, and national insurance rates. Life expectancy may rise due to the advances in treatments and healthy lifestyle choices, or it may fall over time due to disease outbreaks. In the UK, the average life expectancy for both men and women has been projected to increase (Office for National Statistics, 2018), meaning an increased dependency of the pensioners on the workforce. This increased reliance places a strain on the government's budget for welfare spending while collecting less revenue. By identifying age-specific risk factors for health and mortality, the government can predict the expected age and duration of retirement, which in turn can inform the expected participation in the workforce at each age, the reasonable age of retirement for the population, as well as the expected dependency by the retired. It is, therefore, crucial for the government to identify the accurate survival trends in the general population and to understand the survival variation in order to sustain the economy. A well developed and up-to-date survival model can provide information about the risk factors that influence mortality direction over time, allowing for better allocation of government funds and, as a result reducing pressure on welfare spending, increasing tax revenues, and boosting pension savings.

Previous published survival models estimated the impact of HRT on the

risks of mortality or morbidity, mostly using self-reported survey data. Randomised control trials of longevity are seldom conducted as it involves long time and high expenses. Furthermore, clinical trials are often restricted to a relatively small number of participants. On the other hand, the increasing volume of electronic health data from UK primary care now allows for long-term follow-up of thousands of patients and investigation of a wide range of risk factors. Mortality registration is more up to date in primary care as general practitioners (GPs) are informed if their patients died. These digital records are becoming an invaluable resource for researchers, enabling them to assess the overall risks of mortality or morbidity in various populations. Although the number of studies using electronic primary care records is increasing (Springate et al., 2014), to date, no other published study has modelled the risks of all-cause mortality associated with HRT using UK primary care data, and therefore there is a need to modelling the impact of HRT using a large scale primary care data.

1.3 Existing survival models of HRT

Numerous survival models have been developed in the past to investigate the risks and benefits of hormone replacement therapy (NICE, 2021). Majority of these models estimated the risks of morbidity and cause-specific mortality associated with HRT (Brenner et al., 1994; Christiansen, 1996; Ettinger et al., 1996; Cass and Runowicz, 1998; Zandi et al., 2002; Hodis et al., 2003; Schuetz et al., 2007;

on the Evaluation of Carcinogenic Risks to Humans et al., 2007; Freeman and Sherif, 2007; Schierbeck et al., 2012), and only a small number of studies focused on the risks of overall mortality (Hunt et al., 1990; Folsom et al., 1995; Manson et al., 2017; Malek et al., 2019). However, there were differences in study design, sample sizes, data collection method, and model development procedure in these studies. In this section, a brief overview of previous HRT models is provided, and some research gaps that were identified from those models are mentioned. These studies and their findings will be discussed in greater detail in Chapter 2.

HRT was first made available in 1960s in the form of estrogen-only therapy, and it was first prescribed in the UK in 1965 (Women's Health Concern, 2017). In 1966, a best-selling book by Robert A. Wilson inspired many menopausal and postmenopausal women in western countries to receive estrogen therapy in order to retain their femininity and increase life expectancy (Wilson, 1966). In the 1970s, two small observational studies discovered that taking estrogen-only HRT increased the risk of endometrial cancer (cancer of the lining of the uterus), and the authors recommended that women with an intact uterus should receive combined hormone therapy (Ziel and Finkle, 1975; Smith et al., 1975), which was later supported by a number of larger studies (Persson et al., 1996; Collaborators et al., 2005). Since then combined HRT was offered to women with a uterus, and estrogen-only HRT to women without a uterus.

The Heart and Estrogen/Progestin Replacement Study (HERS) was the first randomised, double blind, placebo-controlled trial that evaluated the im-

pact of combined HRT on prevention of recurring coronary heart disease (CHD) in American postmenopausal women (Grady et al., 1998), and found a reduced CHD mortality in HRT users. In 1993, the Women's Health Initiative (WHI) set up a randomised, placebo-controlled trial in the United States (US) to examine the risks and benefits of both estrogen-only and combined HRT among postmenopausal women (Women's Health Concern, 2017). The Million Women Study (MWS), a large observational study, was set up in the UK in 1996 to investigate the effects of HRT on breast cancer incidence and mortality. More than a million women in the UK had signed up for the MWS (Emily et al., 2003). A publication from the WHI trial in 2002 reported an increased incidence of breast cancer in combined HRT users (Rossouw et al., 2002). In 2003, MWS also found increased risks of breast cancer in both estrogen-only and combined HRT users (Emily et al., 2003). The findings of these two major studies raised concerns about the safety of HRT use, and as a result of this, the number of HRT users dropped from two million to less than one million in the UK between 2002−2007 (Women's Health Concern, 2017). However, a recent analysis of WHI trial found no association of HRT with the risks of overall, cardiovascular, or total cancer mortality during a cumulative follow-up of 18 years (Manson et al., 2017).

A Bayesian meta-analysis of 19 randomised controlled trials and 8 observational studies on HRT and overall mortality in 228,171 postmenopausal women of mean age 55 found a 28% lower risk of mortality in HRT users (Salpeter et al., 2009). Two separate meta-analyses on HRT and all-cause mortality performed

by Boardman et al. (2015) and Benkhadra et al. (2015), and also Manson et al. (2017) study from the WHI trial found no impact of HRT on overall mortality. On the other hand, Malek et al. (2019) reported a 31% increased risk of death from all-cause in postmenopausal US women who took HRT but experienced menopause at the age of 45 years on average.

The effect of HRT on survival prospects was likely to be wrongly estimated in past studies as majority of them did not account for other risk factors. Existing clinical conditions have a strong influence on mortality and morbidity, so adjusting for them in the model is important for obtaining a more accurate estimate of the effect size. Some studies may have introduced a bias in favour of HRT users due to the inclusion of healthy users in comparison to non-users (Folsom et al., 2004; Grodstein et al., 1997). Only two studies have investigated the individual effects of estrogen-only and combined formulations on the risk of all-cause mortality (Stram et al., 2011; Manson et al., 2017): one found a lower risk of mortality in combined HRT users, while the other found no association with both formulations. However, Manson et al. (2017) results from the WHI trial may not be generalisable to all users as each trial in WHI evaluated a single dose, formulation, and route of administration of HRT. Other limitations of previous research include a lack of age-specific information on the use of HRT and its long-term impact on all-cause mortality (Grodstein et al., 1999; Pentti et al., 2006; Schierbeck et al., 2012). Furthermore, little information about missing data and the time-varying hazards was provided in past studies.

This study aims to fill some of the research gaps mentioned above by developing survival models using routinely collected primary care data. The goals and objectives of this HRT research are outlined in the next section.

1.4 Research aims and objectives

The primary goals of this study are to investigate the effects of estrogen-only and combined HRT on the hazards of all-cause mortality in UK women while adjusting for various important risk factors, and also to analyse how age at starting HRT affects mortality. The Health Improvement Network (THIN) database from UK primary care is used to develop survival models. The research focuses on survival prospects of women who started HRT for the first time at the age of 46 years or older, in comparison to their matched non-users of HRT. The main objectives are to develop population-based survival models addressing the following goals:

1. Review existing HRT models for longevity and morbidity to identify research gaps and key risk factors that affect longevity and morbidity.

2. Investigate the overall survival benefits or risks associated with different types of HRT while adjusting for other important risk factors.

3. Investigate the effects of initiating HRT at different ages on survival while controlling for other risk variables.

4. Estimate the overall effect and the effect of starting age of HRT on the hazards of death from all-causes using complete data (patients with complete records only), as well as full data (all patients including those with missing records) using multiple imputation techniques.

5. Investigate how health, lifestyle choices, socioeconomic factors and their interactions, and also clustering by general practice affect mortality.

6. Investigate the risks of developing various chronic medical conditions in the presence or absence of HRT treatment during follow-up.

7. Develop survival models to calculate life expectancy and the differences in life expectancy between subpopulations with various characteristics.

1.5 Research contribution

This study contributes to existing HRT research by developing survival models that estimate the effect of HRT on women's survival as well as the variation in survival due to the influence of various risk factors, and estimate the effect of starting HRT at different ages. This research also contributes to actuarial science by developing a model for calculating women's life expectancy of postmenopausal women.

The newly developed survival models address both the issue of underestimating the effects of HRT in incidence study designs due to the absence of

controls, as well as address the issue of overestimating the effects of HRT in prevalence study designs due to the adjustment by a limited number of risk factors. This study developed well fitted survival models using electronic primary care data that have a long follow-up and retain information on a wide range of risk factors for both cases and controls. Thus this study was able to adjust for a variety of important risk variables such as comorbidities, treatments, lifestyle and socio-economic factors, and also to examine the interaction effects within and between all types of risk factors. As a result, the newly developed models explored the survival variations in greater detail.

The primary care database used for this study is representative of the UK population (Blak et al., 2011). Thus the newly developed survival models are more generalisable than previous models that used survey or clinical trial data, which in turn make these models more applicable in the clinical and actuarial fields. This study will help women and their physicians in making an informed decision about whether to start or continue HRT. This study also will help the actuaries and government bodies with annuity pricing, and financial planning.

The findings of this study were published in the peer-reviewed medical journal, BJOG: An International Journal of Obstetrics and Gynaecology (Akter et al., 2022).

1.6 Thesis outline

This section provides outline of the chapters in the thesis.

Chapter 2 starts with a brief introduction on HRT, including its use and classification. Then a review of existing HRT research is presented. The first part of literature review discusses the major studies that investigated the impact of HRT, and then it presents the findings from major studies on mortality and morbidity.

Chapter 3 begins by introducing different types of clinical data and explaining the importance of primary care data in building survival models. Then it describes the structure of The Health Improvement Network database, and the patient selection procedures from the database.

Chapter 4 explains the statistical methodologies for survival analyses. First, it describes the various types of model assumptions used in survival analyses. After that, the processes for parameter estimation and model selection in a semi-parametric Cox regression model are described. The model evaluation and validation procedures are then discussed. A Weibull-Double-Cox model that was used to handle the non-proportionality of the hazards in the Cox model is introduced. Finally, the methods for dealing with missing data are described.

Chapter 5 presents the distribution and characteristics of the study population extracted from the THIN database. The prevalence of the baseline con-

ditions in the extracted data and its comparison to the prevalence in the UK general population are discussed in this chapter.

Chapter 6 presents the model development procedures to estimate the adjusted hazards of all-cause mortality associated with estrogen-only and combined HRT in the entire age group and the subgroups by first HRT treatment. The goodness-of-fit of the models is also assessed and validated.

Chapter 7 describes the models developed to estimate the probability and the hazards of developing various medical conditions during follow up in the presence and absence of HRT. First, it presents the prevalence of the medical conditions that the study participants were diagnosed with at follow-up, then it presents the Kaplan-Meier survival analysis for estimating survival probabilities in patients by their HRT treatment status. Then univariate Cox proportional models are presented that are fitted to estimate the hazards of developing selected medical conditions at follow-up.

Chapter 8 describes a model developed for calculating patient's life expectancy. First, the model implementation procedures are explained and the results from the model are presented. Then, the calculation of life expectancy based on the model parameter estimates are explained. Finally, some scenario-based life expectancies for women at postmenopausal ages are presented.

Chapter 9 summarises the findings of this study, its strengths and limitations, and draws a conclusion of this thesis.

Chapter 2

Review of hormone replacement therapy

Chapter 1 provides the background of HRT and explains the rationale for developing HRT models to estimate women's longevity after menopause, and the aims and objectives of this study. This chapter first introduces HRT treatments for menopausal symptoms and then it presents research findings from previous studies. Sections 2.1 and 2.2 discuss menopause and its symptoms, as well as diagnosis and treatments for menopausal symptoms. Section 2.3 describes the design and settings of some major HRT studies that investigated its risks and benefits. Results from these major studies as well as some other large studies are described in Section 2.4 and 2.5, respectively.

2.1 Menopause, its symptoms and treatments

Menopause is a natural process in women's bodies that occurs when the menstruation cycle stops, and as a result women are no longer able to become pregnant. Usually, it happens to healthy women at the age between 50 to 55 years, but it can occur early due to premature ovarian insufficiency, surgical bilateral oophorectomy, chemotherapy or many other reasons. The average age for women to reach menopause in the UK is 51 years (National Institute for Health and Care Excellence, 2019). There are three stages of menopause: perimenopause/pre-menopause, menopause, and post-menopause. In perimenopause stage, the menstruation cycle starts to become irregular, and after several consecutive months of irregular periods, it completely stops and eventually women reach menopausal stage. Most women going through menopause suffer from various menopausal symptoms. Some of the symptoms can be severe and have significant impact on daily activities. In the medical literature, menopause is also known as climacteric.

Women going through menopause experience a wide range of symptoms, which may vary greatly from woman to woman. These symptoms usually start from the perimenopausal stage and can persist up to postmenopausal stage. Common symptoms include hot flashes, night sweats, mood swings, sleep disturbances, vaginal dryness, lack of libido, headache, palpitation, and urinary incontinence. Among these, hot flashes and night sweating are commonly faced

by most women, and these symptoms are collectively called vasomotor symptoms. Around 80% of women in the western countries are affected by these menopause-related symptoms (Freeman and Sherif, 2007).

2.1.1 Diagnosis of menopause

Usually, the signs and symptoms of menopause are sufficient to tell women that they have reached menopause. In some cases, doctors may recommend a blood test if women are in their 40s, and are suffering from menopausal symptoms. The blood test measures the levels of Follicle Stimulating Hormones (FSH) and estrogen (Mayo Clinic, 2018). FSH level rises and estrogen level falls in menopausal women. However, these tests are tests not to be offered to women who are taking contraceptives containing estrogen and/or progestogen because the contraceptives change natural FSH and estrogen levels in the blood (GP Notebook, 2016). Vaginal pH level test is another effective method to confirm the menopause. The pH level is around 4.5 in the vagina during the reproductive years. In menopausal stage, it can rises to 6.0 (Healthline, 2018).

2.1.2 Treatment for menopausal symptoms

There are a number of hormonal and non-hormonal treatments available to relieve the distressing menopausal symptoms. Among these treatments, hormone replacement therapy is the most common and effective way to treat menopausal

symptoms and for reducing the long term impact of menopause. Some complementary and alternative therapies, such as herbal remedies are also available to treat menopausal symptoms, but there is not enough scientific evidence to support these treatments.

2.2 Hormone replacement therapy in menopause

Hormone replacement therapy (HRT) is widely acceptable treatment for menopausal symptoms. Different types, routes, forms, and preparations of HRT are available to treat these symptoms, and the type of treatment depends on patients' medical conditions and severity of the symptoms. HRT is also known as menopausal hormone therapy (MHT), estrogen therapy (ET), estrogen replacement therapy (ERT), and hormone therapy (HT).

2.2.1 History of HRT

HRT has been used for more than sixty years to treat menopausal hormone deficiency in women. Although HRT was first introduced in the 1940s, it was widely available for treatment from the 1960s (Panay et al., 2013). Estrogen replacement therapy was first available on the market under the brand name Premarin, which was extracted from the urine of pregnant mares (on the Evaluation of Carcinogenic Risks to Humans et al., 2007). At that time, another form

of female hormone called bio-identical hormone was also available for HRT treatment. These bio-identical hormones are synthesised from soya bean or yam in laboratories and their molecular structures are the same as hormones produced in women's bodies (Cirigliano, 2007). HRT was first used in the United Kingdom in 1965 (Women's Health Concern, 2017). After raising concerns about the use of only oestrogen, which has been linked to endometrial hyperplasia, it was suggested that progesterone be included in HRT (Burkman et al., 2001). The combination of estrogen and progesterone therapy was offered to those women who had not have surgical hysterectomy after this finding. These days there is a wide selection of estrogen and progesterone hormone therapy available. The hormones come from both natural extraction and in synthesised form, and are marketed under various brand names.

2.2.2 Types and routes of administration of HRT

HRT is available in a variety of forms. These can be taken orally as tablets or non-orally like transdermal patches, implants, and by many other alternative ways, including gel, cream, vaginal ring, and injection. HRT comes with estrogen-only or combination of both estrogen and progesterone hormones. Combination of estrogen and progesterone hormones in HRT is known as combined therapy or combined HRT. Types of drugs used in estrogen-only therapy include estradiol, estrone, estriol, 17-β-estradiol, and tibolone. Among these, 17-β-estradiol, estrone, and estriol are the bioidentical form of the hormone estrogen. The

progestogen is a synthetic version of the hormone progesterone and drug types include dydrogesterone, medroxyprogesterone, norethisterone and levonorgestrel (NHS Choices, 2016). Micronized progesterone is a bioidentical form of the hormone progesterone (Cirigliano, 2007). Estrogen-only HRT is used for women who have had a hysterectomy. Combination therapy is used for women with intact uterus because progesterone safeguards the uterus from endometrial cancer. Estrogen and progesterone are marketed under a large number of brand names in different parts of the world. Examples of major brand names in which estrogen has been marketed include Climara, Climen, Dermestril, Divigel, Estrace, Natifa, Estraderm, Estraderm TTS, Estradot, Estreva, Estrimax, Estring, Estrofem, Estrogel, Evorel, Fem7 (or FemSeven), Menorest, Oesclim, Oestrogel, Sandrena, Systen, and Vagifem. For progesterone, the major brand names are Prometrium and Provera. Combined HRT is available under the brand names Activelle, Angelic, Cliane, Femhrt, Prefest, Prempro, Climara pro, Combipatch, Estalis, Eviana, Evorel Conti, Evorel Sequi, Kliogest, Novofem, Sequidot, and Trisequens (FDA, 2018).

Figure 2.1 shows a flow chart of the different routes for the administration of hormone replacement therapy. Table 2.1, lists different classes of HRT, their trade names, and the other drugs available in the same class. Table 2.2 presents the brand and generic names of different drugs class along with their routes of administration.

20
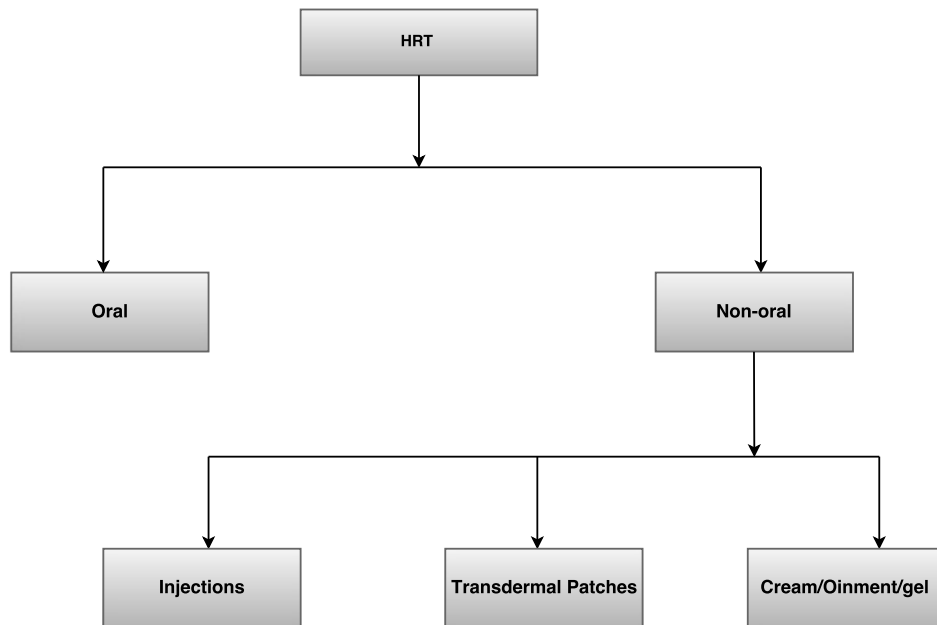
Figure 2.1: Different routes of administration of Hormone Replacement Therapy

Table 2.1: Different classes of drugs used for hormone replacement therapy

Drug class	Trade name	Other drugs in the same class
Estrogen/Oestrogen	Conjugated Equine Estrogen (CEE) Conjugated Estrogen (CE) Premarin	Estradiol/Oestradiol Estriol Estrone Tibolone $17\text{-}\beta\text{-estradiol}$
Progesterone	Progestogen/Progestin	Dydrogesterone Medroxyprogesterone Norethisterone Levonorgestrel Micronized progesterone

Table 2.2: Routes of administration of different classes of drugs used for HRT and their brand names

Routes of administration	Generic name	Brand name
Oral tablets	**Estrogen-only:**	
	Estradiol	Estrofem
	Estradoil acetate	Femtrace
	Estradoil valerate	Progynova
	Micronized estradiol	Estrace
	Progesterone-only:	
	Micronized progesterone	Prometrium
	Medroxyprogesterone acetate	Provera
	Combination of both:	
	Estradiol/Norethindrone acetate	Activella
	Estradiol/Drospirenone	Angeliq
	Ethinyl estradiol/ Norethindrone acetate	Femhrt
	Estradiol/Norgestimate	Prefest
	Conjugated estrogen/ Medroxyprogesterone	Prempro
Transdermal patches	**Estrogen-only:**	Alora
	Estradiol	Climara
		Minivelle
		Vivelle
		Vivelle-Dot
		Menostar
		Estraderm
	Combination of both:	
	Estradiol/Levonrgestrel	Climara Pro
	Estradiol/Norethindrone acetate	Combipatch
Topical gels and oinments	**Estrogen-only:**	Estrogel
	Estradiol	Estrasorb
	Estropipate	Rontagel
	Conjugated estrogens	Divigel
		Elestrin
Nasal sprays	**Estrogen-only:**	Aerodiol
	Estradiol	
Injections	**Estrogen-only:**	
	Estradiol valerate	Delestrogen
	Estradiol undecylate	Progynon Depot

Table 2.2 – *Continued from previous page*

Routes of administration	Generic name	Brand name
	Polyestradiol phosphate	Estradurin
	Estradiol cypionate	Depo-Estradiol
Vaginal ring	**Estrogen-only:**	
	Estradoil	Estring
	Estradoil acetate	Femring

In the following sections, some major HRT studies, their design, and findings from these studies as well as findings from other larger studies are discussed.

2.3 Major studies on HRT

In this section, three large studies on HRT: the Women's Health Initiative Study, the Million Women Study, and the Nurses' Health Study that were highlighted in most HRT related research are described. The results of these studies will be described in later sections of this chapter.

2.3.1 The Women's Health Initiative Study

The Women's Health Initiative (WHI) study, a large and long term national health study in the US was primarily designed to assess some of the most common causes of mortality and morbidity including cancer, cardiovascular disease, and osteoporotic fractures among American postmenopausal women. The WHI was set up in 1993, and enrolled 161,808 healthy postmenopausal women of age

50 to 79 into three randomised control trials (RCTs) and an observational study (OS) at 40 United States clinical centres (WHI Organization). RCTs consisted of three separate arms: The Hormone Therapy, Calcium/Vitamin D, and Dietary Modification trial (Anderson et al., 1998). In Hormone Therapy trial, WHI enrolled 27,347 women (mean age 63 years) between 1993 and 1998 to investigate the risks and benefits of receiving conjugate equine estrogen (CEE), and CEE plus medroxyprogesterone acetate (MPA) therapy on coronary heart disease (CHD), bone fractures and breast cancer. The treatment group was provided daily oral CEE (0.625 mg)-alone to hysterectomised women, and combination of CEE (0.625 mg) with MPA (2.5 mg) to women with intact uterus. In Calcium/Vitamin D and Dietary Modification trial, WHI enrolled 36,282 and 48,835 participants, respectively. There were 68,132 women in total who participated in randomised control trials. The observational study consisted of 93,676 postmenopausal women in total. All participants from the Calcium/Vitamin D trial also participated in the Dietary Modification trial. 8,050 women participated in both the Hormone Therapy and Dietary Modification trials.

At the end of the initial study period in 2005, WHI Extension Studies (2005-2010, 2010-2020) continued to follow-up all women who consented (WHI Organization). WHI studies are considered as the first large, double-blinded, and placebo-controlled clinical trials on HRT in healthy postmenopausal women. Due to perceived increased risk of breast cancer, the CEE plus MPA trial was stopped prematurely (after 5.6 years) in 2002 (Anderson et al., 2004). The CEE-alone
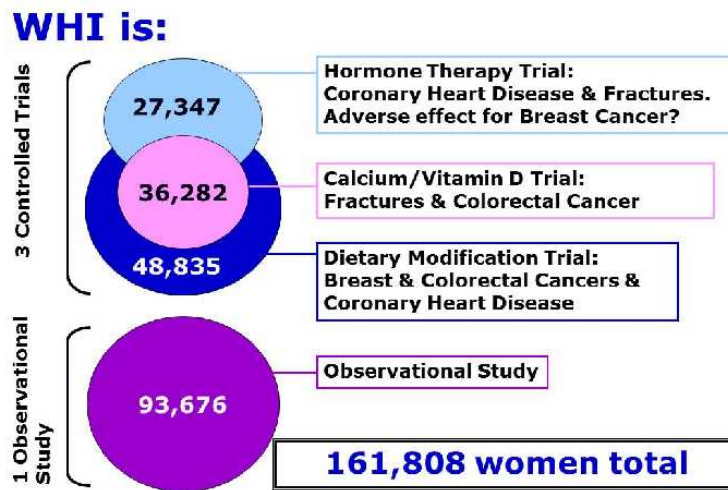
Figure 2.2: Distribution of participants in Women's Health Initiative clinical trial. Figure is reproduced from Women's Health Initiative (2018)

trial also stopped (after 7.2 years) in 2004 due to a perceived increased risk of stroke (Manson et al., 2013). However, the post-intervension follow-up was continued. Several articles have been published on the WHI Hormone Therapy trial (Manson and Martin, 2001; Manson et al., 2013, 2017). Results from these reports will be discussed in the following sections of this chapter. Figure 2.2 shows the distribution of the participants in different trial groups of the WHI.

2.3.2 The Million Women Study

The Million Women Study (MWS), a multi-centre, population-based prospective cohort study was set up in the UK between 1996 and 2001 to investigate the effects of HRT on women's health and specifically, on the incident and fatal breast cancer. It was a collaborative project between Cancer Research UK and

25

the National Health Service (NHS). The MWS invited women aged 50 to 64 years to attend one of 66 NHS Breast Screening Centres, and to participate in the study (Emily et al., 2003). Attendees at these centres were given a study questionnaire, which they were asked to complete and return at the time of breast screening. The questionnaire asked for information about women's sociodemographic status, HRT use, and menstrual history. Around 70% of those attending the programme returned the questionnaires and agreed to take part in the study. Over 1.3 million women enrolled in the study during the study period. Approximately one in every four women in that age group in the UK participated in the study, and this made MWS the world's largest observational study of its kind. A number of reports from the MWS study on HRT and the risks of endometrial, ovarian, and breast cancer had been published (Emily et al., 2003; Collaborators et al., 2005; Beral et al., 2007). These results are discussed in Section 2.5.

2.3.3 The Nurses' Health Study

The Nurses Health Study (TNHS) was a large and long-term prospective cohort study in the USA that began in 1976. Its goal was to investigate the association between diet, smoking, physical activity levels, obesity, oral contraceptive use, hormone therapy, endogenous hormones, dietary factors, and other behaviours and various chronic diseases. The study was divided into three cohorts: TNHS original cohort, TNHS II, and TNHS 3. It recruited registered nurses of ages $30-55$ years from across the different states of USA to respond to a set of baseline

questionnaires (Grodstein et al., 1996). Follow-up questionnaires were sent biennially to update the records of risk factors. The TNHS original cohort started in 1976 and it consisted of 121,700 married women. TNHS II cohort began in 1989 with 116,430 single and married women of ages $25-42$. TNHS 3 cohort started in 2010 and it adds licensed practice nurses and licensed vocational nurses to TNHS II cohort. The enrolment of TNHS 3 cohort is currently open. From the beginning of TNHS original cohort to TNHS 3, more than 280,000 participants enrolled in the programme.

The results of the aforementioned major studies, as well as other large previous HRT models on longevity and morbidity are described in the following sections.

2.4   HRT and its impact on longevity

Numerous observational studies, pooled analyses, and a number of randomised control trials assessed the effects of HRT on mortality. Majority of the mortality investigations were cause-specific. Most HRT studies took place in developed countries, and a large number were in the USA and UK. A significant number of studies took place in some other European countries, such as Finland, Denmark, Sweden, the Netherlands, Poland, and Italy. However, there was a wide variation in study design, the size and composition of the study population, and the length of follow-up. There was also a great variation in results among these studies. In

this section, the results of some of the bigger studies on mortality are discussed. Results of other studies that were reviewed can be found in Appendix A.

Folsom et al. (1995) conducted a prospective cohort study of the association of HRT and overall mortality on 41,070 postmenopausal Iowa women aged 55 to 69, and after 6 year follow-up this study found that HRT users had a reduced risk of all-cause mortality (relative risk (RR), 0.78; 95% confidence interval (CI), 0.65−0.94). Nurses' health study found that current hormone users with coronary heart disease had the largest reduction in mortality (RR, 0.51; 95% CI, 0.45−0.70) (Grodstein et al., 1997). Hodis and Mack (2014) assessed the existing studies on HRT and overall mortality in younger postmenopausal women (less than 60 years at initiation) and found a consistent drop in overall mortality of 30-50% in observational studies and of 19-39% in randomised trials.

In 1996, Persson et al. (1996) performed a cohort study of cancer incidence and mortality in 22,597 Swedish women receiving estrogen-only and combined HRT. The study follow-up was 13 years, and their findings suggested that the use of combined HRT was associated with an increased risk of breast cancer mortality (RR, 1.4; 95% CI, 1.1−1.8). They also found substantially increased risk of endometrial cancer, with the RR of 5.0 and 95% CI, 1.6−5.9 in women who received estrogen-only HRT, but those who received combined therapy had no elevated risk. A longitudinal cohort study of the association of HRT and gynaecological cancer, cardiovascular, and all-cause mortality on 4,544 women of aged 45 to 54 years from England and Wales found a lower risk of overall

mortality among HRT users (RR, 0.56; 95% CI, 0.47−0.66) (Hunt et al., 1990).

A meta-analysis of 26,708 participants from 30 clinical trials by Salpeter et al. (2004) investigated all-cause deaths and deaths due to cardiovascular disease, cancer, or other causes of mortality in younger (mean age < 60 at initiation) and older (mean age > 60) postmenopausal women. Their results showed that HRT reduced mortality in the younger age group (Odds ratio (OR), 0.61; 95%CI, 0.39−0.95), but not in the older age group (OR, 1.03; 95%CI, 0.90−1.18). When they analysed all ages combined, they found that HRT did not significantly affect total mortality (OR, 0.98; 95%CI, 0.87−1.18), the risk for cardiovascular (OR, 1.10; 95%CI, 0.9−1.34) or cancer (OR, 1.03; 95%CI, 0.82−1.39) mortality, but reduced mortality from other causes (OR, 0.67; 95%CI, 0.51−0.88). Five years later, the authors conducted a Bayesian meta-analysis (Salpeter et al., 2009) of 19 randomized control trials consisting of 16,000 women of mean age 55 years at baseline, and found reduced mortality risks (RR, 0.73, 95% CI, 0.52−0.96). When they combined 8 observational studies with 19 trials totaling 228,171 women, their findings did not differ much (RR, 0.72, 95% CI, 0.62−0.82).

Boardman et al. (2015) performed a meta-analysis of 19 randomised control trials of 40,410 postmenopausal women to evaluate the effects of oral HRT on all-cause mortality and on the prevention of cardiovascular disease. They found no effects of HRT on all-cause mortality, cardiovascular death, non-fatal myocardial infraction, angina, or coronary revascularisation. Another meta-analysis on HRT and all-cause mortality performed by Benkhadra et al. (2015) also found

no effect of HRT on mortality (RR, 0.99; 95% CI, 0.94−1.05). Pentti et al. (2006) analysed the association between HRT and mortality in 52 to 70 year old Finnish women focusing into account the duration of its use. Their study found that HRT did not have any impact on death from any cause in any duration. In their studies, the adjusted hazard ratios (HRs) were 1.05 (95% CI, 0.80−1.36) in women who used HRT for less than 5 years, and 1.06 (95% CI, 0.78−1.46) in women who used HRT for greater than 5 years compared to non-users. A recent analysis from WHI trial concluded that both combined (HR, 1.02, 95% CI 0.96−1.08) and estrogen-only (HR, 0.94, 95% CI 0.88−1.01) HRT were not associated with all-cause mortality during the 18 years of cumulative follow-up (Manson et al., 2017). On the contrary, a recent study by Malek et al. (2019) found that all-cause mortality was 31% higher in US postmenopausal women on HRT who had menopause at mean age of 45.

In the next section, existing results of the association of HRT with the incidence of various medical conditions are presented.

2.5   Effect of HRT on morbidity

Numerous observational studies and a number of randomised controlled trials investigated the effects of HRT on the development of various diseases (NICE, 2021). This section provides a review of the impact of HRT on various chronic health conditions. Only studies that comprised larger sample sizes are discussed

in this section. All other studies that were reviewed are summarised in the Appendix A.

2.5.1 Osteoporosis

Osteoporosis is a condition in which the bone mass density decreases and bone becomes more fragile. It is one of the most common causes of morbidity and mortality in postmenopausal women in the western countries. Due to estrogen deficiency, bone density decreases sharply in women, resulting in significant bone mass loss during the postmenopausal years (Gauthier et al., 2011). Women with osteoporosis are highly likely to face injuries such as wrist, hip and vertebral fractures.

A report published by the International Longevity Centre (ILCUK, 2010) on the current management of postmenopausal women aged 55 or over in the UK shows the following alarming results:

- Overall, the number, rate and cost of fractures among women of this age group are rising;

- The level of hospital admissions has increased from 10.4 per 1,000 population in 2004/05 to 11.4 per 1,000 population in 2008/09;

- The tariff cost has risen from approximately £390 million in 2005/06 to over £430 million in 2008/09;

- The total cost to the NHS of hospital stays alone is in excess of £400 million per year for women in this age group;

- Nearly 10% of women aged over 55 years who go into hospital with a fracture die while they are an in-patient − this equates to around 6,000 deaths per year;

- There is a significant regional variation in levels of fracture admissions for women.

Postmenopausal women are often prescribed HRT for osteoporosis treatment. Several epidemiological studies have shown that long term use of HRT provides protection against bone fractures (Christiansen, 1996; Cauley et al., 2003; Salpeter et al., 2009). A WHI trial of estrogen plus progestin showed 34% reduction in hip and clinical vertebral fractures, and 24% reduction in total osteoporotic fractures (Manson et al., 2013). Further analysis from WHI trial showed an overall 33% hip fracture decrease in women who received the CEE plus MPA and CEE-alone compared with the placebo group (Manson et al., 2017). During the WHI study's 13-year follow-up, women assigned to the CEE plus MPA group had fewer bone fractures than the placebo group (HR, 0.81, 95% CI, 0.68-0.97).

2.5.2 Dementia

Dementia is a chronic cognitive disorder that affects communication and performance of a person's daily activities. Alzheimer's disease (AD) is the most

common form of dementia, causing 60%−80% of all dementia cases (Alzheimer's Association, 2018). It specifically affects parts of the brain that control thought, memory and language. Older women are at a greater risk of Alzheimer's disease than men due to postmenopausal estrogen deficiency (Zandi et al., 2002). Over the past three decades, several observational studies (Henderson et al., 1994; Paganini-Hill and Henderson, 1994, 1996; Kawas et al., 1997; Baldereschi et al., 1998; Zandi et al., 2002), randomized control trials (Mulnard et al., 2000; Wang et al., 2000) and meta-analyses (LeBlanc et al., 2001; Henderson, 2014) have been carried out to examine the relationship between HRT treatment and the risk of developing dementia. Some observational studies suggested that early initiation of HRT after menopause and its long term use delays the onset of AD (Paganini-Hill and Henderson, 1996; Kawas et al., 1997; Baldereschi et al., 1998). A meta-analysis of nine randomized clinical trials of estrogen-only HRT conducted by Henderson (2014) found no improvement in cognitive symptoms in AD women who received HRT. Mulnard et al. (2000) found adverse effect of HRT treatment in women with AD. The WHI Memory Study (WHIMS) evaluated the effect of combined HRT on dementia (Shumaker et al., 2003) and found that combined HRT increased the risk of dementia in postmenopausal women aged 65 years or older. They enrolled 4,532 healthy postmenopausal women who were free from probable dementia and aged 65 years and older for the study. Probable dementia as defined in their study included Alzheimer's disease (AD), vascular dementia (VaD) and mixed types (both AD and VaD). After an aver-

age four years of follow up, the study found that the hazard ratio for probable dementia was 2.05 (95% CI, 1.21−3.48) in combined HRT users compared to the placebo group. At the same time, another WHI subgroup study on 2,947 hysterectomized women aged 65−79 years evaluated the effects of daily CEE (0.625 mg)-only on the incidence of probable dementia (Shumaker et al., 2004). After an average follow-up of 5.2 years, the study found that the relative risks of probable dementia for CEE-alone versus placebo was 1.49 (95% Cl, 0.83-2.66).

A recent case-control study conducted on 118,501 women of age 55 and over using CPRD and Qresearch datasets found no overall risks of developing dementia in HRT users (Vinogradova et al., 2021). Imtiaz et al. (2017) conducted an observational study on estrogen-only HRT and Alzheimer's disease based on Finland's hospital register and self-reported data. Their study followed up 8,195 women of ages 47−56 for 20 years, and found no association of estrogen-only HRT with AD risk (HR, 0.92, 95% CI, 0.68−1.2). However, another recent case-control study on HRT and the risk of Alzheimer disease in 83,688 postmenopausal women of age 60 and over in Finland showed that the use of HRT was associated with 9%−17% increased risk of Alzheimer disease (Savolainen-Peltonen et al., 2019). They also found that the risk of the disease did not differ significantly between estrogen-only (OR, 1.09, 95% CI, 1.05−1.14) and combined therapy (OR, 1.17, 95% CI, 1.13−1.21). Results from different studies on HRT and dementia appears to indicate that although early initiation of HRT delays the onset of dementia, late initiation may be detrimental.

2.5.3 Cardiovascular disease

Cardiovascular disease (CVD) is an umbrella term for conditions affecting the heart and blood vessels. Coronary heart disease (CHD), stroke, transient ischaemic attack (TIA), peripheral arterial disease (PAD), myocardial infarction, deep vein thrombosis, and pulmonary embolism are the most common forms of CVD. There are an estimated 3.5 million women living with CVD in the UK, with around 78,000 dying from conditions such as heart attack and stroke each year, accounting for a quarter of all female deaths (British Heart Foundation, 2016). CVD kills more than twice as many women as breast cancer in the UK. It is known that oestrogen helps to reduce the formation of fatty-plaques in women's heart. So, after menopause, women are more likely to develop cardiovascular disease than men. An observational study conducted by Gast et al. (2011b) showed that women with vasomotor symptoms have an increased risk of CHD. Majority of observational studies suggest that HRT may have beneficial effects in lowering the risk of cardiovascular events among postmenopausal women (Grodstein et al., 1994; Grodstein and Stampfer, 1995; Grodstein et al., 1997; Boardman et al., 2015). However, results from the randomised control trials are mixed (Hulley et al., 1998; Hodis et al., 2003; Schierbeck et al., 2012). Hodis and Mack (2014) showed that HRT may protect women from cardiovascular disease if started around the time of menopause, but other studies showed that late initiation could be harmful (Hulley et al., 1998). Gast et al. (2011a) pooled data from a Dutch and Swedish population-based sample of 10,787 women

of age 46 to 64 years who were free of CVD at baseline to study the effect of HRT on CVD, and found no association of HRT with CVD risk. Grodstein and Stampfer (1995) found that women who are receiving HRT treatment currently are at 50% lower risk of occurring CHD than a never user. Heart and Estrogen/progesterone Replacement Study (HERS) found no overall reduction in risk of CHD among postmenopausal women (Hulley et al., 1998). A meta-analysis of 19 randomised trials consisting of 40,410 postmenopausal women showed that HRT had no protective effects against death from CVD (Boardman et al., 2015) but was associated with an increased risk of stroke (RR, 1.24, 95% CI, 1.10-1.41).

2.5.4 Breast cancer

Each year, about 55,000 women are diagnosed with breast cancer in the UK (Macmillan Cancer Support, 2019). It becomes more common in women after menopause. Although the exact cause of breast cancer is not clearly known, it is thought that some factors, such as not having any children, first child birth after the age of 30, not breast feeding the children, early menarche (having period before the age of 12), late menopause (after the age of 55), taking oral contraceptive pill, prior family history of breast cancer, obesity, and so on, increase the risk of breast cancer. To date, the majority of published studies have found an increased risks of breast cancer in HRT users (Manson and Martin, 2001; Rossouw et al., 2002; Emily et al., 2003; Manson et al., 2013). In 2002, a study from the WHI trial found that combined HRT increased the risk of breast cancer

in women by 26% (Rossouw et al., 2002). However, further reports of WHI indicate that estrogen-only HRT decreased the risk of breast cancer by 23% (Manson et al., 2013). The MWS showed that current use of HRT increases the incidence of breast cancer (Emily et al., 2003). They reported that use of HRT reduces the sensitivity of mammography, making it less likely to detect breast cancer. Their findings also showed that the risk was substantially greater for combined HRT. The results are based on 517 deaths in women who had no history of breast cancer at recruitment. Another study showed that estrogen alone does not increase the risk of breast cancer if taken for $5-7$ years and women initiating HRT ten or more years after the menopause experience a 23% reduction in risk (Stefanick et al., 2006). In contrast, a randomised control trial of 1,006 healthy Danish women of ages $45-58$ undertaking combined hormone therapy concluded early initiation and prolonged HRT did not result in an increased risk of breast cancer (Schierbeck et al., 2012). Recently, a meta-analysis of 143,887 individual participant data from 58 studies on HRT and breast cancer risk showed that any type of HRT except vaginal estrogen was associated with a greater risk of breast cancer, which increased steadily with duration of use (Beral et al., 2019). The study found that risk was higher for combined therapy than for estrogen-only preparation. In summery, results from the various studies indicate that HRT is a risk factor for breast cancer, and that combined therapy users have greater risk than the estrogen-only users.

2.5.5 Endometrial and Ovarian cancer

Endometrial cancer is a type of cancer that develops in the tissues of the endometrium (lining of the uterus) in women. It is also known as uterine or womb cancer. According to the WHO, approximately 320,000 women are diagnosed with endometrial cancer worldwide each year and 76,000 die, making it the sixth most common cancer in women (McGuire, 2016). There was a 40% increase in endometrial cancer in the UK between 1993 and 2013 (Galaal et al., 2014). There is a high risk of endometrial hyperplasia in women who use estrogen-only HRT. A meta-analysis of 30 observational studies found an increased risk of endometrial cancer among estrogen-only HRT users compared to non-users (RR, 2.3, 95% CI, 2.1-2.5), but no increased risk of mortality from endometrial cancer. (Grady et al., 1995). Persson et al. (1999) investigated the risk of developing endometrial cancer on a cohort of 8,438 Swedish women. Their findings showed that there was a fourfold increased risk of invasive endometrial cancer in women who used estrogen-only therapy (RR, 4.2; 95% CI, 2.5-8.4), while those who used combined HRT had no significant risk (RR, 1.4; 95% CI, 0.6-3.3).

Ovarian cancer is a type of cancer that develops in the ovaries. Ovarian cancer is the seventh-most common cancer in women and the eighth-most common cause of death from cancer (WHO, 2014). It is more common in Europe and North America than in Africa and Asia (WHO, 2014). A cohort study of 44,241 postmenopausal women on HRT and the risk of ovarian cancer found

an increased risk of developing ovarian cancer in HRT users (RR, 1.6; 95% Cl, 1.2-2.0) (Lacey Jr et al., 2002). They also found that the increasing duration of estrogen-only use was highly associated with the risk of developing ovarian cancer. However, the information on the types of HRT was absent in their report. An observational study conducted by Folsom et al. (2004) on estrogen-only HRT and ovarian cancer among 31,381 postmenopausal Iowa women also showed an elevated risk of ovarian cancer (RR, 1.7; 95% Cl, 1.1-2.8). However, a meta-analysis of 12 case-control studies consisting of 2,197 cases and 8,893 controls did not find an increased risk (RR, 0.9; 95% CI, 0.7-1.3 in hospital-based studies and RR, 1.1; 95% CI, 0.9-1.4 in population-based studies)(Harris et al., 1992).

2.5.6 Lung cancer

Lung cancer is a commonly diagnosed cancer and a leading cause of mortality in women in the UK. Women are more prone to developing lung cancer than men and it kills more women each year than breast cancer, uterine cancer, and ovarian cancer combined. Even though it is believed that smoking is the number one cause of lung cancer in women, a higher percentage of women who develop lung cancer are life-long non-smokers (Verywellhealth, 2018). In 2011, about 19,700 women were diagnosed with lung cancer in the UK, making it the second most common cancer diagnosed in women after breast cancer (NHS Choices, 2015). However, there is a lack of research to understand the association between HRT and lung cancer. A population-based cohort study performed by Adami

et al. (1989) on Swedish postmenopausal women found that lung cancer risks increased in women who received HRT. Hampton (2009) studied the incidence of non-small cell lung cancer and mortality during 5.6 years of intervention with HRT or placebo, and found that past use of combined HRT increased the risk of dying from lung cancer. However, a recent meta-analysis by Yao et al. (2013) of 25 studies with 656,403 participants showed a reduced lung cancer risk in females receiving HRT treatment (OR, 0.91; 95%CI, 0.83-0.99). Their study found that HRT decreases lung cancer risks in the patients with BMI$< 25\ kg/m^2$ and never smokers (OR, 0.65 and 0.86 respectively). Their study also concluded that HRT increased the risk of lung cancer in women with early menopause. Further research may be needed to understand the association between HRT and lung cancer.

2.5.7   Colorectal cancer

Colorectal cancer, often known as bowel or colon cancer is considered the third leading cause of cancer and death in women (Cokkinides et al., 2005). The risk of colorectal cancer starts to rise in the mid-40s and continues to rise as women get older (Burkman et al., 2001). In 2015, there were 18,700 new cases of colorectal cancer and 7,300 deaths resulting from colorectal cancer in women in the UK (Cancer Research UK, 2015). Grodstein et al. (1999) conducted a meta-analysis of 18 epidemiological studies on postmenopausal HRT use and colorectal cancer, and found a 20% reduction in the risk of colon cancer in HRT users compared to

never users (RR, 0.80; 95% CI, 0.74 - 0.86). They further found a 34% reduction in the risk of colorectal cancer in current HRT users (RR, 0.66, 95% CI, 0.59 - 0.74). A cohort study by Calle et al. (1995) on 422,373 cancer free women found significantly decreased risk of colon cancer (RR, 0.71; 95% CI, 0.61 - 0.83) in women who took estrogen-only therapy. They also found decreased risk of mortality from colon cancer among current estrogen-only users, and there was trend of decreasing risk with increasing years of use among all HRT users. Pooled results from the WHI intervention and extended post-intervention follow-up of two hormone therapy trials found no impact of HRT on mortality from colorectal cancer (HR, 1.21; 95% CI, 0.79-1.84 for estrogen-only, and HR, 1.44; 95% CI, 0.97-2.15 for combined HRT) (Manson et al., 2017). It is thought that HRT decreases pancreatic bile acids that may promote colon cancer incidence in women (Burkman et al., 2001). The majority of studies found a reduced risk of colorectal cancer in HRT users, and hence health professionals should consider this benefit when prescribing HRT to women.

2.5.8 Diabetes

Diabetes mellitus is a chronic metabolic disorder that is primarily associated with the imbalance of glucose and insulin levels in the bloodstream. Based on the nature of the diabetes, it is classified as Type 1 and Type 2. Around 10% of the cases are Type 1 and 90% are Type 2. According to the International Diabetic Federation (IDF), there are about 371 million people worldwide who

have diabetes. In recent years, the number of people diagnosed with diabetes in the UK is estimated to be 3.5 million and among them 44% are women (Diabetes UK, 2015). This means that 1 in every 16 women in the UK has diabetes. Margolis et al. (2004) examined the effect of postmenopausal hormone therapy on diabetes incidence and insulin resistance using WHI randomized controlled data on 15,641 postmenopausal women of ages $50-79$, and found that combined HRT reduces the incidence of diabetes (HR, 0.79, 95% Cl, 0.67-0.9). Pooled results of 107 trials showed that HRT reduced the onset of diabetes in women without diabetes (RR, 0.70, 95% Cl, 0.60-0.90) (Salpeter et al., 2006). Results from various studies appear to indicate that HRT helps women to reduce the risk of developing diabetes.

2.6 Concluding remarks

This chapter reviewed the existing studies on hormone replacement therapy. It introduced HRT, its use, and the classification of HRT drugs. Then it presented existing results of association of HRT with mortality and morbidity. HRT is administrated in a variety of forms and routes, as evidenced by existing studies. The majority of all-cause mortality studies found either a lower risks of overall death or no effect on overall mortality in postmenopausal women who initiated HRT at younger ages (average age 55). However, a small number of studies reported an increased risk of all-cause mortality in women who started HRT on

or before the age of 45. This could be the result of complications associated with early menopause. Morbidity studies showed varied impact of HRT on different classes of disease. In summary, the majority of morbidity analyses found reduced risks of osteoporosis, colorectal cancer, diabetes, and cardiovascular disease in all types of HRT users, but an increased risk of breast cancer in combined HRT users, and an increased risk of endometrial cancer in estrogen-only users. Although no reduced risks of dementia were found in HRT users, larger studies reassured that there were no increased risks either.

Chapter 3

Review of primary care data and data extraction

In Chapter 1, the rationale, aims and objectives of developing HRT survival models were described. In Chapter 2, past studies related to HRT were reviewed in details. This chapter introduces the major primary healthcare databases in the UK, specifically The Health Improvement Network (THIN) database, which was used to extract patient information, and develop survival models of HRT. First, it describes different types of clinical data and discusses the significance of primary healthcare data in survival modelling of HRT. Then the structure of the THIN database and the clinical codes for data extraction are explained. Finally, the design of the study and participant selection from the database are described.

3.1 Sources of clinical data

Most research related to medical and health sciences relies on clinical data. This kind of data are either routinely collected during the period of ongoing patients' care or as a part of a clinical trial program or a designed observational study. Clinical data can be classified into six major categories, such as (i) Electronic health records, (ii) Administrative data, (iii) Claims data, (iv) Patient/disease registries, (v) Health surveys and (vi) Clinical trials data (Health Science Library, 2020). Electronic health records are obtained from the hospital, clinic or general practice, and include a broad range of information such as patients' demographics, diagnoses, treatments, drug prescriptions, laboratory tests, hospitalisations etc. Administrative data are primarily hospital discharge data that are reported to a government agency. Claims data comes from the health insured patients who used the healthcare delivery systems. Disease registries are clinical information systems that record only a limited range of data for chronic conditions such as Alzheimer's disease, diabetes, cancer, heart disease, asthma, and others. Health surveys are mainly conducted by the government agencies aiming to provide an accurate evaluation of the population health, and prevalence estimates of various disease conditions. National health surveys are one of the few types of data collected specifically for research purposes, and this type of data are more widely accessible to the researchers. Clinical trials data comes from the experimental study that aims to test new drugs, treatments, or interventions such as randomised control trials.

Clinical trials are designed to measure specific outcomes, risk factors, and exposures. In most cases, the number of participants is small to moderate in clinical trials, and the study usually covers a selected population from a limited region. This kind of data collection is more expensive and takes a longer period to gather information as the study requires a number of trial phases. Health survey data are collected by interviewing patients based on pre-set questionnaires, which is potentially subject to recall, and thus a chance of bias remains. Disease registries are only concerned with particular chronic conditions and hence this type of data contains only a limited range of information.

In routinely collected data, patients' information is recorded by the clinicians or practitioners at the time they visit the GP. Thus the amount of data collection depends on how frequently a person visited the primary or secondary care, and what type of information the clinicians find relevant to record. This type of data may contain a detailed medical history of patients but it might also have a substantial amount of missing entries for the risk factors that the researchers are interested in investigating. For example, women, children, and sicker person tend to visit the hospital or GP more frequently and so their records are more likely to be updated than men and healthy people. Therefore data on blood pressure, alcohol intake, or smoking status are likely to have more missing entries for the men and the healthier people. In addition, the clinicians or practitioners may code the same medical conditions and treatments in different ways in different practices. This could result in underestimation of the number

of patients selected for a specific condition.

To improve the quality of healthcare provided by the general practitioners, the Quality and Outcomes Framework (QOF) was introduced in the UK in 2004 (Gillam et al., 2012). It is a pay scheme which measures the performance of the GPs and other private practices in terms of the management of the most common chronic medical conditions, major public health concerns, and provision of preventive health services such as blood pressure checking or screening. After the initiation of the QOF, data recording has greatly improved in primary care sector (Taggar et al., 2012).

To improve the validity of research, and accuracy of prevalence estimates it is important to use the same set of codes in all medical studies. There is a clinical codes repository available online at ClinicalCodes.org to ensure that medical and health science researchers identify a medical condition or treatment by using the same set of codes (Springate et al., 2014).

In primary care databases, there is a high volume of person-years data available compared to hospital registries or health survey. All-cause mortality is reliably recorded in primary care database because GPs must be informed that their patients have died (BMA, 2013). In addition, as primary care data hold information on the comprehensive medical history rather than only on a set of specific target conditions, new risk factors can be studied which are not usually recorded in disease registers, or prospective trial cohort studies. Since

those potential risk factors are routinely recorded in primary care, there is an increasing use of this kind of data to develop survival models in clinical and epidemiological research.

Results from clinical trials or observational studies that prospectively collected data might not be generalisable due to strict inclusion and exclusion criteria, the relatively small sample size of population, and a small number of medical centres participating in the study. Secondary data and disease registers might only be representative of severe or chronic medical conditions. In primary care, clinicians keep a record of all kinds of treatments which was given to their patients. In the UK, primary care data are representative of the general population as almost all of its residents are registered at a general practice under the National Health Service. GPs are also informed when a patient enters into the secondary care (Hall, 2009). It means that primary care data retains records of both mild and severe conditions.

Although full primary care records would include all the UK population, the existing electronic medical databases include approximately 6% to 10% of all general practices. Hence there could be variation in representativeness of a particular databases by clinical system or geographical level, and therefore it is important to ensure the validity of the risk models on data from different participating general practices. While each type of clinical data has its own strengths and limitations, primary care data is an indispensable resource for estimating the long-term effect of HRT on the survival of women in the UK.

3.2 Primary care databases in the UK

In the UK, there are four independent public healthcare providers: the National Health Service (NHS) in England, NHS Wales, NHS Scotland, and Health and Social Care in Northern Ireland, which is collectively known as NHS. National health services provide free access to all types of primary and secondary health care to all UK nationals. To get this service, a person must be registered to a local GP practice. Patients records are updated electronically by the primary care physicians or nurses while they visit the practice. Since nearly all residents in the UK are registered to a local GP practice, and a wide range of information on the patients including demographics, treatment history, medical conditions are recorded, there is a high volume of follow-up data stored in this system. According to the British Medical Association (2018), there are 7,613 GP practices in England, 958 in Scotland, 454 in Wales, and 349 in Northern Ireland.

There are a number of software systems used by the GPs and nurses to electronically record patients' health information. In England, among 7,526 GP practices, Egton Medical Information Systems (EMIS) was used in 4199 (56%), followed by SystmOne in 2552 (34%) and Vision software in 636 (9%) practices (Kontopantelis et al., 2018). However, there were great regional variability found in all of these systems. Kontopantelis et al. (2018) reported that EMIS covers data from the West of England, London and the South; SystmOne covers the East and some regions in the South; and Vision software includes data from

London, the South, Greater Manchester and Birmingham.

There are a number of primary care databases in the UK storing these longitudinal electronic health records of anonymous patients from different practices. However, not all the practices are connected to the database systems. These databases are increasingly used for various medical research purposes since they record a high volume of data and provide opportunity to access a wide range of health information.

The main three large primary care databases in the UK that store these electronically recorded information are: the QResearch, the Clinical Practice Research Datalink (CPRD), and The Health Improvement Network (THIN) database (Vezyridis and Timmons, 2016). CPRD was previously known as General Practice Research Database (GPRD). QResearch database is linked to the EMIS services and includes medical records from approximately 1,000 practices. THIN and CPRD are connected through the Vision software system and more than 600 practices contributed to the database by 2015 (Vezyridis and Timmons, 2016). There were approximately half of the practices in THIN which overlap with CPRD database (Seminara et al., 2011). All of these databases keep records of anonymised patients who have been registered at some point at the participant GP practice, and thus also include patients who are not active or transferred out or died. The databases retain a wide range of information on patients including demographics, diagnoses, treatments, consultations, and lifestyle choices. The validity of these clinical information was investigated by systematic reviews and

external comparisons (MacDonald and Morant, 2008; Blak et al., 2011).

Access to all of the major UK primary care databases mentioned above is potentially costly and requires approval from the scientific review committee. Approval is based on submitting an application with a clear research plan and the source of funding. This study made use of THIN database because the IFOA funding covered the cost of accessing it. The study was approved by THIN scientific review committee on 27 September 2018 (approval number: 16THIN095). A detailed description of THIN database is given in the following subsections.

3.2.1 Overview of THIN database

In this study, a subset of The Health Improvement Network (THIN) database comprising patients born on or before 1960 has been used. THIN database stores a collection of longitudinal health records of anonymised patients from various general practices in the UK. The database was set up by In Practice Systems (INPS) in 2003 which collaborates with IMS Health (IMS Health Incorporated, 2015b). INPS developed and maintains VISION software to store registered patients data from various general practices, and the IMS Health provides access to the electronically recorded data for medical research purposes (Wijlaars, L., 2015). Individual patient's information is recorded in THIN in the same way across the different operating clinical systems. Currently, THIN database retains the electronic medical records of 17 million non-identifiable patients equivalent

to 92 million person-years data (IQVIA Medical Research Data, 2017). There are 3.1 million actively registered patients in the database from over 770 general practices, covering 6.2% of the UK population (IQVIA Medical Research Data, 2017). All data in THIN are fully anonymised, processed and validated by CSD Medical Research UK. Data from THIN is made accessible to external researchers conducting protocol driven studies via IQVIA under a sub-license or research agreement approved by THIN Scientific Review Committee. At present, THIN is a registered trademark of Cegedim SA in the United Kingdom and other countries. Reference made to the THIN database is intended to be descriptive of the data asset licensed by IQVIA.

3.2.2 Structure of THIN database

In THIN database, patients' data are arranged by individual general practice. Each practice has an unique alphanumeric code with one lowercase letter followed by a four digit number. The information on different medical conditions are catalogued using hierarchical Read codes. These codes are also alphanumeric and grouped into different themed chapters (e.g. cancer), and include terms relating to symptoms, diagnoses, procedures, and laboratory tests (Wijlaars, L., 2015). Prescription of drugs are currently entered using Multilex codes, which can be easily linked to British National Formulary (BNF) codes in THIN. Most practices now have laboratory data which can be automatically transferred to the electronic medical record.

Table 3.1: Main file types in THIN database. Adapted from Wijlaars, L. (2015)

File Name	Descriptions
PATIENT	Age, sex, registration date when entering the practice, and date when leaving the practice
MEDICAL	Medical diagnoses, date of diagnosis, and location (e.g., GPs office, hospital, consultant) of the event and an option for adding free text referrals to hospitals and specialists
THERAPY	All prescriptions along with the date issued, formulation, strength, quantity, and dosing instructions, indication for treatment for all new prescriptions, and events leading to withdrawal of a drug or treatment
ADDITIONAL HEALTH DATA (AHD)	Vaccinations and prescription contraceptives; miscellaneous information such as smoking, height, weight, immunizations, pregnancy, birth, death, and laboratory results
POSTCODE VARIABLE INDICATORS (PVI)	Postcode linked area based socio-economic, ethnicity and environmental indices
CONSULTATION	Date, time and duration of consultation
STAFF	Gender and roles of staff who entered the data

THIN data is structured by seven ASCII (American Standard Code for Information Interchange) standardised files. In these seven files are names Patient, Medical, Therapy, Additional Health Data (AHD), Postcode Variable Indicator, Consultation, and Staff. AHD file is linked with Patient file by patient ID. In Table 3.1, the main file names and description of the content of the files in THIN database system are presented.

3.2.3 Representativeness of THIN data

To build a good statistical model, the sample should well reflect the targeted population, a high number of follow-up data must be available, and the important risk factors should be adjusted for to minimise the residual confounding. Thus it is essential to assess at what extent THIN is comparable to the general population. A number of studies assessed the representativeness of THIN to the UK general population (Hippisley-Cox et al., 2008; MacDonald and Morant, 2008; Maguire et al., 2009; Blak et al., 2011). Blak et al. (2011) compared the demographics, deprivation status (Townsend score), Quality and Outcomes Framework (QOF) conditions prevalence and deaths records from THIN with the national statistics and QOF 2006/2007 data, and concluded that THIN is generalisable to the UK population in terms of demographics and crude prevalence of chronic medical conditions. They also reported that the death rates in THIN and the UK national death rates are similar when adjusted for demographics and deprivation. However, Hippisley-Cox et al. (2008) reported that there are slightly more patients in THIN from affluent areas. It is therefore important to adjust the deprivation index in model development in order to obtain representative estimates of the UK population. Blak et al. (2011) also reported that THIN includes slightly fewer people of age under 25 years, but this will not affect this study as it focused on women who are aged 46 or over at study entry. The prevalence of the baseline characteristics in selected study population was calculated and compared with the UK national statistics of disease prevalence.

This will be discussed in Chapter 5 in details.

3.3 Clinical codes

Clinical codes are the symbolic form of medical terminology that the clinicians and health service providers use when recording patient's data. Thus, knowledge of clinical codes is required to extract patient's information from the electronic medical record (EMR) databases. It is also needed to establish the validity of research using EMR databases. In this section, two types of clinical codes, the (i) British National Formulary codes and (ii) Read codes are introduced.

3.3.1 The British National Formulary codes

The British National Formulary (BNF) is a UK based drug reference book that contains a list of coded information on medicines prescribed by the GPs and healthcare professionals (EBM DataLab, 2014). BNF provides information on the dosages, side effects, and indications for over 70,000 medicines. It displays medicines in a hierarchical order, and the NHS Business Services Authority (BSA) classify pseudo-codes to drugs and chemicals using a legacy version of the BNF hierarchy (NHS Booklet, 2017). This pseudo-classification is used as a unique identifier in the Practice Level Prescribing Data to show what was prescribed. These BNF codes give a lot of information about a drug or appli-

ance. The first character indicates which part of the BNF a drug is from. For example, drugs in BNF Chapter 4 (Central Nervous System) always begin with "04" (EBM DataLab, 2014). The code then further subdivided into sections. For example, Section 3 of Chapter 4 in BNF contains Antidepressant Drugs, all starting with"0403". The last few characters of the BNF code provide more detailed information about any specific drug, such as whether the product is generic or branded, and about the presentation of the drug (e.g. whether it is a capsule or tablet, and the strength of the drug). Each year, in March and September, the BNF updates their codes (NHS Digital, 2017). NHS Prescription Services update their BNF classifications once a year. In general, BNF codes for the drugs are organised in the following way:

- Characters 1 and 2 show the BNF chapter

- Characters 3 and 4 show the BNF section

- Characters 5 and 6 show the BNF paragraph

- Character 7 shows the BNF sub-paragraph

- Characters 8 and 9 show the chemical substance

- Characters 10 and 11 show the product

- Characters 12 and 13 show the strength and formulation

- Characters 14 and 15 show the equivalent

The 'equivalent' is defined in the following way:

1. If the product is a generic, the 14th and 15th character will be the same as the 12th and 13th character.

2. If the product is a brand, the 14th and 15th character will match that of the generic equivalent.

3.3.2 Read codes

Read codes are a comprehensive list of clinical terminology that are used by the GPs and health professionals in the UK to record the treatment or care given to their patients. These includes a wide range of clinical entities, such as, drugs, treatment, surgery, diagnosis, signs and symptoms, and a variety of administrative items. These codes were first introduced by Dr James Read in 1982 following the initiation of a GP-based computer system, and thereby named after him (Springate et al., 2014). Later in 1985, it was recognised by the NHS, and since then Read codes are used throught the services (NHS Digital, 2018a). There are two versions of Read codes available: Version 2 (v_2) and Version 3 ($CTV3$ or v_3). Both v_2 and v_3 provide a standard vocabulary for health professionals to enter patients' information and procedures in computer systems across primary and secondary care in the UK. To improve the validity and reproducibility of medical research, lists of Read codes for various medical conditions are available online at ClinicalCodes.org (Springate et al., 2014).

3.4 Study design and participant selection

In this section, the design of the study, patient selection process, and inclusion/exclusion criteria for patients for the development of the HRT survival model are explained first. After that, full data extraction procedure from THIN is described.

3.4.1 Design of HRT study

This study focused on estimating the long-term effect of estrogen-only and combined HRT on the hazards of all-cause mortality in healthy women. A matched cohort study was conducted to estimate the association between hazards of death and HRT treatment using retrospectively collected health records from THIN database. Patients who started any formulation of oral or transdarmal HRT for the first time at the age of 46 years or above were selected as the exposed group. The reason for selecting patients aged 46 and onward is that the perimenopausal stage usually starts around the age of 46 in most healthy women in the UK (National Institute for Health and Care Excellence, 2019). The unexposed group were never users of HRT or any type of drug containing estrogen and/or progesterone. To create a balanced cohort, patients in the exposed group were matched with patients up to three in the unexposed group by birth year and general practice. A balanced cohort ensures that the groups compared are of similar characteristics except for the exposure of interest (Buring, 1987). A

balanced cohort increases the statistical power and efficiency of a model as well (Greenland and Morgenstern, 1990). Matching by year of birth allows for consideration of possible medical advancements over time (Kleinbaum and Klein, 2012). When patients from multiple medical practices are included in a study, the medical practice should also be matched because patients from the same practice are more likely to be comparable with each other (Rasbash et al., 2012).

The study entry point for patients in the exposed group was the date when they were prescribed the first HRT, and for the unexposed group it was the study entry date of their matched HRT users. Use of the HRT start date as the study entry for the matched non-users point allowed this study to estimate the hazards in both groups within the same follow-up time frame. The study cut-off point was 1st January 2017. Participants were followed-up from the study entry until death, transfer to another practice, or the study end date, whichever came first. Patients were only eligible to be included in the study if at the time of study entry they had been registered at a general practice as an active patient for at least one year, and their health records had been accessed at least once within the past ten years. Patients with a previous history of any kind of cancer, acute myocardial infarction, severe heart failure, stroke (except transient ischaemic attack), chronic kidney disease (stage 3 to 5), dementia, surgically induced menopause before 45 years of age, premature ovarian insufficiency, and premature menopause were excluded from the study at baseline to assess the long-term impact of HRT on healthy postmenopausal women.

HRT users were classified as either estrogen-only or combined HRT users. Patients were grouped as combined HRT users if they received estrogen and progesterone in a single prescription or in two separate prescriptions. The baseline characteristics for inclusion in the survival models were age at first HRT treatment, birth year, type 2 diabetes, osteoporosis, peripheral arterial/vascular disease (PAD/PVD), coronary heart disease (CHD), oophorectomy, hysterectomy, hypertension, systolic and diastolic blood pressure (SBP/DBP), hypercholesterelomea, anti-hypertensive drugs, smoking, body mass index (BMI), and deprivation status. Baseline characteristics were chosen based on their importance as determined by literature review and expert knowledge with the team. The Standard Query Language (SQL) server 2016 was used to extract data from THIN. The full data extraction process is described in the next subsection.

3.4.2 Data extraction

The subset of THIN data used for this study retains the electronic medical records of 3,515,292 patients who were born on or before the year 1960 and followed up to January 2017. Among them 1,664,457 (47.4%) patients were male and 1,850,835 (52.6%) female. The selection of HRT users, non-users, and the matching process are described in the following subsections.

Selection of HRT users

Female patients with a record of HRT prescription were first selected as the possible exposed group. These patients were identified using the BNF codes of drugs containing oestrogen and/or progesterone. There are four such types of BNF codes containing these drugs: 06040101 (oestrogen), 06040102 (oestrogen and progesterone combined), 08030100 (oestrogen in malignant disease), and 08030200 (progesterone in malignant disease). Among these codes, only 06040101 and 06040102 codes correspond to HRT drugs for menopausal treatments. The BNF codes 08030100 and 08030200 refer to certain types of cancer treatment drugs, and thereby were excluded when selecting patients prescribed HRT. Thus, using the BNF codes (06040101, 06040102) for menopausal treatment, 496,145 women were selected in total. They constituted 26.8% of female population in THIN subset. Next, the active patients were identified by considering those who have valid medical records for at least twelve months after their registration date in the corresponding general practice, and whose health record had been accessed at least once in the past ten years. Actively registered patients who were prescribed HRT after the acceptable mortality reporting (AMR) date of the corresponding general practice were included in the study. AMR date is a starting date from which practice-recorded mortality was close to age- and sex-standardised national mortality rates. However, active patients whose first HRT prescription was issued before the AMR date of the corresponding practice were also selected in this study if they continued HRT after the AMR date.

Excluding patients with invalid records, 369,716 patients remained to be eligible for inclusion in the exposed group, which is approximately 20% of the total number of women patients in the subset of THIN database. Active patients who were at least 46 years of age or above at their first HRT prescription were then selected. There were 281,003 (15.2%) women in total who started HRT for the first time at 46 years or above between 1984 and 2017. After that, patients with any type of cancer, and patients who have had acute myocardial infraction (AMI), severe heart failure, stroke (except TIA), dementia, surgically induced menopause before the age of 45, chronic kidney disease (CKD) stage 3-5, premature menopause, premature ovarian insufficiency, and women undergoing in vitro fertilization (IVF) treatment were excluded. This left 151,683 (8.2%) female patients to be eligible for inclusion in the exposed group. After excluding patients with invalid data entries, such as those with a negative time length from the first HRT prescription to the death, transferred or study end date, there were 135,663 (7.3%) women who were finally eligible for inclusion in the exposed group (see Figure 3.1).

Selection of HRT non-users

After selecting all female patients who took HRT, there were 1,354,690 (73.2%) women left in the THIN subset, who were possibly eligible to be enrolled in the unexposed group. Among them, women who received any kind of hormonal drugs containing estrogen and/or progesterone, including hormonal contracep-

tives, were excluded. After excluding these patients, there were 973,701 (52.6%) women who remained in the unexposed group. Excluding patients with inactive status, invalid records, and all of the medical conditions that were excluded from the exposed group, there were 610,628 patients left in the unexposed group who were eligible to be the possible non-users, which is 33% of total female population in THIN subset. These women were eligible for matching with the selected patients in the exposed group.

Matching

Patients in the exposed group were matched to patients in the unexposed group based on their year of birth and general practice. The justification of matching patients by these factors was explained in subsection 3.4.1. As there were more unexposed patients than those in the exposed group, the matching ratio was set to 1:3, however, in a number of cases, the maximum number of non-users that matched per HRT users was less than 3. Matching patients in the exposed group with up to ten patients in the unexposed group is optimal, although matching five or more unexposed patients hardly improves the statistical efficiency (Raboud and Breslow, 1989). For this reason, the majority of studies match fewer than five unexposed participants with one exposed participant (Cepeda et al., 2003).

There were 112,354 patients in total in the exposed group, each of whom got at least one, two, or three matched non-users, and the total number of
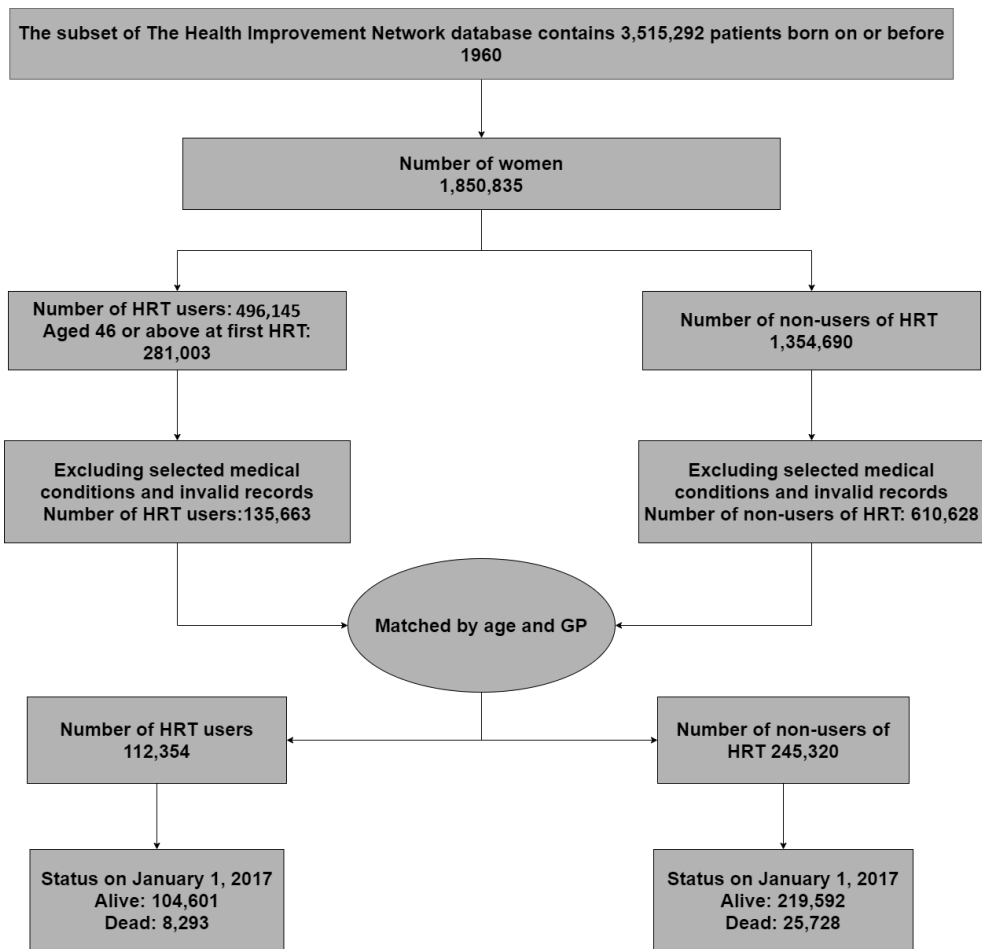
Figure 3.1: Selection procedure of study participants. HRT users were matched with non-users by year of birth and general practice.

matched non-users was 245,320. Out of the initial 135,663 selected HRT users, 23,309 patients were excluded because there were no exact matches based on the matching criteria. Among 112,354 HRT users, 94,132 patients received combined HRT and 18,222 patients received estrogen-only HRT. Figure 3.1 presents a flow-chart of the selection process of the study participants.

3.5 Summary

In this chapter, the types and classification of clinical data, their availability, and the importance in developing survival models for HRT using primary care data are explained first. Then the structure of The Health Improvement Network electronic database and clinical codes for data extraction were discussed. Finally, the design of the study and the full data extraction process are explained. The full dataset selected for this research contains 112,354 HRT users and 245,320 matched non-users of HRT, and these patients were considered for model development.

Chapter 4

Review of statistical methods in survival analysis

In Chapter 1, the background, rational, aims and objectives of HRT study were discussed. Chapter 2 provided a review of the HRT literature, and Chapter 3 discussed the importance of primary care data in developing HRT models, as well as the data extraction process from the THIN database. This chapter discusses the statistical procedures and techniques used to model the survival from the primary care data. Section 4.1 introduces the common terminologies and notation used in survival analysis. Then the parametric and non-parametric survival models are discussed in Section 4.2 and 4.3, respectively. In the subsequent sections, the semi-parametric Cox regression model, parameter estimation from the Cox regression model, and test-statistics to validate the model are presented.

Finally, in Section 4.9, the technique for multiple imputation and the pooling estimated parameters from imputed models are discussed.

Survival analysis is a collection of statistical techniques for analysing the time to an event of interest (Kleinbaum and Klein, 2012). The event can be death, relapse from remission, or a particular disease incidence. For this research, the primary outcome of interest is survival time from the study entry to the event of death from all-cause for the HRT users compared to non-users. Survival analysis is widely used in medical sciences, biological sciences, engineering, and many other areas of research. The main goals of survival analysis are to estimate and interpret survival and/or hazard functions, to compare survival and/or hazard functions across subgroups, and to assess the relationship of explanatory/predictor variables to survival time of the subject. The capability to handle censored survival times makes survival analysis most appealing in medical research. The survival analysis techniques which are relevant to this study are explained below.

4.1 Terminology and notation

In this section, commonly used terminology and notation in survival analysis, such as the survival function, hazard function, cumulative hazard, probability density function, and censoring are described.

4.1.1 Survival time

Survival time is the time elapsed from the study entry until an event of interest occurred. The survival time is a random variable denoted by T, and let t be any specific value of T. Since survival time cannot be negative, the values of T are always non-negative, i.e., $T \geq 0$.

4.1.2 Probability density function

The different values that the random variable T can take have a probability distribution, and the probability density function (p.d.f) of T is denoted by $f(t)$. It is defined as the limit of the probability that an individual faces an event in the short time interval $[t, t + \Delta t)$ per unit width Δt, i.e.,

$$f(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t}. \tag{4.1}$$

Thus the p.d.f, $f(t)$, gives the instantaneous potential for an event to occur at t per unit time. The probability distribution of the survival time T can be defined through the survivor/survival function, the hazard function or the cumulative hazard function. The survival function, $S(t)$, gives the probability that a person survives longer than some specified time t, i.e., the random variable T exceeds the specified time t (Kleinbaum and Klein, 2012). It is defined by:

$$S(t) = P(T > t) = 1 - F(t) = \int_{t}^{\infty} f(u)du, \tag{4.2}$$

where $F(t)$ is the cumulative distribution function (c.d.f); the probability that the random variable T is less than or equal to the specified time t, i.e.,

$$F(t) = P(T \leq t) = \int_0^t f(u)du. \tag{4.3}$$

Survivor functions are always monotonically decreasing as the probability of survival decreases with time, i.e., $S(t) \leq S(u)$ for all $t \geq u$. Initially, at time $t = 0$, it is assumed that the survival probability, $S(0) = 1$, and as time increases, $\lim_{t \to \infty} S(t) = 0$, so that a possibility of eternal life is excluded. In survival analysis, the survival time is also referred to as event failure time.

4.1.3 Hazard function

The hazard function, denoted by $h(t)$, gives the instantaneous potential per unit time for the event to occur, given that the individual has survived up to time t. It is defined by (Kleinbaum and Klein, 2012):

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{1 - F(t)}. \tag{4.4}$$

This could be thought of as the probability of experiencing the event in the next instant given that the event has not happened yet. The hazard function is always non-negative and it has no upper bound. It is also termed the hazard rate, the conditional failure rate, the instantaneous death rate, the intensity rate, or the force of mortality. The hazard function is particularly useful to describe the chance of an event occurring over time. An increasing hazard of mortality

could be an effect of natural ageing while a decreasing hazard of mortality could be a result of recovering after a surgery or receiving a specific treatment. A bathtub shaped hazard function is often appropriate to explain the hazard of death for populations followed for an entire lifetime.

The cumulative hazard function is the integral of the hazard function from time 0 to t. It is denoted by $H(t)$ and defined by:

$$H(t) = \int_0^t h(u)du = \int_0^t \frac{f(u)}{1 - F(u)}du = -\log(1 - F(t)) = -\log(S(t)). \quad (4.5)$$

4.1.4 Relation between survival function and hazard function

There is a clearly defined relationship between the survival function, $S(t)$ and the hazard function, $h(t)$. One can be derived from another using the following formulae (Kleinbaum and Klein, 2012):

$$S(t) = \exp\left[\log(S(t)\right] = \exp\left[-\int_0^t h(u)du\right] = \exp[-H(t)]; \quad (4.6)$$

and

$$h(t) = -\frac{d}{dt}[\log S(t)] = -\left[\frac{dS(t)/dt}{S(t)}\right] = -\left[\frac{d(1 - F(t))/dt}{S(t)}\right] = \frac{f(t)}{S(t)}, \quad (4.7)$$

where $f(t)$ is the first derivative of the cumulative distribution function $F(t)$, i.e., $f(t)$ is the probability density function.

4.1.5 Censoring

In time to event data, individual's true survival time may not be exactly known. Most survival analyses encounter this key analytical problem, and it is termed as censoring (Kleinbaum and Klein, 2012). Several reasons cause censoring, such as an individual did not experience the event during the follow-up period, lost during the follow up, or died (if the event of interest is not death) before the study ends. The most well known types of censoring are left, interval, and right censoring (Prinja et al., 2010). In left censoring, the event happens before the start of the study. In interval censoring, the event occurs within a time interval rather than at any certain time (Hosmer et al., 2011). In right censoring, the event did not occur before or during the study period, and thus the true survival time is longer than the censoring time. The main causes of right censoring are that the subject is no longer in the study or the study ends before the subject has faced the event. In survival data, right-censoring is the most common type of censoring. In this study, all censored observations are right-censored, where the censoring occurred due to patients being alive at the end of the study or transferred to another general practice during the study period. In survival data, censoring and failures are usually coded with a dichotomous censoring indicator δ_i, where $\delta_i = (0, 1)$. If the subjects are censored, then $\delta_i = 0$, and if the event of interest occurred then $\delta_i = 1$. The observations which are not censored are called complete observations.

In the next section, different types of parametric models used in survival analysis are described.

4.2 Parametric survival models

If the probability density function $f(t)$, of the survival time t, follows a particular distribution then it could be modelled by parametrically. A number of parametric families of distributions are used in the analysis of survival data. In this section, some of the most important parametric survival distributions, such as Exponential, Weibull, Log-normal, Log-logistic, Gompertz, and Gompertz$-$Makenham distributions are introduced. These parametric models are widely used in medical and actuarial applications.

4.2.1 Exponential distribution

If the hazard function is equal to a positive constant, so that $h(t) = \lambda$, then the survival time follows an exponential distribution. Thus using Eq. (4.6), the survivor function for exponential distribution can be written as:

$$S(t) = \exp\left[-\int_0^t \lambda du\right] = \exp(-\lambda t), \qquad (4.8)$$

and therefore, the probability density function is given by

$$f(t) = \lambda \exp(-\lambda t), \qquad (4.9)$$

for a parameter $\lambda > 0$.

The exponential distribution was the first widely used lifetime distribution model. However, the application of this model is very limited as a constant hazard is very unusual in real life.

4.2.2 Weibull distribution

The Weibull distribution is one of the most commonly used lifetime distributions in biological and medical applications. In Weibull distribution, the hazard function takes the form

$$h(t) = p\lambda^p t^{p-1}, \tag{4.10}$$

where λ is a scale parameter and p is a shape parameter of the distribution. The Exponential distribution is a special case of Weibull distribution when the shape parameter $p = 1$ in Eq. (4.10).

The survivor function for the Weibull distribution can be written as:

$$S(t) = \exp\left[-p\lambda^p \int_0^t u^{p-1} du\right] = \exp(-\lambda t)^p. \tag{4.11}$$

The probability density function of the Weibull distribution is

$$f(t) = p\lambda^p t^{p-1} \exp(-\lambda t)^p, \tag{4.12}$$

for the shape parameter $p > 0$ and scale parameter $\lambda > 0$.

The log-hazard of the Weibull distribution can be written as:

$$\log h(t) = p \log(\lambda) + \log(p) + (p-1) \log(t), \qquad (4.13)$$

which is a linear function of $\log(t)$ with constant intercept, $(p \log(\lambda) + \log(p))$ and slope $(p-1)$. Thus the hazard increases monotonically if $p > 1$, constant if $p = 1$, and decreases monotonically if $p < 1$.

The Weibull model is fairly flexible, and has been found to provide a good fit for a wide range of lifetime data.

4.2.3   The Log-Normal distribution

If logarithm of the survival time T is normally distributed with mean μ and variance σ^2 i.e., $\log(T) \sim N(\mu, \sigma^2)$ then the survival time T is log-normally distributed. The survival function of the log-normal distribution can be written

$$S(t) = 1 - \Phi \left(\frac{\log(t) - \mu}{\sigma} \right), \qquad (4.14)$$

where μ is a location parameter, the parameter σ is the shape parameter and $\Phi(\cdot)$ is the cumulative probability distribution function of a standard normal distribution. The formula for the hazard function of the log-normal distribution is

$$h(t) = \left(\frac{1}{\sigma t} \right) \phi \left(\frac{\log t}{\sigma} \right) \left[\Phi \left(\frac{-\log t}{\sigma} \right) \right]^{-1}, \qquad (4.15)$$

where $\phi(\cdot)$ is the probability density function of the standard normal distribution.

The probability density function of log-normal distribution is given by

$$f(t) = \left(\frac{1}{\sigma t \sqrt{2\pi}}\right) \exp\left(-\frac{(\log(t) - \mu)^2}{2\sigma^2}\right) \tag{4.16}$$

4.2.4   Gompertz distribution

A random variable T follows the Gompertz distribution if the survivor function takes the following form:

$$S(t) = \exp\left[-\frac{\alpha}{\beta}(\exp(\beta t) - 1)\right], \tag{4.17}$$

where α is a scale parameter and β is a shape parameter, with $\alpha > 0$ and $\beta > 0$.

The hazard and log-hazard functions of the Gompertz distribution are given by:

$$h(t) = \alpha \exp(\beta t), \tag{4.18}$$

and

$$\log h(t) = \log(\alpha) + \beta t, \tag{4.19}$$

The cumulative hazard function of the Gompertz distribution is:

$$H(t) = \frac{\alpha}{\beta}(\exp(\beta t) - 1); t > 0. \tag{4.20}$$

The log hazard (4.19) of the Gompertz distribution is a linear function with an intercept $\log(\alpha)$ and slope β. The probability density function of the Gompertz distribution is:

$$f(t) = \alpha \exp(\beta t) \exp\left[-\frac{\alpha}{\beta}(\exp(\beta t) - 1)\right]. \tag{4.21}$$

Actuaries, biologists and demographers often use the Gompertz distribution to model the lifespan of adult population.

4.2.5   Gompertz−Makeham distribution

Gompertz−Makeham distribution is the extension of the Gompertz distribution, where the hazard function takes the form:

$$h(t) = \alpha \exp(\beta t) + \lambda. \tag{4.22}$$

Here α, β and λ are positive real parameters. The survival and the probability distribution function of the Gompertz−Makeham distribution are:

$$S(t) = \exp\left[-\lambda t - \frac{\alpha}{\beta}(\exp(\beta t) - 1)\right]. \tag{4.23}$$

$$f(t) = (\alpha \exp(\beta t) + \lambda) \exp\left[-\lambda t - \frac{\alpha}{\beta}(\exp(\beta t) - 1)\right]. \tag{4.24}$$

Gompertz−Makeham distribution is widely used in the actuarial tables to describe human mortality.

In the next section, the non-parametric survival analysis for unknown distribution is described.

4.3   Non-parametric statistical methods

Non-parametric or distribution-free methods are frequently used to describe time-to-event data. These methods are easy to understand and implement. In this

section, most widely used non-parametric methods of estimating survival function namely the Kaplan-Meier method, and commonly used non-parametric tests for comparing survival functions are discussed.

4.3.1 Kaplan-Meier method

The Kaplan-Meier (KM) method is a non-parametric statistical procedure to estimate the survival probabilities, and other distribution characteristics, such as the median survival time from the observed survival data (Kaplan and Meier, 1958). KM method allows estimation of the survival curves from censored observations and it is particularly useful when the data are right-censored.

Let $0 < t_{(1)} < t_{(2)} < \ldots < t_{(m)}$ be the distinct ordered observed survival times that are sorted in ascending order, d_j be the number of deaths that occurred at time $t_{(j)}$ where $1 \leq j \leq m$, and n_j be the number of subjects at risk of dying just before time $t_{(j)}$ excluding subjects who are censored at time interval $[t_{(j-1)}, t_{(j)})$. Then the KM method estimates the survival function at time $t_{(j)}$ as:

$$\widehat{S}(t_{(j)}) = \widehat{S}(t_{(j-1)}) \left(1 - \frac{d_j}{n_j}\right). \tag{4.25}$$

Substituting the survival probabilities at times $t_{(j-1)}, t_{(j-2)}, \ldots, t_{(1)}$ respectively in equation (4.25), the KM estimator of survival function at time t can be obtained as:

$$\widehat{S}(t) = \prod_{j:t_{(j)} \leq t} \left(1 - \frac{d_j}{n_j}\right). \tag{4.26}$$
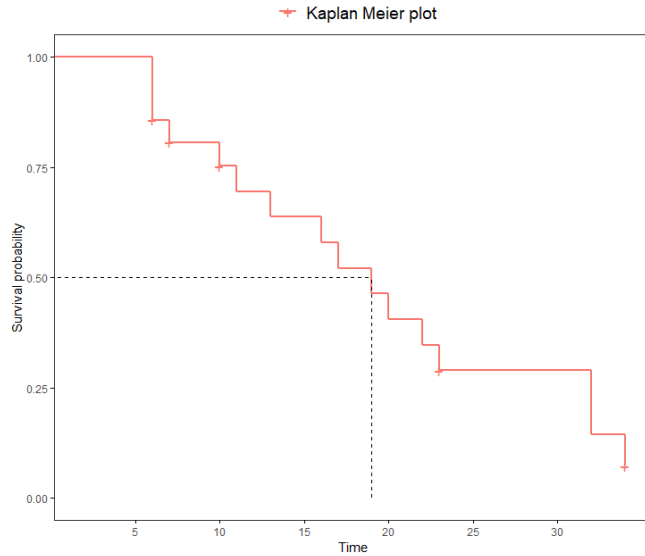
Figure 4.1: A Kaplan-Meier curve of the estimated survival function plotted using the example data shown in Table 4.1. The "+" sign indicates the time when subjects are censored. At time 6, 7, 10, 23, and 34 there were censored observations. The vertical jumps show the time points when events occurred.

The ratio $\dfrac{d_j}{n_j}$ gives the probability of experiencing an event at time $t_{(j)}$, given that the individuals have not experienced the event before that time. The KM estimator $\widehat{S}(t)$ is a decreasing right-continuous step function with jumps at death times. It is also known as the "product-limit estimator" as the survival function in Equation (4.26) is the product of the estimated probabilities of surviving in small successive time intervals, $[t_{(j-1)}, t_{(j)})$ up to time t. The subjects who are censored are excluded from the number at risk of dying in each time interval. The length of the jumps depends on the total number of events observed and the number of observations censored. The estimated median survival time in KM method is the survival time at which the estimated survival function reaches 0.5. An example of survival data is shown in Table 4.1, and Figure 4.1 is the KM plot for these data.

Table 4.1: An example data and calculations used to construct the KM curve in Figure 4.1

Ordered event time $t_{(i)}$	Number at risk n_i	Number of events d_i	Number of censored observations at $[t_{(i)}, t_{(i+1)})$	Estimated survival probability $\hat{S}(t_i)$
0	21	0	0	1
6	21	3	1	0.85
7	17	1	1	0.80
10	15	1	1	0.75
11	13	1	0	0.69
13	12	1	0	0.63
16	11	1	0	0.57
17	10	1	0	0.51
19	9	1	0	0.45
20	8	1	0	0.39
22	7	1	0	0.33
23	6	1	1	0.27
32	4	2	0	0.14
34	2	1	1	0.10

The variance of the estimated survival function, $\widehat{S}(t)$ is calculated by Greenwoods's formula (Greenwood, 1926):

$$\widehat{\text{var}}\left[\widehat{S}(t)\right] = \left[\widehat{S}(t)\right]^2 \sum_{j:t_{(j)} \leq t} \frac{d_j}{n_j(n_j - d_j)} \tag{4.27}$$

and the square root of the estimated variance gives the standard error. The KM estimator $\widehat{S}(t)$ is asymptotically normally distributed for a large number of observations (Hosmer et al., 2011). Therefore, the $100(1 - \alpha)\%$ confidence interval for the survival function can be calculated as:

$$\widehat{S}(t) \pm z_{(1-\alpha/2)}\sqrt{\widehat{var}\left[\widehat{S}(t)\right]}, \tag{4.28}$$

where $z_{(1-\alpha/2)}$ denotes the upper $\alpha/2$ percentile of a standard normal distribution, and $\sqrt{\widehat{var}\left[\widehat{S}(t)\right]}$ gives the standard error. However, the endpoints of the interval

can be negative or greater than one. To form a confidence interval in the range 0 to 1, Kalbfleisch and Prentice (2002) suggest to obtain a confidence interval for the log-log survival function and then transform back to the confidence interval for the survival function. If $\widehat{C}_l$ and $\widehat{C}_u$ denote the lower and upper endpoints of the confidence interval for the log-log survival function, respectively then the confidence interval for $S(t)$ is

$$\left(\exp[-\exp(\widehat{C}_u)], \exp[-\exp(\widehat{C}_l)] \right).$$

These endpoints lie in the interval $(0, 1)$ and most statistical software packages report the confidence interval suggested by Kalbfleisch and Prentice (2002).

4.3.2 Non-parametric tests for comparing survival distributions

It is often of great interest to compare the survival prospects for different groups of patients, for example patients receiving different treatments. By using the KM estimators for the survival probabilities across the groups and plotting them together, it may be clear that there is a difference between the groups. However, from the pictorial representation, it might not be clear whether the difference has occurred only by chance or if there is actually a statistically significant difference. Hypothesis testing is a widely used and effective way to test whether there are some differences between groups.

Comparing two groups

Suppose that there are two types of treatments for a condition. Suppose that n_1 patients were randomly selected to receive treatment 1, and n_2 patients were randomly selected to receive treatment 2, for a total of $n = n_1 + n_2$ patients in the study. Let the survival functions of the two groups be $S_1(t)$ for treatment 1 and $S_2(t)$ for treatment 2.

The hypotheses for comparing two groups of patients are

$$H_0 : S_1(t) = S_2(t)$$

vs

$$H_1 : S_1(t) \neq S_2(t).$$

When there is no censoring, the above hypotheses can be tested using standard non-parametric tests. For example, to compare the survival experience of two groups, the Wilcoxon rank sum test or the Mann-Whitney U test can be used. In the presence of censoring, a number of tests have been suggested in the literature (Marubini and Valsecchi, 2004). The most frequently used of them is the Mantel-Haenszel (log-rank) test.

Combine two samples of sizes n_1 and n_2 into a single sample of size $n = n_1 + n_2$. Let $t_{(1)} < t_{(2)} < \ldots < t_{(m)}$ denote the m distinct ordered failure times in the combined sample, and d_j is the number of failures at time $t_{(j)}$ for $j \leq m$. Also let n_j be the number of patients at risk of failing just prior to time $t_{(j)}$. Further let n_{1j} and n_{2j} be the number of patients at risk just before $t_{(j)}$ in group

1 and group 2, respectively, and d_{1j} and d_{2j} are the number of failures in group 1 and group 2, respectively at time $t_{(j)}$. The data at time $t_{(j)}$ $(j = 1, 2, \ldots, m)$ can be arranged into a 2×2 contingency table as on Table 4.2.

Table 4.2: 2×2 contingency table at observed time $t_{(j)}$

Event	Samples		Total
	1	2	
Failed	d_{1j}	d_{2j}	d_j
Survived	$n_{1j} - d_{1j}$	$n_{2j} - d_{2j}$	$n_j - d_j$
Total	n_{1j}	n_{2j}	n_j

Mantel and Haenszel (1959) suggested a method of comparing the observed cell frequencies with the corresponding expected cell frequencies, where expected cell frequencies are computed under the condition of fixed row and column totals. Since the marginal totals in a 2×2 contingency table are fixed, only one cell frequency can be assigned randomly, say d_{1j}.

Under the null hypothesis and given the equal failure of probabilities, the distribution of d_{1j} is hyper-geometric (Marubini and Valsecchi, 2004) with parameters n_j, d_j and n_{1j}. Hence the mean and variance of d_{1j} are

$$E(d_{1j}) = n_{1j}\frac{d_j}{n_j} \quad \text{and} \quad \text{Var}(d_{1j}) = \frac{n_{1j}n_{2j}d_j(n_j - d_j)}{n_j^2(n_j - 1)},$$

respectively.

The log-rank statistic (without continuity correction) suggested by Mantel and Haenszel (1959) is

$$\chi^2_{M-H} = \frac{\{\sum_{j=1}^m [d_{1j} - E(d_{1j})]\}^2}{\sum_{j=1}^m \text{Var}(d_{1j})}, \tag{4.29}$$

which under H_0, approximately follows a chi-square distribution with 1 degree of freedom.

There are several tests that belong to the same family as the Mantel-Haenszel (log-rank) test, such as the Gehan, Tarone-Ware and Prentice tests (Marubini and Valsecchi, 2004). The Mantel-Haenszel chi-square test equally weights all the differences between the observed and the expected cell frequencies, while the other above mentioned tests weight the observed minus the expected cell frequencies differently at different points of failure.

Suppose w_j represents the weight at failure time $t_{(j)}$, then the statistic

$$\chi^2 = \frac{\{\sum_{j=1}^m w_j[d_{1j} - E(d_{1j})]\}^2}{\sum_{j=1}^m w_j^2 \text{Var}(d_{1j})}$$

is distributed as chi-square with 1 degree of freedom under the null hypothesis of equality of two survival distributions.

The Mantel-Haenszel statistic considers unweighted differences between the observed and the expected cell frequencies, i.e., $w_j = 1$ $(j = 1, 2, \ldots, m)$. However, when the assumption of Cox proportional hazards over the follow-up period is violated, log-rank test is less powerful than the Gehan or Tarone-Ware (Tarone and Ware, 1977) tests. The Gehan (1965) statistic (χ_G^2) uses n_j as the weight at time point $t_{(j)}$. That is, the Gehan statistic gives more emphasis to earlier failures. This test assumes that the censoring distributions of both groups are the same. Tarone and Ware (1977) showed that when censoring distributions widely differ then Gehan's test performs poorly. The weights used

in the [Tarone and Ware (1977)](#) statistic (χ^2_{T-W}) are the geometric means of the weights considered in χ^2_{M-H} and χ^2_G statistics.

Comparing more than two groups

Suppose, we have a sample from each of G $(G > 2)$ treatment groups (populations). We are interested in comparing the survival experience of G populations, i.e., our null hypothesis is

$$H_0 : S_1(t) = S_2(t) = \ldots = S_G(t) \tag{4.30}$$

against the alternative that at least two of the survival functions $S_g(t)$ are unequal, i.e.,

$$H_1 : S_g(t) \neq S_l(t) \quad \text{for some} \quad g \neq l.$$

The following test is an extension of the Mantel-Haenszel chi-square test for two groups. Suppose the G samples are combined into a single sample of size $n = n_1 + n_2 + \ldots + n_G$, and $t_{(1)} < t_{(2)} < \ldots < t_{(m)}$ are the ordered failure times in the combined sample. Instead of using a 2×2 contingency table at $t_{(j)}$ for comparing two groups, as in the Mantel-Haenszel chi-square test, we need to form a $2 \times G$ contingency table at $t_{(j)}$ for comparing G populations.

The notation used in the $2 \times G$ contingency Table [4.3](#) is similar to the notation used in Table [4.2](#).

The log-rank test for G populations also compares the observed and ex-

Table 4.3: $2 \times G$ contingency table at observed time $t_{(j)}$

Event	Samples						Total
	1	2	$\ldots$	g	$\ldots$	G	
Failure	d_{1j}	d_{2j}	$\ldots$	d_{gj}	$\ldots$	d_{Gj}	d_j
Survive	$n_{1j} - d_{1j}$	$n_{2j} - d_{2j}$	$\ldots$	$n_{gj} - d_{gj}$	$\ldots$	$n_{Gj} - d_{Gj}$	$n_j - d_j$
Total	n_{1j}	n_{2j}	$\ldots$	n_{gj}	$\ldots$	n_{Gj}	n_j

pected cell frequencies of the number of failures for any $(G-1)$ cells out of G cells. The distribution of the observed number of failures follows a $(G-1)$ dimensional multivariate hyper-geometric distribution under the null hypothesis and fixed marginal totals (Marubini and Valsecchi, 2004). The expected number of failures at $t_{(j)}$ in the g-th sample is

$$E(d_{gj}) = n_{gj} \frac{d_j}{n_j}.$$

Hence the sum of the differences between d_{gj} and $E(d_{gj})$ for the g-th sample is (summing over all time points)

$$U_g = \sum_{j=1}^{m} [d_{gj} - E(d_{gj})].$$

The $(G-1) \times 1$ vector with elements U_g, using the G-th sample as the reference, is

$$\boldsymbol{U} = (U_1, U_2, \ldots, U_g, \ldots, U_{G-1})^\top.$$

The test statistic requires the $(G-1) \times (G-1)$ dimensional variance-covariance matrix $\boldsymbol{\Sigma}(t_{(j)})$ of $\boldsymbol{U}$ at $t_{(j)}$. The (g,l)-th component of $\boldsymbol{\Sigma}(t_{(j)})$ is

$$\boldsymbol{\Sigma}_{g,l}(t_{(j)}) = \begin{cases} \dfrac{n_{gj}(n_j - n_{gj})d_j(n_j - d_j)}{n_j^2(n_j - 1)}; & g = l \\[3mm] -\dfrac{n_{gj}n_{lj}d_j(n_j - d_j)}{n_j^2(n_j - 1)}; & g \neq l \end{cases}$$

where $g, l = 1, 2, \ldots, (G-1)$. Then the variance-covariance matrix of $\boldsymbol{U}$ is (summing over k failure times)

$$\boldsymbol{\Sigma} = \sum_{j=1}^{m} \boldsymbol{\Sigma}(t_{(j)}).$$

The Mantel-Haenszel test statistic for comparing G populations is a quadratic form

$$\chi^2_{M-H} = \boldsymbol{U}^{\top} \boldsymbol{\Sigma}^{-1} \boldsymbol{U}. \tag{4.31}$$

This statistic approximately follows a chi-square distribution with $(G-1)$ degrees of freedom under the null hypothesis. This test statistic reduces to test statistic given in Equation (4.29) if $G = 2$.

Similar to the log-rank test for comparing two populations this general version of the log-rank test is one of a family of chi-square tests for comparing G groups of populations. The alternative test statistics are formed by calculating the weighted differences between the observed and the expected number of failures. Suppose $w_1, w_2, \ldots, w_m$ are the weights for the time points $t_{(1)}, t_{(2)}, \ldots, t_{(m)}$. Then the test statistic for comparing G populations becomes

$$\chi^2_{W} = \boldsymbol{U}_W^{\top} \boldsymbol{\Sigma}_W^{-1} \boldsymbol{U}_W \tag{4.32}$$

where $U_{(W)g} = \sum_{j=1}^{m} w_j [d_{gj} - E(d_{gj})]$ and $\boldsymbol{\Sigma}_W$ is the variance-covariance matrix of $\boldsymbol{U}_W$. This statistic also approximately follows a chi-square distribution with $(G-1)$ degrees of freedom.

4.4 Cox proportional hazards model

If two or more groups of subjects in the study differ only by the treatment of interest, then non-parametric methods such as Kaplan-Meier estimation and log-rank test are useful to compare the survival probabilities among groups. However, in most cases, the subjects in the groups differ by some other additional characteristics, such as demographical variables (e.g., age, sex, socio-economic status, or education), behavioral variables (e.g., smoking history, alcohol consumption, dietary habits, physical activity level) or physiological variables (e.g., blood pressure, blood glucose level, heart rate) that may affect their outcome. In these cases, such variables may be used as covariates of interest in explaining the response variable. Once these potential covariates are adjusted for, the comparison of survival times among groups should be less biased and more precise than a simple comparison such as non-parametric method. In this section, a semiparametric regression model, the Cox proportional hazards model which was used in this study is introduced (Cox, 1972). This model allows to quantify the association between the time to event and a set of specified predictor/explanatory variables.

The Cox proportional hazards (PH) model is a widely applied semiparametric statistical model for analyzing survival data. In medical research, it is commonly used to investigate the relationship between the survival time of patients, and one or more predictor variables. The model allows to assess the effect of several risk factors on survival time simultaneously. The formula for the Cox PH model

is usually written in terms of the hazard function, $h(t, \boldsymbol{X})$ as:

$$h(t, \boldsymbol{X}) = h_0(t)\exp\left(\sum_{k=1}^{p} \beta_k X_k\right). \qquad (4.33)$$

Here, the covariate vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)$ is a collection of time independent predictor variables that are used to predict an individual's hazard associated with those variables. The unknown coefficients $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_p)^{\top}$ measure the effect of the predictor variables that we want to estimate on the log-hazard scale. The $h_0(t)$ is the unspecified baseline hazard function that reflects the changes in hazard function, $h(t, \boldsymbol{X})$ over time, and the term $\exp\left(\sum_{k=1}^{p} \beta_k X_k\right)$ characterizes how hazard function changes with the covariates. The Cox PH model is termed a semiparametric model because of its unknown baseline hazard function $h_0(t)$. The model (4.33) provides an expression for the hazard for an individual with the given set of explanatory variables at time t. Let $X_{i1}, X_{i2}, \ldots, X_{ip}$ be the set of values of covariates for the ith individual. The hazard function for this individual can be written as:

$$h_i(t, \boldsymbol{X}) = h_0(t)\exp(\beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip}). \qquad (4.34)$$

The linear component, $(\beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip})$ in the model is known as risk score or prognostic index for the *ith* individual. According to the Cox PH model, the hazard ratio for any two individuals with covariate vectors $\boldsymbol{X} = (X_1, X_2, ..., X_p)$ and $\boldsymbol{X}^* = (X_1^*, X_2^*, ..., X_p^*)$ does not depend on time. This property is called the proportional hazards (PH) assumption in the Cox model. Hazard ratio can be estimated from the risk scores. Calculation of hazard ratio will be discussed in the next subsection.

Explanatory variables on which a hazard function may depend, are subdivided into covariates and factors. A covariate is a continuous predictor variable that takes the numerical values, such as age, body mass index or systolic blood pressure. A factor is a categorical variable that takes a limited set of values, which are known as the levels of the factor. For example, smoking might be a factor with three levels, such as ex-smoker, current smoker, and non-smoker. Covariates and their combinations, are readily incorporated in the linear component in the Cox PH model. Factors having two levels are also directly used in the model, whereas factors having more than two categories need to be converted into dummy or indicator variables.

The baseline hazard, $h_0(t)$ in the Cox PH formula (4.34) is a function of time t, but it does not involve the covariates $\boldsymbol{X}$. In contrast, the exponential expression involves the explanatory variables $\boldsymbol{X}$, but it does not involve the time t. So, the covariates $\boldsymbol{X}$ and their effects are time-independent. If all the covariates $\boldsymbol{X}$ are equal to zero, i.e., $X_1 = X_2 = ... = X_p = 0$ then the Cox model (4.33) reduces to the baseline hazard function $h_0(t)$. This property of the Cox model is the reason why $h_0(t)$ is called the baseline hazard function. One important feature of the Cox PH model is that the vector of parameters $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_p)^{\top}$ can be estimated without specifying the baseline hazard function $h_0(t)$. Thus, using the Cox PH model, with a minimum of assumptions, the primary information desired from a survival analysis can be obtained.

An important advantage of the Cox PH model is that when the PH as-

sumptions are satisfied, it gives reliable results, and the users do not need to worry about the choice of a parametric model. That is, the Cox PH model is robust for the parametric distribution. This is why it is referred to as a semi-parametric model. The exponential form of the hazard function ensures that the fitted model always provides positive estimated hazard.

4.4.1   Interpreting the hazard ratios

A hazard ratio (HR) in the Cox PH model is the ratio of the hazard function for one individual to the hazard function for the other individual. If $\boldsymbol{X}^* = \left(X_1^*, X_2^*, ..., X_p^*\right)$ is the set of predictor variables for one individual in the exposed group, and $\boldsymbol{X} = (X_1, X_2, ..., X_p)$ for another individual in the unexposed group, then the hazard ratio is defined by

$$\text{HR} = \frac{h(t, \boldsymbol{X}^*)}{h(t, \boldsymbol{X})} = \frac{h_0(t) \exp\left[\sum_{k=1}^p \beta_k X_k^*\right]}{h_0(t) \exp\left[\sum_{k=1}^p \beta_k X_k\right]}. \tag{4.35}$$

Thus, once the values for any two sets of predictors, $\boldsymbol{X}^*$ and $\boldsymbol{X}$ are specified, the value of the estimated hazard ratio is constant and it does not depend on time. Equation (4.35) states the proportional hazard assumption in the Cox model. Hence, the hazard ratio is estimated as

$$\widehat{\text{HR}} = \exp\left[\sum_{k=1}^p \widehat{\beta}_k (X_k^* - X_k)\right]. \tag{4.36}$$

An important feature of the hazard ratios in Cox PH model is that they are adjusted for all other covariates in the model.

If the two individuals differ by only one covariate X_k which denotes the exposure status, i.e., $X_k^* = 1$ and $X_k = 0$, then

$$\widehat{\text{HR}} = \exp(\widehat{\beta_k}), \tag{4.37}$$

where $\widehat{\beta_k}$ is the estimated log-hazard ratio of X_k in the model. If $\widehat{\beta_k}$ is negative, the value of the hazard ratio is less than one, then the exposure is associated with decreased risk and longer survival time for the exposed group. If $\widehat{\beta_k}$ is zero, $\widehat{\text{HR}}$ is equal to one, then there is no difference in survival between groups. When $\widehat{\beta_k}$ is positive, the hazard ratio is greater than one, and then the exposed group has the increased risk and shorter survival time. However, it is important to note that the hazard ratio is a comparison between groups only and thus it does not give any indication of exactly how long it will take for an individual in either group to experience the event.

4.4.2 Parameter estimation

The parameters in the general Cox PH model in Equation (4.33) are denoted by β_k's. The estimates of these parameters, $\widehat{\beta_k}$'s are obtained through the maximum likelihood estimation (MLE). The estimated parameters $\widehat{\beta_k}$'s are the estimated effects of the respective covariates in the model. The MLE of the Cox PH model parameters is performed by maximizing the partial likelihood function $L(\boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is the collection of unknown parameters β_k. The likelihood function used in the inference for the Cox model is called "partial" likelihood as the

likelihood only includes probabilities for the subjects who are not censored, that means the subjects who only got the event (Kleinbaum and Klein, 2012).

Suppose that there are n individual observations in the study. Let $t_{(1)} < t_{(2)} < \ldots < t_{(m)}$ denote the m distinct ordered times at which the events happened, and $\boldsymbol{X}_{(j)} = (X_{j1}, X_{j2}, \ldots, X_{jp})$ be the values of the explanatory variables for the individual who experienced the event at the jth ordered time $t_{(j)}$ $(j \leq m)$. First assume that there are no ties in the event times, i.e., not more than one subject has experienced the event at the same time point. This means that m subjects experienced the event, and the remaining $(n-m)$ observations are right censored. Then the jth term of the partial likelihood function takes the form

$$
\begin{aligned}
L_j(\boldsymbol{\beta}) &= \frac{\exp\left(\sum_{k=1}^{p} \beta_k X_{(j)k}\right)}{\sum_{l \in R(t_{(j)})} \exp\left(\sum_{k=1}^{p} \beta_k X_{lk}\right)} \\
&= \frac{\exp(\boldsymbol{\beta}^\top \boldsymbol{X}_{(j)})}{\sum_{l \in R(t_{(j)})} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_l)},
\end{aligned}
\tag{4.38}
$$

where $\boldsymbol{\beta}^\top$ is the transposed vector of regression coefficients β_k, and $\sum_{l \in R(t_{(j)})} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_l)$ is the sum of the hazards for members of the risk set $R(t_{(j)})$. The risk set $R(t_{(j)})$ contains the subjects who are still at risk at time $t_{(j)}$, i.e., the subjects who are not censored before that time. Then the partial likelihood function used to obtain the maximum likelihood estimator is written as

$$
L(\boldsymbol{\beta}) = \prod_{j=1}^{m} L_j(\boldsymbol{\beta}) = \prod_{j=1}^{m} \frac{\exp(\boldsymbol{\beta}^\top \boldsymbol{X}_{(j)})}{\sum_{l \in R(t_{(j)})} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_l)}.
\tag{4.39}
$$

The maximum likelihood estimation of $\boldsymbol{\beta}$ using the partial likelihood function is then obtained by maximizing the natural logarithm of the partial likeli-

hood function $L(\boldsymbol{\beta})$ (Hosmer et al., 2011):

$$\log L(\boldsymbol{\beta}) = l(\boldsymbol{\beta}) = \sum_{j=1}^{m} \left[\boldsymbol{\beta}^{\top} \boldsymbol{X}_{(j)} - \log \left(\sum_{l \in R(t_{(j)})} \exp(\boldsymbol{\beta}^{\top} \boldsymbol{X}_l) \right) \right]. \qquad (4.40)$$

The maximization process is carried out by taking partial derivatives of $l(\boldsymbol{\beta})$ with respect to β_k's $(k = 1, 2, \ldots, p)$, setting all the derivatives equal to zero, and then solving a system of p partial differential equations for the unknown parameters. The partial derivative of (4.40) with respect to $\boldsymbol{\beta}$ is called the score vector, and denoted by

$$\boldsymbol{U}(\boldsymbol{\beta}) = \frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}. \qquad (4.41)$$

The kth element of the score vector $\boldsymbol{U}(\boldsymbol{\beta})$ is written as:

$$
\begin{aligned}
U_i(\boldsymbol{\beta}) &= \frac{\partial l(\boldsymbol{\beta})}{\partial \beta_k} \\
&= \sum_{j=1}^{m} \left(X_{(j)k} - \frac{\sum_{l \in R(t_{(j)})} X_{lk} \exp(\sum_{k=1}^{p} \beta_k X_{lk})}{\sum_{l \in R(t_{(j)})} \exp(\sum_{k=1}^{p} \beta_k X_{lk})} \right) ; k = 1, 2, \ldots, p \\
&= \sum_{j=1}^{m} \left(X_{(j)k} - \frac{\sum_{l \in R(t_{(j)})} X_{lk} \exp(\boldsymbol{\beta}^{\top} \boldsymbol{X}_l)}{\sum_{l \in R(t_{(j)})} \exp(\boldsymbol{\beta}^{\top} \boldsymbol{X}_l)} \right) \\
&= \sum_{j=1}^{m} \left(X_{(j)k} - A_{(j)k} \right),
\end{aligned} \qquad (4.42)
$$

where

$$A_{(j)k} = \frac{\sum_{l \in R(t_{(j)})} X_{lk} \exp(\boldsymbol{\beta}^{\top} \boldsymbol{X}_l)}{\sum_{l \in R(t_{(j)})} \exp(\boldsymbol{\beta}^{\top} \boldsymbol{X}_l)}. \qquad (4.43)$$

The solution of score equation $\boldsymbol{U}(\boldsymbol{\beta}) = \boldsymbol{0}$ is carried out using iterative methods, such as Newton-Raphson technique which requires the score vector, $\boldsymbol{U}(\boldsymbol{\beta})$, and observed information matrix $\mathbf{I}(\boldsymbol{\beta})$. The score vector $\boldsymbol{U}(\boldsymbol{\beta})$ in Equation (4.41) is of order $p \times 1$, and information matrix $\mathbf{I}(\boldsymbol{\beta})$ is the $p \times p$ matrix of negative second

derivatives of the partial log-likelihood, i.e $\mathbf{I}(\boldsymbol{\beta})$ is given by

$$\mathbf{I}(\boldsymbol{\beta}) = -\frac{\partial^2 l(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}. \tag{4.44}$$

The (k, h)th element of the information matrix in Equation (4.44) is given by

$$
\begin{aligned}
I_{k,h}(\boldsymbol{\beta}) &= -\frac{\partial^2 l(\boldsymbol{\beta})}{\partial\beta_k\partial\beta_h} \\
&= \sum_{j=1}^{m} \left[\frac{\sum_{l\in R(t_{(j)})} X_{lk}X_{lh}\exp(\boldsymbol{\beta}^\top \boldsymbol{X}_l)}{\sum_{l\in R(t_{(j)})}\exp(\boldsymbol{\beta}^\top \boldsymbol{X}_l)} - A_{(j)k}A_{(j)h} \right] ; (k, h) = (1, 2, \ldots, p)
\end{aligned}
\tag{4.45}
$$

where $A_{(j)k}$ is given by Equation (4.43).

The Newton-Rapson method starts with a reasonable estimate of $\boldsymbol{\beta}$, say $\boldsymbol{\beta}_0$, for the solution, and then it successively modifies the estimate until a final solution of $\hat{\boldsymbol{\beta}}$ is obtained.

The first improved estimate of $\boldsymbol{\beta}$ is:

$$\widehat{\boldsymbol{\beta}}_1 = \boldsymbol{\beta}_0 + [\mathbf{I}(\boldsymbol{\beta}_0)]^{-1}\boldsymbol{U}(\boldsymbol{\beta}_0). \tag{4.46}$$

Then the second improved estimate of $\boldsymbol{\beta}$ is:

$$\widehat{\boldsymbol{\beta}}_2 = \widehat{\boldsymbol{\beta}}_1 + [\mathbf{I}(\widehat{\boldsymbol{\beta}}_1)]^{-1}\boldsymbol{U}(\widehat{\boldsymbol{\beta}}_1). \tag{4.47}$$

So for the the nth iteration:

$$\widehat{\boldsymbol{\beta}}_n = \widehat{\boldsymbol{\beta}}_{(n-1)} + [\mathbf{I}(\widehat{\boldsymbol{\beta}}_{(n-1)})]^{-1}\boldsymbol{U}(\widehat{\boldsymbol{\beta}}_{(n-1)}). \tag{4.48}$$

The above procedure is repeated until it converges to $|\widehat{\boldsymbol{\beta}}_n - \widehat{\boldsymbol{\beta}}_{(n-1)}| < \epsilon$, i.e., the successive values of the estimated coefficients $\widehat{\boldsymbol{\beta}}$ differ only by a very small

number ϵ. The estimate of the variance of $\widehat{\boldsymbol{\beta}}$ is obtained as the inverse of the observed information matrix, i.e.,

$$\widehat{\text{var}}(\widehat{\boldsymbol{\beta}}) = [\mathbf{I}(\widehat{\boldsymbol{\beta}})]^{-1}. \tag{4.49}$$

Although the baseline hazard function in the Cox PH model is not specified, parametrically, it can be estimated once the parameters are estimated from the partial likelihood function by using the following formula (Breslow, 1974):

$$\widehat{h_0}(t_{(j)}) = \frac{1}{\sum_{l \in R(t_{(j)})} \exp(\hat{\boldsymbol{\beta}}^\top \boldsymbol{X}_l)}. \tag{4.50}$$

However, in practice, the estimator of the cumulative baseline hazard function is widely used as it provides a more stable estimate than the estimated baseline hazard function (Hosmer et al., 2011). The Breslow's estimator of the baseline cumulative hazard function is:

$$\widehat{H_0}(t) = \sum_{t_{(j)} \leq t} \frac{1}{\sum_{l \in R(t_{(j)})} \exp(\hat{\boldsymbol{\beta}}^\top \boldsymbol{X}_l)}. \tag{4.51}$$

When dealing with a large dataset, it is more likely to have ties in events times. An exact expression for partial likelihood in case of tied failures is derived by Kalbfleisch and Prentice (2002). The exact partial likelihood of Kalbfleisch and Prentice involves the permutation of tied failures at particular time points and hence makes the computation very time consuming. However, several approximations to the partial likelihood that are obtained by modifying the partial likelihood for non-tied case have been proposed, for example Breslow (1974) and Efron (1977).

Let $d_{(j)}$ individuals fail at time $t_{(j)}$ and $D_{(j)}$ denotes the set of these individuals who fail at time $t_{(j)}$. Let $\boldsymbol{X}_{(j+)}$ be the sum of the covariate vectors over $d_{(j)}$ failures at $t_{(j)}$, i.e., $\boldsymbol{X}_{(j+)} = \sum_{j \in D_{(j)}} \boldsymbol{X}_{(j)}$. Then the approximate partial likelihood proposed by Breslow (1974) is

$$L_1(\boldsymbol{\beta}) = \prod_{j=1}^{m} \frac{\exp(\boldsymbol{\beta}^\top \boldsymbol{X}_{(j+)})}{\left[\sum_{l \in R(t_{(j)})} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_l) \right]^{d_{(j)}}}. \tag{4.52}$$

This approximation works well when the number of ties are few in number (Klein and Moeschberger, 2003).

An alternative approximate partial likelihood suggested by Efron (1977) is

$$L_2(\boldsymbol{\beta}) = \prod_{j=1}^{m} \frac{\exp(\boldsymbol{\beta}^\top \boldsymbol{X}_{(j+)})}{\prod_{u=1}^{d_{(j)}} \left[\sum_{l \in R(t_{(j)})} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_l) - \frac{u-1}{d_{(j)}} \sum_{l \in D_{(j)}} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_l) \right]}. \tag{4.53}$$

This likelihood is closer to the exact partial likelihood than Breslow's likelihood (Klein and Moeschberger, 2003).

The maximum likelihood estimator of $\boldsymbol{\beta}$ can be obtained in the same manner as in the non-tied case, i.e., by using the Newton-Raphson iterative procedure. The expressions for the score vector and observed information matrix (as in equations (4.41) and (4.44), respectively) can be obtained using either Breslow's or Efron's partial likelihood. The default partial likelihood used in most statistical software is Efron's likelihood (Therneau and Lumley, 2020).

The baseline hazard and cumulative baseline hazard function for the time-

tied event are estimated by using the Breslow's estimator as follows:

$$\widehat{h}_0(t_{(j)}) = \frac{1}{\sum_{l \in R(t_{(j)})} \exp(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_l)]^{d_{(j)}}}, \tag{4.54}$$

and

$$\widehat{H}_0(t) = \sum_{t_{(j)} \leq t} \frac{d_{(j)}}{[\sum_{l \in R(t_{(j)})} \exp(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_l)]^{d_{(j)}}}, \tag{4.55}$$

respectively.

4.4.3 Inference in the Cox model

In the previous section the parameter estimation procedures of the Cox regression model were discussed. The parameters are estimated by maximizing the partial likelihood function. Let $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_1, \ldots, \widehat{\beta}_p)^\top$ denote the (partial) maximum likelihood estimators of the regression parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$, and $\boldsymbol{\beta}_0 = (\beta_1^0, \ldots, \beta_p^0)^\top$ be a set of particular values of the regression parameters. In this subsection, the test procedure for all parameters and a subset of parameters in the Cox model are discussed.

Global tests

There are three commonly used methods of testing the global (overall) hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ in a regression model, namely the Wald's test, the likelihood ratio test and the score test. These tests are also applicable to test the hypotheses about parameters $\boldsymbol{\beta}$ of the Cox regression model when parameters are estimated using partial likelihood function (Klein and Moeschberger, 2006).

The first test, the Wald's test, is based on the large sample properties of the (partial) maximum likelihood estimators of $\boldsymbol{\beta}$. For large samples, $\widehat{\boldsymbol{\beta}}$ is distributed as a random variable from a p-variate normal distribution with mean vector $\boldsymbol{\beta}$ and estimated variance-covariance matrix $\mathbf{I}(\widehat{\boldsymbol{\beta}})^{-1}$, where $\mathbf{I}(\widehat{\boldsymbol{\beta}})$ is the $p \times p$ information matrix evaluated at $\widehat{\boldsymbol{\beta}}$. The Wald's statistic for testing $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ is given by

$$\chi^2_W = (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^{\top} \mathbf{I}(\widehat{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0). \tag{4.56}$$

Under H_0, the Wald's statistic follows a chi-square distribution with p degrees of freedom.

The likelihood ratio test uses the statistic

$$\chi^2_{LR} = 2[l(\widehat{\boldsymbol{\beta}}) - l(\boldsymbol{\beta}_0)], \tag{4.57}$$

where $l(\boldsymbol{\beta}_0)$ denotes the (partial) log-likelihood under H_0 and $l(\widehat{\boldsymbol{\beta}})$ is the (partial) log-likelihood evaluated at $\widehat{\boldsymbol{\beta}}$. The likelihood ratio statistic also follows the chi-square distribution with p degrees of freedom under H_0.

The score test is based on the asymptotic distribution of the score vector $\boldsymbol{U}(\boldsymbol{\beta}) = (U_1(\boldsymbol{\beta}), \ldots, U_p(\boldsymbol{\beta}))^{\top}$, where the kth element $U_k(\boldsymbol{\beta})$ is defined in equation (4.42). For large samples, the score statistic $\boldsymbol{U}(\boldsymbol{\beta})$ has approximately a p-variate normal distribution with mean vector $\mathbf{0}$, and variance-covariance matrix $\mathbf{I}(\boldsymbol{\beta})$. The information matrix $\mathbf{I}(\boldsymbol{\beta})$ is defined by equation (4.44). The score statistic for testing $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ is

$$\chi^2_{SC} = \boldsymbol{U}(\boldsymbol{\beta}_0)^{\top} \mathbf{I}^{-1}(\boldsymbol{\beta}_0) \boldsymbol{U}(\boldsymbol{\beta}_0), \tag{4.58}$$

which also follows the chi-square distribution with p degrees of freedom if H_0 is true.

The p-values of the above mentioned test statistics are compared with the critical values of χ_p^2 distribution at the pre-specified level of significance to make decision about the overall hypothesis. The null hypothesis is usually set as $H_0 : \boldsymbol{\beta} = \mathbf{0}$ as acceptance of this null hypothesis indicates that the covariates are not associated with the survival.

Local tests

Sometimes researchers are interested in making inference about a subset of the parameters rather than the full set of parameters. Suppose the parameters vector $\boldsymbol{\beta}$ is partitioned as $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)^\top$, where $\boldsymbol{\beta}_1$ is a $q \times 1$ vector and $\boldsymbol{\beta}_2$ is a $(p-q) \times 1$ vector. Suppose the hypothesis of interest is $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^0$, where $\boldsymbol{\beta}_1^0$ is a specific $q \times 1$ vector. The vector $\boldsymbol{\beta}_2$ consists of nuisance parameters. The above hypothesis can be tested using the Wald's test, the likelihood ratio test and the score test.

Suppose the corresponding partitioning of (partial) maximum likelihood estimators, score vector and the information matrix are $\widehat{\boldsymbol{\beta}} = (\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2)^\top$, $\boldsymbol{U}(\boldsymbol{\beta}) = (\boldsymbol{U}_1(\boldsymbol{\beta}), \boldsymbol{U}_2(\boldsymbol{\beta}))^\top$ and

$$
\mathbf{I} = \begin{pmatrix} \mathbf{I}_{11} & \mathbf{I}_{12} \\ \mathbf{I}_{21} & \mathbf{I}_{22} \end{pmatrix},
$$

respectively, where $\mathbf{I}_{11}$ is a $q \times q$ matrix, $\mathbf{I}_{22}$ is a $(p-q) \times (p-q)$ matrix, $\mathbf{I}_{12}$ is a

$q \times (p-q)$ matrix and $\mathbf{I}_{21}^\top = \mathbf{I}_{12}$. The inverse of the information matrix can also be partitioned as

$$\mathbf{I}^{-1} = \begin{pmatrix} \mathbf{I}^{11} & \mathbf{I}^{12} \\ \mathbf{I}^{21} & \mathbf{I}^{22} \end{pmatrix}.$$

Suppose that $\widehat{\boldsymbol{\beta}}_2(\boldsymbol{\beta}_1^0)$ is the (partial) maximum likelihood estimator of $\boldsymbol{\beta}_2$ obtained by maximizing the likelihood function $L(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2)$ with respect to $\boldsymbol{\beta}_2$.

Then the test statistics for the above mentioned tests for testing $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^0$ are:

the Wald's test:

$$\chi_W^2 = (\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^0)^\top [\mathbf{I}^{11}(\widehat{\boldsymbol{\beta}})]^{-1}(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^0),$$

the Likelihood ratio test:

$$\chi_{LR}^2 = 2[l(\widehat{\boldsymbol{\beta}}) - l\{\boldsymbol{\beta}_1^0, \widehat{\boldsymbol{\beta}}_2(\boldsymbol{\beta}_1^0)\}], \text{and}$$

the score test:

$$\chi_{SC}^2 = \boldsymbol{U}_1[\boldsymbol{\beta}_1^0, \widehat{\boldsymbol{\beta}}_2(\boldsymbol{\beta}_1^0)]^\top \mathbf{I}^{11}[\boldsymbol{\beta}_1^0, \widehat{\boldsymbol{\beta}}_2(\boldsymbol{\beta}_1^0)] \boldsymbol{U}_1[\boldsymbol{\beta}_1^0, \widehat{\boldsymbol{\beta}}_2(\boldsymbol{\beta}_1^0)].$$

When H_0 is true, all three statistics have large sample chi-square distribution with q degrees of freedom.

To test the significance of a particular covariate X_j, i.e., whether individual covariate is significantly associated with the survival time ($H_0 : \beta_j = 0$), any of the local tests with $q = 1$ can be used. Alternatively, a Z- statistic can be

computed

$$Z = \frac{\widehat{\beta}_j}{\widehat{SE}(\widehat{\beta}_j)}. \tag{4.59}$$

where $\widehat{SE}(\widehat{\beta}_j)$ is the estimated standard error of $\widehat{\beta}_j$. The standard error, $\widehat{SE}(\widehat{\beta}_j)$, is the square root of the jth diagonal element of $\mathbf{I}(\widehat{\boldsymbol{\beta}})$. The Z- statistic has approximately a standard normal distribution, and thus can be used to obtain the p-value associated with the Z test, and can also be used to obtain a confidence interval for β_j (Hosmer et al., 2011). The $100(1 - \alpha)\%$ confidence interval for β_j is

$$(\widehat{\beta}_j - Z_{(1-\alpha/2)}\widehat{SE}(\widehat{\beta}_j), \widehat{\beta}_j + Z_{(1-\alpha/2)}\widehat{SE}(\widehat{\beta}_j)), \tag{4.60}$$

where $Z_{(1-\alpha/2)}$ is the upper $100(1 - \alpha/2)$th percentile of the standard normal distribution. In most statistical packages, Z- statistics and associated p-values are readily implemented.

4.5   Selection of a significant subset of covariates

In clinical research, there could be a large number of possible predictor variables to be included in the prediction model. However, not all of the considered covariates are related to the outcome and hence analysts must follow an appropriate strategy to select the best subset of significant covariates. In this section, various methods of variable selection with the stopping rules that can be applied in model development are introduced.

The process of selecting the subset of covariates from a large pool of covari-

ates compromises between two conflicting objectives (Montgomery et al., 2015): (i) the model should include as many covariates as possible since inclusion of additional covariates usually increase the explanatory power of the model, and (ii) model should have as few covariates as possible because the variance of the predicted outcome increases with the increase in the number of covariates. Also models with more covariates require more computing time and computer memory. According to Montgomery et al. (2015), there are two methods of selecting the subset of covariates: (i) all possible regressions, and (ii) stepwise regression methods. There are three versions of stepwise regression methods: forward selection, backward elimination and forward stepwise selection procedures. These methods can also be used to select covariates in a proportional hazards regression model (Hosmer et al., 2011). Deciding on whether one subset is better than another is based on a selected information criterion or a ststistical rule (Harrell Jr, 2015), and model with the lowest value of the chosen statistics should be selected (Hastie et al., 2015; Montgomery et al., 2015).

The Akaike information criterion (AIC) and Bayesian information criterion (BIC) are two most commonly used information criteria (Montgomery et al., 2015). The AIC and BIC are defined as

$$\text{AIC} = -2\text{log}(L) + 2p \tag{4.61}$$

and

$$\text{BIC} = -2\text{log}(L) + p\text{log}(m) \tag{4.62}$$

where log denotes the natural logarithm, L is the partial likelihood of the Cox

model, p is the number of parameters estimated in the model and m is the number of uncensored events. The modification to use number of uncensored observations in the Cox model instead of number of observations was suggested by Volinsky and Raftery (2000). BIC places a larger penalty for large number of covariates as the sample size increases and usually select simpler models (Hastie et al., 2015).

4.5.1   All possible regressions

All possible regressions method fits all the survival models having one covariate, two covariates, and so on. If there are p covariates, then a total of $2^p - 1$ survival models needs to be fitted and examined to choose the best model. The AIC or BIC is to be calculated for each model and the model having minimum AIC or BIC in a subset model size (models having same number of covariates) is the best among the subset models. The overall best model is the one that has minimum AIC or BIC in that subset model size and increasing the model size decreases AIC or BIC to a negligible amount. The number of models to be fitted increases rapidly with the increase in the number of covariates and this method is computationally demanding. Nowadays, the availability of high-speed computers make the all possible regressions procedure feasible.

4.5.2 Stepwise regression methods

Stepwise regression procedures use some measure of fit to add a covariate to the model or delete a covariate from the model. The most widely used stopping rule is based on the partial likelihood ratio test (Hosmer et al., 2011). There are some other stopping rules that are based on AIC or BIC (Harrell Jr, 2015). The forward selection procedure starts with the null model that is the model that has no covariates and then adds variables one at a time. The first candidate covariate for entry into the model, X_l, is the one that has the biggest additional contribution to the null model. If stopping rule is not reached, X_l is entered into the model and program goes to next step; otherwise it stops and adopt the null model as the best model. The second candidate covariate for entry is the one that has the highest additional contribution to the model containing X_l. Suppose the candidate covariate for entry into the model is X_m. If stopping rule is not reached, it adopts the model having two covariates X_l and X_m and proceeds to next step; otherwise it stops and adopts the model containing only one covariate X_l. The process of adding new candidate covariates to the model continues until the stopping rule is met or no covariate is remaining to add to the model.

The backward elimination techniques is the opposite process to the forward selection procedure. It starts with the full model, i.e, the model containing all covariates and deletes covariate one at a time. The first candidate covariate to be dropped is the one that has the smallest contribution to the model fit.

If stopping rule is not reached, drop that covariate from the model. Next fit a new model with the remaining $p - 1$ covariates and identify the second candidate covariate to be removed that contributes the least to the model fit and the procedure is repeated. The process terminates when no further covariates can be dropped. The forward stepwise selection procedure is the modified version of forward selection procedure. At each step, the previously added covariates are reassessed on whether they should be dropped from the model. All stepwise regression methods have the unavoidable problem of including unimportant covariates and excluding important covariates (Hosmer et al., 2011). Backward elimination has the advantage over other stepwise regression methods in assessment of the joint predictive ability of covariates as the process starts with the full model (Chowdhury and Turin, 2020).

In this research, backward elimination was carried out to select the best model. The initial model contains the main effects of all covariates and also their two-way interaction. The backward selection was performed manually, i.e., the model with all of the selected variables was fitted first. Then the likelihood ratio (LR) χ^2 test was used to check whether a variable should be dropped from the model based on the cut-off point of p-value. If any insignificant variable were found, the model was fitted again after dropping that variable and performing the same test again. The process continues until all remaining effects are found to be significant. The widely accepted choice for the significance level of p-value is 0.05 or 0.10. The range of p-values from 0.15 to 0.20 is also recommended

(Hosmer Jr et al., 2013). However, choosing a higher cut-off point for p-values may select some irrelevant covariates (Chowdhury and Turin, 2020).

4.5.3 Assessment of the proportional hazards assumption

The use and interpretation of proportional hazards model is valid when the hazard between subgroups of subjects are are independent of time. A large number of graphical and numerical procedures for assessing the proportional hazards assumption have been proposed in the literature (Hosmer et al., 2011). The graphical methods involve subjectivity in interpretation (Persson, 2002) and are not informative when a covariate has several levels or is continuous (Xue and Schifano, 2017). The most popular graphical method for assessing proportionality assumption is the Kaplan-Meier plots. The standard diagnostic procedure of proportionality assumption is a residuals-based test by Keele (2010), a test based on the difference between the observed and the fitted response. The common types of residuals in the Cox model are the Schoenfeld residuals, the martingale residuals, the deviance residuals, and the score residuals (Keele, 2010). Most software packages output one or more of these residuals. Among the numerical tests, the Grambsch and Therneau (1994) test which is based on the scaled Schoenfeld residuals is relatively easy to conduct and interpret and hence has been widely used (Persson, 2002; Keele, 2010).

The Schoenfeld residuals are obtained from the first derivative of the par-

tial log-likelihood function. The first derivative for the kth covariate as specified in Equation (4.42) is

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_k} = \sum_{j=1}^{m} \left(X_{j(k)} - A_{j(k)} \right) \tag{4.63}$$

where $A_{j(k)}$ (as mentioned in Equation (4.43)) is the conditional mean of X_k, over the individuals at risk at time $t_{(j)}$. The estimator of the Schoenfeld residual for the kth covariate is obtained by substituting the partial likelihood estimator $\widehat{\boldsymbol{\beta}}$ in Equation (4.63) and is of the form

$$\widehat{r}_i = \sum_{j=1}^{m} \left(X_{j(k)} - \widehat{A}_{j(k)} \right) \tag{4.64}$$

where $\widehat{A}_{j(k)}$ is the estimator of the conditional mean of kth covariate for the individuals who are in risk set at time $t_{(j)}$. Hence the estimated Schoenfeld residual for the jth subject on the kth covariate is given by

$$\widehat{r}_{jk} = X_{j(k)} - \widehat{A}_{j(k)}. \tag{4.65}$$

As the partial likelihood estimators, $\widehat{\boldsymbol{\beta}}$, are obtained by setting the first derivatives of the partial log-likelihood function to zero and solving the equations, hence the sum of these residuals should be zero. The partial likelihood function only considers the information in complete observations and hence the Schoenfeld residuals for the censored observations are all zero implying that these observations are not informative to the model fit and most software packages set missing values for the Schoenfeld residuals of censored observations.

Grambsch and Therneau (1994) proposed new residuals with greater diagnostic power that scale the Schoenfeld residuals. They suggested to use an

estimator of the variance of the Schoenfeld residuals as a scaling factor. Let $\widehat{\boldsymbol{r}}_j = (\widehat{r}_{j1}, \ldots, \widehat{r}_{jk}, \ldots, \widehat{r}_{jp})^\top$ denote a $p \times 1$ vector of estimated Schoenfeld residuals for the jth subject where $\widehat{r}_{jk}$ is missing if the jth observation is censored. The vector of approximate scaled Schoenfeld residuals for the jth observation as suggested by Grambsch and Therneau (1994) is

$$\widehat{\boldsymbol{r}}_j^* = m\widehat{\text{var}}(\widehat{\boldsymbol{\beta}})\widehat{\boldsymbol{r}}_j \tag{4.66}$$

where m is the number of events in the study.

Let $t^* = (t_1, t_2, \ldots, t_m)$ be the ordered survival time in the ascending order of the event's occurrence. Grambsch and Therneau (1994) test statistic for the kth covariate, T_k, is the correlation between the estimated scaled Schoenfeld residual of the kth covariate and the ordered survival time (Therneau and Lumley, 2020). This test statistic asymptotically follows a chi-square distribution with 1 degree of freedom. If T_k is significantly different from zero then the proportionality of hazards assumption for the kth covariate is violated.

4.6 Model validation

In the previous section, the selection procedures for the best subset of regressors, and methods of assessing the proportional hazards assumption have been discussed. In this section, methods that are used to assess the overall performance of the model are described. The goodness-of-fit of the Cox proportional hazards

model can be evaluated by the measurement of Royston's R-square and Harrell's concordance.

4.6.1 Royston's R-square

In multiple regression analysis with continuous outcomes, overall model performance is assessed by R^2, the coefficient of multiple determination. It measures the proportion of the variance in the observations that can be explained by the regression model. The difference between the observed and the fitted outcomes, the residual, is central to measure R^2. The Royston's R-square can be used to measure the proportion of explained variation in case of censored survival data (Royston, 2006). It is the modified version of the measure proposed by O'Quigley et al. (2005) which itself is a modification of the earlier proposal of Nagelkerke (1991).

Nagelkerke (1991) generalized the definition of R^2 to general regression models using the maximum likelihood as a criterion of fit. According to Steyerberg et al. (2010), this measure is the most commonly used measure of goodness of fit for generalized linear models. The R^2 statistic proposed by Nagelkerke (1991) is

$$R^2_{Na} = 1 - \exp[-\frac{2}{n}\{l(\widehat{\boldsymbol{\beta}}) - l(\mathbf{0})\}] = 1 - \exp(-\frac{\chi^2_{LR}}{n}), \qquad (4.67)$$

where n is the sample size, $l(\widehat{\boldsymbol{\beta}})$ and $l(\mathbf{0})$ denote the (partial) log-likelihood of the full model and the null model, respectively. The likelihood ratio statistic

χ^2_{LR} is defined in equation (4.57). As a proportion measure, R^2_{Na} ranges between 0 and 1. The value 0 indicates that the model cannot explain the survival time at all, and value of 1 indicates that the model perfectly explains the survival time. This measure is negatively correlated with the proportion of censored observations and when this proportion tends to 1 the R^2_{Na} tends to 0. Hence, O'Quigley et al. (2005) suggested to replace the sample size n in equation (4.67) by the number of uncensored observation, i.e., by the number of events m. The new statistic proposed by O'Quigley et al. (2005) is

$$R^2_{OQ} = 1 - \exp(-\frac{\chi^2_{LR}}{m}). \qquad (4.68)$$

Like Nagelkerke (1991)'s R^2_{Na}, the new measure, R^2_{OQ}, also ranges from 0 to 1 and has the same interpretation. The main disadvantage of R^2_{OQ} is that it has mild upward bias for larger amounts of censoring and hence Royston (2006) modified R^2_{OQ} to obtain a measure of explained variation. The variance of the linear term $\mathbf{X}\boldsymbol{\beta}$ in the Cox regression is approximated by $R^2_{OQ}/(1 - R^2_{OQ})$ and the residual variance is $\pi^2/6$. Hence the measure of explained variation in proportional hazards model proposed by Royston (2006) is

$$R^2_R = \frac{R^2_{OQ}/(1 - R^2_{OQ})}{\pi^2/6 + R^2_{OQ}/(1 - R^2_{OQ})} = \frac{R^2_{OQ}}{R^2_{OQ} + (\pi^2/6)(1 - R^2_{OQ})}. \qquad (4.69)$$

This statistic also ranges from zero to one and has the same interpretation as the above two measures.

4.6.2 Harrell's concordance

The degree of discriminative ability between an observed outcome and predicted outcome of generalized linear regression models can be assessed by concordance index and this was proposed by Harrell Jr et al. (1996). It is the most commonly used goodness-of-fit measure in generalized linear models (Steyerberg et al., 2010). Concordance is defined as the probability of agreement between an observed outcome and predicted outcome of any observation (Therneau and Atkinson, 2021). The concordance index is defined as the fraction of concordant pairs. For a survival outcome, the number of concordant pairs is the number of pairs of subjects where a subject with the lower predicted risk score has the longer observed survival time. Suppose C is the number of concordant pairs, D is the number of discordant pairs (subjects with lower risk score have shorter survival time) and T is the number of subjects that have tied predicted scores (not necessarily tied with observed survival time). Then the Harrell's concordance is defined as (Therneau and Atkinson, 2020)

$$c_H = \frac{C + T/2}{C + D + T}. \tag{4.70}$$

Since C, D and T are positive integers, the concordance statistic c_H ranges between 0 and 1. A concordance value of 1 indicates the perfect agreement and 0 indicates complete disagreement between observed survival times and predicted scores. For a good-fit model, the concordance value typically lies between 0.6 and 0.7 (Therneau and Lumley, 2019).

4.7 Frailty Model

The standard Cox model assumes that the subjects with the same value of the covariates have the same distribution of the survival time in the population, i.e., the subjects in the population are homogeneous. This implies that the survival times are independent and identically distributed observations from a distribution. The data used in this study comes from a large number of different general practices. The survival experience of patients from the same practice may be more similar than of the patients from different practices. The heterogeneity of survival experience of patients between practices may be due to different provision of GP and to differences between the participating practice of patient care (Collett, 2015). This unobserved heterogeneity results from the fact that it is impossible to include all the covariates in the model that are thought to influence the disease of interest. This unobserved heterogeneity can be accounted for by including random effects in the model. In survival analysis, such random effects are often referred to as frailties and the models that include random effects are called frailty models. In these models, the frailty usually acts multiplicatively on the baseline hazard. The models that include an unobserved random effect to model dependencies of survival experience in clustered data are called shared frailty models (Balan and Putter, 2020). The random effect represent a characteristic whose values are shared by subjects within a subgroup or cluster. In shared frailty model, practices with a large value of the random effect will experience event of interest earlier than the practices with small values of the random

effect and hence the name frailty (Klein and Moeschberger, 2006).

The shared frailty model is an extension of the basic Cox models: modelling data on different levels, i.e., the model is multilevel. In this research, patients are considered as level 1 units and general practices are level 2 units. Suppose there are n patients in the study and gth general practice has n_g patients, where $g = 1, \ldots, G$ and $n = n_1 + \ldots + n_G$. Then the hazard function for the ith patient $(i = 1, \ldots, n_g)$ in the gth practice, given the frailty, is (Klein and Moeschberger, 2006)

$$h_{ig}(t, u_g, \boldsymbol{\beta}, \boldsymbol{X}_{ig}) = h_0(t)u_g\exp(\boldsymbol{\beta}^\top \boldsymbol{X}_{ig}), \tag{4.71}$$

where the random effects u_g are assumed to be independent and identically distributed observations from a distribution having unit mean and unknown variance σ_u^2 and this variance needs to be estimated. A popular choice for the distribution of u_g is a gamma distribution as this results in explicit expressions for marginal likelihoods (Hosmer et al., 2011). A larger value of the variance σ_u^2 in the gamma frailty distribution indicates a greater degree of heterogeneity between practices and a stronger correlation among the patients within a practice. The inter-practice correlation is measured by Kendall's τ and is given by

$$\tau = \frac{\sigma_u^2}{2 + \sigma_u^2} \tag{4.72}$$

and $\sigma_u^2 = 0$ indicates independence of the patients of the same practice (Klein and Moeschberger, 2006). In such case, shared frailty model reduces to the Cox proportional hazard model specified in Equation (4.33).

The parameters of the shared frailty Cox model are estimated by maximizing the partial log-likelihood function. Let the number of events in the gth general practice be $D_g = \sum_{i=1}^{n_g} d_{ig}$ and $\Lambda_0(t)$ be the baseline cumulative hazard. Then the partial log-likelihood is given by

$$
\begin{aligned}
l(\sigma_u^2, \boldsymbol{\beta}) = \sum_{g=1}^{G} & D_g \log(\sigma_u^2) - \log[\Gamma(1/\sigma_u^2)] + \log[\Gamma(1/\sigma_u^2 + D_g)] \\
& - (1/\sigma_u^2 + D_g) \log \left[1 + \sigma_u^2 \sum_{i=1}^{n_g} \Lambda_0(t) \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_{ig}) \right] \\
& + \sum_{i=1}^{n_g} d_{ig} \{ \boldsymbol{\beta}^\top \boldsymbol{X}_{ig} + \log[h_0(t)] \}.
\end{aligned}
\tag{4.73}
$$

The semi-parametric estimates (if $h_0(t)$ has no parametric form) of parameters are obtained by maximizing the partial log-likelihood function using the Expectation-Maximization (EM) algorithm. The algorithm is used to estimate $\boldsymbol{\beta}$ for a set of possible values of unknown parameter σ_u^2 of gamma frailties. For each chosen values of σ_u^2, the EM algorithm has the following steps (Klein and Moeschberger, 2006):

1. Obtain an initial estimate of $\boldsymbol{\beta}$ by fitting the Cox regression model (without frailty) containing the selected covariates.

2. (E-step) Using the current estimate of $\boldsymbol{\beta}$ from the Cox regression model, estimate the expected value of frailty for each cluster, $\widehat{u}_g$, $g = 1, \ldots, G$ (general practice for this study).

3. (M-step) Update the estimate of $\boldsymbol{\beta}$ by fitting the Cox frailty model in Equation (4.71) using the same set of selected variables and estimated

114

values of frailty $\widehat{u}_g$.

4. Repeat steps 2 and 3 until both $\widehat{\boldsymbol{\beta}}$ and $\widehat{u}_g$ converge.

The entire EM algorithm is repeated for all selected values of σ_u^2. The maximum likelihood estimate of σ_u^2 is the value from the set which maximizes Equation (4.73). Like the standard Cox model, the standard errors of the estimates of shared frailty model are obtained from the inverse of the observed information matrix obtained from the partial log-likelihood function (Klein and Moeschberger, 2006).

Significance test for the regression coefficients can be performed by using a Wald test (discussed in Subsection 4.4.3) or a modified likelihood ratio test for shared frailty model that has been suggested by Nielsen et al. (1992). They define likelihood ratio test statistic as twice the difference of the partial log-likelihood for full model and the partial log-likelihood after deleting the covariate of interest for the same value of $\widehat{\sigma}_u^2$. To test whether there is any association between general practices, the hypotheses of interest are $H_0 : \sigma_u^2 = 0$ against $H_0 : \sigma_u^2 \neq 0$ and they can be tested using the likelihood ratio test. The likelihood ratio test for the shared frailty parameter, σ_u^2, has been suggested by Nielsen et al. (1992) and is given by

$$\chi_{LR}^2 = 2[l(\widehat{\sigma}_u^2, \widehat{\boldsymbol{\beta}}) - l(0, \widehat{\boldsymbol{\beta}})], \tag{4.74}$$

where

$$l(0, \widehat{\boldsymbol{\beta}}) = \sum_{g=1}^{G} \sum_{i=1}^{n_g} d_{ig} \{ \boldsymbol{\beta}^\top \boldsymbol{X}_{ig} + \log[h_0(t)] \} - \sum_{g=1}^{G} D_g. \tag{4.75}$$

115

This test statistic has a chi-square distribution with one degree of freedom.

4.8   Double-Cox regression model

The Cox proportional hazards regression model is based on the assumption that the hazard ratios of the covariates are constant over the entire follow-up time. However, in practice, this assumption is violated in many situations, especially for large datasets with a long follow-up. To deal with this problem, a number of conventional techniques, such as stratification, split time model, extended Cox model for defined time-dependent covariates etc. are usually used. In stratified Cox model, the variables which do not satisfy the proportional hazard assumption are not included in the model, and instead these variables are controlled for by stratification. That means that these variables are not adjusted for in the outcome prediction, and hence there is no way to carry out inferences for the stratified variables. Applying this method, some important predictors might remain unobserved. In split time model, the follow-up period is split into a number of sub-intervals where the proportional hazard assumption is not violated. That means different constant hazards in different time intervals. However, by applying this method, one can only estimate the hazard at different time points but it is not possible to see the hazards changes over time.

In the extended Cox model, a time-dependent covariate is created by adding a product term of the form $X \times f(t)$ involving each time-independent

covariate and a function of time in the standard Cox regression (Kleinbaum and Klein, 2012). The function of time could be in various forms depending on the covariates of interest. One choice is to consider $f_k(t) = t$, for the kth covariate X_k that do not satisfy the PH assumption (Bellera et al., 2010). This means that for each covariates X_k $(k = 1, \ldots, p)$ in the model as a main effect, there is a corresponding time-dependent covariate of the form $X_k \times t$. Another choice for the function $f_k(t)$ is to use $\log(t)$, so that the corresponding time-dependent variables will be of the form $X_k \times \log(t)$. The function of time $f(t)$ could also be a binary indicator, known as the "heaviside function", i.e., $f_k(t) = 1$ for $t \geq t_0$ and $f_k(t) = 0$ for $t < t_0$, where t_0 is a specified time point within the study period (Kleinbaum and Klein, 2012). Thus, the hazard ratio expression obtained from the extended Cox model for the exposure variables that do not satisfy the PH assumption provides the time varying hazard.

There is an alternative approach of specifying hazard function that transforms the cumulative baseline hazard function of the study population by adding an extra Cox regression term for the variables with time-variant coefficients. This modification allows to estimate both the shape and scale parameters for the time-varying coefficients, and thereby the model can handle the non-proportional hazards. In a parametric survival model described in Section 4.2, the baseline hazard function follows a parametric distribution corresponding to the exponential, Weibull, Gompertz, or other form. Begun and Kulinskaya (2022) proposed a Double-Cox regression model in which the unspecified baseline hazard function

117

in the Cox model was replaced by the Weibull or Gompertz distribution and an additional Cox regression term was incorporated to estimate the shape parameter for the variables with time-variant coefficients. To apply this model, the type of the underlying cumulative baseline hazard function of the study population is required to be satisfied. The cumulative baseline hazard of this study population follows the Weibull distribution (Figure B.2). The parameter estimation procedures of the Weibull Double-Cox model are described in the following subsection.

4.8.1 Weibull-Double-Cox regression model

The Cox proportional hazard model consists of two components, one is the unspecified baseline hazard function $h_0(t)$, and the second part is a non-negative function of the covariate vector $\boldsymbol{X}$ which is usually takes the form $\exp(\boldsymbol{\beta}^\top \boldsymbol{X})$. In a proportional hazard model under frailty setting, an additional frailty term is included in the model to take into account the influence of any unobserved factors, for example general practice that may have an impact on the survival. If there are G general practice and g-th general practice has n_g patients then the hazard function for the i-th patients in the g-th practice is given by Eq. (4.71).

The frailty term u_g in Eq. (4.71) is considered as an additional regression parameter which is constrained by a penalty function added to the partial log-likelihood. If the frailty has a gamma distribution, then the model can be written

exactly as a penalised likelihood (Therneau et al., 2003).

The general form of the baseline Weibull cumulative hazard function is $H_0(t) = \left(\dfrac{t}{a}\right)^b$; where a is scale parameter, and b is shape parameter. In Weibull-Double-Cox regression model with shared frailty proposed by Begun and Kulinskaya (2022), the baseline shape parameter of the Weibull cumulative hazard function b is extended by adding a separate Cox regression term as $b\exp(\boldsymbol{\beta}_{shape}^{\top}\boldsymbol{X})$ to estimate the shape parameters for the variables that violated the PH assumption, where $\boldsymbol{X}$ is the vector of covariates for which the shape parameter is calculated, and $\boldsymbol{\beta}_{shape}^{\top}$ is the vector of regression coefficients of those covariates. The cumulative hazard function in the Weibull-Double-Cox model is expressed by the following function (Begun and Kulinskaya, 2022):

$$H(t|\mathbf{X}) = \left(\frac{t}{a}\right)^{b\exp(\boldsymbol{\beta}_{shape}^{\top}\boldsymbol{X})}\exp(\boldsymbol{\beta}_{scale}^{\top}\boldsymbol{X}), \qquad (4.76)$$

where $\exp(\boldsymbol{\beta}_{scale}^{\top}\boldsymbol{X})$ is the Cox regression term of the proportional hazard model, which includes the scale parameters for all covariates. The term $b\exp(\boldsymbol{\beta}_{shape}^{\top}\boldsymbol{X})$ includes the shape parameters only for the variables with time-variant coefficients. The conditional hazard function in the Weibull-Double-Cox regression model is

$$h(t|\boldsymbol{X}) = \left(\frac{t}{a}\right)^b \exp[(\boldsymbol{\beta}_{shape}+\boldsymbol{\beta}_{scale})^{\top}\boldsymbol{X}) + \{\exp(\boldsymbol{\beta}_{shape}^{\top}\boldsymbol{X})-1\}\log H_0(t)]. \quad (4.77)$$

Thus the hazard ratio corresponding to two different covariate vectors $\mathbf{X}$

and $\mathbf{X}^*$ is written as

$$\frac{h(t|\boldsymbol{X})}{h(t|\boldsymbol{X}^*)} = \exp\left\{(\boldsymbol{\beta}_{shape} + \boldsymbol{\beta}_{scale})^\top (\boldsymbol{X} - \mathbf{X}^*)\right.$$

$$\left. + \left[\exp\left(\boldsymbol{\beta}_{shape}^\top \boldsymbol{X}\right) - \exp(\boldsymbol{\beta}_{shape}^\top \mathbf{X}^*)\right] \log H_0(t)\right\}. \tag{4.78}$$

Given the set of covariate vectors $\boldsymbol{X}$, the conditional cumulative hazard function with frailty u_g, takes the form:

$$\tilde{H}(t|\boldsymbol{X}, u_g) = u_g H(t|\boldsymbol{X}) = u_g \exp(\boldsymbol{\beta}_{scale}^\top \boldsymbol{X}) \left(\frac{t}{a}\right)^{b \exp(\boldsymbol{\beta}_{shape}^\top \boldsymbol{X})}, \tag{4.79}$$

where $a > 0$ is the scale parameter and b is the shape parameter for the baseline cumulative hazard function, u_g is the gamma distributed frailty with mean 1 and variance σ_u^2.

The frailty or the random effect in the Weibull-Double-Cox model is estimated on a log scale by defining $w = \log(u_g)$. Suppose that there are G general practice and all individuals of the g-th practice share the same frailty u_g. The conditional survival function is given by

$$S(t|\boldsymbol{X}, u_g) = \exp(-\int_0^t \tilde{H}(x|\boldsymbol{X}, u_g)dx), \tag{4.80}$$

and the marginal survival function is given by

$$S(t|\boldsymbol{X}) = \mathrm{E}[S(t|\boldsymbol{X}, u_g)] = (1 + \sigma^2 H(t|\boldsymbol{X}))^{-1/\sigma^2} \tag{4.81}$$

4.8.2 Parameter estimation in the Weibull-Double-Cox model

The parameter estimation in the Weibull-Double-Cox model is carried out using the Expectation-Maximization (EM) algorithm which is an alternative approach to obtain the maximum likelihood (ML) estimates. The algorithm is used to estimate the vector of the Cox-regression parameters and the parameters of baseline hazard function. Let, $\tilde{\eta} = (a, b, \beta_{scale}^{\top}, \beta_{shape}^{\top})^{\top}$ be a set of possible values of unknown parameter σ_u^2. This is an iterative procedure involving the following four steps (Begun and Kulinskaya, 2022):

1. Obtain an initial estimate of $\hat{\beta}_{scale}^{\top}$ by fitting the Cox regression model (without frailty) containing the selected covariates. The set of coefficients associated with the variables that violates the PH assumptions are considered as $\hat{\beta}_{shape}^{\top}$.

2. (E-step) Using the current estimate of $\tilde{\eta}$, estimate the expected value of frailty for each cluster, $\hat{u}_g$, $g = 1, 2, \ldots, G$ and its logarithm $\hat{w}_g$.

3. (M-step) Update the estimate of $\tilde{\eta}$ by fitting the Weibull-Double-Cox frailty model using the same set of covariates and estimated values of frailty $\hat{u}_g$.

4. Repeat step (2) and (3) until convergence is reached.

The entire EM algorithm is repeated for all selected values of σ_u^2. The ML estimate of σ_u^2 is the value from the set which maximises the marginal likelihood

function. The standard errors of the estimates are obtained from the inverse of the observed information matrix obtained from the marginal log-likelihood finction.

The performance of the Weibull-Double-Cox model is assessed by the concordance statistic, loglikelihood, and AIC. Details of these measurements were described in the previous sections. A R program for estimating the parameters of the Weibull/Gompertz Double-Cox model was developed by Begun et al. (2022) and published in GitHub .

4.9 Missing data and multiple imputation

Missing data is a common issue in clinical and epidemiological studies. Missing data can occur for a wide range of reasons such as patients lost for follow-up, missed medical appointments, failures to send questionnaires, lack of clinical measurements, and typographical errors of transferring data from paper records to electronic databases (Pedersen et al., 2017). Individuals with missing data for a particular risk factor may differ from those with no missing information. Thus excluding missing data in the analyses could lead to a potential bias in the estimaton of the parameters. The interpretation of the results could be significantly affected by the missing data. It can also decrease the power of detecting the associations between covariates and event times (Carroll et al., 2020). If missing data are present in studies then researchers should clearly

report the handing process and the assumptions made (Carroll et al., 2020). In this section, the types of missing data and techniques of handling them are explained.

4.9.1   Types of missing data

Missing data was classified into three broad categories depending on the missingness mechanism (Rubin, 1976): missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). Data are said to be MCAR if missingness does not depend on the observed data or the missing data (Carroll et al., 2020). For example, cholesterol measurements may be missing because of breakdown of an apparatus. The MCAR is very restrictive assumption and unlikely to occur in many studies (Blankers et al., 2010; Zarnoch et al., 2010). More often the missingness mechanism may depend only on the observed data, such missing data are said to be MAR (Jakobsen et al., 2017). For example, missing cholesterol measurements may be lower than observed values because younger people are more likely to have missing cholesterol measurements. The MAR assumption is more realistic than the MCAR assumption for many studies. If the missingness mechanism depends on both observed and missing data, the missing data are classified as MNAR (Jakobsen et al., 2017). For example, cholesterol readings are more likely to be missing in patients who do not receive the lipid lowering drugs, which may not be recorded in the data. Although there is no formal test to identify the type of missingness, it may be possible to as-

sess the association between missingness and other variables. For instance, by creating a contingency table of the individuals with missing data against those without, it can be ascertained that the values are missing in a particular group.

This study made use of electronic health data from UK primary care. It is well established that there is a systematic difference between unobserved and observed data in UK primary care (Hippisley-Cox et al., 2008; Marston et al., 2010). Recording of medical, lifestyle, and socio-demographic information is related to poor health conditions. For example, older and sicker people visit general practice more frequently than others, and the general practitioners are more likely to record information about these patients compared to healthier patients (MacDonald and Morant, 2008). For this research, it was assumed that participants who had no records of medical conditions or treatments, did not have the disease or receive the treatment. Thus, there were missing records in the socio-demographic and lifestyle variables only, i.e., Townsend score, BMI, and smoking category. However, due to the use of a new definition for hypertension, which included blood pressure measurement along with the Read codes, missing values were also generated in hypertension status. The study participants with and without the complete medical records were compared to each other. There was a higher proportion of missingness in the younger age group than in the older age group at study entry in both HRT users and non-users (Table B.4). Although there was an increased proportion of missingness in the non-users than the HRT users in BMI, smoking and hypertension in all age groups, the difference

in missingness between HRT users and non-users decreased with age (Table B.4). In most clinical and epidemiological studies, missing values are assumed to be MAR rather than MCAR or MNAR (Pedersen et al., 2017). Based on the pattern of missingness in the study population and the assessment from the observed data, it was assumed that data was MAR in this study.

4.9.2 Methods of dealing with missing data

There could be potential bias if missing values are not properly handled in the analysis (Carroll et al., 2020). To deal with this problem, a number of statistical methods have been developed over time. The most common and simple approaches are: complete case analysis, missing indicator method, single value imputation, and sensitivity analysis incorporating the best-case and worst-case scenarios (Pedersen et al., 2017). Multiple imputation is a relatively new and widely accepted method when dealing with multilevel missing data. The implementation of each method to handle missing values with their advantages and disadvantages are described below.

Complete case analysis

The most popular and most widely used method to handle missing data is the complete case analysis (listwise or casewise deletion) (Blankers et al., 2010; Pedersen et al., 2017). This is the default method of analysis in most software pack-

ages (Pedersen et al., 2017). In complete case analysis, individuals with missing values are excluded from the dataset before analysis. The complete case analysis assumes MCAR and will lead to biased results for other missingness mechanisms. Even under MCAR assumption, due to reduced sample size this method will have a reduced statistical power and hence is not preferential (Blankers et al., 2010). According to van Buuren (2012), this method produces standard errors that are correct for the small subset of complete cases, but are often larger for the entire dataset. However, when dealing with big datasets with a small proportion of missing information, it is still reasonable to perform the complete case analysis, because the risk of bias is low and the statistical precision is comparatively good (Pedersen et al., 2017).

Missing indicator method

The missing indicator method creates a category for the missing observations (Groenwold et al., 2012; Pedersen et al., 2017). If the covariate with missing values is categorical, then a "missing" category is assigned for the missing values. For continuous covariate, missing values are usually set to zero. In the main analytical (multivariate) model an extra indicator (0/1) variable is added to indicate the presence (1) and absence (0) of missing values (Groenwold et al., 2012; Pedersen et al., 2017). The method is popular because it does not exclude any observations and hence maintains statistical power. However, the method produces too small standard errors (van Buuren, 2012; Pedersen et al., 2017) and

can exhibit severe bias when data are MCAR with a large proportion of missing observations (Greenland and Finkle, 1995).

Single imputation

Substituting missing value with a reasonable guess, such as the mean, median, or modal value of the observed data is called single imputation (Carroll et al., 2020). In longitudinal studies, where measurements from each individual are taken repeatedly over a series of planned time-points, missing values are often replaced with the last observed value for a given variable from the same subject (Pedersen et al., 2017). This approach is called the "last observation carried forward". An advanced single imputation method, regression based single imputation of missing values, imputes missing values with the predicted values of that covariate obtained by a regression model using the complete cases (Pedersen et al., 2017). The regression method regresses that covariate on the remaining covariates of the study. Regression imputation is applicable for both MCAR and MAR data (Blankers et al., 2010).

The standard errors of the estimates in single imputation are usually likely to be too small as this method fails to account for the uncertainty of missing data (Greenland and Finkle, 1995; Sterne et al., 2009; Pedersen et al., 2017). If the data are not MCAR, mean estimation will produce biased estimate of the mean (Blankers et al., 2010).

Multiple imputation

Multiple imputation (MI) is a widely accepted method to deal with missing data. It assumes that data is MAR but also can handle MCAR (Pedersen et al., 2017). This method solves the problem of underestimation or overestimation of standard errors obtained using other methods of handling missing data discussed earlier (Pedersen et al., 2017). In this method, missing values are replaced by plausible imputed values multiple times, creating multiple completed datasets. The imputed values are drawn from the posterior predictive distribution of the missing values using an appropriate model (Rubin, 1996; Carroll et al., 2020). According to Zarnoch et al. (2010), five completed datasets are recommended. However, Blankers et al. (2010) suggested to generate 3 to 10 completed datasets. Schafer and Graham (2002) have shown that for a dataset having 50% missing observations, performing 10 imputations is 95% efficient and concluded that additional imputations have little advantage in respect to removing noise from the estimate itself. The reason for imputing multiple times is to reflect the uncertainty in estimating missing data in one imputation. Across several imputations, imputed values are a random sample of the missing values and that results in a valid statistical inference (Rubin, 1987). Multiple imputation yields unbiased and valid estimates of associations of covariates based on information from the available data; i.e., providing estimates similar to those calculated from full data. It affects the estimates of the coefficients for variables with missing data as well as the remaining variables with no missing data (Pedersen et al., 2017).

A number of packages are implemented in most statistical software to do the imputations. In this study, R package "Jomo" was used to impute the missing values. Jomo can handle joint modelling multilevel multiple imputation assuming data is multivariate normally distributed. It is a relatively new package (Quartagno et al., 2019) and becoming increasingly popular because it guaranteed compatibility for clustered level data based on joint modelling imputation that other packages can not handle. In this study, data is multilevel as it includes patients (level 1) and general practices (level 2). Jomo can handle missing data in continuous, binary, or categorical variables. After imputing the missing data multiple times, each imputed dataset was modelled separately and the estimated coefficients were pooled by applying Rubin's rules (Rubin, 1987). The implementation of multiple imputation consists of the following three stages:

1. **Imputation step:** Select all exposures (covariates with missing values), covariates, and outcomes that were considered in the analysis model. Select auxiliary variables that are associated with the values of the incomplete variables. Inclusion of auxiliary variables may reduce bias and increase the precision of the estimates (Pedersen et al., 2017). Then, create multiple copies (say m datasets) of the selected dataset by replacing missing values with imputed values that are simulated from an appropriate posterior predictive distribution, where the individual values may vary between datasets (Rubin, 1996; Pedersen et al., 2017).

   The imputation process for each covariate with missing data depends on

the measurement scale. Continuous covariates are imputed using a linear regression and binary covariates are imputed using logistic regression. Covariates with more than two categories are imputed with a multimonial regression. In multiple imputation, rows (individuals) having the same pattern of missingness are grouped together. If there is only one variable having missing data, this is called univariate missing data and that variable is called the target variable. In this type of missing data, continuous target variable is regressed (linear regression) on the remaining covariate(s) using the observed data. The imputed value for a row is the predicted value obtained using corresponding values of the regressor(s) plus random noise drawn from a normal distribution with zero mean and appropriate standard deviation. Whereas, logistic regression is used to impute the missing values for a binary target variable. A target variable with more than two categories is imputed using a multinomial logistic regression (van Buuren, 2012).

Joint modelling starts with an assumption about the multivariate distribution of the data, the most widely applied is the multivariate normal distribution (van Buuren, 2012). For a specific missingness pattern, fit a multivariate linear regression model considering missing covariates as the responses and the remaining covariate(s) as regressor(s) using the observed data. Then, for each row in that group obtain the predicted value of the responses using corresponding values of regressor(s) and replace the missing

130

value by the predicted values plus random errors drawn from a multivariate normal distribution with zero mean and appropriate variance-covariance matrix (for more details, see Schafer and Graham (2002)). However, the normality assumption in real data is rarely preserved. Binary or ordinal covariates are imputed through latent normal variables in the Jomo package (Quartagno et al., 2019). Right skewed covariates may be transformed to a logarithmic scale before imputation and imputed values are transformed back to the original scale (Schafer and Graham, 2002). A simulation study conducted by Schafer (1997) showed that imputations based on multivariate normal distribution are robust to non-normal data (van Buuren, 2012).

2. **Analysis step:** Analyse each of the m completed datasets using the chosen statistical model of interest. This gives the parameter estimates with corresponding standard errors for each dataset. The estimated coefficients in each of the imputed datasets could vary because of the uncertainty introduced in the imputation of the missing values (Rubin, 1996; Sterne et al., 2009; Pedersen et al., 2017)

3. **Combination step:** The parameter estimates obtained from each completed dataset are pooled using Rubin's rules (Rubin, 1987), with the corresponding standard errors. The pooling of parameter estimates takes account of the between variability of the completed datasets while pooling of standard errors accounts for both the between and within imputation variations (Marshall et al., 2010). The steps of pooling the estimated pa-

rameters using Rubin's rules are described below.

Rubin's Rule

Suppose that $\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_m$ are the estimates of θ from m completed datasets and $\widehat{V}_1, \widehat{V}_2, \ldots, \widehat{V}_m$, respectively, are their associated estimated variances. As described by Rubin (1987), the pooled estimate of θ is

$$\widehat{\theta}_{MI} = \frac{1}{m} \sum_{l=1}^{m} \widehat{\theta}_l.$$

The variance of this combined estimate has two parts: average within-imputation variance and between-imputation variance. The average within-imputation variance is calculated as

$$\overline{W}_{MI} = \frac{1}{m} \sum_{l=1}^{m} \widehat{V}_l$$

and the between-imputation variance is defined as

$$B_{MI} = \frac{1}{m-1} \sum_{l=1}^{m} (\widehat{\theta}_l - \widehat{\theta}_{MI})^2.$$

The total variance of $\widehat{\theta}_{MI}$ is estimated as

$$T_{MI} = \overline{W}_{MI} + (1 + m^{-1}) B_{MI}$$

and the square root of T_{MI} is the combined standard error.

To test whether the combined estimate is significantly different from zero, Rubin (1987) recommended to use t-statistic

$$t = \frac{\widehat{\theta}_{MI}}{\sqrt{T_{MI}}}.$$

132

The t-statistic follows a t-distribution with v degrees of freedom. The number of degrees of freedom v is calculated as

$$v = (m - 1) \left[1 + \frac{\overline{W}_{MI}}{(1 + m^{-1})B_{MI}} \right]^2 .$$

A $100(1 - \alpha)\%$ confidence interval for the parameter θ is given by

$$\overline{\theta}_{MI} \pm t_{v,1-\alpha/2} \sqrt{T_{MI}}$$

where $t_{v,1-\alpha/2}$ is the upper $100(1 - \alpha/2)$th percentile of a t-distribution with v degrees of freedom.

4.10 Summary

This chapter discussed the statistical procedures for analysing time-to-event data. The application of parametric and non-parametric models are discussed first, with the former being used when the underlying distribution of the study population is known and the latter being used when the distribution of the study population is unknown. The semi-parametric Cox proportional hazards regression model, which does not require the distribution of the baseline hazards to estimate the effects of risk factors on the survival, was explained in details. These included the assumptions required for the Cox model, and for testing the overall quality of the models as well as the effects of the individual parameters. A newly developed parametric Double-Cox model that is capable of handling non-proportionality in Cox regression was introduced. Finally, techniques for

dealing with missing data using the multilevel multiple imputation method were described.

Chapter 5

Characteristics of the study population

In Chapter 1, the rationale, aims and objectives of HRT study were presented. Chapter 2 reviewed the findings from the past research on HRT. In Chapter 3, the importance of primary care data in HRT modelling, and patient selection process from the THIN database were described. Chapter 4 explained the statistical methodology for survival analysis. This chapter presents the distribution and characteristics of the covariates selected for HRT modelling in the extracted population. The covariates include a wide range of important medical, socio-demographic, and lifestyle factors. The prevalence of health and lifestyle variables in the extracted data are also compared to the prevalence in UK population.

5.1 HRT prescription by age cohort

The study design and patient selection process of this study were described in details in Chapter 3. There were 112,354 HRT users who were born on or before 1960, and started HRT for the first time at the age of 46 years or older during 1984−2017, and 245,320 matched non-users in the extracted data file. To examine the trends and prevalence of first prescription across different age cohorts, the number and proportion of patients in four different age groups of 46 to 55, 56 to 65, 66 to 75, and 76^{+} at first date of HRT prescription for the exposed patients were calculated. Table 5.1 shows the frequencies (%) of the first HRT prescription in HRT users and their matched (one up to three) non-users by age group. This table shows that the highest number of first HRT prescriptions (70.08%) were in ages between 46 to 55 years; i.e., either at peri-menopausal stage or menopausal transition stage, when most women suffer from menopausal symptoms. The number of first HRT prescriptions falls rapidly after menopausal transition period. There were 23.53% first prescriptions issued to the age group 56 to 65, and 5.33% were issued to the age group 66 to 75. Only 1,183 (1.05%) women were prescribed first HRT at age 76 and over. HRT prescriptions were probably issued to the oldest age cohort for other reasons, such as osteoporosis treatment (Bromley et al., 2004). The distribution of the HRT users and non-users by one-to-one match is presented in Table B.3.

To observe the prescribing patterns over time, the frequencies of the first

Table 5.1: Prevalence of first HRT prescription in exposed group and the distribution of their matched non-users in the extracted population by age group

Age ranges	HRT users (%)	Non-users (%)	Total (%)
46−55	78743 (70.08%)	159200 (64.89%)	237943 (66.53%)
56−65	26436 (23.53%)	65323 (26.63%)	91759 (25.65%)
66−75	5992 (5.33%)	17431 (7.11%)	23423 (6.55%)
76^{+}	1183 (1.05%)	3366 (1.37%)	4549 (1.27%)
Grand total	**112354 (100%)**	**245320 (100%)**	**357674 (100%)**

HRT prescriptions by calendar year among the four age groups were also calculated. Figure 5.1 shows the number of first HRT prescriptions per year in the extracted data within the four age cohorts. Most women started HRT at age between 46 to 55, and this trend increased over years with the highest number (5,574) of first starters in 2000. After 2000, there was a sudden fall in HRT prescriptions within this age group. This decrease is thought to be caused by the influence of WHI study published in 2002 (Rossouw et al., 2002), that reported detrimental effect of HRT. The negative impact of HRT reported by WHI got a wide media coverage creating panic among HRT users, and compelled physicians to follow revised guidelines on issuing HRT prescriptions (Cagnacci and Venier, 2019). The use of HRT among other age groups also followed the similar pattern over time though the peak was much lower within the other groups. The pattern and prevalence of HRT use within different age groups in this study population follows a trend similar to that described in Bromley et al. (2004) who conducted

Number of first HRT prescriptions over years by age group
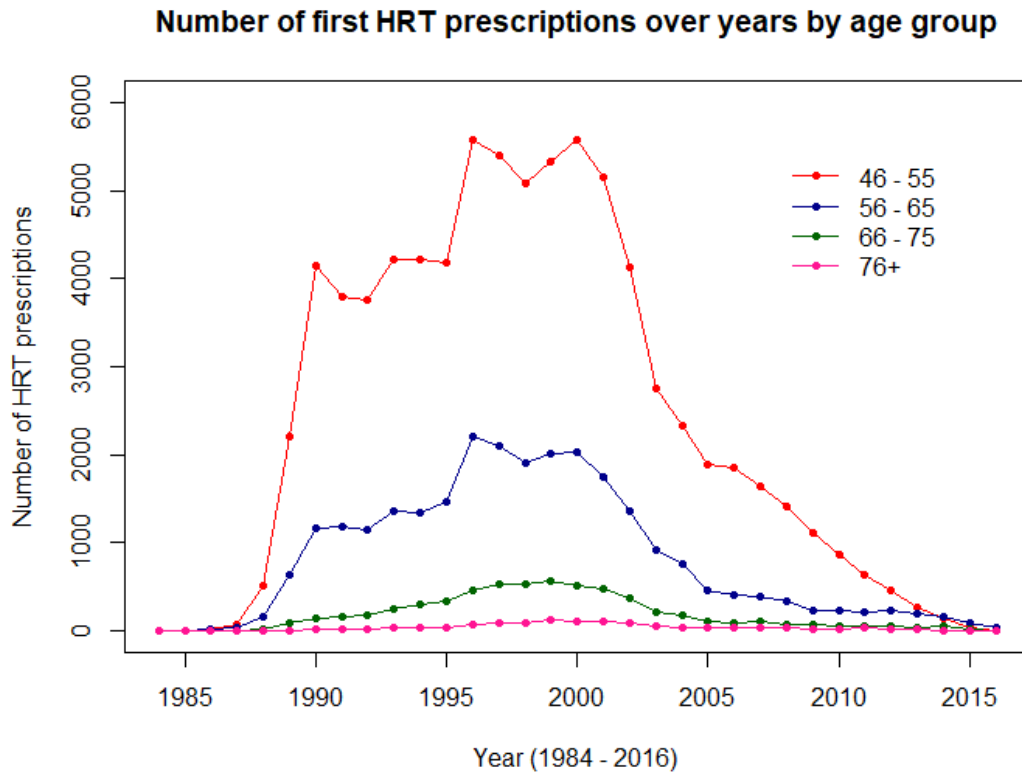
Figure 5.1: Number of first HRT prescriptions in the extracted population by age group and per calendar year in the period 1984−2016.

a descriptive study on HRT utilization in UK women using primary care data.

5.2 Distribution of lifestyle and socio-demographic factors

This section presents the distribution of lifestyle and socio-demographic factors across the study population. The selected covariates are body mass index, smoking status, and Townsend deprivation index. The distribution of these variables

138

Table 5.2: BMI Classification (NICE Obesity Management, 2018)

Classification	BMI (kg/m^2)
Underweight	$<\ 18.5$
Healthy weight	$18.5-24.9$
Overweight	$25.0-29.9$
Obesity I	$30.0-34.9$
Obesity II	$35.0-39.9$
Obesity III	$>\ 40$

in study population is also compared with UK national statistics.

5.2.1   Body mass index

Body mass index (BMI) is a continuous variable which indicates whether a person's body weight is healthy or not. It is calculated by the ratio of the weight of a person to the square of his/her height, and is measured in kg/m². The measurement was first introduced by a Belgian mathematician Adolphe Quetelet, and thus BMI is also known as Quetelet index following his name (Diabetes UK, 2019). For this study, the NICE classification of overweight and obesity measurement scale presented in Table 5.2 was used to group BMI of the study population (NICE Obesity Management, 2018).

To calculate the BMI in the study population, the most recent records of weight and height prior to the study entry date were extracted from the Additional Health Data (AHD) file in THIN. Then the calculated BMI measurements were grouped into five categories, namely underweight, healthy weight, over-

Table 5.3: Distribution of the study population by BMI category at baseline

BMI category	HRT users (%)	Non-users (%)	Total (%)
Underweight	1401 (1.25%)	3090 (1.26%)	4491 (1.26%)
Healthy weight	39322 (35.00%)	64532 (26.30%)	103854 (29.04%)
Overweight	32627 (29.04%)	59615 (24.30%)	92242 (25.79%)
Obese	19622 (17.46%)	47728 (19.46%)	67350 (18.83%)
Missing	19382 (17.25%)	70355 (28.68%)	89737 (25.09%)
Grand total	**112354 (100%)**	**245320 (100%)**	**357674 (100%)**

weight, obese, and missing. All categories of obesity were merged into the obese group. BMI values which are less than 15 kg/m^2 or greater than 50 kg/m^2 were merged into missing category, as these were most likely to happen due to a typographical error. Table 5.3 shows the BMI distributions for the HRT users and non-users in the study population.

The proportions of women with a healthy weight and overweight were higher in the exposed group than the unexposed, accounting for 35.0% and 29.04% for the exposed, 26.3% and 24.3% for the unexposed group, respectively. On the other hand, there were slightly more obese women in exposed group (19.46%) than the unexposed (17.46%). There were a very small number of underweight women in both groups and the proportions were nearly the same (HRT users: 1.25%; non-users: 1.26%), and thereby this group was excluded from the study. The proportion of missing records was higher in non-users (28.68%) than in HRT users (17.25%), which is probably due to the less frequent attendance of the non-users to the GP compared to HRT users.

According to the NHS information centre, in 2008, 32% adult women

were classified as overweight, and 25% adult women were classified as obese in the UK (NHS, 2010). There were 31% overweight and 27% obese women in the UK in 2015 (NHS Digital, 2016). The total percentages of overweight and obese women in the study population are 34.4% and 25.1% respectively, if the missing values are excluded. As the prevalence of overweight and obesity in the UK females are increasing over time (Cancer Research UK, 2019a), and in this study, the highest number of first HRT prescription were around the year 2000, the prevalence of overweight and obese females in the UK were noted around that time. According to a report by Statista (2018), there were 21% obese women in the UK in 2000. If the missing values are excluded, there were 21.1% of obese women in the exposed group and 25.14% in the unexposed group in this study population. Although the prevalence of obesity in HRT users was the same as in the UK general population, the prevalence of obesity in non-users was 4% higher in this study. The higher prevalence of obesity in unexposed group could be due to a larger number of missing entries in that group.

5.2.2   Smoking status

The smoking status information of patients in THIN is recorded in either AHD or medical (MED) files, and coded by the symbols "N", "Y", and "D"; where,"N" means non-smoker,"Y" represents current smoker, and "D" stands for ex-smoker. For this study, the most recent smoking history for both HRT users and matched non-users before the study entry date were extracted. Table 5.4 shows the pro-

portion of different smoking categories in the study population at baseline. The distribution shows that, more than half of the patients in both groups were lifetime non-smokers. The proportions of non-smoking patients in both groups are nearly the same, accounting for 53.68% in exposed and 52.76% in unexposed. However, among the exposed group there were slightly more ex-smokers (HRT users: 16.59%, non-users: 13.07%) and current smokers (HRT users: 20.42%, non-users: 17.23%). Similar to BMI, there were more missing information on smoking in non-users than in the HRT users. According to the Office for National Statistics (ONS), there were 16.5% and 15.3% women smokers in the UK in 2013 and 2015, respectively. The incidence of cigarette smoking among UK women declines over time (Cancer Research UK, 2019b). This habit also decreases as women ages (NHS Digital, 2018b). According to another report published by the Cancer Research UK (2018), there were 27% female smokers between ages 50−59, and 13% above age 60 in Great Britain in 2000. From this report, the percentage of women smokers aged 50 and over was 20% in 2000. In this study, the total number of current smokers is 21.33% if the missing values are excluded. Overall, the smoking patterns in this study population are nearly the same as in the national statistics.

5.2.3 Deprivation status

Townsend score, also known as the Townsend deprivation index, is a census based index of material deprivation which is used in the UK as an indicator of a

Table 5.4: Distribution of the study population by smoking category at baseline

Smoking	HRT users (%)	non-users (%)	Total (%)
Non-smoker	60317 (53.68%)	129441 (52.76%)	189758 (53.05%)
Ex-smoker	18645 (16.59%)	32064 (13.07%)	50709 (14.18%)
Current smoker	22942 (20.42%)	42280 (17.23%)	65222 (18.24%)
Missing	10450 (9.30%)	41535 (16.93%)	51985 (14.53%)
Grand total	**112354 (100%)**	**245320 (100%)**	**357674 (100%)**

person's deprivation status. It was first introduced by Peter Townsend in 1987 (Townsend, 1987). This measure has been widely used in medical, education and crime research to establish the relation with people's social deprivation. Score calculation is based on a combination of the percentage of following four census variable indicators for any geographical area given data is available for that area (UK Data Service, 2019):

- Non-car ownership

- Overcrowded households

- Households not occupying own houses

- Unemployment

These variables are weighted equally in the Townsend score calculation. The unemployment and overcrowded households indicators are log transformed to normalise the skewed results from these variables. Then a standard Z scores is calculated from each of the four indicators. The reason of creating Z scores

Table 5.5: Distribution of the study population by Townsend quintile at baseline

Townsend score	HRT users (%)	non-users (%)	Total (%)
1 (least deprived)	30847 (27.46%)	61886 (25.23%)	92733(25.92%)
2	24761 (22.04%)	52170 (21.27%)	76931 (21.51%)
3	20546 (18.29%)	45030 (18.36 %)	65576 (18.33%)
4	15157 (13.49%)	35156 (14.33%)	50313 (14.07%)
5 (most deprived)	9253 (8.24%)	21894 (8.92%)	31147 (8.71%)
Missing	11790 (10.49%)	29184 (11.90%)	40974 (11.46%)
Grand total	**112354 (100%)**	**245320 (100%)**	**357674 (100%)**

is to standardise each variable so that the extreme values did not affect the overall Townsend score too greatly. Finally, the Townsend score is calculated as a sum of four Z scores. The Z scores are centred around of zero mean, and hence any areas with a Townsend score above zero are above the mean and therefore deprived, whereas areas with scores below zero are affluent (Norman and Darlington-Pollock, 2017). Figure 5.2 provides the steps of Townsend score calculation. Once the Townsend deprivation scores were calculated they split into quintiles. The first quintile represent the least deprived group whilst the fifth quintile represents the most deprived. Table 5.5 provides the distribution of the Townsend quintile scores among HRT users and non-users in this study.

There were missing values in the Townsend score. However, the percentage of missing values in Townsend score was relatively low, and the prevalence of missingness is similar in both HRT users (10.5%) and non-users (11.9%). This is plausible because information related to socio-economic status is recorded when a person is registered with a GP and usually it is not updated further. After ex-

Figure 5.2: Steps of Townsend score calculation. Figure modified from UK Data Service (2019)

cluding the missing data, more than 50% (HRT users: 55.3%, non-users: 52.8%) of women in the study population lived in less deprived areas (Townsend score 1 and 2), and nearly 23% lived in the more deprived areas (Townsend score 4 and 5). According to UK Data Service (2019) report, in the UK, there were 45% of population in less deprived areas, and 35% lived in more deprived areas in 2011. There were no significant changes in Townsend deprivation score in the UK from 2001 to 2011 (UK Data Service, 2019). The higher number of women in this study lived in the less deprived areas, and the lower number of women lived in the more deprived areas than in the UK general population. This is probably due to the fact that more highly educated women were HRT users in the UK (Hunt et al., 1987; Bromley et al., 2004), and/or that THIN retains records of more people from the affluent areas (Hippisley-Cox et al., 2008).

5.3 Prevalence of medical conditions at baseline

In Chapter 2, Section 2.5, results from previous studies on the effects of HRT on different chronic diseases were explained. In this section, the prevalence of various medical conditions at study entry in the study population are discussed. The selection of health conditions was based on their importance identified from literature review, and included hypertension, coronary heart disease, peripheral vascular/arterial disease, hypercholesterolaemia, osteoporosis, type II diabetes, oophorectomy and hysterectomy status. The selected medical conditions that

were recorded prior to the patients' study entry date were extracted from the MED file in THIN using corresponding Read codes. The distribution of these conditions for the HRT users and non-users are calculated for full extracted data and age sub-groups at first HRT treatment. Patients who did not have a record of a particular medical condition were assumed not to have the disease, and hence there were no missing records in the disease variables. Table 5.6 displays the frequencies and prevalence of the selected health conditions for HRT users and matched non-users at baseline for full data and age sub-groups. The prevalences of these medical conditions are also compared with the UK national statistics of disease prevalence among women.

5.3.1 Hypertension

Hypertension (HT) is a long term medical condition caused by an elevated blood pressure in the arteries. It is widely known as high blood pressure (BP). HT is more common among UK men than women (NICE, 2019a). Several factors, such as diabetes, obesity, lifestyle choices, physical activity level, and ageing also increase the risk of having high BP. Hypertension is considered to be a major risk factor for cardiovascular disease in women (Gudmundsdottir et al., 2012). In England, HT is the third biggest risk factor for premature death and disability after smoking and diet, and at least half of all strokes, heart attacks, and coronary heart disease are associated with high BP (Public Health England, 2017). It is also a major risk factor for chronic kidney disease, heart failure,

Prevalence of hypertension in extracted patients at baseline
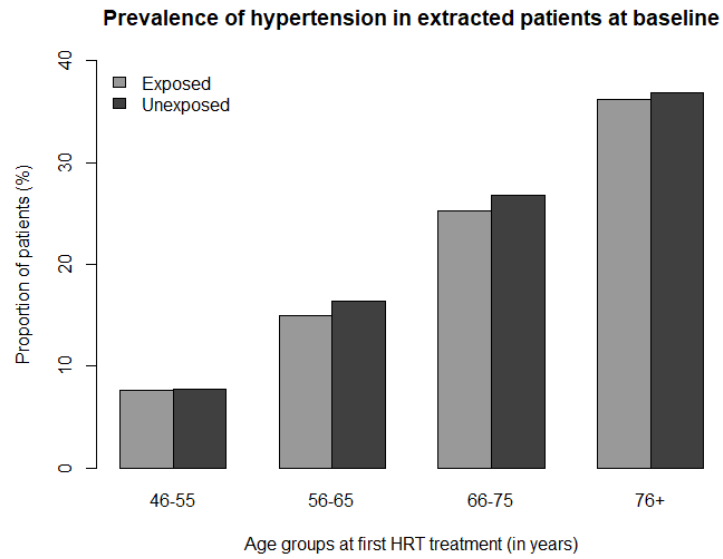
Figure 5.3: Prevalence of hypertension by age group at study entry in HRT users and non-users.

and dementia (Public Health England, 2017). According to the NICE (2019b), a person is diagnosed with hypertension if their multiple blood pressure readings are greater than 140/90 mmHg. The upper reading is for the systolic and the lower for the diastolic blood pressure. The systolic blood pressure (SBP) is the highest level of pressure inside the blood vessels when heart beats, and the diastolic blood pressure (DBP) is the lowest pressure inside the blood vessels when heart rests between beats (Blood Pressure UK, 2017). For this study, the latest records of hypertension among study population before their study entry date were collected from THIN database using the respective Read codes. The overall prevalence of HT in this study was 11.41% at baseline. The prevalence was nearly the same in the exposed and unexposed groups, accounting for 10.58% and 11.79%, respectively. The prevalence was also approximately similar within

each of four age groups, $46-55$, $56-65$, $66-75$, and 76^+ (see Figure 5.3). However, the prevalence of hypertension raised dramatically as age increased in both HRT users and non-users, with the lowest in $46-55$ age group (7.70%) and the highest in 76^+ age group (36.71%). Peri-menopausal women have a lower risk and incidence of hypertension than men but this advantage gradually disappears after menopause (Gudmundsdottir et al., 2012). This could be the reason for low prevalence of HT in the youngest age group ($46-55$) but the highest prevalence in the oldest group (76^+) at first HRT treatment in the study population. Blak et al. (2011) compared the THIN prevalence of various chronic medical conditions to the UK national QOF data (2006/2007) and showed that the prevalence of hypertension in THIN females was 12.7% and in QOF 12.6%. Which is slightly higher than the prevalence of HT in this study. The proportions of HT in exposed group were 07.61%, 14.96%, 25.28%, 36.18%, and in unexposed group, 07.74%, 16.36%, 26.81%, 36.90%, respectively among the youngest to the oldest age groups. In 2003, the prevalences of HT among UK females were 22.8% in the age group $45-54$, 43.2% in $55-64$, 63.5% in $65-74$, and 75.0% in the age group 75^+, respectively (MacDonald and Morant, 2008). This shows a much higher prevalence of HT than in this study population. The main reason of lower prevalence of hypertension in this population compared to the national statistics is that a number of conditions such as severe heart failure, heart attack, and other serious heart related diseases were excluded in this study, and it is known that hypertension is one of the major causes of cardiovascular disease (MacDon-

ald and Morant, 2008). Furthermore, these prevalences were based on the Read codes only, and past study showed that using Read codes only in THIN to calculate prevalence underestimates the actual hypertensive prevalence in the UK (Peng et al., 2016). Thus, a revised definition of hypertension that considered both SBP and DBP measurements along with diagnosis Read code was used in this study. This will be thoroughly explained in Chapter 6 Section 6.6.1.

5.3.2 Coronary heart disease

Coronary heart disease (CHD), also known as coronary artery disease is the most common type of heart problem which is caused by the development of fatty substances in the blood vessels that circulate blood to and from the heart. The probability of developing CHD in menopausal women was estimated to be 46%, and it is also a leading cause of death among postmenopausal women (Grady et al., 1992). The health implications of having CHD were discussed in Chapter 2, Section 2.5.3. The overall prevalence of CHD in HRT users and non-users in this study were 2.19% and 2.46%, respectively (see Table 5.6). The prevalence increased as women aged in both groups similarly except the oldest cohort where the prevalence was slightly higher in HRT users compared to their matched non-users. The prevalence was lowest in the $46 - 55$ age group (HRT users: 0.82%, non-users: 0.75%) and highest in the 76^+ age group (HRT users: 15.7%, non-users: 12.7%). The overall prevalence of coronary heart disease in women in England was 4.0%, and 1.3% in 45 to 54, 3.5% in 55 to 64, 10.0% in 65 to 74,

and 19.3% in 75^+ age group, respectively in 2006 (BHF, 2012). The overall prevalence of CHD in women of Scotland was 5.2% in 2010 (BHF, 2012). The prevalence is slightly lower in this study population than the national statistics probably because of the exclusion of patients with severe heart disease.

5.3.3 Peripheral vascular and artery disease

Peripheral vascular disease (PVD) and peripheral artery disease (PAD) are problems related to inadequate blood flow in the blood vessels which results in poor blood flow in the brain, heart, arms, or legs. As a result of PAD/PVD, the legs and feet are most commonly affected. PAD/PVD also increases the risk of having coronary heart disease and stroke (British Heart Foundation, 2017). People with PAD have a three-fold increased risk of mortality from major cardiovascular events such as heart attack and stroke compared to those without PAD (Fowkes et al., 2013). There are four elevating stages of severity of PAD/PVD; Stage I: asymptomatic, Stage IIa: mild claudication, Stage IIb: moderate to severe claudication, Stage III: ischaemia rest pain, and Stage IV: ulceration or gangrene (BMJ Best Practice, 2018). The overall prevalence of PVD/PAD in the study population is 8.39% and it is nearly equal for the HRT users (8.78%) and the non-users (8.22%).

Figure 5.4 shows the proportion of PAD/PVD in HRT users and non-users by age group for this study at baseline. The prevalence is slightly higher in the
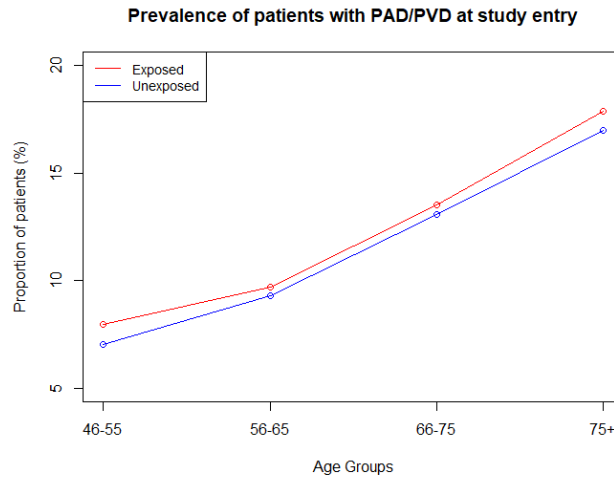
Figure 5.4: Prevalence of Peripheral Vascular and Artery disease by four age groups in HRT users and non-users at study entry. HRT users have a slightly higher prevalence of PAD/PVD compared to the matched non-users.

HRT users than the non-users across all the age groups. For HRT users, the prevalences are 7.97%, 9.71%, 13.50%, 17.84, and for non-users 7.06%, 9.31%, 13.06%, 16.96%, respectively in $46-55$, $56-65$, $66-75$, and 76^+ age groups. The prevalence rises gradually with increased age like in many other medical conditions. PAD/PVD is sometimes undiagnosed and may occur as comorbidities with other diseases such as diabetes and cardiovascular disease. The overall prevalence of PAD/PVD in UK women is not exactly known. A systemic review comprising 34 studies on PAD prevalence performed by Fowkes et al. (2013) showed that the mean prevalences of PAD/PVD in women of high income countries are 5.76%, 8.01%, 11.03%, 18.63% in $45-54$, $55-64$, $65-74$, and 75^+ age groups respectively in the period from 2000 to 2010. This is relatively close to the prevalence of PAD/PVD in this study.

5.3.4   Hypercholesterolaemia

Hypercholesterolaemia is a condition that is characterized by high levels of cholesterol in the bloodstream. Cholesterol is a fatty substance produced in the body or obtained from consumption of fatty foods such as meat, fish, and dairy products. The body needs cholesterol to build cell membranes, make certain kinds of hormones such as progesterone and testosterone, to produce vitamin D, and bile acid. However, too much cholesterol increases the risk of developing coronary heart disease and other cardiovascular disease. Usually, high blood cholesterol levels result from a combination of genetic and environmental risk factors. Lifestyle choices including diet, physical exercise, and smoking strongly influence the cholesterol levels in the blood. Gender, age, and health conditions such as diabetes and obesity also affect the cholesterol levels. A small percentage of people with high cholesterol have an inherited form of hypercholesterolaemia. The most common cause of inherited high cholesterol is a condition known as familial hypercholesterolemia, which results from mutations in the LDLR gene (US National Library of Medicine, 2019).

Table 5.6: Baseline prevalence of the selected medical conditions in the study population by age group at study entry and in total. For each condition, percentages were calculated by taking the ratio of the number of patients with the disease in an age group to the total size of population in that age group.

Age groups	Hypertension			Diabetes (Type II)			Osteoporosis		
	HRT users (%)	non-users(%)	Total(%)	HRT users (%)	non-users(%)	Total(%)	HRT users(%)	non-users(%)	Total(%)
46−55	5993 (7.61)	12323 (7.74)	18316 (7.70)	972 (1.23)	2757 (1.73)	3729 (1.57)	1224 (01.55)	2269 (1.43)	3493 (01.47)
56−65	3955 (14.96)	10684 (16.36)	14639 (15.95)	572 (2.16)	2325 (3.56)	2897 (3.16)	1221 (04.62)	1934 (02.96)	3155 (03.44)
66−75	1515 (25.28)	4674 (26.81)	6189 (26.42)	199 (3.32)	987 (5.66)	1186 (5.06)	861 (14.37)	1080 (6.20)	1941 (8.29)
76+	428 (36.18)	1242 (36.90)	1670 (36.71)	76 (6.42)	243 (7.22)	319 (7.01)	323 (27.30)	495 (14.71)	818 (17.98)
Total	**11891 (10.58)**	**28923 (11.79)**	**40814 (11.41)**	**1819 (1.62)**	**6312 (2.57)**	**8131 (2.27)**	**3629 (3.23)**	**5778 (2.36)**	**9407 (2.63)**

	Hypercholesterolaemia			Oophorectomy			CKD(Stage 1-2)		
	HRT users (%)	non-users(%)	Total(%)	HRT users (%)	non-users(%)	Total(%)	HRT users(%)	non-users(%)	Total(%)
46−55	660 (0.84)	1219 (0.77)	1879 (0.79)	12676 (16.10)	11256 (7.07)	23932 (10.06)	48 (0.06)	131 (0.08)	179 (0.08)
56−65	566 (2.14)	1379 (2.11)	1945 (2.12)	7515 (28.43)	9731 (14.90)	17246 (18.80)	30 (0.11)	143 (0.22)	173 (0.19)
66−75	203 (3.39)	609 (3.49)	812 (3.47)	1938 (32.34)	2787 (15.90)	4725 (20.17)	31 (0.52)	117(0.67)	148 (0.63)
76+	44 (3.72)	105 (3.12)	149 (3.28)	344 (29.10)	499 (14.82)	843 (17.19)	47 (3.97)	119 (3.54)	166 (3.65)
Total	**1473 (1.31)**	**3312 (1.35)**	**4785 (1.34)**	**22473 (20.00)**	**24273 (9.89)**	**46746 (13.07)**	**156 (0.14)**	**510 (0.21)**	**666 (0.19)**

	PAD/PVD			Hysterectomy			CHD		
	HRT users (%)	non-users(%)	Total(%)	HRT users (%)	non-users(%)	Total(%)	HRT users(%)	non-users(%)	Total(%)
46−55	6275 (7.97)	11240 (7.06)	17515 (7.36)	6679 (8.48)	3006 (1.90)	9685 (4.07)	643 (0.82)	1200 (0.75)	1843 (0.77)
56−65	2567 (9.71)	6079 (9.31)	8646 (9.42)	4365 (16.51)	3530 (5.40)	7895 (8.60)	720 (2.72)	1923 (2.94)	2643 (2.90)
66−75	809 (13.50)	2276 (13.06)	3085 (13.17)	1264 (21.10)	1284 (7.37)	2548 (10.88)	469 (7.83)	1349 (7.34)	1818 (7.76)
76+	211 (17.84)	571 (16.96)	782 (17.19)	222 (18.76)	278 (8.26)	500 (11.00)	186 (15.72)	428 (12.72)	614 (13.50)
Total	**9862 (8.78)**	**20166 (8.22)**	**30028 (8.39)**	**12530 (11.15)**	**8098 (3.30)**	**20628 (5.77)**	**2463 (2.19)**	**6032 (2.46)**	**8495 (2.38)**

Table 5.7: Interpretation of cholesterol levels. Reproduced from Todkar et al. (2013)
.

Cholesterol type	mg/dL	mmol/L	Interpretation
Total cholesterol	< 200	< 5.2	Desirable
	200−239	5.2−6.2	Borderline
	> 240	> 6.2	high
LDL cholesterol	< 100	< 2.6	Most desirable
	100−129	2.6−3.3	Good
	130−159	3.4−4.1	Borderline
	160−189	4.1−4.9	High
	> 190	> 4.9	Very high
HDL cholesterol	< 40	< 1.0	Undesirable
	41−59	1.0−1.5	Borderline
	> 60	> 1.55	Good

Cholesterol is produced by the liver and it travels through the bloodstream in the form of lipoproteins, while cholesterol from food consumption is stored in the form of triglycerides. There are two forms of lipoproteins, low density lipoproteins (LDL) and high density lipoproteins (HDL). LDL transport cholesterol from liver to different tissues while HDL is responsible for the removal of excess cholesterol from tissues and brings it back to the liver for removal from the body. LDL is sometimes referred to as bad cholesterol because a high LDL level leads to fat build up in the arteries. The cholesterol level in the blood is typically measured in milligrams per deciliter (mg/dL) or in millimoles per liter (mmol/L). The combination of LDL, HDL, and triglycerides are called total cholesterol. Table 5.7 presents the diagnostic level of measurements of three types of cholesterol both in mg/dL and mmol/L with their interpretation.

The overall percentage of patients with hypercholesterolaemia in this study was 1.34%, with similar prevalences between HRT users (1.31%) and non-users (1.35%). The prevalence raised steadily among HRT users from the youngest to the oldest age cohort, from 0.84% in 46−55 to 3.72% in 76$^+$ age cohort. The distribution was also similar among non-users but there was a slight decrease in 76$^+$ group (3.12%) compared to the 66−75 (3.49%) group (Table 5.6). MacDonald and Morant (2008) compared the hypercholesterolaemia prevalence data of Health Survey for England (HSE) for 1998 and 2003 to the THIN data from 1998 to 2006 and found a big difference in prevalence between HSE and THIN. In 1998, the overall prevalence of hypercholesterolaemia was 7.8% in THIN women and 68% in HSE. In 2003, the prevalence was 17.7 % in THIN and 73% in HSE, respectively. According to MacDonald and Morant (2008), the prevalences of hypercholesterolaemia by age in female THIN patients in 1998 were 13% in 45−54, 20% in 55−64, 21% in 65−74, and 19% in 75$^+$ group, respectively. This is higher than in this study population except the oldest age group.

5.3.5 Osteoporosis

The nature of osteoporosis and related health hazards were discussed in details in Chapter 2, Section 2.5.1. The overall prevalence of osteoporosis in this study was 2.6% at study entry. The prevalence was slightly higher in HRT users (3.2%) than in the matched non-users (2.4%). The prevalence raised greatly in both

groups as age increased with nearly doubled prevalence in HRT users compared to the non-users in 66−75 group (HRT users: 14.4%, non-users: 6.2%) and 76$^+$ group (HRT users: 27.3%, non-users: 14.7%). From literature review of HRT, it is known that HRT is also prescribed for osteoporosis treatment (Bromley et al., 2004). The high prevalence of osteoporosis among older age group of the HRT users at baseline could be a reason for HRT prescription. The age specific prevalence of osteoporosis in UK women is not known. However, it was estimated that the prevalence of osteoporosis rises in women from 2% at age 50 to more than 25% at 80 years (NICE, 2012). The prevalence of osteoporosis in the UK women over 50 years of age was 21.8% in 2015 (International Osteoporosis Foundation, 2017). In 2010, the overall prevalence of osteoporosis in women was 2.5% in the UK (NICE, 2012). The total prevalence of osteoporosis in the study population was similar to the UK national statistics in 2010.

5.3.6 Type II diabetes

In Chapter 2, Section 2.5.8, the types of diabetes mellitus and previous results regarding its association with HRT were discussed. Among all types of diabetes, type II is the most prevalent form, accounting for 90% of all diabetes HRT users (Whicher et al., 2020). The overall prevalence of Type II diabetes in the extracted population was 2.27%, accounting for 1.62% in HRT users and 2.57% in non-users. non-users have slightly higher prevalence of type II diabetes than HRT users in all age categories and the prevalence increased steadily in both

groups as women aged (see Table 5.6). The lowest prevalence was in $46-55$ age group (HRT users:1.23%, non-users 1.73%), and the highest was in 76^+ age group (HRT users: 6.42%, non-users: 7.22%), respectively. Diabetes prevalence in the UK women rises both by calendar year and age (Diabetes UK, 2019). The rate of increase nearly doubled in the latest decade compared to the previous one (Whicher et al., 2020). The overall prevalence of diabetes in UK women was 1.9% in 1994 and 4.2% in 2006 (Diabetes UK, 2010). In 2006, diabetes prevalences were 3.6%, 6.0%, 10.4%, and 10.6% within $45-54$, $55-64$, $65-74$ and 75^+ age groups in women of England (Diabetes UK, 2010). The total prevalence of type II diabetes in this study population was similar to the 1994 UK statistics.

5.3.7   Oophorectomy and hysterectomy

Oophorectemy and hysterectomy are the surgical procedures of removing the ovaries and uterus. Oophorectomy is termed as unilateral or bilateral depending on the removal of one or both ovaries. If the ovaries are removed with fallopian tubes then it is termed as salpingo-oophorectomy. Overall, the proportion of oophorectomy was nearly double (HRT users: 20%, non-users: 9.89%), and hysterectomy was nearly four-fold (HRT users: 11.15%, non-users: 3.30%) in the HRT users compared to non-users in the study population at baseline. Apart from treating menopausal symptoms, HRT is also widely prescribed for oophorectomised and hysterectomised patients (Bromley et al., 2004), and this could be the cause of high number of these patients among HRT users in the study pop-

ulation. Although these conditions were more prevalent in all age subgroups of HRT users at baseline, the highest prevalence of oophorectomy (32.3%) and hysterectomy (21.1%) were found in HRT users in 66−75 age group.

5.3.8 Summary

This chapter presented the distribution and characteristics of the study population in respect to a number of important covariates that were considered for this study. The prevalences of the selected covariates were calculated for full data as well as for different age categories at first HRT treatment to examine the pattern of these condition in this study population and also to compare it with the UK national statistics. It was observed that nearly three-fourths of women took first HRT at age between 46 to 55 years, i.e either at peri-menopausal or menopausal transition age. There were more healthy weight and overweight women in HRT users and more obese women in non-users. The proportion of ex-smokers and current smokers were slightly higher in HRT users. More than half of the study population lived in low deprivation areas. At the time of study entry, among all of the medical conditions considered, hypertension was the most prevalent in the participants. The prevalence of the majority of the medical conditions was similar in HRT users and non-users of all age groups at baseline. However, the prevalence of osteoporosis was greater in HRT users, and in particular, it was much higher in the older HRT users than the non-users at study entry. The number of oophorectomised and hysterectomised patients were nearly double and

four-fold, respectively, in the HRT users compared to the non-users. Additionally, it was observed that the missingness was higher in non-users than in HRT users.

Chapter 6

Survival modelling of hormone replacement therapy

This chapter presents the development of the survival models for hormone replacement therapy (HRT). The models estimated the hazards of all-cause mortality associated with the HRT treatment in women adjusting for a wide range of important risk factors. The covariate selection for HRT models was based on their importance identified from past research, and expert knowledge within the team. This chapter starts with a description of the coded covariates used in the final models. Then it explains the model development procedures in detail. Finally, the results of HRT models are presented and discussed.

6.1 Coding of covariates

The design and patient selection process of HRT study were explained in Chapter 3. The descriptive statistics of the study population for the selected covariates were presented in Chapter 5. Following the initial distribution check, women who started HRT after the age of 65 and/or born between 1900-1920 with their matched non-users were excluded due to a very small proportion. Thus, the data set comprises 105,199 HRT users who were born between 1921 and 1960, and started the HRT treatment at age 46 to 65 in $1984-2017$, and 224,643 matched non-users for model development. The flow diagram 6.1 shows the selected patients and their status at the end of the study period. These patients were used in model development.

This section describes the coding/recoding of the covariates that were adjusted for the final HRT model. Some covariates with similar characteristics were merged to create a new variable, and covariates with a very low proportion ($<1\%$) of exposure were excluded from the analysis to improve statistical efficiency of the models. To minimise the possibility of biases in the imputed data, only socio-demographic and lifestyle variables with low to moderate proportion of missing values were considered for inclusion. After performing the descriptive statistics check, the variables that were selected for inclusion in the model were age at study entry, type of HRT, year of birth, type 2 diabetes, peripheral vascular/arterial disease (PVD/PAD), hypercholesterolaemia, coronary heart disease
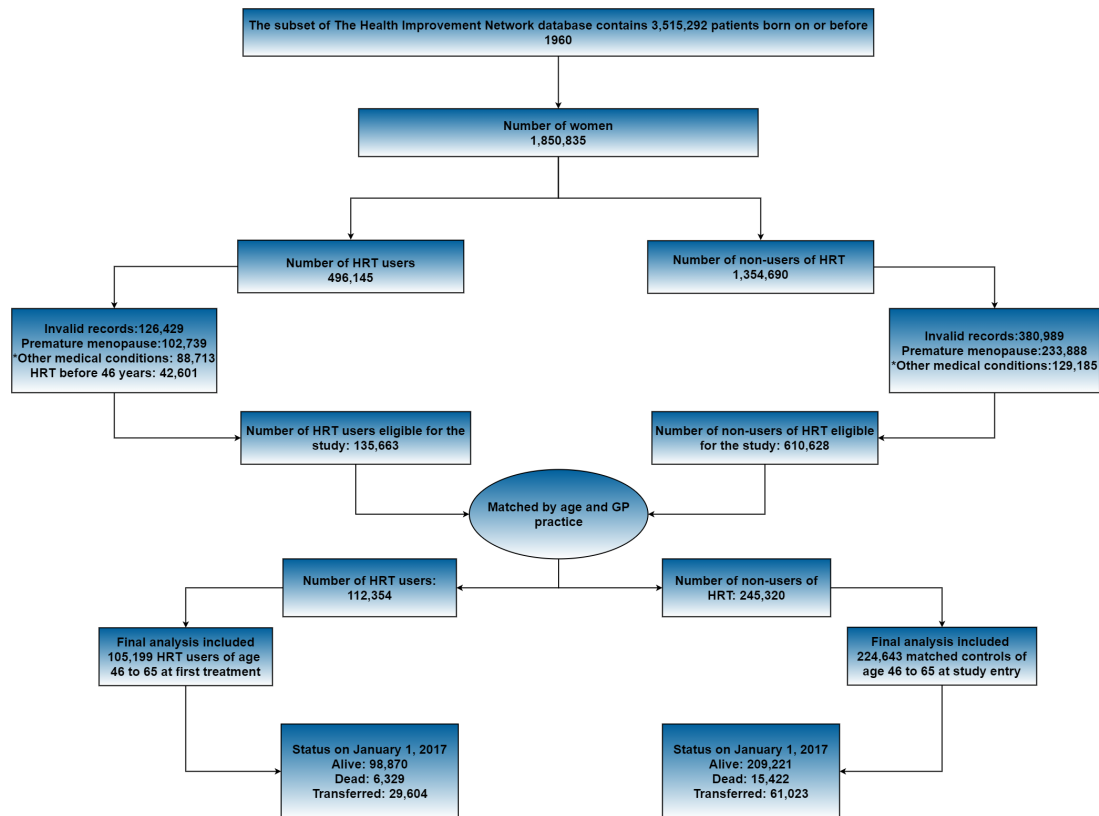
Figure 6.1: Participants selected for the model development. HRT users were matched with the non-users by year of birth an general practice.

(CHD), osteoporosis, hypertension and its treatments, oophorectomy, hysterectomy, categorised body mass index (BMI), smoking, and Townsend deprivation status.

6.1.1 Clinical variables

Hypertension is one of the most important clinical variables as it is strongly related to the development of other chronic conditions, such as stroke, heart disease, and cardiovascular diseases (MacDonald and Morant, 2008). To know the accurate prevalence of hypertension is important for medical research. In

electronic medical records (EMRs), the diagnosis of hypertension is commonly under-coded (Peng et al., 2016). In Chapter 5, Subsection 5.3.1, it was showed that the prevalence of hypertension using Read codes was also much lower in this study population than in the national statistics. Peng et al. (2016) compared the prevalence and treatment rate of hypertension in THIN with the numbers from Health Survey for England (HSE) for adult population in 2011, and found that, the use of diagnosis Read codes only to identify hypertensive patients underestimates the actual hypertension prevalence in the UK. The use of diagnosis Read codes results in the prevalence of 14.49% in THIN females whereas the HSE report showed that there was 28.27% prevalence of hypertension within adult female population of England in 2011. According to the HSE, if any of the following conditions are present: SBP $\geq$ 140 mmHg, DBP $\geq$ 90 mmHg, or prescription of anti-hypertensive drugs, then a person should be considered as hypertensive. The HSE definition was also used in the hypertension survey in Canada and USA (Joffres et al., 2013). The prevalence of hypertension varied greatly by definition in THIN. By considering the combination of anti-hypertensive drug prescriptions and abnormal blood pressure ($\geq$ 140/90 mmHg), Peng et al. (2016) found the prevalence of 41.71%, which is 13.44% higher than the HSE prevalence. However, the use of diagnosis Read codes or two abnormal blood pressure records within 2-year period provided nearly similar (32.81%) prevalence as HSE (28.27%). The prevalence of hypertension in this study population using different definitions was also compared with Peng et al. (2016) study (Table B.1).

Peng et al. (2016) suggested that the diagnosis Read codes or two abnormal blood pressure records within a 2-year period could be used for hypertension surveillance in THIN in order to validate against the HSE. Serumaga et al. (2011) defined hypertension using at least two Read codes and MacDonald and Morant (2008) defined hypertension if any of these conditions were met: a clinical diagnosis Read code, high blood pressure recordings (SBP $\geq$ 140 mmHg or DBP $\geq$ 90 mmHg) or a record of the antihypertensive drug prescription. In this study, the definition suggested by Peng et al. (2016) was used to classify hypertensive patients. That is, if a patient was identified as hypertensive either with a Read code or have had high blood pressure record, then the patient was considered to have hypertension. Among hypertensive patients, those who were treated with antihypertensive drugs, were coded as treated hypertension, and those who were not taking the drugs were coded as untreated hypertension. Patients who were not hypertensive according to this new definition were also classified as having treated hypertension if they had records of antihypertensive drug prescription. Patients who were not identified as hypertensive using the Read codes, and have not had a high blood pressure record (SBP $<$ 140 mmHg and DBP $<$ 90 mmHg), and did not receive anti-hypertensive drugs were coded as no hypertension. Patients with missing values in blood pressure records and no diagnosis of hypertension using Read codes were considered as missing (HRT users: 22.0%, non-users: 31.7%) and handled by multiple imputation. Thus, in the final model, three levels of hypertension category were used: (i) no hypertension, (ii) treated

hypertension, and (iii) untreated hypertension.

The oophorectomy and hysterectomy status was grouped into one variable with four levels; (i) Intact: no history of removal of ovaries and uterus (ii) Hysterectomy with oophorectomy: hysterectomy and at least one ovary removed, (iii) Hysterectomy only: hysterectomy but no history of removal of ovaries, and (iv) Oophorectomy only: no history of hysterectomy, and one or both ovaries removed. Due to a very small proportion of women in the Hysterectomy only category, this level was not included in the final analysis.

Other medical variables adjusted in the model of all-cause mortality were type 2 diabetes, hypercholosterelemea, PAD/PVD, osteoporosis, CHD, and these were all kept at yes or no level.

6.1.2   Demographic and lifestyle variables

In the survival model, Townsend deprivation score, body mass index, and smoking status were considered to be adjusted for. In the final model, Townsend deprivation quintile scores 1 and 2 were recoded as low, score 3 as medium, and score 4 and 5 as high level of deprivation. BMI was initially coded as healthy weight, overweight and obese and further recoded into two levels: healthy weight/overweight and obese as there was no significant survival difference found between healthy weight and overweight women. Smoking status was categorised at three levels: ex-smoker, current smoker, and non-smoker as recorded in THIN.

There were missing values in BMI (exposed: 16.5%, unexposed: 28.0%), smoking (exposed: 9.1%, unexposed: 16.9%), and Townsend deprivation index (exposed: 10.4%, unexposed: 11.8%) and these were estimated by multiple imputation (Table B.2). Missingness were higher among non-users than the HRT users in BMI and smoking status.

Age at study entry were divided into four groups: $46-50$, $51-55$, $56-60$, and $61-65$ years. Year of birth variable was grouped into four decade-long birth cohorts: $1921-1930$, $1931-1940$, $1941-1950$, and $1951-1960$.

6.2 Participants' characteristics and follow-up

The baseline characteristics of the study population used for survival modelling and their follow-up information are presented in Table 6.1. 105,199 HRT users who were born between 1921-1960 and started HRT at ages between 46 to 65 years in $1984-2017$. There were 224,643 age and GP practice matched non-users. The mean ($\pm$ SD) age of women at first treatment was 53 ($\pm$ 5.02) years, and the mean duration of HRT use was 6.0 ($\pm$ 4.8) years. Among HRT users, 17,606 (17%) received estrogen-only and 87,593 (83%) received combined therapy.

The total length of study was 32 years, and the average follow-up of was 13.5 (SD $\pm$ 6.8) years for the exposed group and 13.2 (SD $\pm$ 7.0) years for the unexposed. During follow-up, 21,751 women died in total, of whom 6,329 (6%)

Table 6.1: Selected baseline characteristics and follow-up information for the study participants.

Characteristic	HRT users			Non-users	
	Estrogen-only ($n = 17606$)	Combined HRT ($n = 87593$)	Total ($n = 105199$)	($n = 224643$)	P-value[4]
Death at follow-up	1110 (6.3)	5219 (6.0)	6329 (6.0)	15422 (7.0)	<2.2e-16
Transferred out	5078 (28.8)	24526 (27.9)	29604 (28.1)	61023 (27.2)	7.9e-09
Mean follow-up years (± SD)	13.7 (7.1)	14.0 (6.7)	13.5 (6.8)	13.2 (7.0)	–
Age group at HRT					
46−50	5035 (28.6)	37219 (42.5)	42254 (40.2)	87108 (38.8)	2.7e-14
51−55	6011 (34.1)	30654 (35.0)	36665 (34.9)	72486 (32.3)	<2.2e-16
56−60	4069 (23.1)	13286 (15.2)	17355 (16.5)	40674 (18.1)	<2.2e-16
61−65	2491 (14.1)	6434 (7.30)	8925 (8.5)	24375 (10.9)	<2.2e-16
Birth cohort					
1921−1930	573 (3.3)	1361 (1.6)	1934 (1.8)	5565 (2.5)	<2.2e-16
1931−1940	5450 (31.0)	18940 (21.6)	24390 (23.2)	55047 (24.5)	<2.2e-16
1941−1950	8438 (47.8)	44453 (50.7)	52891 (50.3)	96142 (42.8)	<2.2e-16
1951−1960	3145 (17.9)	22839 (26.1)	25984 (24.7)	67889 (30.2)	<2.2e-16
Hypertension					
No[1]	10017 (56.9)	55266 (63.1)	65283 (62.1)	134337 (59.8)	<2.2e-16
Treated[1]	4419 (25.1)	18657 (21.3)	23076 (22.0)	49421 (22.0)	<2.2e-16
Untreated[1]	3170 (18.0)	13670 (15.6)	16840 (16.0)	40885 (18.2)	0.032
Uterine/ovarian status					
Intact	6779 (38.5)	78214 (89.3)	84993 (80.8)	203625 (90.6)	<2.2e-16
Hysterectomy with Oophorectomy[3]	9945 (56.5)	1067 (1.2)	11012 (10.5)	6502 (2.9)	<2.2e-16
Oophorectomy only	882 (5.0)	8312 (9.5)	9194 (8.7)	14516 (6.5)	<2.2e-16
PAD/PVD	1348 (7.7)	7498 (8.6)	8846 (8.4)	17340 (7.7)	8.9e-12
Diabetes Type II	317 (1.8)	1233 (1.4)	1550 (1.5)	5089 (2.3)	<2.2e-16
CHD	336 (1.9)	1033 (1.2)	1369 (1.3)	3130 (1.4)	<2.2e-16
Osteoporosis	352 (2.0)	2101 (2.4)	2453 (2.3)	4215 (1.9)	0.035
Hypercholesterolaemia	254 (1.4)	972 (1.1)	1226 (1.2)	2605 (1.2)	0.891
Body mass index					
Healthy weight/overweight[1]	13109 (74.5)	69023 (78.8)	82132 (78.1)	161294 (71.8)	<2.2e-16
Obese[1]	4497 (25.5)	18570 (21.2)	23067 (21.9)	63349 (28.2)	<2.2e-16
Smoking status					
Non-smoker[1]	10966 (62.3)	50716 (57.9)	61682 (58.6)	141301 (62.9)	<2.2e-16
Ex-smoker[1]	3187 (18.1)	15854 (18.1)	19041 (18.1)	35269 (15.7)	<2.2e-16
Current smoker[1]	3468 (19.7)	21022 (24.0)	24490 (23.3)	48298 (21.5)	<2.2e-16
Deprivation status					
Low[1]	9648 (54.8)	47738 (54.5)	57386 (54.5)	117488 (52)	<2.2e-163
Medium[1]	3662 (20.8)	17957 (20.5)	21616(20.5)	46950 (20.9)	<2.2e-16
High[1]	4296 (24.4)	21811 (24.9)	26107 (24.8)	60204 (26.8)	<2.2e-16

[1] The reported prevalence of variables with missing values are the mean of ten imputed datasets. Due to missingness in systolic and diastolic blood pressure, missing values were generated in hypertension category.
[2] All values are reported as No. (%) except the mean follow-up time
[3] Hysterectomy and at least one ovary removed
[4] P-values are obtained from a χ^2-test.

were HRT users, and 15,422 (7%) non-users. There were 44 deaths per 10,000 years of follow-up in HRT users compared to 63 deaths in non-users. 29,604 (28%) users and 61,023 (27%) non-users were lost to follow-up.

The model development process was performed in two steps: (i) complete case analysis: the model was fitted for complete data, i.e., by taking out the patients with missing records and their matched counterparts, and (ii) full data analysis: by considering all patients and using the imputed values for participants with missing data. For both complete case and full data analyses, same statistical methods were implemented. In the next section, the model development procedures for HRT study are explained in details.

6.3   Model development

A Cox proportional hazards regression model was initially fitted to estimate the effect of estrogen-only and combined HRT on the hazards of all-cause mortality in women who started treatment first time at age between 46 and 65 years. The outcome variable was time to death from any cause in years; that is, from the date of study entry to the date of death. The model was initially fitted with the following predictors: age group at study entry, birth cohort, type of HRT, PAD/PVD, hypertension and its treatments, hypercholesterolaemia, osteoporosis, CHD, type 2 diabetes, oophorectomy and hysterectomy status, BMI category, smoking and deprivation status. Second order interactions included

HRT with all other predictors, and all of the selected medical conditions with lifestyle variables. Interactions of BMI and smoking with deprivation status, and BMI with smoking status were also included. Missing values in the covariates were estimated by joint modelling multilevel multiple imputation which was described in Chapter 4 in details. The imputation process included all selected variables in their original form and the survival time. There were two levels in the imputation model, patient and practice level, to adjust for the correlation between patients from the same GP practice. The estimation had a burn-in-length of 100 iterations, that means the value of each 100th iteration was registered, and there were 1000 iterations in total resulting in 10 imputed datasets. The distributions of the variables with missing values in the complete and imputed data were similar (Table B.5).

Each imputed dataset contained records of all patients with the imputed missing values and was modelled separately. Backward elimination technique was applied to select the variables at 5% significance level for the main exposures, and 1% significance level for the interaction effects. The variable selection process by backward elimination was described in Chaper 4 Subsection 4.5.2 in detail. The contribution of the covariates in explaining the variation of the hazard in the Cox regression model was assessed by ANOVA (analysis of variance). Grambsch and Therneau's test (Grambsch and Therneau, 1994) was performed to check the non-proportionality of hazards at 5% level of significance and was found to be significant for the variables age group and birth cohort (Table B.7). This test

is based on the correlation between scaled Schoenfeld residuals of the coefficients and log of survival time. Schoenfeld residuals are calculated as the predictors' value for the individual who got the event minus its expected value assuming that the hypothesis of the model holds, and thereby there is a separate residual for each individual for each covariates (Bellera et al., 2010). The non-proportionality was also verified by plotting the residuals of the time-variant coefficient $\beta(t)$ against survival time (Figure B.1). A smooth plot of the Schoenfeld residuals is used to directly visualise the log hazard ratio over time.

To handle the non-proportionality, Weibull-Double-Cox regression model was used. This model is capable of estimating both the shape for the time-variant covariates that violated the PH assumption and the scale parameter. Weibull-Double-Cox model replaces the unspecified baseline hazards function in the Cox regression model by a Weibull baseline hazard function and incorporates an additional Cox-regression term with the shape. Both scale and shape parameters contributed to the estimation of the hazards for the time-variant variables. The Weibull-Double-Cox model was chosen because the underlying baseline hazards of the study population fitted well with the Weibull distribution (Figure B.2). General practice was included in the model as a random effect or frailty to account for the unobserved heterogeneity of patients across practices. The details of parameter estimation using this model were described in Chapter 4.

A subgroup analyses was also performed based on the age at HRT initiation categorised into 5-year intervals to investigate the impact of age at HRT

initiation on the hazards of all-cause mortality. Four separate age subgroup models were fitted for age groups $46-50$, $51-55$, $56-60$, and $61-65$ years. The Weibull-Double-Cox model were fitted for subgroup analyses as the PH assumption was also violated in all subgroup models. For each subgroup model, both complete case and full case (imputed data) analysis were performed, and the same sets of explanatory variables and their interactions were adjusted for.

Results from the models on ten imputed datasets were pooled by Rubin's rules describe in Chapter 4. The goodness-of-fit of the models was evaluated by the concordance statistic, loglikelihood, and AIC (Table B.8). All analyses were performed in the statistical software R (version 3.6.1), using the package 'survival', 'MASS','ucminf', 'rms', and 'hmisc'. R package 'jomo' was used for joint modelling multiple imputation, and package 'sql' was used for re-matching data for complete case analysis. Both complete and full case models for all ages included the following significant scale parameters: age category at study entry, birth cohort, type of HRT, type 2 diabetes, coronary heart disease, hypertension and its treatments, oophorectomy and hysterectomy status, body mass index, smoking, and deprivation status, and the following significant shape parameters: age category at study entry and birth cohorts. There were two significant scale interaction parameters: BMI with smoking and type 2 diabetes with smoking. All explanatory variables and their significant interactions in the model for all ages were also significant in the majority of age subgroup models. There were no significant interactions of HRT with any other covariates in the full case or

subgroup models, meaning that the survival effect of HRT on the risks of all-cause mortality were the same across all sub populations.

In the next section, pooled results for full data (all age combined) and age-subgroup models are discussed. Results from the complete case analyses (without missing data) for all models will be presented in the Appendix B, Figure B.4 - B.8.

6.4 Results of the survival modelling

Results of the final survival models developed for the full dataset and for age subgroups are displayed in forest plots. The adjusted and unadjusted effects of HRT on the hazards of all-cause mortality for the full data and for four age sub-groups at first treatment are presented in Figure 6.2. Results related to surgery, medical conditions, and other treatments are presented in Figures 6.3 and 6.4, respectively. The estimated hazard ratios for all other significant covariates and their interactions for the full data and for subgroup analyses are presented in Appendix B, Figures B.3−B.8.

The adjusted hazard ratios of all-cause mortality associated with HRT were time invariant. Overall, the hazard of death was lower in combined HRT users compared to non-users, and there was no significant impact of estrogen only formulation on mortality. Combined HRT reduced the hazard of all-cause

mortality by 9% (HR, 0.91; 95% CI, 0.88 - 0.94) in women of age between 46 and 65 years at the time of first HRT treatment compared to the non-users of the same age. The hazard ratio (95% CI) of estrogen only HRT users was 0.99 (0.93 - 1.07). In age subgroup analyses, it was found that combined HRT reduced the risks of death by 13%, 12%, and 8% in women who received first treatment at age 51 to 55 (HR, 0.87; 95% CI, 0.82 - 0.92), 56 to 60 (0.88; 0.82 - 0.93), and 61 to 65 (0.92; 0.85 - 0.98), respectively. The effect of combined HRT was not statistically significant in women who started treatment between 46 and 50 years (0.98; 0.92 - 1.04). There was also no significant impact of estrogen-only HRT on mortality in all subgroups (Figure 6.2).

Both oophorectomy and hysterectomy were associated with improved survival prospects. Compared with the intact group, women who have had hysterectomy with oophorectomy had the overall mortality reduction of 24% (HR 0.76; 95% CI, 0.71 - 0.81), and in age subgroups, the highest reduction was in the oldest (61-65) age cohort (0.72; 0.64 - 0.81) and the lowest was in the youngest (46-50) age cohort (0.81; 0.65 - 0.99) at first treatment. In the full model, women who have had oophorectomy only had the hazard of mortality of 0.86 (95% CI, 0.82 - 0.91), and in the age subgroups, the hazards in the youngest to oldest age cohorts were 0.94 (0.82 -1.08), 0.78 (0.70 - 0.86), 0.83 (0.76 - 0.92), and 0.93 (0.85 - 1.03), respectively compared to the intact group. Due to a very small number of hysterectomised women who have not had oophorectomy (< 0.1%), this group was excluded from the analysis.

Age cohort	HRT treatment	Unadjusted HR (95% CI)	Adjusted HR (95% CI)
Overall	No (reference group)		
	Combined HRT	0.79 (0.76-0.81)	0.91 (0.88-0.94)
	Estrogen-only	0.87 (0.82-0.93)	0.99 (0.93-1.07)
46-50	No		
	Combined HRT	1.02 (0.96-1.09)	0.98 (0.92 - 1.04)
	Estrogen-only	0.93 (0.80-1.08)	1.01 (0.84 - 1.21)
51-55	No		
	Combined HRT	0.84 (0.79-0.89)	0.87 (0.82-0.92)
	Estrogen-only	0.83 (0.74-0.93)	1.03 (0.89-1.18)
56-60	No		
	Combined HRT	0.81 (0.76-0.87)	0.88 (0.82-0.93)
	Estrogen-only	0.79 (0.71-0.89)	0.98 (0.86-1.12)
61-65	No		
	Combined HRT	0.87 (0.81-0.94)	0.92 (0.85 - 0.98)
	Estrogen-only	0.74 (0.65-0.83)	0.93 (0.81 - 1.07)

Unadjusted Hazard Ratio: 0.5 0.7 0.9 1.1 1.3 1.5
Adjusted Hazard Ratio: 0.5 0.7 0.9 1.1 1.3 1.5

Figure 6.2: Unadjusted and adjusted hazard ratios of all-cause mortality associated with the use of HRT by age at first treatment. The age categories included patients who started HRT at that age and their matched non-users. The hazard ratios (95% confidence intervals) were adjusted for age at first HRT, birth cohorts, type of HRT, oophorectomy/hysterectomy status, type II diabetes, coronary heart disease (CHD), hypertension and its treatments, deprivation status, body mass index, and smoking status. General practice was also included in the model as frailty.

Age cohort	Oophorectomy/ hysterectomy status		Adjusted hazard ratio (95% CI)
Overall	Intact (reference group)		
	Hysterectomy with oophorectomy		0.76 (0.71-0.81)
	Oophorectomy-only		0.86 (0.82-0.91)
46-50	Intact		
	Hysterectomy with oophorectomy		0.81 (0.65-0.99)
	Oophorectomy-only		0.94 (0.82-1.08)
51-55	Intact		
	Hysterectomy with oophorectomy		0.74 (0.65-0.85)
	Oophorectomy-only		0.78 (0.70-0.86)
56-60	Intact		
	Hysterectomy with oophorectomy		0.78 (0.69-0.87)
	Oophorectomy-only		0.83 (0.76-0.92)
61-65	Intact		
	Hysterectomy with oophorectomy		0.72 (0.64-0.82)
	Oophorectomy-only		0.93 (0.84-1.02)

0.6 0.8 1
Adjusted Hazard Ratio

Figure 6.3: The adjusted hazard ratios of all-cause mortality associated with a history of oophorectomy and hysterectomy. The age categories included patients who started HRT at that age and their matched non-users. The hazard ratios (95% confidence interval) were adjusted for age at first HRT, birth cohort, type of HRT, type II diabetes, coronary heart disease (CHD), hypertension and its treatments, deprivation status, body mass index, and smoking status. General practice was also included in the model as frailty.

There were increased hazards of mortality in women with both treated and untreated hypertension. Overall, the hazard ratios of death from all-cause in the treated and untreated hypertensive women were 1.51 (95% CI, 1.43 - 1.59), and 1.31 (1.24 - 1.38), respectively compared to women without hypertension. In the age subgroup models, these findings did not differ much both in the treated and untreated groups (Figure 6.4). Patients who have had coronary heart disease (CHD) also had increased risks of mortality. Overall, women with CHD had 1.5 times higher hazard of death than women without the condition (HR, 1.52; 95% CI, 1.41 - 1.64). Similar to hypertension the hazards of CHD did not vary much in age subgroups (Appendix B Figures B.4 - B.8).

Survival also significantly differed by patients' socioeconomic status. The hazards of all-cause death were greatest in women with high level of deprivation, and the risks was greater in women at medium level of deprivation compared to those at low deprivation. Overall, women who lived in highly deprived areas faced 42% increased hazards (HR, 1.42; 95% CI, 1.38 - 1.47) of death compared to the women in low deprivation areas. In the age subgroups, the hazards in high deprivation area were greatest in the youngest age cohort (1.53; 1.42 - 1.64) and lowest in the oldest age cohort (1.30; 1.20 - 1.36) at first treatment. In women at medium level of deprivation, the hazard of death in the full data was 1.17 (1.13 - 1.21), and in age subgroups the lowest hazard was in 51-55 age group and highest was in 61-65 age group, respectively compared to the group with low deprived (Figure B.3).
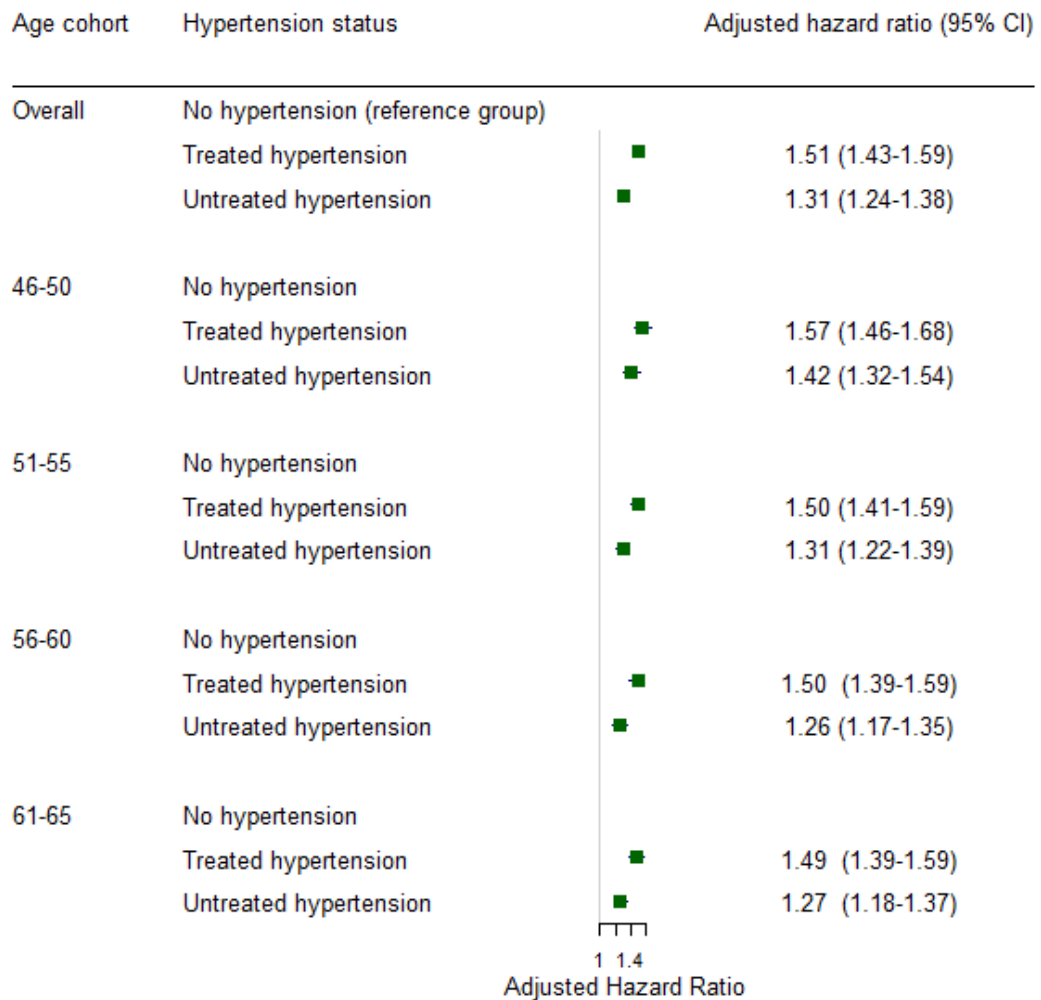
Figure 6.4: The adjusted hazards of all-cause mortality associated with hypertension and its treatments. Patients who were treated with anti-hypertensive drugs were included in the treated hypertension group. The hazard ratios (95% confidence interval) were adjusted for age at first HRT, birth cohort, type of HRT, type II diabetes, coronary heart disease (CHD), oophorectomy/hysterectomy status, body mass index, deprivation status, and smoking status. General practice was also included in the model as frailty.

The interaction of BMI and smoking had the most considerable impact on the survival (Appendix B Figures B.4 - B.8). The hazard ratios of all-cause mortality in current smokers compared to non-smokers were higher in healthy-weight/overweight women than in obese women in all age cohorts. In ex-smokers compared to non-smokers, the risk was higher in obese women compared to the healthyweight/overweight women. In the full model, an extra interaction was found between smoking and type 2 diabetes, where the hazard in current smoker compared to non-smoker was 3.28 times higher (3.28; 2.42 - 3.64) in diabetic patients compared to non-diabetics. In ex-smokers, compared to non-smokers, there were 2.14 times increased risk of death in diabetic women than in non-diabetic.

As the hazards of mortality by birth cohort were time-variant, both shape and scale parameters were estimated for this risk factor. The estimated scale parameter "a" and shape parameter "b" in the Weibull baseline hazard function with the scale and shape parameters of each birth cohort contributed to the time-varying hazards for the four birth cohorts. In Figure 6.5, the cumulative hazards of all-cause mortality in full data by age group at study entry and HRT type for women born in four different birth cohorts are plotted. In each birth cohort, women who were on the combined HRT had reduced hazards of mortality, and oestrogen-only HRT had increased hazards of mortality in comparison to the non-users. However, the effect of estrogen-only HRT was found statistically insignificant in the adjusted model. Comparing the hazards in four birth cohorts,
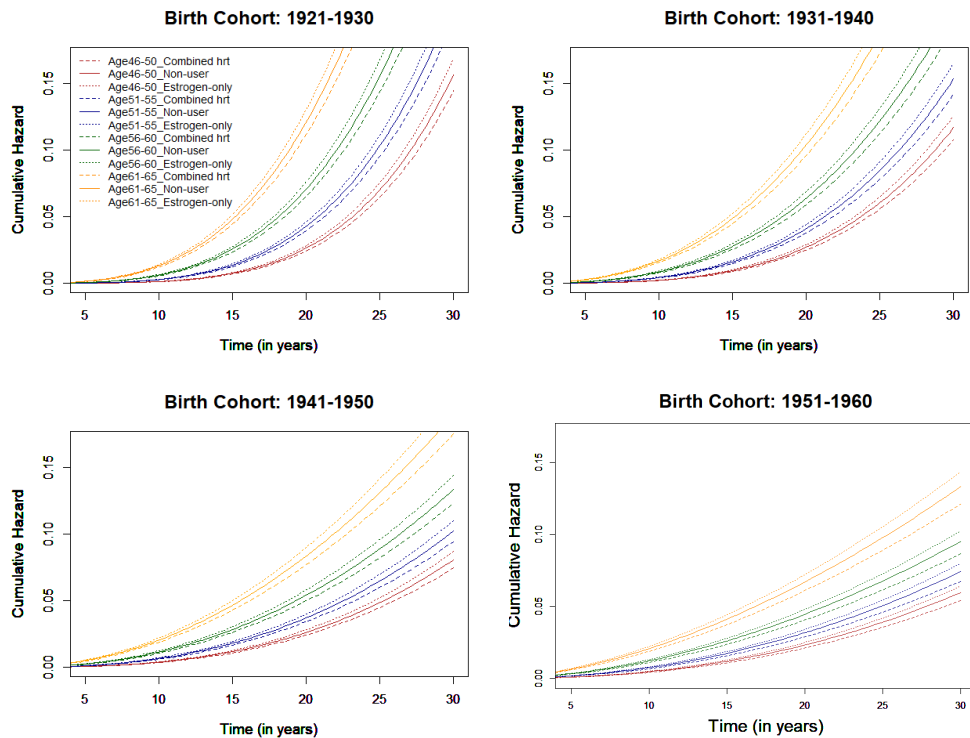
Figure 6.5: Cumulative hazard plots of all-cause mortality for four age sub-groups of 46-50, 51-55, 56-60, and 61-65, respectively, at first HRT treatment by HRT type in four birth cohorts.

it was found that longevity increased in women born in the later birth cohorts compared to women who took HRT at the same age and of same type but were born in the earlier birth cohorts.

Finally, there was significant heterogeneity among patients by general practice. The variance of the frailty was 0.16 (95%CI, 0.12 - 0.17) with standard error of 0.004 in the full data and it ranged from 0.11 (0.002) to 0.16 (0.005) in the age subgroup models.

In age subgroup analyses, the variables that contributed the most to the survival differences were type of HRT, oophorectomy/hysterectomy, and depri-

vation status, and the variables that contributed the least to the variation of the survival were hypertension and its treatments, and coronary heart disease.

6.5   Model performance

The concordance between the estimated hazards of all-cause mortality and patient's survival time in model for full data (all ages) was 0.68 (standard error, 0.002), and for the full subgroup models, its values range from 0.76 (0.005) in $46-50$ age group to 0.81 (0.004) in $61-65$ groups, indicating a good-fit (Therneau and Atkinson, 2020). The loglikelihoods were higher in the imputed models than in the complete case models meaning that the imputed models are more robust. See Table B.8 for all results.

The survival models that included only HRT (unadjusted models) estimated lower hazards of mortality than the adjusted models (Figure 6.2). Unadjusted models also had the lower loglikelihood compared to the adjusted model. The difference in the estimated hazards between the unadjusted and adjusted models, and the corresponding model performance demonstrate the importance of adjusting for important confounders when estimating the effects of medical conditions or treatments. The fact that loglikelihood in full case models is lower than in the complete case model indicates that importance of considering more data to obtain a robust model.

181

Chapter 7

Morbidity analyses at follow-up

This chapter presents the analyses of some chronic medical conditions that the study population developed after study entry. Non-parametric survival analysis techniques were used to estimate the probabilities of diagnosis of a medical condition at follow-up for both HRT users and matched non-users. The prevalence of various medical conditions at follow-up were calculated and presented for both groups. The unadjusted Cox proportional hazard models were fitted to estimate the hazards of developing a number of chronic medical conditions for four age groups by first HRT treatment. Section 7.1 of this chapter describes the prevalence of selected medical conditions for HRT users and non-users at follow-up, and Section 7.2 presents the Kaplan-Meier analysis of some of the selected chronic conditions. Results from the age group-wise unadjusted Cox PH models are described in Section 7.3.

7.1 Incidence of medical conditions

The records of diagnosis of some chronic medical conditions, namely dementia, osteoporosis, hypertension, hypercholesterolaemia, CHD, PAD/PVD, heart failure, type II diabetes, oophorectomy, hysterectomy, myocardial infarction (heart attack), and breast cancer among study participants after study entry were identified using the corresponding clinical Read codes. The incidence of these conditions at follow-up was calculated for HRT users and non-users, and by type of HRT, separately. Among the study participants, those who had the above mentioned conditions at baseline were excluded to obtain the incidence of the disease after study entry. However, this does not apply to breast cancer, dementia, and chronic heart disease as the study population were free from these conditions at study entry.

Table 7.1 presents the incidence of the selected medical conditions that patients were diagnosed with during follow-up period. Hypertension (HRT users, 40.7%, non-users, 37.8%) was the most common condition developed for both HRT users and non-users, with almost 11% higher incidence in estrogen-only (48.8%) users compared to non-users (37.8%). Peripheral vascular disease (HRT users, 23.8%, non-users, 19.3%) was the second most prevalent condition at follow-up for both groups, around 5% higher in estrogen-only HRT users than in non-users. Osteoporosis (HRT users, 14.3%, non-users, 13.8%) was the next most common condition in HRT users and non-users with nearly equal rate of

Table 7.1: Incidence of selected medical conditions that the study population developed at follow-up by HRT status.

Medical conditions	No.(%) of patients[a]				
	HRT users			Non-users	
	Estrogen-only ($n = 17606$)	Combined HRT ($n = 87593$)	Total ($n = 105199$)	($n = 224643$)	P-value[b]
Hypertension	4893 (48.8)	21701 (39.3)	26594 (40.7)	50866 (37.8)	2.2e-16
PVD/PAD[c]	3912 (24.1)	19041 (23.8)	22953 (23.8)	39931 (19.3)	2.2e-16
Osteoporosis	2605 (15.1)	12102 (14.2)	14707 (14.3)	30396 (13.8)	8.3e-06
CKD[d]	1957 (11.1)	7649 (8.7)	9606 (9.2)	19665 (8.8)	2.2e-16
Diabetes Type II	1454 (8.4)	5739 (6.7)	7193 (7.0)	17030 (7.8)	2.2e-16
Hypercholesterolaemia	1418 (8.2)	6491 (7.5)	7909 (7.6)	13563 (6.1)	2.2e-16
CHD[e]	1227 (7.1)	4582 (5.3)	5809 (5.6)	10327 (4.7)	2.2e-16
Breast cancer	723 (4.1)	4172 (4.7)	4895 (4.6)	6243 (2.8)	2.2e-16
Oophorectomy	890 (16.7)	18146 (23.4)	19036 (23.0)	23812 (12.0)	2.2e-16
Hysterectomy	586 (9.3)	6893 (8.0)	7479 (8.1)	8308 (3.8)	2.2e-16
Dementia	516 (2.9)	2222 (2.5)	2738 (2.6)	5042 (2.3)	1.8e-11
Heart failure	433 (2.4)	1714 (2.0)	2147 (2.0)	5100 (2.3)	3.2e-08
Myocardial infarction	390 (2.2)	1803 (2.1)	2193 (2.1)	4427 (2.0)	4.0e-02

[a] All values are reported as No. (%), percentages were calculated by the number of conditions patients developed at follow-up over the number of patients who did not have that condition at baseline.
[b] P-values are obtained from a χ^2- test.
[c] Peripheral vascular or arterial disease   [d] Chronic kidney disease   [e] Coronary heart disease

diagnosis in both groups. Both estrogen-only and combined HRT users had comparatively more oophorectomies than the non-users at follow-up. The proportion was nearly 5% higher in estrogen-only, and doubled in combined HRT users than in non-users. There were also more hysterectomies in both estrogen-only and combined HRT users compared to non-users but the proportion is much lower than that of ooporectomised women in all groups. There was slightly higher proportion of CHD events recorded in estrogen-only HRT users compared to the non-users. Although the proportion of diagnosis of breast cancer (HRT users, 4.6%, non-users, 2.8%) was much lower than other conditions such as, hypertension, PAD/PVD, and osteoporosis at follow-up, the HRT users were diagnosed with more breast cancer than the non-users, 1.3% higher in estrogen-only and nearly 2% higher in combined HRT users. The incidence of dementia, heart failure, and myocardial infarction was comparatively low in both groups at follow-up with similar percentages of diagnoses across all groups.

The incidence of the majority of the selected medical conditions was slightly higher in HRT users than in non-users group at follow-up. However, it should be taken into account that there were more missing records in non-users than in HRT users (Table B.2). The higher proportions of diagnosis of the selected medical conditions among HRT users during follow-up could be because the HRT users received the HRT treatment and for that reason they visited the GP practice more frequently than the non-users, and hence their health status was checked and updated more often than in the non-user group.

7.2 Non-parametric survival analyses

The Kaplan-Meier (KM) survival analyses of selected chronic medical conditions were conducted to estimate the probabilities of development of the conditions for the estrogen-only, combined HRT, and non-users groups during follow-up. The non-parametric Kaplan-Meier survival analysis procedures were described in Chapter 4 Section 4.3 in details. Figure 7.1 presents the Kaplan-Meier estimates of the time-to-diagnosis of six chronic medical conditions, such as, Type II diabetes, osteoporosis, dementia, breast cancer, CHD, and heart failure at follow-up.

The Kaplan-Meier curves show that, the probability of diagnosis of Type II diabetes was lower in combined HRT users than both in non-users and in estrogen-only users over the entire study follow-up period, and estrogen-only users developed slightly more Type II diabetes compared to non-users after around 10 years of follow-up. Until 10 years of follow-up, there were no differences in heart failure diagnosis among the three groups. However, after 10 years, the probability was lower in combined HRT users than in other two groups. Both estrogen-only and combined HRT users were diagnosed with more breast cancer compared to the non-users over the entire follow-up, and after seven years of follow-up, the proportion started to increase slightly for the combined HRT users. While the proportions of diagnosis of coronary heart disease were nearly the same in combined HRT and non-users, estrogen-only HRT HRT users devel-
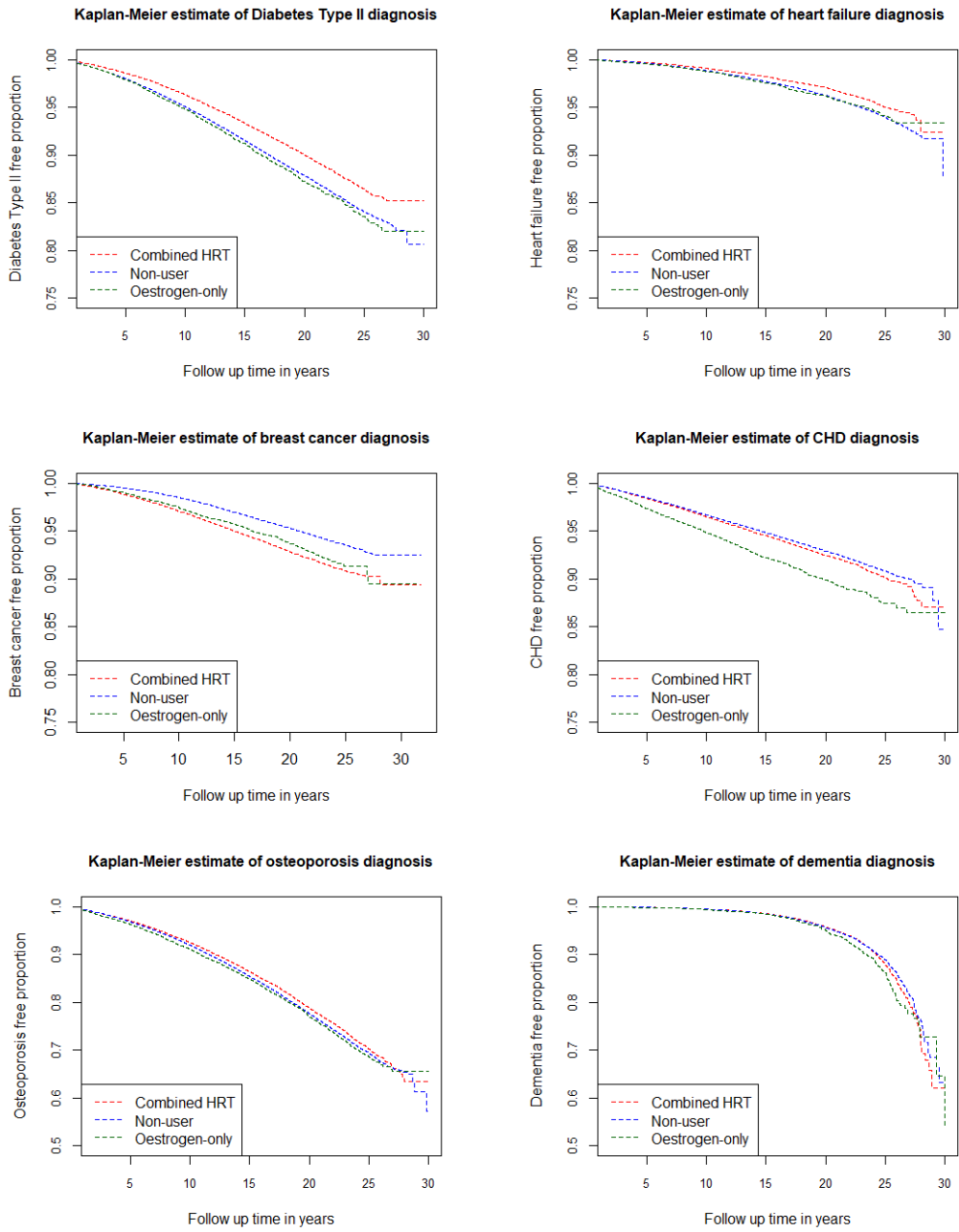
Figure 7.1: Kaplan-Meier survival plots for time to diagnosis of medical conditions in estrogen-only, combined HRT, and non-users at follow-up.

oped relatively higher CHD than the other two groups. A very small increase in CHD events was observed also in combined HRT users compared to the non-users after 13 years of follow-up.

Although there was higher proportions of osteoporosis diagnosis in all groups over time at follow-up, after 10 years combined HRT HRT users had slightly less osteoporosis than the other two groups. The probability of dementia diagnosis remained very low until 15 years of follow-up, but after that there was rapid increase in the diagnosis in all groups, with slightly more diagnosis in estrogen-only HRT users between 18 to 24 years of follow-up.

7.3 Univariate morbidity models of selected medical condition

This section presents results from the unadjusted Cox proportional hazard models that were fitted to estimate the hazards of development of selected chronic medical condition at follow-up for HRT users compared to non-users by age group at study entry. Each univariate model was fitted with shared frailty at general practice, and the outcome was time from study entry to diagnosis of that particular medical condition. The full dataset that was used to develop the survival models in Chapter 6 was used to fit each univariate model. The unadjusted hazards were estimated for four age groups at HRT initiation compared

Conditions	Age group at study entry		Hazard Ratio (95% CI)
PVD	46-50		1.21(1.18-1.25)
	51-55		1.15(1.12-1.19)
	56-60		1.17(1.13-1.22)
	61-65		1.23(1.17-1.30)
Osteoporosis	46-50		1.10(1.04-1.12)
	51-55		0.94(0.91-0.97)
	56-60		0.95(0.91-0.99)
	61-65		1.01(0.96-1.10)
Diabetes (Type II)	46-50		0.94(0.90-0.98)
	51-55		0.80(0.77-0.84)
	56-60		0.80(0.75-0.85)
	61-65		0.82(0.77-0.90)
Breast cancer	46-50		1.56(1.47-1.67)
	51-55		1.61(1.51-1.72)
	56-60		1.62(1.48-1.76)
	61-65		1.64(1.46-1.84)
Dementia	46-50		1.34(1.20-1.50)
	51-55		1.14(1.04-1.25)
	56-60		1.21(1.11-1.32)
	61-65		1.15(1.05-1.26)
Heart failure	46-50		1.11(1.00-1.22)
	51-55		0.90(0.82-0.98)
	56-60		0.81(0.73-0.89)
	61-65		0.74(0.66-0.84)
Myocardial Infarction	46-50		1.40(1.27-1.54)
	51-55		0.96(0.88-1.10)
	56-60		0.84(0.75-0.93)
	61-65		0.97(0.85-1.10)
TIA	46-50		1.47(1.32-1.64)
	51-55		1.28(1.17-1.41)
	56-60		1.23(1.11-1.36)
	61-65		1.21(1.10-1.36)

0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 2
HR and 95% Confidence Interval for conditions at follow-up

Figure 7.2: Unadjusted hazard ratios and 95% confidence intervals of the conditions diagnosed at follow up for the exposed group compared to unexposed group for age category 46-50, 51-55, 56-60, and 61-65 years at study entry.

to the matched non-users of the same age group. In addition, the hazards of breast cancer development after starting HRT were estimated for estrogen-only and combined HRT separately for four different age groups to investigate the impact of these two types of HRT in more details.

Figure 7.2 presents unadjusted hazard ratios and 95% confidence intervals of the selected medical conditions by age-group. In all age groups, the hazards

189

of developing peripheral vascular disease (PVD) were 15−23% higher in HRT users with the similar hazards in age groups 51-55 and 56-60 years, and a little higher in 46-50 and 61-65 age groups. The hazards of developing osteoporosis were mixed, with a somewhat increased risk in the youngest age cohort (46-50), and somewhat reduced hazards in 51-55 and 56-60 age group, but in the oldest age group (61-65) there was no significant benefit or risk. In all age groups, there were reduced risks of developing type II diabetes in HRT users with 20% reduction in 51-55 and 56-60 age group, 18% in 61-65, and 6% in 46-60 years age group.

In each age group, the unadjusted hazards of developing breast cancer at follow-up were more than 1.5 times higher in HRT users than in non-users, and these hazards were somewhat increased as age increased at starting HRT treatment. In all age group, the risks of developing dementia were 14−34% higher in HRT users compared to non-users, where the youngest age group at HRT initiation had the highest hazard and 51-55 group age at first HRT had the lowest hazard.

HRT users had reduced risks of heart failure compared to the non-users at follow-up and this risk decreased with increased age at HRT initiation except for the youngest cohort, where risk was not significant. There were 40% higher risks of myocardial infarction in the youngest age group, and 16% reduction in the age group 56-60, but the other two groups showed non-significant hazards. All HRT users had higher risks of transient ischaemic attack (TIA) with the highest

risks (47%) in the youngest age group at first HRT treatment.

Table 7.2 presents the age group-wise unadjusted hazard ratios and confidence intervals of breast cancer development at follow-up for estrogen-only and combined HRT users compared to non-users, and Figure 7.3 shows the corresponding forest plot. Several studies showed the increased risks of breast cancer associated with combined HRT and our unadjusted analysis agrees with these studies. While there were an increased risks of breast cancer in both estrogen-only and combined HRT users compared to non-users, the risks were higher in combined HRT users than in estrogen-only users except for the youngest age group. The risks did not increase much with increased age at starting HRT for estrogen-only users, but the risks increased significantly in combined HRT users who started HRT at older age. For age 46-50 at first treatment, the risks were the same for both types of HRT users, but for 61-65 age group, combined HRT users had 2-fold higher risks.

7.4 Concluding remarks

This chapter presented the descriptive analysis of various chronic medical conditions that the HRT users and non-users developed at follow-up. The incidence of these diseases was estimated for both groups and non-parametric survival analyses were conducted to estimate the probability of diagnosis of the condition for estrogen-only and combined HRT group, and also for the unexposed group, over

Table 7.2: Unadjusted hazard ratios and 95% confidence intervals of breast cancer development at follow up for estrogen-only and combined HRT

Age group at first HRT	Estrogen only HR (95% CI)	Combined HR HR (95% CI)
46−50	1.56 (1.46−1.68)	1.56 (1.41−1.72)
51−55	1.58 (1.48−1.69)	1.81 (1.57−2.09)
56−60	1.59 (1.46−1.73)	2.02 (1.59−2.55)
61−65	1.61 (1.44−1.81)	2.10 (1.48−2.96)
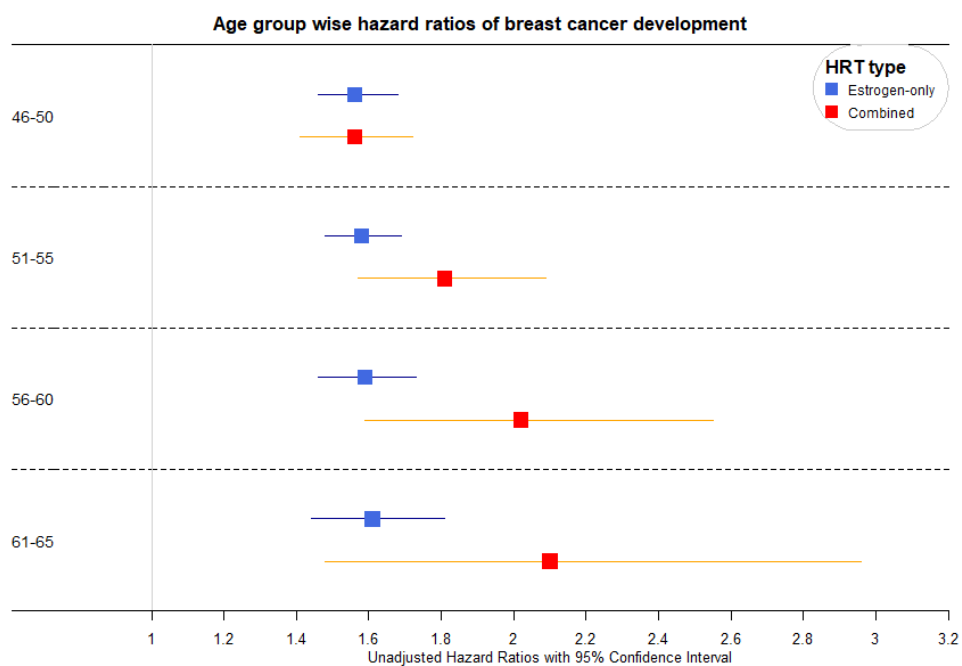
Figure 7.3: Forest plots of unadjusted hazard ratios and their 95% confidence intervals for breast cancer development at follow-up in estrogen-only and combined HRT users compared to non-users.

time. Unadjusted Cox proportional hazard models with frailty estimated the hazards of diagnosis of these conditions at follow-up for HRT users compared to non-users by age group at HRT initiation. Unadjusted models were fitted for each medical condition due to time limitation. The incidence of most medical conditions at follow-up was somewhat higher among HRT users compared to the non-users. This could be due to a higher missingness in non-users, as the HRT users may have visited their GP more frequently than the non-users, allowing their health status to be checked and updated more regularly.

Chapter 8

Survival model for estimating residual life expectancy

This chapter presents a model which is developed for calculating women's life expectancy at postmenopausal ages. First, the meaning of life expectancy and its importance in various sectors are described briefly. Then, the model implementation process and results from the model are presented. Next, the methods for calculating patients' residual life expectancy using the model parameters are explained. Finally, some scenario-based estimates of life expectancy of patients at various ages are presented.

8.1 Life expectancy

Life expectancy is a statistical estimate of the average time a person is expected to survive. More broadly, it is the average number of years of life that one can expect to live at a particular age. Life expectancy ideally depends on a person's birth year, current age, and several other factors including socio-demographic status, lifestyle choices, and health, and it does not always remain stable. Actuaries and insurance industry constantly deal with life expectancy to calculate the annuity or insurance premiums for their clients, and therefore identifying the potential trajectories of life expectancy is crucial for them, as well as it is important to the government for resource planning and distribution.

In Chapter 1, the significance of estimating accurate longevity prospects in different sectors including the actuaries, pension providers, and government, was briefly described. Actuaries and demographers usually use life tables to calculate life expectancy. A life table which is also known as mortality table or actuarial table shows the probability of survival of a person at a particular age while considering their lifestyle choices, medical history, and several other factors. There are two types of life tables used in actuarial science: period life tables and cohort life tables. The period life table depicts the mortality rates of a population over a specified time period. A cohort life table, on the other hand, is used to represent the overall mortality rates of a specific group of population over their entire lifetime. Figure 8.1 shows the period and cohort life expectancy
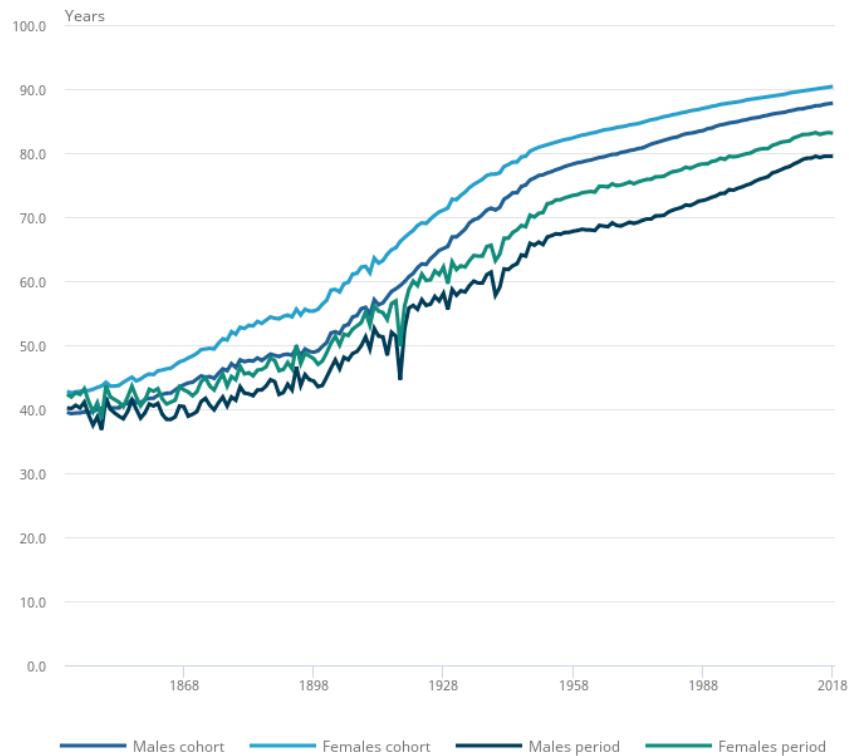
195

Figure 8.1: Period and Cohort life expectancy at birth by sex for England and Wales, 1841 to 2018.

projections at birth in England and Wales for males and females from 1841 to 2018.

Typically, epidemiologists measure the effect of risk factors on individual mortality as hazard ratios, whereas actuaries term this as force of mortality. However, hazard ratios or the force of mortality need to be translated in order to calculate life expectancy. In general, lower hazards mean increased life expectancy, and higher hazards mean decreased life expectancy. Hazard ratios

(HRs) estimated from the Cox regression model are the most commonly used measurement to assess the covariate effects in studies that examine time-to-event outcomes such as survival. However, when proportional hazard assumption in the Cox model is violated, HR is no longer a reliable measurement. Previous studies calculated life expectancy by taking the mean survival difference between two time points, such as at study entry and at the end of follow up, in the case of non-proportional hazards (Royston and Parmar, 2011; Trinquart et al., 2016). More recent studies showed that including an assumption of a parametric distribution of the survival in the study population gives better results when calculating life expectancy (Dehbi et al., 2017; Kulinskaya et al., 2020). Kulinskaya et al. (2020) estimated the changes in individual and period life expectancy due to medical advances and health interventions by translating the hazard ratios from Cox regression model combined with the Weibull or Gompertz baseline distribution. Begun et al. (2019) showed that the use of parametric (Weibull or Gompertz) Double-Cox model provides flexible tools to handle the non-proportionality by allowing for the estimation of the hazards of the covariates that violate the PH assumption. Because of medical advancements and improvements in healthcare and treatments, it is now more common in survival studies with long follow-up that the hazards for many potential risk factors are no longer constant. Thus, calculating life expectancy using methodology from Begun et al. (2019) model fitted to primary care data may provide a more accurate projection of future life expectancy.

In this study, a Weibull-Double-Cox model based on the survival distribution of the selected study population was fitted, and the results from this model were used to estimate women's residual life expectancy (RLE) at various postmenopausal ages. In the next section, the model implementation process is explained first. Following that, the methods for calculating RLE from the estimated model parameters, and the calculated results are presented.

8.2   Model implementation

The study population selected to develop the survival model of HRT for all ages was considered for the development of a survival model for calculating life expectancy. However, a new survival model was implemented as the survival model developed to estimate the hazards of all-cause mortality associated with HRT in Chapter 6 was not suitable for calculating life expectancy. This is because the age at study entry was considered as a categorical variable in the previous survival models. Past research also suggested that age needs to be included as a continuous variable to calculate life expectancy, because the risk of mortality increases monotonically with age, and age is independently associated with overall survival (Liu et al., 2019). Moreover, the age-subgroup analyses of HRT and all-cause mortality in this research found decreased survival with increased age at first HRT treatment.

The study population included patients who had started combined or

estrogen-only HRT between the ages of 46 and 65, and their matched non-users. In full data, there were 105,199 HRT users who started HRT within this age in 1984−2017, and 224,643 matched non-users. A total of 21,751 (case:6,329, control:15,422) deaths were recorded during follow-up. The characteristics of the study population were described in the previous chapters.

Both complete case and full data analysis were performed for the model development. For full data analysis, ten imputed datasets were used and the estimated parameters were pooled using Rubin's rules. The final models included the following significant covariates: age at study entry (continuous), birth cohort (1921-1930, 1931-1940, 1941-1950, 1951-1960), hypertension (yes, no), oophorectomy and hysterectomy status (intact, hysterectomy with oophorectomy, oophorectomy only), coronary heart disease (yes, no), deprivation status (low, medium, high) and the following significant interactions: BMI (healthyweight/overweight, obese) and smoking (non-, ex-, current smoker) and type 2 diabetes (yes, no) and smoking (non-, ex-, current smoker). The significant covariates and their interactions were initially selected from the Cox proportional hazards model at 5% level of significance for the main exposures, and 1% level of significance for the interaction effects. Patients' survival time was the time from first HRT prescription (for non-users, time from study entry) to death from any cause in years.

The violation of PH assumptions in the Cox model was found for the birth cohort only, which was confirmed by the Grambsch and Therneau's test (Gramb-

sch and Therneau, 1994). As the participants' survival follows the Weibull distribution (Figure B.2), the Weibull-Double-Cox model with frailty at general practice level was then fitted to estimate the shape and scale parameters for the respective covariates. The pooled results from the imputed models are presented in Table 8.1.

Hazard ratios for HRT treatments were constant throughout the follow-up, meaning that the survival chances did not differ for patients with the same treatment status. The hazard of mortality in the study population increased 1.11 times (95% CI, 1.10−1.12) with each year of age increase. Compared to patients who lived with low level of deprivation, the hazard of mortality was higher in both medium and high levels of deprivation. Compared to healthyweight/overweight (H/O) and non-smoker (NS) women, the hazards of death were highest in obese and current smoker (CS) women. Similarly, patients who had type 2 diabetes (DM2) and were CS had the greatest hazard compared to patients who were non diabetic and NS. The variance of the frailty in the model was 0.098 (95%CI, 0.076−0.127), indicating that the survival prospects varied greatly depending on the general practice. Finally, the concordance of the model was 0.68 with a standard error of 0.002, indicating that the model provides good fit to the data.

The methods for calculating residual life expectancy from estimated model parameters are described in the following section. Some scenario-based average life expectancies at different postmenopausal ages are presented at the end of next section.

Table 8.1: Parameter estimates from the Weibull-Double-Cox model on imputed data

Variables	Estimate	95% CI	P-value
Weibull Scale (a)	79.89	71.27-89.56	0
Weibull Shape (b)	3.25	3.03-3.50	0
Birth cohort (shape)			
1931-1940	0.88	0.82-0.95	0.0023
1941-1950	0.79	0.73-0.86	0
1951-1960	0.71	0.64-0.79	0
Birth cohort (scale)			
1931-1940	0.57	0.42-0.58	0.0005
1941-1950	0.38	0.28-0.51	0
1951-1960	0.27	0.18-0.39	0
Age (scale)	1.11	1.10-1.12	0
HRT treatment (scale)			
Combined HRT	0.89	0.88-0.97	0.0117
Estrogen-only	1.11	0.95-1.21	0.2286
Deprivation status (scale)			
High	1.46	1.38-1.55	0
Medium	1.22	1.14-1.31	0
Uterine/ovarian status (scale)			
Hysterectomy with oophorectomy	0.76	0.68-0.85	0
oophorectomy only	0.87	0.80-0.95	0.0018
Hypertension (scale)			
Yes	1.45	1.38-1.53	0
CHD (scale)			
Yes	1.61	1.43-1.79	0
BMI & smoking (scale)			
H/O & CS	2.13	1.99-2.27	0
H/O & ES	1.43	1.31-1.56	0
Obese & NS	1.31	1.21-1.42	0
Obese & CS	2.19	2.03-2.37	0
Obese & ES	1.64	1.49-1.80	0
DM2 & smoking (scale)			
No & CS	1.44	1.35-1.54	0
No & ES	1.08	0.99-1.16	0.0661
Yes & NS	3.10	2.72-3.54	0
Yes & CS	3.28	2.84-3.78	0
Yes & ES	2.14	1.78-2.57	0
σ^2	0.098	0.076-0.127	
Concordance (se)	0.68 (0.002)		
Loglik	-122734.3		
AIC	245629.5		

8.3 Calculation of residual life expectancy

If $S(t)$ represents the survival function, then in general, the life expectancy $e(z)$ for an individual at age z is calculated using the following formula:

$$e(z) = \frac{\int_z^\infty S(t)\,dt}{S(z)}. \tag{8.1}$$

Survival function in the Weibull-Double-Cox model is defined by the following formula (Begun and Kulinskaya, 2022):

$$
\begin{aligned}
S(t|a,b,\boldsymbol{\beta}_{shape}^\top, &\boldsymbol{\beta}_{scale}^\top, \sigma^2) \\
&= E[S(t|a,b,\boldsymbol{\beta}_{shape}^\top, \boldsymbol{\beta}_{scale}^\top, \sigma^2, u_g)] \\
&= (1 + \sigma^2 H(t|a,b,\boldsymbol{\beta}_{shape}^\top, \boldsymbol{\beta}_{scale}^\top))^{-1/\sigma^2}
\end{aligned}
\tag{8.2}
$$

where u_g is the gamma distributed frailty with mean 1 and variance σ^2, and the hazard function in the above expression is defined by:

$$H(t|a,b,\boldsymbol{\beta}_{shape}^\top, \boldsymbol{\beta}_{scale}^\top, u_g) = \left(\frac{t}{a}\right)^{b\exp(\boldsymbol{\beta}_{shape}^\top u_g)} \exp(\boldsymbol{\beta}_{scale}^\top u_g), \tag{8.3}$$

where t is the time length from the treatment/diagnosis, a and b are the scale and shape parameters of the Weibull baseline hazard function, $\boldsymbol{\beta}_{shape}^\top$ and $\boldsymbol{\beta}_{scale}^\top$ are the estimates for shape and scale parameters in the Cox regression terms in the model.

Then, the residual life expectancy (RLE) for an individual at age z for Weibull-Double-Cox model estimates takes the following form:

$$\mathbf{RLE} = e(z) = \frac{\int_z^\infty S(t|a,b,\boldsymbol{\beta}_{shape}^\top, \boldsymbol{\beta}_{scale}^\top, \sigma^2, u_g)\,dt}{S(z)}. \tag{8.4}$$

To calculate RLE based on the variables included in the Weibull-Double-Cox model, all different combinations of possible values of the covariates were generated, and there were 15,452 combinations in total. This allowed to calculate life expectancy for any possible combination of factors that applies to an individual. Using the Weibull-Double-Cox model parameter estimates, a R program was developed by Ilyas Bakbergenuly (2021) for calculating LE. This was used to calculate life expectancy for women at postmenopausal ages. Table 8.2 presents the estimated average life expectancy for estrogen-only, combined, and non-users of HRT, and the ratio and difference of RLEs between HRT users and non-users at age 55 and 65, based on their birth cohort and deprivation status. Table 8.3 presents the average life expectancy of women by HRT treatment status at age 55, based on their birth cohort, BMI and smoking status.

On average, combined HRT users starting HRT at age 55 had nearly 2 years longer (combined HRT: 9.30 years, non-users: 7.37 years) LE than non-users in the 1921-1930 birth cohort residing in low deprivation level areas, and this difference was reduced by nearly 6 months for women born in the same birth cohort but living in high deprivation areas. Compared to women from the 1921-1930 birth cohort with low deprivation, for women from the 1951-1960 birth cohort with low deprivation, this difference in life expectancy at age 55 increased by nearly 6 years, and it reduced to 4.6 years for women with high deprivation. However, age 65 at HRT initiation, the difference in life expectancy was reduced to less than one year on average for the 1921-1930 birth cohort and nearly two

Table 8.2: Average residual life expectancy for women starting HRT at age 55 and 65 by birth cohort and deprivation status. Other factors included at the same levels were uterine/ovarian status, hypertension, CHD, and interaction of BMI and type 2 diabetes with smoking

Birth Cohort	Deprivation	Age	LE_1^a	LE_2^b	LE_3^c	LED_{13}^*	LED_{23}^{**}	LER_{13}^{***}	LER_{23}^{****}
1921-1930	Low	55	7.03	9.30	7.37	-0.34	1.93	0.95	1.26
1921-1930	Medium	55	6.31	8.29	6.58	-0.27	1.71	0.96	1.26
1921-1930	High	55	5.55	7.26	5.81	-0.26	1.45	0.95	1.25
1931-1940	Low	55	10.66	14.22	11.20	-0.54	3.02	0.95	1.27
1931-1940	Medium	55	9.46	12.63	9.95	-0.49	2.68	0.95	1.27
1931-1940	High	55	8.30	11.03	8.72	-0.42	2.31	0.95	1.26
1941-1950	Low	55	14.82	19.86	15.58	-0.76	4.28	0.95	1.27
1941-1950	Medium	55	13.14	17.65	13.83	-0.69	3.82	0.95	1.28
1941-1950	High	55	11.45	15.34	12.04	-0.59	3.30	0.95	1.27
1951-1960	Low	55	20.18	27.16	21.24	-1.06	5.92	0.95	1.28
1951-1960	Medium	55	18.76	24.78	19.68	-0.92	5.10	0.95	1.26
1951-1960	High	55	15.51	20.90	16.32	-0.81	4.58	0.95	1.28
1921-1930	Low	65	3.56	4.33	3.68	-0.12	0.65	0.97	1.18
1921-1930	Medium	65	3.34	3.97	3.43	-0.09	0.54	0.97	1.16
1921-1930	High	65	3.10	3.64	3.17	-0.07	0.47	0.98	1.15
1931-1940	Low	65	5.10	6.42	5.28	-0.18	1.14	0.97	1.21
1931-1940	Medium	65	4.68	5.81	4.83	-0.15	0.98	0.97	1.20
1931-1940	High	65	4.25	5.21	4.39	-0.14	0.82	0.97	1.19
1941-1950	Low	65	6.90	8.93	7.20	-0.30	1.73	0.96	1.24
1941-1950	Medium	65	6.26	8.02	6.51	-0.25	1.51	0.96	1.23
1941-1950	High	65	5.62	7.10	5.84	-0.22	1.26	0.96	1.22
1951-1960	Low	65	9.28	12.24	9.72	-0.44	2.52	0.95	1.26
1951-1960	Medium	65	8.34	10.91	8.73	-0.39	2.18	0.95	1.26
1951-1960	High	65	6.39	8.78	6.75	-0.36	2.03	0.96	1.24

[a]Average RLE of estrogen-only HRT users, [b]Average RLE of combined HRT users, [c]Average RLE of HRT non-users, [*]Average RLE difference of estrogen-only HRT users and non-users, [**]Average RLE difference of combined HRT users and non-users, [***]Average RLE ratio of estrogen-only HRT and non-users, [****]Average RLE ratio of combined HRT users and non-users

years for women from the 1951-1960 birth cohort, reflecting that the benefits of HRT diminish at older starting age. Although estrogen-only HRT users seemed to have a little shorter life expectancy than the non-users, the hazard ratio for estrogen-only HRT was not significantly different (HR, 1.11, 95% CI, 0.95-1.21, p-value = 0.23) from the non-users meaning that there was no real difference.

Combined HRT users who started HRT at age 55 had 2.4 years longer life expectancy on average than non-users for the 1921-1930 cohort with no DM2 and NS, and nearly 1 year longer life expectancy than non-users among women born in the same birth cohort with DM2 and CS. In the 1951-1960 birth cohort, this difference in life expectancy increased to 7.37 years for women on combined HRT with no DM2 and NS and 3.29 years for women with DM2 and CS.

The estimation of RLE in this study suggest that in women born in 1921-1930, 1931-1940, 1941-1950, and 1951-1960 birth cohorts, and living with low deprivation, no DM2 and non-smokers, a combined HRT user can expect to live an average of 2, 3-4, 4-5, and 6-7 years longer than the non-users if they start HRT at age 55, and nearly 1.5, 2, 3, and 4.5 years longer if they start HRT at age 55 but live in high deprivation area, or have DM2, and are current smokers. Between 1921-1930 and 1951-1960, the average life expectancy of women increased by up to 13 years for estrogen-only and non-users of HRT, but up to 18 years for combined HRT users at starting age 55.

Table 8.3: Average residual life expectancy for women starting HRT at age 55 by birth cohort and interaction of type 2 diabetes (DM2) and smoking status. Other factors included at the same levels were deprivation status, uterine/ovarian status, hypertension, CHD, and interaction of BMI with smoking

Birth Cohort	DM2 & Smoking	Age	LE_1^a	LE_2^b	LE_3^c	LED_{13}^*	LED_{23}^{**}	LER_{13}^{***}	LER_{23}^{****}
1921-1930	No & NS	55	8.68	11.56	9.12	-0.44	2.44	0.95	1.27
1921-1930	No & CS	55	6.81	9.04	7.15	-0.34	1.89	0.96	1.26
1921-1930	No & ES	55	8.20	10.92	8.61	-0.41	2.31	0.95	1.27
1921-1930	Yes & NS	55	4.32	5.55	4.50	-0.18	1.05	0.96	1.23
1921-1930	Yes & CS	55	4.20	5.38	4.37	-0.17	1.01	0.96	1.23
1921-1930	Yes & ES	55	5.32	6.97	5.57	-0.25	1.40	0.96	1.25
1931-1940	No & NS	55	13.29	17.77	13.98	-0.69	3.79	0.95	1.27
1931-1940	No & CS	55	10.36	13.85	10.88	-0.52	2.97	0.95	1.27
1931-1940	No & ES	55	12.54	16.78	13.19	-0.65	3.59	0.95	1.27
1931-1940	Yes & NS	55	6.32	8.33	6.62	-0.30	1.71	0.95	1.26
1931-1940	Yes & CS	55	6.13	8.05	6.41	-0.28	1.64	0.96	1.26
1931-1940	Yes & ES	55	7.97	10.61	8.36	-0.39	2.25	0.95	1.27
1941-1950	No & NS	55	18.57	24.87	19.53	-0.96	5.34	0.95	1.27
1941-1950	No & CS	55	14.39	19.35	15.15	-0.76	4.20	0.95	1.28
1941-1950	No & ES	55	17.49	23.47	18.41	-0.92	5.06	0.95	1.27
1941-1950	Yes & NS	55	8.62	11.49	9.10	-0.48	2.39	0.95	1.27
1941-1950	Yes & CS	55	8.34	11.10	8.75	-0.41	2.35	0.95	1.27
1941-1950	Yes & ES	55	10.97	14.75	11.56	-0.59	3.19	0.95	1.27
1951-1960	No & NS	55	25.38	34.08	26.71	-1.33	7.37	0.95	1.28
1951-1960	No & CS	55	19.61	26.46	20.65	-1.04	5.81	0.95	1.28
1951-1960	No & ES	55	23.90	32.15	25.17	-1.27	6.98	0.95	1.28
1951-1960	Yes & NS	55	11.57	15.58	12.17	-0.60	3.41	0.95	1.28
1951-1960	Yes & CS	55	11.17	15.04	11.75	-0.58	3.29	0.95	1.28
1951-1960	Yes & ES	55	14.87	20.10	15.66	-0.79	4.44	0.95	1.28

[a] Average RLE of estrogen-only HRT users, [b] Average RLE of combined HRT users, [c] Average RLE of HRT non-users, [*] Average RLE difference of estrogen-only HRT users and non-users, [**] Average RLE difference of combined HRT users and non-users, [***] Average RLE ratio of estrogen-only HRT and non-users, [****] Average RLE ratio of combined HRT users and non-users

8.4   Conclusion

This chapter presented the calculations and results on the residual life expectancy of women of postmenopausal age as estimated by fitting a newly developed Weibull-Double-Cox survival model. The model implementation process was outlined, and the method for calculating life expectancy using the model parameters was explained. At the end, some scenario-based life expectancies from the model based on various characteristics at different postmenopausal ages are presented and explained. This study found that, on average, women's residual life expectancy increased by up to 13 years for estrogen-only and non-users, and up to 18 years for combined HRT users between the birth cohorts of 1921-1930 and 1951-1960. This study also found that the difference in residual life expectancy for women on combined HRT at 55 was from 2 to 6 years longer than non-users across the four birth cohorts, but if starting HRT at age 65 this difference was from less than 1 year to 2.5 years. There were no noticeable differences in life expectancy between estrogen-only HRT users and their non-users for all birth cohorts.

Chapter 9

Discussion

This study investigated the long-term effects of hormone replacement therapy on the survival of women in the United Kingdom by developing survival models using electronically recorded primary care data from The Health Improvement Network Database. This chapter discusses these newly developed survival models, with an emphasis on their validity and usability in clinical and actuarial practise. First, the main findings of this research are summarised, and the contributions to the current clinical evidence are discussed. Then the strengths and limitations of this study are addressed. Next, the implications of this research in medical care and treatment are provided. Finally, the overall conclusions of this thesis are given.

9.1 Main findings

For this research, the electronic medical records of 105,199 healthy HRT users who initiated HRT at ages 46 to 65 between 1984 and 2017 and 224,643 matched healthy non-userss were selected from THIN database. This large population based matched cohort study estimated the long-term adjusted effects of HRT on the hazards of overall and age-specific all-cause mortality by developing five survival models. This study also estimated the hazards of developing various chronic medical conditions after starting HRT, and developed a model for calculating life expectancy of women after menopause.

9.2 Models for HRT and mortality

This study investigated the overall and age-specific hazards of all-cause mortality in women born between 1921 and 1960 who started HRT at ages 46 to 65 between 1984 and 2017. After adjusting for important risk factors, this study found that, estrogen-only HRT was not associated with significantly increased or decreased hazards of all-cause mortality at any age, but combined HRT was associated with an overall decreased risk of death from all causes. Age-specific mortality analysis found that combined HRT users who started treatment between the ages of 51 and 55 had the greatest reduction in the hazards of death, and starting combined HRT at an older age diminishes the survival benefit.

Only a few previous studies on HRT and the risks of all-cause mortality assessed the effects of estrogen-only and combined HRT separately (Manson et al., 2017; Stram et al., 2011). Results of this study agree in respect to estrogen-only HRT effects with the Manson et al. (2017) findings from the WHI trial, and in respect to combined HRT with Stram et al. (2011) report. There were, however, some significant discrepancies when comparing this study to these two investigations. WHI was a randomized non-users trial of 13,816 postmenopausal HRT users versus placebo, and Stram et al. (2011) used survey data of a prospective cohort from California Teachers Study. The average age of women taking HRT in both studies was around 63 years, which was more than a decade away from the menopausal transition age. Furthermore, WHI investigated only one single dose of oral estrogen and progesterone, whereas participants in this study received a variety of doses and preparations of oral and transdermal HRT.

Other mortality studies reported varied results on the total impact of HRT on all-cause mortality (Ettinger et al., 1996; Hunt et al., 1987; Grodstein et al., 1997; Salpeter et al., 2009; Folsom et al., 1995; Malek et al., 2019). Results from this study on combined HRT agree with Hunt et al. (1987), Ettinger et al. (1996), Grodstein et al. (1997), and Salpeter et al. (2009) who also found a decreased risk of death from all-cause in HRT users ranging from 27% to 46%. However, compared to the prior studies, this study found less reduction (9%) of hazards of death from all-causes. There were several possible factors that may have caused this difference. Firstly, this study estimated the hazards using large-scale

primary care data, whereas most other studies relied on survey or register data with a smaller number of participants. Secondly, this study analysed the effect of combined and estrogen-only HRT on all-cause mortality separately, while most other studies estimated the total impact of HRT. Other possible explanations for the lower reduction in hazards in this study compared to others include the absence of age-matched non-userss in the majority of observational studies. Some of these studies were criticised for introducing bias by selecting healthy HRT users compared to non-users. In this study, both HRT userss and non-userss were the same age and had similar health characteristics at baseline. In addition, this study estimated hazards of all-cause mortality by adjusting for a wide range of important risk factors, while most other studies only adjusted for demographical and/or lifestyle factors. However, compared to the adjusted analysis, the unadjusted analysis of this study found a greater reduction in the hazards of all-cause mortality in both types of HRT users (combined HRT: 21%, estrogen-only: 13%). Thus, a new finding is that the reduction in hazards of all-cause mortality associated with combined HRT in the general population is lower than previously estimated.

There were no significant interactions between the type of HRT or age at HRT initiation with other morbidities or lifestyle factors such as hypertension or smoking, meaning that the effect of HRT on the hazards of death from all-causes was consistent across different groups. Women who had both oophorectomy and hysterectomy had significantly improved survival prospects. This study found

a significant interaction between BMI and smoking, where obesity and current status of smoking associated with the greatest hazard of death compared to healthyweight/overweight and non-smoker women. Furthermore, results of this study agree with the findings of Drever and Whitehead (1995) regarding significant survival variation in the UK due to the level of deprivation. Finally, this study found significant variation in the survival of patients by general practice.

9.3   Models for HRT and morbidity

This study investigated the impact of HRT on the hazards of developing some life threatening medical conditions at follow-up by fitting univariate survival models for each condition and estimating the probability of developing these medical conditions using KM survival analysis. This study found that women who took estrogen-only HRT had significantly higher risks of developing hypertension, and slightly higher risks of coronary heart disease, stroke, dementia, and breast cancer than non-users. Women who were on combined HRT had lower risk of developing type 2 diabetes, heart failure, and osteoporosis, but an increased risk of developing breast cancer. Moreover, age-specific analysis showed that starting combined HRT at an older age further increased the risks of developing breast cancer. Results from this study partly agree with the latest review on the effect of HRT by the National Institute of Health and Care Excellence (NICE) (NICE, 2021). According to the current NICE guidelines, the benefits of HRT include

the prevention of osteoporotic fractures, colorectal cancer, and cardiovascular disease, while the risks include a slight increase in CHD and thromboembolic events. The current NICE guidelines also state that estrogen-only HRT is associated with little or no change in the risk of breast cancer and combined HRT is associated with an increased risk of breast cancer.

9.4   Life expectancy in HRT users

This study investigated the life expectancy of women at postmenopausal ages based on various important risk factors. A Weibull-Double-Cox model was implemented as it enabled to calculate the life expectancy considering the variables with both time-variant and time-invariant hazards. This study showed that the average survival increased by up to 18 years for combined HRT users, and up to 13 years for estrogen-only and non-users from the oldest birth cohort to the youngest at age 55. This study also indicates that the average difference in life expectancy between combined HRT users and non-users at 55 years was 2 to 6 years across four birth cohorts, and at age 65 the difference was only 1 to 2.5 years. Starting HRT at the age of 55, the highest difference in life expectancy found between combined HRT and non-users was 6 years in the 1951-1960 cohort with low deprivation, and the lowest difference was 1 year in the 1921-1930 cohort with type 2 diabetes and current smoking status.

9.5   Research strengths and limitations

This research made use of electronic primary care records from The Health Improvement Network Database that are broadly representative of the UK general population when adjusted for demographics and deprivation (Blak et al., 2011; MacDonald and Morant, 2008). Due to the availability of information on all prescribed drugs in primary care records, a large number of anonymised HRT users were able to be selected for this study. Because of the matched cohort study design and the exclusion of the selected medical conditions from both HRT userss and non-userss, this study was able to estimate more precisely the effects of HRT on the survival of healthy users compared to healthy non-users. The availability of a wide range of information in primary care records, including comorbidities, treatment history, lifestyle factors, and demographics, allowed this study to adjust for a large number of important confounders and test their interactions. This meant that the survival variations could be investigated in greater depth, making the results more generalisable to the larger population. The use of multiple imputation techniques to replicate missing records allowed nearly all extracted patients to be included in the analyses. Using a Weibull Double-Cox model, this study was able to estimate hazards for time-varying covariates. In addition, the random effect included in the survival model accounted for the dependence of patients from the same general practice. Finally, this study spanned 32 years, with an average patient follow-up for almost 14 years. Because of the long study period, changes in the hazards of various risk factors, as well as more deaths,

could be observed, and life expectancy could be more accurately estimated.

The study participants selected for this study received a great variety of HRT formulations and doses, so these were not differentiated in the analyses. Although many potential risk factors were adjusted for, there could be residual confounding due to a number of other risk factors, such as age at menarche (first menstruation) and menopause, parity (number of children born), diet, and physical activity. These covariates were not adjusted for in the models because they were not systematically recorded in the database. To avoid immortality bias, the duration of HRT use was not considered in the analyses because prolonged use may be confounded with longer survival. Despite the fact that THIN is broadly representative of the UK general population, due to high regional clustering in THIN (Kontopantelis et al., 2018), further research might be needed to validate the results using data from other large UK primary care databases. More missingness in the non-userss than in the HRT userss in the lifestyle and socio-demographic variables could lead to bias in the imputed data. The more missing records in non-userss and a higher proportion of diagnosis of the selected medical conditions among HRT users compared to non-users for the majority of conditions could be explained by the fact that HRT users visited the practise more frequently than non-users as they were receiving treatment, so their health status was checked and updated more frequently. This could be a potential limitation of the morbidity analyses of this study.

9.6 Research implications

The key findings of this research suggests that the long-term hazards of all-cause mortality associated with combined HRT are lower, and estrogen-only HRT was not associated with any significant changes. This study was based on a large sample of healthy women who were registered with UK general practises and were followed up for many years. As a result, this research is more generalisable and strengthens the emerging consensus that the benefits of long-term HRT outweigh the harms for the vast majority of women. This information may help women in making decisions about whether or not to use HRT, as well as help clinicians in making decisions about prescribing HRT.

Current clinical guidelines from the National Institute of Health and Care Excellence in the UK recommend administering combined HRT to symptomatic women with a uterus, and estrogen-only HRT to women without a uterus after discussing its benefits and risks with them (NICE, 2021). The latest NICE guidelines state that HRT prevents osteoporotic fractures, colorectal cancer, and cardiovascular disease if the treatment starts before the age of sixty, while the risks include slight increase of CHD, stroke, and venous thromboembolism. However, NICE has not released any guidance on all-cause mortality yet. In this study, combined HRT users had lower risks of type 2 diabetes, heart failure, and osteoporosis, and estrogen-only users had higher risk of developing hypertension and CHD events than the non-users at follow-up. Although current NICE guidelines

state that estrogen-only HRT has little or no effect on the risk of breast cancer and that combined HRT can increase the risk, this study found an increased risk of breast cancer for both types of HRT. Nevertheless, this study did not observe an increased mortality risks among HRT users. This could be because women are more likely to die from cardiovascular disease, osteoporosis, and dementia than from breast cancer (Climént-Palmer and Spiegelhalter, 2019), and so benefits in these conditions will outweigh the risks associated with rarer conditions. It is, therefore, important to disseminate a balanced information on the potential benefits and risks of HRT and not to overestimate the possible risks, to allow women and their doctors to make an informed choice.

Other findings of this study indicate that the hazards of death for hypertensive women are higher in both the treated and untreated groups, and the hazard ratios the same with increasing age. This research suggest that women who had both oophorectomy and hysterectomy had considerably better chance of survival. The findings of this study also suggest that the survival varies greatly depending on the level of deprivation, and being obese and smoker makes survival chances worse. Additionally, the findings of this research will assist clinicians when treating patients.

The model for calculating life expectancy can help actuaries in setting annuity pricing and insurance premiums, as well as women in making decisions about health management and retirement planning. This model was transformed into an online calculator, which is accessible through the shiny app (Njabulo

Ncube, 2021). There is also an R package available in GitHub (Ilyas Bakbergenuly, 2021) named "mylongevity". Both this app and the R package allow to estimate the life expectancy based on the given characteristics and thereby could be used by physicians, actuaries and individuals.

9.7   Further research

The main focus of this study was to investigate the impact of long-term HRT on the overall survival of women patients in the UK. However, it is also important to conduct thorough investigation of the long term impact of HRT on other health conditions and mortality from those conditions using primary care data. Due to the limited time of this study, it was not possible to fit an adjusted model for each of the medical conditions separately and this was left for further investigation. Although this study found an overall reduction in hazards of death from all-causes in combined HRT users compared to non-users, and also varying reduced hazards in combined HRT users of age groups 51 to 55, 56 to 60, and 61 to 65, starting combined HRT at an age between 46 to 50 showed no significant benefit. Therefore, further investigation is needed to fully understand the reason behind this.

9.8 Conclusions

The primary goals of this research were to investigate how estrogen-only and combined hormone replacement therapy impacted the hazards of overall and age-specific all-cause mortality in postmenopausal women using a large scale primary care data from UK general practices.

This study compared the hazards of all-cause mortality in women on estrogen-only and combined HRT at starting ages between 46 to 65 in 1984-2017 with age and general practice matched non-users and found that combined HRT was associated with a reduced risk of mortality from all causes during a long follow-up, but estrogen-only HRT had no impact. Compared to previous studies, this study found less reduction in hazards in mortality, which is probably due to the absence of matched non-users, small sample sizes, and little adjustment of other confounders in some prior studies. This study filled several gaps in past research and was the first of its kind to use large primary care data to investigate the hazards of all-cause mortality associated with HRT, thereby to strengthen the evidence that the long-term benefits of HRT outweigh the risks.

The findings of this study were published in the peer-reviewed medical journal BJOG: An International Journal of Obstetrics and Gynaecology (Akter et al., 2022). These findings may help women and their doctors in making decisions about HRT use. However, each woman should make an informed decision about the potential risks and benefits, considering her own clinical condition,

concerns and expectations.

Bibliography

Adami, H.-O., Persson, I., Hoover, R., Schairer, C., and Bergkvist, L. (1989). Risk of cancer in women receiving hormone replacement therapy. *International Journal of Cancer*, 44(5):833–839.

Akter, N., Kulinskaya, E., Steel, N., and Bakbergenuly, I. (2022). The effect of hormone replacement therapy on the survival of uk women: a retrospective cohort study 1984- 2017. *BJOG: An International Journal of Obstetrics & Gynaecology*, 129(6):994–1003.

Alzheimer's Association (2018). About dementia. https://www.alz.org/what-is-dementia.asp. Accessed: 2018-03-10.

Anderson, G., Cummings, S., Freedman, L., Furberg, C., Henderson, M., Johnson, S., Kuller, L., Manson, J., Oberman, A., Prentice, R., et al. (1998). Design of the Women's Health Initiative clinical trial and observational study. *Controlled Clinical Trials*, 19(1):61–109.

Anderson, G. L., Limacher, M., Assaf, A. R., Bassford, T., Beresford, S. A., Black, H., Bonds, D., Brunner, R., Brzyski, R., Caan, B., et al. (2004). Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. *Jama*, 291(14):1701–1712.

Balan, T. A. and Putter, H. (2020). A tutorial on frailty models. *Statistical Methods in Medical Research*, 29(2):3424–3454.

Baldereschi, M., Di Carlo, A., Lepore, V., Bracco, L., Maggi, S., Grigoletto, F., Scarlato, G., Amaducci, L., et al. (1998). Estrogen-replacement therapy and alzheimer's disease in the Italian longitudinal study on aging. *Neurology*, 50(4):996–1002.

Begun, A. and Kulinskaya, E. (2022). A simulation study of the estimation quality in the double-cox model with shared frailty for non-proportional hazards survival analysis. *arXiv preprint arXiv:2206.05141*.

Begun, A., Kulinskaya, E., and MacGregor, A. J. (2019). Risk-adjusted cusum control charts for shared frailty survival models with application to hip re-

placement outcomes: a study using the njr dataset. *BMC medical research methodology*, 19(1):1–15.

Begun et al. (2022). Parametric Gamma-Frailty Models with Non-Proportional Hazard Functions. https://rdrr.io/github/AB5103/doubleCoxr/. Accessed: 2022-02-20.

Bellera, C. A., MacGrogan, G., Debled, M., De Lara, C. T., Brouste, V., and Mathoulin-Pélissier, S. (2010). Variables with time-varying effects and the cox model: some statistical concepts illustrated with a prognostic factor study in breast cancer. *BMC medical research methodology*, 10(1):1–12.

Benkhadra, K., Mohammed, K., Al Nofal, A., Carranza Leon, B. G., Alahdab, F., Faubion, S., Montori, V. M., Abu Dabrh, A. M., Zúñiga Hernández, J. A., Prokop, L. J., et al. (2015). Menopausal hormone therapy and mortality: a systematic review and meta-analysis. *The Journal of Clinical Endocrinology & Metabolism*, 100(11):4021–4028.

Beral, V., Collaborators, M. W. S., et al. (2007). Ovarian cancer and hormone replacement therapy in the Million Women Study. *The Lancet*, 369(9574):1703–1710.

Beral, V., Peto, R., and Pirie, K. (2019). Type and timing of menopausal hormone therapy and breast cancer risk: individual participant meta-analysis of the worldwide epidemiological evidence. *The Lancet*.

BHF (2012). Coronary Heart Disease Statistics. https://www.bhf.org.uk/informationsupport/publications/statistics/coronary-heart-disease-statistics-2012. Accessed: 2019-07-15.

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. *Journal of Innovation in Health Informatics*, 19(4):251–255.

Blankers, M., Koeter, M. W., and Schippers, G. M. (2010). Missing data approaches in eHealth research: simulation study and a tutorial for non-mathematically inclined researchers. *Journal of Medical Internet Research*, 12(5):e54.

Blood Pressure UK (2017). Blood pressure facts and figures. http://www.bloodpressureuk.org/mediacentre/Factsandfigures. Accessed: 2019-03-28.

BMA (2013). Confirmation and certification of death. https://www.bma.org.uk/advice/employment/gp-practices/service-provision/confirmation-and-certification-of-death. Accessed: 2019-09-10.

BMJ Best Practice (2018). Peripheral arterial disease. https://bestpractice.bmj.com/topics/en-gb/431/aetiology. Accessed: 2019-03-28.

Boardman, H. M., Hartley, L., Eisinga, A., Main, C., Roqué i Figuls, M., Bonfill Cosp, X., Gabriel Sanchez, R., and Knight, B. (2015). Hormone therapy for preventing cardiovascular disease in post-menopausal women. *The Cochrane Library*.

Brenner, D. E., Kukull, W. A., Stergachis, A., Van Belle, G., Bowen, J. D., McCormick, W. C., Ten, L., and Larson, E. B. (1994). Postmenopausal estrogen replacement therapy and the risk of alzheimer's disease: a population-based case-control study. *American Journal of Epidemiology*, 140(3):262–267.

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics*, pages 89–99.

British Heart Foundation (2017). HEART MATTERS. https://www.bhf.org.uk/informationsupport/heart-matters-magazine/medical/peripheral-arterial-disease. Accessed: 2019-03-28.

British Heart Foundation (July 2016). CVD Statistics. https://www.bhf.org.uk/-/media/files/.../heart.../bhf-cvd-statistics---uk-factsheet.pdf. Accessed: 2018-01-12.

British Medical Association (2018). General practice in the UK. https://www.bma.org.uk/-/media/files/pdfs/press%20briefings/general-practice.pdf. Accessed: 2019-02-04.

Bromley, S., Vries, C., and Farmer, R. (2004). Utilisation of hormone replacement therapy in the United Kingdom. *BJOG: An International Journal of Obstetrics & Gynaecology*, 111(4):369–376.

Buring, J. E. (1987). *Epidemiology in medicine*, volume 515. Lippincott Williams & Wilkins.

Burkman, R. T., Collins, J. A., and Greene, R. A. (2001). Current perspectives on benefits and risks of hormone replacement therapy. *American Journal of Obstetrics & Gynecology*, 185(2):S13–S23.

Cagnacci, A. and Venier, M. (2019). The controversial history of hormone replacement therapy. *Medicina*, 55(9):602.

Calle, E. E., Miracle-McMahill, H. L., Thun, M. J., and Heath Jr, C. W. (1995). Estrogen replacement therapy and risk of fatal colon cancer in a prospective cohort of postmenopausal women. *JNCI: Journal of the National Cancer Institute*, 87(7):517–523.

Cancer Research UK (2018). Trends over time for tobacco statistics. https://www.cancerresearchuk.org/health-professional/cancer-statistics/risk/tobacco#heading-Three. Accessed: 2019-01-28.

Cancer Research UK (2019a). Overweight and obesity statistics. https://www.cancerresearchuk.org/health-professional/cancer-statistics/risk/overweight-and-obesity#heading-Three. Accessed: 2019-11-13.

Cancer Research UK (2019b). Smoking prevalence trends over time. https://www.cancerresearchuk.org/health-professional/cancer-statistics/risk/tobacco#heading-Three. Accessed: 2019-11-13.

Cancer Research UK (July, 2015). Bowel Cancer Statistics. http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer#heading-One. Accessed: 2018-03-19.

Carroll, O. U., Morris, T. P., and Keogh, R. H. (2020). How are missing data in covariates handled in observational time-to-event studies in oncology? a systematic review. *BMC medical research methodology*, 20:1–15.

Cass, I. and Runowicz, C. (1998). Menopause and controversies in hormone replacement therapy. *American Journal of Managed Care*, 4(2):257–262.

Cauley, J. A., Robbins, J., Chen, Z., Cummings, S. R., Jackson, R. D., LaCroix, A. Z., LeBoff, M., Lewis, C. E., McGowan, J., Neuner, J., et al. (2003). Effects of estrogen plus progestin on risk of fracture and bone mineral density: the Women's Health Initiative randomized trial. *Jama*, 290(13):1729–1738.

Cepeda, M. S., Boston, R., Farrar, J. T., and Strom, B. L. (2003). Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *American journal of epidemiology*, 158(3):280–287.

Chowdhury, M. Z. I. and Turin, T. C. (2020). Variable selection strategies and its importance in clinical prediction modelling. *Family Medicine and Community Health*, 8(1).

Christiansen, C. (1996). Hormone replacement therapy and osteoporosis. *Maturitas*, 23:S71–S76.

Cirigliano, M. (2007). Bioidentical hormone therapy: a review of the evidence. *Journal of Women's Health*, 16(5):600–631.

Climént-Palmer, M. and Spiegelhalter, D. (2019). Hormone replacement therapy and the risk of breast cancer: How much should women worry about it?

Cokkinides, V., Albano, J., Samuels, A., Ward, M., and Thum, J. (2005). American cancer society: Cancer facts and figures. *Atlanta: American Cancer Society.*

Collaborators, M. W. S. et al. (2005). Endometrial cancer and hormone-replacement therapy in the Million Women Study. *The Lancet*, 365(9470):1543–1551.

Collett, D. (2015). *Modelling Survival Data in Medical Research.* CRC press.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202.

Dehbi, H.-M., Royston, P., and Hackshaw, A. (2017). Life expectancy difference and life expectancy ratio: two measures of treatment effects in randomised trials with non-proportional hazards. *BMJ*, 357.

Diabetes UK (2010). Key statistics on diabetes. https://www.diabetes.org.uk/resources-s3/2017-11/diabetes_in_the_uk_2010.pdf. Accessed: 2019-05-21.

Diabetes UK (2015). Facts and stats. Diabetes UK, London.

Diabetes UK (2019). facts figures. https://www.diabetes.org.uk/professionals/position-statements-reports/statistics. Accessed: 2020-03-17.

Diabetes UK (2019). How to Calculate BMI. https://www.diabetes.co.uk/bmi/how-to-calculate-bmi.html. Accessed: 2019-05-21.

Drever, F. and Whitehead, M. (1995). Mortality in regions and local authority districts in the 1990s: exploring the relationship with deprivation. *Population Trends*, (82):19–26.

EBM DataLab (2014). Prescribing Data: BNF Codes. https://ebmdatalab.net/prescribing-data-bnf-codes/. Accessed: 2018-07-15.

Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association*, 72(359):557–565.

Emily et al., V. (2003). Breast cancer and hormone-replacement therapy in the Million Women Study. *The Lancet*, 362(9382):419–427.

Ettinger, B., Friedman, G. D., Bush, T., and Quesenberry Jr, C. P. (1996). Reduced mortality associated with long-term postmenopausal estrogen therapy. *Obstetrics & Gynecology*, 87(1):6–12.

FDA (July, 2018). Menopause Hormone Therapy. https://www.fda.gov/ForConsumers/ByAudience/ForWomen/ucm118627.htm. Accessed: 2018-03-03.

Folsom, A. R., Anderson, J. P., and Ross, J. A. (2004). Estrogen replacement therapy and ovarian cancer. *Epidemiology*, 15(1):100–104.

Folsom, A. R., Mink, P. J., Sellers, T. A., Hong, C.-P., Zheng, W., and Potter, J. D. (1995). Hormonal replacement therapy and morbidity and mortality in a prospective study of postmenopausal women. *American Journal of Public Health*, 85(8_Pt_1):1128–1132.

Fowkes, F. G. R., Rudan, D., Rudan, I., Aboyans, V., Denenberg, J. O., McDermott, M. M., Norman, P. E., Sampson, U. K., Williams, L. J., Mensah, G. A., et al. (2013). Comparison of global estimates of prevalence and risk factors for peripheral artery disease in 2000 and 2010: a systematic review and analysis. *The Lancet*, 382(9901):1329–1340.

Freeman, E. and Sherif, K. (2007). Prevalence of hot flushes and night sweats around the world: a systematic review. *Climacteric*, 10(3):197–214.

Galaal, K., Al Moundhri, M., Bryant, A., Lopes, A. D., and Lawrie, T. A. (2014). Adjuvant chemotherapy for advanced endometrial cancer. *The Cochrane Library*.

Gast, G.-C. M., Pop, V. J., Samsioe, G. N., Grobbee, D. E., Nilsson, P. M., Keyzer, J. J., Wijnands-van Gent, C. J., and van der Schouw, Y. T. (2011a). Hormone therapy and coronary heart disease risk by vasomotor menopausal symptoms. *Maturitas*, 70(4):373–378.

Gast, G.-C. M., Pop, V. J., Samsioe, G. N., Grobbee, D. E., Nilsson, P. M., Keyzer, J. J., Wijnands-van Gent, C. J., and van der Schouw, Y. T. (2011b). Vasomotor menopausal symptoms are associated with increased risk of coronary heart disease. *Menopause*, 18(2):146–151.

Gauthier, A., Kanis, J. A., Jiang, Y., Martin, M., Compston, J. E., Borgström, F., Cooper, C., and McCloskey, E. V. (2011). Epidemiological burden of postmenopausal osteoporosis in the uk from 2010 to 2021: estimations from a disease model. *Archives of Osteoporosis*, 6(1-2):179–188.

Gillam, S. J., Siriwardena, A. N., and Steel, N. (2012). Pay-for-performance in the United Kingdom: impact of the quality and outcomes framework: a systematic review. *The Annals of Family Medicine*, 10(5):461–468.

GP Notebook (2016). Menopause. https://www.gpnotebook.co.uk/simplepage.cfm?ID=1288699910. Accessed: 2018-02-12.

Grady, D., Applegate, W., Bush, T., Furberg, C., Riggs, B., Hulley, S. B., Group, H. R., et al. (1998). Heart and estrogen/progestin replacement study (HERS): design, methods, and baseline characteristics. *Controlled Clinical Trials*, 19(4):314–335.

Grady, D., Gebretsadik, T., Kerlikowske, K., Ernster, V., and Petitti, D. (1995). 95039031 hormone replacement therapy and endometrial cancer risk: A meta-analysis. *Maturitas*, 22(3):265.

Grady, D., Rubin, S. M., Petitti, D. B., Fox, C. S., Black, D., Ettinger, B., Ernster, V. L., and Cummings, S. R. (1992). Hormone therapy to prevent disease and prolong life in postmenopausal women. *Annals of Internal Medicine*, 117(12):1016–1037.

Grambsch, P. M. and Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81(3):515–526.

Greenland, S. and Finkle, W. D. (1995). A critical look at methods for handling missing covariates in epidemiologic regression analyses. *American Journal of Eidemiology*, 142(12):1255–1264.

Greenland, S. and Morgenstern, H. (1990). Matching and efficiency in cohort studies. *American Journal of Epidemiology*, 131(1):151–159.

Greenwood, M. (1926). A report on the natural duration of cancer. *Reports on Public Health and Medical Subjects*, (33):1–26.

Grodstein, F., Colditz, G. A., and Stampfer, M. J. (1994). Postmenopausal hormone use and cholecystectomy in a large prospective study. *Obstetrics & Gynecology*, 83(1):5–11.

Grodstein, F., Newcomb, P. A., and Stampfer, M. J. (1999). Postmenopausal hormone therapy and the risk of colorectal cancer: a review and meta-analysis. *The American Journal of Medicine*, 106(5):574–582.

Grodstein, F. and Stampfer, M. (1995). The epidemiology of coronary heart disease and estrogen replacement in postmenopausal women. *Progress in Cardiovascular Diseases*, 38(3):199–210.

Grodstein, F., Stampfer, M. J., Colditz, G. A., Willett, W. C., Manson, J. E., Joffe, M., Rosner, B., Fuchs, C., Hankinson, S. E., Hunter, D. J., et al. (1997). Postmenopausal hormone therapy and mortality. *New England Journal of Medicine*, 336(25):1769–1776.

Grodstein, F., Stampfer, M. J., Manson, J. E., Colditz, G. A., Willett, W. C., Rosner, B., Speizer, F. E., and Hennekens, C. H. (1996). Postmenopausal estrogen and progestin use and the risk of cardiovascular disease. *New England Journal of Medicine*, 335(7):453–461.

Groenwold, R. H., White, I. R., Donders, A. R. T., Carpenter, J. R., Altman, D. G., and Moons, K. G. (2012). Missing covariate data in clinical research: when and when not to use the missing-indicator method for analysis. *CMAJ*, 184(11):1265–1269.

Gudmundsdottir, H., Høieggen, A., Stenehjem, A., Waldum, B., and Os, I. (2012). Hypertension in women: latest findings and clinical implications. *Therapeutic Advances in Chronic Disease*, 3(3):137–146.

Hall, G. C. (2009). Validation of death and suicide recording on the thin uk primary care database. *Pharmacoepidemiology and Drug Safety*, 18(2):120–131.

Hampton, T. (2009). Lung cancer mortality higher in women who used combination hormone therapy. *Jama*, 302(6):615–616.

Harrell Jr, F. E. (2015). *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer.

Harrell Jr, F. E., Lee, K. L., and Mark, D. B. (1996). Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, 15:361–387.

Harris, R., Whittemore, A. S., Itnyre, J., and Group, C. O. C. (1992). Characteristics relating to ovarian cancer risk: collaborative analysis of 12 US case-control studies: Iii. epithelial tumors of low malignant potential in white women. *American Journal of Epidemiology*, 136(10):1204–1211.

Hastie, T., Tibshirani, R., and Friedman, J. (2015). *The elements of statistical learning: data mining, inference, and prediction*. Springer Series in Statistics.

Health Science Library (2020). Data Resources in the Health Sciences. https://guides.lib.uw.edu/hsl/data/findclin. Accessed: 2020-12-26.

Healthline (2018). Physical Exam. https://www.healthline.com/health/menopause/tests-diagnosis. Accessed: 2018-03-06.

Henderson, V. W. (2014). Alzheimer's disease: review of hormone therapy trials and implications for treatment and prevention after menopause. *The Journal of Steroid Biochemistry and Molecular Biology*, 142:99–106.

Henderson, V. W., Paganini-Hill, A., Emanuel, C. K., Dunn, M. E., and Buckwalter, J. G. (1994). Estrogen replacement therapy in older women: comparisons between alzheimer's disease cases and nondemented control subjects. *Archives of Neurology*, 51(9):896–900.

Hippisley-Cox, J., Coupland, C., Vinogradova, Y., Robson, J., and Brindle, P. (2008). Performance of the qrisk cardiovascular risk prediction algorithm in an independent uk sample of patients from general practice: a validation study. *Heart*, 94(1):34–39.

Hodis, H. N. and Mack, W. J. (2014). Hormone replacement therapy and the association with coronary heart disease and overall mortality: clinical application of the timing hypothesis. *The Journal of Steroid Biochemistry and Molecular Biology*, 142:68–75.

Hodis, H. N., Mack, W. J., Azen, S. P., Lobo, R. A., Shoupe, D., Mahrer, P. R., Faxon, D. P., Cashin-Hemphill, L., Sanmarco, M. E., French, W. J., et al. (2003). Hormone therapy and the progression of coronary-artery atherosclerosis in postmenopausal women. *New England Journal of Medicine*, 349(6):535–545.

Hosmer, D. W., Lemeshow, S., and May, S. (2011). *Applied survival analysis: Regression modeling of time-to-event data*. John Wiley & Sons.

Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). *Applied logistic regression*, volume 398. John Wiley & Sons.

Hulley, S., Grady, D., Bush, T., Furberg, C., Herrington, D., Riggs, B., Vittinghoff, E., Heart, Group, E. R. S. H. R., et al. (1998). Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. *Jama*, 280(7):605–613.

Hunt, K., Vessey, M., and McPHERSON, K. (1990). Mortality in a cohort of long-term users of hormone replacement therapy: an updated analysis. *BJOG: An International Journal of Obstetrics & Gynaecology*, 97(12):1080–1086.

Hunt, K., Vessey, M., MCPHERSON, K., and Coleman, M. (1987). Long-term surveillance of mortality and cancer incidence in women receiving hormone replacemen therapy. *BJOG: An International Journal of Obstetrics & Gynaecology*, 94(7):620–635.

ILCUK (2010). Osteoporosis in the UK. https://ilcuk.org.uk/osteoporosis-in-the-uk-at-breaking-point/. Accessed: 2021-03-06.

Ilyas Bakbergenuly (2021). Calculation of life expectancy. https://rdrr.io/github/ilyasstatistics/mylongevity/. Accessed: 2021-12-03.

Imtiaz, B., Tuppurainen, M., Rikkonen, T., Kivipelto, M., Soininen, H., Kröger, H., and Tolppanen, A.-M. (2017). Postmenopausal hormone therapy and alzheimer disease a prospective cohort study. *Neurology*, 88(11):1062–1068.

International Osteoporosis Foundation (2017). Facts and Statistics. https://www.iofbonehealth.org/facts-statistics. Accessed: 2019-03-04.

IQVIA Medical Research Data (2017). Data collection arrangements. https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/the-health-improvement-network-thin-database/. Accessed: 2020-07-28.

Jakobsen, J. C., Gluud, C., Wetterslev, J., and Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials–a practical guide with flowcharts. *BMC Medical Research Methodology*, 17(1):1–10.

Jang, Y.-C., Huang, H.-L., and Leung, C. Y. (2019). Association of hormone replacement therapy with mortality in colorectal cancer survivor: a systematic review and meta-analysis. *BMC cancer*, 19(1):1–10.

Joffres, M., Falaschetti, E., Gillespie, C., Robitaille, C., Loustalot, F., Poulter, N., McAlister, F. A., Johansen, H., Baclic, O., and Campbell, N. (2013). Hypertension prevalence, awareness, treatment and control in national surveys from england, the usa and canada, and correlation with stroke and ischaemic heart disease mortality: a cross-sectional study. *BMJ open*, 3(8):e003423.

Kalbfleisch, J. and Prentice, R. (2002). *The statistical analysis of failure time data*. John Wiley & Sons.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481.

Kawas, C., Resnick, S., Morrison, A., Brookmeyer, R., Corrada, M., Zonderman, A., Bacal, C., Lingle, D. D., and Metter, E. (1997). A prospective study of estrogen replacement therapy and the risk of developing alzheimer's disease the baltimore longitudinal study of aging. *Neurology*, 48(6):1517–1521.

Kawecka-Jaszcz, K., Czarnecka, D., Olszanecka, A., Rajzer, M., and Jankowski, P. (2002). The effect of hormone replacement therapy on arterial blood pressure and vascular compliance in postmenopausal women with arterial hypertension. *Journal of Human Hypertension*, 16(7):509–516.

Keele, L. (2010). Proportionally difficult: Testing for nonproportional hazards in Cox models. *Political Analysis*, 18(2):189–205.

Klein, J. P. and Moeschberger, M. L. (2003). *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media.

Klein, J. P. and Moeschberger, M. L. (2006). *Survival analysis: Techniques for censored and truncated data*. Springer Science & Business Media.

Kleinbaum and Klein (2012). *Survival Analysis:A Self-Learning Text*. Springer.

Kontopantelis, E., Stevens, R. J., Helms, P. J., and Edwards (2018). Spatial distribution of clinical computer systems in primary care in england in 2016 and implications for primary care electronic medical record databases: a cross-sectional population study. *BMJ open*, 8(2):e020738.

Kulinskaya, E., Gitsels, L. A., Bakbergenuly, I., and Wright, N. R. (2020). Calculation of changes in life expectancy based on proportional hazards model of an intervention. *Insurance: Mathematics and Economics*, 93:27–35.

Lacey Jr, J. V., Mink, P. J., Lubin, J. H., Sherman, M. E., Troisi, R., Hartge, P., Schatzkin, A., and Schairer, C. (2002). Menopausal hormone replacement therapy and risk of ovarian cancer. *Jama*, 288(3):334–341.

LeBlanc, E. S., Janowsky, J., Chan, B. K., and Nelson, H. D. (2001). Hormone replacement therapy and cognition: systematic review and meta-analysis. *Jama*, 285(11):1489–1499.

Liu, W.-X., Shi, M., Su, H., Wang, Y., He, X., Xu, L.-M., Yuan, Z.-Y., Zhang, L.-L., Wu, G., Qu, B.-L., et al. (2019). Effect of age as a continuous variable on survival outcomes and treatment selection in patients with extranodal nasal-type nk/t-cell lymphoma from the china lymphoma collaborative group (clcg). *Aging (Albany NY)*, 11(19):8463.

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.

Macmillan Cancer Support (2019). Breast Cancer. https://your.macmillan.org.uk/cancer-information-and-support/breast-cancer/risk-factors-for-breast-cancer. Accessed: 2019-09-10.

Maguire, A., Blak, B. T., and Thompson, M. (2009). The importance of defining periods of complete mortality reporting for research using automated data from primary care. *Pharmacoepidemiology and Drug Safety*, 18(1):76–83.

Malek, A. M., Vladutiu, C. J., Meyer, M. L., Cushman, M., Newman, R., Lisabeth, L. D., Kleindorfer, D., Lakkur, S., and Howard, V. J. (2019). The association of age at menopause and all-cause and cause-specific mortality by race, postmenopausal hormone use, and smoking status. *Preventive Medicine Reports*, 15:100955.

Manson, J. E., Aragaki, A. K., Rossouw, J. E., Anderson, G. L., Prentice, R. L., LaCroix, A. Z., Chlebowski, R. T., Howard, B. V., Thomson, C. A., Margolis, K. L., et al. (2017). Menopausal hormone therapy and long-term all-cause and cause-specific mortality: the Womens Health Initiative randomized trials. *Jama*, 318(10):927–938.

Manson, J. E., Chlebowski, R. T., Stefanick, M. L., Aragaki, A. K., Rossouw, J. E., Prentice, R. L., Anderson, G., Howard, B. V., Thomson, C. A., LaCroix, A. Z., et al. (2013). Menopausal hormone therapy and health outcomes during the intervention and extended poststopping phases of the Womens Health Initiative randomized trials. *Jama*, 310(13):1353–1368.

Manson, J. E. and Martin, K. A. (2001). Postmenopausal hormone-replacement therapy. *New England Journal of Medicine*, 345(1):34–40.

Mantel, N. and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22(4):719–748.

Margolis, K., Bonds, D., Rodabough, R., Tinker, L., Phillips, L., Allen, C., Bassford, T., Burke, G., Torrens, J., Howard, B., et al. (2004). Effect of oestrogen plus progestin on the incidence of diabetes in postmenopausal women: results from the Womens Health Initiative hormone trial. *Diabetologia*, 47(7):1175–1187.

Marshall, A., Altman, D. G., Royston, P., and Holder, R. L. (2010). Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study. *BMC Medical Research Methodology*, 10(1):1–16.

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.

Marubini, E. and Valsecchi, M. G. (2004). *Analysing survival data from clinical trials and observational studies*, volume 15. John Wiley & Sons.

Mayo Clinic (2018). Menopause: Diagnosis and Treatment. https://www.mayoclinic.org/diseases-conditions/menopause/diagnosis-treatment/drc-20353401. Accessed: 2018-02-12.

McGuire, S. (2016). World cancer report 2014. geneva, switzerland: World health organization, international agency for research on cancer, who press, 2015. *Advances in Nutrition: An International Review Journal*, 7(2):418–419.

Menopause support UK (2021). Menopause support. https://menopausesupport.co.uk/?page_id=60. Accessed: 2021-03-28.

Montgomery, D. C., Peck, E. A., and Vining, G. G. (2015). *Introduction to linear regression analysis*. John Wiley & Sons.

Mulnard, R. A., Cotman, C. W., Kawas, C., van Dyck, C. H., Sano, M., Doody, R., Koss, E., Pfeiffer, E., Jin, S., Gamst, A., et al. (2000). Estrogen replacement therapy for treatment of mild to moderate alzheimer disease: a randomized controlled trial. *Jama*, 283(8):1007–1015.

Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, 78(3):691–692.

National Institute for Health and Care Excellence (2019). Menopause: diagnosis and management. https://www.nice.org.uk/guidance/ng23. Accessed: 2020-03-06.

NHS (2010). The information centre for health and social care. https://files.digital.nhs.uk/publicationimport/pub00xxx/pub00206/obes-phys-acti-diet-eng-2010-rep.pdf. Accessed: 2019-10-21.

NHS Booklet (2017). BNF Classification and Pseudo Classification. https://www.nhsbsa.nhs.uk/sites/default/files/2017-04/BNF_Classification_Booklet-2017_0.pdf. Accessed: 2018-10-02.

NHS Choices (July, 2015). Lung cancer in women. https://www.nhs.uk/Livewell/Lungcancer/Pages/Womenandlungcancer.aspx. Accessed: 2018-01-12.

NHS Choices (July, 2016). HRT hormones. https://www.nhs.uk/conditions/hormone-replacement-therapy-hrt/types/. Accessed: 2018-01-12.

NHS Digital (2016). Health Survey for England 2015 Adult overweight and obesity . https://files.digital.nhs.uk/publicationimport/pub22xxx/pub22610/hse2015-adult-obe.pdf. Accessed: 2019-10-21.

NHS Digital (2017). Practice level prescribing . https://digital.nhs.uk/data-and-information/areas-of-interest/prescribing/practice-level-prescribing-in-england-a-summary/practice-level-prescribing-glossary-of-terms#practice-code-and-names. Accessed: 2018-10-02.

NHS Digital (2018a). Read Codes. https://digital.nhs.uk/services/terminology-and-classifications/read-codes. Accessed: 2018-10-02.

NHS Digital (2018b). Statistics on Smoking - England, 2006. https://digital.nhs.uk/data-and-information/publications/statistical/statistics-on-smoking/statistics-on-smoking-england-2006. Accessed: 2019-10-21.

NICE (2019a). Hypertension in adults: diagnosis and management. https://www.nice.org.uk/guidance/ng136/chapter/Context. Accessed: 2020-02-12.

NICE (2019b). Hypertension in adults: diagnosis and management. https://www.nice.org.uk/guidance/ng136/chapter/Recommendations#diagnosing-hypertension. Accessed: 2020-10-21.

NICE (2021). Long-term benefits and risks of hormone replacement therapy. https://www.nice.org.uk/consultations/672/10/long-term-benefits-and-risks-of-hormone-replacement-therapy. Accessed: 2021-09-10.

NICE (August, 2012). Osteoporosis: assessing the risk of fragility fracture. https://www.nice.org.uk/guidance/cg146/evidence/osteoporosis-fragility-fracture-full-guideline-186818365. Accessed: 2018-03-03.

NICE Obesity Management (2018). Obesity: identification, assessment and management. https://www.nice.org.uk/guidance/cg189/chapter/1-recommendations#generic-principles-of-care. Accessed: 2018-10-21.

Nielsen, G. G., Gill, R. D., Andersen, P. K., and Sørensen, T. I. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics*, 19:25–43.

Njabulo Ncube (2021). Actuarial Translation [Longevity and Morbidity Risks]. https://njabulo-ncube.shinyapps.io/actuarialtranslation/. Accessed: 2021-12-28.

Norman, P. and Darlington-Pollock, F. (2017). The changing geography of deprivation in great britain: Exploiting small area census data, 1971 to 2011.

Office for National Statistics (2018). Past and projected period and cohort life tables, 2018-based, UK: 1981 to 2068. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/bulletins/pastandprojecteddatafromtheperiodandcohortlifetables/1981to2068. Accessed: 2021-10-11.

on the Evaluation of Carcinogenic Risks to Humans, I. W. G., Organization, W. H., and for Research on Cancer, I. A. (2007). Combined estrogen-progestogen contraceptives and combined estrogen-progestogen menopausal therapy. volume 91. World Health Organization.

ONS UK (2019). Leading causes of death, UK: 2001 to 2018. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/causesofdeath/articles/leadingcausesofdeathuk/2001to2018. Accessed: 2020-09-10.

O'Quigley, J., Xu, R., and Stare, J. (2005). Explained randomness in proportional hazards models. *Statistics in Medicine*, 24(3):479–489.

Paganini-Hill, A. and Henderson, V. W. (1994). Estrogen deficiency and risk of alzheimer's disease in women. *American Journal of Epidemiology*, 140(3):256–261.

Paganini-Hill, A. and Henderson, V. W. (1996). Estrogen replacement therapy and risk of alzheimer disease. *Archives of Internal Medicine*, 156(19):2213–2217.

Panay, N., Hamoda, H., Arya, R., and Savvas, M. (2013). The 2013 British menopause society & Womens Health Concern recommendations on hormone replacement therapy. *Menopause International*, 19(2):59–68.

Pedersen, A. B., Mikkelsen, E. M., Cronin-Fenton, D., Kristensen, N. R., Pham, T. M., Pedersen, L., and Petersen, I. (2017). Missing data and multiple imputation in clinical epidemiological research. *Clinical Epidemiology*, 9:157.

Peng, M., Chen, G., Kaplan, G. G., Lix, L. M., Drummond, N., Lucyk, K., Garies, S., Lowerison, M., Weibe, S., and Quan, H. (2016). Methods of defining hypertension in electronic medical records: validation against national survey data. *Journal of Public Health*, 38(3):e392–e399.

Pentti, K., Honkanen, R., Tuppurainen, M. T., Sandini, L., Kroger, H., and Saarikoski, S. (2006). Hormone replacement therapy and mortality in 52-to 70-year-old women: the kuopio osteoporosis risk factor and prevention study. *European Journal of Endocrinology*, 154(1):101–107.

Persson, I. (2002). *Essays on the assumption of proportional hazards in Cox regression*. PhD thesis, Acta Universitatis Upsaliensis.

Persson, I., Weiderpass, E., Bergkvist, L., Bergström, R., and Schairer, C. (1999). Risks of breast and endometrial cancer after estrogen and estrogen–progestin replacement. *Cancer Causes & Control*, 10(4):253–260.

Persson, I., Yuen, J., Bergkvist, L., and Schairer, C. (1996). Cancer incidence and mortality in women receiving estrogen and estrogen-progestin replacement therapylong-term follow-up of a swedish cohort. *International Journal of Cancer*, 67(3):327–332.

Prinja, S., Gupta, N., and Verma, R. (2010). Censoring in clinical trials: review of survival analysis techniques. *Indian journal of community medicine: official publication of Indian Association of Preventive & Social Medicine*, 35(2):217.

Public Health England (2017). Health matters: combating high blood pressure. https://www.gov.uk/government/publications/health-matters-combating-high-blood-pressure/health-matters-combating-high-blood-pressure. Accessed: 2019-03-28.

Quartagno, M., Grund, S., and Carpenter, J. (2019). Jomo: a flexible package for two-level joint modelling multiple imputation. *R Journal*, 9(1).

Raboud, J. and Breslow, N. (1989). Efficiency gains from the addition of controls to matched sets in cohort studies. *Statistics in medicine*, 8(8):977–985.

Rasbash, J., Steele, F., Browne, W., Goldstein, H., Charlton, C., and Browne, W. (2012). A users guide to mlwin, v2. 26. centre for multilevel modelling. *University of Bristol*.

Rossouw et al., G. (2002). Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *Jama*, 288(3):321–333.

Royston, P. (2006). Explained variation for survival models. *The Stata Journal*, 6(1):83–96.

Royston, P. and Parmar, M. K. (2011). The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Statistics in Medicine*, 30(19):2409–2421.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3):581–592.

Rubin, D. B. (1987). Multiple imputation for survey nonresponse.

Rubin, D. B. (1996). Multiple imputation after 18^+ years. *Journal of the American statistical Association*, 91(434):473–489.

Salpeter, S., Walsh, J., Ormiston, T., Greyber, E., Buckley, N., and Salpeter, E. (2006). Meta-analysis: effect of hormone-replacement therapy on components of the metabolic syndrome in postmenopausal women. *Diabetes, Obesity and Metabolism*, 8(5):538–554.

Salpeter, S. R., Cheng, J., Thabane, L., Buckley, N. S., and Salpeter, E. E. (2009). Bayesian meta-analysis of hormone therapy and mortality in younger postmenopausal women. *The American Journal of Medicine*, 122(11):1016–1022.

Salpeter, S. R., Walsh, J. M., Greyber, E., Ormiston, T. M., and Salpeter, E. E. (2004). Mortality associated with hormone replacement therapy in younger and older women. *Journal of General Internal Medicine*, 19(7):791–804.

Savolainen-Peltonen, H., Rahkola-Soisalo, P., Hoti, F., Vattulainen, P., Gissler, M., Ylikorkala, O., and Mikkola, T. S. (2019). Use of postmenopausal hormone therapy and risk of alzheimers disease in finland: nationwide case-control study. *BMJ*, 364:l665.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. CRC press.

Schafer, J. L. and Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7(2):147–177.

Schierbeck, L. L., Rejnmark, L., Tofteng, C. L., Stilgren, L., Eiken, P., Mosekilde, L., Køber, L., and Jensen, J.-E. B. (2012). Effect of hormone replacement therapy on cardiovascular events in recently postmenopausal women: randomised trial. *BMJ*, 345:e6409.

Schuetz, F., Diel, I. J., Pueschel, M., von Holst, T., Solomayer, E. F., Lange, S., Sinn, P., Bastert, G., and Sohn, C. (2007). Reduced incidence of distant metastases and lower mortality in 1072 patients with breast cancer with a history of hormone replacement therapy. *American Journal of Obstetrics & Gynecology*, 196(4):342–e1.

Seminara, N., Abuabara, K., Shin, D., Langan, S., Kimmel, S., Margolis, D., Troxel, A., and Gelfand, J. (2011). Validity of The Health Improvement Network (THIN) for the study of psoriasis. *British Journal of Dermatology*, 164(3):602–609.

Serumaga, B., Ross-Degnan, D., Avery, A. J., Elliott, R. A., Majumdar, S. R., Zhang, F., and Soumerai, S. B. (2011). Effect of pay for performance on the management and outcomes of hypertension in the United Kingdom: interrupted time series study. *BMJ*, 342:d108.

Shumaker, S. A., Legault, C., Kuller, L., Rapp, S. R., Thal, L., Lane, D. S., Fillit, H., Stefanick, M. L., Hendrix, S. L., Lewis, C. E., et al. (2004). Conjugated equine estrogens and incidence of probable dementia and mild cognitive impairment in postmenopausal women: Women's Health Initiative Memory Study. *Jama*, 291(24):2947–2958.

Shumaker, S. A., Legault, C., Rapp, S. R., Thal, L., Wallace, R. B., Ockene, J. K., Hendrix, S. L., Jones III, B. N., Assaf, A. R., Jackson, R. D., et al. (2003). Estrogen plus progestin and the incidence of dementia and mild cognitive impairment in postmenopausal women: the Women's Health Initiative Memory Study: a randomized controlled trial. *Jama*, 289(20):2651–2662.

Smith, D. C., Prentice, R., Thompson, D. J., and Herrmann, W. L. (1975). Association of exogenous estrogen and endometrial carcinoma. *New England Journal of Medicine*, 293(23):1164–1167.

Springate, D. A., Kontopantelis, E., Ashcroft, D. M., Olier, I., Parisi, R., Chamapiwa, E., and Reeves, D. (2014). Clinicalcodes: an online clinical codes repository to improve the validity and reproducibility of research using electronic medical records. *PloS one*, 9(6).

Statista (2018). Obesity prevalence among individuals in England from 2000 to 2017, by gender. https://www.statista.com/statistics/334126/obesity-prevalence-by-gender-in-england-uk/. Accessed: 2019-01-24.

Stefanick, M. L., Anderson, G. L., Margolis, K. L., Hendrix, S. L., Rodabough, R. J., Paskett, E. D., Lane, D. S., Hubbell, F. A., Assaf, A. R., Sarto, G. E., et al. (2006). Effects of conjugated equine estrogens on breast cancer and mammography screening in postmenopausal women with hysterectomy. *Jama*, 295(14):1647–1657.

Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M., and Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*, 338:b2393.

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M. J., and Kattan, M. W. (2010). Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology (Cambridge, Mass.)*, 21(1):128.

Stram, D. O., Liu, Y. O., Henderson, K. D., Sullivan-Halley, J., Luo, J., Saxena, T., Reynolds, P., Chang, E. T., Neuhausen, S. L., Horn-Ross, P. L., et al. (2011). Age-specific effects of hormone therapy use on overall mortality and ischemic heart disease mortality among women in the california teachers study. *Menopause (New York, NY)*, 18(3):253.

Taggar, J. S., Coleman, T., Lewis, S., and Szatkowski, L. (2012). The impact of the quality and outcomes framework (qof) on the recording of smoking targets in primary care medical records: cross-sectional analyses from The Health Improvement Network (THIN) database. *BMC Public Health*, 12(1):329.

Tarone, R. E. and Ware, J. (1977). On distribution-free tests for equality of survival distributions. *Biometrika*, 64(1):156–160.

Therneau, T. and Atkinson, E. (2020). 1 the concordance statistic.

Therneau, T. M. and Atkinson, E. (2021). Concordance. https://cran.r-project.org/web/packages/survival/vignettes/concordance.pdf. Accessed: 2020-02-25.

Therneau, T. M., Grambsch, P. M., and Pankratz, V. S. (2003). Penalized survival models and frailty. *Journal of Computational and Graphical Statistics*, 12(1):156–175.

Therneau, T. M. and Lumley, T. (2019). Package survival.

Therneau, T. M. and Lumley, T. (2020). Package survival.

Todkar, M., Bhat, N., Fernandes, G., Pujari, H., Jain, I., and Gaikwad, M. (2013). Original study.

Townsend, P. (1987). Deprivation. *Journal of Social Policy*, 16(2):125–146.

Trinquart, L., Jacot, J., Conner, S. C., and Porcher, R. (2016). Comparison of treatment effects measured by the hazard ratio and by the ratio of restricted mean survival times in oncology randomized controlled trials. *Journal of Clinical Oncology*, 34(15):1813–1819.

UK Data Service (2019). 2011 UK Townsend Deprivation Scores. https://www.statistics.digitalresources.jisc.ac.uk/dataset/2011-uk-townsend-deprivation-scores. Accessed: 2019-05-08.

US National Library of Medicine (2019). Genetic Home Reference. https://ghr.nlm.nih.gov/condition/hypercholesterolemia#sourcesforpage. Accessed: 2019-03-07.

van Buuren, S. (2012). *Flexible imputation of missing data*. CRC press.

Verywellhealth (2018). Lung Cancer in Women. https://www.verywell.com/symptoms-of-lung-cancer-in-women-2249393. Accessed: 2018-03-12.

Vezyridis, P. and Timmons, S. (2016). Evolution of primary care databases in uk: a scientometric analysis of research output. *BMJ open*, 6(10):e012785.

Vinogradova, Y., Dening, T., Hippisley-Cox, J., Taylor, L., Moore, M., and Coupland, C. (2021). Use of menopausal hormone therapy and risk of dementia: nested case-control studies using QResearch and CPRD databases. *BMJ*, 374.

Volinsky, C. T. and Raftery, A. E. (2000). Bayesian information criterion for censored survival models. *Biometrics*, 56(1):256–262.

Wang, P., Liao, S., Liu, R., Liu, C., Chao, H., Lu, S., Yu, H., Wang, S., and Liu, H.-C. (2000). Effects of estrogen on cognition, mood, and cerebral blood flow in ad a controlled study. *Neurology*, 54(11):2061–2066.

WHI Organization. Welcome to the Women's Health Initiative. https://www.whi.org/SitePages/WHI%20Home.aspx. Accessed: 2018-03-10.

Whicher, C., ONeill, S., and Holt, R. G. (2020). Diabetes in the UK: 2019. *Diabetic Medicine*, 37(2):242–247.

WHO (2014). *Global status report on alcohol and health, 2014*. World Health Organisation.

Wijlaars, L. (2015). The Health Improvement Network (THIN) research team. https://www.ucl.ac.uk/pcph/research-groups-themes/thin-pub/database. Accessed: 2018-05-04.

Wilson, R. A. (1966). Md. feminine forever. *New York: M. Evans*.

Women's Health Concern (November, 2017). HRT: The history. https://www.womens-health-concern.org/help-and-advice/factsheets/hrt-the-history/. Accessed: 2018-01-12.

Women's Health Initiative (2018). Welcome to the Women's Health Initiative. https://www.whi.org/SitePages/WHI/Home.aspx. Accessed: 2018-02-06.

Xue, Y. and Schifano, E. D. (2017). Diagnostics for the Cox model. *Communications for Statistical Applications and Methods*, 24(6):583–604.

Yao, Y., Gu, X., Zhu, J., Yuan, D., and Song, Y. (2013). Hormone replacement therapy in females can decrease the risk of lung cancer: a meta-analysis. *PLoS One*, 8(8):e71236.

Zandi, P. P., Carlson, M. C., Plassman, B. L., Welsh-Bohmer, K. A., Mayer, L. S., Steffens, D. C., Breitner, J. C., Investigators, C. C. M. S., et al. (2002). Hormone replacement therapy and incidence of alzheimer disease in older women: the cache county study. *Jama*, 288(17):2123–2129.

Zárate, S., Stevnsner, T., and Gredilla, R. (2017). Role of estrogen and other sex hormones in brain aging: Neuroprotection and DNA repair. *Frontiers in Aging Neuroscience*, 9:430.

Zarnoch, S. J., Cordell, H. K., Betz, C. J., and Bergstrom, J. C. (2010). Multiple imputation: An application to income nonresponse in the national survey on recreation and the environment. *US Department of Agriculture Forest Service, Southern Research Station*, 49:1–15.

Ziel, H. K. and Finkle, W. D. (1975). Increased risk of endometrial carcinoma among users of conjugated estrogens. *New England Journal of Medicine*, 293(23):1167–1170.

Appendices

Appendix A

Literature review of HRT on mortality and morbidity

The major HRT studies that were conducted in the past, and their results are described in Chapter 2. In this appendix, the research summary, data, and findings of various published papers on HRT which were considered for literature review are tabulated. The designs and settings, study period, sample size, statistical methods used to estimate the results, and the outcomes are presented in the table in chronological order.

Table A.1: Past research publications on Hormone Replacement Therapy

Lead author (year published)	Design and setting	Study period	Sample size	Statistical method	Results
Hunt et al. (1987)	Women ages $45 - 54$ were recruited at 21 specialist menopause clinics around Britain to monitor mortality and cancer incidence. Past medical history, height, weight, blood pressure, marital status, occupation and smoking with a detailed HRT history was documented at the start of the study.	$1974 - 1983$	4,544	Logistic regression	Overall mortality was significantly lower than the national mortality rates (Relative Risk (RR) =0.58; 95% Confidence Interval (CI), 0.49-0.70) among HRT users apart from the ovarian cancer (RR=1.43; 95% Cl, $0.62 - 2.82$).
Colditz et al. (1987)	US women of ages $30 - 55$ years were recruited to determine the effects of HRT on CHD.	$1976 - 1982$	121,700	Cox proportional hazards model	Bilateral oophorectomy increases the risk of CHD.
Stampfer et al. (1991)	A cohort of female registered nurses in the US completed a set of mailed questionnaire about estrogen use and were followed up for 10 years with two years periodic resurvey Prior history of cardiovascular	1976-1986	48,470	Cox proportional hazards model	Current estrogen use was associated with a reduction in the incidence of CHD as well as mortality from CVD, but it was not associated with any change in the risk of stroke. The overall HR of CHD in women currently

Continued on next page

Table A.1 – *Continued from previous page*

Lead author (year published)	Design and setting	Study period	Sample size	Statistical method	Results
	disease and cancer were excluded at baseline.				taking estrogen was 0.56 and 95% CI, 0.40 − 0.80.
Paganini-Hill and Henderson (1994)	A prospective cohort study of residents of Leisure World Laguna Hills, a retirement community in Southern California was conducted by mailed questionnaires to evaluate the estrogen deficiency and risk of AD in women.	1981 − 1993	8,877 (3760 died)	Univariate and multivariate regression techniques	The risk of Alzheimer's disease and related dementia was less in estrogen users relative to nonusers (odds ratios (OR)= 0.69, 95% CI, 0.46 − 1.03). The risk decreased significantly with increasing estrogen dose and with increasing duration of estrogen use.
Folsom et al. (1995)	The Iowa Women's Health Study was performed by mailed questionnaires.	1986 − 1991	41,070	Cox proportional hazards regression model	The multivariate adjusted HR of current hormone users compared to non-users are: total mortality (HR = 0.78, 95% CI, 0.65 − 0.94), CHD (HR = 0.74, 95% CI, 0.48 − 1.12), endometrial cancer (HR = 4.3, 95% CI, 2.7 − 6.9), breast cancer (HR = 1.23, 95% CI, 0.99 − 1.55), colon cancer (HR = 0.72, 95% CI, 0.46 − 1.12), hip fracture (HR = 0.53, 95% CI,

Continued on next page

Lead author (year published)	Design and setting	Study period	Sample size	Statistical method	Results
					$0.31 - 0.91$).
Paganini-Hill and Henderson (1996)	A prospective cohort study of residents of Leisure World Laguna Hills, was conducted by mailed questionnaires to evaluate the effects of different estrogen preparation, varying doses of estrogen, and duration of estrogen replacement therapy on the risk of developing AD.	$1981 - 1995$	8,877 (3760 died)	Cox proportional hazards model	The risk of AD and related dementia was significantly reduced in estrogen users compared with non-users (HR = 0.65, 95% CI, $0.49 - 0.88$). The risk decreased significantly with both increasing doses ($P = 0.01$) and and increasing duration ($P = 0.01$) in women taking oral CEE.
Ettinger et al. (1996)	Cohort of women born between 1900 and 1915 was selected to compare all-cause and cause-specific mortality rates due to long-term use of ERT.	1969-1973	1,110	Cox-proportional hazards model	For death from any cause, the age-adjusted HR = 0.54 and associated 95% CI = $0.38 - 0.76$ in estrogen users compared to non-users.
Kawas et al. (1997)	A prospective study of ERT and the risk of developing Alzheimers disease in the Baltimore Longitudinal Study of	16 years of follow-up	472	Cox-proportional hazards model	The HR for AD in ERT users compared to nonusers was 0.46 (95% CI, $0.209 - 0.997$) after adjusting for education.

245

Lead author (year published)	Design and setting	Study period	Sample size	Statistical method	Results
	Aging categorized women into oral and transdermal ERT users.				
Grodstein et al. (1997)	The Nurses Health Study, where female registered USA nurses of ages $30 - 55$ completed a set of mailed questionnaires including information on menopause, CVD, and cancer.	$1976 - 1992$	121,700	Logistic regression	Current hormone users had a lower risk of death (OR = 0.63, 95% CI = $0.56 - 0.70$) than lifetime non-users. Survival benefit decreases with longer duration (> 10) of use.
Sourander et al. (1998)	Women born between 1923 and 1930 in Turku, Finland, were invited to participate in a mammography screening with questionnaires including the use of hormone therapy.	$1987 - 1995$	7,944 women contributed to 53,305 person-years follow-up	Cox proportional hazards model	Current ERT did not increase the risk of breast cancer (HR = 0.57, 95% Cl, $0.27 - 1.20$) compared to non-users.
Baldereschi et al. (1998)	The Italian Longitudinal Study of aging, a population based, multicentre survey examined the association	Not known	2,816	Cox Proportional hazards model	ERT is associated with a reduced prevalence of AD (HR = 0.28, 95% CI, $0.08 - 0.98$) after adjusting for age, education, age at menarche, age at menopause, smoking and

Lead author (year published)	Design and setting	Study period	Sample size	Statistical method	Results
	of ERT and AD in the postmenopausal women of age 65 – 84 years				alcohol habits, body weight at the age of 50 years, and the number of children.
Mulnard et al. (2000)	A randomized, double-blind, placebo non-userled clinical trial to examine the effects of ERT on treatment of mild to moderate AD.	1995 – 1999	120	Linear and logistic regression model	ERT did not show disease progression nor did it improve global, cognitive, or functional outcomes in women with mild to moderate AD.
Hedblad et al. (2002)	An urban cohort of peri/postmenopausal women of median age 55.4 in Sweden was followed up for nine years.	1983 – 1992	5,721	Kaplan-Meier	Women using HRT had a lower incidence of myocardial infraction (MI).
Lacey et al. (2002)	A cohort study of former participants in the Breast Cancer Detection Demonstration Project, a nationwide breast cancer screening programme at twenty-nine US clinical centres. The average age of women was 56.6 years.	1979 – 1998	44,241	Poisson regression model	329 women developed ovarian cancer during follow-up. Time dependent analysis adjusted for age, menopause type, oral contraceptive use, ever use of estrogen-only was significantly associated with ovarian cancer (RR = 1.6; 95% Cl, 1.2 – 2.0). Increasing duration of estrogen-only use was significantly associated with ovarian cancer.

Lead author (year published)	Design and setting	Study period	Sample size	Statistical method	Results
	Incidence of ovarian cancer was the main outcome measure.				
Hodis et al. (2003)	A double blinded, placebo non-userled trial on women with mean age 63.5 and at least one coronary-artery lesion. Participants were randomly assigned to either non-user group, estrogen-only (17β - estradiol), or estrogen-progestin group (17β - estradiol and medroxyprogesterone acetate).	1995 – 2000	226	ANOVA and chi-square test	In older postmenopausal women with established coronary-artherosclerosis, estrogen-only or combined HRT had no significant effect on the progression of atherosclerosis.
Beral et al. (2003)	A breast screening programme conducted by NHS invites all women in the UK of ages 50 – 64 years for routine screening together with postal questionnaire about use of HRT. The questionnaires	1996 – 2001	1,084,110 person-year follow-up	Cox proportional hazards regression model	HRT users at recruitment were more likely to develop breast cancer than never users (HR = 1.66; 95% CI, 1.58 – 1.75, $P < 0.0001$) and die from it (HR = 1.22; 95% CI, 1.00 – 1.48, $P = 0.05$). The risk was significantly increased for women receiving HRT preparation

248

Table A.1 – *Continued from previous page*

Lead author (year published)	Design and setting	Study period	Sample size	Statistical method	Results
	are asked to be returned before screening.				containing oestrogen-progestogen (HR = 2.00; 95% CI, 1.88 – 2.12, $P < 0.0001$).
Margolis et al. (2004)	A randomised, double-blinded WHI trial on postmenopausal women of ages 50 to 79 with intact uterus were recruited to examine the incidence of diabetes among HRT user.	1993 – 2001	15,641	Cox proportional hazards regression model	Combined therapy reduces the incidence of diabetes, possibly mediated by a decrease in insulin resistance unrelated to body size. The cumulative incidence of treated diabetes was 3.5% in the HRT groups and 4.2% in the placebo group (HR = 0.79, 95% Cl, 0.67 – 0.93, $P = 0.004$).
Paganini-Hill et al. (2006)	Cohort study of residents of a California retirement community was performed by a postal health survey including details on HRT use.	1981 – 2003	122,203 person-years of follow-up	Cox proportional hazards regression model	Older women undergoing estrogen therapy treatment had significantly increased longevity (HR = 0.91; 95% CI, 0.87 – 0.96). Hazards of death were lowest among long-term (≥ 15 years) users (HR = 0.83; 95% CI, 0.74 – 0.93 for 15 – 19 years) and (HR = 0.87; 95% CI, 0.80 – 0.94 for 20^+ years). Lower dose users (0.625 mg) had a slightly better survival rate than higher dose users (HR = 0.84; 95% CI, 0.78 – 0.91 vs

Table A.1 – *Continued from previous page*

Lead author (year published)	Design and setting	Study period	Sample size	Statistical method	Results
					HR = 0.91; 95% CI, 0.83 − 0.97).
Schuetz et al. (2007)	The Breast Unit of the Department of Gynecology and Obstetrics at the University Hospital of Heidelberg recruited patients with breast cancer aged 45-70 years.	1990 − 1999	1,072	Cox proportional hazards regression model, Kaplan-Meier	The use of HRT before the diagnosis of breast cancer results in more favourable primary tumours, with a lower incidence of recurrences and a better overall survival rate. 5-year survival was 92% (HR = 0.37; 95% Cl, 0.24 − 0.57).
MacGregoret al. (2007)	A meta-analysis of 103 studies considered the effects of menopause and hormone replacement therapy on headache and migraine.	1950 − 2007	1,436	Hierarchical Bayesian random-effects model	There is an increased longevity in younger postmenopausal women taking hormone therapy compared to those who are not taking the therapy.
Gast et al. (2011)	Data from Dutch and Swedish women of age 46 − 64 years who are free from CHD, stroke, venous thrombosis/ pulmonary embolism or cancer at baseline. Information on HRT, vasomotor symptoms (VMS)	1997 − 2007	91,310 pearson-years	Cox proportional hazards regression model	After multivariate adjustment, HRT use was not associated with the risk of CHD among women with or without intense VMS.

Continued on next page

Table A.1 – *Continued from previous page*

Lead author (year published)	Design and setting	Study period	Sample size	Statistical method	Results
	and potential confounders were collected by questionnaires.				
Hunter et al. (2011)	Cross-sectional study contributed by the UK collaborative trial of Ovarian Cancer Screening (UKCTOCS) cohort. Women without oophorectomy and aged 54 – 65 years completed a follow-up questionnaire.	2001 – 2008	10,418	Binary logistic regression	Women who had taken HT in the past and discontinued the treatment were more likely to have hot flashes and night sweats.
Schierbeck et al. (2012)	Randomised, open label trial of healthy Danish women aged between 45 – 58 and followed up to death, CVD and cancer.	1990 – 1993	2,016	Cox proportional hazards model	Early initiation and prolonged HRT significantly reduces the risk of the combined endpoint of mortality, myocardial infraction or heart failure and does not result in an increased risk of breast cancer or stroke.
Anderson et al. (2012)	Hysterectomised women of ages 50-79 years were selected from the WHI randomised, double-	1993 – 1998	10,739	Cox proportional regression model	The use of oestrogen was associated with lower incidence of invasive breast cancer (151 HRT users, 0.27% per year) compared with placebo (199 HRT users,

Continued on next page

Table A.1 – *Continued from previous page*

Lead author (year published)	Design and setting	Study period	Sample size	Statistical method	Results
	blinded, placebo-non-userled trial at 40 US clinical centres to examine the effect of oestrogen (0.625 mg) on invasive breast cancer incidence, tumour characteristics, and mortality.				0.35% per year); (HR 0.77, 95% CI $0.62 - 0.95$: $P = 0.02$) after a median of 5.9 years of follow-up.
Manson et al. (2013)	Two WHI hormone therapy trials (estrogen-only vs. placebo and combined HRT vs. placebo) were used to find out the health outcomes in Intervention and extended post-stopping Phases. Intervention lasted a median of 5.6 years in combined HRT trial and 7.2 years in estrogen-only trial with 13 years of cumulative follow-up until September 30, 2010.	$1993 - 1998$	27,347	Cox proportional hazards regression model	The number of CHD HRT users were 196 for combined vs 159 for placebo (HR = 1.18, 95% CI, $0.95 - 1.45$), and 206 vs 155 for invasive breast cancer (HR = 1.24, 95% CI, $1.01 - 1.53$). For CEE alone, younger women (aged $50 - 59$) had more favourable results for all-cause mortality and myocardial infraction.

Continued on next page

Table A.1 – *Continued from previous page*

Lead author (year published)	Design and setting	Study period	Sample size	Statistical method	Results
Boardman et al. (2015)	A meta-analysis of 19 RCTs to assess the effects of oral HRT for the prevention of cardiovascular disease in postmenopausal women to ascertain differential effects between use in primary or secondary prevention	Not known	40,410		Results show no evidence that hormone therapy provides any protective effects against death from any cause, death from cardiovascular disease, non-fatal heart attack or angina, either in healthy women or women with pre-existing heart disease.
Imtiaz et al. (2017)	A prospective study from the Kuopio Osteoporosis Risk Factor and Prevention study cohort in Finland to explore the association between postmenopausal HRT and AD. Self-administered questionnaires were sent to all women aged 47 – 56 years in every 5th year, starting from 1989. Register-based	1989 – 2009	8,195	Cox proportional hazards model	Postmenopausal estrogen use was not associated with the risk of AD in register-based or self-reported data (HR = 0.92, 95% CI, 0.68 - 1.2, and HR = 0.99, 95% CI, 0.75 - 1.3, respectively). Long-term self-reported postmenopausal HRT was associated with reduced AD risk (HR = 0.53, 95% CI, 0.31 - 0.91).

Continued on next page

Table A.1 – *Continued from previous page*

Lead author (year published)	Design and setting	Study period	Sample size	Statistical method	Results
	information on HRT prescriptions was available since 1995.				
Manson et al. (2017)	WHI trial on postmenopausal women of age 50 to 79 were used to examine the total and cause specific cumulative mortality, including during the intervention and extended postintervention follow-up up to December 31, 2014.	1993 – 1998	27,347 (7,489 died)	Cox proportional hazards model	During the cumulative 18 years follow-up, all-cause mortality was 27.1% in the HRT group and vs 27.6% in the placebo group (HR = 0.99, 95% CI, 0.94 – 1.03). The use of combined HRT for a median of 5.6 years or estrogen-only for a median of 7.2 years was not associated with the risks of all-cause, cardiovascular, or cancer mortality.

Appendix B

Survival models for hormone replacement therapy

Table B.1: The prevalence of hypertension using alternative definitions in this study population and the Peng et al. (2016)[1] study.

Definition	46−55 HRT user	46−55 non-user	56−65 HRT user	56−65 non-user	66+ HRT user	66+ non-user	Total HRT user	Total non-user	Peng et al. (2016)
Diagnosis Read code	7.62	7.75	15.02	16.41	27.07	28.50	10.58	11.79	14.49
High blood pressure	19.65	21.26	32.95	37.89	49.74	53.20	24.72	28.55	32.47
Antihypertensive drug	17.38	14.04	24.73	22.75	43.24	41.57	20.73	18.66	31.86
Diagnosis Read code or antihypertensive drug	18.72	15.60	27.19	25.86	46.32	44.83	22.45	20.78	32.17
Diagnosis Read code or high BP	24.93	27.29	42.87	50.00	64.16	68.39	31.69	37.00	32.81
High BP or antihypertensive drug	35.76	35.41	52.60	57.34	76.59	79.76	42.37	45.78	41.71
Diagnosis Read code or high BP or antihypertensive drug	36.68	36.62	54.25	59.88	78.32	81.62	43.51	46.84	41.75

[1]Peng et al. (2016) compared the prevalence of hypertension in THIN with the Health Survey England (HSE) using different definitions. In this table the prevalences calculated by Peng et al. (2016) from THIN are compared with the prevalence in the extracted study population.

Table B.2: Proportion (%) of missingness in BMI, smoking, Townsend deprivation score, and hypertension status for HRT users and non-users in full data.

| Covariates | HRT users | | | Non-users | Total |
| | Oestrogen-only | Combined | Total | | |
	(17606)	(87593)	(105199)	(224643)	(329842)
BMI	3356(19.1)	14016(16.0)	17372(16.5)	62629(28.0)	80001(24.3)
Smoking	2111(12.0)	7422(8.5)	9533(9.1)	37909(16.9)	47442(14.4)
Deprivation	1828(10.4)	9093(10.3)	10921(10.4)	26604(11.8)	37525(11.4)
Hypertension[1]	4629(26.3)	18463(21.1)	23092(22.0)	71201 (31.7)	94293(28.6)

[1]Missing values were generated in hypertension category due to missingness in systolic and diastolic blood pressure records.

Table B.3: Distribution of the HRT users and non-users by one to one match

Age group	HRT users (%)	Non-users (%)
46−55	27507 (81.30)	27507 (81.30)
56−65	6030 (17.82)	6030 (17.82)
66−75	271 (0.80)	271 (0.80)
75$^+$	24 (0.07)	24 (0.07)
Total	**33832 (100)**	**33832 (100)**

Table B.4: Proportion (%) of missingness in BMI, smoking, Townsend deprivation score, and hypertension status for HRT users and non-users in full data by age at study entry.

| Covariates | HRT users | | | Non-users | Total |
	Oestrogen-only (17606)	Combined (87593)	Total (105199)	(224643)	(329842)
BMI					
46-50	1195(6.8)	4555(5.2)	5750(5.4)	21480(9.6)	27217(8.3)
51-55	949(5.4)	4117(4.7)	5066(4.8)	17662(7.9)	22743(6.9)
56-60	738(4.2)	3416(3.9)	4154(3.9)	13008(5.8)	17144(5.2)
61-65	474(2.7)	1928(2.2)	2402(2.3)	10479(4.7)	12897(3.9)
Smoking					
46-50	810(4.6)	2715(3.1)	3525(3.4)	13723(6.1)	17248(5.2)
51-55	668(3.8)	2257(2.6)	2925(2.8)	10538(4.7)	13463(4.1)
56-60	439(2.5)	1576(1.8)	2015(2.0)	7413(3.3)	9428(2.8)
61-65	194(1.1)	874(1.0)	1068(1.0)	6240(2.8)	7308(2.2)
Deprivation					
46-50	651(3.7)	3085(3.5)	3736(3.6)	9485(4.0)	13221(4.0)
51-55	546(3.1)	2658(3.0)	3204(3.0)	7938(3.4)	11142(3.3)
56-60	456(2.6)	2376(2.7)	2832(2.7)	6635(2.9)	9467(2.8)
61-65	175(1.0)	974(1.1)	1149(1.1)	2546(1.1)	3695(1.1)
Hypertension[1]					
46-50	1531(8.7)	6300(7.2)	7831(7.4)	23809(10.6)	31640(9.6)
51-55	1285(7.3)	5689(6.5)	6974(6.6)	19542(8.7)	26516(8.1)
56-60	1038(5.9)	3761(4.3)	4799(4.5)	14824(6.4)	19623(6.0)
61-65	775(4.4)	2713(3.1)	3506(3.3)	13026(5.9)	16532(5.1)

[1]Missing values were generated in hypertension category due to missingness in systolic and diastolic blood pressure records.

Table B.5: Distribution[1] of the covariates with missing values in the complete data and imputed data.

|	Complete data[2] (108255)			Imputed data[3] (329842)		
	Oestrogen-only (6539)	Combined (36296)	Non-user (65420)	Oestrogen-only (17606)	Combined (87593)	Non-user (224643)
Hypertension						
No	3523(53.9)	22552(62.1)	37904(57.9)	10021(56.9)	55267(63.1)	134276(59.8)
Treated	1785(27.3)	8031(22.1)	15182(23.2)	4416(25.1)	18658(21.3)	49329(22.0)
Untreated	1231(18.8)	5713(15.7)	12334(18.9)	3169(18.0)	13668(15.6)	41038(18.3)
Deprivation						
Low	3646(55.8)	20001(55.1)	35326(54.0)	9648(54.8)	52057(54.5)	127761(52.3)
Medium	1327(20.3)	7388(20.4)	13457(20.6)	3662(20.8)	19574(20.5)	51088(20.9)
High	1566(23.9)	8907(24.5)	16637(25.4)	4078(24.4)	23811(24.9)	65596(26.8)
Smoking						
Non-smoker	4150(63.5)	21058(58.0)	41858(64.0)	10968(62.3)	55378(57.9)	152996(62.9)
Ex-smoker	1218(18.6)	6892(19.0)	10714(16.4)	3186(18.1)	17317(18.1)	38079(15.7)
Current smoker	1171(17.9)	8346(23.0)	12848(19.6)	3468(19.7)	22914(24.0)	52240(21.5)
Body mass index						
Healthy weight/overweight	4734(72.4)	28371(78.2)	46299(70.8)	14170 (74.5)	74740(78.8)	172613 (71.8)
Obese	1805(27.6)	7925(21.8)	19121(29.2)	4840 (25.5)	20067(21.2)	67826(28.2)

[1]Values are reported as number (%)
[2]Patients with the complete records only
[3]The reported prevalences is the mean of ten imputed datasets.

Table B.6: Selected baseline characteristics of the study population by age subgroup at HRT initiation

| Characteristics	46-50			51-55			56-60			61-65		
	Oestrogen-only	Combined	Non-users	Oestrogen-only	Combined	Non-users	Oestrogen-only	Combined	Non-users	Oestrogen-only	Combined	Non-users
Hypertension												
No	3399(67.5)	25831(69.4)	60144(69.1)	3531(58.7)	19144(62.5)	43364(59.8)	2060(50.6)	7257(54.6)	20212(49.7)	1079(43.3)	2993(46.5)	10530(43.2)
Treated	932(18.5)	6804(18.3)	14539(16.7)	1387(23.1)	6539(21.3)	15592(21.5)	1208(29.7)	3411(25.7)	11052(27.2)	855(34.3)	1998(31.1)	8055(33.1)
Untreated	704(14.0)	4584(12.3)	12425(14.3)	1093(18.2)	4971(16.2)	13530(18.7)	801(19.7)	2618(19.7)	9410(23.1)	557(22.4)	1443(22.4)	5790(23.8)
Uterine/ovarian status												
Intact	2231(44.3)	34623(93.0)	83158(95.5)	2233(37.2)	27112(88.5)	65141(89.9)	1412(34.7)	11165(84.0)	34928(85.9)	903(36.3)	5314(82.6)	20398(83.7)
Hysterectomy with oophorectomy	2579(51.2)	286(0.8)	870(1.0)	3451(57.4)	359(1.2)	2119(3.0)	2460(60.5)	254(2.0)	1974(4.9)	1455(58.4)	168(2.6)	1539(6.3)
Oophorectomy only	225(4.5)	2310(6.2)	3080(3.5)	327(5.4)	3183(10.4)	5226(7.2)	197(4.8)	1867(14.1)	3772(9.3)	133(5.3)	952(14.8)	2438(10.0)
PAD/PVD	317(6.3)	2270(6.1)	5139(5.9)	390 (6.7)	1992 (6.5)	4494(6.2)	317 (7.8)	1009(7.6)	2969(7.3)	241(9.7)	611(9.5)	2242(9.2)
Diabetes Type II	66(1.3)	361(1.0)	1158(1.3)	98(1.6)	451(1.5)	1604(2.2)	85(2.1)	260(2.0)	1300(3.2)	68(2.7)	161(2.5)	1027(4.2)
CHD	31(0.6)	223(0.6)	445(0.5)	91(1.5)	306(1.0)	762(1.1)	117(2.9)	275(2.1)	919(2.3)	97(3.9)	229(3.6)	1004(4.1)
Osteoporosis	65(1.3)	409(1.1)	784(0.9)	80(1.5)	429(1.4)	942(1.3)	98(2.9)	79(3.2)	358(2.7)	107(4.3)	302(4.7)	658 (2.9)
Hypercholesterolaemia	372(7.4)	2619(7.0)	4299(5.0)	464(7.7)	2258(7.4)	4468(6.2)	365(9.0)	1060(8.0)	3025(7.4)	217(8.7)	554(8.6)	1771(7.3)
Body mass index												
Healthy weight/overweight	3714(73.7)	29193(78.4)	63101(72.4)	4420(73.5)	24045(78.4)	51277(70.7)	3045(74.8)	10621(79.9)	28979(71.3)	1926(77.3)	5160(80.0)	17533(71.9)
Obese	1321(26.2)	8026(21.5)	24007(27.5)	1591(26.5)	6609(21.5)	21209(29.3)	1024(25.2)	2665(20.1)	11695(28.7)	565(22.7)	1274(19.8)	6842(28.1)
Smoking status												
Non-smoker	2953(58.6)	20360(54.7)	55149(63.3)	3736(62.1)	18185(59.3)	45164(62.3)	2621(64.4)	8121(61.1)	25405(62.5)	1657(66.5)	3986(61.9)	15538(63.7)
Ex-smoker	828(16.4)	6179(16.6)	12291(14.1)	1032(17.2)	5682(18.5)	11488(15.8)	788(19.4)	2639(19.8)	6862(16.9)	510(20.5)	1383(21.5)	4507(18.5)
Current smoker	1254(24.9)	10680(28.7)	19668(22.6)	1243(20.7)	6787(22.1)	15834(21.8)	660(16.2)	2526(19.0)	8407(20.7)	324(13.0)	1065(16.5)	4330(17.7)
Deprivation status												
Low	2766(55.0)	20129(54.1)	46550(53.4)	3331(55.4)	16901(55.1)	37981(52.4)	2202(54.1)	7334(55.2)	20809(51.2)	1373(55.1)	3490(54.2)	12275(50.4)
Medium	989(19.6)	7461(20.1)	17679(20.3)	1257(20.9)	6241(20.4)	15055(20.8)	839(20.6)	2717(20.5)	8462(20.8)	506(20.3)	1307(20.3)	5198(21.3)
High	1280(25.4)	9629(25.8)	22879(26.3)	1423(23.7)	7512(24.5)	19450(26.8)	1028(25.3)	3235(24.3)	11403(28.0)	612(24.6)	1637(25.4)	6902(28.3)

Table B.7: Results of the Grambsch and Therneau test in the Cox PH model. A significant P-value (< 0.05) is an indication of the violation of PH assumption.

Variable	Rho	Chi-square	P-value
HRT treatment			
No (Ref)			
Combined HRT	-0.002697	0.07441	0.78513
Estrogen-only	-0.000159	0.00075	0.97817
Age group at study entry			
46-50 (Ref)			
51-55	-0.031085	10.05726	0.00152
56-60	-0.027581	7.98915	0.00471
61-65	-0.040494	17.26878	0.00003
Birth cohort			
1921-1930 (Ref)			
1931-1940	-0.026251	7.13106	0.00757
1941-1950	-0.046838	22.71298	<0.00001
1951-1960	-0.042161	18.07075	0.00002
Hypertension			
No (Ref)			
Treated	-0.014381	2.17870	0.13993
Untreated	-0.000269	0.00075	0.97817
Uterine/ovarian status			
Intact (Ref)			
Hyeterectomy with oophorectomy	-0.008301	0.71952	0.3963
Oophorectomy only	-0.001018	0.01070	0.9175
Diabetes Type II			
No (Ref)			
Yes	-0.006143	0.37215	0.35913
Body mass index			
Healthyweight/overweight (Ref)			
Obese	-0.004197	0.110294	0.73980
Smoking status			
Non-smoker (Ref)			
Ex-smoker	-0.002385	0.285390	0.51398
Current-smoker	-0.005167	0.710294	0.83980
Deprivation status			
Low (Ref)			
Medium	-0.004281	0.188900	0.66383
High	-0.007716	0.615187	0.73980
Global	NA	103.6045	< 0.00001
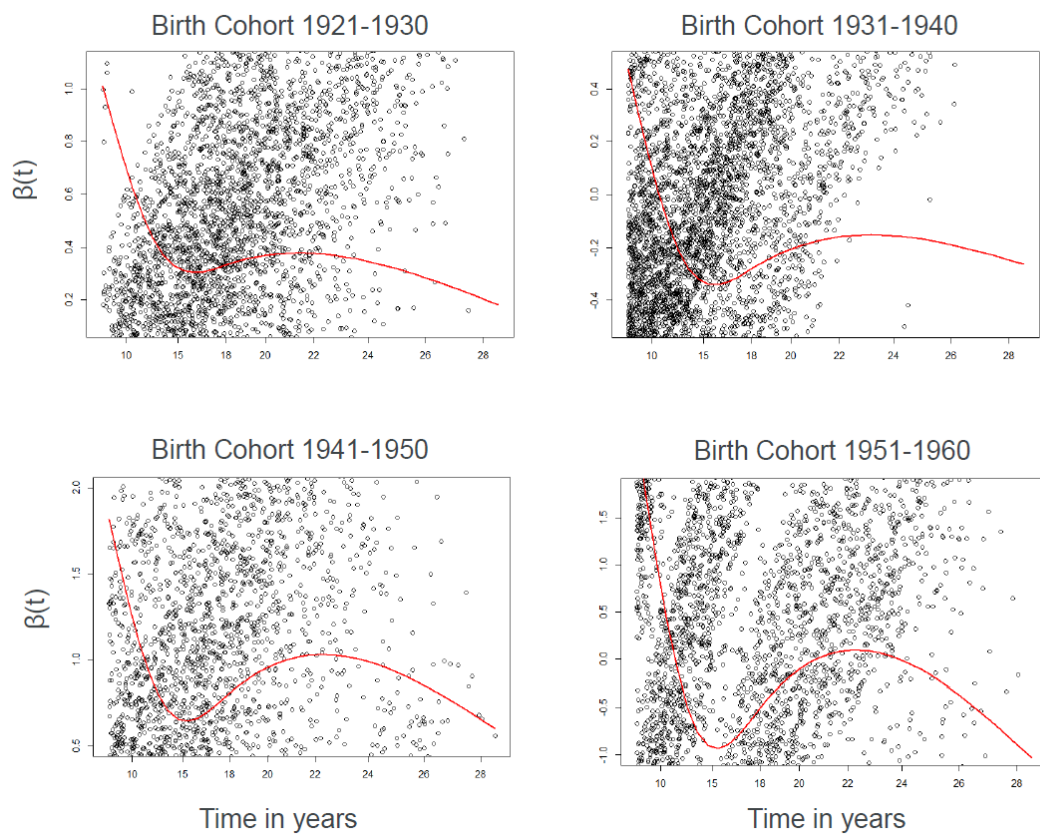
Figure B.1: Scaled Schoenfeld residual plots of the coefficients against event time for four birth cohorts. The smoothed red curve in each plot is an estimate of the regression coefficient in log hazard scale for the birth cohort over time. Non-linear curves (non-zero slopes) are the indication of the violation of proportional hazards assumptions.
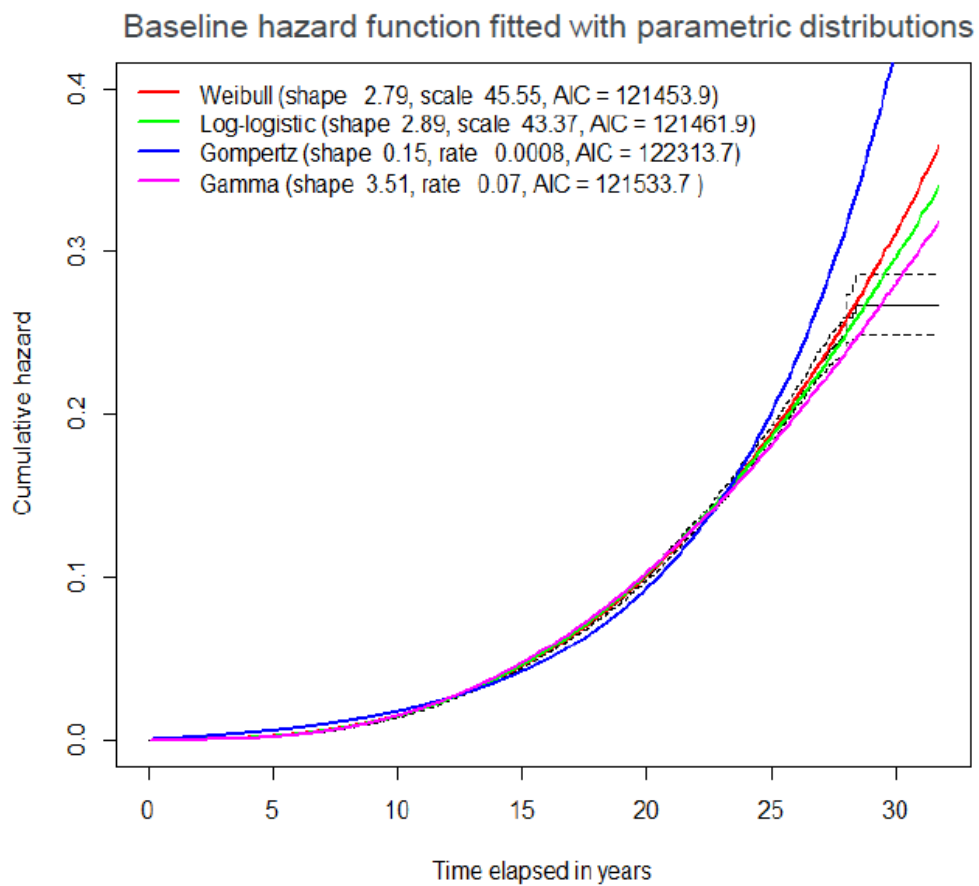
Figure B.2: The estimated baseline cumulative hazard function of the study population and cumulative hazards fitted using different parametric distributions. Weibull distribution fits well to the underlying baseline hazards.

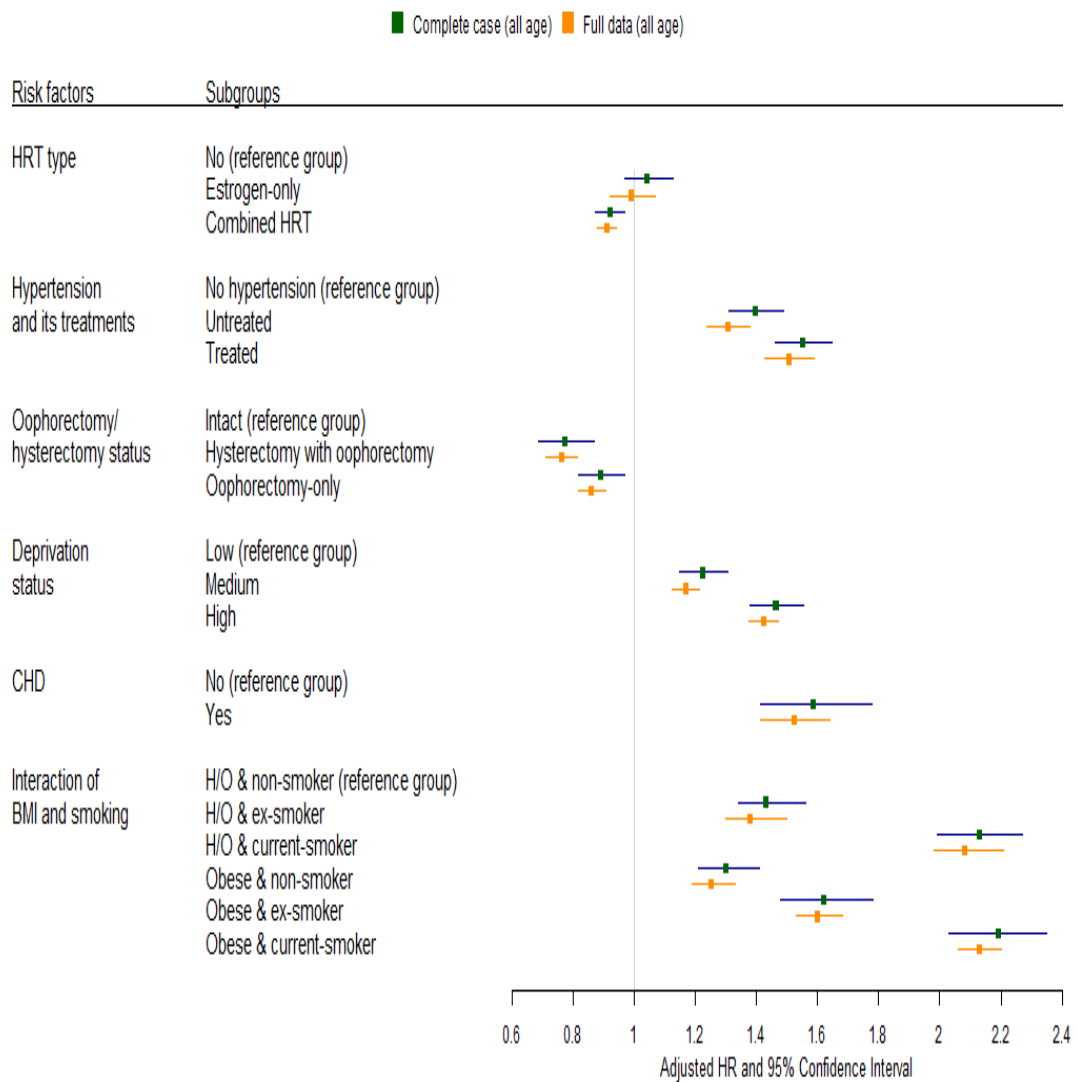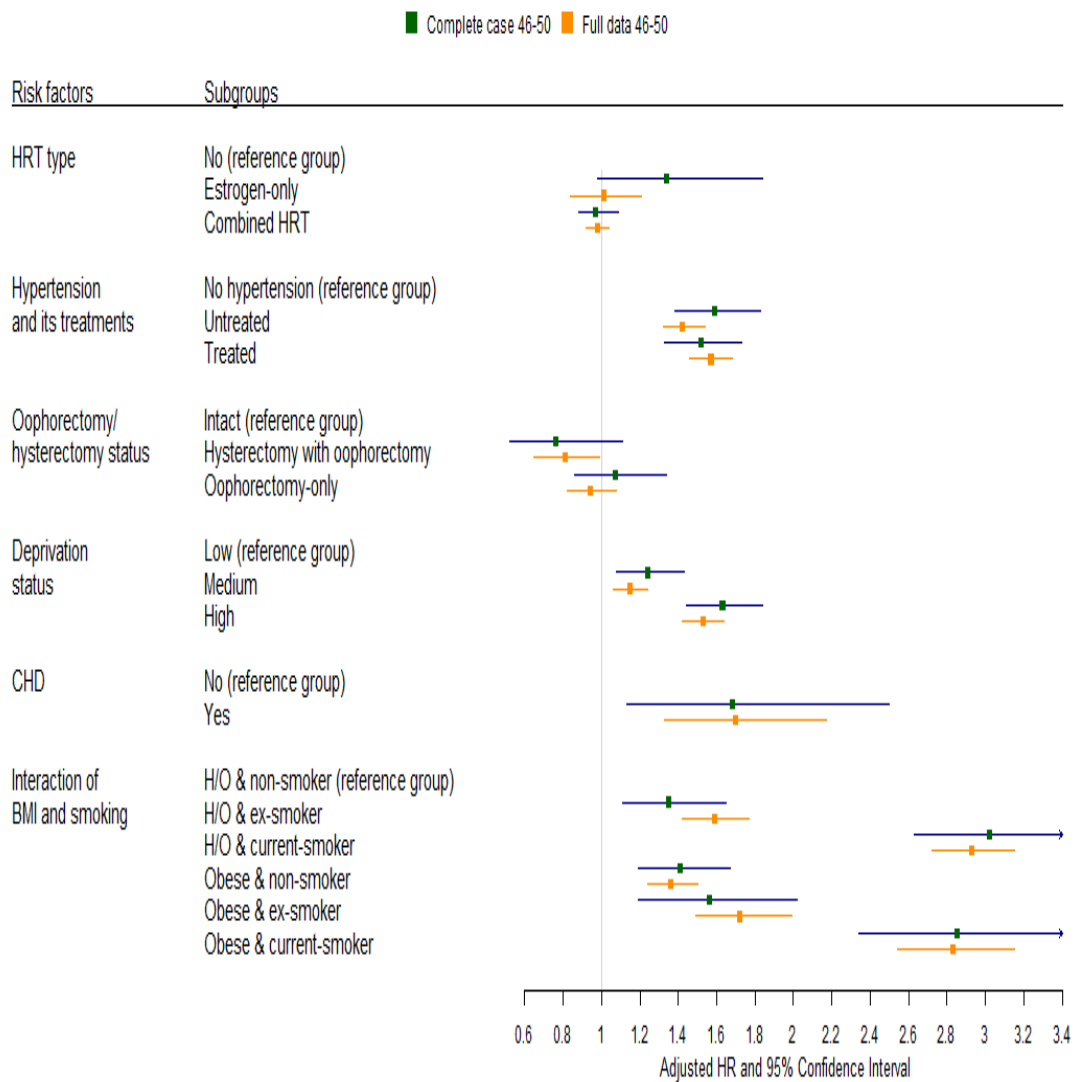| Age cohort	Deprivation status		Adjusted hazard ratio (95% CI)
Overall	Low (reference group)		
	Medium		1.17 (1.13-1.21)
	High		1.42 (1.37-1.47)
46-50	Low		
	Medium		1.15 (1.06-1.24)
	High		1.53 (1.42-1.63)
51-55	Low		
	Medium		1.12 (1.05-1.20)
	High		1.46 (1.38-1.56)
56-60	Low		
	Medium		1.21 (1.13-1.30)
	High		1.39 (1.30-1.48)
61-65	Low		
	Medium		1.13 (1.05-1.22)
	High		1.27 (1.18-1.36)

0.8 1 1.2 1.4 1.6 1.8 2
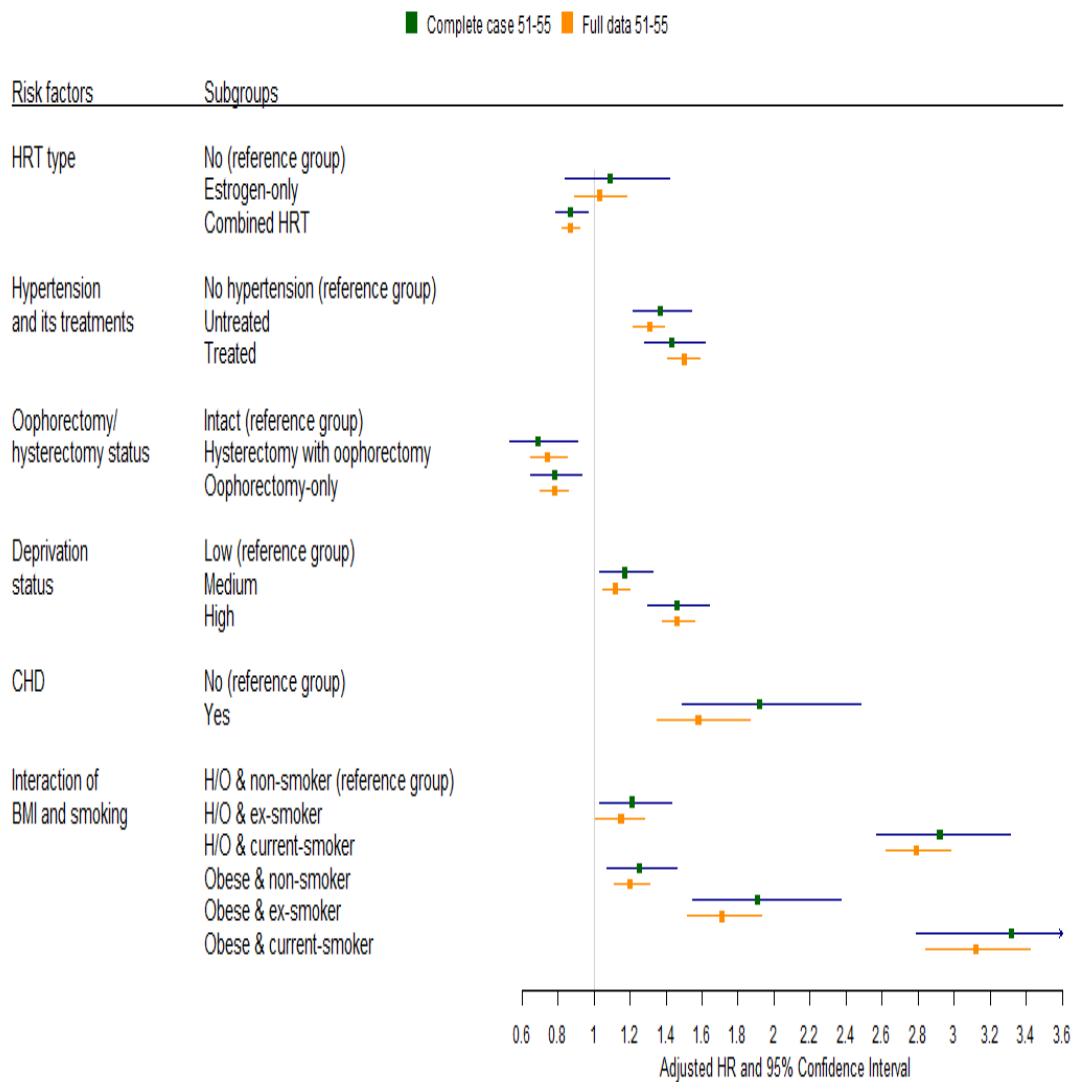Adjusted Hazard Ratio
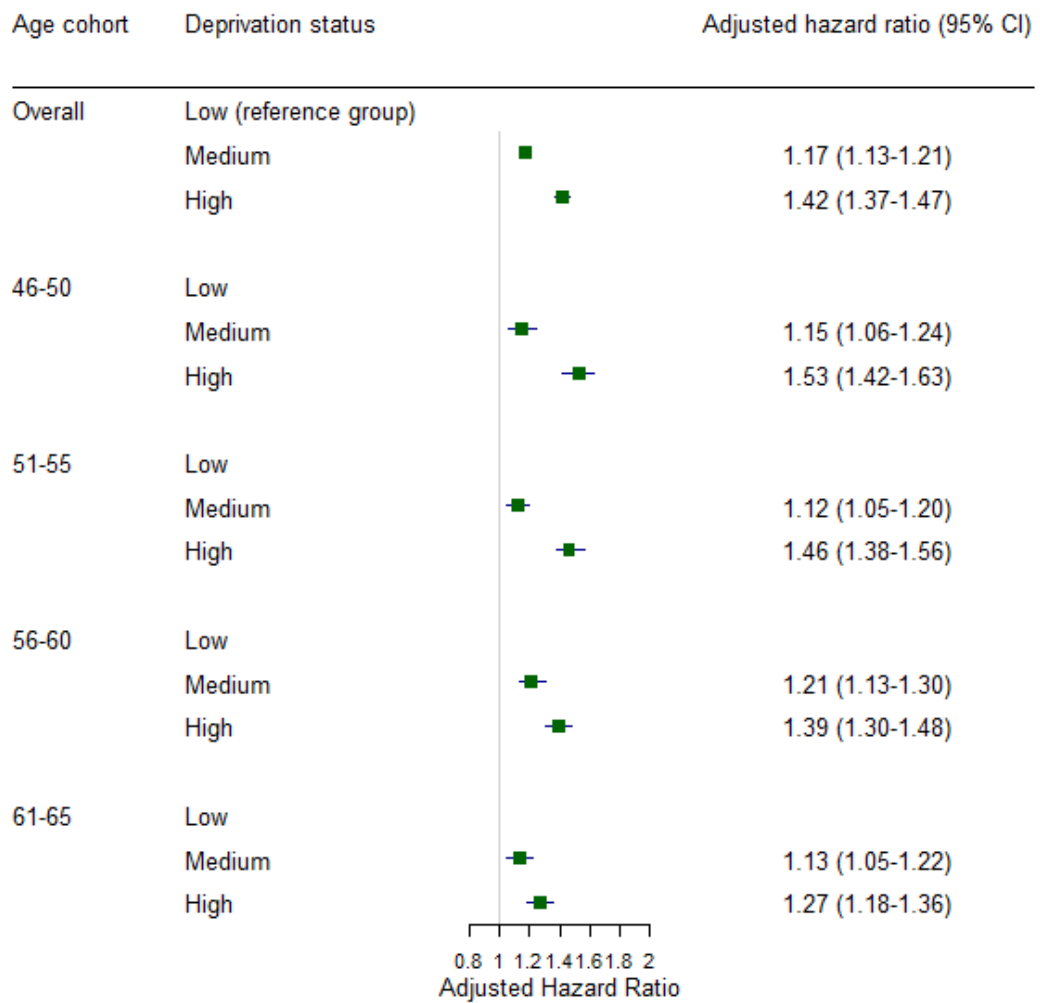
Figure B.3: The adjusted hazard ratios of all-cause mortality associated with deprivation status in full data (all ages) and four age subgroups at first HRT prescription. The hazards ratios (95% confidence interval) were adjusted for age at first HRT, birth cohort, type of HRT, type II diabetes, coronary heart disease (CHD), oophorectomy/hysterectomy status, body mass index, hypertension and its treatments, and smoking status. General practice was also included in the model as frailty.

Figure B.4: The adjusted hazard ratios of all-cause mortality associated with HRT type, hypertension and its treatments, oophorectomy/hysterectomy status, deprivation status, coronary heart disease, and interaction of BMI and smoking in complete and full HRT user model for all ages (46-65).

Figure B.5: The adjusted hazard ratios of all-cause mortality associated with hypertension and its treatments, oophorectomy/hysterectomy status, deprivation status, coronary heart disease, and interaction of BMI and smoking in complete and full HRT user model for age 46 to 50 at first treatment.
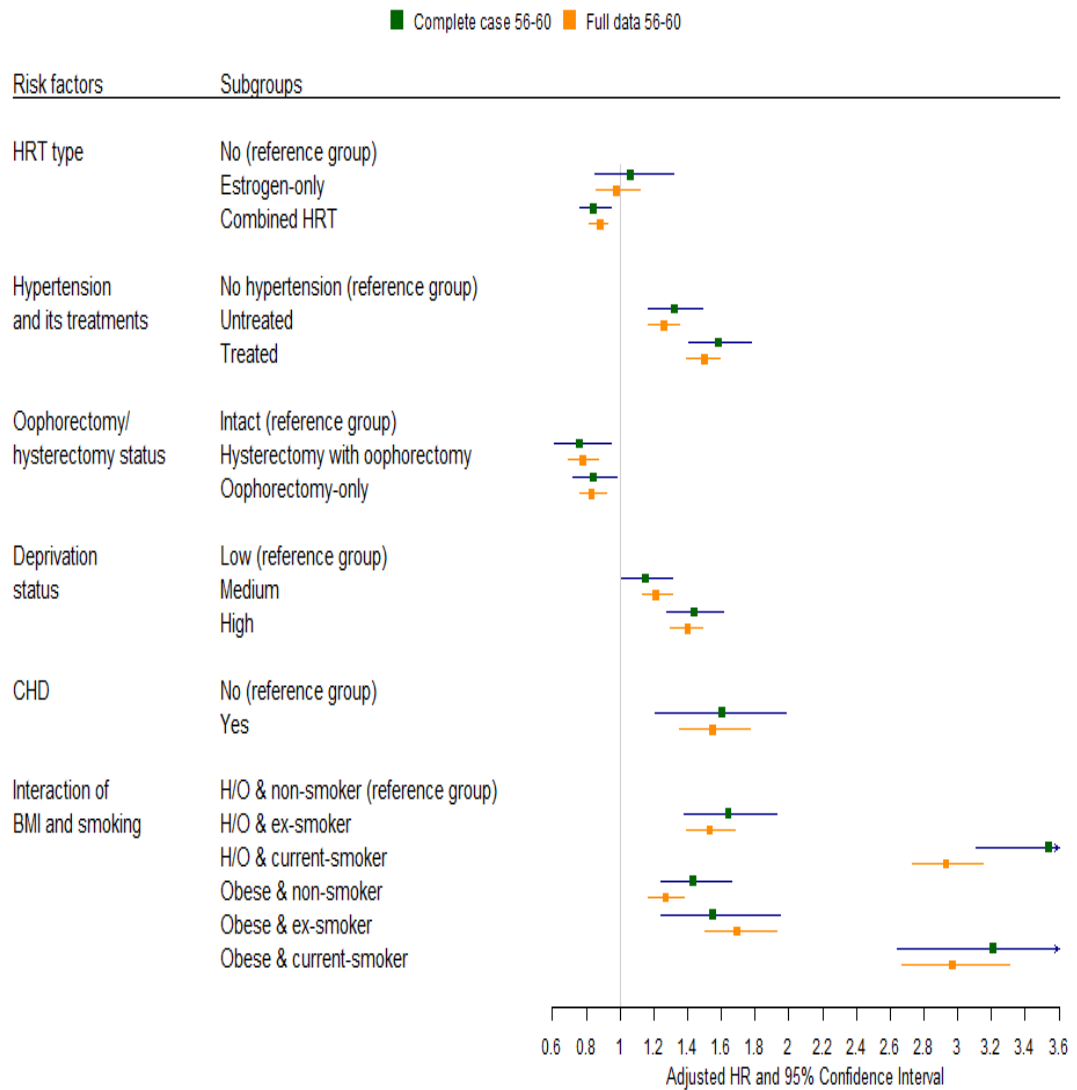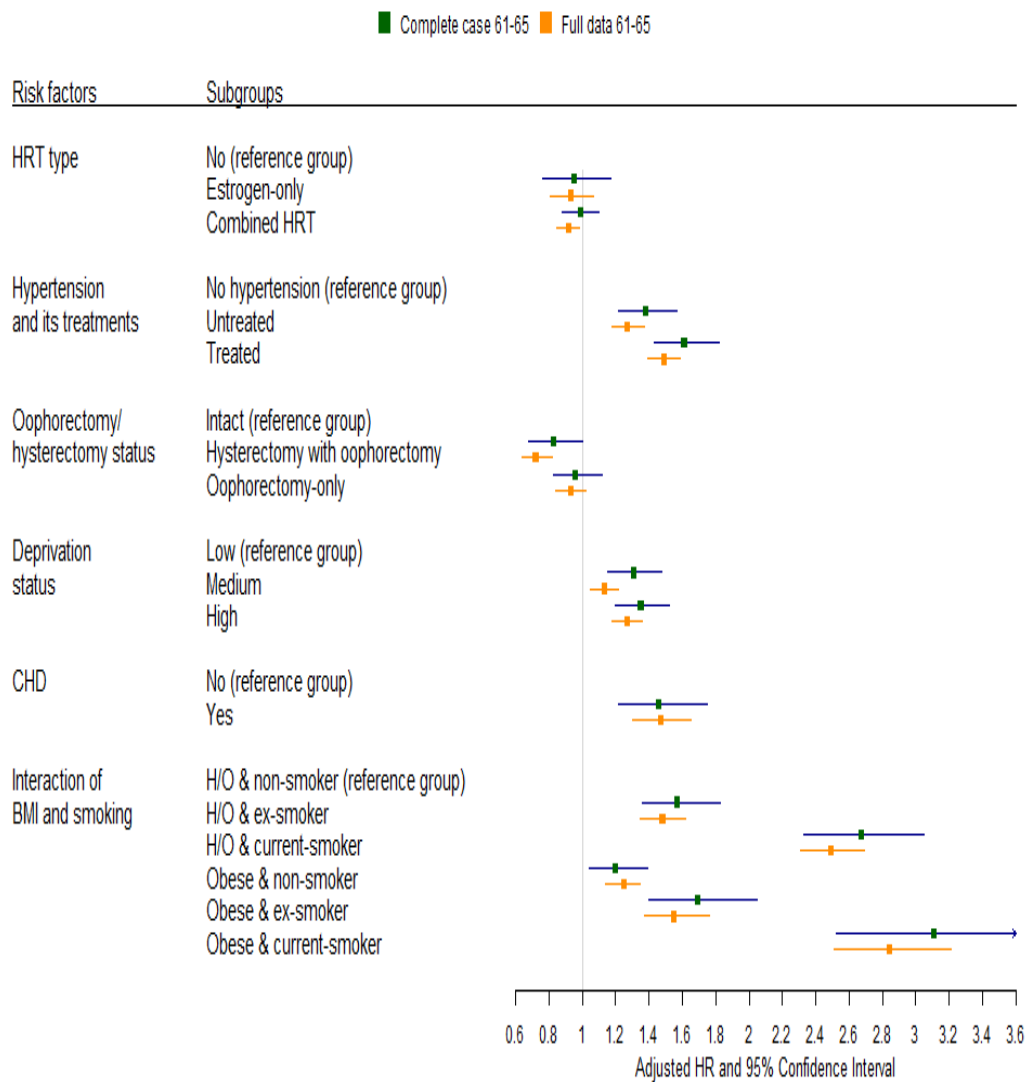
Figure B.6: The adjusted hazard ratios of all-cause mortality associated with hypertension and its treatments, oophorectomy/hysterectomy status, deprivation status, coronary heart disease, and interaction of BMI and smoking in complete and full HRT user model for age 51 to 55 at first treatment.

Complete case 56-60 ■ Full data 56-60

Risk factors	Subgroups
HRT type	No (reference group)
	Estrogen-only
	Combined HRT
Hypertension and its treatments	No hypertension (reference group)
	Untreated
	Treated
Oophorectomy/ hysterectomy status	Intact (reference group)
	Hysterectomy with oophorectomy
	Oophorectomy-only
Deprivation status	Low (reference group)
	Medium
	High
CHD	No (reference group)
	Yes
Interaction of BMI and smoking	H/O & non-smoker (reference group)
	H/O & ex-smoker
	H/O & current-smoker
	Obese & non-smoker
	Obese & ex-smoker
	Obese & current-smoker

0.6 0.8 1 1.2 1.4 1.6 1.8 2 2.2 2.4 2.6 2.8 3 3.2 3.4 3.6
Adjusted HR and 95% Confidence Interval

Figure B.7: The adjusted hazard ratios of all-cause mortality associated with hypertension and its treatments, oophorectomy/hysterectomy status, deprivation status, coronary heart disease, and interaction of BMI and smoking in complete and full HRT user model for age 56 to 60 at first treatment.

Figure B.8: The adjusted hazard ratios of all-cause mortality associated with hypertension and its treatments, oophorectomy/hysterectomy status, deprivation status, coronary heart disease, and interaction of BMI and smoking in complete and full HRT user model for age 61 to 65 at first treatment.

Table B.8: Model performance statistics for complete records and imputed data in survival models for all ages and age subgroups at first HRT treatment.

Age group[1]	Concordance (SE)		Logliklihood		AIC	
	Complete HRT user	Imputed data	Complete HRT user	Imputed data	Complete HRT user	Imputed data
All age	0.66 (0.004)	0.68 (0.002)	-37462.5	-122833.3	74989.0	245730.6
46-50	0.75 (0.009)	0.76 (0.005)	-9202.3	-30861.3	18448.6	61766.2
51-55	0.75 (0.008)	0.76 (0.004)	-10144.7	-36342.3	20333.5	72728.6
56-60	0.73 (0.008)	0.77 (0.004)	-9096.5	-29284.2	18236.9	58612.5
61-65	0.79 (0.008)	0.81 (0.005)	-8849.7	-25221.5	17739.4	50487.0

[1]The age groups included HRT users who received HRT treatment at that age and non-users matched on age and general practice. Patients with incomplete medical records had missing observations at least in one of the following covariates: BMI, smoking, Townsend deprivation score, and hypertension status.