Quantitative Approaches in Monitoring Population Quality Of Life

Maria Gheorghe

Quantitative Approaches in Monitoring Population Quality Of Life

Maria Gheorghe

Funding
Some of the studies performed in this thesis were financially supported by the Netherlands Organization for Health, Research and Development (Chapters 2 and 5)

ISBN: 978-94-6169-944-2

Cover design: Optima Grafische Communicatie, Rotterdam, The Netherlands

Layout and printed by: Optima Grafische Communicatie, Rotterdam, The Netherlands

Quantitative Approaches in Monitoring Population Quality Of Life

Kwantitatieve methoden voor het monitoren van kwaliteit van leven in de bevolking

Thesis

to obtain the degree of Doctor from
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. H.A.P. Pols

and in accordance with the decision of the Doctorate Board.
The public defense shall be held on
Thursday 15th of September 2016 at 11:30 hours

by

Maria Gheorghe

born in Sibiu, Romania

Erasmus University Rotterdam

Doctoral committee

Promotor:	Prof.dr. W.B.F. Brouwer
Other members:	Prof.dr.ing. P.H.C. Eilers
	Prof.dr. H.C. Boshuizen
	Prof.dr. E.K.A. van Doorslaer
Copromotor:	Dr. P.H.M. van Baal

Contents

Pentru Mama

Chapter 1

General introduction

Background

During the past century, most countries in the world have witnessed a substantial increase in life expectancy which was mainly attributable to declines in old-age mortality rates (Eggleston and Fuchs 2012). Consequently, the proportion of elderly has increased and is expected to increasese even further; a demographic phenomenon commonly referred as population aging. Worldwide, people aged 60 and older make up over 11 per cent of the global population, and is expected that by 2050, this number will rise to about 22 per cent (United Nations 2013). For facing these changes in the structure of their populations, world nations need to be equipped with the right set of economic policies. In that context, extensive societal, political and scientific debates have been associated with population ageing. Important examples include discussions on the affordability of growing healthcare expenditures (HCE) and those on raising the statutory retirement age in order to increase labour force participation of the elderly. The rationale for and exact consequences of policy decisions in these areas crucially depend on the extent to which the increases in life expectancy are accompanied by concomitant increases in life years spent in good health. Therefore, monitoring the level of population health and its changes over time and by population subgroups is a key component for determining whether these proposed policy changes are necessary, possible and will have the desired societal effects. The aim of this thesis is to investigate changes in population health and the way in which such changes may be used in economic evaluations of interventions aimed to improve health. In doing so, it is important to use a measure of population health that can capture the complex nature of human health.

In 1946, the World Health Organization (WHO) defined health as "the state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity" (WHO 1946). Although this indicates that health is a broad concept, traditionally health has been measured rather narrowly. For example, it has been measured by using all-cause and disease-specific mortality based indicators or health measures such as self-rated health or disability. However, as will be shown in what follows, these measures either ignore (e.g. mortality based indicators) or inadequately capture (e.g. self-rated health and disability measures) the impact of non-fatal diseases on health. Therefore, more recently, the concept of health-related quality of life (HRQoL) has been proposed in the context of monitoring population health (Dolan 2000). Estimating population health using a multidimensional measure of health such as HRQoL is important in at least two contexts. First, it is important to use a health measure that captures various health domains (e.g. physical, mental and social) in order to adequately monitor the level of population health and its changes over time and by population subgroups (Perenboom et al. 2004, Water et al. 1996, Picavet and Hoeymans 2002, Majer et al. 2013,

Bruggink et al. 2009). This is particularly relevant in aging societies for informing public policies associated with population aging. Second, estimating changes in population health using HRQoL is important as these estimates can be used to assess the benefits in economic evaluations of health care interventions (Drummond et al. 2005). Then, these estimates directly affect the results of economic evaluations and potentially influence the allocation of limited resources in the healthcare sector. The purpose of this thesis is to use HRQoL in exploring changes in population health in these two contexts, while investigating various methodological aspects related to modelling the HRQoL outcome.

Population Health Measures

Developing a composite measure of overall health status that would allow comparisons among populations, countries, regions and over time is challenging. To date, there is no single accepted indicator of population health; instead, various measures of health are currently in use. Until recently, all-cause, disease-specific mortality based measures were used as basic indicators of population health. Although these indicators allow various comparisons between population subgroups, in time, by region and across countries (Leon 2011, Wilmoth 2000), they provide insufficient information to fully describe the health of a population as they do not capture the impact of non-fatal diseases on health (Murray et al. 2000). In order to overcome that shortcoming, mainly two other indicators – i.e. self-rated health and disability – have been used.

Measurements of self-perceived or 'self-rated' health use a single question that asks individuals to rate their own health (commonly by choosing one of the five possible answers ranging from very poor to excellent). However, despite this indicator's popularity and its ease of use, it has been questioned whether this is a suitable measure for tracking population health. For example, a US chapter using four national surveys identified strong inconsistencies in self-ratings among surveys and in certain population subgroups (Salomon et al. 2009).

Disability measures usually consist of several questions targeted at measuring specific aspects of health status, especially aspects referring to physical health. Similar to self-rated health, disability measures health in one dimension by using a binary outcome (i.e. a person is either disabled/not-disabled). Nevertheless, such relatively insensitive, one-dimensional measures are unlikely to sufficiently capture the complex, broad aspects of health. Therefore, the concept of HRQoL, mainly developed by psychologists and health economists, may be used to overcome these issues.

HRQoL can be viewed as a multidimensional or as a one-dimensional concept. As a multidimensional concept, HRQoL encompasses at least three health domains, i.e. physical,

mental and social, each covered in one or more questions. People can indicate how they score on the different domains by answering these questions, in this way HRQoL measures are able to capture a wide variety of health states. As a one-dimensional concept, HRQoL attempts to capture each health state in just one value referred to as 'utility' which generally is anchored on the state 'dead' (with value zero) and the state perfect health (with value 1). Most imperfect health states receive a score between 0 and 1, although negative scores can be possible as well, reflecting health states valued as being worse than dead. The health state valuations are typically obtained from a sample of the general population using recognized valuation techniques such as time trade off (TTO, (Attema et al. 2013)) or standard gamble (SG, (Gafni 1994)). The fact that health states are valued using preference-based valuation techniques is perhaps one of the most notable advantages of HRQoL measures compared to both self-rated health and disability measures. Valuation functions have been developed for instruments such as: the EuroQoL EQ-5D (Dolan 1997), the Health Utility index HUI (http://fhs.mcmaster.ca/hug/) and the SF-6D (Brazier et al. 2002). These instruments are referred to as generic preference-based HRQoL measures. A further distinction can be made between 'generic' and 'disease-specific' HRQoL. While a generic measure aims at capturing all health aspects relevant to quality of life and allows comparing the health status of various populations and across different disease and healthcare areas, a disease-specific measure focuses on specific health domains and problems associated with a particular disease. Similar to generic questionnaires, disease-specific questionnaires are usually multidimensional, but unlike the former, the latter do not aim to cover all relevant aspects of human health. Compared to the number of generic questionnaires, the number of disease-specific ones is very high: for almost each common disease, a disease-specific questionnaire has been constructed. Through-out this thesis the terms health-related quality of life (HRQoL) and quality of life (QoL) will be used interchangebly.

Summary measures of population health have been developed as an extension of the basic metrics described above; these measures 'combine information on mortality and nonfatal health outcomes to represent the health of a particular population as a single numerical index' (Murray et al. 2000). A variety of summary population health measures are available that differ in the type of health indicator they use for capturing the impact of non-fatal diseases on health. Given that this thesis uses HRQoL as an indicator that measures the impact of non-fatal disease on health, quality-adjusted life expectancy (QALE) is used as a summary measure of population health. QALE is defined as an equivalent of years lived in full health (Murray et al. 2000).

Economic evaluations

Economic evaluations aim to inform decisions regarding resource allocation in the field of healthcare with the aim to improve human welfare. Typically, this is operationalised in such a way that economic evaluations assist in maximizing health benefits from available resources (Brouwer et al. 2008). Economic evaluation has been defined as the 'comparative analysis of alternative courses of action in terms of both their costs and consequences' (Drummond et al. 2005). Since they offer a systematic and transparent framework for deciding which interventions among alternatives to fund from a restricted budget, economic evaluations are reasonably well accepted in the decision making process within the systems of different countries such as the UK, The Netherlands, Canada and Australia. In health care, the most common form of economic evaluation and commonly proposed as the reference case in decision making processes, is cost-utility analysis (CUA). This form is required by important health technology assessment agencies such as the National Institute for Clinical Excellence (NICE, (Longworth et al. 2014)) in the UK and the Dutch Health Care Institute (Zinl, (Zorginstituut Nederland 2016)) in The Netherlands.

A cost-utility analysis evaluates at least two alternative interventions in terms of their incremental benefits and costs and summarizes the results in an incremental cost-effectiveness ratio (ICER). Hence, the ICER represents the additional costs per additional health unit induced by an intervention in comparison to a relevant comparator, such as usual care. In a cost-utility analysis health benefits are assessed in terms of quality-adjusted life years (QALYs), an index comprising both length of life and health-related quality of life (HRQoL). The basic idea underlying QALY calculation is simple; it assumes that one year lived in perfect health equals 1 QALY (1 year of life times 1 HRQoL utility) and that a year of life lived in a health state less then perfect is worth less than one. Note that, QALE, a second acronym for QALY, is used in the population health literature as a summary measure of population health status. As mentioned above, QALE represents life expectancy weighted for the quality of surviving years and so is actually measured in QALYs.

The result of an economic evaluation, the ICER, is summarized in incremental costs per QALYs gained, which enables outcomes (i.e. ICERs) to be compared across therapeutic areas and evaluations. Throughout this thesis the terms economic evaluation, cost-effectiveness analysis and cost-utility analysis will be used interchangeably (as is commonly done). Note that cost-effectivness analysis represents a generic form of cost-utility analysis in which effects are expressed in natural (clinically relevant) units such as number of successfully treated patients, life years gained.

In the next sections, the topics covered in this thesis will be further introduced. First, we address the issue of estimating changes in population health as measured by HRQoL. Next, we address the issues of the appropriate methods to be used when modelling the HRQoL outcome. Finally, we present the implications of both changes in population health and HRQoL methodological considerations for economic evaluations.

Topics Considered In This Thesis

Estimating changes population health as measured by HRQoL

In an aging society, it is important to establish whether the observed increases in life expectancy are accompanied by concomitant improvements in health. This is particularly relevant for the ongoing debates regarding extended labour-force participation by the elderly and regarding raising of statutory retirement ages. In those contexts, observing increases in peoples' health levels around the legal or practical retirement or withdrawal ages are especially important.

Many empirical studies have estimated changes in population health using various populations, health indicators and calendar periods (Perenboom et al. 2004, Water et al. 1996, Picavet and Hoeymans 2002, Majer et al. 2013, Bruggink et al. 2009). Such studies often analysed changes over time in the age and gender-specific health patterns using either self-rated health or disability based measures for monitoring population health. Potentially due to differences in health measures, calendar periods and included age groups, previous research reached diverse conclusions regarding the trends in population health, with some studies indicating an increase in the number of years lived in good health or free of disabilities (Bruggink et al. 2009) while other pointing towards the reverse situation (Salomon et al. 2012). A high number of empirical studies have also analysed time patterns stratified by age, gender and socioeconomic status (SES); many of which showed that health inequalities by SES have persisted and widened over time (Turrell and Mathers 2001, Martikainen et al. 2001, Mackenbach et al. 2003, Singh and Siahpush 2006, Mackenbach et al. 2008, Meara et al. 2008, van Kippersluis et al. 2010, Maki et al. 2013). Hence, less educated people do not only have a shorter life expectancy but also live more years in poor health (Maki et al. 2013, Kunst et al. 2005, Majer et al. 2011). This has been confirmed for many European countries including Belgium, Denmark, France and the Netherlands, as well as for New Zealand and the US (Van Oyen et al. 2011, Cambois et al. 2001, Bronnum-Hansen and Baadsgaard 2008, Bruggink 2009, Crimmins and Saito 2001, Davis et al. 1999). Notably, the vast majority of the above studies used health measures based either on self-rated health or disability. Using large survey data, in this thesis, we investigate changes in the health of the Dutch population as measured

by quality-adjusted life expectancy (QALE). These changes were explored over time and by population groups, i.e. by age, gender and education level.

Besides the above descriptive studies, empirical studies using econometric techniques investigated population aging in relation to healthcare use and health (Fryback et al. 2007, Getzen 1992, Zweifel et al. 1999). Some showed that with advancing age, population health deteriorates (Fryback et al. 2007) and health care use increases (Bos and von Weizsacker 1989, OECD 1988) suggesting that further increases in life expectancy would likely increase the average number of years lived in poor health and the aggregate healthcare use. However, others showed that such straightforward conclusions may be flawed. For example, it has been demonstrated that healthcare expenditures (HCE) are centred in the last phase of life (Seshamani 2004, Seshamani 2004, Zweifel et al. 2004, Werblow et al. 2007, Wanless 2004). This would suggest that age itself is not the main driver of the observed HCE patterns, but rather that proximity to death or time to death (TTD) is associated with high HCE. In other words, higher average health care costs at higher ages are mainly caused by the fact that more people die at higher ages and the period before dying is associated with high healthcare use and costs. This implies that an increase in life expectancy postpones the expensive last period of life, which suggests that aging of the population per se might have a more limited impact on HCE than generally believed. Although this mechanism has been demonstrated using large cost databases from insurance companies and hospital registries, less is known whether a similar mechanism may explain the relationship between population aging and health. If similar to healthcare demand, health losses are centred in the last phase of life then further increases in life expectancy would not necessarily translate into additional years spent in poor health. On the contrary, increases in life expectancy then could go hand in hand with improvements in health. Using survey data for the Dutch population and appropriate econometric methods for modelling the HRQoL outcome, the relationship between HRQoL, age and TTD has been investigated in this thesis. Furthermore, the obtained results have been translated to the economic evaluation context, in particular for estimating QALYs gained in economic evaluations of life prolonging interventions.

Statistical modelling of HRQoL values From a statistical point of view, the HRQoL outcome variable can be treated either as a collection of discrete variables or as a continuous bounded variable (i.e. the utility). Therefore, for modelling HRQoL data, two main categories of methods have been proposed in the literature, depending on whether the outcome variable of interest is the utility score or the probability that a respondent selects a particular level of functioning on each question of the HRQoL questionnaire. In this thesis, we will refer to the former as *utility score modelling* (Franks et al. 2003, Fryback

et al. 1997, O'Brien et al. 2003, Versteegh et al. 2012) and to the latter as *probability modelling* (Longworth et al. 2014, Le and Doctor 2011, Gray et al. 2006).

Utility score modelling is the most frequently used approach and includes methods such as OLS, Tobit like models (Austin 2002), censored least absolute deviation models (CLAD, (Sullivan and Ghushchyan 2006), hurdle models (Mullahy 1986), beta regressions (Basu and Manca 2012, Hunger et al. 2011, Hunger et al. 2012), and finite mixture models (FMM, (Hernandez Alava et al. 2012, Hernandez Alava et al. 2014, Coca Perraillon et al. 2015)). Nevertheless, HRQoL utility scores present certain non-standard features: such data typically has mass points at one of the boundaries (i.e. at one), is skewed, exhibits discontinuity and is heteroscedastic given that the variance will approach zero as the mean approaches either boundary point (Kieschnick and McCullough 2003). Due to these characteristics, although different methods have been proposed for modelling these data, there is no commonly preferred approach.

Probability modelling involves modelling the responses to the different HRQoL questions rather than their overall utility score and, subsequently apply the valuation functions to the estimated probabilities. The commonly used method in this class is the multinomial logit (MNL) model proposed by Gray and colleagues (Gray et al. 2006). More complex methods have also been proposed, e.g. Lee and Doctor used Bayesian Networks (BNs) for predicting the probability of each response level for all HRQoL domains obtained from a Bayesian updating process.

Although, as discussed above, a large variety of methods has been used for modelling HRQoL data, it remains unclear which method performs best and for which HRQoL instrument. The majority of these methods have been used for modelling EQ-5D HRQoL data, although some methods such as beta regressions have been shown to perform better for other HRQoL instruments such as the SF-6D. This thesis contributes to the literature referring to HRQoL methodologies in two ways. First, it highlights the use of beta regressions for modelling the SF-6D data. A beta regression model is used for estimating changes in population health using cross-sectional SF-6D data. Furthermore, mixed beta regressions have been previously proposed in the literature for modelling longitudinal SF-6D data (Hunger et al. 2012). In this thesis we extend this latter approach by developing a mixed beta regression model estimated using the Bayesian paradigm for modelling the longitudinal SF-6D utility outcome. Compared to maximum likelihood approaches, the Bayesian estimation procedure makes it possible to develop more complex models that can better address the idiosyncrasies of HRQoL data, i.e. heteroskedasticity and missing data.

Second, in this thesis a large number of the currently existing methodologies for modelling the EQ-5D data will be compared in an EQ-5D mapping exercise (Longworth et al. 2014). Mapping is used in situations when generic preference-based HRQoL measures such as EQ-5D were not included in clinical trials but disease-specific instruments were. Since most reimbursement agencies require generic (EQ-5D) values in reimbursement dossiers, this poses a problem. In those situations, a common solution is to develop a model for predicting or 'mapping' the generic HRQoL scores from the disease-specific questionnaire using a dataset that includes both of these measures. The developed model which specifies how the disease specific scores translate into generic (e.g. EQ-5D) scores, can subsequently be used to predict generic HRQoL at the patient level in situation where this was not directly observed. Generally, the generic HRQoL instrument considered in mapping studies is the EQ-5D. Here, we investigate which methods for mapping perform best. The methods compared in terms of fit and prediction error are: OLS, linear mixed effect models, Tobit models, beta regressions, finite mixture models, multinomial logit models and Bayesian Networks.

Implications for economic evaluations

In this thesis we will address two issues regarding HRQoL and economic evaluations. First, we investigate how changes in population HRQoL affect the estimation of QALYs (and consequently of ICERs) in life prolonging interventions. Second, we will address the impact of various methodologies used to model the HRQoL data on economic evaluations' output. In what follows, we will present both issues in more detail.

Economic evaluations of life prolonging interventions often use modelling techniques to estimate health benefits expressed in quality-adjusted life years (QALYs). That is especially the case when an intermediate effect (for instance: weight loss, newly detected cases through screening) is connected to causally related events (such as the incidence of diseases and/or death) which were not directly observed within the trial period of the intervention because the follow-up period is too short (Buxton et al. 1997). Although modelling is a powerful tool for estimating the health benefits as it enables to synthetize evidence from different sources, often certain assumptions are required. For example, for estimating QALYs gained of life prolonging interventions, economic evaluation analysts need to decide on the assigned HRQoL value(s) in the added years of life. Because guidelines are missing, economic evaluation analysts make diverse methodological choices in estimating QALY gains. Nevertheless, these choices may have a substantial impact on the final economic evaluation results (i.e. ICER). In standard practice of economic evaluation, two choices are typically made. First, the vast majority of economic evaluations assume that HRQoL in the added years of life equals perfect health (with the value of one). That is, the absence of the disease under study translates into perfect health

(de Kok et al. 2009). However, to assume that, for example, a 70 year-old woman whose death is postponed by 20 years due to a life prolonging intervention will experience no HRQoL losses in added life years and hence remains in perfect health until death is rather unrealistic. Other researchers recognize this and propose using a population average age-specific HRQoL estimate (Anonychuk et al. 2009, Schousboe et al. 2011, Tosteson et al. 2008) during gained life years. In that way they attempt to account for the usually observed decrease in population health by age (Fryback et al. 2007). However, if the observed changes in population health by age can be explained by the relationship between health (here as measured by HRQoL) and proximity to death or mortality, then estimates of QALY gains (and therefore ICERs) can be further improved by making use of this relationship.

In this thesis we hypothesized that the observed relationship between HRQoL and age can be explained to a large extent by a relationship between increasing age-specific mortality and low HRQoL associated with the period near death. For example, population average HRQoL at age 80 may be lower than that at age 60 because there are many more individuals in their last year of life at age 80 than at age 60. Interventions that postpone death therefore to some extent also postpone health losses. In other words, if HRQoL values correlate with mortality and depend strongly on TTD, postponement of death will result in postponement of HRQoL losses and only the last years of life will be spend in poor health. Therefore, using age- and gender-specific population HRQoL estimates would result in an underestimation of QALYs gained due to these interventions compared to the situation in which HRQoL stratified by age, gender and TTD would have been used. Consequently, the ICER would be overestimated. Therefore, understanding what drives the observed changes in population health by age provides us with better tools for accurately estimating the health gains in economic evaluations of life prolonging interventions.

The second issue addressed refers to the impact that various methodologies have on estimating the outcomes in economic evaluations that use mapping functions. Mapping is commonly accepted by reimbursement agencies as a last resort solution in the situation in which EQ-5D or other generic HRQoL measures are not available in the clinical trial data. Nevertheless, guidelines indicating appropriate methodologies when developing the mapping models are not available. This is despite the fact that the method of choice for developing the mapping model is of critical importance as it has a direct effect on estimating HRQoL and QALY changes in those economic evaluations that need to use such models.

Aim Of This Thesis

The overall aim of this thesis is to investigate changes in population health, to explore a mechanism that can explain these changes and to assess their consequences and implications for economic evaluations of health interventions. In doing so, several methodological challenges associated with modelling HRQoL data were investigated. The issues addressed can be separated in several subheadings: the issue of using HRQoL as a measure of population health, the issue of investigating changes in population health over time and by population subgroups in an aging population and the issue of using these changes in economic evaluations, and the issue of employing appropriate techniques for modelling cross-sectional as well as longitudinal HRQoL data. Using various datasets for the Dutch population, including survey data and clinical trial data, this thesis provides empirical insights into the above-mentioned issues. In particular, it yields insights in terms of better understanding changes in population health and highlights their implications for economic evaluations. By better understanding what drives changes in population health, this thesis contributes to inform health policies targeted at addressing population aging as well as to improve the standard practice of economic evaluations of life prolonging interventions. Furthermore, this thesis adds to the literature of methodologies used for modelling the HRQoL outcome by exploring a large variety of methods applied for modelling the HRQoL data.

Research Questions And The Outline Of This Thesis

Considering the background presented in the previous sections, specific research questions can be formulated. These research questions contribute to the overall aim of this thesis as mentioned above:

- Are (Dutch) people living longer lives in better or worse health?
- Compared to the lower educated, are the higher educated (Dutch) people living more or less years in good health?
- What is the relationship between HRQoL, age and time to death (TTD)?
- What are the implications of the relationship between HRQoL, age and TTD for economic evaluations of life prolonging interventions?
- What methods should be used for modelling HRQoL data?

This thesis includes five chapters, each providing answers to particular research questions.

Chapters 2 and 3 of this thesis aim to answer the first and second research question, respectively, by establishing whether people live longer lives and in better health. Using survey data sets for the Dutch population, these chapters investigated population health trends for the Dutch population covering the calendar period between 2001 and 2008 and that between 2001 and 2011. While chapter 2 illustrates HRQoL trends by age and gender, chapter 3 presents HRQoL trends stratified by age, gender and level of education thus exploring whether there are differences in health trends by socioeconomic status, i.e. for the lower educated compared to the higher educated. Estimating such changes in population health has important implications for public policies targeted at addressing population aging. Examples include public policies referring to increasing the official retirement age or those referring to the affordability of healthcare.

Chapter 4 aims to answer the third research question by proposing a conceptual model to further explain the observed changes in population health by age. In doing so, the relationship between HRQoL, age and TTD is investigated using a longitudinal dataset linked to the mortality registry.

Chapter 5 focuses on the fourth research question by exploring the implications of the observed relationship between HRQOL, age and TTD on estimating health benefits in economic evaluations of life prolonging interventions. Therefore, this chapter projects the results from chapter 4 into the context of economic evaluations and exemplifies this mechanism for life prolonging preventive interventions. The results of this chapter are relevant for the standard practice of economic evaluation of life prolonging interventions.

To some extent the last research question is addressed in each chapter of this thesis. Chapters 2 and 3 and 5 used a beta regression approach for modelling cross-sectional HRQoL data. Chapter 4 proposed a mixed beta regression estimated using the Bayesian approach for establishing the relationship between TTD and HRQoL. Finally, in chapter 6 a large number of methods are compared and assessed for their predictive performance. This is presented in the context of a 'mapping' study which aims at mapping or predicting a generic HRQoL instrument (here EQ-5D) from a disease-specific questionnaire. Two randomized clinical trials were used for performing the analyses in this chapter.

Chapter 7 focuses on the main discussion points and policy recommendations that follow from the studies presented in this thesis. Finally, a summary is included in English and in Dutch.

Chapter 2

Did the health of the Dutch population improve between 2001 and 2008? Investigating age- and gender-specific trends in quality of life.

With Werner Brouwer and Pieter Van Baal

European Journal of Health Economics; 2015 Nov;16(8):801-811.

Abstract

Although many countries' populations have experienced increasing life expectancy in recent decades, quality of life (QoL) trends in the general population have yet to be investigated. This chapter investigates whether QoL changed for the general Dutch population over the period 2001-2008. A beta regression model was employed to address specific features of the QoL distribution (i.e. boundedness, skewness and heteroskedasticity), as well non-linear age and time trends. Quality-adjusted life expectancy (QALE) was calculated by combining model estimates of mean QoL with mortality rates provided by Statistics Netherlands. Changes in QALE were decomposed into those changes caused by QoL changes and those caused by mortality-rate changes. The results revealed a significant increase in QoL over 2001-2008 for both genders and most ages. For example, QALE for a man/woman aged 20 was found to have increased by 2.3/1.9 healthy years, of which 0.6/0.8 was due to QoL improvements.

Introduction

During the 20th century, the populations of Western European countries experienced a substantial increase in life expectancy (Cutler et al. 2006). Whereas in the early decades of that century life-expectancy increases mainly resulted from decreases in young-age mortality, in the past few decades most gains in life expectancy are attributable to decreases in old-age mortality. For example, only in the period 2001-2008, in the Netherlands, life expectancy at birth has increased by 1.57 years for women and 2.52 years for men (Statistics Netherlands. 2011), mainly due to decreases in 65+ mortality (Mackenbach et al. 2011). An important question that arises is to what extent these extra years of life are spent in good or bad health (Gruenberg 1977, Fries 2000). One of the most relevant economic implications of this increased longevity among the elderly is the effect on labour-force participation (Eggleston and Fuchs 2012). In this context, it is becoming increasingly important to estimate whether population ageing is accompanied with a concomitant increase in the quality of life. This chapter aims at doing so for the Netherlands. Further we will present a short review of the most commonly used population health measures and highlight the benefits of using quality of life (QoL) for monitoring population health.

Until recently, health had been measured solely by mortality-based indicators. Life expectancy and all-cause, disease-specific and infant mortality were compared in time, by region and across countries (Leon 2011, Wilmoth 2000). Although mortality-based indicators are useful, they provide insufficient information to fully describe the health of a population as they do not capture the impact of non-fatal diseases on health (Murray et al. 2000). To date, mainly two indicators – i.e. self-rated health and disability – have been used to monitor the impact of non-fatal diseases on population health.

Self-rated health measures health by using a single question that asks individuals to rate their own health (commonly by choosing one of the five possible answers ranging from very poor to excellent). However, despite this indicator's popularity, it is still unclear to what extent changes in self-rated health reflect differences in, or perceptions of, health (Layes et al. 2012). Disability measures usually consist of several questions targeted at measuring specific aspects of health status, especially aspects referring to physical health. Because neither disability nor self-rated health is sufficiently generic for evaluating the health of a population, economists and psychologists proposed the concept of Quality of Life (QoL, (Dolan 2000)). QoL measures health by including multiple questions referring to various health dimensions such as physical health, mental health and social functioning. Furthermore, as a continuous variable generally defined on a scale from 0 (death) to 1 (full health), QoL measures are able to capture a wide variety of

health states. Perhaps one of the most notable advantages of QoL measures compared to both self-rated health and disability is that, the measured health states are valued using preference weights. These weights are commonly obtained from a sample of the general population using recognized valuation techniques such as time trade off (TTO) or standard gamble (SG). In this chapter, we will use the SF-6D index which is able to distinguish 7500 health states. Each of these health states is assigned a value between 0 and 1 using the valuation algorithm developed by Brazier and Roberts (Brazier and Roberts 2004). These valuations were derived using the SG technique in which respondents from the general population were asked to choose between remaining in a state of ill health (defined by SF-6D) for a period of time or a medical intervention which would either restore perfect health or result in death.

QoL measurements are widely used in clinical trials and cost-effectiveness analyses (Drummond et al. 2005, Drummond et al. 2005). Although influential health economists have proposed using QoL to monitor population health (Cutler and Richardson 1997, Williams 1999), few studies have been reported in this area. A cross-country comparison in QoL and quality-adjusted life expectancy (QALE) has been published (Heijink et al. 2011), and trends in QALE for the American population have been investigated (Jia et al. 2011). More recently, trends in healthy life expectancy between 1990 and 2010 were reported for 187 countries by combining disease prevalences and corresponding disease-specific disability weights (Salomon et al. 2012). However, none of these studies looked at QoL trends in the general population and possible influences on QALE. This is mainly due to the lack of QoL data measured in the general population. Attempting to fill that gap, the present chapter investigates whether QoL changed for the Dutch population in the period 2001-2008 and seeks to interpret any such changes as gains or losses of healthy years lived (in addition to gains due to decreased mortality rates). The data analysed here is taken from an annual health survey in the Netherlands, which includes the SF-12 from which the SF-6D was derived and valued using the algorithm developed by Brazier and Roberts (Brazier and Roberts 2004). There are two main steps in our analysis. First, we consider how age- and gender-specific patterns in QoL changed over the period in question. Second, we interpret changes in QoL as additional gains or losses of healthy years lived by estimating how QALE changed over the period in question, thereby decomposing changes in QALE into those due to mortality-rate changes and those due to QoL changes.

In order to draw meaningful conclusions when investigating QoL trends in 2001-2008, we need to properly model the non-standard QoL distribution. Important concerns when modelling SF-6D QoL are: a bounded distribution between 0 and 1; a strongly left (i.e. negatively) skewed distribution; and heteroscedasticity, which is unsurprising for

such bounded variables. For these reasons, it is inappropriate to model QoL by assuming a normal distribution. Obviously, a bounded outcome such as QoL is not normally distributed because it is not defined for negative values. Moreover, the fact that the QoL is observed over a closed interval has two major implications. First, the conditional expectation function must be nonlinear, and, second, the conditional variance must be a function of the mean (Kieschnick and McCullough 2003). Clearly, both of these conditions would be violated if a normal distribution were to be assumed for the QoL outcome. In contrast, previous research has shown that these conditions are satisfied by beta regression models, which provide a flexible method for modelling both cross-sectional QoL data (Basu and Manca 2012, Hunger et al. 2011) and longitudinal QoL measurements (Hunger et al. 2012). In this chapter, we will use this state of the art and model the SF-6D score using a beta distribution assumption.

This chapter is organized as follows. Section 2 describes the data used in the present chapter. Section 3 presents in greater detail the methods used to model changes in QoL over time. Section 4 illustrates the main results and findings of our analyses. Finally, Section 5 draws conclusions and discusses the most salient points.

Data

This chapter was based on the Permanent Survey of Living Conditions (POLS: Permanent Onderzoek Leef Situatie) for the years 2001-2008. POLS is an on-going annual cross-sectional survey. The survey is sampled on records from a centralized municipal registry and does not include the institutionalized population. The POLS health survey monitors developments in lifestyle, health, medical consumption, preventive behaviour and well-being in the Netherlands and since 2001 has included the SF-12 questionnaire. The Health Module of the survey is collected using a face-to-face interview and a written questionnaire. The interviewer visits the participants at home, askes for informed consent, conducts an interview and then leaves a written (drop-off) questionnaire, which includes the SF-12 form. This questionnaire contains 12 items for measuring health across 8 dimensions such as: physical functioning, social functioning, mental health, bodily pain, role limitations-physical, role limitations-emotional, general health and vitality. By excluding the question regarding general health and by combing the two questions referring to role limitations, the SF-6D questionnaire (which has 6 questions) was obtained from the SF-12. Some interviewees failed to return the written questionnaire, so approximately 20-25% of the SF-12 items (used to obtain the SF-6D) were missing. For those who did respond, the mean self-rated health was slightly higher than for those who did not indicating that the missing completely at random (MCAR) assumption is violated. In

practice, it is difficult to distinguish between the other two missing data mechanisms as described by Rubin (Rubin 1976)(Rubin 1976)(Rubin, 1976), i.e. the missing at random (MAR) and missing not at random (MNAR) assumptions. For imputing the missing items of the SF-12 questionnaire, we used here multiple imputation (MI, (van Buuren et al. 1999, van Buuren et al. 2006, Raghunathan, T.E., Lepkowski, J.M., Van Hoewyk, J. 2001)) which yields valid estimates under the MAR assumption. Although it is difficult to test for departures from this assumption, Peyre and colleagues showed that, regardless of the missing data mechanism, MI is superior to other methods when imputing missing data in quality of life questionnaires (Peyre et al. 2011). Appendix 1 illustrates details regarding our imputation model. It is worth mentioning that the imputation resulted in ten different datasets which were first analysed separately. Subsequently, the estimates of interest from the ten analyses were combined using standard MI rules (Rubin 1987).

Utility valuation functions have been derived for SF-6D, by using the preference-based valuation algorithm developed by Brazier and Roberts. The analysed sample included 57612 individuals: 27804 men and 29808 women. *Error! Reference source not found.* below shows descriptive statistics of this sample including details of the socio-demographics variables employed: age and gender. For reasons that will become obvious later, in this chapter we focus only on these two socio-demographic variables and we excluded others such as education or income. *Error! Reference source not found.* shows that both men and women have similar ages in our sample. Furthermore, mean QoL is larger for men compared to women which is consistent with previous knowledge about QoL for men and women (Heijink et al. 2011). It is worth mentioning that mean QoL has been calculated from the 10 imputed datasets used. *Error! Reference source not found.* shows also a list with assigned SF-6D values that have high frequency in our dataset: 0.922, 0.863, 0.8 and 1. These are imperfections resulting from the SF-6D valuation process and are beyond our control. Compared with other QoL instruments such as the EQ-5D (Pullenayegum et al. 2010) or the HUI (Austin 2002), the distribution of the SF-6D index does not show the strong ceiling effect at the value of one: only approximately 5% of respondents have an SF-6D utility score of one.

Figure 2-1 shows that the observed SF-6D distribution is left-skewed for both genders; in this case, it is slightly more skewed for men than for women. Moreover, the distribution is bounded with values ranging from 0.345 to 1. A simple glance at Figure 2-1 suggests that methods based on normality distribution assumptions are unlikely to be suitable for modelling the asymmetric and bounded SF-6D outcome.

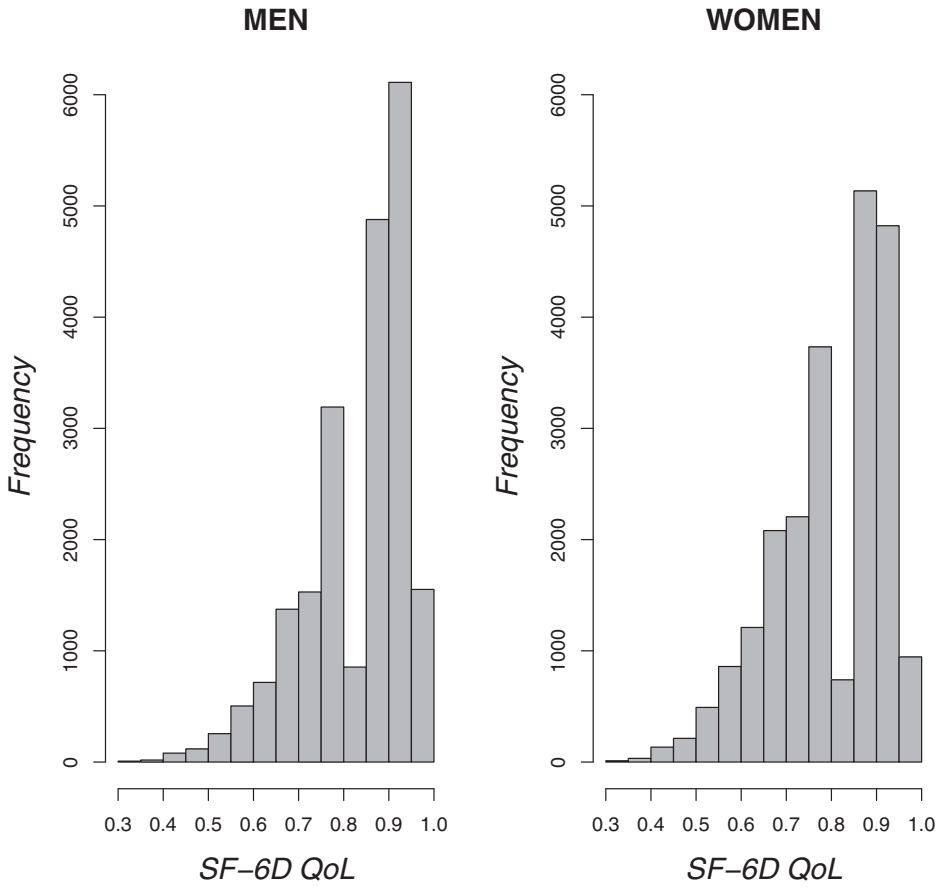

Figure 2-1: The distribution of the observed SF-6D.

Methods

Beta regression

Let us consider a variable Y following a beta density which is described by two parameters: the location or mean parameter denoted by $\mu(0< \mu <1)$ and the dispersion parameter also called the precision parameter denoted by $\phi(\phi > 0)$. Typically, in a beta regression framework, μ is modelled using a regression structure, that is as a function of various explanatory variables, whereas the precision parameter can either be assumed constant over observations (Ferrari 2004, Smithson and Verkuilen 2006) or modelled using a regression structure (Smithson and Verkuilen 2006). As QoL data are usually characterised by the presence of heteroscedasticity, we opted to explicitly model not

only the mean parameter but also the precision parameter ϕ as a function of explanatory variables

Note that, the beta distribution is defined on the closed interval between 0 and 1, hence, including the end points. However, in the beta regression context, the beta distribution needs to exclude the two end points. This is mainly because, for defining a regression structure on the parameters, a link function needs to be specified for mapping the observed interval onto the real line. For example, for beta regression, Cox tested a number of link functions for the mean parameter and found that logit link function works best (Cox 1996). Hence, for beta regression, the conditional mean parameter is commonly modelled using a logit link function. Such a link cannot be defined at 0 or 1. For data that is observed between 0 and 1 including 0 and 1, Ospina and Ferarri developed a general class of regression models that include a mixture of a continuous and a discrete distribution (Ospina 2010). The continuous part is described by a beta distribution defined on (0,1) and the discrete part is defined by a Bernoulli distribution at either point. These are referred to as beta inflated models by Ospina and Ferarri. Because, we had observations at 1, in this chapter, we used the beta inflated at one (BEINF1) distribution to model the SF-6D data. Hence, besides the two typical parameters that define a beta distribution, i.e. μ and ϕ, the BEINF1 distribution will have one more parameter, the inflation parameter, denoted by ν that models the observations at 1. This distribution is implemented in the package gamlss which is freely available in R (R. A. Rigby, et al. 2010, Stasinopoulos and Rigby 2007). The interested reader can find more details regarding the parameterization of the BEINF1 in gamlss elsewhere (R. A. Rigby, et al. 2010).

The expectation of a random variable Y that follows a BEINF1 distribution, as parameterized in gamlss, is $E[Y] = \frac{\mu + \nu}{1 + \nu}$. Hence, the inflation parameter has an effect on the mean estimate. The variance of E[Y] will be estimated using the delta method (Cramer 1946) with the first-order Taylor expansion (for details, see Appendix). Furthermore, the variance of a variable Y that follows a BEINF1 distributions is a function of both the mean parameter μ and of the precision parameter ϕ (Ospina 2010).

Previous research has shown that the relationship between QoL and age is typically non-linear and it is modelled using either a polynomial of order 2 (Austin 2002) or dummy variables (Hunger et al. 2011, Li and Fu 2009). The disadvantage of using dummies is that the gradient is not smooth and there is an inherent arbitrariness in defining age categories. Although polynomials are smooth functions, they have the limitation of being global functions. To overcome these problems when modelling the non-linear relation between QoL and age, we will use smoothers such as P-splines (Eilers 1996).

When modelling QoL changes between 2001 and 2008, the observed age-specific QoL trajectory may be different for different years. For generating different P-splines curves for each year in 2001-2008, we used varying coefficients models (VCM, (Hastie and Tibshirani 1993)). Within gamlss, the R function for using P-splines in conjunction with VCM models is denoted by pvc(b, by=c, df), where b is generally a continuous variable, c denotes either a continuous or a discrete variable and df are the degrees of freedom that define the smoothing parameter. Hence, the complexity of a P-spline is defined by the corresponding df Model selection and; therefore, the optimal df were obtained here by minimizing the Akaike information criterion (AIC, (Akaike 1973))[1] .

A common problem for bounded variables such as QoL is the presence of heteroscedasticity because the variance changes with the mean (Kieschnick and McCullough 2003). Hence, it is important to model not only the mean or location parameter but the entire outcome distribution. This has been achieved here by modelling all three parameters that described the beta inflated at one (BEINF1) distribution as functions of explanatory variables. Note that, as commented above, the inflation parameter models just the probability at one. By modelling the other two parameters of the BEINF1 distribution (μ and ϕ), we implicitly model the shape of the QoL distribution and hence, the skewness and the variation. Hence, each parameter of the BEINF1 distribution has been modelled using the pvc function in gamlss. We used log link functions for both the precision parameter ϕ and the inflation parameter ν. Because men and women have different QoL patterns and for ease of computation, we developed separate regression models for men and women. The final developed models for men and women are denoted by equations (1)-(3) and (4)-(6), respectively:

$$\text{logit}(\mu)=\text{pvc}(a,\text{by}=y,\text{df}=5.6) \quad (2.1)$$

$$\text{logit}(\phi)=\text{pvc}(a,\text{by}=y,\text{df}=4.5) \quad (2.2)$$

$$\log(\nu)=\text{pvc}(a,\text{by}=y,\text{df}=2.63), \quad (2.3)$$

1 In gamlss a generalized Akaike information criterion (GAIC) is implemented GAIC=GD+$\lambda\times$df, where λ is the penalty, GD is the global fitted deviance GD=−2L with L is the fitted log likelihood function. The user can chose various penalties, however the most widely used ones are λ=2 and λ=log(n). Whereas for λ=2, Akaike criterion is derived, for λ=log(n) with n being the sample size, the Schwartz Bayesian criterion (SBC) is obtained. In this chapter, we chose to use AIC over SBC because the AIC splines were smoother and gave a more convincing visual fit to the data.

$$\text{logit}(\mu)=\text{pvc}(a,by=y,df=4) \quad (2.4)$$

$$\log(\phi)=\text{pvc}(a,by=y,df=1.7) \quad (2.5)$$

$$\log(\nu)=\text{pvc}(a,by=y,df=1.68), \quad (2.6)$$

where a denotes the age variable and y denotes the calendar year.

As we used MI to impute the missing items of the SF-12, the developed models for each gender were applied to each of the ten imputed data sets. The estimates of interest, i.e. the parameters of the QoL distribution together with the mean QoL for persons aged 20-80, were derived for each model. Then, final results of the estimates were obtained by applying the standard MI rules defined in (Rubin 1987) and adapted by (Schafer and Olsen 1998). The ultimate step in our analysis was to calculate the QALE.

Decomposing quality adjusted life expectancy (QALE)

QALE is an equivalent of years lived in full health (Murray et al. 2000) and was computed by combining estimates of mean QoL from the regression models developed using GAMLSS with mortality rates from Statistics Netherlands. QALE was calculated for every gender, age and year, using the Sullivan life table approach (Sullivan 1971):

$$QALE(a,t)=\frac{\sum_{i=0}^{max} L(a+i,t)\times QoL(a+i,t)}{l(a,t)} \quad (2.7)$$

where $QALE(a, t)$ denotes the number of healthy years lived by a person of age a in year t, $L(a, t)$ are the number of years lived at age a in year t, $QoL(a, t)$ is the mean QoL at age a in year t, and $l(a,t)$ is the total number of survivors at age x in year t. As stated in the introduction (above), changes in QALE between 2001 and 2008 are to be decomposed into those due to QoL changes and those due to mortality-rate changes. We therefore calculated QALE in two situations. Firstly, for each gender and year in the period 2001-2008 we computed QALE by assuming that the mean QoL was constant over time (QALE (QoL=ct)) and equal to the mean QoL estimated from the regression models for year 2001. In this way, we observed changes in QALE over 2001-2008 due to trends in mortality rates only. Secondly, we computed QALE by using the estimated mean QoL from the beta models for all years in the period 2001-2008 (QALE (QoL≠ct)) and thus revealed changes in QALE due to both mortality-rate and QoL changes. The difference between QALE (QoL=ct) and QALE (QoL≠ct) indicates the healthy years gained or lost due to changes in QoL only.

Results

While smoothers such as P-splines aid in modelling the unknown relationship between QoL and age, the estimated smooth curves' coefficients cannot be tested for significance in the p-value sense. However, p-splines are an excellent tool for prediction purposes which was, in fact, the aim here. Our modelling approach enables us to examine the entire estimated distributions of the SF-6D for various ages. Figure 2-2 illustrates the trends of the estimated mean QoL (the vertical line) and those of the estimated SF-6D QoL distributions for men and women aged 20, 65 and 75, respectively. Over 2001-2008, for both genders, mean QoL shows an increasing trend. Moreover, the shape of the QoL distribution changes over this period. Although variability increases with age, it decreases

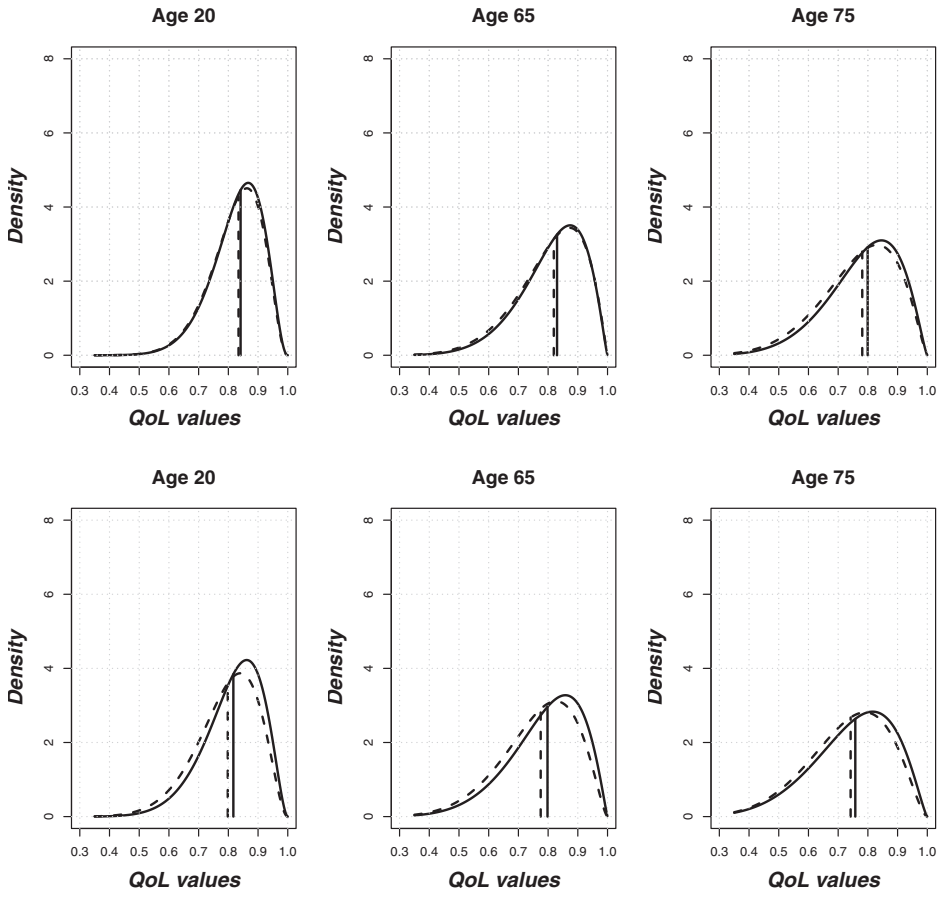

Figure 2-2: Trends of the estimated QoL distributions and the estimated mean QoL (vertical line) for different ages: top — a model for men; bottom — a model for women. The dashed line is for year 2001, whereas the continuous line is for year 2008

over this period. However, all these trends over 2001-2008, both in the mean and in the shape of the QoL distribution, are small. For interpreting such small changes, in particular those in the mean QoL, we use QALE, which, as stated in section 3.2, combines estimates of mean QoL with mortality rates from Statistics Netherlands. Figure 2-3 illustrates the two ingredients used in calculating the QALE: the mean QoL estimated from the regression models, and the mortality rates. Figure 2-3 suggests that QoL increased more for women than for men. Although it is not easily observable from Figure 2-3, mortality rates as provided by Statistics Netherlands, decreased more for men than for women.

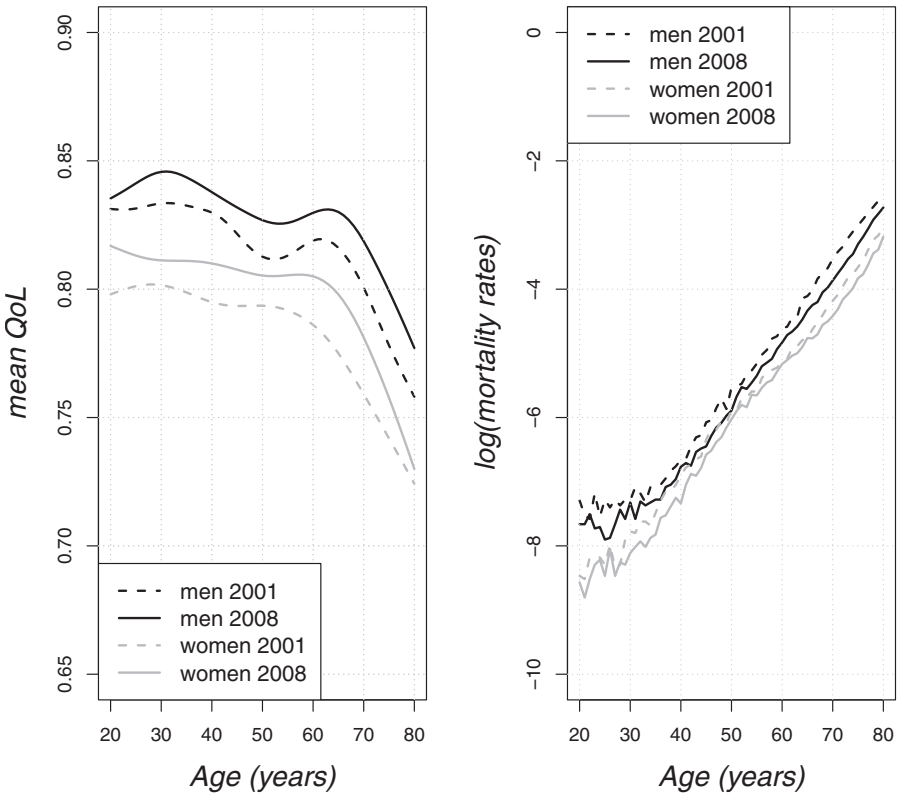

Figure 2-3: Mean QoL (left) and mortality rates (right), in 2001 and 2008, for men and women

Using the method described in Subsection 3.2 (above), we calculated QALE (QoL=ct) and QALE (QoL≠ct). In this way, we decomposed changes in QALE over the period 2001-2008 due to mortality-rate changes and QoL changes. *Table 2-2* shows life expectancy (LE) and QALE for the years 2001 and 2008 for an average man and woman aged 20 and 65, respectively. We chose to illustrate the results for these ages in order to juxtapose the healthy life years for the younger and the aging population. As *Table 2-2* shows, over 2001-2008, LE at age 20 and 65, increased by approximately 2.3 and 1.7 years for men,

respectively. Over the same period, LE at age 20 and 65, increased by approximately 1.4 and 1.2 years for women, respectively. Table 2 indicates that although, in 2001, QALE (QoL=ct) was slightly higher for an average woman aged 20 than for an average man of the same age, by 2008 the situation was the converse; that is, QALE (QoL=ct) showed an increase of approximately 1.8 healthy years for an average man aged 20 and an increase of 1.1 healthy years for an average woman of the same age. Perhaps even more interestingly, QALE (QoL≠ct) indicates that over 2001-2008 the increase in QoL induced approximately 0.6 and 0.8 healthy years for a man and a woman aged 20, respectively.

Table 2-1: Descriptive statistics of the POLS sample

VARIABLE	MEN	WOMEN	BOTH
mean Age ±SE	47.16±0.2	47.96 ±0.1	47.6±0.15
mean SF-6D QoL ±SE	0.82 ±0.001	0.79 ±0.001	0.8±0.001
% QoL values of 1	7%	3.8%	5%
% QoL values of 0.922	27%	19.45%	24%
% QoL values of 0.863	19%	20%	20%
% QoL values of 0.8	6.5%	7.4%	7%
Year: -2001	3395 (12.21%)	3639 (12.21%)	7034 (12.21%)
-2002	3418 (12.29%)	3691 (12.38%)	7109 (12.34%)
-2003	3467 (12.47%)	3666 (12.30%)	7133 (12.38%)
-2004	3905 (14.04%)	4168 (13.98%)	8073 (14.01%)
-2005	3732 (13.42%)	3912 (13.12%)	7644 (13.27%)
-2006	3465 (12.46%)	3764 (12.63%)	7229 (12.55%)
-2007	3070 (11.04%)	3341 (11.21%)	6411 (11.13%)
-2008	3352 (12.06%)	3627 (12.17%)	6979 (12.11%)

Table 2-2: LE, QALE (QoL=ct) and QALE (QoL≠ct) with 95% confidence intervals for a man and a woman aged 20 and aged 65, respectively

		MAN		WOMAN	
		2001	2008	2001	2008
Age 20	LE	57.06	59.33	61.81	63.24
	QALE(QoL≠ct)	46.87[46.30-47.44]	49.20[48.65-49.74]	48.42[47.84-49.00]	50.28[49.68-50.88]
	Gain in QALE		2.33		1.86
	Gain in QALE due to mortality		1.76		1.06
	Gain in QALE due to QoL		0.57		0.8
Age 65	LE	15.90	17.64	19.72	20.88
	QALE(QoL≠ct)	12.71[12.46-12.96]	14.30[14.05-14.55]	14.90[16.62-15.18]	15.92[15.63-16.22]
	Gain in QALE		1.59		1.02
	Gain in QALE due to mortality		1.33		0.85
	Gain in QALE due to QoL		0.26		0.17

Smaller changes over this period were found at age 65. At that age, over 2001-2008, QALE (QoL=ct) increased by approximately 1.3 and 0.8 healthy years for an average man and an average woman, respectively. Changes in QoL in the same period induced approximately 0.3 and 0.2 healthy years for an average man and woman, respectively.

Figure 2-4, which illustrates QALE gains for persons aged 20-80 years, shows that over the analysed period, QALE gains decreased with age for both genders. Over 2001-2008, QALE (QoL=ct) gains were generally smaller for women than for men for most ages in the range 20-80, whereas QALE gains due to QoL changes (the cross-hatched region) were higher for women than for men aged 20-65 years and higher for men than for women ages 65+. However, for persons aged 65+, QALE gains due to changes in QoL were small (less than 0.2 healthy years).

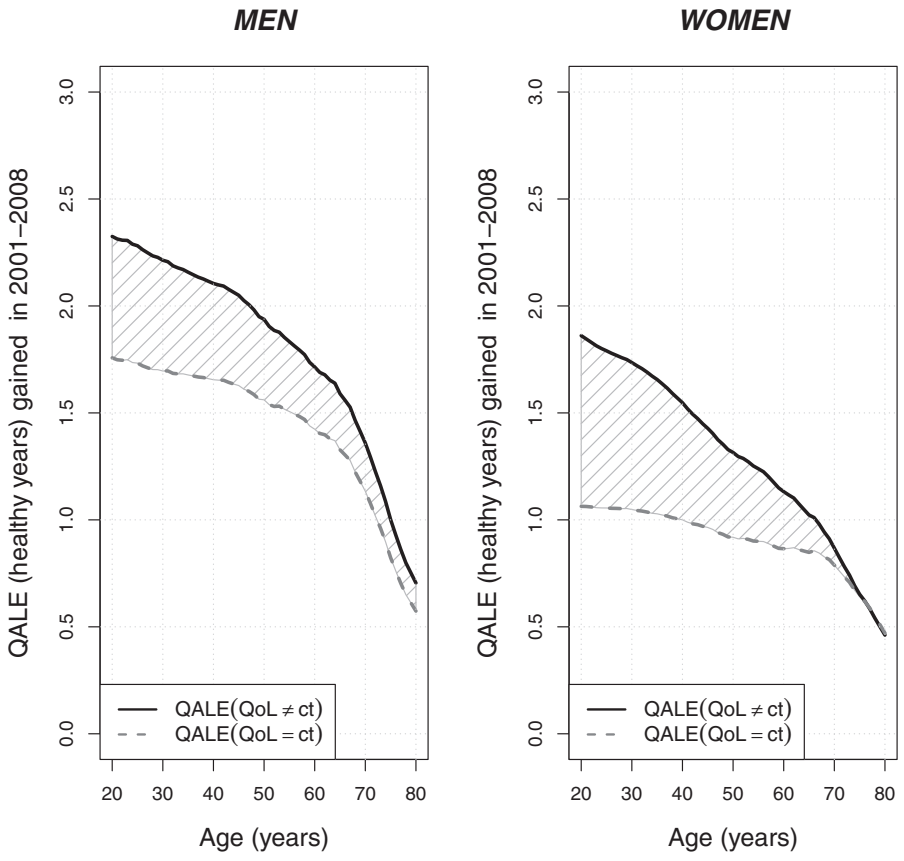

Figure 2-4: QALE gains over 2001-2008 for persons aged 20-80 decomposed into: QALE gains due to mortality rate changes (the red dotted line); QALE gains due to both mortality rate and QoL changes (the black solid line). The cross-hatched region represents QALE gains

Conclusions And Discussion

Our results indicate that, over the analysed period of 2001-2008, the health of the Dutch population improved for both men and women. More broadly and arguably of greater importance, we observed small QoL improvements in 2001-2008, i.e. at most ages, QoL increased with approximately 0.02 for women and 0.01 for men. However, as indicated by quality-adjusted life expectancy (QALE), in 2001-2008, these apparent small changes resulted in approximately 0.8 and 0.6 heathy years gained by a men and a women aged 20. These results indicate that, small improvements in QoL among the general population may have a sizeable impact on population health as measured by QALE. Although reduced mortality rates contribute more to the improvement in QALE than does the increase in QoL, our results indicate that, in recent years, the Dutch population has gained both in length and quality of life. Interestingly, for most ages, Dutch women benefited more from QoL changes than Dutch men did, while the converse is true of mortality-rate changes.

As shown by Statistics Netherlands, in 2001-2008, life expectancy at all ages increased more for men than that for women (Statistics Netherlands. 2011). With respect to QoL changes, we are not aware of any chapter reporting on QoL trends for the Dutch population. However, there are several studies reporting on trends in healthy expectancy (HE) (Perenboom et al. 2004, Water et al. 1996, Picavet and Hoeymans 2002, Majer et al. 2013, Bruggink et al. 2009). Because these studies used various health measures, calendar periods and age groups, they reached diverse conclusions on the trends. For example, Bruggink and colleagues reported that the number of years lived without disability has decreased in 1980-2010 for both men and women, but slightly more for men. On the other hand, Perenboom et.al. reported similar results for both men and women, i.e., in 1989-2000, the number of years with severe and moderate disability has decreased while the number of years with minor disability has increased. However, more recent research that investigated trends in HE for 187 countries including the Netherlands indicated that the number of years lived without disability has increased in 1990-2010 for the Dutch population and this increase was larger for Dutch women compared to Dutch men (Salomon et al. 2012).

This chapter has several strengths. Firstly, it confirms that beta distribution is suitable for modelling the SF-6D score due to its flexibility in modelling highly skewed outcomes (Hunger et al. 2012, Figueroa-Zúñiga and Arellano-Valle, Reinaldo B., Ferrari, Silvia L.P. 2013). Moreover, an issue often rose when modelling HRQoL data is the presence of heteroscedasticity, which is a typical characteristic of such bounded outcome variables. Our modelling approach provided an explicit solution to this issue by modelling both

parameters that describe the beta distribution, i.e. the location and the precision parameter, as functions of explanatory variables such as age and gender. In doing so, we implicitly modelled the shape of the QoL distribution and hence, the skewness and the variation. Note that, for beta distributed outcomes, the conditional variance is a function not only of the precision parameter but also of the mean. This is particularly important for bounded outcomes, because for these response variables the variance changes with the mean (Kieschnick and McCullough 2003). This modelling approach enabled us to explore trends not only in the mean QoL but also in the shape of the QoL distribution. We found that with advancing age, variation increased while skewness decreased. Furthermore, over the analysed period of 2001-2008, the QoL variation decreased whereas skewness increased. An additional advantage of our approach is the use of VCM and P-splines to model the relationship between QoL, age and calendar year. P-splines proved to be useful in modelling the relationship between age and QoL, which was unknown a priori and is often difficult to model.

This chapter has a number of limitations. Firstly, in comparison with estimating single parameters, the interpretation of the smooth curves' coefficients is more difficult. To address this limitation, we used QALE. Therefore, changes in mean QoL over the period 2001-2008 were interpreted as gains or losses of healthy years lived. Secondly, it could be argued that QoL should have been modelled using other variables that had previously been proven to predict QoL well, such as education or income (Muennig et al. 2005, Luo et al. 2005, Cherepanov et al. 2010). In the present chapter, models including the explanatory variables of age, year and education were developed (results not shown). Although QoL differed significantly for various educational brackets, health gains were observed for both the low and the highly educated. Nevertheless, when the results were averaged on age and calendar year to compute QALE in a Sullivan Life Table, they were similar to those given above in this chapter. It should be noted that mortality rates for different education or income levels were not available. Thirdly, the analyses presented above include only the non-institutionalized population, which may lead to biased QoL estimates. In order to quantify the impact that the institutionalized population may have had on the QoL estimates, we performed a calculation involving the percentage of the institutionalized Dutch population. For each age, we assumed that the institutionalized persons had the lowest SF-6D index. Our calculations showed that the QoL trend estimates were not changed by the inclusion of the institutionalized population. Fifthly, the results presented above are derived from the analyses performed for the imputed data sets. It should be noted that analyses of the complete data set were also performed (results not shown). Due to a relatively small percentage of missing data (20-25%), results from the complete data and from the imputed data were similar, though QoL was slightly overestimated using the complete data (differences between 0.01 and 0.07). In

addition, in the sample used, many respondents had the same SF-6D utility score. For example, the values 0.922, 0.863 and 0.8 were assigned to approximately 24%, 20% and 7% of the individuals, respectively. This makes the SF-6D variable quasi-discrete at some intervals. Previous research indicates that, generally, the beta distribution is robust to such violations of the continuity assumption (Tamhane et al. 2002). Finally, in this chapter we used the SF-6D derived from the SF-12. It should be noted that the observed range of the SF-6D was between 0.345 and 1. In one chapter (Fryback et al. 2007), the authors showed that for a national survey sample of non-institutionalized adults, both the range and the mean of a number of QoL indices (e.g. HUI2, HUI3, EQ-5D, SF-6D derived from the SF-36) differ significantly. For example, the minimum observed value for the EQ-5D was -0.11 whereas for the HUI3 it was -0.34 and for the SF-6D (derived from SF-36) it was 0.3. The range discrepancies between various QoL indices suggest that our results may be sensitive to the QoL instrument used. In addition, a recently published chapter (Luo et al. 2012) indicates that the SF-6D derived from the SF-36 is more discriminative than that derived from the SF-12; hence, the former is preferable for use in population health surveys.

In conclusion, this chapter provides compelling evidence that Dutch people are not only tending to live longer than they previously did but are also living more healthily. From a policy perspective, our results can be considered important in various contexts. Firstly, the on-going debates on mounting healthcare expenditures have raised controversial questions regarding the efficiency of such expenditure. In that context, the benefits of healthcare - including both length and quality of life - need to be rigorously investigated. Our approach and results may be useful in justifying increased healthcare expenditures, though in the context of the present chapter we cannot draw any conclusions regarding the causal relationship between the increase in QALE that we have observed and increased healthcare spending. Secondly, that we have observed (moderate) quality of life improvements at most ages is also relevant to the on-going debates regarding extended labour-force participation by the elderly and the raising of pension ages. In those respects, observing increases in peoples' health levels around the legal or practical retirement ages are especially important. Obviously, translating our current results into answers to these policy questions requires further research. The current chapter represents an important first step in demonstrating, at least for the Netherlands, that in recent years the general public has been living longer and more healthily than was previously the case.

Appendix

Multiple imputation model

We used a fully conditional specification (FCS), also known as multivariate imputation by chained equations (MICE), proposed by various authors (van Buuren et al. 1999, van Buuren et al. 2006, Raghunathan, T.E., Lepkowski, J.M., Van Hoewyk, J. 2001) for each SF-12 item with missing values, conditional on all other variables in an imputation model. We developed the imputation model based on variables from the POLS face-to-face interview, in particular: year, age, educational level, marital status, self-rated health, general practitioner (GP) visits, smoking status, number of sport activities per week, happiness, number of working hours per week, number of church visits per month. The number of imputations used was established following pre-defined guidelines (Graham 2007). Given the percentage of missing data (around 20-25% of the SF-12 items) and the computer power necessary, ten imputations were used to impute the SF-12 missing values.

Analysing multiple imputed data involves two steps: first a standard method is applied to each simulated data set, then the estimates of interest from each data are combined to obtain a final result using the rules defined in (Rubin 1987) and adapted from (Schafer and Olsen 1998).

Delta method

The delta method estimates the variance of a non-linear function of one or more variables by using the Taylor expansion around the mean of the variables. Therefore, if x_0 and y_0 are the mean values of μ and ν, respectively, the first order Taylor expansion of $f(\mu,\nu)=\frac{\mu+\nu}{1+\nu}$ about the values (x_0, y_0) is:

$$\begin{aligned} f(\mu,\nu) &\approx f(x_0,y_0)+\frac{\partial f(\mu,\nu)}{\partial \mu}\Big|_{(x_0,y_0)}(\mu-x_0)+\frac{\partial f(\mu,\nu)}{\partial \nu}\Big|_{(x_0,y_0)}(\nu-y_0) \\ &= f(x_0,y_0)+\frac{1}{1+y_0}(\mu-x_0)+\frac{1-x_0}{(1+y_0)^2}(\nu-y_0) \end{aligned} \tag{2.8}$$

Then,

$$\begin{aligned} Var[f(\mu,\nu)] &= \frac{1}{(1+y_0)^2} Var[\mu] + \frac{(1-x_0)^2}{(1+y_0)^4} Var[\nu] + 2\frac{1-x_0}{(1+y_0)^3} cov(\mu,\nu) \\ &= \frac{1}{(1+y_0)^2} Var[\mu] + \frac{(1-x_0)^2}{(1+y_0)^4} Var[\nu] + 2\frac{1-x_0}{(1+y_0)^3} \rho(\mu,\nu)\sigma_\mu\sigma_\nu \end{aligned} \quad (2.9)$$

where p is the product moment correlation between μ and ν. If we assume ρ=0, the estimated variance of the f(μ,ν) is:

$$Var[f(\mu,\nu)] = \frac{1}{(1+y_0)^2} Var[\mu] + \frac{(1-x_0)^2}{(1+y_0)^4} Var[\nu] \quad (2.10)$$

Although many countries' populations have experienced increasing life expectancy in recent decades, quality of life (QoL) trends in the general population have yet to be investigated. This chapter investigates whether QoL changed for the general Dutch population over the period 2001-2008. A beta regression model was employed to address specific features of the QoL distribution (i.e. boundedness, skewness and heteroskedasticity), as well non-linear age and time trends. Quality-adjusted life expectancy (QALE) was calculated by combining model estimates of mean QoL with mortality rates provided by Statistics Netherlands. Changes in QALE were decomposed into those changes caused by QoL changes and those caused by mortality-rate changes. The results revealed a significant increase in QoL over 2001-2008 for both genders and most ages. For example, QALE for a man/woman aged 20 was found to have increased by 2.3/1.9 healthy years, of which 0.6/0.8 was due to QoL improvements.

Chapter 3

Health inequalities in the Netherlands: trends in quality adjusted life expectancy (QALE) by educational level

With Parida Wubulihasimu, Frederik Peters, Wilma Nusselder and Pieter Van Baal

European Journal of Public Health, 2016.

Abstract

Quality-adjusted life expectancy (QALE) has been proposed as a summary measure of population health because it encompasses multiple health domains as well as length of life. However, trends in QALE by education or other socio-economic measure have not yet been reported. This chapter investigates changes in QALE stratified by educational level for the Dutch population in the period 2001-2011. Using data from multiple sources, we estimated mortality rates and health-related quality of life (HRQoL) as functions of age, gender, calendar year and educational level. Subsequently, predictions from these regressions were combined for calculating QALE at ages 25 and 65. QALE changes were decomposed into effects of mortality and HRQoL. In 2001-2011, QALE increased for men and women at all educational levels, the largest increases being for highly educated resulting in a widening gap by education. In 2001, at age 25, the absolute QALE difference between the low and the highly educated was 7.4 healthy years (36.7 vs. 44.1) for men and 6.3 healthy years (39.5 vs. 45.8) for women. By 2011, the QALE difference increased to 8.1 healthy years (38.8 vs. 46.9) for men and to 7.1 healthy years (41.3 vs. 48.4) for women. Similar results were observed at age 65. Although the gap was largely attributable to widening inequalities in mortality, widening inequalities in HRQoL were also substantial. In the Netherlands, population health as measured by QALE has improved, but QALE inequalities have widened more than inequalities in life expectancy alone.

Introduction

In recent decades, inequalities in health status by educational level or other measures of socio-economic status (SES) have persisted and widened (Turrell and Mathers 2001, Martikainen et al. 2001, Mackenbach et al. 2003, Singh and Siahpush 2006, Mackenbach et al. 2008, Meara et al. 2008, van Kippersluis et al. 2010, Maki et al. 2013). Health expectancy calculations have consistently shown that less educated persons not only have a shorter life expectancy but also live more years in poor health (Maki et al. 2013, Kunst et al. 2005, Majer et al. 2011), (Van Oyen et al. 2011, Cambois et al. 2001, Bronnum-Hansen and Baadsgaard 2008, Bruggink 2009, Crimmins and Saito 2001, Davis et al. 1999). In the Netherlands, for example, Statistics Netherlands (CBS) reported a difference of approximately seven years in life expectancy at birth between the groups with the lowest and the highest educational levels, whereas this difference was as high as 14 years for disability-free life expectancy (Bruggink 2009). These findings may be explained through a variety of mechanisms running from education to health but also vice-versa (Cutler and Lleras-Muney 2010) e.g. important channels through which education influences health are life-style related risk factors such as smoking, alcohol consumption and physical inactivity but also financial resources, housing and work conditions and access to care.

To compare the health of different populations, summary health measures that combine information on both non-fatal and fatal health outcomes are recommended (Murray et al. 2000). A variety of summary population health measures are available that differ in how they account for the health impact of non-fatal diseases. Hence, various definitions of disability and self-rated health (SRH) measures are used extensively. However, there are concerns that these measures inadequately measure the health impact of non-fatal diseases. Firstly, SRH is measured using a single question: individuals are asked to rate their own health (usually by choosing one of five possible answers ranging from very poor to excellent). Despite this indicator's popularity, it is still unclear to what extent changes in SRH reflect actual differences in health, or just perceptions thereof (Salomon et al. 2009, Layes et al. 2012). Secondly, disability measures usually focus on aspects of physical health and disregard other important dimensions such as mental health. Hence, one concern is that such one-dimensional measures are unlikely to capture the complex multidimensional concept that is human health (Crimmins 1996).

With the aim of broadly measuring aspects of health, economists and psychologists proposed the concept of health-related quality of life (HRQoL) (Dolan 2000). Although HRQoL has been extensively used for monitoring health in clinical trials, its use for measuring population health is still limited, mainly due to its unavailability in large-

scale surveys. HRQoL is a generic multidimensional measure of health; in other words, it measures health with multiple questions that refer to both physical and mental health resulting in a large number of health states. Perhaps, the most distinctive feature of HRQoL compared to the other health measures is that the HRQoL health states are valued by measuring the strength of individuals' preferences for these health states. For this valuation step representative samples of the general population are used. The explicit valuation of different health states in terms of a common metric makes it possible to calculate QALE in combination with mortality rates.

Few studies have been reported using HRQoL for monitoring population health: examples include a cross-country comparison in HRQoL and QALE (Heijink et al. 2011), and time trends in QALE (Jia et al. 2011). Recently, QALE has been proposed as a more useful indicator of health inequalities than all other healthy life expectancy measures (Collins 2013); however, to our knowledge, trends in QALE by education or other SES measures have not yet been reported in the literature.

The aim of this chapter is to investigate, for the Dutch population over the period 2001-2011, trends in health inequalities, measured in QALE, that are related to educational level. Furthermore, QALE trends by education level were decomposed those trends into effects of disparities in HRQoL and mortality. The focus will be on whether health inequalities by education as measured by QALE have winded or narrowed in 2001-2011.

Methods

Datasets

This chapter used several data sources. Firstly, to estimate mortality by education for the period 2001-2011, we used the Dutch Labour Force Survey (LFS) linked to mortality registry. In LFS, each year about 60.000 households are added which participate in multiple question rounds within one year. Because education is not a time dependent variable, we used only the baseline round of LFS of the years 1997-2011 from which we extracted education level; therefore, in fact we did not make use of the longitudinal design of this data. Information on mortality rates was obtained by linking the LFS survey to mortality registries for the period 2001-2011. Therefore, our final sample consisted of all individuals that were interviewed between 1997 and 2011 and died between 2001 and 2011. This means that the follow-up time of the respondents ranged between 15 years and 1 year. Thus, the average follow-up time was by design larger for more recent years. However, because follow-up was included in the models developed for estimating

mortality rates by education it is unlikely that the differences in follow-up time by year would bias our results.

Previous studies using data collected in the Netherlands generally classified educational level into four categories (Bruggink 2009, Kulhanova et al. 2014); however, due to restricted sample sizes we have pooled the upper two categories, resulting in the following categories: *low educated*, which includes primary education (ISCED 0 and 1); *medium educated*, which includes pre-vocational education (ISCED 2) and *highly educated*, which includes secondary and tertiary education (ISCED 3-6).

Secondly, for estimating HRQoL by age and educational level we used data from the POLS (*Permanent Onderzoek Leef Situatie*, in english Permanent Research into Living Conditions) survey over the period 2001-2011. POLS is sampled on records from the municipal population registries and does not include the institutionalised population. Between 2001 and 2009 the self-reported health data was collected by interviews and written questionnaires, whereas since 2010 it was collected via the Internet and by telephone. The number of respondents in POLS used in our analyses was relatively similar between 2001 and 2011, i.e. it ranged from 7000 to 8000 respondents with about 50% men and 50% women, respectively.

In POLS, information on health is obtained by using the SF-6D questionnaire, which comprises six questions on health aspects such as physical functioning, pain, vitality, social functioning, role limitations and mental health (see Appendix). SF-6D incorporates 241 health states, which in this chapter are valuated with the algorithm developed by Brazier and Roberts (Brazier and Roberts 2004) for which respondents from the UK general population were asked to choose between remaining in a state of ill health (as defined by the SF-6D) for a certain period of time or a medical intervention that would either restore them to perfect health or result in death. The resulting SF-6D HRQoL value ranges between 0.345 and 1.

For our analyses, we restricted the sample to persons aged 25+ and discarded all observations from which data was missing *Table 3-1* presents descriptive statistics of the data used in our analysis.

Table 3-1: Descriptive statistics of the datasets used

		Men	Women	Total
LFS data	Population aged 25+	389527	399902	789429
	Mean age (years)±SE	47.45±0.006	47.31±0.004	47.38±0.005
	Highly educated %	71.2	63.2	67.1
	Medium educated %	20.4	25.8	23.1
	Low educated %	8.4	11	9.7
POLS data	Population aged 25+	27268	28985	56253
	Mean Age (years)±SE	51.41±0.092	50.37±0.091	50.88±0.065
	Mean HRQoL)±SE	0.83±0.001	0.8±0.001	0.81±0.001
	Highly educated %	59.7	49.7	54.6
	Medium educated %	28.8	34.6	31.8
	Low educated %	11.5	15.7	13.6

\SE denotes standard error

Data analyses

The analyses performed in this chapter involved three steps. First, we estimated mortality rates by age, gender and educational level using the LFS. Second, we estimated the HRQoL by age, gender and educational level using POLS. Finally, we combined these two outcomes in a lifetable to estimate QALE. We decomposed changes in QALE over time and by education into those due to mortality-rate changes and those due to HRQoL changes.

Mortality rates by education

To estimate mortality rates stratified by age, gender and educational level, we constructed a panel dataset whereby the annual number of deaths of all persons ever interviewed for the LFS is obtained from the death registry and the number of person-years is estimated as the sum of the number of people surveyed in a particular year and the number of survivors from the previous year. A Poisson regression model with the number of person-years as offset and the expected number of deaths by age, gender, year and educational level as outcome variable was fitted on that dataset. Predictor variables were dummy variables indicating educational level and interactions thereof with age and calendar year (both as continuous variables). Dummies for each calendar year and age were included to control for confounding. In addition, a variable measuring the length of follow-up time in the LFS and an interaction term of follow-up time with age were added to the model. This is intended to control for selection effects in the LFS registry. Finally, mortality rates estimates were calibrated to be consistent with total mortality rates for the Dutch population (see Appendix for further details).

HRQoL by education

The SF-6D was modeled as a function of age, gender, year and educational level. We have followed recommendations of recent research for modelling the SF-6D outcome using beta regression (Basu and Manca 2012, Hunger et al. 2011, Hunger et al. 2012) which has alos been used in chapter 2 of this thesis. Furthermore, the unknown non-linear relationship between HRQoL and age was modelled here using penalised-splines (P-splines (Eilers 1996)). These smoothers have excellent numerical properties and have been proposed in chapter 2 for modelling the non-linear relationship between HRQoL and age. Furthermore, we also used dummies for education and interactions thereof with age and calendar year.

Quality-Adjusted Life Expectancy (QALE) and decomposition

QALE is an equivalent of years lived in full health (Murray et al. 2000) and was computed in this chapter using the Sullivan method (Sullivan 1971). QALE used model estimates of HRQoL and mortality by age, gender, year and educational level:

$$QALE(a,g,t,e) = \frac{\sum_a L(a,g,t,e) \times HRQoL(a,g,t,e)}{l(a,g,t,e)} \quad (3.1)$$

where $L(a,g,t,e)$ is the number of person-years lived and $l(a,g,t,e)$ is the total number of survivors.

For analysing how QALE varies over time and across different educational levels, we disentangled contributions to differences between two QALE estimates into effects of mortality changes and of HRQoL changes. For example, we decomposed the estimated QALE difference between years 2001 and 2011, as well as the QALE difference between low educated and highly educated (and between the medium educated and the highly educated), into effects of mortality and of HRQoL. In addition, QALE differences at ages 25 and 65, by education and over time were decomposed also by age by assessing the contribution of all ages to the observed changes in QALE at a specific age. The age 65 was chosen based on the fact that, at the moment, this is the official statuary retirement age in many European countries. Therefore, it is particularly important to observe changes in population health and possible inequalities in health around retirement.

Because mortality and HRQoL are non-linear functions of age and are therefore non-additive, we used a decomposition algorithm that takes this into account (Andreev et al. 2002), (see Appendix).

Results

Trends in QALE between 2001 and 2011

Table 3-2 indicates that, in 2001-2011, LE and QALE increased for both men and women with a greater increase for men than for women. Furthermore, for men and women aged 65, the higher the education level, the greater the gain in both LE and QALE. For men aged 25, the greatest gain in both LE and QALE was for the medium educated whereas for women aged 25, the greatest gain was for the highly educated. Hence, as time progressed, the more highly educated Dutch lived not only longer but also in better health than those of lower educational levels which indicates that health inequalities by education have widened in 2001-2011. Also, at both ages 25 and 65, LE increased more than QALE for men, whereas the converse was true for women (with the exception of medium educated women). This results in a compression of morbidity for low and highly educated women and in an expansion of morbidity for men at all educational levels. Furthermore, *Table 3-2* indicates that inequalities in QALE widened more than those in mortality as indicated by LE. For example, at age 25, the LE/QALE gap between the low and the highly educated widened by approximately 0.45/0.72 for men and by approximately 0.63/0.84 for women, respectively. At age 65, the LE/QALE gap between the low and the highly educated widened by approximately 0.8/0.8 for men and approximately 0.52/0.6 for women, respectively.

Table 3-2: LE and QALE in 2001 and 2011

Gender	Educational level	Year	Age 25				Age 65			
			LE	LE gain 2001-2011	QALE	QALE gain 2001-2011	LE	LE gain 2001-2011	QALE	QALE gain 2001-2011
Men	High	2001	53.83		44.08		16.81		13.42	
		2011	56.71	2.88	46.93	2.85	19.29	2.48	15.59	2.17
	Medium	2001	50.70		40.82		15.29		12.06	
		2011	53.92	3.22	43.83	3.01	17.53	2.24	13.97	1.91
	Low	2001	47.85		36.71		14.08		10.73	
		2011	50.28	2.43	38.84	2.13	15.76	1.68	12.10	1.37
Women	High	2001	58.75		45.79		21.00		15.66	
		2011	60.53	1.78	48.43	2.64	22.64	1.64	17.40	1.74
	Medium	2001	56.82		44.2		20.08		15.12	
		2011	58.35	1.53	45.77	1.57	21.49	1.41	16.33	1.21
	Low	2001	53.57		39.51		18.26		13.09	
		2011	54.72	1.15	41.31	1.80	19.38	1.12	14.23	1.14

Decomposing gains in QALE over time

Table 3-3 shows that the observed QALE gains were due not only to diminishing mortality rates but also to improvements in HRQoL, though the former had a larger impact than the latter especially for women aged 25 years, approximately half of the QALE gain over the period 2001-2011 was due to HRQoL improvements, whereas at age 65 approximately one-third of the QALE gain was due to HRQoL improvements. For men, less than one-fifth of the QALE gain in 2001-2011 was due to HRQoL improvements at both ages 25 and 65, respectively.

Table 3-3: Decomposition of QALE gains over the period 2001-2011

Gender	Educational level	Age 25			Age 65		
		QALE gains due to mortality	QALE gains due to HRQoL	Total QALE gains	QALE gains due to mortality	QALE gains due to HRQoL	Total QALE gains
Men	High	2.29	0.56	2.85	1.97	0.20	2.17
	Medium	2.55	0.46	3.01	1.76	0.15	1.91
	Low	1.85	0.28	2.13	1.29	0.08	1.37
Women	High	1.32	1.32	2.64	1.2	0.54	1.74
	Medium	1.13	0.44	1.57	1.04	0.17	1.21
	Low	0.82	0.98	1.80	0.78	0.36	1.14

Decomposing educational inequalities in QALE

Table 3-4 indicates that in 2001-2011, the gaps between the low and the highly educated and between the medium and the highly educated widened for most population subgroups (i.e. men and women aged 25 and 65), the exception being medium educated men aged 25. For example, at age 25, the gap between the low and the highly educated, widened by approximately 0.72 healthy years (mortality effect: 0.38; HRQoL effect: 0.34) for men and by 0.84 healthy years (mortality effect: 0.57; HRQoL effect: 0.27) for women. Similar results are observed for the elderly: at age 65 the gap had widened by approximately 0.8 healthy years (mortality effect: 0.64; HRQoL effect: 0.16) for men and by approximately 0.6 healthy years (mortality effect: 0.44; HRQoL effect: 0.16) for women. These results again indicate that the widening gap in QALE was due to widening inequalities in both mortality and HRQoL, with the former apparently contributing more than the latter for most subgroups. Nevertheless, the widening inequalities in HRQoL were substantial, especially at age 25. For example, for men aged 25 the increase in HRQoL inequalities accounted for approximately half of the increase in QALE inequalities, whereas for women of the same age the increase in HRQoL inequalities accounted for approximately one-third of the increase in QALE inequalities. As a result, inequalities as measured by QALE widened even more than inequalities in mortality as measured by LE.

Table 3-4: Decomposition of inequalities in QALE by education (baseline: highly educated) in 2001 and 2011

Gender	Educational level	Year	Age 25			Age 65		
			Difference in QALE due to mortality	Difference in QALE due to HRQoL	Total difference in QALE	Difference in QALE due to mortality	Difference in QALE due to HRQoL	Total difference in QALE
Men	Medium	2001	-2.49	-0.76	-3.25	-1.19	-0.18	-1.37
		2011	-2.23	-0.87	-3.10	-1.39	-0.23	-1.62
	Low	2001	-4.69	-2.68	-7.37	-2.11	-0.57	-2.68
		2011	-5.07	-3.02	-8.09	-2.75	-0.73	-3.48
Women	Medium	2001	-1.45	-0.14	-1.59	-0.68	0.14	-0.54
		2011	-1.67	-0.99	-2.66	-0.86	-0.21	-1.07
	Low	2001	-3.78	-2.5	-6.28	-1.95	-0.63	-2.58
		2011	-4.35	-2.77	-7.12	-2.39	-0.79	-3.18

Discussion

This chapter has shown that, in 2001-2011, both life duration and the number of years lived in good health as indicated by QALE increased for the Dutch population. At both the ages 25 and 65, respectively, over the investigated period, LE increased more than QALE for men, whereas the converse was true for women; therefore in 2001-2011, men experienced an expansion of morbidity while women experienced a compression of morbidity. Furthermore, for both genders, LE and QALE increased more for the highly educated than for the low and medium educated with the exception of medium educated men. Our decomposition analysis shows that improvements in QALE over time were driven mainly by declining mortality rates, though HRQoL improvements were not negligible, especially at age 25 and for women. Finally, we found that health inequalities by education as indicated by QALE widened for most population subgroups. Most of the increase in QALE inequalities was due to widening inequalities in mortality; however, widening inequalities in HRQoL caused inequalities as measured by QALE to widen even more.

Our results are in line with the literature that indicates widening health inequalities by SES in many European countries including Belgium, Denmark and France but also in New Zealand and the US (Van Oyen et al. 2011, Cambois et al. 2001, Bronnum-Hansen and Baadsgaard 2008, Bruggink 2009, Crimmins and Saito 2001, Davis et al. 1999). To our knowledge, QALE trends by educational level for the Dutch population have not been previously reported. However, a chapter investigating trends in disability-free life expectancy (DFLE) for the Dutch population by education reported that inequalities in

terms of that health measure were constant in various periods, i.e. between 1997 and 2000, as well as between 2005 and 2008 (Bruggink 2009). However, various measures of health result in different estimates. This is one of the reasons why it is important to use an adequate measure of population health. Disability based health measures such as disability-free life expectancy (DFLE) focus on measuring physical aspects of health and they capture the health status in one dimension (e.g. disabled/not-disabled). On the other hand, health-related quality of life (HRQoL) based measures such as quality-adjusted life expectancy (QALE) have the advantage of measuring multiple aspects of health including not only physical domains but also mental and social functioning domains. Therefore, such measures are multidimensional; i.e. HRQoL defines a variety of health states usually ranging between 0 (death) and 1 (full health). Hence, in our view, compared to other health measures such as DFLE, QALE offers a generic multi-dimensional measure of health and, when possible, should be used for reporting the magnitude of health and health inequalities in scientific and lay audiences.

Our findings may be related to changes in health care expenditures and wider societal developments such as the economic crisis. Previous research suggested that, in the Netherlands, an increase in the healthcare expenditures at the beginning of the investigated period here, i.e. in year 2001, led to increases in life expectancy (Wubulihasimu et al. 2015). Another line of research has argued that the global financial recession starting in 2008 led to severe cuts in public spending which may have widened inequalities in health, (Stuckler et al. 2011, McKee et al. 2012).

The institutionalized population was not included in POLS and was only partially included in LFS since persons might have been transiting to, for example, nursing homes before death. However, in general, the percentage of the institutionalised Dutch population remained constant over the period 2001-2011. Hence, it is unlikely that the exclusion of the institutionalized would have significantly affected the overall QALE trends in our chapter.

A second limitation of our chapter is the participation rates in the LFS and POLS data, especially because a selection bias associated with underestimating the effect of low socio-economic status on poor health was observed (Visscher 1997, Lorant et al. 2007). However, as both participation rates were fairly constant in 2001-2011, we would not expect this to have significantly affected the reported QALE trends.

This chapter has several noteworthy strengths. Firstly, we estimated both components – mortality and health status – disaggregated by SES. This goes beyond studies that solely estimated trends in health status or mortality by SES.

A second strength of our chapter is the use of QALE as a population health measure. As QALE was based on the SF-6D questionnaire which incorporates 241 possible health states, QALE is a broader measure of population health than life expectancy or other healthy life expectancy measures that are based on binary variables (i.e. disabled or not).

Investigating the underlying causes of the observed health trend inequalities for the Dutch population was beyond the scope of our chapter. This is, however, an interesting topic that merits further research.

Concluding, our findings that the highly educated tend to live longer and in better health than the less educated are especially relevant to the ongoing debates on extended labour participation for the elderly and on the raising of pension ages. On the one hand, as the more highly educated live longer, they place greater demands on pension resources. On the other hand, as they live a greater number of years in good health, they may participate in the labour force for longer. Social policies aimed at raising the legal retirement age should be adopted with caution, as not all population subgroups make equal demands on pension resources or have the same health status around the legal retirement age. Obviously, further research is required before our results can be translated into answers to such policy questions.

Appendix

We estimated mortality rates by education in two steps. First, we used the LFS data for estimating relative risks by age, calendar year and education level (denoted RR(a,t,e) which equals the mortality rate of educational class e, age a, year t divided by the mortality rate of the reference educational class age a, year t). To estimate RR(a,t,e) we fitted a Poisson regression model with the exposure as offset and the expected number of deaths by year, age, education class and calendar year as outcome variable:

$$E\left[D\middle|a,e,t,y\right]=e^{\theta' X} \tag{3.2}$$

where D denotes the number of deaths, y denotes calendar year, X a vector of predictor variables and θ the vector of coefficients that need need to be estimated. Predictor variables were dummy variables indicating educational class and interactions thereof with age and calendar year (both as continuous variables). To control for confounding a set of dummy variables for each year and age were added to the model. Furthermore, to account for selection bias by education in the LFS data, a variable measuring the length of follow-up time in the LFS and an interaction thereof with age were added to the model. Second, these relative risks were then used to decompose mortality rates from the total population by exploiting the following relationship:

$$m\left(a,t,e\right)=RR\left(a,t,e\right)\times\frac{m\left(a,t\right)}{\sum_e RR\left(a,t,e\right)\times p\left(e\middle|a,t\right)} \tag{3.3}$$

Where m(a,t,) denotes total mortality rates for age a, calendar year t as published by Statistics Netherlands and m(a,t,e) denotes mortality rates for age a, calendar year t and education level e. Equation (2) states that mortality rates in a particular year at a particular age are the weighted average of the mortality rates of the different educational subgroups. Note that p(a|t,e) denotes the proportion of individuals at an education subgroup at a particular age in a given year and was taken directly from the LFS data.

Chapter 4

Health losses at the End of Life. A Bayesian mixed beta regression approach

With Monique Verschuren, Susan Picavet, Werner Brouwer and Pieter Van Baal.

Journal of the Royal Statistical Society, Series A, Statistics in Society, 2016.

Abstract

The relationship between ageing, health, and healthcare expenditures (HCE) is of central importance to academics and public policymakers. Generally, it is observed that with advancing age, health deteriorates and HCE increase. This seems to imply that increases in life expectancy would strongly increase both the demand for HCE and the number of years lived in poor health. Previous research has shown that such straightforward conclusions may be flawed. For example, it has been established that not age but 'time to death (TTD)' is the main driver of increased HCE at advanced ages. This chapter aims to extend this line of research by investigating the relationship between age, TTD, and health, the latter being longitudinally measured via a health-related quality of life (HRQoL) questionnaire. We propose an approach for modelling the HRQoL outcome that accounts for both the non-standard nature of this response variable (e.g. bounded, left-skewed, heteroscedastic) and the panel structure of the data. Analyses were performed within a Bayesian framework. We found that health losses are centred in the final phase of life, which indicates that future increases in longevity will not necessary increase life years spent in poor health. This may alleviate the consequences of population aging.

Introduction

Over the past few decades, developed countries have seen a substantial increase in life expectancy (Cutler et al. 2006), which is mainly attributable to decreasing mortality rates at advanced ages (Eggleston and Fuchs 2012). As a result, the proportion of elderly persons in such countries' populations has also increased, a demographic phenomenon commonly referred to as 'population ageing'. Much societal, political and scientific debate has centred on population ageing, key questions being whether the rapidly growing healthcare expenditures (HCE) are affordable, and whether the statutory retirement age should be raised in order to increase labour force participation of the elderly. The rationales for, and consequences of, policy decisions in these areas depend on the extent to which the increased number of life years (due to increased longevity) would be spent in good health.

In general, it is observed that, with advancing age, health deteriorates and healthcare use increases (Fryback et al. 2007, Getzen 1992). This implies that increases in life expectancy would increase the number of years lived in poor health, which may limit scope for extending working lives and increase healthcare utilization. However, such seemingly obvious conclusions may be misleading. For example, several researchers have shown that healthcare use is centred in the final phase of life (Seshamani 2004, Seshamani 2004, Zweifel et al. 2004, Werblow et al. 2007, Wanless 2004). Using large datasets (e.g. healthcare insurers' claims databases or hospitals' registry data) including information on healthcare use of individuals followed until death, age (i.e. time since birth) was shown to poorly predict HCE when 'time to death' (TTD) was taken into account. This has an important implication: population ageing may have only limited impact on future HCE growth, since ageing implies the postponement of these expensive final years of life (Zweifel et al. 1999, Seshamani 2004, Zweifel et al. 2004). In this chapter, we argue that population ageing may have a limited impact not only on future HCE use but also on average population health state. For the purposes of the present research, health is measured using a health-related quality of life (HRQoL, (Dolan 2000)) questionnaire. HRQoL indicates health dimensions including physical health, mental health, and social functioning. An important feature of HRQoL is that health states are assigned numerical values using preference weights commonly obtained from samples of the general population, resulting in a single HRQoL index generally ranging from 0 (dead) to 1 (perfect health).

Previous research has shown that population average HRQoL decreases with age Heijink et al. 2011). This has alos been shown in chpater 2 and 3 of this thesis. We hypothesize that the observed relationship between HRQoL and age is largely attributable to a rela-

tionship between increasing age-specific mortality and the low HRQoL associated with the period prior to death. For example, it is reasonable to assume that the population average HRQoL at age 80 will be lower than that at age 60, because at age 80 there are many more individuals in their final year of life than at age 60. If this hypothesis were to be confirmed, there would be important implications for the economics of ageing populations, as a strong relationship between TTD and HRQoL indicates that increases in longevity go hand in hand with increases in HRQoL as shown in chapters 2 and 3. Consequently, this would have important implications for the ongoing political debate on raising the official retirement age.

This chapter aims to test the hypothesis stated in the previous paragraph by investigating the relationship between age, TTD and HRQoL. In doing so, we address several important methodological issues that arise due to the non-standard HRQoL response and the panel structure of the data; that is, we propose modelling the HRQoL outcome using the mixed beta regression and the flexibility of the Bayesian estimation. Note that previous studies that investigated the relationship between TTD and HCE used large longitudinal datasets extracted from hospital registries or insurers' claims databases, which do not include HRQoL measures. Generally, HRQoL data is scarcer than HCE data. For the main analyses of this chapter, we analysed a longitudinal dataset including information on 356 individuals who were followed for 16 years, with a maximum of four measurement rounds.

Modelling the HRQoL outcome can be problematic due to its non-standard distribution: that is, bounded (usually defined between 0 and 1) and strongly (left) skewed. Much research on modelling HRQoL relies upon the robustness and computational ease derived from the normality assumption (Austin 2002, Pullenayegum et al. 2010, Pullenayegum et al. 2012). However, normality and homoscedasticity are unlikely to hold when the dependent variable is bounded. Because the domain of the normal distribution (i.e. the domain of values for which its density is defined) is unbounded, predictions outside of the HRQoL domain, i.e. outside of the interval (0, 1] are possible. Recent research has proposed using a beta distribution when modelling HRQoL data (Basu and Manca 2012, Hunger et al. 2011, Hunger et al. 2012). A beta regression approach has been used in chpaters 2 and 3 as well which also utilized a particularly attractive feature of the beta distribution; that is, its recognition of a relationship between the mean and the variance that typically occurs for bounded variables (Kieschnick and McCullough 2003). Whereas a normally distributed variable can have any variance, a beta distributed variable with the mean close to one of its boundary values has a smaller variance than a beta distributed variable with the mean at the midpoint of the interval. Although it is possible to model the variance of a normally distributed variable (for example, as a function of explanatory

variables), using maximum likelihood approaches, estimating the variance as a function of the mean is not straightforward. Moreover, modelling HRQoL over time poses additional problems, as multiple observations over time for the same individual are correlated. A recent chapter proposed using a longitudinal beta regression estimated using the maximum likelihood approach for modelling HRQoL over time (Hunger et al. 2012). We expand on previous literature by estimating a mixed beta regression model within the Bayesian paradigm using Markov chain Monte Carlo (MCMC) methods to model the HRQoL data. Due to its flexibility in modelling longitudinal bounded outcomes, this approach has been recently proposed for modelling longitudinal data defined between 0 and 1 (Figueroa-Zúñiga 2013); however, to our knowledge this has not yet been used to model HRQoL data.

Our modelling approach is appealing for various reasons. First, we performed main analyses on a rather small sample size: 356 individuals followed until death. Unlike maximum likelihood approaches, Bayesian estimation enables small sample sizes to be accounted for, by employing prior information on model's parameters. Second, the MCMC methods permit great flexibility in specifying complex non-standard models that would be more difficult and time-consuming to estimate according to the classical maximum likelihood approach; for example modelling the heteroscedasticity in a natural and straightforward way by estimating the dispersion parameter as a function of explanatory variables and possibly of random effects. Another example consists in specifying models that assess the impact of different sources of bias (e.g. non-ignorable missing data) on the estimated parameters. Such sensitivity analyses will be exemplified in this chapter. Hence, compared to maximum likelihood approaches, the Bayesian estimation procedure makes it possible to develop more complex models that are easier to implement and estimate. Section 2 presents the data used in this chapter; Section 3 presents the methods employed to estimate HRQoL as a function of age and TTD; Section 4 illustrates the main results and findings of our analyses; finally, Section 5 draws conclusions from, and discusses the implications of, our findings.

Data

The data analysed here is from the Dutch Doetinchem Cohort Chapter, which investigates, using responses to SF-36 health questionnaires at five-year intervals, how changes in lifestyle and biological factors affect health aspects including the incidence of chronic diseases, physical and cognitive functioning, and health-related quality of life (Verschuren et al. 2008). Doetinchem is a municipality in the east of the Netherlands. The institutionalized population is excluded from the cohort. We analysed data from

four successive measurement rounds: Round 2 (1993 – 1997) to Round 5 (2008 – 2011) which included the HRQoL measurement. This data was linked to Statistics Netherlands' mortality registries in order to identify deceased individuals. In line with previous research on the relationship between HCE and TTD using samples of deceased individuals, for the main analyses of this chapter we used the subset of deceased individuals (aged 50+) from the Doetinchem data; that is, approximately 10% of the entire group. There are multiple measurements for about half of those who died and for about 30% of those who survived. The response rates for the deceased were: 43.34% for Round 2, 63.43% for Round 3, 50.17% for Round 4, and 35.05% for Round 5. The cases including missing responses were discarded from the dataset used for the main analyses of this chapter.

In the Doetinchem Cohort Chapter, health-related quality of life (HRQoL) is measured using the SF-6D questionnaire, which is an abbreviated version of the SF-36 health questionnaire. SF-6D has six health dimensions (i.e. questions), each with four to six levels. These questions monitor: physical functioning (6 levels), role limitations (4 levels), social functioning (5 levels), pain (6 levels), mental health (5 levels) and vitality (5 levels). (See Appendix A1 for a full description of the SF-6D questionnaire). A health state as defined by the SF-6D is obtained by selecting one statement from each dimension, starting with physical functioning and ending with vitality. In this way, 18000 different health states can be defined. For example, the state 111111 indicates perfect health, as the best (healthiest) answer was provided for each health dimension. In health economics, it is common to choose a smaller set of representative health states and use various techniques to attach utility values to those indicating individual preferences for them. In this case, we used the valuation algorithm developed by Brazier and Roberts (Brazier et al. 2002). With this algorithm utility values were attached to 249 selected states by using the standard gamble (SG) technique, whereby 611 respondents from the UK general population were asked to choose between remaining in an ill state (one of the 249 selected states) for a period of time or a medical intervention which would either restore them to perfect health or result in death. Each respondent was asked to value six health states. Consequently, this algorithm attaches a weight ranging between 0 (dead) and 1(full health) to each health state. In our sample, the observed SF-6D ranges between 0.388 and 1 (full health). For our purposes, TTD is defined as the time between the observed SF-6D measurement and death. In our sample, TTD ranges from less than one month to 193 months (approximately 16 years), while age ranges from 50 to 81 years. At baseline measurement the maximum observed age was 66 years; at the end of follow-up the maximum observed age was 81. *Table 4-1* compares descriptive statistics of deceased individuals aged 50+ with those of survivors aged 50+. Note that for survivors, TTD is not observed within the observation time, i.e. TTD is censored. *Table 4-1* indicates that, average TTD was significantly larger, and mean age at death slightly

larger, for women than for men. However, for both men and women the mean age at death is lower than that reported for the entire Dutch population between 1993 and 2011 (Statistics Netherlands. 2011). This is due to the observed follow-up time; obviously, if the follow-up had been longer, we would have observed more deaths at older ages. *Table 4-1* also shows that the mean SF-6D utility value is smaller for the deceased than for the survivors, the difference being approximately 0.02 for men and 0.03 for women. Moreover, mean SF-6D HRQoL values are larger for men than for women, which are consistent with previous findings (Heijink et al. 2011). In addition, *Table 4-1* indicates that the percentage of observations having an SF-6D value of 1 (full health) was generally small, lower for deceased than for survivors, and lower for women than for men.

Table 4-1: Descriptive statistics of the complete-case Doetinchem sample for those aged 50+

		Deceased	Survivors
men	*Number*	221	1746
	mean HRQoL± S.E. [1]	*0.78 ±0.005*	*0.80 ±0.001*
	mean age (years)± S.E.	*67.8 ±0.482*	66.3±0.183
	mean TTD (years) ± S.E.	*4.7±0.241*	4.8±0.084[2]
	% one HRQoL measurement	*56*	*31*
	% two HRQoL measurements	*27*	*28*
	% three HRQoL measurements	*17*	*33*
	% four HRQoL measurements	*0*	*8*
	% full health	*2.2*	*2.5*
women	*Number*	*135*	*1924*
	mean HRQoL± S.E.	*0.74 ±0.007*	*0.77 ±0.001*
	mean age (years)± S.E.	*68.6±0.620*	66.4±0.179
	mean TTD (years)± S.E.	*5.8 ±0.346*	5.3±0.089
	% one HRQoL measurements	62	34
	% two HRQoL measurements	27	31
	% three HRQoL measurements	10	30
	% four HRQoL measurements	1	5
	% full health	0.9	1.3

[1] S.E represents the standard error.

[2] For survivors, TTD is censored as it is not observed within the observation time.

Figure 4-1 shows how mean SF-6D HRQoL varies with age for both deceased and survivors and that, at most ages, it is lower for the former. The relationship is approximately linear for deceased and survivors, the greater wiggliness of the pattern for deceased being due to a substantially smaller sample size for that group. Furthermore, Figure 4-2 shows that the HRQoL distribution of deceased is more skewed and has greater variation than that of survivors.

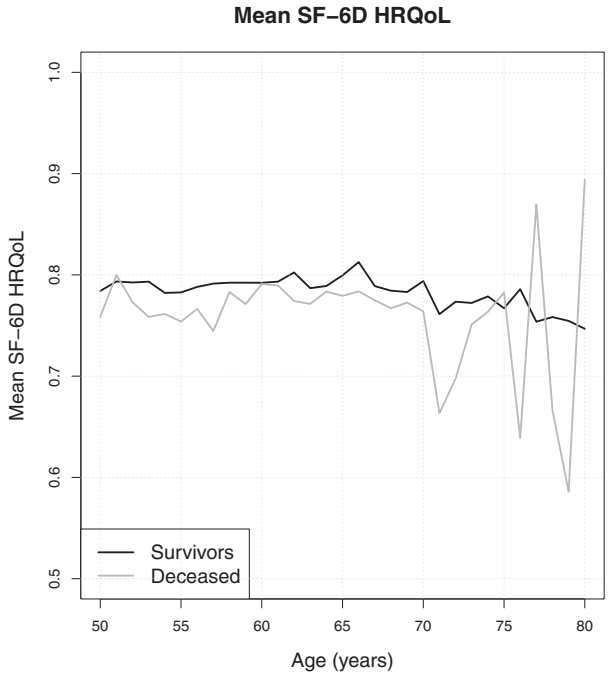

Figure 4-1: SF-6D HRQoL mean pattern by age for deceased and survivors aged 50+

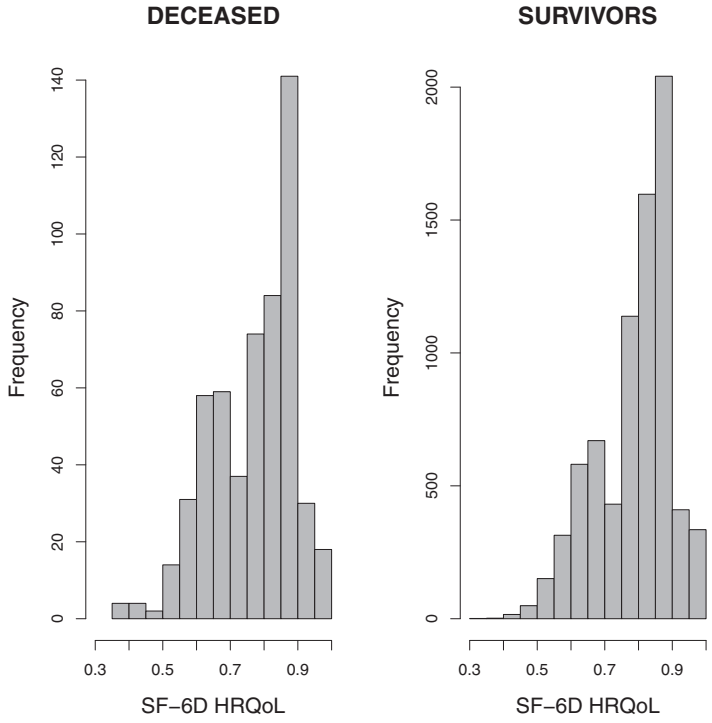

Figure 4-2: SF-6D HRQoL distribution for deceased and survivors aged 50+

Methods

We propose using a linear mixed model assuming a beta distribution for modelling the HRQoL outcome variable (i.e. a mixed beta regression, (Figueroa-Zúñiga 2013)) to handle the skewed and bounded nature of the HRQoL distribution and to properly model the correlation between and within subjects over time. In the present chapter this model is estimated using the Bayesian paradigm and the Markov chain Monte Carlo (MCMC) sampling methods available in WinBUGS (Spiegelhalter et al. 2003). In what follows, we present the main model which is the mixed beta regression model with model specification, prior specifications as well as robusteness to prior specifications in the Bayesian estimation. Furthemore, we present model extensions that account for various sources of uncertinity caused by ignoring the missing data or the sample of survivors (i.e. ignoring the TTD censoring) in the main analyses.

Bayesian mixed beta regression model

The probability density function of a variable y following a beta density described by the location/mean parameter $\mu(0<\mu<1)$ and precision parameter $\phi(\phi>0)$ is given by:

$$f(y|\mu,\phi)=\frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)}y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1},\ 0<y<1 \qquad (4.1)$$

where $\Gamma(\bullet)$ denotes the gamma function, $E[y]=\mu$ and $Var[y]=\frac{\mu(1-\mu)}{1+\phi}$. In this context, ϕ is interpreted as a precision parameter because, for each fixed value of the mean μ, $1+\phi$ is inversely proportional to the variance of y. If y has the above beta density, then $y\sim Beta(\mu\phi,(1-\mu)\phi)$. The beta distribution is defined on the open interval (0,1); therefore, for fitting a beta model, the observed values of 1 in the SF-6D index need to be transformed. Because in our dataset only 1.8% of the SF-6D measurements attained the maximum value of 1, we opted to transform the boundary point from 1 to 0.99, (Smithson and Verkuilen 2006, Verkuilen and Smithson 2012). An alternative approach for situations in which HRQoL values of 1 are more frequent is to use a two-part model in which one part models the probability mass at 1 and the other models the HRQoL using beta regression. Such an approach has been developed for beta regression models by Ospina and Ferrari (Ospina 2010).

In a beta regression framework, μ is modelled using a regression structure; that is, as a function of various explanatory variables, whereas the precision parameter can be either assumed to be constant over observations (Ferrari 2004, Smithson and Verkuilen

2006) or modelled using a regression structure (Smithson and Verkuilen 2006). In this chapter, we model parameters μ and ϕ using separate regression structures. A known issue when modelling any bounded variable is the presence of heteroscedasticity. As stated above, for bounded outcomes typically the variance changes with the mean, the variance of the beta distribution is determined by both μ and ϕ. Hence by modelling these two parameters as functions of explanatory variables we can explicitly address heteroscedasticity when modelling HRQoL. Because μ and ϕ do not restrict each other, they can be modelled separately.

Let us define a longitudinal design where $i=1,\ldots,n$, observations being clustered within $j=1,\ldots,m$ subjects. Also, let y_{ij} denote HRQoL for observation i and subject j, and let $y_{ij} \sim Beta(\mu_{ij}\phi_{ij},(1-\mu_{ij})\phi_{ij})$. In order to model changes in the individual trajectories over time, individually varying intercept and slopes are used when modelling μ_{ij} and ϕ_{ij}. Therefore, for each parameter, each individual will be assigned two random effects: one for the intercept and another for the observation time of the SF-6D measurement, which in this case is the HRQoL measurement round, which takes values 1, 2, 3 and 4. The longitudinal trajectory for μ_{ij} is given by equation (2), while that for ϕ_{ij} is given by equation (3).

$$G(\mu_{ij}(t_{ij}))=X(t_{ij})^{T}{}_{ij}\beta+R_{ij}\delta+b_{j0}+b_{j1}\times t_{ij} \quad (4.2)$$

$$H(\phi_{ij}(t_{ij}))=P(t_{ij})^{T}{}_{ij}\alpha+M_{ij}\gamma+c_{j0}+c_{j1}\times t_{ij} \quad (4.3)$$

where β and a are vectors of fixed effects with corresponding time-varying design matrices $X_{ij}^{T}(t_{ij})$ and $P_{ij}^{T}(t_{ij})$; δ and γ denote vectors of time-invariant covariates with corresponding design matrices R_{ij} and M_{ij}; b_j and C_j denote vectors of subject-specific random effects for individual *j,* with B_{j0} (or C_{j0}, respectively) denoting the random effects for the intercepts and b_{jl} (or c_{jl}, respectively) denoting the random effects for the observation time t_{ij}. Furthermore, *G* and *H* denote the logit and the log link functions that map the intervals (0,1) and $(0,\infty)$ onto the real line.

Specification of prior distributions

In order to complete the Bayesian representation, prior distributions need to be specified for all unknown parameters. For this purpose we considered a standard mixed effects model; hence we used the guidelines defined in Gelman & Hill (Gelman, A. and Hill, J. 2007) for specifying normal prior distributions with very large variances for the fixed-effects parameters; that is, $\alpha \sim N(0,100)$, $\beta \sim N(0,100)$, $\delta \sim N(0,100)$ and $\gamma \sim N(0,100)$, which indicates that we expect these fixed effects coefficients to be in the range $(-100,100)$. Furthermore, in a typical mixed model random, effects are assumed to be normally distributed:

$$\beta_j = (\beta_{j0}, \beta_{j1})^T \sim N(0, D) = N\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & \rho_{12}\sigma_1\sigma_2\\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}\right] \quad (4.4)$$

$$c_j = (c_{j0}, c_{j1})^T \sim N(0, \Sigma) = N\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\tau_1^2 & \theta_{12}\tau_1\tau_2\\ \theta_{12}\tau_1\tau_2 & \tau_2^2\end{pmatrix}\right] \quad (4.5)$$

where D and Σ denote the variance-covariance matrices of the random effects; σ_1^2 and σ_2^2 (or τ_1^2 and τ_2^2, respectively) are the variances for b_{j0} and b_{j1} (or c_{j0} and c_{j1}, respectively) or the between-subject variances and the within-subject variances, respectively; $\rho_{12}\sigma_1\sigma_2$ (and $\theta_{12}\tau_1\tau_2$, respectively) denotes the covariance between b_{j0} and b_{j1} (or between c_{j0} and c_{j1}, respectively); and ρ_{12} (or θ_{12}, respectively) denotes their correlations. Here, for computational reasons, we modelled the variance-covariance matrices *D* and Σ using a Wishart prior distribution, for example, $D \sim Wishart(\Omega, df)$, where Ω is a scale matrix and df represent the degree of freedom; we used an identity matrix prior for Ω and we set the degrees of freedom to 3 (one more than the dimension of Ω) in order to induce a uniform prior on the correlation.

In Section 3.3, using the above prior specifications, we select the model that produces the best short-term predictions of HRQoL data. Note that in the above formulation the precision parameter ϕ was modelled using a regression structure; however, in the next section, for illustrative purposes, ϕ is also be considered constant over observations. For those situations, we specify a prior distribution of the form $\phi = 1/U^2$ with $U \sim U(0, p)$, with P large (*p=100,* for example), which was shown to be less informative than the commonly used inverse gamma prior distribution for the precision parameter (Gelman, A. and Hill, J. 2007). Section 3.4 considers model robustness to prior specifications.

Model specification

To explore the relationship between HRQoL and TTD we developed two mixed beta regression models: one including the variables age, gender and TTD, the other including only the variables age and gender. This chapter refers to these models as 'the TTD model' and 'the age-specific model', respectively. From the model specifications we deliberately excluded disease indicators. This is because both age and TTD can be regarded as proxy variables that 'borrow' their explanatory power in a statistical model from other determinants of HRQoL, of which diseases processes are the most important. The more adjustments for various diseases made in the statistical analysis, the less the variables age and TTD are expected to matter. For this reason, we also excluded socio-economic status variables from the model specification because it has been shown, for example, that education is a good predictor of mortality (Cutler et al. 2006, Lleras-Muney 2005).

Our selection of models was based on a minimum loss of Bayesian Deviance Information Criterion (DIC, (Spiegelhalter et al. 2002)). Model specification was investigated in two steps. First, we considered the mixed beta regression for the location parameter μ_{ij} and a common precision parameter ϕ for each observation y_{ij}. Second, we considered the mixed beta regression for the location parameter μ_{ij} and a different precision parameter for each observation; hence, at this step we also modelled ϕ using a regression structure. *Table 4-2* and *Table 4-3* present model selection results for the age-specific model and for the TTD model, respectively. For both models, DIC values indicate that adding random effects substantially improves the model fit. Furthermore, modelling the precision parameter using a regression structure reduces DIC by about 278 units in the age-specific model and 190 units in the TTD model. For the model specification of μ and ϕ, we fitted models that used terms such as age^2, TTD^2 and various interactions between the variables age, gender and TTD. However, DIC values indicated that such specifications would not improve the model fit: DIC reductions of less than 5 units were not considered significant/relevant (Spiegelhalter et al. 2002). Hence, we opted for the simpler model specification in those situations.

Table 4-2: Model selection for the age-specific model

Model	Nr.	logit($\mu_{i(j)}$)	$\varphi_{i(j)}$	*DIC*
Model selection for $\mu_{i(j)}$	(1)	$\delta_0+\delta_1 g_i+\beta_1 a_i$	$\varphi=1/U^2, U\sim U(0,100)$	-820
	(2)	$\delta_0+\delta_1 g_i+\beta_1 a_i+\beta_2 a_i\times g_i$	$\varphi=1/U^2, U\sim U(0,100)$	-822
	(3)	$\delta_0+\delta_1 g_i+\beta_1 a_i+\beta_2 a_i^2$	$\varphi=1/U^2, U\sim U(0,100)$	-822
	(4)	$\delta_0+\delta_1 g_i+\beta_1 a_i+b_{j0}$	$\varphi=1/U^2, U\sim U(0,100)$	-1051
	(5)	$\delta_0+\delta_1 g_i+\beta_1 a_i+b_{j0}+b_{j1}\times t_{ij}$	$\varphi=1/U^2, U\sim U(0,100)$	-1114
Model selection for $\varphi_i(j)$	(6)	$\delta_0+\delta_1 g_i+\beta_1 a_i+b_{j0}+b_{j1}\times t_{ij}$	$\log(\varphi_i)=\gamma_0+\alpha_1 a_i$	-1145
	(7)	$\delta_0+\delta_1 g_i+\beta_1 a_i+b_{j0}+b_{j1}\times t_{ij}$	$\log(\varphi_i)=\gamma_0++\gamma_1 g_i+\alpha_1 a_i$	-1146
	(8)	$\delta_0+\delta_1 g_i+\beta_1 a_i+b_{j0}+b_{j1}\times t_{ij}$	$\log(\varphi_i)=\gamma_0+\alpha_1 a_i+\alpha_3 a_i\times g_i$	-1140
	(9)	$\delta_0+\delta_1 g_i+\beta_1 a_i+b_{j0}+b_{j1}\times t_{ij}$	$\log(\varphi_{ij})=\gamma_0+\alpha_1 a_i+c_{j0}$	-1201
	(10)	$\delta_0+\delta_1 g_i+\beta_1 a_i+b_{j0}+b_{j1}\times t_{ij}$	$\log(\varphi_{ij})=\gamma_0+\alpha_1 a_i+c_{j0}+c_{j1}\times t_{ij}$	-1335

*a denotes age, g denotes gender and t denotes measurement time (or measurement round).

The final age-specific model is described by model specification number (10) in *Table 4-2* and the final TTD model is showed by model specification number 12 in Table 4-3. For the remainder of this chapter, these models will be referred to as 'baseline models'.

Table 4-3: Model selection for the TTD model

Model	Nr.	logit($\mu_{i(j)}$)	$\varphi_{i(j)}$	*DIC*
Model selection for $\mu_{i(j)}$	(1)	$\delta_0+\delta_1 g_i+\beta_1 TTD_i$	$\varphi=1/U^2, U\sim U(0,100)$	-823
	(2)	$\delta_0+\delta_1 g_i+\beta_1 a_i+\beta_2 TTD_i$	$\varphi=1/U^2, U\sim U(0,100)$	-823
	(3)	$\delta_0+\delta_1 g_i+\beta_1 a_i+\beta_2 TTD_i+\beta_3 a_i\times g_i$	$\varphi=1/U^2, U\sim U(0,100)$	-826
	(4)	$\delta_0+\delta_1 g_i+\beta_1 a_i+\beta_2 TTD_i+\beta_3 a\times TTD_i$	$\varphi=1/U^2, U\sim U(0,100)$	-825
	(5)	$\delta_0+\delta_1 g_i+\beta_1 a_i+\beta_2 TTD_i+\beta_3 TTD_i^2$	$\varphi=1/U^2, U\sim U(0,100)$	-827
	(6)	$\delta_0+\delta_1 g_i+\beta_1 a_i+\beta_2 TTD_i+b_{j0}$	$\varphi=1/U^2, U\sim U(0,100)$	-1061
	(7)	$\delta_0+\delta_1 g_i+\beta_1 a_i+\beta_2 TTD_i+b_{j0}+b_{j1}\times t_{ij}$	$\varphi=1/U^2, U\sim U(0,100)$	-1125
Model selection for $\varphi_{i(j)}$	(8)	$\delta_0+\delta_1 g_i+\beta_1 a_i+\beta_2 TTD_i+b_{j0}+b_{j1}\times t_{ij}$	$\log(\varphi_i)=\gamma_0+\alpha_1 TTD_i$	-1156
	(9)	$\delta_0+\delta_1 g_i+\beta_1 a_i+\beta_2 TTD_i+b_{j0}+b_{j1}\times t_{ij}$	$\log(\varphi_i)=\gamma_0+\alpha_1 a_i$	-1157
	(10)	$\delta_0+\delta_1 g_i+\beta_1 a_i+\beta_2 TTD_i+b_{j0}+b_{j1}\times t_{ij}$	$\log(\varphi_i)=\gamma_0+\alpha_1 TTD_i+\alpha_2 a_i$	-1158
	(11)	$\delta_0+\delta_1 g_i+\beta_1 a_i+\beta_2 TTD_i+b_{j0}+b_{j1}\times t_{ij}$	$\log(\varphi_{ij})=\gamma_0+\alpha_1 TTD_i+c_{j0}$	-1289
	(12)	$\delta_0+\delta_1 g_i+\beta_1 a_i+\beta_2 TTD_i+b_{j0}+b_{j1}\times t_{ij}$	$\log(\varphi_{ij})=\gamma_0+\alpha_1 TTD_i+c_{j0}+c_{j1}\times t_{ij}$	-1310

*a denotes age, g denotes gender and t denotes measurement time (or measurement round).

The developed models were estimated using Markov Chain Monte Carlo (MCMC) simulation methods implemented in WinBUGS. In our MCMC simulation, for each model, we used 30 000 iterations with a burn-in of 5000 iterations and three chains with different starting values. Furthermore, all the necessary diagnostic tests for assessing convergence – e.g. autocorrelation, trace and history – were performed and desirable results were observed in the chains (i.e., the chain for each parameter was stationary and uncorrelated). Moreover, convergence was assessed using Gelman and Rubin's convergence diagnostic (Gelman and Rubin 1992) for the three chains that had different starting values. For brevity, detailed numerical results of the above procedures are not shown.

In the Appendix we compare the above models with equivalent specifications under the normality assumption; that is, the linear mixed effects model and the simple linear model, which were estimated using WinBUGS assuming non-informative prior distributions. Often, researchers attempt to address some of the issues surrounding bounded outcomes by transforming the dependent variable. A common approach uses the normal distribution for estimation after performing a logistic transformation on the dependent variable. While this transformation removes the problem of obtaining expected values outside of permissible bounds, it also assumes that the variance would be stabilized, which can be of concern given that other distributional models such as beta and simplex distributions imply that this transformation will not stabilize the variance. That said, in the present chapter we considered also estimating the linear

model and the linear mixed effects model after logistic transformation of the HRQoL data. All of these models estimated under the normality assumption were compared with the mixed beta regression in terms of their predictive capacity as indicated by DIC. Furthermore, QQ plot comparisons were investigated. DIC indicated that the mixed beta regression offers a more flexible model, which predicts the atypical HRQoL outcome better than all the models that assume a normal distribution (including those estimated on the transformed data).

Robustness to prior specifications

In our developed model specifications we considered standard prior distributions for a linear mixed effects model; that is, normal distributions with large variances for the fixed effects parameters and multivariate normal distributions for the random effects. However, the impact of the scale choice under the normal model may not be neglected, especially when measurements present outliers. In such situations, a t-distribution may be more appropriate. Compared to the normal distribution, the t-distribution has heavier tails and may be more suitable for producing values that fall far from its mean. For testing the robustness of the normal prior specifications in our models, we modelled all the fixed effects parameters using a t-distribution; for example, we used $\beta \sim t(\mu_\beta, \tau_\beta, df_\beta)$, where μ_β denotes the location parameter, τ_β is the precision parameter and df_β are the degrees of freedom. For specifying prior distributions for μ_β, τ_β and df_β, we used the guidelines defined in Gelman and Hill (Gelman, A. and Hill, J. 2007); hence, we used a normal distribution with large variance for the location parameter, i.e. $\mu_\beta \sim N(0,100)$, for the precision parameter we used a prior distribution of the form $\tau_\beta = 1/U^2$ with $U \sim U(0,100)$ and for the degrees of freedom we used $df_\beta = 1/df\_inv$ with $df\_inv \sim U(0,.5)$. Similar specifications were used for the other fixed effects a, δ and γ. Furthermore, we also modelled the random effects using a multivariate t-distribution as specified above, with the distinction that we used a Wishart prior distribution for modelling the variance-covariance matrix. *Table 4-4* compares DIC results of models under these different prior specifications. Based on the DIC values illustrated in *Table 4-4*, we note that prior specifications using the t-distribution do not seem to produce better models than those using the normal distribution (differences of 3-4 DIC units between models using t-distributed priors and those using normal distributed priors),. Hence, we opted to use the normal and multivariate normal distributions for specifying priors for the unknown parameters in our models.

Table 4-4: DIC for models with different prior specifications

Model	Priors for fixed effects	*Priors for random Effects*	*DIC*
$\text{logit}(\mu_{ij})=\delta_0+\delta_1 g_i+\beta_1 a_i+b_{j0}+b_{j1}\times t_{ij}$ $\log(\varphi_{ij})=\gamma_0+\alpha_1 a_i+c_{j0}+c_{j1}\times t_{ij}$	Normal distribution	Multivariate normal distribution	-1335
	t-distribution	Multivariate normal distribution	-1339
	t-distribution	Multivariate t-distribution	-1332
$\text{logit}(\mu_{ij})=\delta_0+\delta_1 g_i+\beta_1 a_i+\beta_2 TTD_i+b_{j0}+b_{j1}\times t_{ij}$ $\log(\varphi_{ij})=\gamma_0+\alpha_1 TTD_i+c_{j0}+c_{j1}\times t_{ij}$	Normal distribution	Multivariate normal distribution	-1310
	t-distribution	Multivariate normal distribution	-1313
	t-distribution	Multivariate t-distribution	-1311

Model extensions for multiple bias adjustments

We performed sensitivity analyses for assessing the uncertainty caused by various sources of bias, including non-ignorable missing data and censoring. These various sources of bias were investigated thoroughly as suggested by Professor Greenland (Greenland 2005) by extending the baseline models with sub-models that include the bias parameters. Such extensions can be incorporated straightforwardly by performing a full Bayesian analysis as in (Greenland 2009). In our analysis we model both separately and concomitantly three potential sources of bias: that due to non-ignorable missing response data, that due to censoring, and that due to small sample size. Note that when performing these sensitivity analyses, we used different samples of our dataset:

To model the selection bias due to missing response data, we used the entire sample of deceased individuals aged 50+; that is, all individuals that did or did not respond to the HRQoL questionnaire. This sample includes 453 individuals (and 1142 observations). Of these individuals, 75% had at least one missing measurement and about 21% had all HRQoL measurements missing.

To model the selection bias due to censoring, we used the sample of deceased and survivors aged 65+, consisting of 1158 survivors and 156 deceased (and a total of 2165 observations). We disregarded the cases with missing response data. Using the sample of individuals aged 65+ instead of those aged 50+ was necessary in this situation in order to achieve model convergence in WinBUGS. Due to increased model complexity and additional data, even in this situation model results were obtained after running one such model for approximately 20 hours.

To model the selection bias due to both missing response data and censoring, we used the sample of deceased and survivors aged 65+, including all individuals that did or did not respond to the HRQoL questionnaire. This sample includes 2020 individuals (and 3803 observations), of which 1761 were survivors and 259 deceased. For deceased and survivors, approximately 60% and 53% of individuals had at least one HRQoL measurement missing, while about 40% and 34% had all HRQoL measurements missing, respectively.

For brevity, we present the model developments in detail for each of the above-described situations for the TTD model (the age-specific model being obtained by excluding the TTD variable from the model specification).

Selection bias due to missing response data

The results obtained with the baseline models are valid under the missing at random (MAR) assumption for the HRQoL response. While testing for the underlying missing-data mechanism is impossible with the data at hand, a non-ignorable missing-data mechanism or departures from MAR assumption may introduce bias into the analyses. In fact, it has been observed that, in general, sicker individuals tend not to respond to questionnaires (Ibrahim and Molenberghs 2009). Therefore, it may well be that the HRQoL missingness depends on the unobserved HRQoL values, which suggests a non-ignorable missing data mechanism (i.e. missing not at random (MNAR)). To investigate the extent to which the conclusions of our chapter would be affected by a potential non-ignorable missing data mechanism, we have adopted the full probability modelling approach (Mason et al. 2012).

Let us partition the HRQoL response denoted by y_{ij}(for subject j, observation i) into observed y_{ij}^{obs} and missing values y_{ij}^{miss} , i.e. $y_{ij}=(y_{ij}^{obs},y_{ij}^{miss})$. Furthermore, let $m=m_{ij}$ denote a missing data indicator such that:

$$m_{ij} = \begin{cases} 0, y_{ij} & \text{observed} \\ 1, y_{ij} & \text{missing} \end{cases} \tag{4.6}$$

Under MAR, estimating the missing response y_{ij}^{miss} is equivalent to posterior predictions from the model fitted to the data including the complete cases only. To model the MNAR mechanism, we extended the TTD baseline model (as indicated in Section 3.3) with a sub-model for the missingness mechanism that relates the probability of non-response, denoted by p_i, to both observed and unobserved variables in the dataset:

$$mi \sim Bernoulli(pi)$$
$$logit(p_i)=\delta_2+\beta_3*a_i+\delta_3*g_i+\beta_4*TTD_i+s(y_i) \quad (4.7)$$

where s denotes the function that links the model that analyses the response data to the non-response sub-model. The function s indicates how the distribution of the response among respondents relates to that among those that did not respond to the HRQoL questionnaire. Therefore, s quantifies the influence of the HRQoL outcome on the non-response. When $s=0$ the outcome is said to be MAR, indicating that the distribution of the missing data would be similar to that of the complete case data, whereas when, the outcome is said to be MNAR, indicating that two distributions would be different. We refer to s as to 'the selection bias function'. Obviously, many choices can be made for s. Some argue that, in the absence of prior knowledge, a linear relationship between the probability of missingness and the response may be assumed (Mason et al. 2012). Others indicate that since missing data is often more frequent among the sicker, a non-linear form for s may be more appropriate. Here, we use the non-linear function of the form $s(y_i)=\eta log(y_i)$, as proposed in (Scharfstein et al. 2003), with η denoting the selection bias parameter. The parameter η is interpreted as the log odds ratio of missingness between subjects who differ by one unit of $log(y)_i$. Hence, $\eta<0$ (>0), indicates that, compared to respondents, the distribution of the response among those who did not respond to the HRQoL questionnaire is heavily weighted towards the high values of y. Note that other parameterizations may be considered to reflect the nonlinearity, e.g. polynomial functions. When fitting the above models, we used non-informative priors for the fixed effects parameters. i.e. $\delta_2,\delta_3,\beta_3,\beta_4,\eta \sim N(0,40)$.

Selection bias due to censoring

The baseline models were fitted on the dataset for the deceased who ignores the subset of survivors and, therefore, ignores the bias possibly introduced by ignoring the censored nature of the TTD variable. Estimation of a model with censored regression may be approached as an estimation problem with missing data, as described in (Roderick J. A. Little 1992). In other words, censored values can be treated as missing. Here, we treated the censored TTD values as missing and we extended the baseline models with a sub-model that imputes the missing TTD values. Similarly to response missing, we adopted a full probabilistic Bayesian approach. We extended the baseline TTD model in such a way as to account for non-ignorable censoring; that is, for missing TTD that is MNAR, assuming that the TTD missingness depends on other observed covariates but also potentially on the unobserved TTD itself. The baseline model was extended with a sub-model for imputing the missing values of the TTD covariate.

We assume $TTD_i \sim Normal(\omega_i,\zeta)$. Here we can either assume vague priors for ω_i,ζ, or build a regression model relating TTD to other covariates such as age and gender and

assume vague priors only for the variance parameter ζ. Since mortality depends on age and gender, we modelled ω_i as a function of age and gender:

$$\omega i=\delta 4+\beta 5*ai+\delta 5*gi \quad (4.8)$$

and we used vague priors for the variance parameter $\zeta=1/U^2, U\sim U(0,100)$ and for the parameters $\delta_4,\delta_5, \beta_5\sim N(0,100)$.

Sub-model for the TTD missingness mechanism

Let $Cens_i$ denote a censoring indicator, such that:

$$Cens_i = \begin{cases} 1, y_{ij} & observed \\ 0, y_{ij} & censored \end{cases} \quad (4.9)$$

$$Censi\sim Bernoulli(q_i)$$
$$logit(q_i)=\delta_6+\beta_6*a_i+\delta_7*g_i+Q(TTD_i) \quad (4.10)$$

where q_i denotes the probability of observing TTD and Q indicates the selection bias function that links the missing TTD (i.e. the censoring mechanism) with the analysis model. Here, we assume a linear function for Q, i.e. $Q(TTD_i)=\beta_7*TTD_i$. Nevertheless, other forms (e.g. nonlinear functions) may be investigated. However, a model with additional complexities would be troublesome to estimate in this case. Note that, when $\beta_7=0$, the missing TTD covariate is MAR while for $\beta_7\neq 0$, the missing covariate is MNAR. We used vague priors for the parameters $\delta_6,\delta_7,\beta_6, \beta_7,\sim N(0,40)$.

Informative priors

A recent chapter that investigated the relationship between TTD and HRQoL reported some model parameter estimates that can be used in this analysis in the form of informative priors as shown in chapter 5 of this thesis. In particular, coefficient estimates of age (β_1)and TTD (β_2)for the location parameters were reported from both an age-specific and a TTD model. It should be noted that the above chapter used a larger dataset (i.e. about 1600 individuals) that was cross-sectional. Furthermore, that chapter reports lower variance for the age coefficient than for the TTD coefficient, providing a more informative prior for the age coefficient. Therefore, although we performed analyses that incorporated these informative priors, the results should be treated with caution.

The following priors were used as reported in chapter 5: for the age-specific model, we used the prior $\beta_1\sim N(-0.008,0.001^2)$ and for the TTD model we used the

priors$\beta_1 \sim N(-0.008, 0.0018^2)$ for the age coefficient and $\beta_2 \sim N(0.06, 0.0316^2)$ for the TTD coefficient, respectively.

Multiple bias adjustments were estimated by combining the models described in Sections (a), (b), and (c). For example, a joint model that accounts for uncertainty caused by ignoring the informative missing data mechanism and by ignoring the TTD censoring mechanism was developed by combining the baseline model specification and the submodels from Sections (a) and (b).

Results

This section is structured in two parts. The first part presents results of the baseline models. The second part present results that show to what extent the findings indicated by the baseline models are robust to various sources of bias such as non-ignorable missing data and censoring.

Baseline models

Table 4-5 displays the mean posterior estimates together with 95% equal-tailed credible intervals of all parameters (except the random-effects parameters) from the age-specific and the TTD models. Note that the estimates of the location parameter μ are on the logit scale, whereas those of the precision parameter ϕ are on the log scale. *Table 4-4* also presents the Monte Carlo (MC) error, which measures the accuracy of our simulations, each estimated parameter having an MC error of less than 5%, as previously suggested in the literature (Spiegelhalter et al. 2003).

For the coefficients that model the location parameter μ, the negative age coefficient β_1 indicates that mean HRQoL decreases with age in both the age-specific and the TTD model. Although in both models the 95% credible interval for β_1 includes 0, the absolute value of the coefficient estimate in the TTD model is approximately 80% less than the value in the age-specific model. In other words, when TTD is included in the model, the strength of the age effect on mean HRQoL decreases considerably and becomes nearly negligible. Furthermore, the positive TTD coefficient β_2 indicates that the greater TTD is, the higher mean HRQoL becomes.

For the coefficients that model the precision parameter ϕ, the negative age coefficient α_1 in the age-specific model indicates that, for fixed values of μ, HRQoL variance increases with advancing age. By contrast, in the TTD model the positive TTD coefficient α_1 indicates that the greater the time to death is, the lower the HRQoL variance becomes.

With respect to the random intercepts and random slopes, for both μ and ϕ, similar variances of the intercepts (σ_1^2 and τ_1^2) and the slopes (σ_2^2 and τ_2^2) were observed in

Table 4-5: Posterior parameter estimates with 95% credible intervals

	Age-specific model				TTD model			
Posterior mean estimates	mean	95% credible interval		MC error	mean	95% credible interval		MC error
δ_0 (intercept)	1.280	1.206	1.354	0.00017	1.284	1.211	1.356	0.00029
β_1(age)	-0.009	-0.018	0.0002	0.00018	-0.002	-0.012	0.009	0.00004
δ_1 (gender) 0=men, 1=women	-0.201	-0.350	-0.052	0.00319	-0.229	-0.377	-0.082	0.00055
β_2(TTD)	-	-	-	-	0.019	0.010	0.036	0.00001
γ_0 (intercept)	3.887	3.539	4.277	0.00002	3.828	3.503	4.190	0.00190
a_1 (age)	-0.027	-0.070	0.015	0.00032	-	-	-	-
α_1(TTD)	-	-	-	-	0.060	-0.024	0.132	0.00003
σ_1^2	0.323	0.247	0.409	0.00107	0.307	0.234	0.390	0.00023
$\rho_{12}\sigma_1\sigma_2$	-0.017	-0.061	0.027	0.00009	-0.029	-0.074	0.016	0.00017
σ_2^2	0.075	0.046	0.116	0.00016	0.076	0.047	0.116	0.00012
τ_1^2	0.636	0.182	1.360	0.00010	0.559	0.156	1.236	0.00428
$\theta_{12}\tau_1^2\tau_2^2$	-0.096	-0.461	0.208	0.00007	-0.083	-0.432	0.200	0.00211
τ_2^2	0.354	0.105	0.882	0.00242	0.326	0.101	0.813	0.00271
ρ_{12}	-0.109	-0.374	0.174	0.00115	-0.187	-0.451	0.103	0.00109
θ_{12}	-0.188	-0.727	0.498	0.00222	-0.176	-0.734	0.516	0.00425

the age-specific model and the TTD model, respectively. Moreover, for both models, the correlation between the intercepts and the slopes has similar negative values.

To further explore the effect of TTD on HRQoL, we compare the mean HRQoL predictions from the age-specific and the TTD model. Our modelling approach enables us to obtain two types of HRQoL predictions; that is, predictions for individuals included in the dataset (which use both the fixed effects coefficients and the individual specific random effects) and predictions for new or average individuals (which use only the fixed effect coefficients). Note that, in fact, given the logit transformation, predictions for new individuals are not the same as average predictions or predictions for the average individual.

Figure 4-3 shows estimated mean HRQoL for all ages 50+ from the age-specific model, the TTD model with TTD fixed at 1 month, and the TTD model with TTD fixed at 16 years. We chose these values to cover the entire observed range of TTD in our dataset. We note that, for both men and women, when controlling for TTD, age has a nearly negli-

gible impact on predicting mean HRQoL. Indeed, in the TTD model, when TTD is fixed, mean HRQoL decreases by only approximately 0.01 between the ages of 50 and 80. On the other hand, when TTD decreases from 16 years to one month, the mean HRQoL decreases by approximately 0.07. Furthermore, in accordance with previous research findings, Figure 4-3 illustrates that women have a lower mean HRQoL than men.

We further focus on both types of HRQoL predictions from the TTD model: those for the

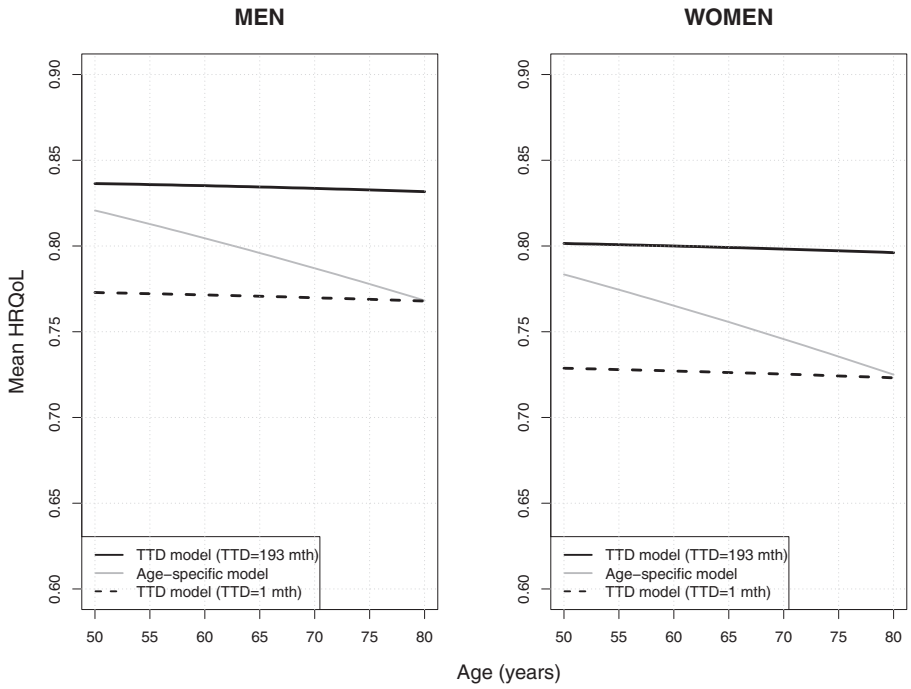

Figure 4-3: Mean HRQoL predictions from the age-specific and the TTD model (predictions for all ages 50+ when TTD is fixed at 1 and 193 months (mth) respectively

average individuals (i.e. new individuals) and new predictions for individuals in the dataset. The upper three graphs in Figure 4-4 show HRQoL trajectories before death for three individuals, including prediction intervals. The leftmost of these three graphs shows predictions for an average individual (a new individual) aged 80 at death, whereas the other two graphs include new predictions at one month before death for two subjects selected for having similar ages of death. Figure 4-4 (top-left) shows that mean HRQoL for a typical man who dies at age 80 decreases from 0.83 [0.79, 0.87] at 193 months before death to 0.77 [0.73, 0.81] at one month before death. Figure 4-4 (top: middle and right) presents different mean HRQoL predictions for the two selected individuals in the dataset who die at ages 80 and 75, respectively. Figure 4-4 shows that, although they are

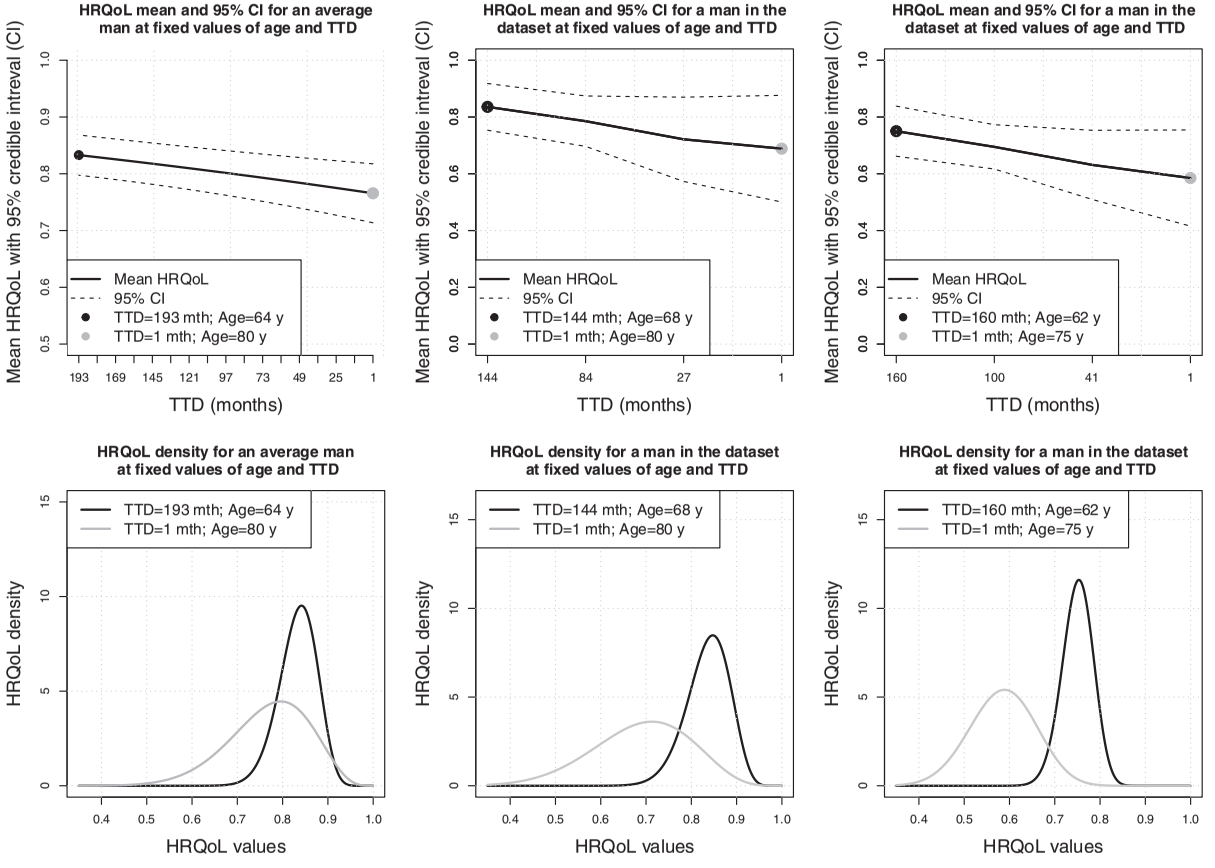

Figure 4-4: Top-Mean HRQoL trajectories (with 95% credible intervals (CI)) for fixed values of age measured in years (y) and time to death (TTD) measured in months (mth) for an average individual and new HRQoL predictions for two individuals in the dataset.; Bottom-HRQoL distributions at fixed values for age and TTD

of the same gender and die at similar ages, the two selected individuals have quite different HRQoL trajectories over time. Although decreasing mean HRQoL with decreasing TTD is evident in all cases, the slope is different for each individual whom is to be expected in a random intercept and random slope model. Furthermore, as these are new predictions for the individuals in the dataset use not only the fixed effects but also the random effects, they exhibit larger variance than those for the average individuals. In Figure 4-4 (bottom), for each subject we investigated changes in the posterior HRQoL distribution as death approaches. Note that the HRQoL distribution is determined not only by the mean parameter but also by the precision parameter ϕ; hence, we used the mean values of both μ and ϕ to obtain the HRQoL distribution (i.e. $HRQoL \sim Beta(\mu\phi,(1-\mu)\phi)$). We compared the HRQoL distribution at the first observation time (for the subjects in our dataset) or at 193 months before death (for an average individual) with the estimated HRQoL distribution at one month before death. Figure 4-4 (bottom) shows that the decrease in mean HRQoL as death approaches is associated with important changes in the HRQoL distribution, which becomes increasingly left-shifted and shows substantially increased variation (reflected by decreased skewness and kurtosis). Hence, these results suggest that the lower the mean HRQoL, the greater the uncertainty.

Robustness to bias adjustments

Table 4-6 shows results for the age and TTD coefficients, obtained by accounting for various sources of bias, with those from the baseline model. In general, we observe that, compared to the baseline model, in the scenarios accounting for various sources of bias, the uncertainty of the main estimates for both the age and the TTD coefficients increases. The mean TTD estimate increases from 0.019[0.010, 0.036] in the baseline model to 0.052 [0.018, 0.086] in the model that accounts for bias caused by non-informative missing response and censoring. Nevertheless, in general, for most of the sensitivity analyses, the mean age-coefficient estimate decreases by about 50% when the TTD variable is included in the model. Due to small variance in the informative priors used, coefficient estimates from models that use these informative priors are more precise, especially those for age. Therefore, when estimating the age coefficient, the information provided by the prior dominates that provided by the dataset. For these reasons, in those situations the age coefficient decreases by about 20% when TTD is included in the model specification.

We conclude that in various scenarios that account for multiple sources of bias such as missing response and censoring, the relationship between age and HRQoL becomes less evident when TTD is included in the model specification. Therefore, the different sensitivity analyses conducted here indicate that, although greater uncertainty around the mean estimates is observed when accounting for various sources of bias, the main

conclusions of these analyses do not change. On the contrary, accounting for missing responses in the HRQoL data and for censoring of the TTD variable results in stronger evidence in favour of the TTD.

Table 4-6: Model estimates accounting for various sources of bias

	β1(coefficient age)						β2(coefficient TTD)		
	Age model			TTD model			TTD model		
Model	Mean	95% credible interval		Mean	95% credible interval		Mean	95% credible interval	
Baseline model [a]	-0.009	-0.018	0.0002	-0.002	-0.012	0.009	0.019	0.010	0.036
Model adjusting for selection bias due to missing response data [b]	-0.011	-0.024	-0.001	-0.006	-0.017	0.005	0.028	0.011	0.045
Model adjusting for selection bias due to censoring [c]	-0.020	-0.027	-0.013	-0.011	-0.021	0.000	0.032	0.006	0.059
Models using informative priors [a]	-0.008	-0.010	-0.006	-0.007	-0.010	-0.006	0.018	0.005	0.024
Models adjusting for selection bias due to missing response and censoring [d]	-0.042	-0.051	-0.033	-0.021	-0.037	-0.006	0.052	0.018	0.086
Models adjusting for selection bias due to missing response and censoring using informative priors [d]	-0.010	-0.011	-0.008	-0.008	-0.010	-0.006	0.051	0.038	0.066

[a] models fitted to the sample of deceased aged 50+ discarding missing HRQoL response
[b] models fitted to the sample of deceased aged 50+ including the missing HRQoL response
[c] models fitted to the sample of survivors and deceased aged 65+ discarding the missing HRQoL response
[d] models fitted to the sample of survivors and deceased aged 65+ including the missing HRQoL response

Discussion And Conclusions

This research suggests that HRQoL and TTD are strongly related and that HRQoL mainly depends on proximity to death rather than age. Our findings indicate that when TTD is included in the model specification, the effect of age on HRQoL becomes negligible. Hence, HRQoL losses induced by decreasing TTD are substantially larger than those induced by increasing age. Furthermore, these results were found robust to various sources of bias such as non-ignorable missing data, TTD censoring or small sample size. Our finding that TTD explains health losses better than age does is consistent with previous research showing that TTD explains healthcare expenditures (HCE) better than age does (Seshamani 2004, Seshamani 2004, Zweifel et al. 2004, Werblow et al. 2007); hence, to some extent our results were to be expected. Similarly to analyses of HCE, the

relationship between HRQoL and TTD can be used to improve forecasts of population health (van Baal and Wong 2012a), building on projections of future life expectancy (Oeppen and Vaupel 2002).

This chapter also contributes to the available methodologies for modelling HRQoL by building on recent research that has proposed beta regression for modelling the non-standard HRQoL data. We have demonstrated that the mixed beta regression estimated using the MCMC methods is useful in modelling longitudinal SF-6D HRQoL data. Compared to methods that use the normal distribution assumption, our approach yielded models with better predictive capabilities and enabled us to draw more precise conclusions when investigating the relationship between HRQoL and TTD. It should be noted that the HRQoL index is frequently used as an important element in cost-effectiveness analysis of clinical trial data. The approach proposed here may be employed in such settings; the Bayesian estimation could prove useful, especially given the relatively small sample sizes of such clinical trial data and the possibility of incorporating informative priors. Bayesian MCMC methods have been proposed for modelling the cost element over time in a cost-effectiveness analysis (Cooper et al. 2007). Other, non-Bayesian approaches previously proposed for modelling end-of-life costs are also interesting for modelling end-of-life HRQoL. For example, (Chan and Wang 2010) propose an approach for estimating end-of-life medical costs by including a systematic change point for the slope in the period preceding death. They exemplified that method with Medicare data in which complete trajectories of healthcare use were known. Their method is relevant to situations in which similar databases are available for HRQoL. To encourage economic evaluation modelers to use the mixed beta regression model for modelling HRQoL over time, we provide the WinBUGS code for some of our models in the Appendix.

This chapter has several limitations. Perhaps its major limitation is in relation to the dataset used for performing the main analyses. First, the ultimate sample of deceased individuals used in the main analyses of this chapter was relatively small and consisted of rather young individuals (whose mean age at death was lower than that reported by Statistics Netherlands). This may explain why, although – in terms of HRQoL losses – TTD has a considerably larger impact than age, those losses are rather small. We hypothesize that a follow-up observed for longer may show a larger effect of TTD on HRQoL. The effect of TTD can be expected to be stronger at more advanced ages, when a far greater number of people are close to death and experience poor health states. Despite this apparent limitation of our dataset, which would normally diminish the strength of the TTD effect on HRQoL, these two variables were found to be strongly related. That said, future research using larger longitudinal datasets with longer follow-up is recommended.

Second, the main analysis presented above was performed only on the sample of deceased individuals, which constituted approximately 10% of all the available data. This is in line with previous research chaptering the relationship between HCE and TTD (Zweifel et al. 1999, Seshamani 2004, Seshamani 2004). However, in doing so a large portion of our dataset was excluded from the main analyses presented above. To confirm the TTD effect on HRQoL, we performed the same analysis on the sample of both deceased and survivors aged 65+. In this way we assessed the uncertainty caused by ignoring the censored nature of the TTD variable. In other words, we explicitly modelled the selection bias due to censoring by treating the TTD variable as a covariate with missing data. Note that, in a different context, survival analysis techniques have been developed to accommodate for censoring (Chan and Wang 2010). However, in this context, treating censoring as a covariate with missing data is more applicable to our current Bayesian model. We found that the parameter of selection bias due to censoring was significant and had important implications for most model' estimates; nevertheless, in this case, the main conclusions regarding the relationship between age, TTD and HRQoL remained unchanged. Third, the dataset did not include institutionalized individuals. Unfortunately, information on the percentage of the Dutch population institutionalized prior to death was not available. However, as the closer individuals come to death, the stronger HRQoL and TTD are related, we suspect that if very ill individuals were to be taken into account, the effect of TTD on HRQoL would become even larger. However, individuals who are very close to death and have a low HRQoL are unlikely to be included in any survey, whether they are institutionalized or not.

Fourth, about half of individuals responded to only one measurement round, resulting in data imbalance. However, similarly to all random effect models, the mixed beta regression model accommodates to any degree of imbalance in the data. Furthermore between 35% and 65% of the HRQoL data was missing in the sample for deceased individuals aged 50+ used for the main analyses of this chapter. We have performed sensitivity analyses in order to account for the selection bias caused by ignoring the missing data mechanism. We found that the selection bias parameter was significant and negative, suggesting that the HRQoL response is MNAR and that the distribution among those who did not respond to the HRQoL questionnaire is more weighted towards the low values of HRQoL. Furthermore, these sensitivity analyses provided stronger evidence that TTD is more important than age; this is to be expected, since subjects in poorer health are closer to death and tend not to respond to health questionnaires.

Finally, in this chapter we used the SF-6D HRQoL index, which was observed to lie between 0.388 and 1. Fryback et al. have shown that for a national survey sample of non-institutionalized adults, both the range and the mean of several HRQoL indices (e.g.

HUI2, HUI3, EQ-5D, and SF-6D) differ significantly (Fryback et al. 2007): the minimum observed range for HRQoL was -0.34 (for HUI3) to 0.3 (for the SF-6D). These range discrepancies between various HRQoL indices suggest that our results may be sensitive to the HRQoL instrument used. For example, given that some HRQoL instruments (e.g. HUI, EQ-5D) have a larger range than SF-6D, which was used in our chapter, we suspect that the TTD effect would be even larger if the former instruments were to be used instead of the SF-6D. More generally, even though we used a multidimensional measure of health to quantify health losses, our findings cannot be generalized to all domains of health. Furthermore, for some consequences of population ageing it may be worthwhile to investigate the influence of TTD on specific health domains.

The present chapter also has several noteworthy strengths. First, although small, the longitudinal dataset enabled us to account for possible HRQoL trends for at least half of those deceased, while analysing the relationship between HRQoL and TTD, which is why such an analysis should be preferred to a cross-sectional one.

Second, due to the non-standard nature of the HRQoL distribution, improper modelling techniques may lead to invalid results when modelling the relationship between age, TTD and HRQoL. An important strength of this chapter is the use of a mixed beta regression estimated under the Bayesian paradigm for modelling the longitudinal HRQoL data. This approach allows the error term to be straightforwardly modelled as a function of various explanatory variables. In other words, heteroscedasticity, a common issue for any bounded outcome, can be explicitly addressed in a natural way. Here we modelled both parameters that describe the variance of the beta distribution, i.e. the location and the precision parameter, as functions of explanatory variables such as age or TTD. Plots of the residuals against TTD and other covariates indicated that heteroscedasticity was not a problem in our models. Another advantage of our approach is the possibility of estimating complex models that account for the uncertainty caused by various potential sources of bias such as non-ignorable missing data and TTD censoring. Note that in another context, for handling irregular distributions and a ceiling effect, other non-Bayesian estimations for semiparametric extensions of generalized linear models (GLM) have been previously proposed (Luo and Tsai 2012, Chan 2013).

The model we proposed here can be extended in various ways. For example, one can use non-linear terms such as splines to model the relationship between some of the explanatory variables (e.g. age) and HRQoL. Given the approximately linear patterns observed in our data, this was not the case in the present chapter; however, this can be easily implemented using WinBUGS. Furthermore, in this chapter the number of measurements that attained the maximum observed value of 1 was approximately 1.9%;

hence, we opted to transform the boundary point at 1 to a slightly smaller value (Smithson and Verkuilen 2006). We conducted sensitivity analyses to investigate to what extent the chosen cut-off point would affect our findings. For this purpose, we ran models with the cut-off fixed at 0.9 and 0.999. We found that moving the cut-off point to these values does not seem to have a strong impact on the results, although this is probably due to the small number of observations that were assigned the cut-off value in the present dataset. Note that this aspect may have more serious consequences, especially when a larger number of observations are subject to cut-off. In that case, a two-part model specification may be more appropriate (Mullahy 1986)." Another notable advantage of the Bayesian approach is the possibility of including informative prior distributions, which would enable analysts to incorporate multiple sources of evidence in a single model. This is potentially important when HRQoL data from multiple trials is available. Using coefficient estimates from chapter 5, we have exemplified the use of informative priors in our analyses.

This chapter has two important implications. In terms of method, this chapter showed that mixed beta regression models and Bayesian estimation provide flexibility in modelling longitudinal HRQoL data. From a policy perspective, our results may have important implications for policies aimed at dealing with population ageing. This research represents a first step in showing that increases in life expectancy may not necessarily translate into additional years lived in poor health, as health losses are postponed towards the end of life (Fries 1980). This suggests that the demand for health care may increase less markedly than would otherwise be expected. Furthermore, if health losses are mainly concentrated in the final phase of life, linking retirement age to life expectancy seems to be a viable policy option. Hence, better projections of future population health continue to be highly important from both a scientific and a societal perspective.

In summary, using HRQoL as a measurement of health, we found that health losses are centred in the final phase of life. This has important implications for projecting future population health, as well as for policies aimed at reducing the societal consequences of ageing populations.

Appendices

A1: The SF-6D questionnaire

Table 4-7: The domains of the SF-6D questionnaire, source (Brazier, Roberts, et al., 2002)

Level	Physical functioning	Role limitations	Social Functioning	Pain	Mental Health	Vitality
1	Your health does not limit you in vigorous activities	You have no problem with your work or other regular daily activities as a result of your physical health or any emotional problems	Your health limits your social activities none of the time	You have no pain	You feel tense or downhearted and low none of the time	You have a lot of energy all of the time
2	Your health limits you a little in vigorous activities	You are limited in the kind of work or other regular daily activities as a result of your physical health	Your health limits your social activities a little of the time	You have pain but it does not interfere with your normal work (both outside of home and housework) a little bit	You feel tense or downhearted and low a little of the time	You have a lot of energy most of the time
3	Your health limits you a lot in vigorous activities	You accomplish less than you would like as a result of emotional problems	Your health limits your social activities some of the time	You have pain but it does not interfere with your normal work (both outside of home and housework) moderately	You feel tense or downhearted and low some of the time	You have a lot of energy some of the time
4	Your health limits you a little in bathing and dressing	You are limited in the kind of work or other regular daily activities as a result of your physical health and accomplish less than you would like as a result of emotional problems	Your health limits your social activities most of the time	You have pain but it does not interfere with your normal work (both outside of home and housework) quite a bit	You feel tense or downhearted and low most of the time	You have a lot of energy a little of the time
5	Your health limits you a lot in bathing and dressing		Your health limits your social activities all of the time	You have pain but it does not interfere with your normal work (both outside of home and housework) extremly	You feel tense or downhearted and low all of the time	You have a lot of energy none of the time

A2: Comparison of mixed beta with linear mixed effects and linear model

Table 4-8 compares DIC for the models used in this chapter with the linear model and the linear mixed effects model, which are based on the normal distribution assumption. For comparative purposes, in this chapter we also considered linear and linear mixed effects models that use the logistic transformation of the HRQoL outcome under the normal distribution assumption. For comparative purposes, we estimated all of these models using WinBUGS, although these models can also be estimated using the classical maximum-likelihood-based approach and software such as R or STATA. In fact, running these models with non-informative priors in WinBUGS gave similar estimates as the R functions lm() and lme() for the linear model and for the linear mixed effects model, respectively. As these models use the normal distribution assumption, ϕ is not the precision parameter but the parameter that models the standard deviation often known as σ. Although, in the Bayesian context, σ can be modelled using a regression structure, for comparison with more typical approaches available in other statistical software we did not perform such analyses here (for example, this cannot be done with the packages available in R). *Table 4-8* indicates that our approach resulted in substantially smaller DIC than models based on normality distribution assumption, including models that use logistic transformation of the HRQoL data. *Table 4-8* compares the QQ plot of the TTD models referred to above. We observe that, compared to the QQ plots for the linear mixed effect model and for the linear model, the QQ plot for the mixed beta model displays an almost straight line. Moreover, although they are inferior in terms of their predictive capacity, the linear model and the linear mixed effects model that use the transformed HRQoL data display QQ plots similar to those of the mixed beta model. It should be noted that for Figure 4-5 we used the mean of the estimated residuals as obtained from WinBUGS.

Table 4-9 compares posterior estimates of these models for the fixed effects coefficients of the location parameter. We observe that, in all models, the 95% credible interval for the age coefficients (β_1) includes 0 and that the absolute value of the age coefficient (β_1) decreases considerably when including TTD in the estimation. *Table 4-9* shows that, by accounting for TTD, the age coefficient decreases by approximately 78% in the mixed beta regression approach, 88% in the linear model and 91% in the linear mixed effects model. The models using transformed data indicate that by accounting for TTD, the age coefficient decreases by approximately 70% in the linear mixed effects model and 75% in the linear model.

Table 4-8: DIC values

	Nr	Model	Model specification	*DIC*
Age-specific model	(1)	$y_{ij}\sim N(\mu_i,\varphi)$	$\mu_{ij}=\delta_0+\delta_1 g_i+\beta_1 a_i$ $\varphi=1/U^2, U\sim U(0,100)$	-770
	(2)	$y_{ij}\sim N(\mu_i,\varphi)$	$\text{logit}(\mu_{ij})=\delta_0+\delta_1 g_i+\beta_1 a_i$ $\varphi=1/U^2, U\sim U(0,100)$	-771
	(3)	$y_{ij}\sim N(\mu_{ij},\varphi)$	$\mu_{ij}=\delta_0+\delta_1 g_i+\beta_1 a_i+b_{j0}+b_{j1}\times t_{ij}$ $\varphi=1/U^2, U\sim U(0,100)$	-1179
	(4)	$y_{ij}\sim N(\mu_{ij},\varphi)$	$\text{logit}(\mu_{ij})=\delta_0+\delta_1 g_i+\beta_1 a_i+b_{j0}+b_{j1}\times t_{ij}$ $\varphi=1/U^2, U\sim U(0,100)$	-1197
	(5)	$y_{ij}\sim Beta(\mu_{ij}\varphi_{ij},(1-\mu_{ij})\varphi_{ij})$	$\text{logit}(\mu_{ij})=\delta_0+\delta_1 g_i+\beta_1 a_i+b_{j0}+b_{j1}\times t_{ij}$ $\log(\varphi_{ij})=\gamma_0+\alpha_1 a_i+c_{j0}+c_{j1}\times t_{ij}$	-1350
TTD model	(6)	$y_{ij}\sim N(\mu_i,\varphi)$	$\mu_{ij}=\delta_0+\delta_1 g_i+\beta_1 a_i+\beta_2 TTD_i+b_{j0}+b_{j1}\times t_{ij}$ $\varphi=1/U^2, U\sim U(0,100)$	-776
	(7)	$y_{ij}\sim N(\mu_i,\varphi)$	$\text{logit}(\mu_{ij})=\delta_0+\delta_1 g_i+\beta_1 a_i+\beta_2 TTD_i+b_{j0}+b_{j1}\times t_{ij}$ $\varphi=1/U^2, U\sim U(0,100)$	-776
	(8)	$y_{ij}\sim N(\mu_{ij},\varphi)$	$\mu_{ij}=\delta_0+\delta_1 g_i+\beta_1 a_i+\beta_2 TTD_i+b_{j0}+b_{j1}\times t_{ij}$ $\varphi=1/U^2, U\sim U(0,100)$	-1185
	(9)	$y_{ij}\sim N(\mu_{ij},\varphi)$	$\text{logit}(\mu_{ij})=\delta_0+\delta_1 g_i+\beta_1 a_i+\beta_2 TTD_i+b_{j0}+b_{j1}\times t_{ij}$ $\varphi=1/U^2, U\sim U(0,100)$	-1203
	(10)	$y_{ij}\sim Beta(\mu_{ij}\varphi_{ij},(1-\mu_{ij})\varphi_{ij})$	$\text{logit}(\mu_{ij})=\delta_0+\delta_1 g_i+\beta_1 a_i+\beta_2 TTD_i+b_{j0}+b_{j1}\times t_{ij}$ $\log(\varphi_{ij})=\gamma_0+\alpha_1 TTD_i+c_{j0}+c_{j1}\times t_{ij}$	-1310

** a denotes calendar age. g denotes gender and t denotes observation time.*

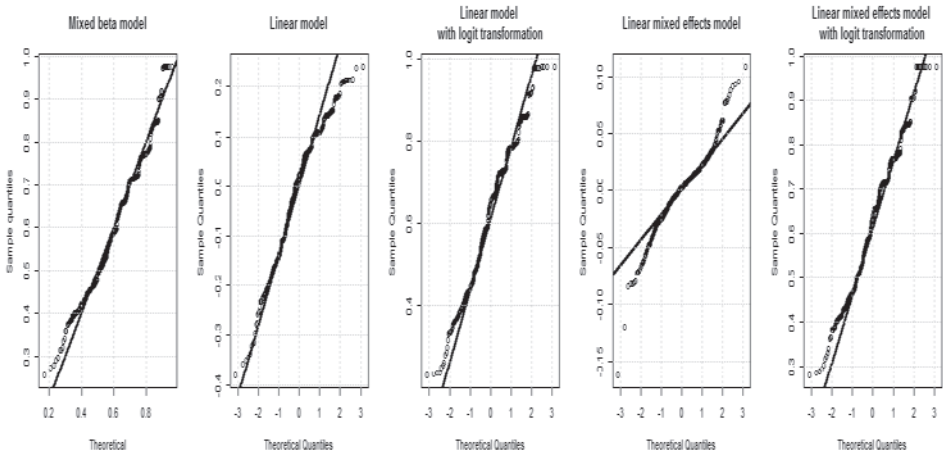

Figure 4-5: QQ Plots for various models

Table 4-9: Comparison of posterior estimates for the fixed effects coefficients

Model	Coefficient	Age-specific model				TTD model			
		mean	95% credible interval		MC error	mean	95% credible interval		MC error
Linear model	δ_0 (intercept)	0.767	0.757	0.777	0.00003	0.767	0.757	0.777	0.00002
	δ_1 (gender)	-0.035	-0.056	-0.014	0.00006	-0,037	-0,058	0,016	0.00006
	β_1(age)	-0.0005	-0.0021	0.0008	0.00004	-0.00006	-0.0015	0,0014	0.00004
	β_2(TTD)	-	-	-	-	0,0024	0,00084	0,0048	0.000006
Linear mixed effects	δ_0 (intercept)	0,768	0,754	0,783	0,00006	0,769	0,754	0,784	0.00005
	δ_1 (gender)	-0,04	0,07	0,01	0,00012	-0,42	-0,07	0,01	0.0001
	β_1(age)	-0,00098	-0,003	0,001	0,000008	-0,00009	-0,002	0,002	0.000009
	β_2(TTD)	-	-	-	-	0,0036	0,0048	0,0006	1E-07
Linear model with logit transformation	δ_0 (intercept)	1.198	1.142	1.256	0.00017	1.2	1.144	1.258	0.000017
	δ_1 (gender)	-0.199	-0.313	-0.083	0.0003	0.210	-0.326	-0.095	0.00003
	β_1(age)	-0.004	-0.012	0.004	0.00002	-0.001	-0.010	0.004	0.00002
	β_2(TTD)	-	-	-	-	0.0192	0.0048	0.0336	0.00006
Linear mixed effects model with logit transformation	δ_0 (intercept)	1.295	1.22	1.371	0.0004	1.301	1.227	1.376	0.0002
	δ_1 (gender)	-0.208	-0.360	-0.057	0.0008	-0.223	-0.374	0.073	0.0006
	β_1(age)	-0.01	-0.012	0.00005	0.00005	-0.003	-0.014	0.004	0.0006
	β_2(TTD)	-	-	-	-	0.020	0.006	0.036	0.0004
Mixed beta	δ_0 (intercept)	1.280	1.206	1.354	0.00017	1.284	1.211	1.356	0.00029
	δ_1 (gender)	-0.201	-0.350	-0.052	0.00319	-0.229	-0.377	-0.082	0.00055
	β_2(age)	-0.009	-0.018	0.0002	0.00018	-0.002	-0.012	0.009	0.00004
	β_2(TTD)	-	-	-	-	0.019	0.010	0.036	0.00001

Chapter 5

Quality of life and time to death: Have the Health Gains of Preventive Interventions Been Underestimated

With Werner Brouwer and Pieter Van Baal

Medical Decision Making 2015; 35(3):316-327

Abstract

This chapter explores the implications of the relation between quality of life (QoL) and time to death (TTD) for economic evaluations of preventive interventions. By using health survey data on QoL for the general Dutch population linked to the mortality registry, we quantify the magnitude of this relationship. For addressing specific features of the non-standard QoL distribution such as boundness, skewness and heteroscedasticity, we modeled QoL using a generalized additive model for location, scale and shape (GAMLSS) with a beta inflated outcome distribution. Our empirical results indicate that QoL decreases when approaching death suggesting that there is a strong relationship between TTD and QoL. Predictions of different regression models revealed that ignoring this relationship results in an underestimation of the QALY gains for preventive interventions. The underestimation ranged between 3% and 7% and depended on age, the number of years gained from the intervention and the discount rate used.

Introduction

Modelling techniques are frequently applied in economic evaluations of life prolonging preventive interventions in order to estimate the effects of these interventions in terms of quality-adjusted life years (QALYs) gains. Examples of such preventive interventions are smoking cessation (Hoogenveen et al. 2008), weight loss (van Baal et al. 2008), screening (Heijnsdijk et al. 2012), vaccination (de Kok et al. 2009), medical treatment of risk factors (Mihaylova et al. 2006). By using modelling, an intermediate effect (for instance: weight loss, newly detected cases through screening, number of smokers quitted) is connected to causally related events (such as the incidence of diseases and/or death) that cannot be observed within the trial period of the intervention because the follow-up period is often too short (Buxton et al. 1997). Although modelling is a powerful tool to estimate the benefits of preventive interventions by synthetizing evidence from different sources, caution must be taken in estimating QALYs gained when preventive interventions are assumed to extend length of life. In that case, it is important to go beyond quality of life (QoL) losses due to the disease of interest. If an intervention adds years to the life of an individual, he/she becomes exposed to other diseases that may also result in quality of life losses. Recent research illustrates the relevance of this issue as the estimated health benefits of prostate cancer screening depended crucially on QoL in life years gained (Heijnsdijk et al. 2012). In that chapter, the authors assumed various utility values such as 1, 0.95 and 0.93 in life years gained. These different utility values resulted in 72, 56 and 6 QALYs gained, respectively, in a population of 1000 men aged 55+ at time of screening. However, these values can be considered high given that the added life years as a result of prostate cancer screening are spent at an older age. For instance, Fryback and colleagues showed that, an average American man in the age range 55-64 has a mean EQ-5D utility of 0.86, a mean HUI3 utility of 0.78 and a mean SF-6D index of 0.79 (Fryback et al. 2007) . If these values were used in the analyses by Heijnsdijk et al. it is likely that prostate cancer screening would result in QALYs lost rather than QALYs gained. Therefore, it matters which assumptions are made regarding the quality of life people will experience during life years gained.

Many economic evaluations of life prolonging interventions choose to ignore the impact of competing diseases on quality of life altogether and account only for the QoL losses due to the disease under chapter (de Kok et al. 2009). The implicit assumption then is that if an intervention prolongs life of individuals, they experience perfect health in all gained years. In other words, the absence of the disease under study translates into perfect health. However, to assume that, for example, a woman whose death is postponed due to HPV vaccination (de Kok et al. 2009) will experience no quality of life losses in added life years and hence remains in perfect health until death is unrealistic. Indeed,

empirical studies generally found quality of life to decrease with age (Fryback et al. 2007, Heijink et al. 2011, Fryback and Laurence 1997). Some economic evaluations incorporate these empirical findings in their models and assume that QoL in life years gained equals the age-specific mean in the population (Anonychuk et al. 2009, Schousboe et al. 2011, Tosteson et al. 2008). Although this method may lead to more realistic estimates of QALY gains of preventive interventions, we argue that estimates can be improved by making use of the relation between QoL and time to death (TTD).

TTD has been investigated mainly in relation to health care expenditures (HCE) (Zweifel et al. 1999, Seshamani 2004, Seshamani 2004, Zweifel et al. 2004, Werblow et al. 2007, Wanless 2004, Polder et al. 2006, Stearns and Norton 2004)) but recently also with disability (Klijs et al. 2011). The first and probably most influential chapter on TTD and HCE was published by Zweifel and colleagues (Zweifel et al. 1999). This chapter analysed the relationship between age and HCE using Swiss sick fund data, and found that the magnitude of HCE is explained to a greater extent by TTD rather than age. Therefore, higher average health care costs at higher ages are mainly caused by the fact that more people die at higher ages and the period before dying is associated with high healthcare use. This implies that an increase in life expectancy postpones the expensive last period of life, which suggested that aging of the population per se might have a more limited impact on HCE than generally believed. More importantly, it has been highlighted that cost-effectiveness analyses overestimate the costs and consequently the incremental cost-effectiveness ratio (ICER) of life prolonging preventive interventions when the relation between TTD and HCE is not explicitly modelled (Gandjour and Lauterbach 2005). Although the impact of TTD on QoL has never been studied before, many studies showed that QoL is a good predictor for survival in persons with a chronic disease (Fan et al. 2002). Furthermore, in the general population it was found that a lower QoL is associated with a higher mortality risk (Kaplan et al. 2007).

The aim of this chapter is to a) quantify the relation between QoL and TTD in the general population and b) show the relevance of this relation for the economic evaluation of preventive interventions. We hypothesize that a part of the decrease in QoL of the elderly is the result of higher mortality risks of the elderly, which are accompanied by lower QoL values. Previous research has shown that population average QoL decreases with age (Fryback et al. 2007). We argue that the observed relationship between QoL and age is in fact a relationship between increasing age-specific mortality and low QoL associated with the period close to death. For example, population average QoL at age 80 may be lower than that at age 60 because there are many more individuals in their last year of life at age 80 than at age 60. Preventive interventions postponing death therefore to some extent also postpone the losses in QoL. Hence, modelling QoL

values exclusively depending on age will suggest that life years gained by a preventive intervention are spent in poorer health than actually will be the case. In other words, if QoL values correlate with mortality and depend strongly on TTD, postponement of death will result in postponement of QoL losses and only the last years of life will be spend in poor health. We hypothesize that not accounting for this mechanism will result in an underestimation of QALY gains of preventive interventions. As a consequence, this leads to an overestimation of the cost-effectiveness ratio. Here, we will demonstrate this empirically and compare our proposed approach (i.e., projecting age-specific QoL estimates stratified by TTD for all years from the intervention until death) with the best used technique up to date in the practice of economic evaluations (i.e. using age-specific QoL estimates in life years gained).

Our empirical results confirmed that ignoring the relation between QoL and TTD results in an underestimation of QALYs gained for preventive interventions that extend life. The level of this underestimation ranges between 3% and 7% and depends mostly on the discount rate used but also on variables such as age and the number of years gained due to a preventive intervention.

Methods

Data

The present chapter was based on the Permanent Survey of Living Conditions (POLS: *Permanent Onderzoek Leef Situatie*) for years 2001-2008 that was linked to mortality registry. POLS is an on-going yearly cross-sectional survey. It started in 1981 and is coordinated by Statistics Netherlands. The survey is sampled on records from a centralized municipal registry, and does not include the institutionalized population. The POLS health survey monitors developments in lifestyle, health, medical consumption, preventive behaviour, and well-being in the Netherlands and starting from 2001 it includes the SF-12 questionnaire. The Health Module of the survey is collected both in a face-to-face interview and a written questionnaire. The interviewer visits the participants at home, asks for informed consent, conducts an interview and leaves a written (drop-off) questionnaire that includes the SF-12 questionnaire. Not everyone who completed the interview returned the written questionnaire, so approximately 20-25% of the SF-12 items were missing. We analysed a complete data set obtained by deleting the records corresponding to the missing fields of the SF-12. The SF-6D was derived from the SF-12 using the algorithm developed by (Brazier and Roberts 2004). QoL scores such as the SF-6D are widely used in clinical trials and cost-effectiveness analysis for measuring QALYs (Drummond et al. 2005).

The POLS sample used in this chapter includes individuals from years 2001-2008 who were linked to the mortality registry and were followed-up for approximately 10 years, i.e. for the period 2001-2010. In the POLS data, for each individual in 2001-2008 there is only one measurement of the SF-12. We used a POLS sample of individuals aged 50+. *Table 5-1* below presents descriptive statistics of the POLS dataset. Within the POLS survey, for ages 50+, in 2001-2010, 1633 persons (952 men and 681 women) have deceased and 17664 individuals (8726 men and 8938 women) have survived. From the deceased, 426 (45%) men and 264 (38.7%) women died within three years of the measurement. For the deceased, TTD will be defined as the length of time from the SF-12 measurement until death. *Table 5-1* shows that both mean TTD and mean age at death is larger for women than for men. This is consistent with the observation knowledge that Dutch women have lower mortality rates than the Dutch men (Statistics Netherlands. 2011). Furthermore, *Table 5-1* shows that the mean SF-6D QoL is lower for the group of deceased compared to that of survivors: we observe differences of approximately 0.06 for men and 0.09 for women. In addition, *Table 5-1* shows that the percentage of individuals that were assigned a utility value of 1 (indicating perfect health) is larger for the groups than have in general higher QoL scores, i.e. for the survivors compared to the deceased and for men compared to women.

Table 5-1: Descriptive statistics of the POLS dataset

Mean ± S.E.	*Men*	*Women*	*Both*
Number deceased	*952*	*681*	*1633*
Number survivors	*8726*	*8938*	*17664*
SF-6D QoL deceased	*0.76±0.01*	*0.70±0.01*	*0.74±0.007*
SF-6D QoL survivors	*0.82±0.002*	*0.79±0.002*	*0.8±0.002*
Mean TTD (months)	*44.6±1.7*	*47.8±2*	*46±1.33*
Mean age at death(years)	*71.6±0.6*	*73±0.7*	*72.2±0.46*
% deceased in full health	*5.7%*	*3%*	*4.5%*
% survivors in full health	*8.9%*	*5.4%*	*7.2%*

The relation between QoL and mortality is usually investigated with survival analysis with QoL entered as a predictor variable (Kaplan et al. 2007). Since in that case the outcome is mortality, it would not be possible to produce QoL estimates stratified by TTD. Therefore, in this chapter, regression models will be fitted using the SF-6D QoL as an outcome variable and age, gender, TTD and interactions between these variables as predictor variables. In the model specification we will deliberately exclude disease indicators. This is because just as age, TTD is a proxy variable that borrows its explanatory power in a statistical model from different determinants of QoL of which diseases processes are the most important ones. The more adjustments for various diseases used in the statistical analysis, the less the variables age and TTD are expected to matter.

QoL scores are typically difficult to model due to the non-standard nature of their distribution. Major concerns when modelling QoL are: (a) QoL values are bounded (generally defined between 0 and 1); (b) the QoL distribution is strongly negative (left) skewed; (c) the distribution is heteroscedastic, i.e. the constant variance assumption is violated because it has been shown that QoL variance usually increases as mean QoL decreases (Basu and Manca 2012). For addressing these problems when modelling QoL, various approaches have been proposed in the literature including the classical Tobit model and robust variants of the classical Tobit models such as symmetrically least squares models and censored least absolute deviations models (CLAD) (Austin 2002). However, a simulation chapter that compared the performance of these models with ordinary least squares (OLS) in terms of bias and confidence intervals found that both Tobit and its robust variants models were biased and recommend the use of OLS with robust standard errors for modelling the QoL scores (Pullenayegum et al. 2010). Furthermore, more recent studies found that beta regression models outperform OLS in terms of fit and efficiency when modelling both cross-sectional QoL data (Hunger et al. 2011) and longitudinal QoL data (Hunger et al. 2012). In fact, due to its flexibility in modelling highly skewed data, statisticians proposed the use of beta regression for modelling distributions with similar characteristics as the QoL scores (Kieschnick and McCullough 2003). Because beta regression is defined in the open interval (0,1) and the QoL data has measured utilities at one we used a beta inflated distribution at one (Ospina 2010) in which the inflation parameter models the probability mass at one. The estimation of this model was facilitated by using the GAMLSS approach (R. A. Rigby, et al. 2010, Stasinopoulos and Rigby 2007, Rigby and Stasinopoulos 2005, Rigby and Stasinopoulos 2010). A major advantage of GAMLSS models is that they allow modelling not only the mean or location parameter but also the parameter involving the dispersion of the QoL distribution thus offering an explicit and natural way for modelling the heteroscedasticity in the QoL data.

Generalized Additive Models for Location, Scale and Shape (GAMLSS)

Because the SF-6D index is a continuous variable defined on the interval (0, 1] we used the beta inflated at one *BEINF1* distribution assumption to model the QoL using the GAMLSS approach (R. A. Rigby, et al. 2010, Stasinopoulos and Rigby 2007, Rigby and Stasinopoulos 2005, Rigby and Stasinopoulos 2010). *BEINF1* is a special case of the class of inflated models. The word *inflation* is used to indicate that the probability mass is exceeded at the boundary of a certain parametric distribution; in this case, that of the beta distribution. *Inflation* can be associated with any parametric distribution. Therefore, *BEINF1* is a mixture of a continuous beta distribution defined on the interval (0,1) and a degenerate distribution, which gives non-negative probabilities at 1. *BEINF1* is defined by three parameters: the location parameter denoted by μ, the dispersion parameter denoted by σ and the parameter that models the probability mass at one denoted by ν.

This approach is in fact similar with a two part model in which the first part is modelled through the inflation parameter ν and the second part is modelled through parameters μ and σ with a beta distribution. Generally, various shapes of the beta distribution can be obtained for various values of the parameters μ and σ. Using the parameterization implemented in GAMLSS, the probability density function of a *BEINF1*(μ,σ,ν) is:

$$f_Y(y|\mu,\sigma,\nu)=\begin{cases}(1-p_1)f(y|\mu,\sigma), & 0<y<1\\ p_1, & y=1,\end{cases} \tag{5.1}$$

where p is the probability mass at 1 and represents the probability of observing 1 and $f(y\mu,\sigma)$ is the beta distribution defined by:

$$f\left(y|\mu,\sigma\right)=\frac{1}{B\left(\alpha,\beta\right)}y^{\alpha-1}\left(1-y\right)^{\beta-1}, \tag{5.2}$$

where the relation between the parameters $(\mu,\sigma)$and (α,β) is given by $\alpha=\frac{\mu\left(1-\sigma^2\right)}{\sigma^2}$, $\beta=\frac{\left(1-\mu\right)\left(1-\sigma^2\right)}{\sigma^2}$ and $B(\alpha,\beta)$ is the beta function. Details regarding the parameterization of the *BEINF1* in GAMLSS can be found elsewhere (R. A. Rigby, et al. 2010). If $Y\sim BEINF1(\mu,\sigma,\nu)$ then $E[Y]=\frac{\mu+\nu}{1+\nu}$ and $Var[Y]=\sigma^2\mu(1-\mu)$. Therefore, the inflation parameter influences the mean while the variance is described by the dispersion and the mean parameters. Note that the variance is not determined solely by the dispersion parameter; also the mean parameter has an influence.

Model selection

Model selection in GAMLSS has been performed by minimizing the generalized Akaike information criterion (GAIC) for different penalties: $GAIC=-2\times l(\theta)+df\times\#$, where $-2\times l(\theta)$ is the fitted deviance, df denotes the total degrees of freedom and # is the penalty for each degree of freedom used in the model. For example, when $\#=2$, the original Akaike information criterion (AIC) is obtained and when $\#=\log(n)$ with *n* denoting the sample size, the Schwarz Bayesian information Criterion (SBC) is obtained. As the original AIC may be too generous and the SBC may be too restrictive with respect to the number of optimal parameters selected in a model, we have selected our models using GAIC with various penalties, e.g. #=*2; 2.5; 3; 3.5; 4.* We developed two separate models: one model that includes variables age, gender and TTD and another model that includes only variables age and gender. Throughout this chapter, the first model will be called the TTD approach while the latter model will be named the age-specific approach. The TTD model is described by equations (5.3)-(5.5) and the age-specific model is shown by equations (5.6)-(5.8), respectively. For the model specification of each parameter, we fitted models that used terms such as age^2, TTD^2 and various interactions between variables age,

gender and TTD. However, in terms of GAIC, only the model that included TTD^2 for the mean parameter appeared to be superior; all the other models were equivalent with the models illustrated here. For brevity of exposition, we did not present these results here.

$$\text{logit}(\mu)=\alpha_{10}+\alpha_{11}\text{age}+\alpha_{12}\text{gender}+\alpha_{13}\text{TTD}+\alpha_{14}\text{TTD}^2 \quad (5.3)$$

$$\log(\sigma)=\beta_{10}+\beta_{11}\text{TTD} \quad (5.4)$$

$$\log(\nu)=\delta_{10}+\delta_{11}\text{gender} \quad (5.5)$$

and the age-specific model is:

$$\text{logit}(\mu)=\alpha_{20}+\alpha_{21}\text{age}+\alpha_{22}\text{gender} \quad (5.6)$$

$$\log(\sigma)=\beta_{20}+\beta_{21}\text{age} \quad (5.7)$$

$$\log(\nu)=\delta_{20}+\delta_{21}\text{gender} \quad (5.8)$$

Prevention, QoL and TTD

In order to assess the relevance of TTD when evaluating life prolonging preventive interventions we calculated QALYs gained of a hypothetical preventive intervention in different scenarios. In each scenario we calculated incremental QALY gains using predictions from the TTD model and compared these to incremental QALY gains using predictions from the age-specific model.

For example, we assume that a person has undergone a preventive intervention at age *i* and assuming that without the intervention he/she would have died at age *x* while with intervention death occurred at age *y*. Hence, *y*-*x* represents the life years gained due to the preventive intervention. The incremental QALY gains due to the intervention calculated using QoL predictions from the age-specific model are:

$$\Delta\text{QALY}(i,x,y,g)=\sum_{j=i}^{y-1}\text{QoL}(\text{age}=j,g)-\sum_{j=i}^{x-1}\text{QoL}(\text{age}=j,g)=\sum_{j=x}^{y-1}\text{QoL}(\text{age}=j,g) \quad (5.9)$$

where g denotes gender.

If we use the TTD approach, the incremental QALY gained from the intervention becomes:

$$\Delta\text{QALY}(i,x,y,g,\text{TTD})=\sum_{j=i}^{y-1}\text{QoL}(\text{age}=j,g,\text{TTD}=y-j)-\sum_{j=i}^{x-1}\text{QoL}(\text{age}=j,g,\text{TTD}=x-j) \quad (5.10)$$

From (5.9) and (5.10) we observe an important distinction between the two approaches: the approach including TTD involves a projection of the QoL utility from the start of intervention until death whereas the comparison method uses only the QoL weights in the life years added due to the intervention.

In order to assess the impact of TTD on the incremental QALYs gained due to a preventive intervention we compute the percentage change of $\Delta QALY(i,x,y,g,TTD)$ relative to $\Delta QALY(i,x,y,g)$:

$$\%\Delta QALY\_change = \frac{\Delta QALY(i,x,y,g,TTD) - \Delta QALY(i,x,y,g)}{\Delta QALY(i,x,y,g)} \times 100 \qquad (5.11)$$

Results

Table 5-2 presents the regression parameter coefficients of the developed GAMLSS models. As indicated, we used a beta inflated distribution for modelling the relationship between the QoL outcome and TTD which is described by three parameters. *Table 5-2* presents the regression coefficients of each of these parameters. The coefficients are presented on the scale of the link functions used: logit, logit and log. Although the use of these link functions in generalized linear models impedes straightforward interpretation of the regression coefficients, we can still analyze the sign of these regression coefficients. The location parameter μ can be thought as similar to the mean parameter in linear regression. By examining the coefficients of μ, we observe that variables age and gender were significant in both the age-specific and the TTD model. The negative sign

Table 5-2: Parameter's coefficients from the age-specific model and from the TTD model

		Age-specific model		TTD model	
Model parameter	Variable	Coefficient	S.E.	Coefficient	S.E.
logit (μ)	constant	1.912*	0.138	1.757*	0.14
	age	-0.008*	0.001	-0.008*	0.0018
	gender (men=1, women=0)	-0.231*	0.034	-0.241*	0.0346
	TTD	-	-	0.005*	0.0022
	TTD^2	-	-	-0.000019	0.00002
logit (Φ)	constant	-0.638*	0.166	-0.753*	0.042
	age	-0.002	0.002	-	-
	ttd	-	-	-0.001	0.0007
log (ν)	constant	-0.212*	0.363	-0.212*	0.363
	gender (men=1, women=0)	-0.686*	0.266	-0.686*	0.2667

of the age effect suggests that an increase of age triggers a decrease in the predicted μ. The sign of the TTD coefficient for μ suggests that QoL decreases when approaching death. The negative sign of the TTD squared coefficient implies that the impact of TTD on QoL becomes stronger if one is closer to death (as values of TTD approach 0). The interpretation of the coefficients of the precision parameter is more difficult since both parameters, μ and ϕ, determine the variance of the outcome distribution. The coefficients of the inflation parameter μ can be considered similar to the part of a two part model that models the probability mass at 1. The regression coefficients of ν indicates that, compared with men; women have a lower probability of being assigned a QoL utility value of one. Please note that all the results presented in this section are based on the sample of deceased.

For investigating the effects of the relationship between QoL and TTD on estimating QALYs gained due to preventive interventions, we will compare QoL predictions from the age-specific model with those from the TTD model. In Figure 5-1, we predicted mean QoL values when TTD was fixed at approximately the maximum observed TTD in our dataset, i.e. at 10 years before death and close to the minimum observed TTD, at 1 year before death. This enabled us to investigate the entire range of QoL predictions by proximity to death and to compare these predictions with those from the age-specific

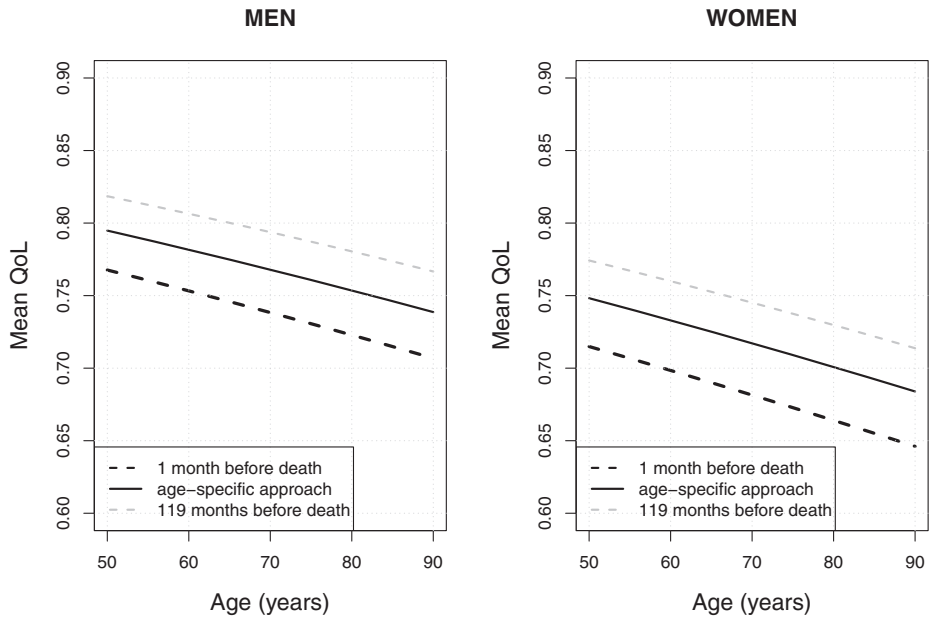

Figure 5-1: Mean QoL predictions from the age-specific regression model and from the TTD regression model for men and women (predictions for 1 year before death and for 10 years before death)

model. As expected, in all 3 situations, the estimated mean QoL decreases with age. Moreover, for all ages 50+, the predicted mean QoL at 1 year before death is lower than the one at 10 years before death and the predicted mean QoL from the age-specific model lies between the other two predictions from the TTD model.

An advantage of using GAMLSS is that by having estimated each parameter of the *BEINF1* distribution we can look at the entire QoL distribution for various ages and values of TTD. Figure 5-2 shows that especially TTD but also age plays an important role in determining the shape and variation of the QoL distribution. Substantial changes in the QoL distribution with TTD have been noted: regardless of age and gender, skewness increased with approximately 0.3 between 1 and 10 years before death. Moreover, for both men and women, between 1 and 10 years from death, kurtosis decreased with approximately 0.2 for most ages. Furthermore, although not easily observed by a visual inspection, for fixed times to death, changes in the QoL distributions have been noticed with age: skweness increases with approximately 0.2 between age 50 and age 90 while kurtosis decreases with approximately 0.1 between these ages. These results indicate that, lower mean QoL values are associated with QoL distributions that have more variation showed by an increase in skewness and a decrease in kurtosis. In other words, lower mean QoL values have higher uncertainty than higher mean QoL values have.

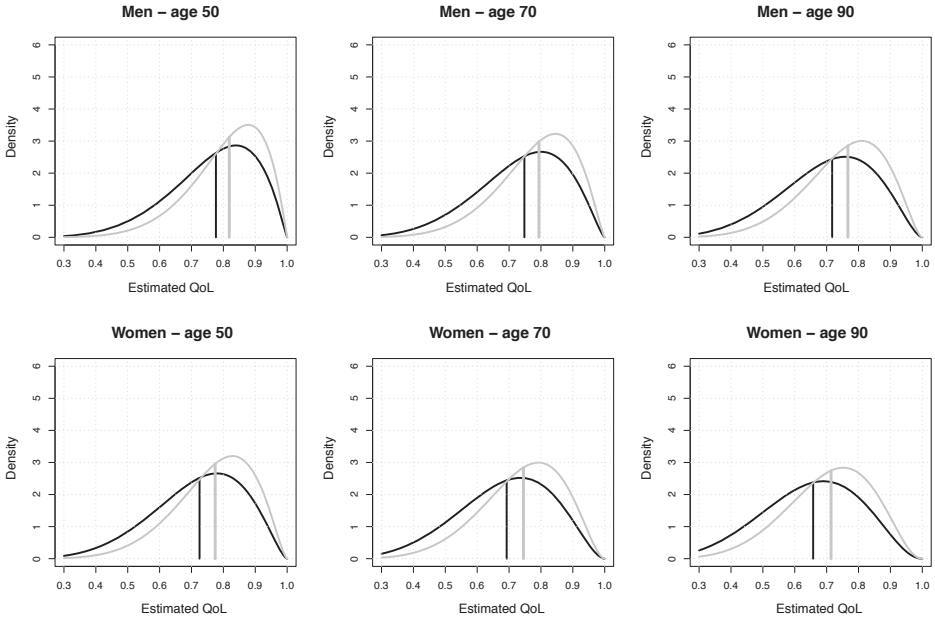

Figure 5-2: The estimated QoL distribution for various ages using the TTD model at 1 year before death (black line) and the TTD model at 10 years before death (grey line). The vertical lines represent the estimated mean QoL corresponding to the 2 situations.

For Figure 5-3 let us assume that a hypothetical woman had a screening intervention at age 70 years and that she was diagnosed with breast cancer. Further, we assume she received early treatment and she lived until the age of 85 years. Had she not undergone the screening, she would have discovered the cancer later and she would have lived only until the age of 80 years. Therefore, due to the screening intervention she gained 5 life years. Figure 5-3 shows the estimated mean QoL with and without the intervention using the age-specific approach (left) and the the TTD approach (right). Assuming that the preventive intervention was not performed and that in this case death would have occurred at age 80, at this age of death the estimated mean QoL from the age-specific model was 0.7 and from the TTD model was 0.66, respectively. On the other hand, with intervention, when death would have occurred at age 85; the age model estimated a QoL of 0.69 whereas the TTD estimated a QoL of 0.66. This example shows that the effect of TTD is stronger when closer to death as QoL decreases strongly in the period preceding death. When calculating QALYs gained due to the hypothetical intervention, QALYs calculation using QoL predictions from the age-specific model involves only QoL utilities in the life years gained. On the contrary, QALYs calculation using QoL predictions from the TTD model incorporates QoL utilities in all the years from the intervention until death. Therefore, as opposed to the age-specific approach, the TTD approach projects the QoL weight from the intervention until death. In the example illustrated above, QALYs gained using the age-specific approach were 3.47 and those gained using the TTD model were 3.62. The relative change suggests that QALYs gained were underestimated with 4.3% when the relation between TTD and QoL was left out. When computing these values, we ignored the discounting rate. For example, a discount rate of 1.5% (as applied in the Netherlands) yields an underestimation of 4.8%. Furthermore, a discount rate of 3.5% (as applied in the UK) and of 5% results in an underestimation of 5.7% and of 6.4%, respectively. Therefore, if the relation between QoL and TTD is ignored, the higher the discount rate, the higher the underestimation of QALYs gained. Hence, for this example, depending on the discount rate used, QALYs underestimation ranges between 4.3% and 6.4%.

Predicted QoL at the baseline age in the example, i.e. at age 70, is higher for the TTD model as TTD values in the example are higher than observed in the data at age 70. As indicated by table 1, the average TTD for all ages is about 4 years. Moreover, at age 70, the mean TTD in our data was approximately 6 years. Hence, in Figure 5-3 (left), the estimated QoL value from the age-specific model (0.72) at age 70 takes into account this average TTD value available in our data. On the other hand, in Figure 5-3 (right) the TTD model predicts mean QoL counting back 10 years before death. Consequently, when compared with the age-specific approach, the estimated mean QoL value predicted from the TTD model (0.75) is and should be larger because the TTD model predicts for a TTD value that is larger than the average observed in our dataset.

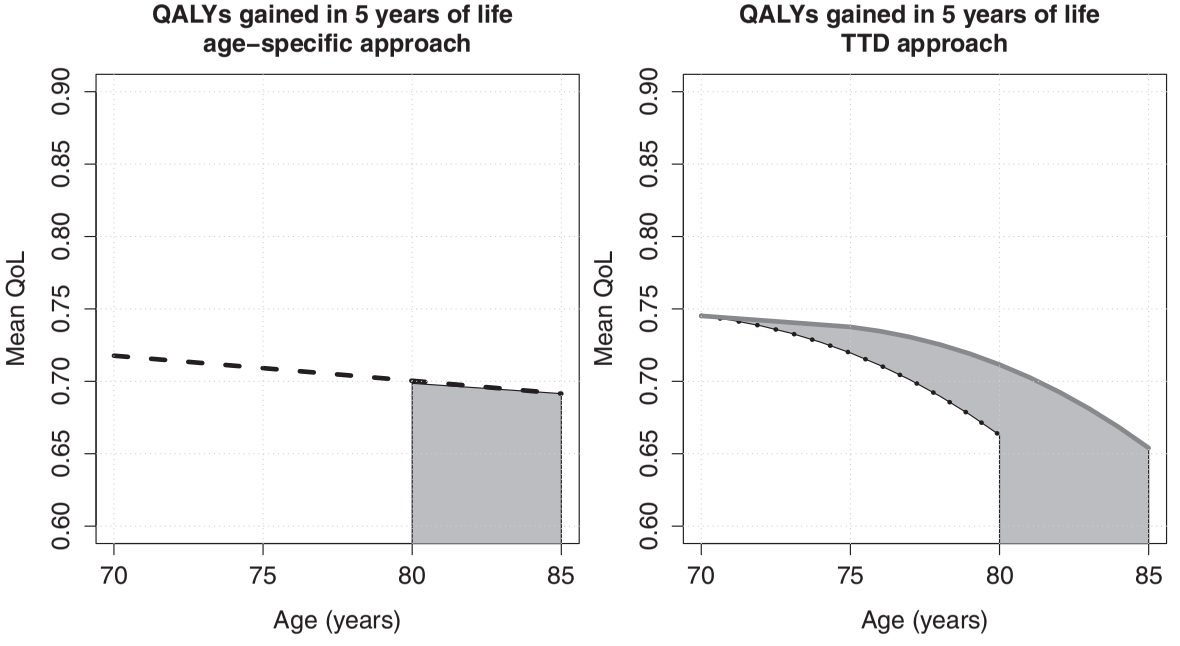

Figure 5-3: QALY gains for 5 extra years of life using the age-specific approach (left) and the TTD approach (right). The grey area indicates the QALYs that are gained.

Furthermore, we have investigated the changes in the level of the QALYs gained underestimation by exploring various potentially influential factors such as: discount rate, gender, age, and number of years gained from the preventive intervention. We found that QALYs underestimation increases with age, number of life years gained due to a preventive intervention and with the discount rate used. Furthermore, it is worth noting that the underestimation tends to be slightly higher for women than for men (differences of approximately 1%). We found that, depending on the above mentioned factors, the QALYs gained underestimation due to a preventive intervention ranges between 3% and 7%. Note that, larger underestimation than 7% could be attained for higher discount rates than 5%.

Discussion

In this chapter, we showed that QoL is related to TTD and that this relationship has important implications when estimating QALYs gained of life prolonging preventive interventions. In order to quantify the effect of this relationship when estimating QALYs gains, we compared QALYs gained calculated using average age-specific QoL estimates stratified by TTD in the life years following a preventive intervention up until death with QALY gains using QoL estimates stratified by age only.

Our empirical findings illustrate that ignoring the relation between TTD and QoL results in QALYs underestimation compared with the case in which age-specific QoL weights would be used. This underestimation increases with advancing age and with the discount rate used. Similar results were observed for both genders with a slight tendency of a higher underestimation for women than for men. Hence, the results reveal that, depending on age, gender, number of years gained from a preventive intervention and the discount rate, QALY gains underestimation in economic evaluations of preventive interventions ranges between approximately 3% and approximately 7%, respectively. Because most modelling studies model effects of interventions through shifts of survival curves, we also investigated how QALYs gains would be underestimated in that case. We did that by decreasing mortality rates for the general Dutch population with varying percentages for ages 50 to 80. Here, we found the same results: depending on age at which life years were gained, and the discount rate used, the shift in the survival curves resulted in an underestimation of QALYs gained that ranges between 3.5% and 6% when information on mortality is not used for estimating QoL weights in life years gained.

Taken all together, our results indicate that ignoring the relationship between QoL and TTD when estimating the health gains in economic evaluations of life prolonging

preventive interventions results in an overestimation of the ICER in these economic evaluations with about 3% to 7%. For interventions that results in an ICER estimate close to the threshold this might be a relevant difference. For example, applying a QALYs underestimation of 3-7% to an estimated ICER of 20000 Euro/QALY calculated using the age-specific QoL weights results in an ICER estimated between 19400 and 18700 cost/QALYs.

Because proximity to death acts as a proxy variable for all kind of disease processes and disabilities that decrease QoL, the approach we proposed can be applied broadly to any life prolonging intervention. However, we chose to illustrate the effects that the relationship between QoL and TTD has on the estimation of QALYs gained due to preventive intervention because of the dataset we had at our disposal, i.e. we used data from the general Dutch population. As preventive interventions are targeted at the general public, this dataset was suitable to draw conclusions about preventive type of interventions, in particular when evaluating primary prevention type of interventions which prevent disease onset. In secondary or tertiary prevention, the patient may not be cured and may suffer from other disease related QoL losses. The approach we proposed can be easily implemented to other life prolonging interventions with appropriate datasets available. Because, all models used in economic evaluations of life prolonging interventions predict survival, all these modelling studies could potentially exploit the relation between QoL and TTD to improve their QALYs estimation.

To our knowledge, this was the first attempt to use the relation between QoL and TTD. However, several studies indicate that QoL is a good predictor for mortality the general populations (Fan et al. 2002, Kaplan et al. 2007). For example, researchers have shown that for the Canadian population the effect of QoL on mortality was strong and significant: an increase of one unit in QoL caused a decrease in the mortality risk with about 53% (Kaplan et al. 2007). TTD has been previously studied in relation to other measures of health such as health care costs (Gandjour and Lauterbach 2005). It has been found that cost-effectiveness analyses overestimate the incremental cost-effectiveness ratio (ICER) of preventive interventions when the relation between TTD and HCE is not explicitly modelled (Gandjour and Lauterbach 2005, Gandjour and Lauterbach 2005, van Baal et al. 2011). Our research shows evidence of ICERs for preventive interventions being overestimated (thus leading to relatively unfavourable ICERs) also when the relation between TTD and QoL is ignored.

Our chapter has a number of limitations. First, in our empirical analysis we used a cross-sectional data; therefore, for each individual, in the period 2001-2008 there is only one measurement of the SF-12. The main advantage of using such a cross-sectional dataset

was the sample size. However, only one SF-12 measurement was available for each person in the data. Future research using longitudinal datasets is encouraged, as it would allow accounting not only for the variation between subjects but also for variation over time within subjects.

Second, we only used the complete case data by deleting the records corresponding to the missing items of the SF-12. This could have biased our results. Still, in chapter 5 we observed that for this dataset QoL in the imputed data was similar with that from the complete case data.

Third, a limitation of the POLS data is that it does not include the institutionalized population. However, probably, persons that are very close to death and have a low QoL would not be in a survey whether they are institutionalized or not. Nevertheless, because the relation between QoL and TTD is stronger when one is closer to death, we suspect that, if accounting for the institutionalized population, the effect of TTD on QoL would become even larger resulting in an even larger QALYs gain underestimation.

Fourth, in our analysis we did not account for other factors, such as the use of future technologies that may affect the health gains of preventive measures. However, with our dataset it is impossible to control for these factors. More importantly, if these other factors are expected to be important when assessing the health gains of a particular intervention, these should be explicitly modelled by (for instance, by extending the number of states in a Markov model). Our QoL estimates stratified by age and TTD should be used in a similar fashion by modellers as 'other cause' mortality rates. Just as any model used to estimate the consequences of a life prolonging intervention includes an 'other causes' mortality our estimates could be used to estimate QoL losses due to 'other causes'.

Finally, we used in this chapter the SF-6D derived from the SF-12; the observed range of the SF-6D was from 0.345 to 1. Fryback and colleagues showed that for a national survey sample of non-institutionalized adults, both the range and the mean of a number of QoL indices (e.g. HUI2, HUI3, EQ-5D, SF-6D derived from the SF-36) differ significantly (Fryback et al. 2007). For example, the minimum observed value for the EQ-5D was -0.11 while for the HUI3 was -0.34 and for the SF-6D (derived from SF-36) was 0.3. The range discrepancies between various QoL indices suggest that our results may be sensitive to the QoL instrument used. For example, given the larger range of EQ-5D and of HUI compared to SF-6D, we suspect that the QALYs gain underestimation would be higher when the former instruments would be used compared to the SF-6D.

This chapter has a number of strengths. First, all models used in economic evaluations of life prolonging interventions predict survival and use a lifetime horizon which makes it easy to calculate TTD; therefore, our proposed approach could be easily applied to other interventions and populations of interest. Importantly, the empirical results presented in this chapter can be used in the practice of economic evaluations of preventive interventions. Our QoL estimates stratified by age and TTD should be used in a similar fashion by modellers as 'other cause' mortality rates. Just as any model used to estimate the consequences of a life prolonging intervention includes an 'other causes' mortality our estimates could be used to estimate QoL losses due to 'other causes'.

Second, in order to address the methodological challenges associated with modelling the QoL utility, we used GAMLSS with *BEINF1* assumption for the response variable. In particular, GAMLSS accounted for specific features of the QoL distribution such as boundness, skewness and heteroscedasticity. One advantage of our modelling approach is that it allowed us to directly model the variance of the QoL distribution. Our empirical results indicated important changes in the QoL distribution for various ages and times until death. A similar modelling approach would have not been possible with more traditional regression models like generalized additive models (GAM) or ordinary least squares (OLS).

This research has important implications for the practice of economic evaluations of life prolonging preventive interventions. This chapter brings upfront the issue of estimating QALYs gained due to life prolonging preventive interventions. In the practice of economic evaluations, the lack of standards regarding the estimation of QALYs gained due to life prolonging interventions results in ad-hoc choices. Nevertheless, these choices may have a high impact on the final outcome of economic evaluations. In order to realistically estimate the health benefits in cost-effectiveness studies, the economic evaluations analysts need to account for quality of life losses caused by competing risks in added years. We propose to use QoL utilities stratified by age, gender and proximity to death when estimating QALYs gained in economic evaluations of life prolonging preventive interventions. This approach can be easily incorporated in the standard practice of economics evaluations for preventive interventions. The use of this approach will result in more accurate ICERs for life prolonging preventive interventions and; therefore, will contribute to assist decision makers in making better decisions.

Chapter 6

Predicting patient-reported outcomes from disease-specific questionnaires: an extensive comparison of existing methods

With Anca Hanea, Melinde Boland, Bettina Grun and Maureen Rutten van Molken .

Under review.

Abstract

Patient-reported outcome measures (PROMs) such as the EuroQol-5 dimensions (EQ-5D) questionnaires are important in health economics, especially for generating evidence to support reimbursement decisions. When these measures are not available from clinical studies, usually they are predicted or 'mapped' from other measures such as disease-specific ones. For this purpose, two classes of methods have been proposed and these can be distinguished by whether the dependent variable is the EQ-5D utility or the probability that a respondent selects a particular level for each question: (i) utility mapping methods (OLS, Tobit, linear mixed effects (LME), beta regression (BM), finite mixture regression (FMM)) (ii) response mapping methods (multinomial logit (MNL), Bayesian Networks (BNs)). The aim of this study is to perform a comparison of these methods. Answers on a disease-specific questionnaire and the EQ-5D were collected in two trials (N1=5157/N2=366) which were separately used for fitting and validating the models. We found that FMM fitted the EQ-5D data better than the other methods. Although, the differences in the metrics of out-of-sample prediction accuracy among all the methods was small, on average errors were lowest for FMM (MAE=0.187 and MAE=0.176 for the two trials). Furthermore, out-of-sample performance depended largely on whether the sample population used to develop the models included a high number of respondents in poor health. Concluding, for generating external predictions, with the present datasets, none of the existing mapping methods strongly outperformed the others on the entire range of the EQ-5D utility; however, on average FMM produced slightly smaller errors.

Introduction

Patient-reported outcome measures (PROMs) which include health-related quality of life (HRQoL) instruments have a number of applications in health care decision making, ranging from being used as indicators of hospital performance (Devlin et al. 2010), to measuring and describing population health status (Parrish 2010) and capturing the effect of alternative health treatments on patient's morbidity and mortality (Brazier et al. 1999). Here we will refer to the context in which PROMs are used for evaluating the impact of alternative treatments on patients' health; therefore, contributing in generating evidence for supporting reimbursement decisions of medical technologies.

PROMs have certain classifications. Depending on whether the valuation method used to derive their summary value or index score is consistent with economic theory, these outcomes can be preference-based measures (PBM) or non-preference-based measures (NPBM). PBM refer to the fact that each patient's score incorporates values that reflect the preferences of the general public. A further distinction relates to generic versus disease-specific depending on whether PROMs evaluate general HRQoL or disease-specific aspects of health. For supporting resource allocation decisions, PROMs need to be preference-based and generic as only these provide a common ground for comparing health outcomes across various clinical studies. However, often such outcomes are not included in clinical studies instead disease-specific ones are. In such situations, a common solution accepted is to develop a model for 'mapping' or predicting the generic HRQoL from another disease-specific measure using a dataset that includes both of these measures for developing a model that can be further used to predict generic HRQoL at the patient level when this data is unavailable. Therefore, the choice of the mapping model is of critical importance as it can potentially affect the reimbursement decisions.

One of the most commonly used preference-based generic HRQoL is EuroQoL-5D (EQ-5D) developed by the EuroQol Group (www.euroqol.org); this has been indicated as the instrument of choice by the National Institute for Health Care and Excellence (NICE) to support reimbursement decisions. EQ-5D consists of five dimensions of health that describe the current health states of patients: mobility, self-care, usual activity, pain/discomfort, anxiety/depression. Each of these dimensions is measured with one item and respondents have to respond on a three-point scale (no, some or extreme problems), hence 243 (3^5) health states can be identified by a five-digit number. The key characteristic of the EQ-5D instrument is that it enables each health state to be converted into a single summary value by applying a formula that attaches a score on each of the levels in each dimension. This formula is obtained through a valuation process performed using samples from the general populations.

Societal preferences values for each of the EQ-5D health states have been obtained using sample populations of many countries including the UK (Dolan 1997), the Netherlands (Lamers et al. 2006) and the US (Shaw et al. 2005). The most widespread used tariff is that of the UK and this will also be employed in the present chapter. An important guide to the developed value sets is presented in (EuroQol Group 2007) and useful indications regarding the international comparison tariffs can be found in (Norman et al. 2009). Most of these surveys estimated the EQ-5D index score in two steps. First, by using a representative sample population for each country, the time trade-off method was used to elicit preferences for 42 out of 243 EQ-5D health states. Second, regression methods were used for estimating values for the remaining EQ-5D health states. In this way, all EQ-5D health states are described with a single summary index score usually ranging from bellow 0 (for patients reporting severe problems in all five dimensions) to 1 (full health). In the UK, the score ranges from -0.595 and 1. Due to its widespread use and also because it is the recommended instrument by NICE, the majority of mapping studies focused on predicting EQ-5D either from other generic HRQoL instruments (Franks et al. 2003, Fryback et al. 1997, Gray et al. 2006, Sullivan and Ghushchyan 2006) or from disease-specific ones (Versteegh et al. 2012, Arnold et al. 2015). This chapter focuses on mapping a disease-specific measure onto EQ-5D.

Two main categories of methods have been proposed in the literature for predicting EQ-5D data from other disease-specific measures. These can be distinguished by whether the dependent variable is the utility score or the probability that a respondent selects a particular level for each question, i.e. researchers proposed using either utility score mapping also known as direct mapping (Franks et al. 2003, Fryback et al. 1997, O'Brien et al. 2003, Versteegh et al. 2012) or probability mapping also labeled indirect mapping (Longworth et al. 2014, Le and Doctor 2011, Gray et al. 2006). Because the utility tariff is applied after the mapping model is estimated, compared to utility mapping, probability mapping has the advantage of being easily accommodated to tariffs for various countries.

The EQ-5D utility data presents certain non-standard features which are a result of its bounded nature, i.e. such data typically has mass points at boundaries (in this case at one), is skewed, exhibits discontinuity (e.g. between 0.883 and 1) and is heteroscedastic given that the variance will approach zero as the mean approaches either boundary point (Kieschnick and McCullough 2003). Due to these characteristics, there is no commonly standard accepted utility mapping method. However, compared to response mapping, previous research indicates that utility mapping is by far the most frequently used approach (Dakin 2013). This chapter showed that about 80% of the mapping studies used the ordinary least square method (OLS, (Franks et al. 2003, Fryback et al. 1997, O'Brien et al. 2003, Versteegh et al. 2012)), often compared with Tobit like models such as the

classical Tobit model (Austin 2002) or censored least absolute deviation models (CLAD, (Sullivan and Ghushchyan 2006)). Furthermore, for modeling the probability mass at one, hurdle models were sometimes used (Mullahy 1986). Given the longitudinal structure of the clinical studies that include PROMs, random effects or linear mixed effects models (LME) have also been proposed (Soini et al. 2012). Previous research has pointed out that the above methods which are based on the normal distribution assumption are likely to result in biased and inconsistent estimates of the mapping coefficients, in predictive values that may be outside of the domain, and generally in poor prediction performance (Basu and Manca 2012, Mortimer and Segal 2008). To address these issues more complex model specifications have been recently proposed, e.g. finite mixture models (FMM). Some found FMM to be superior when compared to OLS and Tobit models in mapping exercises (Hernandez Alava et al. 2012, Hernandez Alava et al. 2014, Coca Perraillon et al. 2015) while others indicated they did not show improvements when compared to multinomial logistic model (Kent et al. 2015). Furthermore, it has been shown that beta regression models (BM) offer superior predictive performance when compared to OLS and Tobit like models (Khan and Morris 2014).

To avoid modeling directly the non-standard EQ-5D utility distribution, others proposed probability mapping by predicting the instruments' response questions rather than its utility score (Le and Doctor 2011, Gray et al. 2006, Conigliani et al. 2015) and apply the utility tariff to the estimated probabilities. For example, Gray and colleagues employed a series of multinomial logit (MNL) regressions to estimate the probability that a respondent selects a particular severity level from each EQ-5D question and then applied the UK tariff to these probabilities (Gray et al. 2006). More complex methods have been proposed for response mapping, e.g. Lee and Doctor used Bayesian Networks (BNs) for predicting the probability of each response level of the five EQ-5D domains obtained from a Bayesian updating process and found that the BNs mapping model was superior in terms of predictive accuracy when compared to the MNL model but also to OLS or CLAD. The superior predictive performance of BNs has also been confirmed in (Borchani et al. 2012). To account for both the likely dependence between the EQ-5D item responses at the patient level and the fact that the EQ-5D item responses are naturally ordered, recently ordered probit models have been highlighted in a mapping exercise that predicts the EQ-5D from another preference based measure (Conigliani et al. 2015).

This chapter will consider two issues. First, the promising mapping approaches, e.g. FMM, beta regression models and BNs have only been compared separately to the commonly used mapping methods such as OLS and Tobit like models; therefore, it is still unclear which of these approaches predicts better the EQ-5D data. This can only be established by performing an extensive comparison among the above mentioned methods; to our

knowledge, this has not yet been conducted. Therefore, the aim of this chapter is to compare (i) utility mapping methods: OLS, Tobit models, LME, BM and FMM with (ii) response mapping methods: MNL and BNs. Second, because of data unavailability, a common issue when validating mapping models is that such models are not sufficiently tested for their out-of-sample prediction performance. Many of the above studies have shown that the proposed mapping models perform well within their validation sample taken from the same sample population but often these models remain untested for their performance in other populations. In this chapter, we will use two longitudinal clinical studies that include both a disease-specific questionnaire and the EQ-5D. We will use interchangeably the data from one trial to fit the model and the data from the other one to validate it for its out-of-sample performance. The disease-specific questionnaire considered here is the chronic obstructive pulmonary disease questionnaire (CCQ); an instrument developed for measuring health status in chronic obstructive pulmonary disease (COPD) patients. CCQ includes ten questions monitoring general health dimensions and also questions specific for COPD (e.g. about coughing). For mapping the CCQ onto EQ-5D, a model has been recently developed using OLS and Tobit methods (Boland et al. 2015).

This chapter is organized as follows. Section 2 describes the data used in the present chapter. Section 3 presents an overview of the methods used for developing the mapping model. Section 4 illustrates the main results and findings of our analyses. Finally, Section 5 draws conclusions and discusses the most salient points.

Data

This chapter uses data coming from two different randomized clinical trials: the RECODE trial (n=6516 observations and 1086 individuals), a two-year, cluster randomized controlled trial recruited from general practice and the GO-AHEAD (n=382 observations and 166 individuals) trial, a 3 month, multi-center randomized trial with 166 patients hospitalized for COPD exacerbation. Patients included in these trials completed both the CCQ and the EQ-5D questionnaire at six and three time points, respectively. About 20% and 4% of observations were missing in the RECODE and GO-AHEAD trial; in order to perform our analyses we have discarded the observations including missing data.

Figure 6-1 describes the distributions of each of the five EQ-5D responses and shows that a small number of observations have an EQ-5D response equal to three (this is corresponding to the 'extreme problems' scale). Figure 6-2 shows the derived EQ-5D utility for both the RECODE and the GO-AHEAD trial and indicates features usually observed for this data as indicated in the literature (Boland et al. 2015), i.e. the EQ-5D score has some

mass at one (i.e. full health), is bimodal and discontinuous as it has no mass between 0.883 and 1. A simple visual inspection indicates that both the response distributions and the utility scores are different in RECODE compared to GO-AHAED suggesting worse health states for patients in GO-AHAED than for those in RECODE which is to be expected since all patients in GO-AHEAD were hospitalized for a COPD exacerbation. *Table 6-1* illustrates descriptive statistics of the EQ-5D scores in the two trials confirming the difference in patient's health as indicated by EQ-5D. For example, about 23% of patients in GO-AHEAD and about 15% in RECODE had a utility score less than 0.5, respectively.

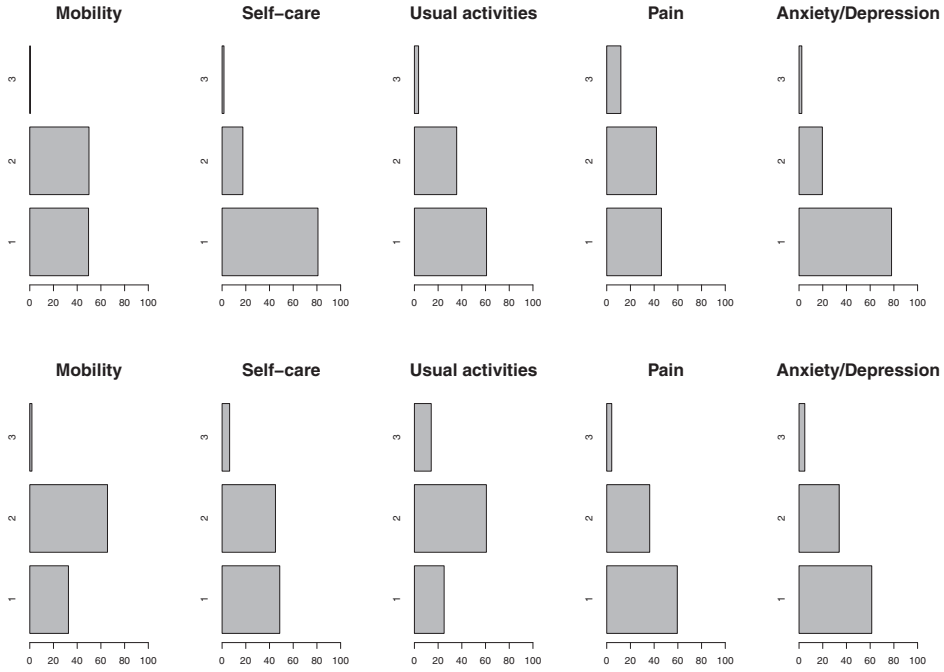

Figure 6-1 Distribution of responses across the five EQ-5D domains for the RECODE trial (top) and for GO-AHEAD trial (bottom)

Table 6-1: Descriptive statistics for EQ-5D utility scores in RECODE and GO-AHEAD

	RECODE (n=5157)	GO-AHEAD (n=366)
Mean	0.706	0.615
Minimum	-0.594	-0.594
Maximum	1	1
EQ-5D<0.25 (%)	13.0	15.6
0.25<=EQ-5D<0.5 (%)	2.5	7.7
0.5<=EQ-5D<0.75 (%)	31.7	44.3
0.75<=EQ-5D<=0.883 (%)	27.3	20.4
EQ-5D=1 (%)	25.5	12

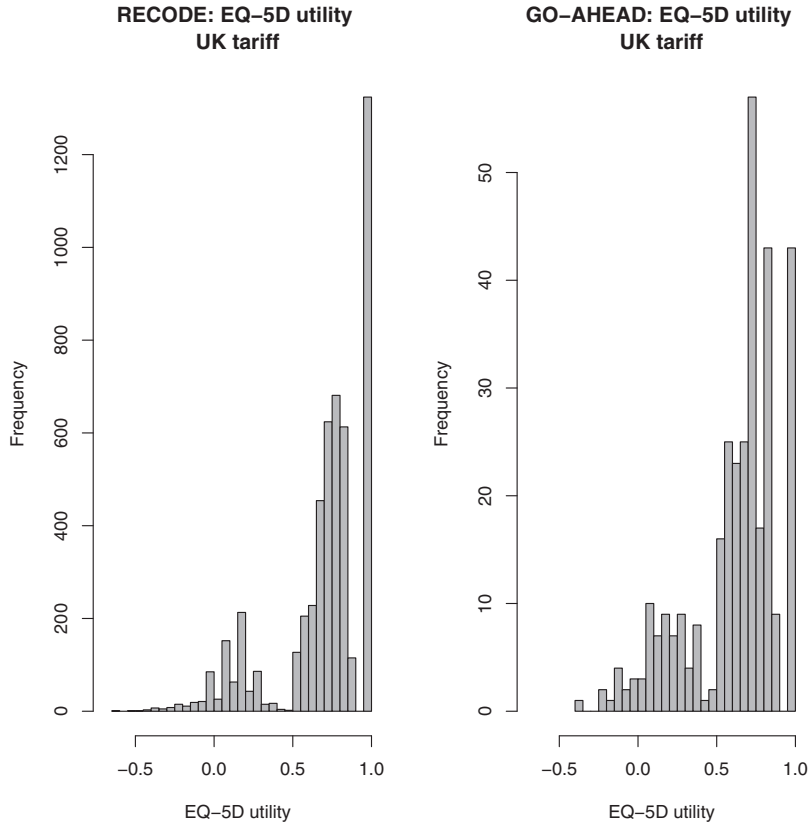

Figure 6-2: Distribution of EQ-5D utility scores for the RECODE trial (left) and for GO-AHEAD trial (right)

CCQ is an instrument used to measure HRQoL for patients diagnosed with COPD; this questionnaire includes ten questions and meaures aspects of health on three domains, i.e. symptoms, functional and mental state. Respondents are asked to respond on each item (question) on a seven point scale resulting in 282 million possible health states. Response options for questions 1-6 are: never/hardly ever/a few times/several times/ many times/a great many times/almost all the time whereas response options for questions 7-10 are: not limitted at all/very slightly limited/slightly limited/ moderate limited/ very limited/extremly limited/totally limited or unable to do. Therefore, the scale of each CCQ question ranges from the best health state (0) to the worst health state (6). Figure 6-3 illustrates the CCQ questions frequencies as indicated in the RECODE and GO-AHAED trial. In concordance with responses observed in the EQ-5D questionnaire, we observe that patients in the GO-AHEAD trial give responses that suggest a worse health state compared to those in RECODE.

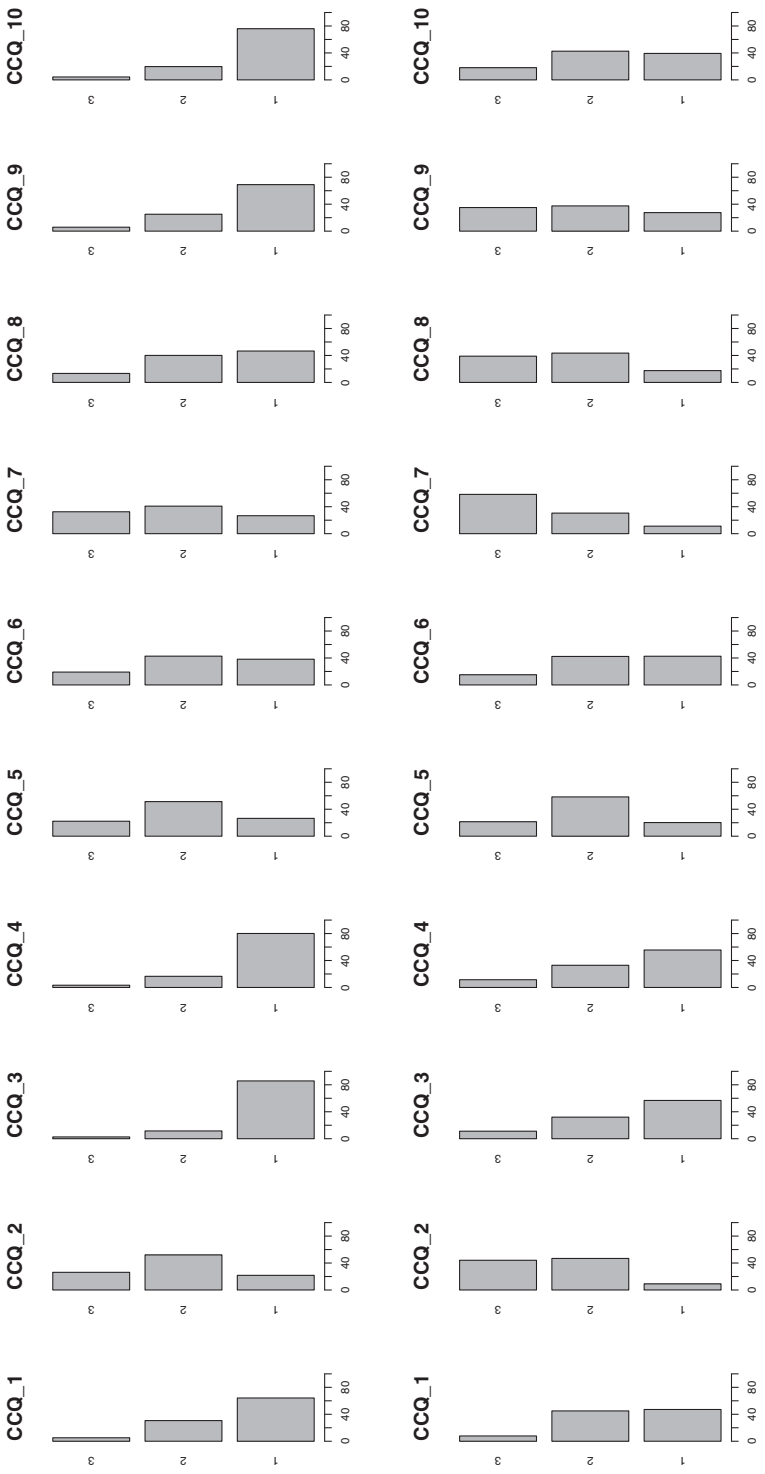

Figure 6-3: The distribution of the ten questions included in the CCQ questionnaire in RECODE trial (top) and GO-AHEAD trial (bottom)

Methods

In what follows, we will briefly present the various mapping methods used for predicting either the EQ-5D score (i.e. utility score mapping) or the EQ-5D response (i.e. response mapping). Here we will not consider the problem of verifying what sets of covariates is preferable or what is the optimal functional form of these covariates; rather we focus on comparing the various methods from a general point of view. For enhancing comparability between methods, in regression models, we opt for modelling EQ-5D as a function of the ten CCQ domains and we excluded interaction terms between the CCQ domains while in the BNs we developed a graph separately for each EQ-5D question together with the CCQ variables. Furthermore, for ease of estimation using response mapping, we have collapsed the seven-point response scale of each CCQ question onto a three-point scale by combining the response options one and two, three and four, and five, six and seven as in (Boland et al. 2015). Therefore, in all the models developed, CCQ variables have been entered as dummy variables with three levels.

Utility mapping methods

Ordinary Least Squares (OLS)

Mainly due to its ease of use, OLS is frequently used for modelling the EQ-5D data. Therefore, any extensive comparison between various EQ-5D mapping methods must include an OLS model:

$$y_i=x_i^T\beta+\varepsilon_i,\ \varepsilon_i\sim N(0,\sigma^2) \qquad (6.1)$$

where y_i is the EQ-5D tariff score for individual *i*, β is a vector of coefficients and x_i^T is a row vector of covariates; here these covariates are represented by the ten CCQ questions. OLS assumes that the errors are identical, independent and normally distributed.

Linear mixed effects model (LME)

Generally, observations in clinical trial data are collected on multiple time points for patients; therefore, each patient will have multiple measurements. For addressing these data features we will use a linear mixed effects model (LME). Such a model contains two parts, a fixed and a random effects part: the fixed effects have the same interpretation as in the linear regression whereas the random effects are interpreted as how a subset of regression parameters for example, for the *j*-th subject deviates from those in the population. A flexible model specification for a mixed effects model consists in a random intercept and random slope model allowing each individual to have a different trajectory over time:

$$y_{ij}=x_i^T\beta+b_{i0}+b_{i1}t_{ij}+\varepsilon_i,\ b_i \sim N(0,D),\ \varepsilon_i \sim (0,\sigma^2) \quad (6.2)$$

where y_{ij} is the EQ-5D tariff score for individual *i* at time point *j;* $b_i=(b_{i0},b_{i1})$ denoting the vector of random effects for each subject *i*. The random effects are assumed to be normally distributed with mean zero and variance-covariance matrix *D* and are assumed independent of the error terms ε_i. This model enables obtaining both marginal predictions (i.e. for the average individual) and new predictions for individuals in the dataset. Note that, for out-of-sample prediction, LME marginal estimates need to be used.

Tobit model

For dealing with the bounded nature of the utility score outcome, Tobit like models including the classical Tobit model (Austin 2002) or censored least absolute deviations (CLAD), (Sullivan and Ghushchyan 2006, Wijeysundera et al. 2011) were proposed previously for mapping the EQ-5D utility. However, Tobit-type models are intended to be used in a manner more applicable to censored dependent variables and this clearly is not the case with the EQ-5D tariffs. Here we will consider the classical Tobit model. In the Tobit model the independent variable y is assumed right censored at one:

$$y_i^*=x_i^T\beta+\varepsilon_i,\ \varepsilon_i \sim N(0,\sigma^2) \quad (6.3)$$

$$y_i = \begin{cases} y_i^*, y_i^* < 1 \\ 1, y_i^* \geq 1 \end{cases} \quad (6.4)$$

where y_i^* is assumed a latent ('unobserved') variable. Similar to OLS, the classic Tobit model also assumes that the errors are normally distributed. Previous studies have shown that the classical Tobit model performs poorly for departures from normality or in the presence of heteroscedasticity (Maddala 1983, Greene 2000).

Beta regression model (BM)

Since the EQ-5D utility is defined between 0.5 and 1, it therefore, exceeds the definition interval for the beta regression model (BM), i.e. (0,1). Hence, certain steps are required before fitting a beta regression model. First, we used a transformation of the form $y_i^*=(y_i-a)/(b-a)$where *a* is the minimum utility, *b* is the maximum utility and y_i is the observed EQ-5D utility. In this way the transformed EQ-5D utility will be defined on the closed interval between 0 and 1, i.e. $y_i^* \epsilon [0,1]$. Note that only one observation was actually equal to 0, we took a slightly larger value than 0 for that observation (0.0000001). Nevertheless, the transformed EQ-5D data had about 25% and 12% of observations equal to one in RECODE and GO-AHEAD trial, respectively. Second, we developed a BM using a two-part model specification (Mullahy 1986). The first part models the probabil-

ity that an individual has an EQ-5D utility of one using a logistic regression. The second part, applied the beta regression model to EQ-5D data defined between zero and one.

$$f(y_i) = \begin{cases} \pi_1, & y_i = 1 \\ (1-\pi_1) f(y|\mu,\phi), & 0 < y_i < 1 \end{cases} \quad (6.5)$$

where π_1 is the probability of observing 1; $f(y\mu,\phi)$ denotes the beta density parametrized by the location/mean parameter $\mu(0<\mu<1)$ and precision parameter $\phi(\phi>0)$:

$$f(y_i|\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}, \; 0 < y < 1 \quad (6.6)$$

with $\Gamma(\bullet)$ denoting the gamma function, $E[y]=\mu$ and $Var[y] = \frac{\mu(1-\mu)}{1+\phi}$. In this context, ϕ is interpreted as a precision parameter because, for each fixed value of the mean μ, $1+\phi$ is inversely proportional to the variance of y. Hence, this model is defined by three parameters; we will model each of these parameters, i.e. (μ,ϕ,π_1) as functions of the ten CCQ domains using the link functions *logit*, *log*, *log* respectively. Such models have been implemented using maximum likelihood techniques by (Ospina 2010) in the GAMLSS package in R (Stasinopoulos and Rigby 2007, Rigby and Stasinopoulos 2010).

The expected value of the estimated transformed EQ-5D utility data will be calculated as follows:

$$E[y_i^*]=\pi_1 E[y_i^* | y_i^*=1]+(1-\pi_1)E[y_i^* | y_i^*<1]=\pi_1+(1-\pi_1)E[y_i^* | y_i^*<1] \quad (6.7)$$

For assessing model performance and comparing this model with the other approaches, the expected value indicated in (6.7) was back-transformed to the original scale of the EQ-5D data.

Finite mixture models (FMM)

A finite mixture model (FMM) assumes that the probability density generating the observed outcome is a combination of *K* different densities:

$$f(y_i) = \prod_{j=1}^{K} p_j f_j(y_i|\mu_j,\sigma_j) \quad (6.8)$$

where p_j are nonnegative quantities that sum to one, $0<p_j\leq 1$, $\sum_{j=1}^{K} p_j = 1$, p_j are mixing proportions or weights and f_j are component densities of the mixture, *K* denotes the number of components or classes, μ_j denotes the mean for each component and σ_{2j} denotes the variance for each component. Therefore, each density in the mixture is characterized by its own mean and variance. In a FMM, densities can be discrete or continuous or combina-

tions of the two. Furthermore, in a FMM, probabilities p_j can either be considered constant (equal for each observation) or can be modelled separately as a function of explanatory variables. In the FMM specifications considered here:

- μ_j with $\mu_j=x_i^T\beta_j$ were estimated as linear functions of the ten CCQ questions;
- σ_j^2 were estimated as constants for each component *k* (i.e. equal for each observation and component)
- p_j were estimated as functions of the ten CCQ questions using multinomial logistic regressions

In what follows we will consider various FMM where densities will be considered either combinations of densities of continuous variables or combinations of densities of continuous and discrete variables. The number of classes included in the mixture models was chosen by means of a two-fold cross-validation. We sampled observations regardless of their patient ID and we evaluated the cross-validations by using the smallest prediction error criterion. For brevity these results are not shown here; however, it is worth mentioning that we considered up to five classes and, we found that mixture densities with two classes fitted better the EQ-5D utility data than mixtures with more components. Therefore, we developed the following mixture models: a Gaussian mixture model (GMM) containing two classes of normal distributions, inflated Gaussian mixture model (IGMM) including three classes: two classes of normal distributions and one class with a fixed value at one for modelling the mass at one and beta mixture models (BMM) containing two classes of beta distributions. These models were fitted in R using the packages *flexmix* (Grün and Zeileis 2012) and *betareg (Grün et al. 2012)*.

(a) Gaussian mixture model (GMM)

For addressing the bimodality observed in the EQ-5D utility data, a FMM with two classes of normal distributions was used:

$$f(y_i)=p_1\phi_1(y_i|\mu_1,\sigma_1)+p_2\phi_2(y_i|\mu_2,\sigma_2) \quad (6.9)$$

where p_1 and p_2 denote the probability that an observation belongs to class 1 and class 2, respectively. $\phi_j(y_i|\mu_j,\sigma_j),j=1,2$ denote univariate Gaussian distributions with mean $\mu_j=x_i^T\beta_j$ and variance σ_j^2. Therefore, a total of 35 parameters were estimated with this model.

(b) Inflated Gaussian mixture model (IGMM)

In order to address the discontinuity observed between the EQ-5D values of 0.883 and 1, a special mixture model with three classes where one class has a mean with a fixed value at one and the other two follow a Normal distribution was developed:

$$f(y_i)=p_1\phi(y_i|\mu_1,\sigma_1)+p_2\phi(y_i|\mu_2,\sigma_2)+(1-p_1-p_2)\phi(y_i|1,\sigma_3) \quad (6.10)$$

where p_1 denotes the probability that an observation belongs to class one, p_2 denotes the probability that the entity belongs to class 2 and $(1-p_1-p_2)$ denotes the probability of observing one. In order to estimate this FMM, we implemented a *flexmix* driver (the code for this driver is included in the Appendix). Compared to the model denoted by equation (9), this model will estimate 11 more parameters resulting in a total of 46 parameters.

(c) Beta mixture models (BMM)

Similar to the beta model illustrated in section 3.1.4, we used a two-part model specification for developing a model which includes a mixture of beta distributions instead of one beta distribution for the second part.

$$f\left(y_i\right)=\begin{cases}\pi_1, & y_i=1\\ \left(1-\pi_1\right)\left(p_1B\left(y_i|\mu_1,\phi_1\right)+\left(1-p_1\right)B\left(y_i|\mu_2,\phi_2\right)\right), & 0<y_i<1\end{cases} \quad (6.11)$$

where π_1 denotes the probability of observing an EQ-5D utility equal to one, $p_1$and $1-p_1$ denote the probability that an observation belongs to class one and two of the mixture model, respectively; $B(y_i|\mu_j,\phi_j)$, $j=1,2$ denote univariate beta distributions with mean $\mu_j=x_i^T\beta_j$ and variance $\frac{\mu_j\left(1-\mu_j\right)}{1+\phi_j}$.

This model will estimate a total of 46 parameters.

(d) Predictions with FMM

In this chapter two types of predictions with FMM will be generated: in-sample and out-of-sample predictions. To obtain in-sample predictions with FMM, the EQ-5D data was estimated as the sum of the predictions for each component weighted by the probability of falling into a cluster given the observed EQ-5D data and the covariates x_i. To generate out-of-sample predictions with FMM, we estimated two types of predictions. First, unobserved EQ-5D data was estimated as the sum of the predictions for each component weighted by the estimated mixing probabilities given the covariates x_i. We refer to this type of predictions as to the 'weighted average (WA)' predictions. Second, unobserved EQ-5D data was assigned to a class based on the maximum of the estimated mixture probabilities given x_i and then the predicted EQ-5D was the prediction corresponding to the assigned class; we refer to this type of prediction as to 'maximum probability clustering (MPC)'.

Response mapping methods

Response mapping involves predicting the probability levels of each EQ-5D question. However, for comparing these approaches with score based mapping methods, the predicted probabilities need to be converted into EQ-5D scores by applying the EQ-5D tariffs. For converting the estimated EQ-5D probabilities into the EQ-5D scores we will use the Expected-Utility (EU) method and the Most-Likely Probability (MP) approach as proposed in (Le and Doctor 2011). In the EU method, the EQ-5D utility is obtained by taking the expectation of the estimated probabilities and the associated utilities for each EQ-5D domain in part. In the MP method, the maximum estimated probability for each EQ-5D domain is assigned with the corresponding utility. Details about these methods are provided in (Le and Doctor 2011).

Multinomial logistic regression (MNL)

Let y_{il} be the EQ-5D response for individual *i* to dimension *l (l*=1,…,5) and let β_{rl} be the vector of coefficients for level *r (r=1, 2, 3)*. The probability of observing outcome *r* for each individual *i* in each dimension *l* is:

$$P(y_{il} = r) = \frac{\exp(x_i^T \beta_{rl})}{\sum_{s=1}^{3} \exp(x_i^T \beta_{sl})} \qquad (6.12)$$

where, for identification purposes, for each EQ-5D dimension$\beta=0_{1l}$. Therefore, the remaining coefficients β_{2l}, β_{3l} represent changes relative to $y=1_{il}$.

Bayesian networks (BNs)

BNs are a form of graphical models with focus on structure discovery, i.e. determining an optimal graphical structure from the observed data. Compared to regression methodologies, in the BNs framework there is no need for a distinction between response or outcome variables and explanatory variables, depending on the situation, any variable can be considered as outcome or explanatory variable. Recall that within a regression framework the covariates are assumed independent of each other while a BNs model enables to explicitly model relationships between all covariates and is therefore intuitively more reasonable than regression analysis.

BNs belong to the graphical models structures known as directed acyclic graphs (DAG) and enable an effective representation of the joint probability distribution over a set of variables. A structure of a DAG is represented by two sets: the set of nodes which represents random variables and the set of edges which indicates potential statistical dependencies between variables. In particular, an edge form a node X_i to another node X_j indicates that X_i influences X_j with X_i being referred to as a parent of X_j (which is referred to as the child node). An extension of these relationships includes the set of descendants or ancestor which represent the set of nodes that can be reached through a direct path from the child node (the parents of the parents of the child node). A BN model is based on the conditional independence relationships between the variables i.e. each variable is conditional independent of its ancestors given its parents. This property is necessary for potentially reducing drastically the number of parameters that are needed to estimate the joint probability distribution. BNs were developed for both continuous and discrete variables; in what follows we will only cover discrete BNs. With a set of only discrete variables X_i, *i=1,..,n*, the joint probability distribution is factorized as follows:

$$P\left(X_1,...X_n\right)=\prod_{i=1}^{n}P\left(X_i\middle|Pa\left(X_i\right)\right), \tag{6.13}$$

where $Pa(X_i)$denotes the set of parents of node X_i.

Both the structure of the BNs and their parameters can be learned from available data. As such, these graphical models are widely used in many areas e.g. statistics, machine learning and artificial intelligence with many methods being developed for learning the structure and the corresponding parameters.

Structure learning was performed here using both constraint-based methods (Spirtes et al. 2001) as in (Le and Doctor 2011) but also score-based methods such a greedy search algorithms (Bouckaert 1995). For most of the constrained-based methods the

BNs' structure is searched starting with a saturated graph which is fully connected. Subsequently, edges between nodes are removed if independence tests, e.g. χ^2 tests are rejected with various levels of significance. The constraint-based methods used were PC (after its authors Peter and Clark (Spirtes et al. 2000)) and grow-shrink algorithms (GS, (Margaritis 2003)).

Greedy search methods assess the BN structure using goodness of fit measures such as the Bayesian Information Criterion (BIC). With these methods the search starts with an empty graph and then edges between two variables are added, deleted or reversed until no score improvements can be achieved. In this way, the BN structure with the best score (i.e. best fit) is selected. Examples of algorithms included in this category and used here include hill climbing and tabu search (Bouckaert 1995).

For learning the structure of the BN, we used the *bnlearn* package available in R (Scutari 2010). For ease of computation, we developed five separate BNs models, each including the ten CCQ questions and one EQ-5D domain. Because various methods result in different BNs structures, we selected the optimal BNs structure by means of a two-fold cross-validation analysis in which the model that resulted in the smallest prediction error for the EQ-5D question was selected. A description of the finally learnt BNs structure for each EQ-5D domain is illustrated in the Appendix. After the BNs structure was learned, the corresponding conditional probability parameters were estimated from the data by means of a Bayesian estimation in which each child was estimated as a function of its parents using a logistic regression model.

Once the structures and parameters of the BNs were learned, they were used to estimate the probabilities of the response levels for each EQ-5D domain; this process is called 'probabilistic inference' and requires the use of special algorithms that need to be used in this updating process. We used an algorithm for exact inference called 'junction tree' (Nagarajan et al. 2013) which has been implemented in package *gRai*n in R. An overview of BNs and their implementation and potential use in R can be found in (Nagarajan et al. 2013).

Comparison between different models

Because the aim of the mapping is to predict EQ-5D at the patient level when these data are unavailable, the accuracy of the prediction is a key aspect of the performance of various mapping methodologies. Note that, all of the presented models will be compared in terms of their estimated utility scores. For assessing the quality of the forecast in a mapping model, it has been recommended (Longworth and Rowen 2011) that the mean absolute error (MAE) and the mean square error (MSE) between the estimated and the observed utilities should always be reported. Note that, compared to MAE, MSE is more

sensitive to outliers and large changes. These measures will be used for both within sample validation and out-of-sample validation. Furthermore, to assess in-sample model fit, we will also report R^2 which indicates the percentage of explained variance by the models. The two trials analysed here differ in sample size and in the extent of observations in poor health. Due to these differences, we will present measures of validation for each trial separately instead of averages between the two.

Results

In-sample validation

Table 6-2 shows in-sample prediction errors of all fitted models for both the RECODE and GO-AHEAD trials as indicated by MAE, MSE and R^2. *Table 6-2* indicates that generally, utility mapping methods result in better in-sample performance than response mapping approaches, i.e. lower MAE and MSE and higher R^2. Furthermore, the models trained with GO-AHEAD trial resulted in smaller errors than those trained with RECODE. This may be due to the inclusion of more individuals in poor health in GO-AHEAD compared to RECODE.

Among the utility mapping methods, FMM fitted the EQ-5D utility data best while Tobit and OLS models provided the worst fit to the EQ-5D data regardless of which trial data was used to develop the models. Compared to the other approaches, all FMM appear superior in fitting the EQ-5D utility data with the IGMM having the smallest errors. The errors of the LME and BM were in-between FMM; and, Tobit and OLS.

Among the response mapping methods, MNL and BNs fitted relatively similar the RECODE data but MNL fitted better the GO-AHEAD sample. However, for both of these response mapping methods, compared to MP, EU provided better accuracy.

For *Table 6-3*, we have calculated MAE and MSE for various EQ-5D intervals. *Table 6-3* shows that IGMM and GMM strongly outperformed the other methods on all EQ-5D intervals.

Figure 6-4 presents observed versus predicted values for the GO-AHEAD trial suggesting that in general FMM provide superior fit compared to the other approaches. Figure 6-5 illustrates the fitted distributions for the GO-AHEAD trial indicating once more that the FMM fit the EQ-5D data substantially better than the other approaches. Similar results for RECODE trial are enclosed in Appendix.

Table 6-2: In-sample validation.

Fitted/Predicted		RECODE (N=5157)/ RECODE (N=5157)[a]						GO-AHEAD (N=366)/ GO-AHEAD (N=366) [b]					
Mapping method	Model	Min	Max	Mean	MAE	MSE	R^2	Min	Max	Mean	MAE	MSE	R^2
Utility mapping	Data	-0.594	1.000	0.706	-	-	-	-0.594	1.000	0.615	-	-	-
	OLS	0.018	0.886	0.706	0.179	0.060	0.31	0.093	1.001	0.615	0.173	0.050	0.42
	LME	-0.037	1.010	0.706	0.125	0.030	0.65	-0.028	1.038	0.615	0.122	0.026	0.70
	Tobit	0.011	1.0000.761		0.178	0.065	0.28	0.092	1.000	0.634	0.175	0.052	0.40
	BM	-0.064	0.871	0.700	0.180	0.059	0.32	-0.038	0.938	0.613	0.169	0.049	0.43
	GMM	-0.101	0.911	0.706	0.098	0.014	0.84	-0.103	0.943	0.624	0.089	0.013	0.85
	IGMM	-0.085	1.000	0.706	0.052	0.006	0.95	-0.088	1.000	0.624	0.065	0.009	0.90
	BMM	-0.087	0.902	0.667	0.100	0.018	0.78	-0.058	0.934	0.598	0.082	0.012	0.83
Response mapping	MNL (EU)	-0.085	0.863	0.695	0.181	0.060	0.32	-0.178	0.924	0.604	0.161	0.044	0.49
	MNL (MP)	-0.126	1.000	0.823	0.182	0.079	0.22	-0.213	1.000	0.677	0.166	0.057	0.37
	BN (EU)	0.159	0.833	0.692	0.185	0.061	0.30	0.242	0.869	0.587	0.184	0.054	0.38
	BN (MP)	0.015	1.000	0.815	0.182	0.079	0.21	0.196	1.000	0.721	0.229	0.107	0.10

[a] models fitted and predicted to RECODE trial ; [b] models fitted and predicted to GO-AHEAD trial

Table 6-3: In-sample MAE and MSE by EQ-5D utility tariff intervals.

Interval/Model	RECODE (N=5157) /RECODE (N=5157) [a]							GO-AHEAD (N=366)/GO-AHEAD (N=366) [b]									
	<0.25		0.25-0.5		0.5-0.75		>0.75		<0.25		0.25-0.5		0.25-0.5		>0.75		
	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MSE
OLS	0.486	0.270	0.346	0.155	0.109	0.020	0.138	0.028	0.343	0.142	0.217	0.061	0.122	0.025	0.153	0.038	
LME	0.309	0.123	0.244	0.087	0.094	0.015	0.093	0.014	0.235	0.070	0.159	0.037	0.088	0.013	0.106	0.019	
TOBIT	0.509	0.303	0.388	0.203	0.139	0.030	0.110	0.021	0.346	0.145	0.228	0.069	0.131	0.028	0.139	0.035	
BM	0.653	0.445	0.471	0.242	0.123	0.020	0.090	0.011	0.487	0.259	0.336	0.125	0.105	0.016	0.097	0.015	
GMM	0.086	0.015	0.129	0.020	0.105	0.015	0.097	0.012	0.110	0.019	0.092	0.013	0.064	0.006	0.100	0.015	
IGMM	0.087	0.016	0.156	0.029	0.064	0.006	0.032	0.003	0.141	0.028	0.102	0.015	0.046	0.004	0.049	0.005	
BMM	0.161	0.048	0.180	0.041	0.100	0.014	0.064	0.006	0.144	0.032	0.121	0.024	0.062	0.006	0.068	0.007	
MNL (MP)	0.472	0.258	0.339	0.149	0.109	0.020	0.144	0.030	0.289	0.110	0.212	0.057	0.111	0.022	0.155	0.039	
MNL (EU)	0.594	0.394	0.478	0.268	0.138	0.034	0.093	0.019	0.340	0.170	0.320	0.118	0.121	0.030	0.105	0.025	
BN (MP)	0.489	0.268	0.324	0.141	0.103	0.017	0.153	0.032	0.361	0.156	0.168	0.040	0.119	0.023	0.191	0.049	
BN (EU)	0.616	0.409	0.454	0.246	0.112	0.026	0.103	0.021	0.600	0.388	0.359	0.144	0.128	0.027	0.160	0.071	

[a] models fitted and predicted to RECODE trial ; [b] models fitted and predicted to GO-AHEAD trial

Out-of-sample validation

Table 6-4 shows the out-of-sample measures of predictive performance, i.e. MAE and MSE for models trained with RECODE and validated with GO-AHEAD and for the reverse situation, i.e. for models trained with GO-AHEAD and validated with RECODE.

Table 6-4 shows that, in general, regardless of whether the models were trained on RECODE or GO-AHEAD, when used for predicting external data, the prediction errors produced with utility mapping methods are similar to those produced with response mapping methods. While compared to the other methods investigated, the FMM fitted the EQ-5D utility data substantially better, when used for out-of-sample prediction, they showed only modest improvements. Furthermore, not all FMM developed here showed better prediction performance than the other methods. In fact, in this case, the out-of-sample prediction error of the FMM depended largely on whether the dataset used to train the models included more or less individuals in poor health. Our results showed that FMM performed better when models were trained using a datasets that included a larger number of individuals in poor health.

When the models were developed for RECODE and validated with GO-AHEAD (see *Table 6-4*), MAE showed that, among the utility mapping methods, on average, the prediction error is smallest for beta mixture models (BMM), especially for BMM with maximum probability clustering (MPC). Furthermore, MAE indicates that BNs with the maximum probability (MP) and LME are the second best approaches.

Table 6-4: Out-of-sample validation

Fitted/ Predicted		RECODE(N=5157)/GO-AHEAD(N=366)[a]					GO-AHEAD(N=366)/RECODE(N=5157)[b]				
Mapping method	Model	Min	Max	Mean	MAE	MSE	Min	Max	Mean	MAE	MSE
Utility mapping	OLS	0.045	0.874	0.514	0.206	0.066	0.033	1.001	0.755	0.181	0.067
	LME	0.211	0.829	0.558	0.196	0.058	0.047	0.970	0.739	0.180	0.065
	Tobit	0.044	1.000	0.537	0.203	0.065	0.034	1.000	0.803	0.185	0.074
	BM	0.032	0.871	0.509	0.209	0.067	-0.381	0.966	0.770	0.181	0.069
	GMM (MPC)	-0.074	0.903	0.593	0.204	0.085	-0.192	0.976	0.812	0.187	0.079
	GMM (WA)	0.100	0.863	0.520	0.213	0.067	0.072	0.965	0.773	0.176	0.067
	IGMM (MPC)	-0.061	1.000	0.606	0.199	0.082	-0.197	1.000	0.828	0.187	0.084
	IGMM (WA)	0.142	0.888	0.521	0.216	0.067	0.142	0.975	0.79	0.178	0.069
	BMM (MPC)	-0.103	0.918	0.677	0.187	0.073	-0.334	0.954	0.778	0.181	0.073
	BMM (WA)	0.270	0.895	0.595	0.191	0.056	0.012	0.948	0.737	0.179	0.063
Response mapping	MNL (EU)	0.000	0.856	0.499	0.214	0.071	-0.192	0.930	0.751	0.181	0.068
	MNL (MP)	-0.074	1.000	0.619	0.205	0.079	-0.323	1.000	0.829	0.193	0.088
	BN(EU)	0.158	0.833	0.518	0.209	0.065	0.242	0.869	0.730	0.178	0.063
	BN(MP)	-0.056	1.000	0.652	0.195	0.071	0.015	1.000	0.753	0.269	0.138

[a] models fitted and predicted with RECODE and predicted with GO-AHEAD trial ; [b] models fitted with GO-AHEAD and predicted to RECODE trial

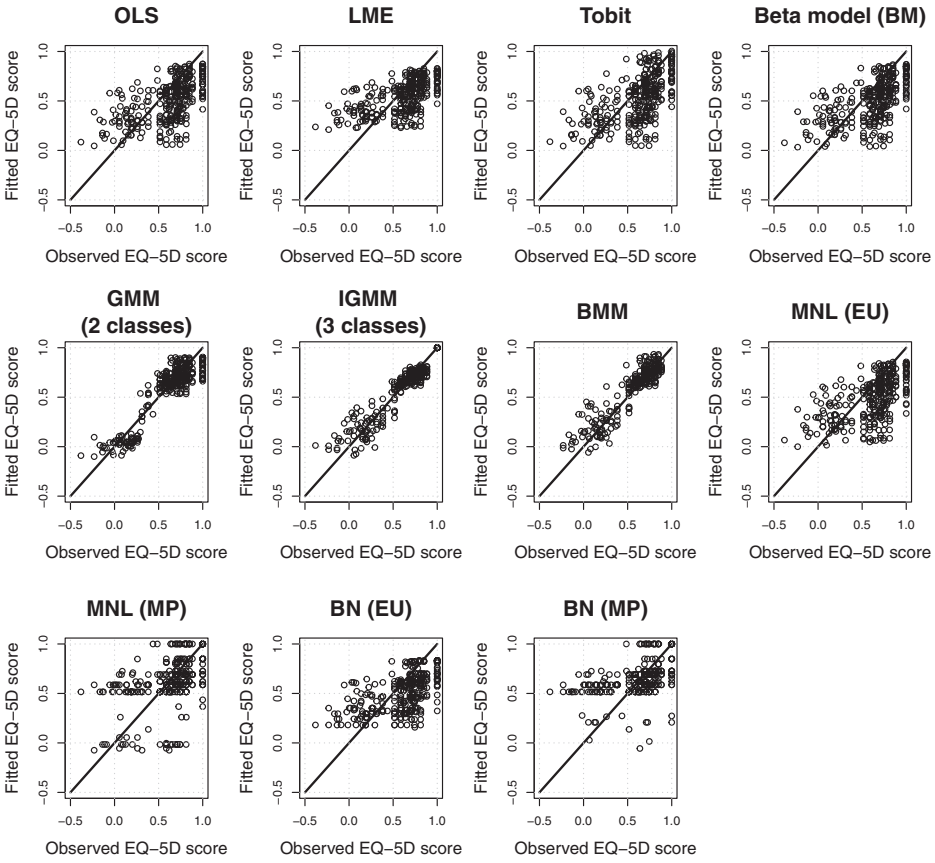

Figure 6-4: Observed versus predicted EQ-5D scores for GO-AHEAD trial

When models were developed with GO-AHEAD and validated with RECODE, MAE indicates that, on average GMMs are superior to both the other utility mapping as well as response mapping methods. Furthermore, BNs resulted in very similar errors to these. In fact, compared to utility mapping, response mapping methods (i.e. MNL and BNs) seem to benefit more if the dataset used to develop such models includes a large number of respondents in good health.

In general, we found that WA and EU were superior to MPC and MP. One explanation for this may be because the misclassification error is more present for individuals in poor health as that observed sample size is the smallest in both trials.

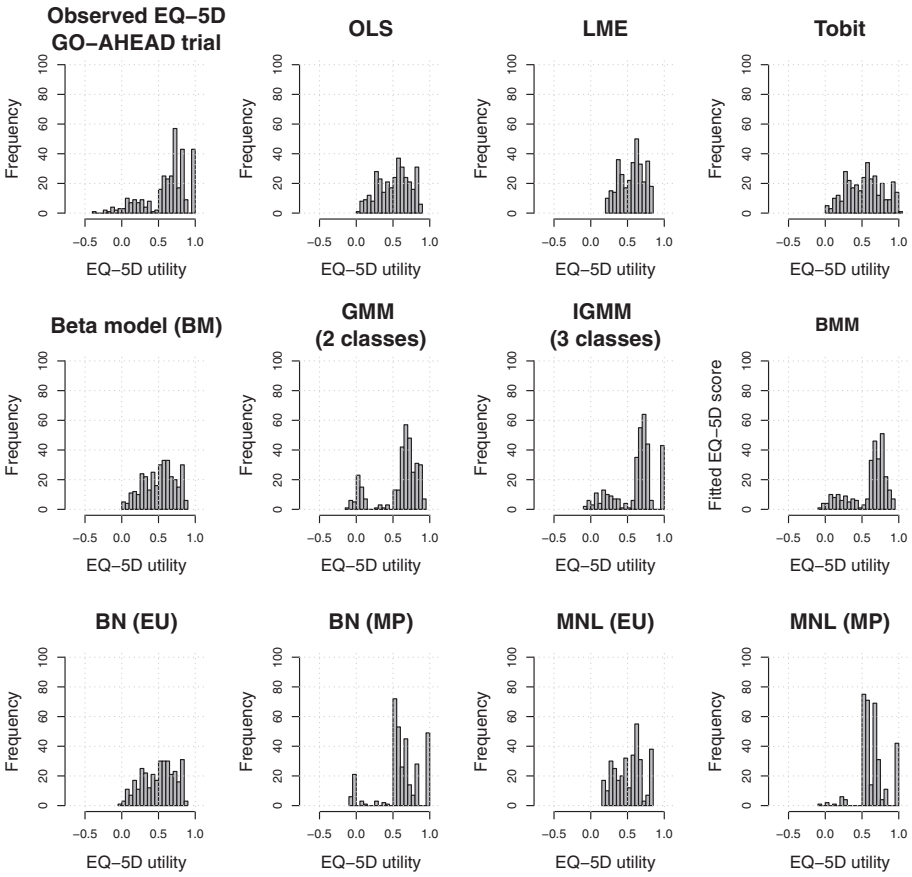

Figure 6-5: Fitted distributions for the GO-AHEAD trial

Table 6-5 shows that there is not one single method that predicts the EQ-5D data better than the others on all EQ-5D intervals. Furthermore, we observe that regardless of the trial data used to learn the model, BNs tend to perform better when EQ-5D is defined between 0.25 and 0.75 while FMM perform better for individuals that are in good health (i.e. when EQ-5D>0.75). Regarding those that are in the poorest health (i.e. EQ-5D<0.25) all methods result in substantially higher errors than those for the other intervals.

Table 6-5: Out-of-sample MAE and MSE by EQ-5D utility tariff intervals

Interval/ Model	RECODE (N=5157)/GO-AHEAD (N=366) [a]								GO-AHEAD (N=366)/RECODE (N=5157) [b]							
	<0.25		0.25-0.5		0.5-0.75		>0.75		<0.25		0.25-0.5		0.5-0.75		>0.75	
	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
OLS	0.246	0.089	0.144	0.033	0.191	0.059	0.223	0.071	0.561	0.345	0.396	0.192	0.111	0.019	0.119	0.021
LME	0.337	0.136	0.156	0.037	0.143	0.033	0.212	0.060	0.549	0.331	0.379	0.177	0.103	0.017	0.127	0.023
TOBIT	0.249	0.092	0.157	0.042	0.200	0.062	0.195	0.062	0.584	0.378	0.433	0.237	0.138	0.029	0.103	0.018
BM	0.244	0.087	0.144	0.033	0.194	0.061	0.228	0.073	0.572	0.362	0.417	0.210	0.122	0.022	0.109	0.018
GMM (MPC)	0.372	0.190	0.334	0.116	0.171	0.076	0.144	0.044	0.636	0.442	0.485	0.252	0.140	0.028	0.091	0.012
GMM (WA)	0.269	0.103	0.138	0.030	0.196	0.058	0.226	0.072	0.581	0.368	0.420	0.205	0.111	0.018	0.105	0.016
IGMM (MPC)	0.398	0.208	0.328	0.112	0.145	0.064	0.155	0.042	0.635	0.441	0.506	0.281	0.137	0.032	0.091	0.018
IGMM (WA)	0.282	0.107	0.134	0.028	0.196	0.056	0.230	0.073	0.587	0.377	0.433	0.222	0.122	0.022	0.090	0.015
BMM (MPC)	0.423	0.255	0.332	0.123	0.145	0.043	0.108	0.022	0.598	0.390	0.465	0.233	0.115	0.025	0.104	0.016
BMM (WA)	0.384	0.173	0.184	0.047	0.135	0.027	0.183	0.047	0.520	0.307	0.382	0.181	0.119	0.022	0.122	0.023
MNL (EU)	0.234	0.081	0.153	0.035	0.203	0.067	0.234	0.079	0.551	0.340	0.393	0.193	0.118	0.023	0.118	0.023
MNL (MP)	0.382	0.191	0.278	0.092	0.159	0.059	0.164	0.048	0.618	0.429	0.477	0.275	0.167	0.043	0.099	0.021
BN(EU)	0.277	0.103	0.118	0.023	0.179	0.050	0.239	0.077	0.548	0.324	0.371	0.167	0.086	0.012	0.133	0.025
BN(MP)	0.465	0.247	0.254	0.074	0.115	0.028	0.160	0.043	0.653	0.476	0.464	0.257	0.198	0.052	0.207	0.101

[a] models fitted and predicted with RECODE and predicted with GO-AHEAD trial ; [b] models fitted with GO-AHEAD and predicted to RECODE trial

Discussion And Conclusions

This chapter aimed to predict the most often used preference-based PROMs, the EQ-5D from the CCQ, a questionnaire including ten questions specific for COPD patients. For this purpose, we presented an extensive comparison among previously proposed mapping approaches including the commonly used mapping models such as OLS, LME, Tobit model and MNL as well as some of the new and promising approaches such as BNs and FMM. Furthermore, these methods are classified into utility mapping methods and response mapping. Therefore, this chapter compared eight model classes and a total of 14 models. The focus here was on out-of-sample prediction by using two independent trial datasets that differ in sample size and the extent of which they included individuals in poor health.

Various conclusions regarding the performance of the investigated methods can be drawn from this chapter:

- In general, regardless of the dataset used to develop the models, utility score mapping methods fitted the EQ-5D data better than response mapping approaches. Clearly, as shown by the in-sample prediction errors, the FMM provided the best fit to the EQ-5D data. The second best method was LME while the other utility mapping such as OLS, Tobit and Beta regression resulted in similar results with the response mapping methods (i.e. MNL and BNs)
- Out-of-sample performance of the different methods depended on whether the EQ-5D data included a large number of participants in poor and/or good health. In general, models trained on a dataset that included a larger number of respondents in poor health resulted in smaller out-of-sample prediction errors
- For both trials considered here, FMM provided the smallest out-of-sample prediction errors. However, compared to the other approaches, the improvements obtained with these models were rather modest. This was mainly because the misclassification error for FMM was larger for those respondents in poor health.
- BNs resulted in similar prediction errors with FMM when the models were trained on the dataset that had more respondents in poor health.
- In general, response mapping appeared more appropriate when using a dataset with more respondents in poor health while FMM resulted in more accurate predictions for the reverse situation
- Regardless of the dataset used, none of the methods outperformed the other approaches on the entire range of EQ-5D data, with errors being substantially larger for individuals in the poorest heath (i.e. when EQ-5D<0.25).

Similar to previous research (Longworth et al. 2014, Versteegh et al. 2012), we found that all the mapping methods investigated over-predicted the EQ-5D at the bad health states: both utility mapping methods such as OLS, Tobit, LME, BM, FMM and the response mapping such as MNL and BNs had substantially higher errors for EQ-5D intervals lower than 0.5. In addition, we found that BNs tend to perform better in predicting the EQ-5D data in the middle of the EQ-5D interval (i.e. 0.25<EQ-5D<0.75) while FMM tend to perform better when EQ-5D≥0.75.

There is conflicting evidence from previous research regarding the prediction performance of FMM when modeling the EQ-5D data. While some showed that Gaussian mixture models outperform methods such as OLS and Tobit (Hernandez Alava et al. 2012, Hernandez Alava et al. 2014, Coca Perraillon et al. 2015), others showed that they did not demonstrate superior predictive accuracy compared to MNL when applied to external datasets (Kent et al. 2015). In addition, previous research indicated that FMM may be particularly useful for populations in good health (Coca Perraillon et al. 2015). Our findings are to some extent in line with this in that we found FMM to be on average superior to the other mapping methods but that was mainly due to their better accuracy for individuals that were in good health.

Furthermore, we found that, BM models performed better than OLS, Tobit, MNL and BNs but only in-sample and not out-of-sample. With respect to their in-sample performance our results are similar to previous findings that indicated BM to be superior to OLS and Tobit models when used in mapping exercises (Khan and Morris 2014).

Our finding that MNL did not necessarily perform better than OLS especially when these were developed using a dataset with a small number of respondents in poor health is similar to what others showed (Longworth et al. 2014, Brazier et al. 2010). Furthermore, previous research illustrated that BNs are superior to OLS, Tobit and MNL (Le and Doctor 2011, Borchani et al. 2012). To some extent, our results are in line with this since we found that BNs, especially BNs using the EU method, produced predictions with better accuracy than OLS, LME, Tobit, BM, MNL; and similar errors as FMM, when the model was trained on the dataset that included more respondents in poor health. Therefore, response mapping methods, particularly BNs, may be suitable when the mapping models are developed using populations in poorer health. Further research investigating this would be worthwhile.

There are a number of areas in which the work reported here could be extended. First, the methods are compared using point by point error estimates such as MAE and MSE and the uncertainty surrounding the compared methods was not reported here. It has

been suggested that reporting mean errors such as MAE and MSE may not be ideal since for providing suitable measures of uncertainty predictions should be probabilistic, i.e. they should take the form of probability distributions over future quantities (Dawid 1984). Such an approach would be more suitable to a Bayesian estimation and may be worthwhile for future research. Second, the aim here was to compare the existing mapping methods especially when used to predict external EQ-5D data; therefore, we did not consider the problem of verifying what sets of covariates is preferable or what is the optimal functional form of these covariates. This issue may be considered in future research.

The main advantage of this chapter is with respect to the availability of two distinct trial datasets that differed in their sample size and EQ-5D distributions. This enabled to investigate the performance of various methods in external dataset and to test their robustness to various EQ-5D distributions. In this way, more insights regarding the performance of the currently existing mapping methodologies with different datasets can be gained. Furthermore, although the variety of methods investigated here was tested in the context of a mapping exercise, our results may be useful for other health care applications that use EQ-5D data (e.g. for monitoring population health as shown in chapters 2, 3 and 4).

Concluding, this research shows that, when used for external EQ-5D predictions, none of the currently existing mapping methodologies produces the most accurate predictions for the entire range of the EQ-5D data regardless of the sample population used to train the model. A number of recommendations follow from this research. First, FMM may be suitable when the sample population used to develop the mapping model includes a large number of respondents in good health as the misclassification rate is lower in that situation. Second, response mapping, particularly Bayesian Networks (BNs) may be suitable when the sample population used to train the model includes a large number of individuals in poor health. The current chapter represents a first step in providing an extensive comparison of the currently existing mapping methodologies and identifying their differences when used with various datasets. Obviously, translating our results into definite answers regarding the suitability of the current mapping methods for disease-specific questionnaires in general requires further research. Therefore, an important recommendation that follows from this chapter is to test the performance of the above methods using other disease-specific questionnaires.

Appendices

A1: The structure of the Bayesian Models

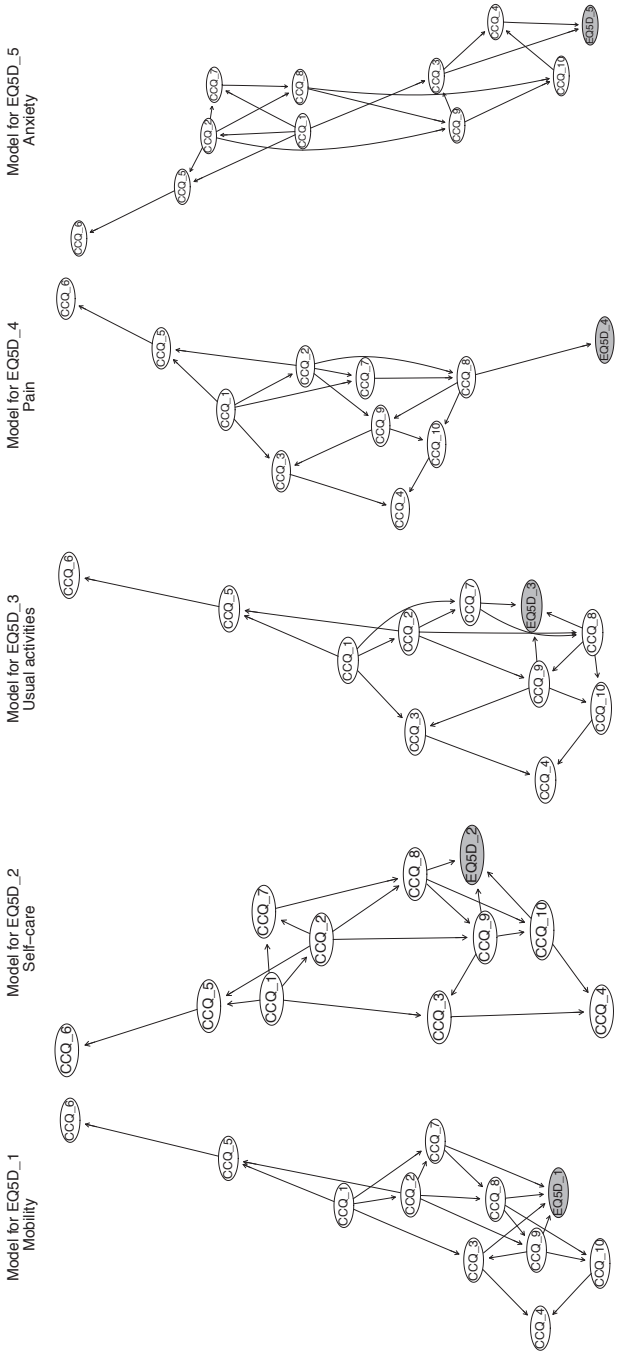

Figure 6-6: BNs learned model structures

A2: Observed versus fitted EQ-5D scores for RECODE trial

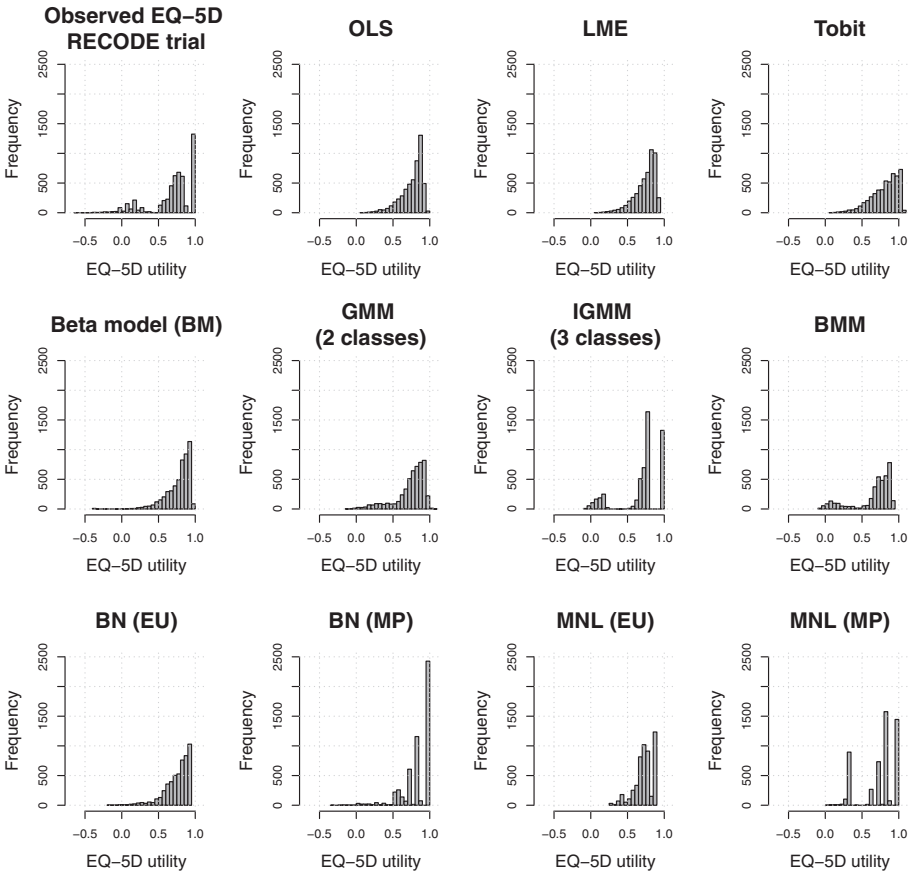

Figure 6-7: Fitted distributions for the RECODE trial

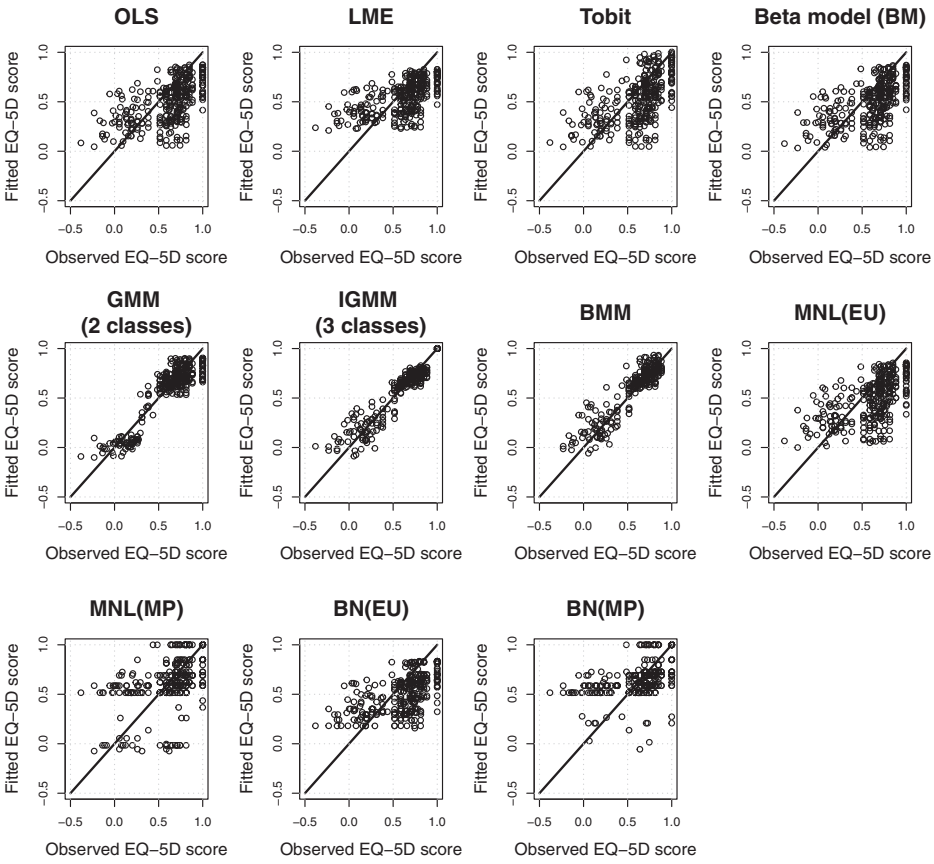

Figure 6-8: Observed versus predicted EQ-5D scores for RECODE trial

A3: R code for the driver extending the flexmix package for fitting the Inflated Gaussian mixture model (IGMM)

```
library("flexmix")
setClass("FLXMRmgglm", representation(mean = "numeric", sigma = "ANY"), contains="FLXMRglm")

FLXMRmgglm <- function(formula = . ~ ., mean, sigma = NULL, ...) {
  new("FLXMRmgglm", FLXMRglm(formula, family = "gaussian", ...),
    name = paste("FLXMRmgglm", "gaussian", sep =":"), mean = mean, sigma = sigma)
}
setMethod("FLXgetModelmatrix", signature (model ="FLXMRmgglm"),function (model, data, formula,
lhs=TRUE, ...)
  {
    model <- callNextMethod (model, data, formula, lhs)
    if (attr (terms (model@fullformula), "intercept") == 0)
      stop ("please include an intercept")
    model
  })
setMethod ("FLXremoveComponent", signature(model = "FLXMRmgglm"), function (model, nok, ...)
  {
    if (1 %in% nok) as(model, "FLXMRglm") else model
  })
setMethod("FLXmstep", signature (model = "FLXMRmgglm"), function (model, weights, components, ...)
  {
    coef <- c(model@mean, rep(0, ncol(model@x) - 1))
    names(coef) <- colnames(model@x)
    sigma <- if (is.null(model@sigma)) sqrt(mean(weights[, 1] * (model@y - model@mean)^2)/mean(weights[, 1]))
        else model@sigma
    comp.1 <- with(list(coef = coef, sigma = sigma, df = sum(is.null(model@sigma)), offset = NULL,
              family = model@family), eval(model@defineComponent))
    c(list(comp.1),
     FLXmstep(as(model, "FLXMRglm"), weights[, -1, drop = FALSE],
          components[-1]))
  })
### model fit

Model <- FLXMRmgglm(mean = 1, sigma = 0.0001)

mod_rec<-flexmix(qol~CCQ_1+CCQ_2+CCQ_3+CCQ_4+CCQ_5+CCQ_6+CCQ_7+CCQ_8+CCQ_9
+CCQ_10,concomitant=FLXmultinom(~CCQ_1+CCQ_2+CCQ_3+CCQ_4+CCQ_5+CCQ_6+CCQ_7
CCQ_8+CCQ_9+CCQ_10), k = 3, data =data,model = Model, control = list(minprior = 0, iter.max = 100))
```

General discussion

Findings Of This Thesis

The aim of this dissertation was to investigate changes in population health, to explore a mechanism that can explain these changes and to assess their consequences and implications for economic evaluations of health care interventions. In doing so, several methodological challenges associated with modelling HRQoL data were addressed.

The following sections present a discussion of our main findings and answer the research questions posed in the introduction. The research questions can be grouped in the following themes: changes in population health, the relationship between HRQoL and time to death (TTD), and methodological considerations when modelling HRQoL data.

Changes in population health

In an aging society, it is important to establish whether increases in life expectancy are accompanied by concomitant improvements in health. The results presented in chapters 2 and 3 showed that Dutch people are not only tending to live longer than they previously did but also enjoy a higher quality of life than before. In addition, our results indicate that, although the overall trend is found in all population groups, the higher educated gain more in length and quality of life than the lower educated, thus widening the socio-economic gap in health.

The empirical results presented in chapter 2 showed that the health of the Dutch population as measured by quality-adjusted life expectancy (QALE) has improved in the period 2001 to 2008. This suggests that, for the Dutch population, the observed declines in mortality rates over recent periods were accompanied by concomitant improvements in HRQoL at all ages. For example from 2001 to 2008, QALE for a man/woman aged 20 years increased by 2.3/1.9 healthy years, of which 0.6/0.8 was due to HRQoL improvements.

Chapter 3 showed that in the period 2001 to 2011, both life expectancy (LE) and QALE have increased more for highly educated individuals compared to low and medium educated individuals. For example, in 2001, at age 25, the absolute QALE difference between the low and the highly educated was 7.4 healthy years (36.7 vs. 44.1) for men and 6.3 healthy years (39.5 vs. 45.8) for women. By 2011, the QALE difference had increased to 8.1 healthy years (38.8 vs. 46.9) for men and to 7.1 healthy years (41.3 vs. 48.4) for women. Similar inequalities in health have been observed at older ages. Therefore, our results show that, at least in the Netherlands, in recent periods, population health as measured by QALE has improved in all groups, but, nonetheless, QALE inequalities have widened, even more so than inequalities in life expectancy alone.

The relationship between HRQoL and TTD

In this thesis the relationship between HRQoL and TTD has been investigated in two contexts. First, in the context of population aging, we investigated whether the relationship between HRQoL and TTD may explain the observed changes in population health by age. Second, in the context of economic evaluations, we explored the potential consequences and implications of the relationship between HRQoL and TTD for economic evaluations of life prolonging interventions. In other words, we investigated how the relationship between HRQoL and TTD can be exploited to improve estimates of QALY gains in economic evaluations of life prolonging interventions. These two issues have been studied in chapters 4 and 5, respectively.

Applying statistical methods to a dataset for the Dutch population, chapter 4 showed that the observed relationship between health as measured by HRQoL and age can be explained to a large extent by the relationship between HRQoL and TTD. Our empirical results indicate that when TTD is accounted for, the effect of age on HRQoL becomes negligible. Hence, HRQoL losses induced by decreasing TTD are substantially larger than those induced by increasing age. These findings suggest that further increases in life expectancy will not necessarily result in more years spent in poor health. Therefore, to a large extent, this mechanism is aligned with the observed health improvements for the Dutch population, as indicated in chapters 2 and 3 of this thesis.

Chapter 5 showed that the relationship between HRQoL and TTD has important implications for estimating gains in Quality Adjusted Life Years (QALYs) in economic evaluations of life prolonging interventions. The fact that HRQoL depends strongly on TTD and that health losses are centred in the last phase of life can be used in estimating QALYs gained due to life prolonging interventions. That is because interventions that extend life, to a large extent postpone the HRQoL losses towards the end of life. Hence, not all years gained due to a life prolonging intervention may be spent in poor health, but mainly the last years of life. In chapter 5 we found that ignoring the relation between HRQoL and TTD would result in an underestimation of QALY gains due to life prolonging interventions and, consequently, to an overestimation of the cost-effectiveness ratio compared to the situation in which this relationship is adequately accounted for. We found that the level of this underestimation of QALY gains ranges between 3% and 7% and depends mostly on the discount rate used.

Statistical modelling of HRQoL data

Due to its non-standard distribution (bounded, skewed, heteroskedastic and with discontinuity points) modelling HRQoL may be complicated. Therefore, careful consideration needs to be taken when modelling the various HRQoL instruments. In this thesis

we modelled two generic preference-based HRQoL indices: the SF-6D (derived from the SF-12 and SF-36) and the EQ-5D, respectively. To some extent, each chapter of this thesis posed the question which methods should be used for modelling the HRQoL data.

Chapters 2, 3 and 5 modelled cross-sectional SF-6D data using beta regression applied within a generalized additive models for location, scale and shape (GAMLSS) framework. This approach enabled modelling both parameters of the beta distribution (i.e. location and precision) as function of explanatory variables, therefore allowing to address some of the issues associated with modelling HRQoL data such as heteroscedatisicty. To model longitudinal SF-6D data, chapter 4 used a mixed beta regression and highlighted the benefits of using the Bayesian estimation. Chapter 6 compared a large number of currently used methodologies for modelling EQ-5D data.

The results showed in chapters 2, 3 and 5 indicate that GAMLSS and beta distribution are suitable for modelling cross-sectional SF-6D data. The results presented in Chapter 4 showed that mixed beta regression and Bayesian estimation is appropriate for modelling longitudinal HRQoL data. Our findings showed that the beta distribution enabled fitting the non-standard SF-6D outcome while Bayesian estimation allowed estimating complex models that would otherwise be more difficult to estimate. For example, Bayesian estimation permits to make use of prior information. Furthermore, with a Bayesian approach it was possible to straightforwardly account for various sources of bias specific to HRQoL data in observational studies, such as non-ignorable missing data. Note that, although the benefits of using Bayesian estimation were highlighted here in a case study that modelled the longitudinal SF-6D outcome, they apply largely to other HRQoL instruments such as the EQ-5D and to situations in which HRQoL data may be cross-sectional rather than longitudinal.

Chapter 6 compared a large number of existing methods for developing a prediction model that allows mapping EQ-5D outcome from a disease specific-questionnaire. By using a dataset that included both the EQ-5D and a disease-specific questionnaire, a model was developed that can be used in future studies to predict generic HRQoL at the patient level when this data is unavailable. The findings from this chapter showed that, compared to the other existing methods currently in use for modelling EQ-5D data such as OLS, Tobit, beta regression, multinomial logit models and Bayesian networks, finite mixture models (FMM) appear to fit best the EQ-5D data. However, we found that, when used for out-of-sample prediction, these methods showed only slight improvements compared to the other considered approaches. Our empirical results indicated that, when used for external predictions, none of the investigated methods outperforms the others on all EQ-5D intervals.

Synthesis Of Findings

Various connections between the results obtained in the chapters of this thesis can be established. First, it is interesting to investigate to what extent the relationship between HRQoL and TTD (as shown in chapter 4) can be used to estimate the changes in population health as measured by QALE over recent calendar periods in the Netherlands (as shown in chapters 2 and 3). We calculated that about 90% and 70% of the QALE gain in males and females aged 20, respectively (as highlighted in chapter 2) can be explained by using the relationship between HRQoL and TTD as shown in chapter 4. Furthermore, we found that in this QALE gain, about 50% and 25% was due to HRQoL improvements for men and women, respectively. Therefore, to a large extent, the results obtained in the descriptive studies illustrated in chapters 2 and 3 can be deduced by using solely the mechanism presented in chapter 4.

Second, chapters 4 and 5 investigated the relationship between HRQoL and TTD using different datasets. Therefore, it is interesting to observe whether the results from these two studies are consistent. The model developed in Chapter 4 showed that the TTD coefficient was 0.019 [0.010, 0.036] while that developed in Chapter 5 indicated that the TTD coefficient estimate was 0.06 [0.012, 0.108]. Note that in Chapter 4 we performed a Bayesian analysis for investigating the relationship between HRQoL and TTD. In doing so, the results of the TTD coefficient from chapter 5 have been used in the form of prior information in the Bayesian analysis performed in chapter 4; the combined results for the TTD coefficient was 0.018 [0.005,0.024]. Therefore, we conclude that, the results reported in chapter 5 did not appear to influence much those obtained in chapter 4.

Limitations

Each chapter of this thesis includes a specific list of limitations. We will not repeat those here. Instead, we will highlight several general limitations. First, it has been acknowledged that all chapters of this thesis used datasets containing information about Dutch individuals. Therefore, the empirical results presented here are particularly relevant for the Netherlands. While clearly the results illustrating descriptive changes in population health as indicated in chapters 2 and 3 are relevant for the Dutch population only, the mechanism describing the relationship between HRQoL, age and TTD can be generalized to other populations as well. It is likely that a similar relationship between HRQoL and TTD will be observed in other European countries as well. In fact, a similar mechanism has been demonstrated for healthcare expenditures using data sets from Switzerland

(Zweifel et al. 1999), the United Kingdom (Seshamani 2004, van Baal and Wong 2012a, Seshamani and Gray 2004) and the Netherlands (van Baal and Wong 2012a).

Second, for investigating changes in population health, most of the chapters of this dissertation (chapters 2-5) use the SF-6D HRQoL instrument. Note that SF-6D scores range between 0.345 and 1. Therefore, when compared to other HRQoL instruments such as EQ-5D or HUI, which typically exceed the boundaries of the interval defined between zero and one, it is more difficult to detect small changes in health with SF-6D. It is possible that if HRQoL instruments defined on a wider range would have been used instead of the SF-6D (e.g. EQ-5D), larger health changes over time and by age would have been observed. Therefore, due to these range discrepancies between various HRQoL indices, the empirical results presented here may be sensitive to the HRQoL instrument used.

Implications And Future Research

The presented results in this dissertation are relevant from both a theoretical and a practical point of view and can be useful to researchers, policy makers and to improve the methodology of economic evaluations of health technologies.

We hope that the insights of this thesis will inspire other researchers in the field of health economics to conduct further investigations in some of the areas covered in this thesis. For example, investigating changes in population health and the mechanism of the relationship between TTD and HRQoL using empirical datasets relevant for other populations besides that of the Netherlands would be worth pursuing. The proposed approach may be adapted to any life prolonging interventions also in different populations (e.g. diseased populations or populations of a specific socio-economic group) if data is available to estimate the impact of TTD on HRQoL. That is because TTD can be viewed as a proxy variable for diseases and processes that increase mortality risk and that are not explicitly modelled in an economic evaluation. By using TTD when estimating QALY gains due to an intervention that extends life, we implicitly account for these other unobserved diseases in the added years of life. Such an adjustment is possible as all modelling studies of life prolonging interventions include information on mortality. In doing so, using other HRQoL measures defined on wider intervals, for example EQ-5D or HUI, would lead to better insight into the role of choice of HRQoL instrument.

From a policy perspective, our results can be considered important in the context of population aging. Our finding that Dutch people live longer lives in better health indicate that, in recent periods in the Netherlands, there has been a compression of mor-

bidity (Fries 2000, Fries 2002). This may be important in alleviating the consequences of population aging in various contexts. One example refers to the ongoing debates regarding extended labour-force participation by the elderly and raising of the pension age. According to a press release made by Statistics Netherlands, in the Netherlands, the Dutch government decision to raise the current retirement age from 65 years to 67 years (which should be the case in 2022) will reduce the number of pensioners by half a million by 2025 (Statistics Netherlands 2014). The same document reports that from 2022 onwards the entitlement age for state old-age pension in the Netherlands will be linked to the increase in life expectancy. However, these social policies alone may be insufficient for reducing the financial burden induced by population aging if individuals chose to cease work due to ill-health. Previous research showed that health is the main determinant of labour supplyof older workers. For example, Lindeboom, in a comprehensive review, argued that a number of studies indicated that poor health is the main cause of labour force exit among older workers (Lindeboom 2012). Therefore, the results from chapters 2 and 3, showing that in recent periods the health of the Dutch population has improved, are useful for informing reforms targeted at systematically increasing the official retirement age in the Netherlands. Nonetheless, in chapter 3 it was argued that such restructuring should acknowledge the differences in health between socio-economic groups. Chapter 3 showed that health inequalities favouring the higher educated exist at all ages. However, such differences are rarely acknowledged by policy makers. In our view efforts should be made in diminishing health inequalities and perhaps in studying the possibility of making the retirement age flexible by allowing for differential retirement age by educational or socio-economic status (De Waegenaere et al. 2014, Bovenberg et al. 2006).

In chapter 4 it has been shown that health losses are concentrated in the last phase of life. This confirmed the results obtained in the descriptive studies presented in chapters 2 and 3. They imply that in the Netherlands there is a compression of morbidity. These results may also be important in explaining the demand for healthcare. Given that in many OECD countries aging accelerates, there are fears that HCE may drastically increase. Our results indicating that health losses are centred in the last phase of life suggest that this may not necessarily be the case. This finding is in line with previous research showing that similar to health losses, the demand for health care use is concentrated in the last phase of life because individual HCE can be better explained by TTD than age (Zweifel et al. 1999, Seshamani 2004, Zweifel et al. 2004, Werblow et al. 2007). This is to be expected since both age and TTD act as proxies for morbidity and disability which are the real drivers of individual HCE. Note that the majority of studies investigating the effect of TTD on HCE used individual level data but aimed at informing healthcare use at the macro level. For example, using individual level data, several studies have shown that if

life expectancy increases, excluding TTD from the analyses results in an overestimation of macro-level HCE (Polder et al. 2006, Stearns and Norton 2004). However, these results were subject to critique as they did not account for the difference in determinants of individual health care use and macro HCE (Getzen 1992, Getzen 2000). Using macro-level HCE data and accounting for growth factors by unidentified causes (e.g. medical technology advances), it has been shown that, including TTD in forecast models does not lower future HCE projections (van Baal and Wong 2012b). Similar to future HCE projections, TTD may be used for projecting future population health, e.g. for forecasting QALE. Note that unlike for healthcare costs, the disctinction between micro and macro HRQoL is not relevant when forecasting population QALE.

In our view, by stimulating the compression of morbidity, to some extent, the consequences of population aging may be alleviated. Examples of such economic benefits include the fact that people may be able to remain in the labour force for longer and that individual HCE may increase less markedly. Given that nowadays morbidity is especially driven by the presence of chronic diseases (e.g. diabetes, cancer), governments may seek to enforce programs that enable the postponement of onset of these chronic diseases though strategies such as screening and prevention.

The evidence provided in this thesis illustrates two issues with implications for the practice of economic evaluations. First, chapter 5 showed that, compared to the situation in which age- and gender-specific HRQoL would be used, ignoring the relationship between HRQoL and TTD results in an underestimation of QALY gains (and consequently an overestimation of the ICER) due to life prolonging interventions. This is in line with previous research chaptering the relationship between other health measures such as healthcare expenditures (HCE) and TTD, and its implications for life prolonging preventive interventions. It has been shown that ignoring the effect of TTD on HCE would result in an overestimation of the costs (and consequently of the ICER) in those economic evaluations (Gandjour and Lauterbach 2005). With respect to the estimation of QALY gains due to life prolonging interventions, our results highlight that the assumption made regarding HRQoL in life years gained due to these interventions has a direct impact on the final cost-effectiveness results. Nevertheless, often economic evaluation analysts assume that the absence of the disease under investigation equals people being in full health. The lack of standards and guidelines regarding the estimation of HRQoL in life years gained due to life prolonging interventions leaves room for such unrealistic assumptions in the daily practice of economic evaluations. In our view, a set of guidelines for including the effect of competing risks in the added years of life due to life prolonging interventions is required. The approach illustrated in this thesis, i.e.

using HRQoL stratified by age, gender and TTD in life years gained due to life prolonging interventions, can be used to assist such guidelines.

Second, in practical economic evaluations, methodological recommendations regarding HRQoL can be useful, for example, for developing prediction or 'mapping' models. In this context, using inappropriate methodologies for modelling EQ-5D (or other HRQoL) data can have a significant impact on the final outcomes of an economic evaluation. Regarding the methods used to model HRQoL data, our findings indicate that there is no all-purpose best method. For example, certain methods may be more suitable for modelling SF-6D than EQ-5D data. This highlights the need for a clear, case by case justification of methodological choices in this context. Furthermore, analysts may consider whether the purpose of the model is to *explain* or to *predict* HRQoL data. In addition, when developing HRQoL models, analysts should be aware of and preferably account for the various sources of bias which may occur, e.g. selection bias due to missing data.

Concluding, this thesis explored a wide range of quantitative approaches for monitoring changes in population health while emphasising how these changes may be used to improve health estimates in economic evaluations of life prolonging interventions. All in all, we hope that this thesis contributes to the health economics literature by enabling a better understanding of the consequences of aging on population health, by improving the methodology of economic evaluations, and by highlighting a number of statistical methods for modelling HRQoL data in various contexts.

Summary

Nowadays, across the world people are living longer than previous generations did. The proportion of elderly in the total population has also increased, a demographic phenomenon commonly referred as population aging. Currently, people aged 60 and older make up over 11 per cent of the global population. It is expected that by 2050, this number will double to about 22 per cent. These changes in the structure of their populations pose new challenges and require nations to be equipped with the right set of economic policies. In that context, extensive societal, political and scientific debates have been associated with population ageing. Important examples include debates on the affordability of growing healthcare expenditures (HCE), and those on raising the statutory retirement age in order to increase labour force participation of elderly. The rationale for and exact consequences of policy decisions in these areas crucially depend on the extent to which the increases in life expectancy are accompanied by concomitant increases in life years spent in good health. Therefore, monitoring the level of population health and its changes over time and by population subgroups is a key component for determining whether: (a) policy changes associated with population aging are necessary and (b) if necessary, these policy changes will have the desired societal effects. The aim of this thesis was to investigate changes in population health and the way in which these changes may be included in economic evaluations. In doing so, health-related quality of life (HRQoL) was used as a measure of population health. Estimating population health using HRQoL is challenging because of the non-standard distribution (bounded, skewed, heteroskedastic and with discontinuity points) of the HRQoL outcome. To some extent, each chapter of this thesis posed the question which methods should be used for modelling the HRQoL data. Therefore, this thesis, investigates various statistical methods for modelling different types of HRQoL data.

In ***Chapters 2 and 3*** we studied changes in the health (as measured by HRQoL) of the Dutch population, in the period between 2001 and 2008, for various population subgroups. We found that Dutch people are not only tending to live longer than they previously did but are also living more healthily. In addition, Chapter 3 showed that, although this seems to be the case for all population subgroups, compared to the lower educated, the higher educated Dutch people profited most in terms of longevity and years in good health. In other words, between 2001 and 2008 health has improved for all population groups in the Netherlands, but more for the highly educated. This indicates that health inequalities by education widened for the Dutch population. Our finding that Dutch people live longer lives in better health indicates that, in recent periods in the Netherlands, there has been a compression of morbidity. This is important in assessing the consequences of population aging in various contexts. One example refers to the ongoing debates regarding extended labour-force participation of the elderly and raising the pension age.

It has been documented that, with advancing age, health deteriorates and healthcare expenditures (HCE) increase. This could imply that increases in life expectancy increase the number of years lived in poor health, which may limit the scope for extending working lives and increase HCE. However, such seemingly obvious conclusions may be misleading. It has been extensively shown that healthcare utilization is centred in the final years of life and, therefore, population aging may have limited effect on future healthcare use. That is because time to death (TTD) is a stronger predictor of HCE than age is. In explaining the observed changes in population health as indicated in Chapters 2 and 3, this thesis explores a mechanism that investigates the relationship between HRQoL and TTD. ***Chapter 4*** showed that the observed relationship between health as measured by HRQoL and age can be explained to a large extend by the relationship between HRQoL and TTD. Similar to HCE, we found that when TTD is accounted for, the effect of age on HRQoL becomes negligible. Hence, HRQoL losses induced by decreasing TTD are substantially larger than those induced by increasing age. The fact that we found health losses to be centred in the last years of life has important implications for the challenges commonly related to population aging. Indeed, these results suggest that further increases in life expectancy will not necessarily result in more years spent in poor health.

Using a different dataset, ***Chapter 5*** confirmed that the observed changes in HRQoL by age can be explained by the relationship between HRQoL and TTD. Furthermore, we found that this relationship has important implications for estimating health gains in economic evaluations of life prolonging interventions. That is because interventions that extend life, to a large extent postpone HRQoL losses towards the end of life. Hence, not all years gained due to a life prolonging intervention are spent in poor health, but mainly the last years of life. The empirical results presented in Chapter 5 showed that ignoring the relationship between HRQoL and TTD results in an underestimation of the health gains of preventive interventions that extend life.

Throughout this thesis, we applied various statistical methods for modelling HRQoL utility data, in particular we modelled two utility scores: the SF-6D and the EQ-5D. ***Chapters 2, 3 and 5*** showed that for modelling the cross-sectional SF-6D data, a beta regression applied within a generalized additive model for location, scale and shape (GAMLSS) provides a flexible approach. ***Chapter 4*** showed that, mixed beta regression and the Bayesian estimation is attractive for modelling longitudinal SF-6D data. Our findings indicate that a beta distribution enabled fitting the non-standard SF-6D outcome, while Bayesian estimation allowed estimating complex models. Hence, it was possible to straightforwardly account for various sources of bias specific to observational studies and HRQoL data, such as non-ignorable missing data and censoring.

Although compulsory for performing an economic evaluation, in practice, information on HRQoL data such as EQ-5D data is often not available. In ***Chapter 6***, a number of statistical methods (i.e. OLS, Tobit, beta regression, multinomial logit models and Bayesian networks, finite mixture models) were compared for estimating EQ-5D scores when this data is absent. Hence, we developed prediction models that allow predicting EQ-5D data for patients using other aspects of their health (i.e. a disease specific-questionnaire). Our findings showed that finite mixture models fitted the EQ-5D data best. However, for out-of-sample prediction, none of the investigated methods outperformed the others on the entire EQ-5D domain.

In this thesis we attempted to contribute to the health economics literature in various ways. From a policy perspective, it enables to better understand the consequences of aging on population health while illustrating important issues for the practice of economic evaluation particularly for improving the methodology of economic evaluations. While addressing these topics, a number of statistical methods for modelling HRQoL data in various contexts were highlighted.

Samenvatting

De levensverwachting is in veel landen de afgelopen decennia flink gestegen. Het aandeel ouderen in de wereld is mede daardoor toegenomen, een demografisch fenomeen wat ook wel vergrijzing wordt genoemd. Wereldwijd is de proportie personen boven de 60 jaar momenteel 11 procent en naar verwachting is dit percentage in 2050 verdubbeld. Deze veranderingen in de structuur van de bevolking zorgen voor nieuwe uitdagingen in economisch beleid. Belangrijke uitdagingen in veel landen betreffen de betaalbaarheid van de alsmaar groeiende uitgaven aan de gezondheidszorg en het verhogen van de pensioensleeftijd zodat ouderen langer onderdeel kunnen uitmaken van de beroepsbevolking. De consequenties van politieke beslissingen op deze gebieden hangen voor een belangrijk deel af van de mate in hoeverre langer leven gepaard gaat met een goede gezondheid. Daarom is het van belang om de volksgezondheid en de veranderingen daarin over de tijd in kaart te brengen. Het doel van dit proefschrift is om veranderingen in de gezondheid van de Nederlandse bevolking in kaart te brengen en te onderzoeken of deze veranderingen relevant zijn voor economische evaluaties. Hierbij is gezondheidsgerelateerde kwaliteit van leven (health related quality of life HRQoL) gebruikt als meetinstrument voor de gezondheid van de populatie. Daarnaast is in dit proefschrift de vraag gesteld welke methoden gebruikt moeten worden voor het modelleren van HRQoL data. Dit proefschrift onderzoekt dan ook verschillende statistische methoden voor het modelleren van verschillende typen HRQoL data.

In hoofdstuk 2 en 3 hebben we veranderingen in gezondheid geschat van de Nederlandse bevolking, tussen 2001 en 2008, voor verschillende subgroepen van de bevolking. We zagen dat Nederlanders niet alleen langer leven dan vroeger, maar ook in een betere kwaliteit van leven. Ondanks dat dit het geval lijkt voor alle subgroepen in de bevolking, laat hoofdstuk 3 zien dat in vergelijking tot de laagopgeleiden, de hoogopgeleide Nederlanders hiervan het meest profiteren. In andere woorden, tussen 2001 en 2008 is de gezondheid van alle bevolkingsgroepen in Nederland verbeterd, maar deze is meer verbeterd voor de hoogopgeleiden. Dit betekent dat ongelijkheden in gezondheid naar opleiding groter zijn geworden in Nederland. Onze bevinding dat Nederlanders langer leven in betere kwaliteit van leven, indiceert ook dat er in Nederland in de afgelopen periode een compressie van morbiditeit heeft plaatsgevonden. Dit is van belang bij het bepalen van de consequenties van vergrijzing in verschillende contexten.

Vele studies hebben reeds laten zien dat naarmate mensen ouder worden, de gezondheid achteruit gaat en het gebruik van zorg toeneemt. Dit suggereert dat langer leven niet noodzakelijkerwijs leidt tot langer werken maar vooral tot extra zorgvraag. Echter, zulke schijnbaar duidelijke conclusies kunnen misleidend zijn. Zo heeft bijvoorbeeld ander onderzoek laten zien dat het gebruik van gezondheidszorg gecentreerd is in de jaren voorafgaand aan de dood en dat daarom een ouder wordende bevolking een

gelimiteerd effect heeft op toekomstig zorggebruik. Dit komt doordat de tijd tot dood (TTD) een sterkere voorspeller is voor zorggebruik dan leeftijd. Om de geobserveerde veranderingen in de volksgezondheid te verklaren zoals deze zijn uiteengezet in hoofdstuk 2 en 3, onderzoekt dit proefschrift een mechanisme tussen HRQoL en TTD. Hoofdstuk 4 laat zien dat de geobserveerde relatie tussen gezondheid, zoals gemeten door HRQoL en leeftijd voor een groot deel verklaard kan worden door de relatie tussen HRQoL en TTD. Net als bij de gezondheidsuitgaven hebben we aangetoond dat wanneer er rekening wordt gehouden met TTD, het effect van leeftijd op HRQoL verwaarloosbaar wordt: HRQoL verliezen veroorzaakt door verlaging in TTD zijn aanzienlijk hoger dan de verliezen veroorzaakt door een verhoogde leeftijd. Deze resultaten suggereren dat een verdere stijging van de levensverwachting niet per se resulteert in meer jaren gespendeerd in slechte gezondheid. Hoofdstuk 5 laat zien dat wanneer een andere dataset werd gebruikt, de geobserveerde veranderingen in HRQoL door leeftijd, ook het resultaat zijn van een sterke relatie tussen HRQoL en TTD. Daarnaast zagen we dat deze relatie belangrijke implicaties heeft voor het schatten van gezondheidswinst in economische evaluaties van levensverlengende interventies. Dit komt doordat deze levensverlengende interventies voor een groot gedeelte de HRQoL verliezen uitstellen naar het einde van het leven. Daarom worden niet alle jaren die gewonnen worden door levensverlengende interventies, gespendeerd in slechte gezondheid. De empirische resultaten die in hoofdstuk 5 gepresenteerd worden, laten zien dat het negeren van de relatie tussen HRQoL en TTD resulteert in een onderschatting van de gezondheidswinst van preventieve levensverlengende interventies.

In dit proefschrift zijn verschillende statistische methoden toegepast om HRQoL utiliteiten te modelleren waarbij voornamelijk de utiliteitsscores van de SF-6D en de EQ-5D gemodelleerd werden. Hoofdstuk 2, 3 en 5 lieten zien dat een beta regressie toegepast in een gegeneraliseerd additief model voor locatie, schaal en vorm (GAMLSS) een flexibele benadering was voor het modelleren van cross-sectionele SF-6D data. Hoofdstuk 4 liet zien dat voor het modelleren van longitudinale SF-6D data een Bayesiaanse schattingsmethode het meest aantrekkelijk was. Onze bevindingen indiceren dat een beta distributie ervoor heeft gezorgd dat de niet normale SF-6D uitkomst gemodelleerd kon worden, terwijl de Bayesiaanse schatting ervoor zorgt dat complexe modellen geschat kunnen worden. Op deze manier was het mogelijk om op een directe manier te corrigeren voor verschillende bronnen van bias specifiek voor observationele studies en HRQoL data, zoals non-ignorable missende data en censoring.

Alhoewel voor economische evaluaties HRQoL data zoals de EQ-5D benodigd zijn, is in de praktijk dit type data vaak niet beschikbaar. In hoofdstuk 6 worden verschillende

statistische methoden (i.e. OLS, Tobit, beta regressie, multinomiale logit modellen and Bayesiaanse netwerken, finite mixture modellen) vergeleken om EQ-5D utiliteiten te schatten wanneer deze data niet aanwezig is. In dit hoofdstuk manier hebben we predictiemodellen ontwikkeld voor patiënten door gebruik te maken van andere aspecten van de gezondheid op basis van ziekte specifieke vragenlijsten. Onze bevindingen toonden aan dat de finite mixture modellen, het beste bij onze data pasten. Echter, voor voorspellingen buiten de sample was er geen model dat beter presteerde dan de andere modellen voor het gehele EQ-5D domein.

In dit proefschrift hebben we geprobeerd om op verschillende manieren bij te dragen aan de gezondheidseconomische literatuur. Vanuit een beleidsperspectief draagt dit proefschrift bij aan een beter begrip van de gevolgen van ouder worden voor de volksgezondheid terwijl belangrijke zaken ten aanzien van de praktijk van economische evaluaties geïllustreerd worden, in het bijzonder door het verbeteren van de methodologie van economische evaluaties. Terwijl deze verschillende thema's aan bod komen, wordt er tegelijkertijd dieper ingegaan op de statistische methoden voor het modelleren van HRQoL data.

Acknowledgements

Acknowledgements

I am very proud to admit that I did not go through the experience of being a young researcher at Erasmus University all by myself. There are many people that have supported, helped and influenced me all these years, who must be thanked.

I shall begin with people that were there for me from a professional point of view. I was very lucky for having two wonderful supervisors for my PhD. I would like to express my enormous gratitude to Pieter Van Baal; first of all for giving me the opportunity to be a PhD student under his supervision; and second for his advices and unconditional support. His flair for research combined with his enthusiasm gave contour to this thesis. Over the years I felt that Pieter gave me the wings and thought me how to fly. Without his trust that our papers will become 'hits' that will take over the world my time as a PhD student would have been a lot duller. I would like to acknowledge my promotor Werner Brouwer for the inspiring talks and constant support during my PhD. Werner, I felt that you were always there for me! Thank you for your contribution to this book!

Special thanks go to some of my collaborators: Job Van Exel and Matthijs Versteegh (for helping me write my first project proposal), Melinde Boland and Maureen Rutten Van Moelken (for sharing your experience about mapping and COPD).

Shifting my attention onto more personal acknowledgements, I would like to thank my friends and family. At iBMG I enjoyed an environment with many young research colleagues from which a learned a lot not only about health economics but also about the Dutch culture (now I like bitterballen!). My gratitude goes to all of you. Special acknowledgements go to my Romanian group of friends from Delft (or who use to live in Delft) with whom I shared so many joyful moments (I don't need to list your names here because you know exactly who you are.).

When one moves abroad some friends become family. I am lucky to have such special friends in the Netherlands whom are deeply missed now that I moved back to my home country. Parida Wubulihasimu - many thanks for being a very good friend and colleague (during my master and PhD). Your companionship in the Netherlands meant a lot to me. Special thanks go to Maartje Goorden for her friendship and support during my PhD years. Thank you for all the fun and good times (great squash games!).

Writing this thesis, being as happy as I am now, would have not been possible if it weren't for my family. Thank you Mella for being a marvellous sister. I am forever in debt for your patience in listening me talking about my 'big future plans'. Thank you for your friend-

ship and faith in me, for giving me the most sincere, and helpful advices and for trying to bring me 'down to earth' when necessary. That brings me to my mother, words fail if I try to express my enormous gratitude for my mother. Draga mama, cuvintele palesc daca incerc sa imi exprim recunostinta fata de tine. Prietenia, sprijinul si increderea ta constanta in mine, m-au facut sa merg inainte spre culmi nebanuite. Sper ca toata lumea sa aiba o mama ca tine! Tie iti dedic aceasta carte.

Marius, you have always been there for me as I turned from a first year master student to the author of this book. Before I even thought about working in research, you saw the match between me and research. My dearest companion in life, thank you, for making everything possible, beautiful and special. Everything would be worthless without you!

A wholehearted *thank you* to all of you.

Maria Gheorghe

PhD portfolio

Name PhD candidate: Maria Gheorghe

Erasmus University Rotterdam

Department: Institute for Health Policy and Management (iBMG), Erasmus University

PhD period: 2011-2016

Promotor: Prof. Dr. Werner Brouwer

Supervisor: Dr. Pieter Van Baal

PhD training

2011: Missing Values in Clinical Research, Rotterdam

2011: Quality of life measurements, Rotterdam

2012: Bayesian statistics, Leiden University, Leiden

2013: Decision Analytic modelling for economic evaluations: advanced course, Glasgow

2013: Joint modelling of longitudinal and time to event data, Rotterdam

2015: Network meta-analysis and mixed treatment comparison, Leicester

International conferences

2012: 21st European Workshop on Econometrics and Health Economics, Lund

2013: Health Economists' Study Group (HESG), Coventry: *Quality of life and time to death: have the health gains of preventive interventions been underestimated?* Study discussed.

2014: International Society for Pharmacoeconomics and Outcomes Research (ISPOR), Amsterdam: *Modeling health-related quality of life longitudinally (HRQoL). A Bayesian mixed beta regression approach* (poster presentation). Poster presentation.

2014: International Health Economics Association (iHEA) Dublin: *Quality of life and time to death: have the health gains of life prolonging preventive interventions been underestimated?* Oral presentation.

2015: International Biometrics Society (IBS), Nijmegen: Using Bayesian Networks (BNs) for predicting patient reported outcomes from disease-specific questionnaires-podium presentation. Oral presentation.

National conferences

2011: 'Low Lands Health Economists' Study Group (LoLa HESG), Netherlands: *Did the health of the Dutch population change between 2001 and 2008?* Study discussed.

2012: 'Low Lands Health Economists' Study Group (LoLa HESG), Netherlands: *Quality of life and time to death: have the health gains of preventive interventions been underestimated?* Study discussed.

2013: 'Low Lands Health Economists' Study Group (LoLa HESG), Netherlands: *Health losses at the end of life.* Study discussed.

2014: 'Low Lands Health Economists' Study Group (LoLa HESG), Netherlands: *Health Losses at the end of life.* Study discussed.

2015: 'Low Lands Health Economists' Study Group (LoLa HESG), Netherlands: *Predicting patient reported outcomes measures (PROMs) from disease-specific questionnaires: a comparison of existing methods.* Study discussed.

Peer reviewed publications

Gheorghe M., Brouwer W. B. F., van Baal P. H. M. Did the health of the Dutch population improve between 2001 and 2008? Investigating age- and gender-specific trends in quality of life. European Journal of Health Economics; 2015 Nov;16(8):801-11.

Gheorghe M., Brouwer W. B. F., van Baal P. H. M. Quality of Life and Time to Death: Have the Health Gains of Preventive Interventions Been Underestimated? Medical Decision Making 2015 Apr; 35(3):316-27.

Gheorghe M., Wubulihasimu P., Peters F., Nusselder W., van Baal P. H. M. Health inequalities in the Netherlands: trends in quality adjusted life expectancy (QALE) by educational level, European Journal of Public Health, 2016, Apr 16.

Gheorghe M., Picavet S., Verschuren M., Brouwer W. B. F., van Baal, P. H. M. Health losses at the end of life. A Bayesian mixed beta regression. Journal of the Royal Statistical Society, Series A, Statistics in Society.

Wubulihasimu P., Gheorghe M., Slobbe L., Polder J., van Baal, P. H. M. Trends in Dutch hospital spending by age and disease 1994-2010. Health Policy, 2014.

Hanea A., Gheorghe M., Hanea R., Ababei D. Non-parametric Bayesian networks for parameter estimation in reservoir simulation: a graphical take on the ensemble Kalman filter (part I). Computational Geosciences 2013 12/01; 17(6):929-949.

Reports

2015: Health-related constraints to raising Retirement Ages in the EU: A probabilistic Markov-Model of age-related disability rates for selected disease causes and related impacts on public payer cash benefit expenditures. European Union (EU)
http://ec.europa.eu/health/systems_performance_assessment/docs/retirement_ages_en.pdf

2013: Estimating and forecasting costs of illness in the Netherlands. RIVM

Teaching

2012-2013: Lecturer at the Public Health Economics (PHE) course

2012-2015: Master thesis contact person for HEPL and HE students

2014-2015: Master thesis supervisor for HEPL and HE students

Awards

2015: Dutch Society for Technology Assessment in Healthcare (NTVAG) award for the best paper published by a young researcher

Gheorghe M., Brouwer W. B. F., van Baal P. H. M. Quality of Life and Time to Death: Have the Health Gains of Preventive Interventions Been Underestimated? Medical Decision Making 2015 Apr; 35(3):316-27.

About the author

Maria Gheorghe (1983, Sibiu, Romania) holds a B.Sc. in Economics and Statistics (2002-2007) from the Academy of Economic Studies in Bucharest, Romania. After her graduation, Maria worked as a Life Actuary at Vienna Insurance Group in Bucharest (2007-2008). In 2008, Maria enrolled in the master of Applied Mathematics at TU Delft in the Netherlands. Her study at Delft was financially supported through s scholarship from TU Delft. From 2011 to 2016, Maria worked as a PhD researcher at the (institute of Health Policy & Management of the) Erasmus University Rotterdam, focusing on applying various quantitative methods for estimating population quality of life. During this period, she worked on projects for the Netherlands Organization for Health, Research and Development (ZonMw), the Dutch National Institute for Public Health and the Environment (RIVM) and the European Union (EU). In 2015, Maria won the Dutch Society for Technology Assessment in Healthcare (NTVAG) award for the best paper published by a young researcher.

On May 2016, Maria moved back to Bucharest, where she now works for Sanofi. In her new position, she enables patient access to new treatments in Romania. Moreover, in her spare time, she continues her academic work with former colleagues.

References

(1) Eggleston KN, Fuchs VR. The New Demographic Transition: Most Gains in Life Expectancy Now realized Late in Life. Journal of economic perspectives 2012; 26(3):137-156.

(2) United Nations. Departament of Economic and Social Affairs, Population division: World population aging 2013.

(3) WHO. World Health Organization (WHO): Preamble to the constitution of the World Health Organization. 1946.

(4) Dolan P. Handbook of Health Economics Volume 1 Chapter 32 The measurement of health-related quality of life for use in resource allocation decisions in health care. Handbook of health economics 2000;1: 1723-1760.

(5) Perenboom RJ, Van Herten LM, Boshuizen HC, Van Den Bos GA. Trends in disability-free life expectancy. Disabil Rehabil 2004 Apr 8; 26(7):377-386.

(6) Water VD, Boshuizen HC, Perenboom RJM. Health expectancy in the Netherlands 1983-1990. Eur J Public Health 1996 03/01; 6(1):21-28.

(7) Picavet HS, Hoeymans N. Physical disability in The Netherlands: prevalence, risk groups and time trends. Public Health 2002 Jul;116(4):231-237.

(8) Majer IM, Stevens R, Nusselder WJ, Mackenbach JP, van Baal PH. Modeling and forecasting health expectancy: theoretical framework and application. Demography 2013 Apr;50(2):673-697.

(9) Bruggink JW, Garssen J, Lodder B, Kardal M. Trends in gezonde levensverwachting [Trends in health expectancy]. In Bevolkingstrend, Den Haag, The Netherlands: Centraal Bureau voor de Statistiek. 2009.

(10) Drummond MF, Sculpher MJ, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. third ed.: Oxford University Press; 2005.

(11) Leon DA. Trends in European life expectancy: a salutary view. Int J Epidemiol 2011Apr;40(2):271-277.

(12) Wilmoth JR. Demography of longevity: past, present, and future trends. Exp Gerontol 2000; 35(9-10):1111.

(13) Murray CJ, Salomon JA, Mathers C. A critical examination of summary measures of population health. Bull World Health Organ 2000; 78(8):981-994.

(14) Salomon JA, Nordhagen S, Oza S, Murray CJ. Are Americans feeling less healthy? The puzzle of trends in self-rated health. Am J Epidemiol 2009 Aug 1;170(3):343-351.

(15) Attema AE, Edelaar-Peeters Y, Versteegh MM, Stolk EA. Time trade-off: one methodology, different methods. The European Journal of Health Economics 2013 07/31; 14:53-64.

(16) Gafni A. The standard gamble method: what is being measured and how it is interpreted. Health Serv Res 1994 06; 29(2):207-224.

(17) Dolan P. Modeling valuations for EuroQol health states. Med Care 1997 Nov;35(11):1095-1108.

(18) Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ 2002 Mar; 21(2):271-292.

(19) Brouwer WB, Culyer AJ, van Exel NJ, Rutten FF. Welfarism vs. extra-welfarism. J Health Econ 2008 Mar; 27(2):325-338.

(20) Longworth L, Yang Y, Young T, Mulhern B, Hernandez Alava M, Mukuria C, et al. Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modelling and survey. Health Technol Assess 2014 Feb;18(9):1-224.

(21) Zorginstituut Nederland. Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg. 2016.

(22) Salomon JA, Wang H, Freeman MK, Vos T, Flaxman AD, Lopez AD, et al. Healthy life expectancy for 187 countries, 1990-2010: a systematic analysis for the Global Burden Disease Study 2010. Lancet 2012 Dec 15; 380(9859):2144-2162.

(23) Turrell G, Mathers C. Socioeconomic inequalities in all-cause and specific-cause mortality in Australia: 1985-1987 and 1995-1997. Int J Epidemiol 2001 Apr; 30(2):231-239.

(24) Martikainen P, Makela P, Koskinen S, Valkonen T. Income differences in mortality: a register-based follow-up study of three million men and women. Int J Epidemiol 2001 Dec; 30(6):1397-1405.

(25) Mackenbach JP, Bos V, Andersen O, Cardano M, Costa G, Harding S, et al. Widening socioeconomic inequalities in mortality in six Western European countries. Int J Epidemiol 2003 Oct;32(5):830-837.

(26) Singh GK, Siahpush M. Widening socioeconomic inequalities in US life expectancy, 1980-2000. Int J Epidemiol 2006 Aug; 35(4):969-979.

(27) Mackenbach JP, Stirbu I, Roskam AJ, Schaap M, Menvielle G, Leinsalu M, et al. Socioeconomic inequalities in health in 22 European countries. N Engl J Med 2008 Jun 5;358(23):2468-2481.

(28) Meara ER, Richards S, Cutler DM. The gap gets bigger: changes in mortality and life expectancy, by education, 1981-2000. Health Aff (Millwood) 2008 Mar-Apr; 27(2):350-360.

(29) van Kippersluis H, O'Donnell O, van Doorslaer E, Van Ourti T. Socioeconomic differences in health over the life cycle in an Egalitarian country. Soc Sci Med 2010 Feb; 70(3):428-438.

(30) Maki N, Martikainen P, Eikemo T, Menvielle G, Lundberg O, Ostergren O, et al. Educational differences in disability-free life expectancy: a comparative study of long-standing activity limitation in eight European countries. Soc Sci Med 2013 Oct; 94:1-8.

(31) Kunst A, Bos V, Lahelma E, Bartley M, Lissau I, Regidor E, et al. Trends in socioeconomic inequalities in self-assessed health in 10 European countries. Int J Epidemiol 2005 Apr; 34(2):295-305.

(32) Majer IM, Nusselder WJ, Mackenbach JP, Klijs B, van Baal PH. Mortality risk associated with disability: a population-based record linkage study. Am J Public Health 2011 Dec;101(12):e9-15.

(33) Van Oyen H, Charafeddine R, Deboosere P, Cox B, Lorant V, Nusselder W, et al. Contribution of mortality and disability to the secular trend in health inequality at the turn of century in Belgium. Eur J Public Health 2011 Dec; 21(6):781-787.

(34) Cambois E, Robine JM, Hayward MD. Social inequalities in disability-free life expectancy in the French male population, 1980-1991. Demography 2001 Nov; 38(4):513-524.

(35) Bronnum-Hansen H, Baadsgaard M. Increase in social inequality in health expectancy in Denmark. Scand J Public Health 2008 Jan; 36(1):44-51.

(36) Bruggink JW. Development in the (healthy) life expectancy by education. in Dutch: Ontwikkelingen in (gezonde) levensverwachting naaropleidingsniveau. Bevolkingstrends 2009; 57:71-5.

(37) Crimmins EM, Saito Y. Trends in healthy life expectancy in the United States, 1970-1990: gender, racial, and educational differences. Soc Sci Med 2001 Jun; 52(11):1629-1641.

(38) Davis P, Graham P, Pearce N. Health expectancy in New Zealand, 1981-1991: social variations and trends in a period of rapid social and economic change. J Epidemiol Community Health 1999 Sep; 53(9):519-527.

(39) Fryback DG, Dunham NC, Palta M, Hanmer J, Buechner J, Cherepanov D, et al. US norms for six generic health-related quality-of-life indexes from the National Health Measurement study. Med Care 2007 Dec;45(12):1162-1170.

(40) Getzen TE. Population aging and the growth of health expenditures. J Gerontol 1992 May; 47(3):S98-104.

(41) Zweifel P, Felder S, Meiers M. Ageing of population and health care expenditure: a red herring? Health Econ 1999 -09; 8(6):485-496.

(42) Bos D, von Weizsacker RK. Economic consequences of an aging population. European Economic Review 1989; 33:345-355.

(43) OECD. Ageing population: the social policy implications. 1988.

(44) Seshamani M, Gray A. Time to death and health expenditure: an improved model for the impact of demographic change on health care costs. Age and Ageing 2004 Sep;33(6):556-561.

(45) Seshamani M, Gray A. Ageing and health-care expenditure: the red herring argument revisited. Health Econ 2004 Apr;13(4):303-314.

(46) Zweifel P, Felder S, Werblow A. Population Ageing and Health Care Expenditure: New Evidence on the "Red Herring". Geneva papers on risk and insurance. Issues and practice 2004 Oct; 29(4):652-666.

(47) Werblow A, Felder S, Zweifel P. Population ageing and health care expenditure: a school of 'red herrings'? Health Econ 2007 Oct;16(10):1109-1126.

(48) Wanless D. Population Ageing and Health Care Expenditure. Annu Rev Public Health 2004; 25:457.

(49) Franks P, Lubetkin EI, Gold MR, Tancredi DJ. Mapping the SF-12 to preference-based instruments: convergent validity in a low-income, minority population. Med Care 2003 Nov; 41(11):1277-1283.

(50) Fryback DG, Lawrence WF, Martin PA, Klein R, Klein BE. Predicting Quality of Well-being scores from the SF-36: results from the Beaver Dam Health Outcomes Study. Med Decis Making 1997 Jan; 17(1):1-9.

(51) O'Brien BJ, Spath M, Blackhouse G, Severens JL, Dorian P, Brazier J. A view from the bridge: agreement between the SF-6D utility algorithm and the Health Utilities Index. Health Econ 2003 Nov; 12(11):975-981.

(52) Versteegh M, Leunis A, Luime J, Boggild M, Uyl-de Groot CA, Stolk EA. Mapping QLQ-C30, HAQ, and MSIS-29 on EQ-5D. Med Decis Making 2012 Jul; 32(4):554-568.

(53) Le QA, Doctor JN. Probabilistic mapping of descriptive health status responses onto health state utilities using Bayesian networks: an empirical analysis converting SF-12 into EQ-5D utility index in a national US sample. Med Care 2011 May;49(5):451-460.

(54) Gray A, Rivero-Arias O, Clarke PM. Estimating the association between SF-12 responses and EQ-5D utility values by response mapping. Med Decis Making 2006 Jan; 26(1):18-29.

(55) Austin PC. A Comparison of Methods for Analyzing Health-Related Quality-of-Life Measures. Value in health 2002 -07;5(4):329; 329-337; 337.

(56) Sullivan PW, Ghushchyan V. Mapping the EQ-5D index from the SF-12: US general population preferences in a nationally representative sample. Med Decis Making 2006 Jul; 26(4):401-409.

(57) Mullahy J. Specification and testing of some modified count data models. Journal of Econometrics 1986; 33(3):341-365.

(58) Basu A, Manca A. Regression estimators for generic health-related quality of life and quality-adjusted life years. Med Decis Making 2012 Jan; 32(1):56-69.

(59) Hunger M, Baumert J, Holle R. Analysis of SF-6D index data: is beta regression appropriate? Value Health 2011 Jul; 14(5):759-767.

(60) Hunger M, Doring A, Holle R. Longitudinal beta regression models for analyzing health-related quality of life scores over time. BMC Med Res Methodol 2012 Sep 17; 12:144.

(61) Gheorghe M, Brouwer WB, van Baal PH. Did the health of the Dutch population improve between 2001 and 2008? Investigating age- and gender-specific trends in quality of life. Eur J Health Econ 2015 Sep 14; 16(8):801-811.

(62) Gheorghe M, Brouwer WB, van Baal PH. Quality of Life and Time to Death: Have the Health Gains of Preventive Interventions Been Underestimated? Med Decis Making 2015 Oct 23; 35(3):316-327.

(63) Hernandez Alava M, Wailoo AJ, Ara R. Tails from the peak district: adjusted limited dependent variable mixture models of EQ-5D questionnaire health state utility values. Value Health 2012 May;15(3):550-561.

(64) Hernandez Alava M, Wailoo A, Wolfe F, Michaud K. A comparison of direct and indirect methods for the estimation of health utilities from clinical outcomes. Med Decis Making 2014 Oct;34(7):919-930.

(65) Coca Perraillon M, Shih YC, Thisted RA. Predicting the EQ-5D-3L Preference Index from the SF-12 Health Survey in a National US Sample: A Finite Mixture Approach. Med Decis Making 2015 Oct;35(7):888-901.

(66) Kieschnick R, McCullough BD. Regression analysis of variates observed on (0, 1): percentages, proportions and fractions. Statistical modelling 2003 -10;3(3):193-213.

(67) Buxton MJ, Drummond MF, Van Hout BA, Prince RL, Sheldon TA, Szucs T, et al. Modelling in economic evaluation: an unavoidable fact of life. Health Econ 1997 May;6(3):217-227.

(68) de Kok IMCM, van Ballegooijen M, Habbema JDF. Cost-Effectiveness Analysis of Human Papillomavirus Vaccination in the Netherlands. JNCI : Journal of the National Cancer Institute 2009 Aug;101(15):1083-1092.

(69) Anonychuk AM, Bauch CT, Merid MF, Van Kriekinge G, Demarteau N. A cost-utility analysis of cervical cancer vaccination in preadolescent Canadian females. BMC Public Health 2009 Oct 31;9:401-2458-9-401.

(70) Schousboe JT, Kerlikowske K, Loh A, Cummings SR. Personalizing mammography by breast density and other risk factors for breast cancer: analysis of health benefits and cost-effectiveness. Ann Intern Med 2011 Jul 5;155(1):10-20.

(71) Tosteson AN, Stout NK, Fryback DG, Acharyya S, Herman BA, Hannah LG, et al. Cost-effectiveness of digital mammography breast cancer screening. Ann Intern Med 2008 Jan 1;148(1):1-10.

(72) Cutler D, Deaton A, Lleras-Muney A. The determinants of mortality. Journal of economic perspectives 2006;20(3):97-120.

(73) Statistics Netherlands. Statline. 2011.

(74) Mackenbach JP, Slobbe L, Looman CW, van der Heide A, Polder J, Garssen J. Sharp upturn of life expectancy in the Netherlands: effect of more health care for the elderly? Eur J Epidemiol 2011 Dec;26(12):903-914.

(75) Gruenberg EM. The Failures of Success. The Millbank Quaterly 1977;55:3.

(76) Fries JF. Compression of morbidity in the elderly. Vaccine 2000;18(16):1584.

(77) Layes A, Asada Y, Kepart G. Whiners and deniers - what does self-rated health measure? Soc Sci Med 2012 Jul;75(1):1-9.

(78) Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. Med Care 2004 Sep;42(9):851-859.

(79) Cutler DM, Richardson E. Measuring the health of the US population. Brookings papers on economic activity.Microeconomics 1997;1997:217.

(80) Williams A. Calculating the global burden of disease: time for a strategic reappraisal? Health Econ 1999 -02;8(1):1-8.

(81) Heijink R, van Baal P, Oppe M, Koolman X, Westert G. Decomposing cross-country differences in quality adjusted life expectancy: the impact of value sets. Popul Health Metr 2011 Jun 23;9(1):17-7954-9-17.

(82) Jia H, Zack MM, Thompson WW. State Quality-Adjusted Life Expectancy for U.S. adults from 1993 to 2008. Qual Life Res 2011 Aug;20(6):853-863.

(83) Rubin DB. Inference and missing data. Biometrika 1976;63(3):581; 581-592; 592.

(84) van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. Stat Med 1999 Mar 30;18(6):681-694.

(85) van Buuren S, Brand JPL, Groothuis-Oudshoorn CGM, Rubin DB. Fully conditional specification in multivariate imputation. Journal of statistical computation and simulation 2006 Dec;76(12):1049-1064.

(86) Raghunathan TE, Lepkowski JM., Van Hoewyk J. A multivariate technique for multiply imputing missing values using a sequence of regression models. Surv Methodol Jun 2001, 27(1); 85-95.

(87) Peyre H, Leplege A, Coste J. Missing data methods for dealing with missing items in quality of life questionnaires. A comparison by simulation of personal mean score, full information maximum likelihood, multiple imputation, and hot deck techniques applied to the SF-36 in the French 2003 decennial health survey. Qual Life Res 2011 Mar;20(2):287-300.

(88) Rubin DB. Multiple imputations for non-response in surveys. New York: John Qiley & Sons; 1987.

(89) Pullenayegum EM, Tarride JE, Xie F, Goeree R, Gerstein HC, O'Reilly D. Analysis of health utility data when some subjects attain the upper bound of 1: are Tobit and CLAD models appropriate? Value Health 2010 Jun;13(4):487-494.

(90) Ferrari S, Francisco C. Beta Regression for Modelling Rates and Proportions. Journal of applied statistics 2004 Aug;31(7):799; 799-815; 815.

(91) Smithson M, Verkuilen J. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. Psychol Methods 2006 Mar;11(1):54-71.

(92) Cox C. Nonlinear quasi-likelihood models: applications to continuous proportions. Comput Stat Data Anal 1996 -04;21(4):449; 449-461; 461.

(93) Ospina R, Ferrari SLP. Inflated beta distributions. Statistical papers (Berlin, Germany) 2010 Jan;51(1):111-126.

(94) Rigby RA, Stasinopoulos DM. A flexible regression approach using GAMLSS in R. 2010.

(95) Stasinopoulos DM, Rigby RA. Generalized additive models for location scale and shape (GAMLSS) in R. Journal of statistical software 2007;23(7):1.

(96) Cramer H. Mathematical Methods of Statistics. Princeton, NJ: Princeton University Press; 1946.

(97) Li L, Fu AZ. Some methodological issues with the analysis of preference-based EQ-5D index score. Health Serv Outcomes Res 2009 -09;9(3):162-176.

(98) Eilers PHC, Marx BD. Flexible smoothing using B-splines and penalties. Statistical science 1996; 11(2):89-121.

(99) Hastie T, Tibshirani R. Varying-Coefficient Models. Journal of the Royal Statistical Society.Series B (Methodological) 1993;55(4):pp. 757-796.

(100) Akaike ,Hirotogu. Information Theory and an Extension of the Maximum Likelihood Principle. 1973.

(101) Schafer JL, Olsen MK. Multiple Imputation for Multivariate Missing-Data Problems: A Data Analyst's Perspective. Multivariate behavioral research 1998 Oct;33(4):545-571.

(102) Sullivan DF. A single index of mortality and morbidity. HSMHA Health Rep 1971 Apr;86(4):347-354.

(103) Figueroa-Zúñiga JI, Arellano-Valle RB, Ferrari SLP. Mixed beta regression: A Bayesian perspective. Comput Stat Data Anal 2013 May;61: 137-147.

(104) Muennig P, Franks P, Jia H, Lubetkin E, Gold MR. The income-associated burden of disease in the United States. Soc Sci Med 2005 Nov;61(9):2018-2026.

(105) Luo N, Johnson JA, Shaw JW, Feeny D, Coons SJ. Self-reported health status of the general adult U.S. population as assessed by the EQ-5D and Health Utilities Index. Med Care 2005 Nov;43(11):1078-1086.

(106) Cherepanov D, Palta M, Fryback DG, Robert SA. Gender differences in health-related quality-of-life are partly explained by sociodemographic and socioeconomic variation between adult men and women in the US: evidence from four US nationally representative data sets. Qual Life Res 2010 Oct;19(8):1115-1124.

(107) Tamhane A, Ankenmana B, Yang Y. The beta distribution as a latent response model for ordinal data (I): Estimation of location and dispersion parameters. Journal of statistical computation and simulation 2002 Jan;72(6):473-494.

(108) Luo N, Wang P, Fu AZ, Johnson JA, Coons SJ. Preference-based SF-6D scores derived from the SF-36 and SF-12 have different discriminative power in a population health survey. Med Care 2012 Jul;50(7):627-632.

(109) Graham JW. How Many Imputations are Really Needed? Some Practical Clarifications of Multiple Imputation Theory. Prevention science 2007 Aug;8(3):206-213.

(110) Cutler DM, Lleras-Muney A. Understanding differences in health behaviors by education. J Health Econ 2010 Jan;29(1):1-28.

(111) Crimmins EM. Mixed trends in population health among older adults. J Gerontol B Psychol Sci Soc Sci 1996 Sep;51(5):223-225.

(112) Collins B. Using a survey to estimate health expectancy and quality-adjusted life expectancy to assess inequalities in health and quality of life. Value Health 2013 Jun;16(4):599-603.

(113) Kulhanova I, Hoffmann R, Eikemo TA, Menvielle G, Mackenbach JP. Educational inequalities in mortality by cause of death: first national data for the Netherlands. Int J Public Health 2014 Oct;59(5):687-696.

(114) Andreev EM, Shkolnikov VM, Begun AZ. Algorithm for decomposition of differences between aggregate demographic measures and its application to life expectancies, healthy life expectancies, parity-progression ratios and total fertility rates. Demographic research 2002 Jul;7(14):499-522.

(115) Wubulihasimu P, Brouwer W, van Baal P. The Impact of Hospital Payment Schemes on Healthcare and Mortality: Evidence from Hospital Payment Reforms in OECD Countries. Health Econ 2015 Jun 16; 25 (8):1005-1019.

(116) Stuckler D, Basu S, Suhrcke M, Coutts A, McKee M. Effects of the 2008 recession on health: a first look at European data. Lancet 2011 Jul 9;378(9786):124-125.

(117) McKee M, Karanikolos M, Belcher P, Stuckler D. Austerity: a failed experiment on the people of Europe. Clin Med (Lond) 2012 Aug;12(4):346-350.

(118) Visscher G. De blinde vlek van het CBS : systematische vertekening in het opleidingsniveau: De nonrespons in de Enquête Beroepsbevolking. Sociologische gids 1997;44(3):155-179.

(119) Lorant V, Demarest S, Miermans PJ, Van Oyen H. Survey error in measuring socio-economic risk factors of health status: a comparison of a survey and a census. Int J Epidemiol 2007 Dec;36(6):1292-1299.

(120) Pullenayegum EM, Wong HS, Childs A. Generalized Additive Models for the Analysis of EQ-5D Utility Data. Med Decis Making 2013 Feb;33(2):244-51.

(121) Verschuren WM, Blokstra A, Picavet HS, Smit HA. Cohort profile: the Doetinchem Cohort Study. Int J Epidemiol 2008 Dec;37(6):1236-1241.

(122) Spiegelhalter D, Thomas A, Best N, Lunn D. WinBUGS user manual: Version 1.4. 2003.

(123) Verkuilen J, Smithson M. Mixed and mixture regression models for continuous bounded responses using the beta distribution. Journal of educational and behavioral statistics 2012;37(1):1-32.

(124) Gelman A and Hill J. Data analysis using regression and multilevel/hierarchical models: Cambridge University Press; 2007.

(125) Lleras-Muney A. The Relationship Between Education and Adult Mortality in the United States. The Review of Economic Studies 2005 Jan;72(1):189-221.

(126) Spiegelhalter DJ, Best NG, Carlin BP, Linde Avd. Bayesian Measures of Model Complexity and Fit. Journal of the Royal Statistical Society. Series B (Statistical Methodology) 2002;64(4):583-639.

(127) Gelman A, Rubin DB. Inference from Iterative Simulation Using Multiple Sequences. Statistical Science 1992 11;7(4):457-472.

(128) Greenland S. Multiple-bias modelling for analysis of observational data. Journal of the Royal Statistical Society: Series A (Statistics in Society) 2005;168(2):267-306.

(129) Greenland S. Bayesian perspectives for epidemiologic research: III. Bias analysis via missing-data methods. Int J Epidemiol 2009 Dec; 38(6):1662-1673.

(130) Ibrahim JG, Molenberghs G. Missing data methods in longitudinal studies: a review. Test (Madr) 2009 May 1;18(1):1-43.

(131) Mason A, Richardson S, Plewis I, Best N. Strategy for Modelling Nonrandom Missing Data Mechanisms in Observational Studies Using Bayesian Methods. Journal of Official Statistics 2012;28(2):279-302.

(132) Scharfstein DO, Daniels MJ, Robins JM. Incorporating prior beliefs about selection bias into the analysis of randomized trials with missing outcomes. Biostatistics 2003 Oct; 4(4):495-512.

(133) Little RJA. Regression with Missing X's: A Review. Journal of the American Statistical Association 1992; 87(420):1227-1237.

(134) van Baal PH, Wong A. Time to death and the forecasting of macro-level health care expenditures: some further considerations. J Health Econ 2012 Dec; 31(6):876-887.

(135) Oeppen J, Vaupel JW. Broken limits of life expectancy. Science 2002; 296(3):6-13.

(136) Cooper NJ, Lambert PC, Abrams KR, Sutton AJ. Predicting costs over time using Bayesian Markov chain Monte Carlo methods: an application to early inflammatory polyarthritis. Health Econ 2007 Jan; 16(1):37-56.

(137) Chan KCG, Wang MC. Backward Estimation of Stochastic dgt with Failure Events as Time Origins. 4(3): 1602-1620. Annals of Applied Statistics 2010;4(3):1602-1620.

(138) Luo X, Tsai WY. A proportional likelihood ratio model. Biometrika 2012;99(1):211-222.

(139) Chan KCG. Nuisance parameter elimination for proportional likelihood ratio models with nonignorable missingness and random truncation. Biometrika 100 (1): 269-276. Biometrika 2013;100(1):269-276.

(140) Fries JF. Aging, natural death, and the compression of morbidity. New England Journal of Medicine 1980;303:130-5.

(141) Hoogenveen RT, van Baal PH, Boshuizen HC, Feenstra TL. Dynamic effects of smoking cessation on disease incidence, mortality and quality of life: The role of time since cessation. Cost Eff Resour Alloc 2008 Jan 11; 1186/1478-7547-6-1.

(142) van Baal PH, van den Berg M, Hoogenveen RT, Vijgen SM, Engelfriet PM. Cost-effectiveness of a low-calorie diet and orlistat for obese persons: modeling long-term health gains through prevention of obesity-related chronic diseases. Value Health 2008 Dec;11(7):1033-1040.

(143) Heijnsdijk EAM, Wever EM, Auvinen A, Hugosson J, Ciatto S, Nelen V, et al. Qulity-of-Life Effects of Prostate-Specific Antigen Screening. New England Journal of Medicine 2012; 367:595-605.

(144) Mihaylova B, Briggs A, Armitage J, Parish S, Gray A, Collins R. Lifetime cost effectiveness of simvastatin in a range of risk groups and age groups derived from a randomised trial of 20,536 people. BMJ 2006 Dec 2;333(7579):1145.

(145) Fryback DG, Laurence WF. Dollars May Not Buy as Many QALYs as We Think: A Problem with Defining Quality-of-life Adjustments. Medical Decision Making 1997; 45(12):1162-1170.

(146) Polder J, Barendregt J, van Oers H. Health care costs in the last year of life-the Dutch experience. Soc Sci Med 2006 Oct; 63(7):1720-1731.

(147) Stearns SC, Norton EC. Time to include time to death? The future of health care expenditure predictions. Health Econ 2004 -Apr; 13(4):315-327.

(148) Klijs B, Mackenbach JP, Kunst AE. Future disability projections could be improved by connecting to the theory of a dynamic equilibrium. J Clin Epidemiol 2011 Apr; 64(4):436-443.

(149) Gandjour A, Lauterbach KW. Does prevention save costs? Considering deferral of the expensive last year of life. J Health Econ 2005 Jul; 24(4):715-724.

(150) Fan VS, Curtis JR, Tu SP, McDonell MB, Fihn SD, Using quality of life to predict hospitalization and mortality in patients with obstructive lung diseases. Ambulatory Care Quality Improvement Project Investigators. 2002 Aug; 122(2):429-436.

(151) Kaplan MS, Berthelot JM, Feeny D, McFarland BH, Khan S, Orpana H. The predictive validity of health-related quality of life measures: mortality in a longitudinal population-based study. Qual Life Res 2007 Nov;16 (9):1539-1546.

(152) Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape. Journal of the Royal Statistical Society: Series C (Applied Statistics) 2005; 54(3):507-554.

(153) Rigby B, Stasinopoulos M editors. A flexible approach using GAMLSS in R, 2010.

(154) van Baal PH, Feenstra TL, Polder JJ, Hoogenveen RT, Brouwer WB. Economic evaluation and the postponement of health care costs. Health Econ 2011 Apr; 20(4):432-445.

(155) Devlin NJ, Parkin D, Browne J. Patient-reported outcome measures in the NHS: new methods for analysing and reporting EQ-5D data. Health Econ 2010 Aug; 19(8):886-905.

(156) Parrish RG. Measuring Population Health Outcomes. Preventiing Chronic Disease 2010; 7(4):A71.

(157) Brazier J, Deverill M, Green C, Harper R, Booth A. A review of the use of health status measures in economic evaluation. Health Technol Assess 1999; 3(9):i-iv, 1-164.

(158) Lamers LM, McDonnell J, Stalmeier PF, Krabbe PF, Busschbach J. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. Health Econ 2006 Oct; 15(10):1121-1132.

(159) Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. Med Care 2005 Mar; 43(3):203-220.

(160) EuroQol Group editor. EQ-5D Value Sets: Inventory, Comparative Review and User Guide. Szende A, Oppe M, Devlin N. ed. Berlin: Springer; 2007.

(161) Norman R, Cronin P, Viney R, King M, Street D, Ratcliffe J. International comparisons in valuing EQ-5D health states: a review and analysis. Value Health 2009 Nov; 12(8):1194-1200.

(162) Arnold DT, Rowen D, Versteegh MM, Morley A, Hooper CE, Maskell NA. Testing mapping algorithms of the cancer-specific EORTC QLQ-C30 onto EQ-5D in malignant mesothelioma. Health Qual Life Outcomes 2015 Jan 23; 13(1):6.

(163) Dakin H. Review of studies mapping from quality of life or clinical measures to EQ-5D: an online database. Health Qual Life Outcomes 2013 Sep 5;11:151

(164) Soini EJ, Hallinen TA, Puolakka K, Vihervaara V, Kauppi MJ. Cost-effectiveness of adalimumab, etanercept, and tocilizumab as first-line treatments for moderate-to-severe rheumatoid arthritis. J Med Econ 2012; 15(2):340-351.

(165) Mortimer D, Segal L. Comparing the incomparable? A systematic review of competing techniques for converting descriptive measures of health status into QALY-weights. Med Decis Making 2008 Jan; 28(1):66-89.

(166) Kent S, Gray A, Schlackow I, Jenkinson C, McIntosh E. Mapping from the Parkinson's Disease Questionnaire PDQ-39 to the Generic EuroQol EQ-5D-3L: The Value of Mixture Models. Med Decis Making 2015 Apr 29; 35(7):902-913.

(167) Khan I, Morris S. A non-linear beta-binomial regression model for mapping EORTC QLQ- C30 to the EQ-5D-3L in lung cancer patients: a comparison with existing approaches. Health Qual Life Outcomes 2014 Nov 12:163.

(168) Conigliani C, Manca A, Tancredi A. Prediction of patient-reported outcome measures via multivariate ordered probit models. Journal of the Royal Statistical Society: Series A (Statistics in Society) 2015;178(3):567-591.

(169) Borchani H, Bielza C, Marti Nez-Marti NP, Larranaga P. Markov blanket-based approach for learning multi-dimensional Bayesian network classifiers: an application to predict the European Quality of Life-5 Dimensions (EQ-5D) from the 39-item Parkinson's Disease Questionnaire (PDQ-39). J Biomed Inform 2012 Dec; 45(6):1175-1184.

(170) Boland MR, van Boven JF, Kocks JW, van der Molen T, Goossens LM, Chavannes NH, et al. Mapping the clinical chronic obstructive pulmonary disease questionnaire onto generic preference-based EQ-5D values. Value Health 2015 Mar;18(2):299-307.

(171) Wijeysundera HC, Tomlinson G, Norris CM, Ghali WA, Ko DT, Krahn MD. Predicting EQ-5D utility scores from the Seattle Angina Questionnaire in coronary artery disease: a mapping algorithm using a Bayesian framework. Med Decis Making 2011 May; 31(3):481-493.

(172) Maddala GS. Limited-Dependent and QualitativeVariables in Econometric. Cambridge University Press ed. New York: Cambridge University Press; 1983.

(173) Greene WH. Econometric Analysis. 4th edition ed.: Upper Saddle River, NJ: Prentice Hall; 2000.

(174) Grün B, Zeileis A. FlexMix Version 2: Finite Mixtures with Concomitant Variables and Varying and Constant Parameters. Journal of statistical software 2012; 28(4):1-35.

(175) Grün B, Kosmidis I, Zeileis A. Extended Beta Regression in R: Shaken, Stirred, Mixed, and Partitioned. Journal of statistical software 2012; 48(11):1-25.

(176) Spirtes P, Glymour C, Scheines R. Causation, prediction, and search, 2nd ed. Cambridge: MIT Press; 2001.

(177) Bouckaert R. Bayesian belief networks: from construction to inference. Dissertation, 1995.

(178) Spirtes P, Glymour C, Scheines R. Causation, Prediction and Search. 2nd edition: MIT press; 2000.

(179) Margaritis D. Learning Bayesian network model structure from data. 2003, PhD Thesis.

(180) Scutari M. Learning Bayesian Networks with the bnlearn R Package. Journal of statistical software 2010; 35(3).

(181) Nagarajan R, Scutari M, Lèbre S. Bayesian Networks in R. New York: Springer Science and Business Media; 2013.

(182) Longworth L, Rowen D. The use of mapping methods to estimate health state utility values. NICE DSU TECHNICAL SUPPORT DOCUMENT 10: 2011.

(183) Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. Eur J Health Econ 2010 Apr; 11(2):215-225.

(184) Dawid AP. Present Position and Potential Developments: Some Personal Views: Statistical Theory: The Prequential Approach. Journal of the Royal Statistical Society.Series A (General) 1984; 147:278-292.

(185) Seshamani M, Gray AM. A longitudinal study of the effects of age and time to death on hospital costs. J Health Econ 2004 Mar; 23(2):217-235.

(186) Fries JF. Aging, natural death, and the compression of morbidity. Bull World Health Organ 2002; 80(3):245.

(187) Statistics Netherlands. Increasing life expectancy to push up pension age further from 2022. 2014.

(188) Lindeboom M. Health and work of older workers. The Elgar Companion to Health Economics. Edward Elgar Publishing; 2012.

(189) De Waegenaere A, Ying Y, Bertrand M. Linking Retirement Age to Life Expectancy Effects on Healthy Life Expectancy Before and After Retirement. Social science research network (SSRN) 2014.

(190) Bovenberg H, Mackenbachen J, Mehlkopf R. Een eerlijk en vergrijzingbestendig ouderdomspensioen. ESB 2006.

(191) Getzen TE. Health care is an individual necessity and a national luxury: applying multilevel decision models to the analysis of health care expenditures. J Health Econ 2000 Mar; 19(2):259-270.

(192) van Baal PH, Wong A. Time to death and the forecasting of macro-level health care expenditures: some further considerations. J Health Econ 2012 Dec; 31(6):876-887.

ISBN 9789461699442

institute of
Health Policy
& Management