Mortality forecasting in the context of non-linear past mortality trends

an evaluation

Lenny Stoeldraijer

Mortality forecasting in the context of non-linear past mortality trends

an evaluation

Lenny Stoeldraijer

Explanation of symbols

Empty cell	Figure not applicable
.	Figure is unknown, insufficiently reliable or confidential
*	Provisional figure
**	Revised provisional figure
2017–2018	2017 to 2018 inclusive
2017/2018	Average for 2017 to 2018 inclusive
2017/'18	Crop year, financial year, school year, etc., beginning in 2017 and ending in 2018
2015/'16–2017/'18	Crop year, financial year, etc., 2015/'16 to 2017/'18 inclusive

Due to rounding, some totals may not correspond to the sum of the separate figures.

Printed by: Altavia Sumis

Prepress: Statistics Netherlands, CCN Creatie

Design/lay-out: Edenspiekermann

ISBN (printed version): 978-94-034-1236-8
ISBN (electronic version): 978-94-034-1235-1

Language editing
Miriam Hils (Chapters 1 and 6)
Gijsbert van Dalen (Nederlandse samenvatting)

rijksuniversiteit
groningen

Mortality Forecasting in the Context of Non-linear Past Mortality Trends: an Evaluation

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. E. Sterken
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

donderdag 7 februari 2019 om 16.15 uur

door

Lenny Stoeldraijer

geboren op 20 maart 1985
te Terheijden

Promotores
Prof. dr. F. Janssen
Prof. dr. L.J.G. van Wissen

Beoordelingscommissie
Prof. dr. K. Antonio
Prof. dr. N.W. Keilman
Prof. dr. C.H. Mulder

Voorwoord

Wanneer ik precies aan mijn promotietraject ben begonnen, weet ik niet meer. In ieder geval voor 11 juli 2012, want op die datum is de 'Acceptance letter' van de Universiteit van Groningen gedateerd. Nu, ruim 6 jaar later, is het tijd voor de afronding. En bovenal, tijd om iedereen te bedanken die dit proefschrift mogelijk hebben gemaakt.

Allereerst Fanny Janssen: ik had me geen betere begeleider kunnen wensen. Als het tegen zat of wanneer ik mijn motivatie kwijt was, heb je me altijd kunnen stimuleren om door te gaan. De strakke deadlines, de snelle reacties op mijn teksten, je kennis over het onderwerp, je kritische vragen en je voortdurende enthousiasme hebben er voor gezorgd dat ik niet ben afgehaakt en terugkijk op een mooi promotietraject. En ja, af en toe vreesde ik je mails met opmerkingen en rode teksten. Bedankt dat je me hebt gevraagd om te komen promoveren!

Leo van Wissen, mijn andere begeleider zal ik ook zeker niet vergeten. Meestal wat meer op de achtergrond, maar je nuttige adviezen hebben me zeker verder geholpen. Bedankt dat je mijn promotietraject hebt willen ondersteunen met je kennis en kunde.

Daarnaast wil ik de huidige en voormalige collega's bij Demografie bedanken: jullie zijn als een familie! In het bijzonder wil ik een aantal personen bij naam noemen. Wim Leunis en Eric Fokke, bedankt dat jullie me de ruimte hebben gegeven om aan dit proefschrift te werken. Coen van Duin, je creatieve en (op het eerste gezicht) chaotische ideeën brachten mij geregeld op een ander spoor wanneer ik even vast zat. Bedankt voor je begeleiding vanuit het CBS. Peter Meyer, Han Nicolaas, Rob Broekman en Julien Cook, bedankt voor de gezelligheid (en afleiding) op de kamer. Rob Broekman, dankjewel voor het aanmoedigen om te hardlopen: het doorzetten en er achter komen dat ik meer kan dan ik denk te kunnen, heeft me ook zeker geholpen bij de afronding van dit proefschrift. Na onze rondjes in de lunchpauze kon ik altijd weer met een fris (en rood!) hoofd ermee aan de slag.

Daarnaast zijn er heel veel andere collega's van het CBS en daarbuiten waar ik veel van heb geleerd. Indirect heeft dat bijgedragen aan dit proefschrift. Bedankt voor jullie interesse, jullie kennis, het vertrouwen dat ik heb ontvangen, het enthousiasme, jullie kritische blik, de gezelligheid en de fijne samenwerking!

Ook wil ik Luc Bonneux bedanken als coauteur van hoofdstuk 3. Bedankt voor je kritische en waardevolle feedback op mijn manuscripten en voor de prettige en leerzame samenwerking.

Ineke Meuffels, Gerda Polman, Willemijn van den Berg en Inge Magilsen: onze studie econometrie ligt al een hele tijd achter ons, maar we hebben nog steeds frequent contact en het weerzien is altijd gezellig. Het is nu eindelijk tijd voor dat feestje!

Lieve pap en mam, jullie hebben me altijd gestimuleerd om er uit te halen wat er in zit en staan daarmee aan de basis van dit proefschrift. Dank jullie wel voor jullie onvoorwaardelijke steun en liefde. Lieve (schoon)zusjes en (schoon)broer(tje)s, Lisa, Teake, Ruud en Rosalie, bedankt voor alle weekendjes in het oosten en voor jullie afleiding van het werk. Het voelt goed om te weten dat er altijd een veilige haven voor mij is. En als laatste wil ik mijn grootste fan bedanken, mijn neefje Jelte Blom. Voor zijn blije gezichtje als hij me weer ziet.

Table of contents

Voorwoord 5
Overview of chapters 9

1. Introduction 11

2. Impact of different mortality forecasting methods and explicit assumptions on projected future life expectancy: The case of the Netherlands 41

3. The future of smoking-attributable mortality: the case of England & Wales, Denmark and the Netherlands 75

4. An evaluation of methods to coherently forecast mortality based on both quantitative and qualitative criteria 95

5. Comparing strategies for matching mortality forecasts to the most recently observed data: exploring the trade-off between accuracy and robustness 127

6. Conclusion and discussion 157

Annex. Bevolkingsprognose 2012-2060:
model en veronderstellingen betreffende de sterfte 183

English summary 219

Nederlandse samenvatting 227

Acknowledgements 236
About the author 237

Overview of chapters

The four empirical chapters included in this PhD dissertation are reprints of the following publications and manuscripts:

Chapter 2:
Stoeldraijer, L., van Duin, C., van Wissen, L. and Janssen, F. (2013). Impact of different mortality forecasting methods and explicit assumptions on projected future life expectancy: The case of the Netherlands. *Demographic Research* 29(13): 323–354.

Chapter 3:
Stoeldraijer, L., Bonneux, L., van Duin, C., van Wissen, L. and Janssen, F. (2015). The future of smoking-attributable mortality: the case of England & Wales, Denmark and the Netherlands. Addiction 110(2): 336–45.

Chapter 4:
Stoeldraijer, L., van Duin, C., van Wissen, L. and Janssen, F. (2018). A quantitative and qualitative evaluation of methods to coherently forecast mortality. *Submitted*.

Chapter 5:
Stoeldraijer, L., van Duin, C., van Wissen, L. and Janssen, F. (2018). Comparing strategies for matching mortality forecasts to the most recently observed data. Exploring the trade-off between accuracy and robustness. *Genus* 74(16): 1–20.

Annex:
Stoeldraijer, L., van Duin, C. and Janssen, F. (2012). Bevolkingsprognose 2012–2060: model en veronderstellingen betreffende de sterfte. *Bevolkingstrends* 27-6-2013.

Permissions of the copyright holders have been obtained for all chapters and the annex.

1.

Introduction

1.1 Introduction

Against a background of rapid population aging in Western Europe (European Commission 2014), mortality forecasting is becoming increasingly important. Since 1960, life expectancy in Western Europe has risen by around 10 years (from 70 to 80 years) (United Nations 2017). As people are living longer lives and their health needs are expanding, it is not only the structure of the individual life that is changing, but the structure of society as a whole (Bengtsson and Christensen (Eds.) 2006). In particular, social security programs are becoming strained and the sustainability of pension schemes is being called into question (Currie et al. 2004). In order to have some idea of how long individuals will live in the future, what the size and the composition of the older population will be, and how sustainable current pension schemes will be over the long term, it is essential that we have accurate estimates of future mortality by age. Such estimates are usually obtained through mortality forecasts. Since the recent enactment in several Western countries of pension reforms that link the retirement age and/or retirement payments to rapidly increasing life expectancy (OECD 2015; Carone et al. 2016), having accurate and high-quality mortality forecasts has become increasingly important.

As the relevance of mortality forecasts has grown, researchers, statistical offices, and actuarial associations have become increasingly interested in mortality forecasting, especially in Western Europe, where the proportions of older people are high. As a result, numerous models for mortality modelling and forecasting have been developed over the last few decades (for recent reviews, see Booth and Tickle 2008; Cairns et al. 2011). The majority of these new methods of mortality forecasting are extrapolative in nature; that is, they extend a past mortality trend by assuming that both age patterns and trends remain regular over time (Booth and Tickle 2008). Because mortality trends have largely been linear in the majority of Western European countries, this approach generally works well (Booth and Tickle 2008). Compared with other forecasting approaches, the extrapolative methods are highly objective; i.e., they reduce the role of subjective judgment involved in mortality forecasting (Booth and Tickle 2008).

However, particularly in situations in which past trends have been non-linear, the use of an objective extrapolative method will be more problematic. Indeed, in a number of European countries – especially in Nordic countries, the United Kingdom, and the Netherlands, and particularly among men – past mortality trends have been non-linear: in these countries, the increasing trends in life expectancy stagnated over longer periods of time in the 1950s and the 1960s, and then rose

sharply (Janssen et al. 2004; Vallin and Meslé 2004; Kaneda and Scommegna 2011; Crimmins et al 2011). In addition, in the Netherlands and Denmark, clear non-linear trends have been observed among women, as the increasing trend in life expectancy for women in these countries stagnated in the 1980s (Van der Wilk et al. 2001; Lindahl-Jacobsen et al. 2016). If a trend is not linear, the mortality forecasted based on this trend could vary greatly depending on the historical period used in the estimation of the model (Janssen and Kunst 2007).

To ensure the robustness of mortality forecasting, it is essential that we determine the cause of non-linearity in mortality trends by studying past trends for a large number of countries (Janssen and Kunst 2007). The non-linearity in past mortality trends in Western European countries is mainly attributable to smoking (Janssen et al. 2007; Janssen et al. 2013; Lindahl-Jacobsen et al. 2016). As the full impact on mortality of the widespread uptake of smoking did not occur until 30 years later (Lopez et al. 1994), the influence of smoking resulted in a clear non-linear pattern in mortality, particularly among men. Making explicit adjustments for the distorting effects of smoking is likely to improve the accuracy of the overall mortality forecast (Janssen and Kunst 2007; Bongaarts 2014; Peters et al. 2016). Another option for improving mortality forecasts when the past trends are non-linear is to use the more linear trends of other countries as the underlying long-term trend in mortality (Janssen and Kunst 2007). The use of this approach could produce better estimates of the future direction of the mortality trends in a country with less linear trends. These types of methods are referred to as coherent forecasting methods (see, e.g., Li and Lee 2005).

Both approaches to improving mortality forecasts when past mortality trends are non-linear require additional information, such as information on smoking (direct or indirect estimations) or information on mortality trends in other countries. However, adding such information introduces more subjectivity into a mortality forecast because decisions have to be made about how the information will be incorporated into the forecasting method, and what kind of information will be included.

Thus, there is an important debate about whether only "objective" extrapolation methods should be employed even in cases of non-linearity, or whether it is preferable to include additional information, such as information on trends in other countries or smoking, even if doing so introduces additional subjectivity. To address this question, mortality forecasting approaches must be evaluated in the context of non-linear past mortality trends.

Most of the previous evaluation and comparison studies in the field of mortality forecasting did not consider different types of methods or approaches, such as both extrapolation methods and more explanatory approaches that include additional information. Furthermore, in these previous studies, little attention was paid to the effect of explicit assumptions; i.e., to the specific choices that must be explicitly stated in a method, such as the choice of the length of the historical period used in the estimation of the method (fitting period) and of the mortality rates used as the starting values of the mortality forecast (jump-off rates; i.e., the rates observed in the last year(s) or the rates estimated by the underlying mortality model). Moreover, previous evaluation studies assessed the performance of mortality forecasting methods using a quantitative approach that focused solely on their accuracy. It is, however, essential to evaluate these methods based on qualitative criteria as well (Cairns et al. 2011), such as the robustness and the plausibility of the outcomes of the mortality forecasting method. This PhD thesis will include these different approaches when evaluating the performance of mortality forecasting in the context of non-linear past mortality trends.

In addition to contributing to the debate on the degree of subjectivity associated with particular forecasting methods, this PhD thesis will generate results that can be used to improve the mortality forecasts of Statistics Netherlands. Thus, this study will provide important input for the official national population forecasts of Statistics Netherlands. The Netherlands is among the countries where past trends in mortality have been particularly non-linear (Van der Wilk et al. 2001; Janssen et al. 2003). This lack of regularity has made mortality forecasting, and, subsequently, population forecasting, in the Netherlands especially challenging. Previous methods that were employed by Statistics Netherlands were not able to fully deal with the non-linear past trends. Until 2012, mortality was forecasted by making assumptions about separate causes of death. Statistics Netherlands adopted a new method in 2012 based on recent research insights from Janssen and Kunst (2010) and Janssen et al. (2013). This new method makes use of extrapolation, but includes additional information on trends in other countries in Western Europe, and separately forecasts a clear non-linear pattern in smoking-attributable mortality (Stoeldraijer et al. 2012). The current PhD thesis provides a detailed analysis of the different components of this new approach, and the findings of this study can be used to evaluate, validate, and – ultimately – further improve the mortality forecasts, and, subsequently, the population forecasts, of Statistics Netherlands.

1.2 Objective and research questions

The aim of the current PhD research is to evaluate mortality forecasting in the context of non-linear past mortality trends.

The evaluation is comprised of (i) a quantitative and qualitative evaluation of not just different mortality forecasting models, but different mortality forecasting approaches; (ii) an assessment of the sensitivity of future mortality based on different explicit assumptions (e.g., fitting period, jump-off rates); and (iii) an evaluation of different elements of a mortality forecasting approach that deals with non-linear past mortality trends (e.g., the forecasting of smoking-attributable mortality, a model that forecasts mortality coherently).

The study is guided by the following research questions:

1) In a context in which mortality trends are non-linear, how does the choice of the mortality forecasting method and the explicit assumptions affect future forecasted mortality?
2) How can future levels of smoking-attributable mortality be formally estimated?
3) Which model should be used when the goal is to forecast mortality coherently, namely by taking into account the mortality experiences of other countries?
4) How can mortality forecasts be adjusted to take into account more recently observed data?

1.3 Background

1.3.1 Different mortality forecasting approaches

Mortality forecasting refers to the art and science of determining likely future mortality rates for a population. A forecast is an expectation of what is likely to happen; i.e., what is most likely to occur (De Beer 2011). It is primarily based on an assessment of historical trends and of the conditions for the continuation of these trends. There is a noteworthy distinction between a mortality forecast and a mortality projection: a mortality projection is what might occur. A projection is based on a technical calculation of a model that assumes that current trends will continue (De Beer 2011). Projections can also use hypothetical trends to answer "what-if" kinds of questions.

Only three decades ago, the methods used for mortality forecasting were relatively simple and involved a fair degree of subjective judgment. For example, a forecast might have consisted of a projection based on model life tables or data from another "more advanced" population (see Pollard 1987 for a review). But in the last two decades, more sophisticated models have been developed (Tabeau 2001; Wong-Fupuy and Haberman 2004; Booth and Tickle 2008; Cairns et al. 2011). The new models make increasing use of statistical methods drawn not only from demography, but from other fields of research, including epidemiology, actuarial science, spatial analysis, and Bayesian hierarchical modelling (Booth and Tickle 2008).

The mortality forecasting methods currently being used can be roughly divided into three types of approaches: extrapolation, explanation, and expectation (Booth and Tickle 2008). The extrapolation approach makes use of the regularity in age patterns and trends over time. The methods employed in this approach are the most objective; i.e., they reduce the role of subjective judgment by extrapolating historical trends based on the available data. The explanation approach makes use of (measurable) exogenous variables that are known to be related to certain causes of death. Examples of these approaches are extrapolation by cause of death and explanatory models based on mortality determinants. The expectation approach makes use of the subjective opinions of experts. In this approach, qualitative information and other relevant knowledge are incorporated into the forecast, such as the opinions of experts in demography or epidemiology. Setting a target of life expectancy for a date in the future is a commonly-used expectation method.

The majority of the mortality forecasting methods can be classified as extrapolative approaches. The Lee-Carter method (Lee and Carter 1992) is the dominant method of extrapolative mortality forecasting, and is frequently used as a benchmark for other methods that rely on extrapolation. The Lee-Carter method summarises mortality by age and period for a single population into an overall time trend, an age component, and the extent of change over time by age (Lee and Carter 1992). Mortality is forecasted by extrapolating the parameters for the overall time trend using time series methods, such as autoregressive-integrated-moving average (ARIMA) time series models (Box and Jenkins 1976; Tiao and Box 1981). Many studies since Lee and Carter (1992) have tried to improve upon their model by, for instance, adding more principal components, a cohort effect, a poisson-gamma setting, or a Bayesian version (among others: Booth et al. 2006; DeJong and Tickle 2006; Renshaw and Haberman 2006; Delwarde et al. 2007; Yang et al. 2010; Chen and Cox 2009; Li et al. 2009; Li et al. 2011; Deng et al. 2012; Li et al. 2013; Mitchell et al. 2013; Wisniowski et al. 2015; Ševčíková et al. 2016).

The major reason for the success of extrapolative forecasting methods is their congruence with historic trends. In many countries, the decline in mortality rates has been remarkably regular (see as well 1.3.2). Because extrapolation methods must be based on a steady, long-term trend, these methods work well for countries that exhibit such regular trends, and are now the leading approach for mortality forecasting (Tuljapurkar et al. 2000; Oeppen and Vaupel 2002; White 2002; Booth and Tickle 2008).

1.3.2 Past mortality trends in Western Europe

Over the 20th century, life expectancy in low-mortality countries increased enormously. In the early 1900s, the life expectancy at birth in Western Europe and other low-mortality countries was around 50 years (Kinsella 1992). Today, life expectancy in most Western European countries exceeds 80 years (United Nations 2017).

The historical increase in life expectancy is described in Omran's epidemiological transition theory (Omran 1971). According to this original epidemiological transition theory, all countries have experienced (or will eventually experience) three "ages": (1) the "age of pestilence and famine", during which mortality from infectious diseases is very high; (2) the "age of receding pandemics", during which life expectancy increases as mortality from infectious diseases at young ages decreases; and (3) the "age of the degenerative diseases and man-made diseases", during which the decline in mortality at younger ages gradually shifts towards older ages, with degenerative and man-made diseases like cardiovascular disease and cancers becoming the main causes of death. In the last age, life expectancy in all countries tends to converge towards the maximum level that has almost been reached by the most advanced countries. The timing and the duration of this transition vary across countries.

Omran (1971) thus described an overall transition from high levels of mortality from infectious diseases at young ages to high levels of mortality from cardiovascular diseases and cancers at old ages. He attributed the decrease in infectious diseases in low-mortality countries to modernisation, including improved nutrition, improved hygiene, and large-scale public health innovations.

As soon as Omran published his paper in 1971, the increasing life expectancy trends in Western Europe and other low-mortality countries continued. These further gains were due to socio-economic development and medical progress (Omran 1998; Mackenbach 2013). Since the 1970s, declines in mortality from cardiovascular diseases that were made possible by rapid innovations in medical treatments and

prevention have played an increasing role in improving life expectancy in many developed countries (Meslé and Vallin 2006).

Although life expectancy continued to increase in low-mortality countries in the latter decades of the 20th century, there were also signs of stagnation in some European countries, especially in Eastern European countries, which were hit by a health crisis starting in 1975; but also in some North-western European countries in the 1950s and the 1960s (e.g., Vallin and Meslé 2004). In a number of European countries – especially in Nordic countries, the United Kingdom, and the Netherlands; and particularly among men – life expectancy stagnated over longer periods of time in the 1950s and the 1960s. While life expectancy gains stalled in Northern Europe, in Southern European countries, where life expectancy in 1950 was lower than in Northern Europe because the standard of living was generally lower, life expectancy continued to advance. By 1970, the life expectancy gap between North and South was significantly reduced. Around 1980, male life expectancy in most Western European countries started to increase again (Janssen et al. 2004; Vallin and Meslé 2004; Kaneda and Scommegna 2011; Crimmins et al 2011). The gains registered in Western European countries did not, however, spread to Central and Eastern European countries. Due to the health crisis in that region, life expectancy stagnated (or even decreased), especially among men. Thus, by the mid-1990s, there was a huge East-West life expectancy gap in Europe. However, in some Western European countries, like the Netherlands and Denmark, life expectancy for women stagnated in the 1980s (Van der Wilk et al. 2001; Lindahl-Jacobsen et al. 2016).

These signs of stagnation have been described in Vallin and Meslé (2004), who used them as the basis for their convergence-divergence approach to the health transition. Briefly, their theory, which is based on empirical research, states that a succession of divergence-convergence movements will take place at different times from population to population (Vallin and Meslé 2004, 2005). They also posited that Omran's epidemiologic transition is the first stage of a global process of health transition; while the second stage (the cardiovascular revolution) is characterised by innovations in health from which some countries benefit, while others do not. These developments are expected to result in a trend towards divergence, followed by a trend towards convergence as late-entering countries are able to catch up to the pioneers. The authors further observed that progress in life expectancy made in the most advanced countries, especially among women, indicates that some countries are entering a third stage centred on the ageing process, which will initially lead to a new trend towards divergence between countries (again scattered between pioneers and those lagging behind), and then to a new trend towards convergence (after catching up).

The theory of Vallin and Meslé (2004) explains not just the remarkable similarities in life expectancy trends in Western Europe, but the variations in slopes between countries. Furthermore, there is evidence that behaviour and lifestyle factors (and the knowledge thereof) are becoming increasingly important for life expectancy progress in many countries (O'Doherty et al. 2016; Li et al. 2018). Smoking, alcohol consumption, diet, and exercise have all contributed to the success (or failure) of life expectancy advances.

The periods of stagnation and acceleration in mortality trends are more problematic for mortality forecasting, which relies heavily on the extrapolation of past trends. To ensure the robustness of mortality forecasting, it is essential that we determine the causes of the non-linearity in mortality trends by studying past trends for a large number of countries (Janssen and Kunst 2007).

1.3.3 Important role of smoking in past non-linear mortality trends

The unfavourable developments in life expectancy among men in many North-western European countries in the 1950s and the 1960s are related to changes in lifestyle after the Second World War (i.e., smoking) (Vallin and Meslé 2004). Differences between countries in the timing and the size of the smoking epidemic, the lagged effect of smoking on death rates, and the mortality declines following cessation all help to explain the mortality trends and the differences in mortality levels observed among countries since the middle of the 20th century (Janssen et al. 2007; Janssen et al. 2013; Lindahl-Jacobsen et al. 2016). The extended period of relative stagnation in female life expectancy that some countries (Denmark, the Netherlands, and England and Wales) experienced in the 1980s and 1990s is also a legacy of heavy smoking among women in these countries since the Second World War (Lindahl-Jacobsen et al. 2016).

The adverse impact of smoking on health and mortality is well established (CDC 2010; Ezzati et al. 2003; Doll et al. 2004; Jha and Peto 2014; Peto et al. 1992; Peto et al. 2012; Preston, Glei, and Wilmoth 2010a). In addition to being responsible for the large majority of lung cancer deaths worldwide, smoking has been shown to increase mortality from other cancers, cardiovascular diseases, and most other diseases. Furthermore, smoking is the most important preventable risk factor in the European Union (WHO 2009).

In general, as was described in the smoking epidemic model proposed by Lopez et al. (1994), men in Anglo-Saxon countries were the first to take up smoking in the early 20th century. After a rapid rise lasting two or three decades, male smoking prevalence started to decline. Smoking-attributable mortality (i.e., the number of all deaths in a population caused by smoking) followed the increase and the subsequent decline in smoking prevalence some 30–40 years later. The increase in smoking prevalence generally started about 20 years later for women than for men, but, depending on the country, this period may have been shorter or longer. As the maximum levels of female smoking prevalence were considerably lower than those for men, smoking-attributable mortality was also lower among women than among men. It is posited in the last stage of the original smoking epidemic model that declines in smoking prevalence will reach similar levels for men and women, which suggests that smoking-attributable mortality for men and women should converge in the future (McCartney et al. 2011; Lopez et al. 1994). However, smoking-attributable mortality for women has continued to increase during this last stage. Currently, some countries, such as England and Wales, have already experienced the peak in smoking-attributable mortality for women (Thun et al. 2013). In other countries in Northern and Western Europe, such as Denmark and the Netherlands, this peak appears to be approaching, as the peak in smoking prevalence for women has passed (Janssen et al. 2013; Lindahl-Jacobsen et al. 2016).

Patterns of smoking behaviour and the accompanying patterns of smoking-attributable mortality have changed enormously over time. Indeed, smoking has been the most important non-linear determinant of mortality in low-mortality countries in recent decades. Furthermore, patterns of smoking behaviour and, consequently, of smoking-attributable mortality differ greatly by country, and have contributed to the emergence of a large gender gap in mortality (McCartney et al. 2011; Lopez et al. 1994). Ignoring the smoking epidemic yields a bias in the forecast of life expectancy, especially if the method used relies on extrapolation of past observed mortality trends (Janssen & Kunst 2007). Making explicit adjustments for the distorting effects of smoking is likely to improve the accuracy of forecasts (Janssen and Kunst 2007; Bongaarts 2014; Peters et al. 2016).

1.3.4 Dealing with non-linear past mortality trends in mortality forecasting

Non-linear past trends in mortality pose additional challenges when forecasting mortality. If the trend is not linear, the forecasted mortality could be very different depending on the historical period used in the estimation of the model (Janssen and Kunst 2007).

Thus, when dealing with non-linear past mortality trends, it is essential to determine the cause of the non-linearity by studying past trends for a large number of countries (Janssen and Kunst, 2010). When the cause is known (and measurable), it can be incorporated into the forecasting method.

As was detailed in section 1.3.3, past smoking behaviour has been established as an important factor in the non-linearity of past mortality trends in the Netherlands and in many other Western European countries, especially for men. For this reason, a few studies have explicitly adjusted mortality projections to account for the impact of smoking (e.g., Pampel 2005; Bongaarts 2006; Janssen and Kunst 2007; Girosi and King 2008; Wang and Preston 2009; Technical Panel on Assumptions and Methods 2011; Janssen, van Wissen, and Kunst 2013; Preston et al. 2014). The forecasting approaches used in these papers differ. Bongaarts (2006), Janssen and Kunst (2007) and Technical Panel on Assumptions and Methods (2011) employed an approach that looked at developments in mortality and life expectancy without smoking. Pampel (2005) and Preston et al. (2014) used information on smoking prevalence to forecast smoking-related mortality. Girosi and King (2008) and Wang and Preston (2009) included covariates for smoking within the forecasting method of total mortality. Janssen, van Wissen, and Kunst (2013) separately projected smoking- and non-smoking-related mortality. The different approaches were chosen in part based on the availability of adequate data. Because more assumptions are required in a method that incorporates smoking, a trade-off must be made between the advantage of being able to take the impact of smoking into account and the advantage of the objectivity of a pure extrapolation approach based on total mortality.

When the cause of the non-linearity is unknown, or the cause cannot be quantified within the forecasting method, an approach that can be used to account for the non-linearity is coherent mortality forecasting (Janssen and Kunst 2007). Coherent forecasting methods, whereby "coherent" refers to non-divergent forecasts for sub-populations within a larger population (Li and Lee 2005), were introduced to ensure that divergence as a result of individual forecasting does not occur. The

scholars who proposed these methods observed that mortality patterns and trajectories in closely related populations are likely to be similar in some respects, and that differences are unlikely to increase in the long run. Thus, they argued, experiences in other countries can be used to create a broader empirical basis for the identification of the most likely long-term trend (Janssen et al. 2013; Shair et al. 2017). In other words, the approach assumes that countries with more linear mortality trends could provide better information about the future direction of the mortality trends in a country with less linear trends than the country's own past trends.

In coherent forecasting methods, non-divergence is derived by applying constraints to the parameters of individual forecasts of multiple populations. Most existing coherent forecasting methods are based on the Lee-Carter structure (Carter and Lee 1992; Li and Lee 2005; Li and Hardy 2011; Zhou et al. 2012; Zhou et al. 2013; Yang and Wang 2013; Wan et al. 2013; Kleinow 2015), but there are also methods based on the age-period-cohort structure (Dowd et al. 2011; Cairns et al. 2011a; Jarner and Kryger 2011; Börger and Aleksic 2014) and the functional data paradigm (Hyndman et al. 2013; Shang and Hyndman 2016). Other structures are usually more complex. Even within a single structure, these coherent forecasting methods can differ greatly. So far, few of these methods have been compared in terms of the accuracy of their forecasts (Shang 2016; Enchev et al. 2016; Shair et al. 2017).

A method that simultaneously takes into account smoking and the experiences of other countries was proposed by Janssen et al. (2013). The idea behind their methodology is as follows: by first removing smoking from the mortality trends for each country, the actual long-term trend in mortality driven by socio-economic developments and medical care improvements can be identified. This more linear trend of non-smoking-attributable mortality may be expected to converge across countries, and can then be used in the coherent forecasting method. The non-linear past trend in smoking-attributable mortality, which cannot be captured by age-period modelling or projection, must be projected separately, and subsequently combined with the forecast of non-smoking-attributable mortality. The inclusion of epidemiological information can thus generate a more robust long-term trend that may be used as a basis for projection (Janssen et al. 2013), thereby lessening dependence on the historical period.

1.3.5 Mortality forecasting by Statistics Netherlands

Statistics Netherlands regularly publishes a mortality forecast (Gjaltema and Broekman 2002; Stoeldraijer et al. 2017). The mortality forecast is part of the population forecast, which currently follows a three-year cycle. An extensive

population forecast is issued once every three years, with adjustments being made in the intermediate years. In the intervening years, the adjusted population forecast is supplemented with a household forecast in the first year and a population and household forecast on the municipality level in the second year. The adjustments to the mortality forecast made in the intervening years include a re-estimation of the current forecast method based on the most recent data available, but usually include no changes to the method itself.

The mortality forecast published by Statistics Netherlands in 1950 assumed that mortality rates would remain constant (Gjaltema and Broekman 2002). Because it underestimated the development in life expectancy, the 1951 forecast used an extrapolation of the decrease in five-year mortality rates. However, this still underestimated the development in life expectancy: between 1950 and 1970, life expectancy increased 0.3 years per decade for men and 2.0 years per decade for women. In the forecast published in 1965, extrapolation was increased for the initial years of the forecast period, but mortality rates were again kept constant after 15 years of the forecast period. In 1970, a forecast with four causes of death was introduced. Because the added uncertainty associated with the breakdown was estimated to be too large and the increase in life expectancy in that period was minimal (especially for men), the mortality rates used in the 1975 forecast were again kept equal to the observed rates (over the 1971-1974 period), with a small extrapolation for some ages. However, between 1970 and 1980, life expectancy increased 1.7 years for men and 2.7 years for women.

In its 1980 forecast, Statistics Netherlands used a limit for life expectancy at certain ages after 10 years of the forecast period (Gjaltema and Broekman 2002). The limit was set based on a literature review and consultation with experts from the Netherlands and abroad. It was expected that in the near future, the negative impacts on the life span of the population of certain socio-economic, cultural, and technological developments would not outweigh the positive impacts of developments in medicine, hygiene, nutrition, and preventive health care. It was thus assumed that mortality rates would decline further, and that the excess mortality of men would decrease slightly. After the 10-year period, the mortality rates were kept constant. For the forecasts after 1980, the limit was raised a few times in response to increasing life expectancy. In 1996, the limit was determined for 2050 instead of for 10 years in the future. Because it was assumed that achieving additional increases in life expectancy would become more and more difficult, it was anticipated that the increasing trend would level off in the future.

For its 2002 forecast, Statistics Netherlands used an explanatory model based on life expectancy at birth (de Jong 2003). In this model, the effects of underlying

factors on mortality were taken into account to a limited extent. Therefore, in the forecasts it issued between 2004 and 2012, Statistics Netherlands forecasted mortality using the extrapolation of trends by cause of death (de Jong 2005). This made it possible to include determinants and model non-linearities. However, because a very large number of assumptions were required in applying this method, the model was ultimately seen as too time-consuming and lacking in transparency. In addition, it was found that obtaining well-founded expert expectations about future developments per cause of death was difficult when using this method, and that the level of detail required by the model made it hard to include international trends. Yet more fundamental objections to the use of this type of model were also raised, including that it can allow the cause of death with the least favourable development to dominate the overall future trend in mortality (Wilmoth 1995); and that extrapolating trends per cause of death can paint an overly pessimistic picture, especially over the long term.

Because of the problems associated with the use of these approaches (i.e., underestimation of life expectancy and non-linearity in the trend), Statistics Netherlands adopted a new method in 2012. The method is a refinement of the method proposed in Janssen and Kunst (2010) and Janssen et al. (2013), which was described in the last paragraph of the previous section. To reiterate, the new methodology takes into account mortality trends in other European countries, and systematically includes in the calculation information about developments in smoking. The new methodology is in line with existing evidence that smoking plays an important role in mortality trends in the Netherlands, and it places mortality fluctuations not attributable to smoking in an international context. The mortality forecasting method used by Statistics Netherlands is explained (in Dutch) in Stoeldraijer et al. (2012, included in the Annex of this PhD thesis). The method for forecasting smoking-attributable mortality and the jump-off rates were refined.

The new mortality forecasting method used by Statistics Netherlands requires researchers to make a number of explicit choices. The estimation of smoking-attributable mortality is based on the extrapolation of lung cancer mortality through the use of age-period-cohort analyses and the smoking epidemic model (Lopez et al. 1994). An indirect estimation technique is applied to the observed and forecasted levels of lung cancer mortality in order to estimate the observed and forecasted levels of smoking-attributable mortality (Rostron 2010). To coherently forecast non-smoking-attributable mortality, the Li-Lee method (Li and Lee 2005) is used (with Denmark, England and Wales, Finland, France, Germany, Italy, Norway, Spain, Sweden, and Switzerland serving as the main group of countries), following the work of Janssen et al. (2013). The Li-Lee method is essentially the Lee-Carter method, but is then applied twice, first to the group of countries, and then to the

difference between the group and the country of interest. The last observed mortality rates are used as the initial jump-off rates. However, these choices have yet to be evaluated.

1.3.6 Previous evaluation of the performance of mortality forecasting models

As these new mortality forecasting models were being developed, approaches for evaluating their performance were also proposed. Many of the previous evaluation and comparison studies in the field of mortality forecasting considered one method or similar methods within the same approach. For example, the extensions of the Lee-Carter method have been compared with the original method (among others, Wilmoth 1993; Lee and Miller 2001; Booth et al. 2002; Li and Lee 2005; Renshaw and Haberman 2006; Li et al. 2006; Booth et al. 2006; Shang et al. 2011; Li et al. 2013).

Previous studies often assessed the performance of mortality forecasting models using a quantitative approach that focused solely on their accuracy (Cairns et al. 2009). There are several measures that can be used to summarise the accuracy of forecasting methods. Most of these measures are based on the error of the model or forecast compared to the actual values of death rates, life expectancy, and other relevant statistics. Examples of such measures are the explanation ratio (ER), the root mean squared error (RMSE), the Bayes information criterion (BIC), and the mean absolute (percent) error (MA(P)E). Particularly, as coherent forecasting methods are relatively new, few have been compared in terms of forecast accuracy (Shang 2016; Enchev et al. 2016; Shair et al. 2017). Among the other more qualitative criteria that have been used to evaluate forecasting models are biological reasonableness, the plausibility of predicted levels at different ages, and the robustness of the forecasts relative to the sample period used to fit the model (Cairns et al. 2009). Because these criteria are more qualitative, a visual comparison is generally used in these evaluations (Cairns et al. 2009).

Most previous evaluations of mortality forecasting focused purely on the performance of mortality forecasting models. However, recent studies (Booth et al. 2002; Janssen and Kunst 2007) have noted the importance of explicit assumptions. An explicit assumption is a specific choice that must be explicitly stated in a method, such as the choice of the length of the historical period and of the jump-off rates (i.e., the starting values of the actual mortality forecast). Previous research has shown that the historical period used is the main determinant of the

large differences in the outcomes of mortality forecasts (Janssen and Kunst 2007), especially when there is considerable non-linearity in the trends. In coherent forecasting, the choice of the main group of countries influences the outcome, because the main group determines the long-term trend of a specific country in the coherent mortality forecast (Li and Lee 2005). Moreover, while choosing appropriate jump-off rates is a practical consideration in every mortality forecast (regardless of the method used), it is essential for matching the mortality forecast to the most recently observed data, and thus influences the performance of the forecast. Choosing different jump-off rates can improve the accuracy of a single forecast and/or reduce the discontinuity between the last observed death rate and the first forecasted death rate. However, when successive forecasts differ from each other because different jump-off rates were chosen, the robustness of the forecast is affected.

1.4 Approach

The approach used in this PhD thesis is both academic and practical. It is academic because the thesis contributes to the academic debate on degrees of subjectivity in forecasting methods; and because it supports the further development of mortality forecasting approaches and methods, especially in situations in which the trends are not linear. It is practical because the findings of this PhD thesis can be used to improve the mortality forecasts issued by Statistics Netherlands.

The evaluation approach adopted in this PhD thesis differs from those used in previous evaluation studies. In addition to evaluating different mortality forecasting methods, the thesis evaluates different forecasting approaches (i.e., extrapolation, explanation, and expectation). In the course of evaluating the approaches to and the methods for forecasting mortality, both quantitative and qualitative criteria will be examined: i.e., accuracy (fit to historical data), robustness (stability across different fitting periods), and the plausibility of results (smooth continuation of trends from the fitting period) (Cairns et al. 2009). The focus of the evaluation is not only on the performance of the model, but on the sensitivity of the outcome to underlying explicit assumptions, such as the jump-off rates and the main group of countries chosen. Furthermore, the different elements of a mortality forecasting method that deals with non-linear past mortality trends are evaluated.

This PhD research adopts a data-driven approach. First, although the focus of the PhD thesis is on the Netherlands, other Western European countries are also

studied. The inclusion of data from these other countries made it possible to assess how different past trends, especially linear versus non-linear trends, affect the performance of different mortality forecasting approaches and methods; and to relate the differential effects of the explicit assumptions to previously observed national past trends. In addition, by evaluating mortality models for different countries (e.g., models for forecasting smoking-attributable mortality; coherent mortality modelling), we are able to obtain stronger evidence regarding their performance. A second element of the data-driven approach is that it was possible to make ample use of the already observed past trends. In addition to assessing the model fit and to comparing the future outcomes of the forecasts, it was possible to compare the outcomes forecasted with part of the data and the actual observed values. Third, it should be noted that the majority of mortality forecasting approaches that are being evaluated are also data-driven (Booth et al. 2008), and either consist of the pure extrapolation of past trends in age-specific mortality, or include additional data on either the smoking epidemic or past mortality trends in other countries.

By focusing on the evaluation of the different elements and the explicit assumptions of the mortality forecasting approach used by Statistics Netherlands (e.g., the separate projection of smoking-attributable mortality and the coherent forecasting of non-smoking-attributable mortality), this PhD thesis will contribute to the evaluation, validation, and further development of the mortality forecasts issued by Statistics Netherlands.

1.5 Data and methods

To answer the research questions, the PhD thesis employs both a review of existing forecasting approaches and methods, and an actual evaluation of different forecasting approaches and methods.

In the review of the existing forecasting methods, the current methods for forecasting mortality used by statistical offices in Europe and the different national and international forecasts/projections that exist for the Netherlands are outlined.

In the actual evaluation, this PhD thesis uses data on mortality (all-cause and cause-specific; i.e., lung cancer), population exposure data, and data on smoking prevalence. The data are obtained by sex, age, year (between 1950 and 2014), and country.

The analyses are done separately for men and women, except for the analyses in Chapter 5, for which sex was irrelevant (however, sex-specific analyses are included in the Appendix of Chapter 5).

Most of the data are divided into five-year age groups (0, 1-4, 5-9, ..., 90-94, 95+), but single ages are also used (e.g., in Chapter 5). For specific research questions, only some of the ages are analysed: in Chapter 3, the ages at which smoking-attributable mortality is relevant (40+) are analysed; and in Chapter 5, the ages at which pension reforms are relevant are analysed (65+).

The Netherlands is used as a case study, but other Western European countries are also analysed to extend the conclusions more broadly. The focus is on national populations. Results are predominantly presented for Belgium, Denmark, England and Wales / the United Kingdom, Finland, France, Italy, Norway, Spain, Sweden, Switzerland, and West Germany. Chapter 4 uses as well other countries that are included as part of the main group.

The data used in this thesis have been obtained from various sources. Statistics Netherlands is the source for the data from the Netherlands (all-cause, cause-specific, and lung cancer mortality data; and population exposure data). The Human Mortality Database (HMD, www.mortality.org) is the source for the all-cause mortality and population exposure data from all other countries. The WHO Statistical Information System (WHOSIS, http://www.who.int/healthinfo/statistics/mortality_rawdata/en/) is the source for the lung cancer mortality data for all countries. The data on smoking prevalence are obtained from Cancer Research UK, The Dutch Expert Centre on Tobacco Control, the International Smoking Statistics WEB Edition, the Organization for Economic Co-operation and Development Health Data, and the World Health Organization.

This PhD thesis applies different mortality forecasting techniques to these data in order to address the general objective. These techniques include individual forecasting methods: (i) direct linear extrapolation; (ii) the Lee-Carter model (Lee and Carter 1992); (iii) an extension of the Lee-Carter model that includes a cohort dimension (Renshaw and Haberman 2006); and (iv) the method used between 2004 and 2010 in the official forecast issued by Statistics Netherlands (extrapolation by cause-of-death) (De Jong 2004). This thesis also uses the following coherent forecasting methods: (i) the Li-Lee method (Li and Lee 2005); (ii) the co-integrated Lee-Carter method (Li and Hardy 2011; Cairns et al. 2011a); and (iii) the coherent functional data method (Hyndman et al. 2013). To include smoking in the forecast, a model in which smoking-related and non-smoking-related mortality is projected separately (Janssen and Kunst 2010; Janssen et al.

2013) is used. Age-period-cohort (APC) analysis is used for the extrapolation of lung cancer mortality. Indirect estimation techniques are applied to the observed and the projected levels of lung cancer mortality to obtain the observed and the projected levels of smoking-attributable mortality (an adapted and simplified version of the indirect Peto-Lopez method, Peto et al. 1992; Rostron 2010; Preston et al. 2010).

To evaluate these methods and the explicit assumptions chosen, different approaches are employed. The evaluation is comprised of (i) an assessment of the model fit based on past trends from 1950 onwards; (ii) a forecast based on part of the data and a comparison of the outcomes with actual observed values (in-sample forecasting); and (iii) a comparison of the future outcomes (i.e., for the years 2020, 2030, 2040, or 2050) from different forecasts (out-of-sample forecasting). The outcomes – life expectancy at birth or at age 65 up to 2050 – of the different forecasting methods are compared visually, whereas the other comparisons are mostly done in a tabular manner.

The evaluation is based on both quantitative (i.e., focused on accuracy) and qualitative (i.e., focuses on robustness and the plausibility of the results) evaluation criteria (Cairns et al. 2009). To test the degree of accuracy (fit to historical data), the following measures are used: the explanation ratio (ER); the root mean squared error (RMSE); the mean absolute percent error (MAPE) of the log death rates averaged over ages and years; and the mean absolute error (MAE) of the forecasted life expectancy at age 65. To test the degree of robustness (stability across different fitting periods), the standard deviation of the life expectancy at birth (e0) in 2050 resulting from the use of the three fitting periods, averaged over the seven countries and the three selected main country groups, and the standard deviation (SD) in the increase/decrease of the (out-of-sample) life expectancy at age 65 in a given year in the future are used. To evaluate whether the results are plausible (smooth continuation of trends from the fitting period), the following measures are used: (i) the standard deviation of e0 in 2050 resulting from the selection of the three main country groups, averaged over the seven countries and the three fitting periods; (ii) the standard deviation of e0 in 2050 resulting from the mortality forecasts for the seven countries, averaged (unweigthed) over the three main country groups and the three different fitting periods; and (iii) the improvement of the mortality rates by age between the last year of the fitting period and 2050.

1.6 Outline

This thesis consists of six chapters. The current first chapter introduces the topic of this thesis.

Chapters 2 to 5 each answer one of the four different research questions. Chapter 2 reviews the different mortality forecasting methods and their assumptions in Europe, and assesses their impact on projections of future life expectancy for the Netherlands. More specifically, (i) the current methods used in official mortality forecasts across Europe are reviewed; (ii) the outcomes and the assumptions of different projection methods used within the Netherlands are compared; and (iii) the outcomes of different types of methods based on similar explicit assumptions, including the same historical period, are compared for the Netherlands.

In Chapter 3, a formal estimation of future levels of smoking-attributable mortality up to 2050 is presented for the total national populations of England and Wales, Denmark, and the Netherlands. An update and an extension of the descriptive smoking epidemic model are provided in the estimation.

In Chapter 4, different coherent forecasting methods are evaluated in terms of their accuracy (fit to historical data), robustness (stability across different fitting periods), subjectivity (sensitivity to the choice of the group of countries), and plausible outcomes (smooth continuation of trends from the fitting period) for France, Italy, the Netherlands, Norway, Spain, Sweden, and Switzerland up to 2050.

In Chapter 5, an evaluation of six different options for the jump-off rates and an examination of their effects on the robustness and accuracy of the mortality forecast are presented for Belgium, Finland, France, the Netherlands, Norway, Spain, Sweden, and the United Kingdom. The focus of the chapter is on life expectancy at age 65.

Finally, in Chapter 6, the main findings of the PhD thesis as a whole are summarised and discussed. The implications of these findings for mortality forecasting in the Netherlands, mortality forecasting in general, future research, and policy are also explored.

References

Bengtsson, T. and Christensen, K. (eds) (2006). *Perspectives on Mortality Forecasting*. IV. Causes of Death. vol. IV, Swedish Social Insurance Agency, Stockholm.

Bongaarts, J. (2006). "How long will we live?" *Population and Development Review* 32(4): 605–628.

Bongaarts, J. (2014). Trends in Causes of Death in Low-Mortality Countries: Implications for Mortality Projections. *Population and Development Review* 40(2): 189–212.

Booth, H., Maindonald, J. and Smith, L. (2002). Applying Lee-Carter under conditions of variable mortality decline. *Population Studies* 56 (3): 325–336.

Booth, H., Hyndman, R., Tickle, L. and de Jong, P. (2006). Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions. *Demographic Research* 15(9): 289–310.

Booth, H. and Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. *Annals of Actuarial Science* 3(1-2): 3-43. doi:10.1017/S17484 99500000440.

Börger, M. and Aleksic, M-C. (2014). *Coherent Projections of Age, Period, and Cohort Dependent Mortality Improvements*. Paper presented at the Living to 100 Symposium, Orlando, Fla., January 8–10, 2014.

Box, G.E.P. and Jenkins, G.M. (1975). *Time Series Analysis*. Forecasting and Control, San Francisco: Holden Day.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., Ong, A., and Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from England & Wales and the United States. *North American Actuarial Journal* 13(1): 1-35. doi:10.1080/10920277.2009.10597538.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., and Khalaf-Allah, M. (2011). Mortality density forecasts: An analysis of six stochastic mortality models. *Insurance: Mathematics and Economics* 48(3): 355–367. doi:10.1016/j. insmatheco.2010.12.005.

Cairns, A.J.G, D. Blake, L. Dowd, G.D. Coughlan and M. Khalaf-Allah. (2011a). Bayesian Stochastic Mortality Modelling for Two Populations. *Astin Bulletin* 41(1): 29–59.

Carone, G., Eckefeldt, P., Giamboni, L., Laine, V. and Pamies Sumner, S. (2016). *Pension Reforms in the EU since the Early 2000's: Achievements and Challenges Ahead*. European Economy, Discussion Papers 42. December 2016. Brussels. 64pp. DOI: 10.2765/620267.

Carter, L.R. and Lee, R.D. (1992). Modeling and forecasting US sex differentials in mortality *International Journal of Forecasting* 8(3), 393–411. https://doi.org/10.1016/0169-2070(92)90055-E

Centers for Disease Control and Prevention (CDC). 2010. A Report of the Surgeon General: How Tobacco Smoke Causes Disease. Washington, DC: Government Printing Office.

Crimmins, E.M., Preston, S.H. and Cohen, B. (eds) Panel on Understanding Divergent Trends in Longevity in High-Income Countries; Committee on Population; Division of Behavioral and Social Sciences and Education; National Research Council (2011) *Explaining Divergent Levels of Longevity in High-Income Countries*, The National Academies Collection: Reports Funded by National Institutes of Health (Natl Acad Press, Washington, DC).

Currie, I.D., Durban, M. and Eilers, P.H.C. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling* 4: 279–298.

Doll, R., Peto, R., Boreham, J. and Sutherland, I. (2004). Mortality in relation to smoking: 50 years' observations on male British doctors. *British Medical Journal* 328: 1519–1533.

Dowd, K., Blake, D., Cairns, A.J.G., Coughlan, G.D. and Khalaf-Allah, M. (2011). A gravity model of mortality rates for two related populations. *North American Actuarial Journal* 15: 334–356.

Enchev, V., Kleinow, T. and Cairns, A.J.G. (2016). Multi-population mortality models: Fitting, Forecasting and Comparisons. *Scandinavian Actuarial Journal* (forthcoming).

European Commission (2014). *Population ageing in Europe. Facts, implications and policies*. France. doi:10.2777/60452.

Ezzati M. and Lopez, A. (2003). Measuring the accumulated hazards of smoking: global and regional estimates for 2000. *Tobacco Control* 12: 79–85.

Girosi, F. and King, G. (2008). *Demographic Forecasting*. Princeton University Press.

Gjaltema, T. and Broekman, R. (2002). Vijftig jaar bevolkingsprognoses: voorspelling van de sterfte. *Maandstatistiek van de Bevolking 50: 12-24.*

Hyndman, R.J., Booth, H., and Yasmeen, F. (2013). Coherent mortality forecasting: The product-ratio method with functional time series models. *Demography* 50(1): 261-283. doi:10.1007/s13524-012-0145-5.

Jacobsen, R., Von Euler, M., Osler, M., Lynge, E. and Keiding, N. (2004). Women's death in Scandinavia--What makes Denmark different? *European Journal of Epidemiology* 19(2):117–121.

Jacobsen, R., Keiding, N. and Lynge, E. (2006). Causes of death behind low life expectancy of Danish women. *Scandinavian Journal of Public Health* 34(4):432–436.

Janssen, F., Nusselder, W.J., Looman, C.W.N., Mackenbach, J.P. and Kunst, A.E. (2003). Stagnation in mortality decline among elders in The Netherlands. *The Gerontologist* 43: 722-734. doi: 10.1093/geront/43.5.722.

Janssen, F., Mackenbach, J.P. and Kunst, A.E. (2004). Trends in old-age mortality in seven European countries, 1950-1999. *Journal of Clinical Epidemiology* 57(2): 203-216. doi: 10.1016/j.jclinepi.2003.07.005.

Janssen, F. and Kunst, A. (2007). The choice among past trends as a basis for the prediction of future trends in old-age mortality. *Population Studies* 61: 315–326.

Janssen, F., Kunst, A.E. and Mackenbach, J.P. (2007). Variations in the pace of old-age mortality decline in seven European countries, 1950–1999: the role of smoking and other factors earlier in life. *European Journal of Population 23(2): 171-188.*

Janssen, F. and Kunst, A. (2010). De toekomstige levensverwachting. In: Luijben, A.H.P. and Kommer, G.J. (eds.). *Tijd en toekomst*; deelrapport van de VTV 2010 Van gezond naar beter. RIVM-rapport 270061008, Houten: Bohn Stafleu Van Loghum: 13-20.

Janssen, F., van Wissen, L. and Kunst, A. (2013). Including the smoking epidemic in internationally coherent mortality projections. *Demography* 50: 1341–1362.

Jarner, S.F. and Kryger, E.M. (2011). Modelling Adult Mortality in Small Populations: The SAINT Model. *Astin Bulletin* 41(2): 377–418.

Jha, P. and Peto, R. (2014). Global effects of smoking, of quitting, and of taxing tobacco. *New England Journal of Medicine* 370(1): 60–68.

De Jong, A. (2003). Bevolkingsprognose 2002–2050: veronderstellingen. *Bevolkingstrends* 1e kwartaal 2003.

De Jong, A. (2005). Bevolkingsprognose 2004–2050: veronderstellingen. *Bevolkingstrends* 1e kwartaal 2005.

Kaneda, T. and Scommegna, P. (2011). *Today's Research on Aging*. Population Reference Bureau; Washington, DC. Trends in Life Expectancy in the United States, Denmark, and the Netherlands: Rapid Increase, Stagnation, and Resumption.

Kinsella. KG. (1992). Changes in life expectancy 1900-1990. *American Journal of Clinical Nutrition* 55:1196S–1202S. doi: 10.1093/ajcn/55.6.1196S.

Kleinow, T. (2015). A common age effect model for the mortality of multiple populations. *Insurance: Mathematics and Economics* 63: 147–152.

Lee, R.D. and Carter, L.R. (1992). Modelling and forecasting US mortality. *Journal of the American Statistical Association* 87(419): 659–671.

Lee, R. and Miller, T. (2001). Evaluating the Performance of the Lee-Carter Approach to Modeling and Forecasting Mortality. *Demography* 38(4): 537–549.

Li, N.R. and Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography* 42(3): 575–594. doi:10.1353/dem.2005.0021.

Li, S-H, Hardy, M.R. and Tan, K.S. (2006). *Uncertainty in mortality forecasting: an extension of the classical Lee-Carter approach*. University of Waterloo, Ontario N2L 3GI.

Li, J.S-H. and Hardy, M.R. (2011). Measuring Basis Risk in Longevity Hedges. *North American Actuarial Journal* 15(2): 177–200.

Li, N., Lee, R. and Gerland, P. (2013) Extending the Lee-Carter method to model the rotation of age patterns of mortality-decline for long-term projection. *Demography* 50(6):2037–2051. doi: 10.1007/s13524-013-0232-2

Li, Y., Pan, A., Wang, D.D., Liu, X., Dhana, K., Franco, O.H., Kaptoge, S., Di Angelantonio, E., Stampfer, M., Willett, W.C. and Hu, F.B. (2018). Impact of Healthy Lifestyle Factors on Life Expectancies in the US Population. *Circulation*. https://doi.org/10.1161/CIRCULATIONAHA.117.032047

Lindahl-Jacobsen, R., Oeppen, J., Rizzi, S., Moller, S., Zarulli, V., Christensen, K., et al. (2016). Why did Danish women's life expectancy stagnate? The influence of interwar generations' smoking behaviour. *European Journal of Epidemiology* 31(12):1207-1211.

Lopez, A., Collishaw, A., and Piha, T. (1994). A descriptive model of the cigarette epidemic in developed countries. *Tobacco Control* 3: 242–7.

Mackenbach, J.P. (2013). Convergence and divergence of life expectancy in Europe: a centennial view. *European Journal of Epidemiology* 28(3): 229-240.

McCartney, G., Mahmood, L., Leyland, A.H., Batty, G.D. and Hunt, K. (2011). Contribution of smoking-related and alcoholrelated deaths to the gender gap in mortality: evidence from 30 European countries. *Tobacco Control* 20: 166–8. doi:10.1136/tc.2010.037929.

Meslé, F. and Vallin, J. (2006). The health transition: Trends and prospects. In Caselli, G., Vallin, J. and Wunsch, G. (eds.). *Demography: Analysis and Synthesis, Vol. 2.* Elsevier: 247–266.

O'Doherty, M.G., Cairns, K., O'Neill, V., Lamrock, F., Jørgensen, T., Brenner, H., Schöttker, B., Wilsgaard, T., Siganos, G., Kuulasmaa, K., Boffetta, P., Trichopoulou, A. and Kee, F. (2016). Effect of major lifestyle risk factors, independent and jointly, on life expectancy with and without cardiovascular disease: results from the Consortium on Health and Ageing Network of Cohorts in Europe and the United States (CHANCES). *European Journal of Epidemiology* 31: 455–468. doi: 10.1007/s10654-015-0112-8

OECD (2015). *Pensions at a Glance 2015: OECD and G20 indicators*. OECD Publishing, Paris. http://dx.doi.org/10.1787/pension_glance-2015-en

Oeppen, J. and Vaupel, J.W. (2002). Demography. Broken limits to life expectancy. *Science* 296(5570):1029–31.

Olshansky, S.J. and Ault, A.B. (1986). The fourth stage of the epidemiologic transition: the age of delayed degenerative diseases. *Milbank* Q 64(3):355–91.

Omran, A.R. (1971). The epidemiologic transition. A theory of the epidemiology of population change. *Milbank Mem Fund* Q 49(4):509–38.

Omran, A.R. (1998). The Epidemiologic transition theory revisited thirty years later. *World Health Statistics Quarterly* 52: 99–119.

Pampel, F.C. (2005). Forecasting sex differences in mortality in high income nations: The contribution of smoking. *Demographic Research* 13(18): 455–484.

Peto R., Lopez A., Boreham J., Thun M. and Heath Jr C. (1992). Mortality from tobacco in developed countries: indirect estimation from national statistics. *Lancet* 339: 1268–78.

Peto, R., Lopez, A., Boreham, J. and Thun, M. (2012). Mortality from Smoking in Developed Countries 1950–2005 (or later). http: //www.ctsu.ox.ac.uk/~tobacco/.

Pollard, J.H. (1987). Projection of age-specific mortality rates. Population Bulletin of the United Nations 21-22: 55-69.

Preston S., Glei D. and Wilmoth J. (2010). A new method for estimating smoking-attributable mortality in high-income countries. *International Journal of Epidemiology* 39: 430–8.

Preston, S.H., Stokes, A. Mehta, N.K. and Cao, B. (2014). Projecting the effect of changes in smoking and obesity on future life expectancy in the United States. *Demography* 51: 27–49.

Renshaw, A.E. and Haberman, S. (2006). A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Mathematics and Economics* 38: 556-570.

Rostron B. (2010). A modified new method for estimating smoking-attributable mortality in high-income countries. *Demographic Research* 23: 399–420.

Shair, S., Purcal, S. and Parr, N. (2017). Evaluating Extensions to Coherent Mortality Forecasting Models. *Risks* 5(16): 1-20.

Shang, H.L., Booth, H., and Hyndman, R. (2011). Point and interval forecasts of mortality rates and life expectancy: A comparison of ten principal component methods. *Demographic Research* 25(5): 173–214.

Shang, H.L. (2016). Mortality and life expectancy forecasting for a group of populations in developed countries: a multilevel functional data method. *The Annals of Applied Statistics* 10(3): 1639-1672.

Shang, H.L. and Hyndman, R.J. (2016). Grouped functional time series forecasting: An application to age-specific mortality rates, *Journal of Computational and Graphical Statistics* (to appear).

Stoeldraijer, L., van Duin, C. and Janssen, F. (2012). Bevolkingsprognose 2012-2060: model en veronderstellingen betreffende de sterfte. *Bevolkingstrends* 27-6-2013.

Stoeldraijer, L., van Duin, C., van Wissen, L. and Janssen, F. (2013). Impact of different mortality forecasting methods and explicit assumptions on projected future life expectancy: The case of the Netherlands. *Demographic Research* 29(13): 323–354.

Stoeldraijer, L., van Duin, C. and Huisman, C. (2017). Bevolkingsprognose 2017–2060: 18,4 miljoen inwoners in 2060. *Statistische Trends* December 2017.

Tabeau, E. (2001). A review of demographic forecasting models for mortality. In: Tabeau, E., Van Den Berg Jeths, A., and Heathcote, C. (eds.). *Forecasting mortality in developed countries: Insights from a statistical, demographic and epidemiological perspective*. Dordrecht: Kluwer Academic Publishers: 1-32.

Technical Panel on Assumptions and Methods (TPAM) (2011). *Report to the Social Security Advisory Board.* Washington, DC: Social Security Advisory Board.

Thun, M., Peto, R., Boreham, J. and Lopez, A.D. (2013) Stages of the cigarette epidemic on entering its second century. *Tobacco Control* 21: 96–101.

Tiao, G.C. and Box, G.E.P. (1981). Modeling Multiple Time Series With Applications. *Journal of the American Statistical Association* 76: 802-816.

Trias-Llimós, S., Kunst, A.E., Jasilionis, D. and Janssen, F. (2017). The contribution of alcohol to the East-West life expectancy gap in Europe from 1990 onward. *International Journal of Epidemiology*. DOI 10.1093/ije/dyx244.

United Nations (2017). *World Population Prospects: The 2017 Revision*. Department of Economic and Social Affairs, Population Division, New York.

Vallin, J. and Meslé, F. (2004). Convergences and divergences in mortality: A new approach of health transition. *Demographic Research* 2: 11-44.

Wan, C., Bertschi, L. and Yang, Y. (2013). *Coherent mortality forecasting for small populations: an application to Swiss mortality data*. Paper for the AFIR/ERM Colloqium, Lyon, France, June 2014.

Wang, H. and Preston, S.H. (2009). Forecasting United States mortality using cohort smoking histories. *Proceedings of the National Academy of Sciences* 106(2): 393–398.

WHO (2009). *Health in the European Union - trends and analysis*. Copenhagen: World Health Organization - European Observatory on Health Systems and Policies.

van der Wilk, E.A., Achterberg, P.W. and Kramers, P.G.N. (2001). *Long live the Netherlands! An analysis of trends in Dutch life expectancy in a European context*. RIVM Report 271558002.

Wilmoth, J.R. (1993). *Computational methods for fitting and extrapolating the Lee-Carter model of mortality change* (Technical Report). Department of Demography, University of California, Berkeley.

Wilmoth, J.R.(1995). Are mortality projections always more pessimistic when disaggregated by cause of death? *Mathematical Population Studies* 5(4): 293—319. http://dx.doi.org/10.1080/08898489509525409

Wong-Fupuy, C. and Haberman, S. (2004). Projecting mortality trends: Recent developments in the United Kingdom and the United States. *North American Actuarial Journal* 8(2): 56–83. doi:10.1080/10920277.2004.10596137.

Yang, S.S. and Wang, C.W. (2013). Pricing and Securitization of Multi-Country Longevity Risk with Mortality Dependence. *Insurance: Mathematics and Economics* 52: 157–169.

Zhou, R., Wang, Y., Kaufhold, K., Li, J.S-H. and Tan, K.S. (2012). *Modeling Mortality of Multiple Populations with Vector Error Correction Models: Applications to Solvency II*. Paper for the AFIR/ERM Colloqium, Lyon, France, June 2013.

Zhou, R., Li, J.S-H. and Tan, K.S. (2013). Pricing Standardized Mortality Securitizations: A Two-Population Model with Transitory Jump Effects. *Journal of Risk and Insurance* 80: 733–774.

2.

Impact of different mortality forecasting methods and explicit assumptions on projected future life expectancy: The case of the Netherlands

Abstract

BACKGROUND

With the rapid aging of the population, mortality forecasting becomes increasingly important, especially for the insurance and pension industries. However, a wide variety of projection methods are in use, both between and within countries, that produce different outcomes.

OBJECTIVE

We review the different mortality forecasting methods and their assumptions in Europe, and assess their impact on projections of future life expectancy for the Netherlands.

METHODS

For the Netherlands, we assess the projections of life expectancy at birth (e0) and at age 65 (e65) up to 2050 resulting from different methods using similar explicit assumptions regarding the historical period and the jump-off rates. We compare direct linear extrapolation, the Lee-Carter model, the Li-Lee model, a cohort model, separate projections of smoking- and non-smoking-related mortality, and the official forecast.

RESULTS

In predicting mortality, statistical offices in Europe mostly use simple linear extrapolation methods. Countries with less linear trends employ other approaches or different assumptions. The approaches used in the Netherlands include explanatory models, the separate projection of smoking- and non-smoking-related mortality, and the projection of the age profile of mortality. There are clear differences in the explicit assumptions used, including assumptions regarding the historical period. The resulting e0 in 2050 varies by approximately six years. Using the same historical period (1970–2009) and the observed jump-off rates, the findings generated by different methods result in a range of 2.1 years for women and of 1.8 years for men. For e65, the range is 1.4 and 1.9 years, respectively.

CONCLUSIONS

As the choice of the explicit assumptions proved to be more important than the choice of the forecasting method, the assumptions should be carefully considered when forecasting mortality.

Keywords: mortality forecasting, explicit assumptions, life expectancy

2.1 Introduction

With the rapid aging of the population, mortality forecasts have become more important. Recent reforms in the pension systems in Europe—which were necessary to ensure that pensions remain sustainable—have made the link between pensions and changes in life expectancy more apparent than ever. In general, monthly pension payments are based on remaining life expectancy when people retire. But whereas in some countries benefit levels are linked to life expectancy (Germany, Finland, and Portugal), in others the pension age is set to rise with increasing life expectancy (Denmark, the Netherlands), or the contribution period for pensions is set to be extended as people live longer (France) (OECD 2007). The accurate modelling and projection of mortality rates and life expectancy are therefore of growing interest to researchers.

As mortality forecasts have become increasingly important, numerous models for mortality modelling and forecasting have been developed (for reviews see Pollard 1987; Tabeau 2001; Wong-Fupuy and Haberman 2004; Booth and Tickle 2008). The various methods for mortality forecasting can be divided into three approaches: extrapolation, explanation, and expectation (Booth and Tickle 2008). Extrapolative methods make use of the regularity typically found in both age patterns and trends in time. The explanation approach makes use of structural or epidemiological models of mortality from certain causes of death for which the key exogenous variables are known and can be measured. The expectation approach is based on the subjective opinions of experts involving varying degrees of formality. It should be noted that some mortality forecasting methods include aspects of one or more approaches.

In the past, most methods were relatively simple and were largely based on subjectivity (Pollard 1987). Over time, however, more sophisticated methods that make increasing use of standard statistical methods have been developed and applied (Booth and Tickle 2008). The majority of these methods can be classified as extrapolative approaches, of which the Lee-Carter method has become dominant. This method summarises mortality by age and period for a single population as an overall time trend, an age component, and the extent of change over time by age (Lee and Carter 1992).

One of the strengths of the Lee-Carter method and of extrapolation methods in general is their robustness in situations in which age-specific log mortality rates have linear trends (Booth et al. 2006). However, some countries have less linear trends (e.g., Booth, Maindonald, and Smith 2002 for Australia; Renshaw and

Haberman 2006 for England and Wales; Janssen, Kunst, and Mackenbach 2007 for the Netherlands). It is therefore important to debate whether merely –objective linear extrapolation methods should be employed, despite the non-linearity in the trends, or whether adding information—e.g., by including a cohort effect or trends in other countries, or by using more explanatory models—is preferable, despite the subjectivity this would involve.

One example of a method which includes additional information is coherent forecasting (Li and Lee 2005). This extension of the Lee-Carter model seeks to ensure that the forecasts for related populations maintain certain structural relationships based on commonalities in their historical trends; for example, that forecasts for similar countries are not radically different. The Lee-Carter method has also recently been extended to include a cohort dimension (Renshaw and Haberman 2006), and other stochastic models have been introduced to integrate the cohort dimension in mortality forecasting (see Cairns et al. 2011). Other examples are forecasting methods using valuable medical knowledge and information on behavioural and environmental changes, such as smoking and/or obesity (e.g. Pampel 2005; Olshansky et al. 2005; Bongaarts 2006; Janssen and Kunst 2007; Stewart, Cutler, and Rosen 2009; Wang and Preston 2009; King and Soneji 2011; Janssen, van Wissen, and Kunst 2013). Although these new types of methods have many advantages, the more explanatory methods involve a large element of subjective judgment (see Section 6.5 for a further discussion). Furthermore, as all of the above-mentioned methods are fairly new, the number of times they have been applied is still relatively small.

The advent of new methods has led to a variety of types of methods being used to produce projections within a single country (e.g., Wong-Fupuy and Haberman 2004), which have produced different forecast outcomes. Most existing studies that have compared the outcomes of different methods have focused predominantly on variants within one model, such as the Lee-Carter model and its variants, extensions, or generalisations. These include Booth et al. (2002, 2005, 2006); Li and Lee (2005); Renshaw and Haberman (2006); Hyndman and Ullah (2007); Wang and Liu (2010); and Shang, Booth, and Hyndman (2011). Other studies (CMI 2005, 2006, 2007; Cairns et al. 2011) have compared the Lee-Carter model (and its cohort extension) with other extrapolative statistical models, such as P-splines models and the statistical model CBD (see Cairns, Blake and Dowd, 2006). These studies showed that using different assumptions leads to different outcomes, and that comparing different variants and extensions does not automatically result in the identification of a single best method. The historical period used is the main determinant of large differences in outcomes (see also Janssen and Kunst 2007), especially when there is considerable non-linearity in the trends.

Comparison studies of different types of methods, including both extrapolation and explanatory approaches, are not often undertaken. The comparison of outcomes from different studies is hampered by differences in the explicit assumptions; i.e., in the specific choices that must be explicitly stated in a method, such as the choice of the length of the historical period and of the jump-off rates. It would be helpful to examine, however, whether differences in projection outcomes within a country are caused by different methods, or by the use of the explicit assumptions.

The purpose of this study is to review the different mortality forecasting methods and their assumptions in Europe, and to assess their impact on projections of future life expectancy for the Netherlands.

More specifically, (i) we will review the current methods used in official mortality forecasts in Europe; (ii) compare the outcomes and the assumptions of different projection methods within the Netherlands; and (iii) compare the outcomes of different types of methods for the Netherlands using similar explicit assumptions, including the same historical period.

2.2 Data and methods

2.2.1 Methodology

After first reviewing the current methods for forecasting mortality used by statistical offices in Europe and the different national and international forecasts/projections that exist for the Netherlands, we will show to what extent different methods applied to Dutch data for the period 1970–2009 lead to different future values of life expectancy up to 2050. For the latter, we look at two outcome measures: life expectancy at birth and, in light of pension reforms, life expectancy at age 65. Moreover, we limit our own calculations to methods based on extrapolating the trends in age-specific death rates.

We found mortality forecasting methods used by statistical offices in the Netherlands and Europe in publications, including online publications, up to 2011 by using the following search words: "mortality forecasting", "forecasting", "mortality projection", "population projection", and "projection". Information on the methods is given for Austria, Belgium, Denmark, France, Italy, Ireland, Luxemburg, the Netherlands, Norway, Poland, Portugal, Spain, Sweden, and the UK. Mortality

forecasts for the Netherlands are published by Statistics Netherlands, the Actuarial Society, RIVM (National Institute for Public Health and the Environment), Eurostat (EUROPOP2010), and by four research projects: TOPALS (De Beer 2012); UPE (Alders et al. 2007); Janssen, van Wissen, and Kunst (2013); and the European Demographic Datasheet (VID/IIASA/PRB 2012).

The six different methods we applied to the Dutch data for 1970–2009, separately for both sexes, are:

- direct linear extrapolation;
- the Lee-Carter model (Lee and Carter 1992);
- an extension of the Lee-Carter model that includes the mortality experiences of other countries (Li and Lee 2005);
- an extension of the Lee-Carter model that includes a cohort dimension (Renshaw and Haberman 2006);
- a model in which smoking-related and non-smoking-related mortality is projected separately (Janssen and Kunst 2010; Janssen, van Wissen, and Kunst 2013); and
- the method used between 2004 and 2010 in the official forecast by Statistics Netherlands.

Our review showed that these methods—which represent fundamentally different approaches—were among the mortality forecasting methods used most frequently by the statistical offices in Europe, including in the Netherlands. We have chosen to avoid explicitly applying an expectation approach method because of the high degree of dependence on expert opinion in setting the target; e.g., every outcome can be set. We have also decided to ignore other methods that do not specifically extrapolate trends in age-specific death rates. See Section 2.3 for a more detailed description of the applied methods.

We chose data for 1970–2009 because of the data requirements of the method used by Statistics Netherlands. In addition to using a fixed historical period, we will use the observed values for 2009 as the jump-off rates for all the methods. Whenever possible, the further specifications and assumptions within each framework are also kept the same.

2.2.2 Data

Data on all-cause mortality and population numbers by sex, age (0, 1–4, 5–9, ..., 90–94, 95+), and year (1970–2009) were obtained from Statistics Netherlands. For the Li-Lee model, the same data were also obtained for Denmark, England and

Wales, Finland, France, Italy, Norway, Spain, Sweden, Switzerland, and West Germany from the Human Mortality Database. Lung cancer mortality data and cause-specific mortality data were obtained from Statistics Netherlands for the separate projection of smoking- and non-smoking-related mortality and for the official forecast, respectively.

2.2.3 The models in more detail

The direct linear extrapolation model is given by

$$\ln(m_{x,t}) = a_x + b_x t + \varepsilon_{x,t} \quad (1)$$

where $m_{x,t}$ denotes the central mortality rate at age x and year t, a_x the constant age pattern, b_x the set of age-specific constants that describe relative rate of change at any age and $\varepsilon_{x,t}$ the residual error.

The Lee-Carter model (Lee and Carter 1992) is given by

$$\ln(m_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t} \quad (2)$$

where k_t denotes the underlying time development. a_x is set equal to the average over time of $\ln(m_{x,t})$. b_x and k_t are found using Singular Value Decomposition under the assumptions $\sum_x b_x = 1$ and $\sum_t k_t = 0$. After estimation, we adjusted k_t to fit the observed life expectancy (Lee and Miller 2001) and extrapolated k_t using a random walk with drift.

The Li-Lee method is an extension of the Lee-Carter method, which takes into account the mortality experiences of other populations, e.g. countries (Li and Lee 2005). Short-term differences in mortality are preserved, but in the long term, the age-specific death rates within the group of countries are constrained to a constant ratio to one another. In essence, the Lee-Carter method is applied twice: first to all countries combined ($\ln(M_{x,t}) = A_x + B_x K_t + E_{x,t}$), and then to the residuals ($\ln(\hat{M}_{x,t}) - \ln(m_{x,t,i}) = b_{x,i}^{res} k_{t,i}^{res} + \varepsilon_{x,t,i}^{res}$). $M_{x,t}$ denotes the central mortality rate (at age x and year t) of all countries combined and $m_{x,t,i}$ the central mortality rate of country i. The estimates are combined into one model for the country concerned:

$$\ln(m_{x,t,i}) = a_{x,i} + B_x K_t + b_{x,i}^{res} k_{t,i}^{res} + \varepsilon_{x,t,i} \quad (3)$$

where $a_{x,i}$ equals the average over time of $\ln(m_{x,t,i})$. The time parameter of the residual ($k_{t,i}^{res}$) is extrapolated using an autoregressive model (AR(1)). Other

specifications are the same as in the Lee-Carter method described above. The mortality experiences of ten low-mortality countries surrounding the Netherlands are taken as the experiences of the other populations, i.e. Denmark, England and Wales, Finland, France, Norway, Spain, Sweden, Switzerland and West Germany.

The cohort model represents an extended version of the Lee-Carter model with an extra parameter γ_{t-x} which denotes the underlying cohort effect that is a function of the year of birth $t-x$. (Renshaw and Haberman 2006). The model is given by

$$\ln(m_{x,t}) = a_x + b_x k_t + c_x \gamma_{t-x} + \varepsilon_{x,t} \qquad (4)$$

Thus, this is in fact a model which includes the age effect, the period effect and the cohort effect. Because of the inclusion of cohorts, age groups 0 and 1–4 are combined to have age groups of equal length. Furthermore, cohorts with fewer than four observation years are not taken into account. The parameters are estimated by an iterative process (by maximum likelihood) using the constraints $\sum_x b_x = 1$, $\sum_x c_x = 1$, $\sum_t k_t = 0$ and $\sum_{t-x} \gamma_{t-x} = 0$. After estimation, k_t is adjusted to fit the observed life expectancy and extrapolated using a random walk with drift. To avoid unrealistic future mortality patterns, gamma (γ_{t-x}) is set to zero for the cohorts outside the data. We choose zero because no clear trend for the cohort parameter showed in the Dutch data, and its average over the whole period was close to zero. Because of this constraint, caution is warranted when interpreting the outcomes of this cohort model.

The model which projects non-smoking-related mortality and smoking-related mortality separately is referred to as the "smoking+non-smoking" model. First, non-smoking-related mortality, i.e. mortality after exclusion of deaths caused by smoking, is calculated using etiological fractions. The etiological fractions are the age and sex-specific proportions of total mortality determined by smoking. They are estimated by an adaptation of the indirect Peto-Lopez method (see Janssen and Kunst (2010) and Janssen, van Wissen and Kunst (2013) for more detail). Non-smoking-related mortality is projected using the Lee-Carter method. Assumed future etiologic fractions and the projected non-smoking-related mortality combined give the projected total mortality. The future etiologic fractions are taken from Janssen, van Wissen and Kunst (2013).

Statistics Netherlands publishes a mortality forecast as part of its official population forecast for the Netherlands (Van Duin et al. 2011). Overall mortality is forecasted using decomposition by cause of death (Van Duin et al. 2011; De Jong and Van Der Meulen 2005). The probability of dying from a specific cause-of-death group in a specific age interval is projected for selected sample years. The techniques used are

extrapolation combined with expert opinion, but epidemiological information (smoking) is also used. The all-cause probability of surviving is calculated by multiplying the probability of surviving for each cause of death. Intermediate years are calculated by interpolation. The Brass-logit method (Brass 1971) is used to calculate age-specific probabilities from the probabilities per age interval. Smoothed age-specific probabilities of dying of the last two observed years are used as model curve.

2.3 Forecasting methods in Europe

The approaches currently used by statistical offices in Europe to project future mortality vary considerably (see Table 2.3.1). Extrapolation methods are used most frequently. These methods are either a direct linear extrapolation of the logarithm of the age-specific mortality rates (Austria, Belgium, France, and Spain), or a variant of the Lee-Carter model (Denmark, Italy, Norway, Portugal, and Sweden). Ireland, Luxembourg, Poland, and the UK use a more subjective target approach. For Poland and Luxembourg, information on trends in other countries is included directly in the projection. Statistics Netherlands projects cause-specific mortality using extrapolation techniques combined with expert opinion (Van Duin et al. 2011; De Jong and Van Der Meulen 2005). This is the only national statistical office in our selection that includes epidemiological information directly in its projection. Portugal, France, Ireland, and the UK also include expert opinion in their mortality forecasts. Indirectly, through the knowledge of the experts, this could include trends in other countries and epidemiological information.

In addition to the differences in the methods used, there are also differences in the variants and the extensions employed. Denmark, Italy, Portugal, and Sweden use different variants of the original Lee-Carter method. Norway and Denmark extend the original method. Belgium and Spain extend the direct extrapolation method with a re-estimation after smoothing the age-specific parameter, but use a different period for the re-estimation. Belgium and France both make some adjustments for old-age mortality. Ireland and the UK make a similar assumption about the target value; namely, a constant improvement rate after some year in the future. The UK includes a cohort approach for the convergence because of the apparent cohort effects in that country.

Moreover, the historical period used differs considerably by country. Ireland, Norway, and the UK use long periods (82, 109, and 109 years, respectively),

2.3.1 Methods and assumptions behind mortality forecasting methods of statistical offices in Europe

Country	Reference	Type of method	Assumptions	Historical period	Forecasted period
Austria	Hanika (2010)	Direct extrapolation	Extension: using more recent data for the short-term trend and convergence to a plausible function of age and sex for the long term (Ediev 2008)	1970–2008	2010–2050
Belgium	Bureau fédéral du Plan (2009)	Direct extrapolation	Extension: Old-age adjustment; Extension: 10-year period for re-estimation after smoothing the age-specific parameter	1970–2007	1990–2060
Denmark	Hansen and Stephensen (2010)	Lee-Carter	Variant: adjust k(t) to fit the observed life expectancy (Lee and Miller 2001); Extension: short-term correction to account for the error between estimated jump-off rate and observation (Bell 1997); Extension: smoothing mechanism (De Jong and Tickle 2006)	1990–2009	2010–2100
France	Blanpain and Chardon (2006)	Direct extrapolation, Expert opinion	Age 3-13 no improvement after 2040; Extension: old-age adjustment	1988–2002	2007–2060
Italy	Salvini, Santini and Vignoli (2006)	Lee-Carter	Variant: an autoregressive time-series model with a deterministic time trend	Unknown	2001–2051
Ireland	Central Statistics Office (2008)	Target value, Expert opinion	Long-term rate of 1.5 percent per annum from 2031 onwards; Extension: linear interpolation between mortality declines in 2005 and 2031	1926–2005	2011–2041
Luxembourg	STATEC (2005)	Target value	Assumptions of Eurostat (convergence in 2100 for all EU countries; BMS method (Booth et al. 2002) for total group)	1962–2005	2005–2055
Netherlands	Van Duin et al. (2011)	Cause of death, Direct extrapolation, Lee-Carter, Expert opinion	Different assumptions per cause of death, depending on historical trend	1970–2009	2010–2060
Norway	Keilman and Pham (2005), Brunborg and Texmon (2010)	Lee-Carter	Extension: a quadratic age effect	1900–2008	2010–2060
Poland	Główny Urzad Statystyczny (2009)	Target value	'Catch-up' with developed countries 21-22 years later	1950–2005	2008–2035
Portugal	Instituto Nacional de Estatística (2009)	Lee-Carter, Expert opinion	Variant: Poisson log-bilinear model (Brouhns et al. 2002, Bravo 2007)	1980–2007	2008–2060
Spain	Instituto Nacional de Estadistica (2009)	Direct extrapolation	5-year moving average; Extension: 3-year period for re-estimation after smoothing the age specific parameter	1991–2007	2009–2049
Sweden	Statistiska centralbyrån (2005)	Lee-Carter	Variant: no correction for time component	1990–2002	2003–2050
United Kingdom	Office for National Statistics (2009)	Target value, Expert opinion	Annual rate of improvement converges to 1.0 percent in 2033 and remains constant thereafter; Variant: partly cohort approach for convergence	1900-2008	2008-2083

whereas Denmark, France, Spain, and Sweden use relatively short periods (20, 15, 17, and 13 years, respectively). There is also variation in the length of the forecasted period, ranging from 28 (Poland) to 91 years (Denmark), although this does not seem to correlate with the length of the historical period.

It appears that the observed past trends determine which method and historical period is used. Life expectancy at birth in western Europe has increased by six to 10 years since 1970 (WHO Health Database, Human Mortality Database). All of the countries in western Europe have experienced a rise in life expectancy, albeit at different rates and with periods of stagnation. Countries with a more linear trend (e.g., France and Switzerland) use extrapolation methods with an average historical period, while countries with more non-linear trends (e.g., Denmark, the Netherlands, and Norway) use different approaches in order to take non-linearity into account. Denmark, which has a history of having a less linear trend among women in particular, uses an extrapolation method, but with a short historical period. The Netherlands, which has non-linear trends among both men and women, uses epidemiological information in the forecast. Norway, with a period of stagnation in the 1980s among men, uses a very long period, but includes a quadratic age effect to account for the non-linearity.

2.4 Forecasts for the Netherlands

Like in other European countries, in the Netherlands a large number of national and international projections have been undertaken in recent years. The Netherlands is an exception to the broadly parallel upward trend in life expectancy at birth for men and women in western Europe (WHO Health Database, Human Mortality Database). Dutch life expectancy at birth rose from 76.5 years in 1970 to 82.7 years in 2009 for women, and from 70.8 to 78.5 years for men. The yearly increase in male life expectancy was smaller until 2000, and was larger than in other countries from 2000 onwards. Female life expectancy experienced an extended period of relative stagnation between 1980 and 2000. Overall, mortality improvements in other countries in western Europe in the period 1970–2009 were slightly larger and more linear.

Because of this non-linearity, it is not surprising that various agencies and researchers in the Netherlands have paid particular attention to mortality projection methodology. They used different methods and different assumptions, resulting in a number of different mortality projections for the Netherlands (see Table 2.4.1). International projects, which contain results for the Netherlands (EUROPOP2010, TOPALS, UPE, and European Demographic Datasheet), use a more general method of extrapolation and targeting. TOPALS (De Beer 2012) makes use of a linear spline, which produces a smoother age curve than, for example, the Lee-Carter method. The European Demographic Datasheet (VID/IIASA/PRB 2012)

takes into account that mortality dynamics are characterised by considerable inertia, which generates a more optimistic outlook for future mortality. All four take into account trends in other countries.

Projection methods developed specifically for the Netherlands (Actuarial Society, Statistics Netherlands, RIVM and Janssen, van Wissen, and Kunst (2013)), rely less on direct extrapolation. They use different approaches to account for the non-linearity observed in the trends. The method of the Actuarial Society combines a short-term trend with a long-term trend. The short-term trend (eight years of observations) determines the development in the near future, while the eventual level of the forecast is determined by the long-term trend (20 years of observations) using direct linear extrapolation. The RIVM, Janssen, van Wissen, and Kunst (2013) and Statistics Netherlands use epidemiological information in their models. The two former models also use information on trends in other countries. The differences between these two approaches lie in the different extrapolations of $k_{t,i}^{res}$ in equation (3) and in the use of modelled instead of observed jump-off rates. Statistics Netherlands uses a cause-of-death decomposition.

The different national and international mortality forecasts and projections for the Netherlands produce outcomes for life expectancy at birth in 2050 (see Table 2.4.1) that vary by 5.7 years for women and by 6.6 years for men. This large range may be caused by the different methods and the different explicit assumptions, including the different historical periods, used. The highest life expectancy at birth in 2050 is given by TOPALS 3; namely, 92.1 years for women and 89.1 years for men. The lowest values are given by UPE for both women and men; namely, 86.4 and 82.5 years, respectively. The higher projected outcomes for both the European Demographic Datasheet and TOPALS 3 result from projection methodologies that are different from the projection of trends in age-specific mortality rates. Instead, the European Demographic Datasheet uses the notion of mortality inertia; i.e., that younger cohorts are healthier than their older peers, and their future mortality rates may therefore be lower than those of the currently old cohorts (Ediev 2011). By contrast, TOPALS 3 assumes an acceleration in the decrease in future mortality. Furthermore, TOPALS uses a best-practice level of mortality in which direction the death probabilities move. The speed is determined by a partial adjustment model.

There is no clear difference in the outcomes of the methods used in the international projects and of the methods specifically developed for the Netherlands. The European Demographic Datasheet and TOPALS 3 predict life expectancies at the high end of the range for both men and women, whereas TOPALS 1 and UPE predict life expectancies at the lower end of the range. In addition, the inclusion of trends in other countries generates mixed results. The

2.4.1 Methods, assumptions and outcomes (e0 in 2050) of different national and international mortality forecasts/projections for the Netherlands

Forecast/ projection by	Reference	Type of method	Assumptions	Historical period	e0 2050 NL		
					F	M	Diff.
European Demographic Datasheet 2012	VID/IIASA/PRB (2012), Ediev (2011)	Direct extrapolation, Coherent forecasting, Cohort approach	Extrapolation of exposure-adjusted life table assuming non-divergence and constant mortality conditions. Based on the concept of mortality inertia (Ediev, 2011)	1980–2010	91.8	87.7	4.1
EUROPOP2010	Eurostat (2012), Eurostat (2007)	Lee-Carter, Target approach	Convergence mortality rates in 2100 for all EU countries; Variant: Booth et al. (2002) for all countries combined	1960–2009	88.0	84.0	4.0
TOPALS 1	De Beer (2012)	Target approach, direct extrapolation	Extrapolation of the past trends in the risk ratio (ratio between age-specific probabilities of death and a smooth, standard age schedule, i.e. projected age-specific probability of death of Japanese women) for each country separately	1976–2006	86.6 [1)]	82.9 [1)]	3.7
TOPALS 2	De Beer (2012)	Target approach, Coherent forecasting	Extrapolation of the past trends in the risk ratio of 15 countries in Europe	1976–2006	88.4 [1)]	84.7 [1)]	3.7
TOPALS 3	De Beer (2012)	Target approach, Coherent forecasting	Extrapolation of the past trends in the risk ratio of 15 countries in Europe and the half time will be half of TOPALS 5	1976–2006	92.1 [1)]	89.1 [1)]	3.0
UPE	Alders et al. (2007)	Target approach	The same rate of decline for all countries in 2030 (the eventual rate of decline was empirically estimated using eleven countries in a 30-year period). Extension: the change to the eventual rate is linear	1967/1971–1997/2001	86.4	82.5	3.9
Actuarial Society	Actuarieel Genootschap & Actuarieel Instituut (2010)	Direct extrapolation	Two-year moving average; Extension: old-age adjustment; Extension: correction females	1988–2008	87.3	85.5	1.8
Statistics Netherlands	Statistics Netherlands (2012), Van Duin et al. (2011)	Direct extrapolation, Lee-Carter, Expert opinion, Cause of death	Different assumptions per cause of death, depending on the historical trend	1970–2009	86.6	83.7	2.9
RIVM	Janssen and Kunst (2010)	Explanatory approach, Coherent forecasting	Separate projection of smoking- and non-smoking-related mortality; Including mortality experiences of 10 other European countries	1970–2006	88.1	83.8	4.3
Janssen, van Wissen and Kunst (2013)	Janssen, van Wissen and Kunst (2013)	Explanatory approach, Coherent forecasting	Separate projection of smoking- and non-smoking-related mortality; Including mortality experiences of 10 other European countries	1970–2006	87.4	83.6	3.8

[1)] Results for 2050 obtained from the author

European Demographic Datasheet and TOPALS 3 present a high life expectancy in 2050 for both men and women, but TOPALS 1 and UPE give low values. However, the inclusion of trends in other countries, either by coherent forecasting or by the choice of the target value, produces a greater difference in life expectancy between the sexes in 2050 than the other extrapolation methods.

The historical period used ranges from 20 years (Actuarial Society) to 43 years (EUROPOP2010). Statistics Netherlands includes the most recent data.

2.5 Results of different methods for Dutch mortality

If different methods are applied to the same historical period, a different range of outcomes can be expected. Thus, we apply methods, similar to the ones in Table 2.3.1 and 2.4.1, which are used in Europe and the Netherlands to Dutch mortality data for the period 1970–2009 and compare the outcomes.

To recap, the methods applied to the Dutch data range from simple extrapolation models (direct linear extrapolation and Lee-Carter) to extrapolation models which account for non-linearity in the data, either by including cohort effects or trends in other populations in the Lee-Carter model, or by using more explanatory approaches; i.e., the separate projection of smoking and non-smoking mortality and the projection by cause of death, as is done in the official Dutch forecasts. These methods are all based on the extrapolation of age-specific death rates. See Section 2.2 for more details.

Direct linear extrapolation results in a life expectancy at birth in 2050 of 86.5 years for women and 83.3 years for men (Figure 2.5.2, Table 2.5.1); i.e., an increase of 3.8 years for women and of 4.7 years for men over the next 40 years. The Lee-Carter method gives higher life expectancy values; i.e., 87.4 years for women and 83.8 years for men. The Li-Lee model generates values of 87.7 years for women and 85.0 years for men, which is the highest of the values for men. The cohort model gives a life expectancy at birth of 87.8 years for women and 83.5 years for men. The smoking+non-smoking model, in which smoking-related mortality and non-smoking-related mortality are projected separately, leads to the highest predicted values; i.e., 88.6 years for women and 84.2 years for men. Statistics Netherlands, which uses a cause-of-death decomposition, projects a life expectancy at birth of 86.6 years for women and 83.7 years for men in 2050.

The difference between the models in life expectancy at birth in 2050 is thus 2.1 years for women and 1.8 years for men. The average increase in life expectancy at birth between 2009 and 2050 is 4.8 years for women and 5.4 years for men. The direct extrapolation model results in a lower life expectancy for both men and women than the other models find. The methods which account for the non-linearity generally generate higher outcomes than the simple extrapolation models do.

The increase is almost a straight line for the extrapolation methods, while the cohort model, the smoking+non-smoking model, and the method of Statistics Netherlands are less linear (Figure 2.5.2). The straight line of the extrapolation methods is a result of linear, but slightly declining, increases in life expectancy at birth in the period 2009–2050. The yearly increases of the cohort model, the smoking+non-smoking model, and the method of Statistics Netherlands differ from year to year. They differ not only from the extrapolation methods, but also from each other (see Table 2.5.1 and compare the observation in 2009 and the outcomes in 2030 and 2050). For instance, compared to all other methods, the cohort model predicts a small increase in life expectancy at birth for men in the period 2009–2030 and a relatively large increase in the period 2030–2050. For women, the smoking+non-smoking model predicts larger increases in the first half and smaller increases in the second half of the period, which results in the same increase in the period 2009–2030 as in the period 2030–2050. The method of Statistics Netherlands predicts greater yearly increases in the first half of the period than the other methods, and constant increases in the second half for both men and women.

Three of the six methods—i.e., the Lee-Carter model, the cohort model, and the smoking+non-smoking model—forecast a larger sex difference in life expectancy at birth in 2050 than was observed in 2009.

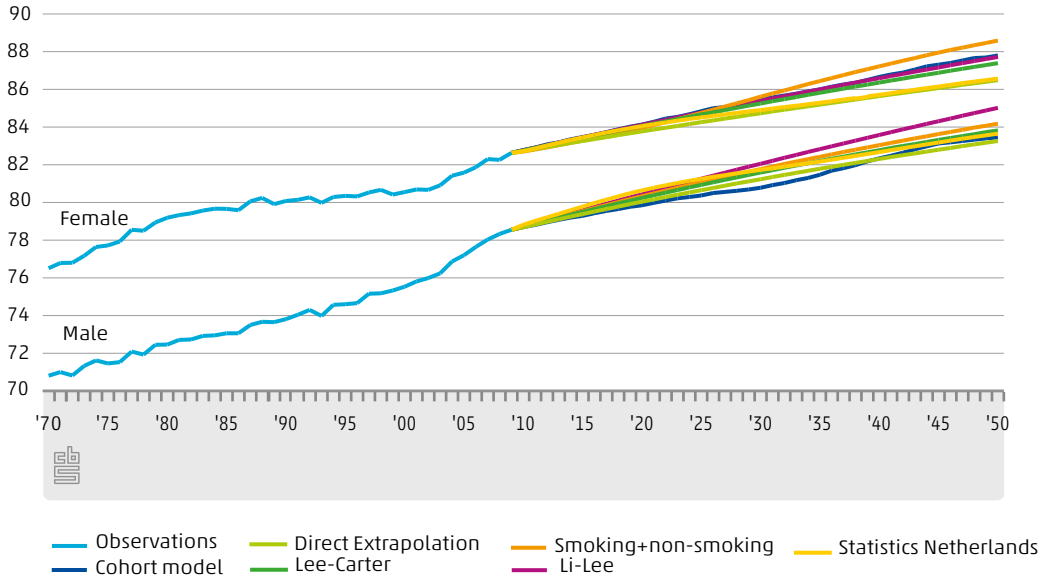
When comparing the forecasted values of remaining life expectancy at age 65 in 2050 according to the different forecasting methods, we find that the differences amount to 1.4 years for women, and 1.9 years for men (Figure 2.5.3, Table 2.5.1). Just as for life expectancy at birth, the smoking+non-smoking model results in the highest remaining life expectancy at 65 for women (25.0 years), and the Li-Lee model results in the highest value for men (22.0 years). The direct extrapolation model results in the lowest value for men (20.2 years) and the second-lowest value for women (23.7 years).

The largest difference between the outcomes at birth and at age 65 is observed for the smoking+non-smoking model. For the short-term, its effect—in terms of a smaller increase in life expectancy—is much more evident for e65 than for e0. Among men, the outcomes at age 65 drift apart more than the outcomes at birth.

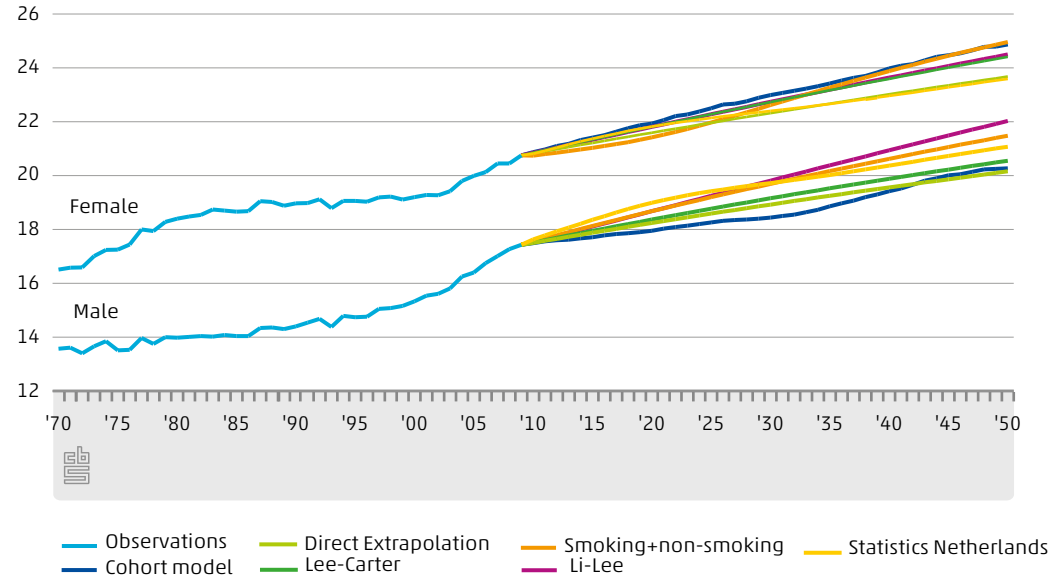
2.5.1 Life expectancy at birth and at age 65: observed (2009) and projected values (2030, 2050) for different mortality forecasting methods, the Netherlands, by sex

	Women		Men		Sex difference	
	e0	e65	e0	e65	e0	e65
Observed values 2009	82.65	20.77	78.53	17.41	4.12	3.36
Projected values 2030						
Direct Extrapolation	84.73	22.32	81.23	18.92	3.50	3.40
Lee-Carter	85.25	22.73	81.59	19.16	3.67	3.57
Li-Lee	85.39	22.74	82.05	19.81	3.34	2.92
Cohort model	85.47	22.99	80.78	18.44	4.69	4.55
"smoking+non-smoking" model	85.61	22.62	81.77	19.69	3.84	2.92
Statistics Netherlands	84.90	22.40	81.73	19.72	3.17	2.68
Projected values 2050						
Direct Extrapolation	86.49	23.68	83.26	20.16	3.23	3.52
Lee-Carter	87.39	24.41	83.82	20.55	4.57	3.86
Li-Lee	87.72	24.50	85.02	22.03	2.70	2.47
Cohort model	87.80	24.87	83.45	20.27	4.35	4.60
"smoking+non-smoking" model	88.59	24.96	84.18	21.48	5.28	4.41
Statistics Netherlands	86.57	23.59	83.65	21.07	2.92	2.52

2.5.2 Life expectancy at birth; observed (1970-2009) and projected (2010-2050) values for different mortality forecasting methods, the Netherlands, by sex



2.5.3 Life expectancy at age 65; observed (1970-2009) and projected (2010-2050) values for different mortality forecasting methods, the Netherlands, by sex



2.6 Discussion

2.6.1 Summary of the results

Mortality projection methods and assumptions clearly differ both between and within countries. In the context of new mortality projection methodologies with a focus on extrapolation, statistical offices in Europe currently predict mortality mainly using extrapolation methods (either direct or Lee-Carter), but they also make use of target values, expert opinion, and cause-specific mortality projections. The method and the historical period used seem to reflect past mortality trends in the country, with simple linear extrapolation being used by countries with gradual increases in life expectancy, and other approaches or different assumptions being applied by countries with non-linear trends.

The approaches used in national and international projections for the Netherlands also include simple linear extrapolation methods, but these are mainly methods that take into account the non-linearity in the observations by including trends in other countries, projecting smoking and non-smoking-related mortality separately,

or making projections based on causes of death. The 10 different projections for the Netherlands resulted in a wide range for life expectancy at birth in 2050 of 5.7 for women and 6.6 for men, which may be caused by both the different methods and the different explicit assumptions, including the length of the historical period, that are used.

When we compare different methods using the same historical period, including simple linear extrapolation and models that account for non-linearity, we can see that the differences in outcomes become smaller. Life expectancy at birth in 2050 then ranges from 86.5 to 88.6 for women and from 83.3 to 85.0 for men. For life expectancy at age 65 in 2050, the values range from 23.6 to 25.0 for women and from 20.2 to 22.0 for men. The models that account for non-linearity in past trends predict a less linear trend for the future as well.

2.6.2 Reflection on the differences in the outcomes

The resulting differences in the outcomes of the different methods using the same explicit assumptions—i.e., 2.1 years for women and 1.8 years for men—are small compared to the differences in the outcomes of the 10 different projections for the Netherlands, which were 5.7 years for women and 6.6 years for men. Although the methods used for the Netherlands and the methods we applied to the Dutch data are not exactly the same, it is clear from these results that using the same historical period and other explicit assumptions result in a smaller range of outcomes. Janssen and Kunst (2007) also found that using different periods may lead to larger differences between the outcomes of the models than the type of models themselves. Moreover, the resulting differences between the outcomes are small compared to the expected average increase in life expectancy at birth between 2009 and 2050, which is 4.8 years for women and 5.4 years for men.

In addition, the range of outcomes, using the same explicit assumptions, is small compared to the range of outcomes for other types of uncertainty. The differences between point forecasts calculated here only describe the uncertainty regarding the type of method. However, there are other types of uncertainty, such as parameter uncertainty (i.e., the uncertainty in the values of the parameters in a given model) and structural uncertainty (i.e., uncertainty because of the stochastic nature of a given model) (Cairns 2000). Parameter and structural uncertainty can be presented by prediction intervals, where parameter uncertainty is very important for long-run forecasts. Statistics Netherlands takes 10 years as a good indicator for the uncertainty of life expectancy at birth in 2050 (Carolina and van Duin 2010). Keilman and Pham (2004) found a 95% prediction interval of life expectancy at

birth in 2050 of 13.1 years for women and 7.7 years for men in the Netherlands. This indicates that all projections for the Netherlands presented in this paper can be rated as acceptable, given the estimated uncertainties for earlier forecasts for the Netherlands. It should be noted, however, that the uncertainty of the projections depends to a large extent on the choice of the explicit assumptions, which are not captured by prediction intervals. Furthermore, the variability in rates, which is used to estimate uncertainty, is, on average, underestimated by most methods (Shang et al. 2011).

2.6.3 Reflection on the explicit assumptions used

In our comparison of the different methods, we used the same explicit assumptions to the greatest extent possible, including the same historical period and the same modelling of . In addition, the most recently observed mortality rates are used as the jump-off rates in all of the projections. These assumptions warrant some attention, however, as they can affect the outcome of the projection, and thus the range of the outcomes.

We used the period 1970–2009 to fit the models. This is the period Statistics Netherlands used in its 2010 official forecast. As we noted above, the choice of the historical period may produce different outcomes. This is expected to influence the outcome of the simple extrapolation methods more than of the more complicated methods, because the latter take into account the possible non-linearity of the data. For instance, the smoking+non-smoking model excludes the non-linear smoking-related mortality trends from all-cause mortality trends. This allows us to obtain a more robust long-term trend that can be used as the projection basis (Janssen, van Wissen, and Kunst 2013). Furthermore, using a shorter historical period as the projection base may reduce jump-off error, but this advantage may be lost after a number of years.

We used the most recently observed mortality rates as jump-off rates to ensure that the first year of the forecast matches smoothly and closely, as well as to account for a possible jump-off error (Lee and Miller 2001). However, it could also be argued that the actual purpose of a forecast should be robustness; i.e., that when the observations are updated in the future, these small changes in the observations result in only modest changes in the forecasts (Cairns et al. 2011). The advantage of using the original Lee-Carter method (Lee and Carter 1992)—and indeed of using many time series approaches—is that it makes it possible to avoid the peculiarities of mortality rates in a particular year by taking the modelled mortality rates as jump-off rates. The two above-mentioned approaches to forecasting (matching the

first year of the forecast versus the robustness of the results) conflict with each other, and may lead to substantial differences in outcomes (Janssen, van Wissen, and Kunst 2013). In addition, the range of outcomes might be different if the modelled death rates are used as jump-off rates. Additional analysis reveals that the outcomes of the simple extrapolation methods and the outcomes for men (because of the large increase in life expectancy since 2002) are influenced the most by the choice of the jump-off rates. For instance, life expectancy at birth in 2050 for men using the direct extrapolation method and modelled jump-off rates is one year lower than with observed jump-off rates. For women, the difference is 0.5 years. The Li-Lee method results in a difference of 0.5 for men and 0.2 for women. Because the effect of using modelled instead of observed mortality rates as jump-off rates is different for the various methods, the range will also be affected.

The modelling of k_t in the Lee-Carter model and its extensions determines part of the future development. We used a random walk with drift, which assumes a linear relationship, and that each forecasted mortality rate changes at a constant exponential rate. Moreover, within the direct linear extrapolation and Lee-Carter model, the rates of decrease at different ages maintain the same ratio to one another over time, because of the assumption of a certain pattern of change in the age distribution of mortality. In practice, the relative speed of decrease at different ages may vary.

The patterns of mortality improvement show a progressive shift of high rates from lower to higher ages in some countries (Andreev and Vaupel 2005). In the past decades, most of the improvement occurred within the younger age groups, and only recently has some improvement been visible at older ages. None of the methods applied here really take into account the potential gains at older ages, as the inertia of past trends dominates the prediction. Estimated mortality improvement at higher ages for the future could therefore be underestimated (see, for instance, the high outcomes of the European Demographic Datasheet in Table 2.5.1, which uses a method that takes into account mortality inertia).

If more information is included in a model, more assumptions need to be made. For instance, the future share of smoking-attributable mortality within the smoking+non-smoking model is also based on assumptions. See Janssen, van Wissen, and Kunst (2013) for a discussion of this issue. In the cohort model we assumed that the cohort effect in equation (4) is zero. Hence, there can be a discontinuity of the last observed and the first fully projected cohort. Alternatives are, for instance, equal to the last observation or the average of the last few observations.

The longer the projection period, the larger the effect of the assumed cohort effect and of future smoking-related shares in the period life expectancy. In our projections we used a projection horizon equal to the historical period of 40 years. In fact, in more general terms, the effect of the (explicit) assumptions will depend on the chosen projection horizon.

2.6.4 Explanation of the observed projection outcomes

While the differences in the outcomes of the six methods applied in analysing the Dutch data are small, they are also clear. For instance, the trend in future life expectancy clearly differs. The simple linear extrapolation methods project an almost straight line, whereas the cohort model, the smoking+non-smoking model, and the method of Statistics Netherlands show a less linear future trend. In addition, the simple linear extrapolation methods generally produce lower outcomes than the methods which account for the non-linearity in the observations. Both can be linked to the non-linearity observed in the past trends. The simple linear extrapolation methods extrapolate the average increase of all-cause mortality over the whole period to the future, and thus result in a straight projection to the future. The cohort model, the smoking+non-smoking model, and the method of Statistics Netherlands include the past non-linear trend, and also extrapolate that trend into the future. In this particular case, including the effects causing non-linearity in a model results in a more optimistic increase over the observed period than extrapolating the average increase in all-cause mortality, and thus in a higher life expectancy in the future.

Because the assumptions about the source of the non-linearity vary between these models, the future non-linearity also differs.

The outcome of the cohort model, which differs from the outcomes of the smoking+non-smoking model and the method of Statistics Netherlands, might result from our assumption regarding the cohort effect for new cohorts (see before). For instance, setting the cohort effects to zero for the cohorts outside the data may also freeze some essential developments in the data. That might also explain why the cohort model predicted different patterns for men and women than the other models.

The smoking+non-smoking model incorporates smoking history, which was not the same for men and women. Among men, smoking had already been decreasing for much of the observation period. Thus, the effect of reduced smoking among men is already reflected in the parameters of the model without the smoking variable.

Among women, the turning point from an increase to a decrease in smoking has not yet been reached (Janssen, van Wissen, and Kunst 2013). Thus, the use of the smoking+non-smoking model influences future mortality improvements among women more than among men.

The trend in future life expectancy predicted by Statistics Netherlands is flattened because the cause of death with the worst future trend will dominate future total mortality. Mortality improvements are expected to be large in the near future, but are expected to be relatively small over the longer term.

Three of the six models forecast an increased sex differential in life expectancy at birth between 2009 and 2050: the Lee-Carter model, the cohort model, and the smoking+non-smoking model. Because of the separate projections of life expectancy for men and women, the past trend causes the increasing gap. The gap between male and female life expectancy in 2050 projected by the smoking+non-smoking model, in particular, is large. As men took up smoking before women, the sex difference increased. As women caught up, the gap decreased. The lag in the process for women means, in short, that the more advanced the stage of the smoking epidemic, the closer the smoking mortality rates of men and women become. Thus, in the (near) future, the gender gap may be expected to narrow because of the smaller differences in the smoking habits of men and women. However, as smoking among women and men moves towards parity, trends in non-smoking mortality become important. Because deaths other than from smoking have risen among men relative to women (Pampel 2002), the gap may increase in the future.

For life expectancy at age 65, we see roughly the same results as for life expectancy at birth: an almost straight line and generally lower outcomes for the simple linear extrapolation methods. Among women, an important difference is that, in the short run, the effect of the smoking+non-smoking model in terms of the increase in life expectancy is much more evident for e65 than for e0. This could be explained by the assumption of an increase in smoking-attributable mortality at older ages, but a decrease at younger ages. A smaller difference is that, among women, the cohort model generates higher life expectancy values at age 65 in 2050 than the other models, relative to the results for life expectancy at birth in 2050. This might again be due to our assumption regarding the cohort effect when projecting e0. Among men, we see greater differences between the smoking+non-smoking model on the one hand, and the linear extrapolation model on the other. This may be related to the ongoing decrease in smoking among older Dutch men.

2.6.5 Forecasting considerations

Mortality forecasting involves a number of decisions. In addition to the forecasting approach, these include the measure to be forecasted, the specification of an underlying data model, and the specific forecasting method. Among the most important issues that must be taken into account when assessing a forecasting method are the amount of subjectivity, robustness, and whether the outcomes will be logical (e.g., Cairns et al. 2009, 2011). An example of a logical outcome is the plausibility of the projected age-sex profiles of the death rates.

The choice of the forecasting approach (extrapolation, explanation, or expectation) may depend on several criteria. For instance, because the extrapolative approach generally requires a lengthy series of data for long-term forecasting, data availability and the projection period are important criteria. Other examples of criteria are the aim and the accuracy of the forecast.

The decision of which measure is to be extrapolated is also important in forecasting mortality. All of the extrapolation methods used in this paper and by the statistical offices in Europe are variants of extrapolation of death rates. Life expectancy may also be used as the measure to be forecast. Oeppen and Vaupel (2002) reported stable trends in the record life expectancy at birth over 160 years, and White (2002) reported a near-perfect fit of average life expectancy of 21 high-income countries to a straight line from 1955 to 1996; an extrapolation of the life expectancy itself would therefore be an alternative to the methods used in this paper. This different approach will generally be more optimistic than an extrapolation based on death rates. Other examples are methods that take into account mortality inertia (see the European Demographic Datasheet above) or an acceleration of the decrease in mortality (see TOPALS above).

Even when the focus is on the extrapolation of trends in age-specific mortality rates, different specific forecasting methods exist. Especially crucial is the choice of either simple linear extrapolation methods or methods that include additional information. Including additional (e.g., epidemiological) information or adding an extra dimension to the data (e.g., cohort) will automatically lead to the use of a more subjective method, but it may also lead to the use of a more robust method (with respect to small changes in the explicit assumptions, such as the historical period used) that generates more logical (i.e., more epidemiologically sound) outcomes.

Models which capture age, period, and cohort effects in mortality will provide a better model fit than age-period models, given that a cohort effect exists in the

mortality data. For some countries, cohort effects are clearly visible, although they are generally much smaller than period effects. However, there is no guarantee that models with a better fit will produce better forecasts. Cohort models in combination with age and period are fairly new and need to be fully tested, particularly in terms of their stability in response to changes in the age range or fitting period (Booth and Tickle 2008).

Epidemiological information can be included in the model in different ways (see, for instance, Pampel 2005; Bongaarts 2006; Janssen and Kunst 2007; Stewart, Cutler, and Rosen 2009; Wang and Preston 2009; King and Soneji 2011; Janssen, van Wissen, and Kunst 2013). Forecasters first have to identify the correlation between the determinant and mortality, and then formulate assumptions about the future. This requires them to have sufficient data on the determinants or indirect techniques to allow them to estimate mortality that can be attributed to a certain risk factor. Currently, no well-defined explanatory models are available, and the determinants are well understood (and measurable) for only a few causes of death. Comorbidities and dependencies among causes make such models even more complex. However, researchers like King and Soneji (2011) have emphasised the potential utility of including risk factors in forecasts, arguing that doing so will improve the quality, accuracy, and transparency of mortality forecasts. A classic example in which the determinant is known and can be measured is the dependence of lung cancer on smoking.

The challenge that arises when using methods that include additional information, such as a cohort effect or epidemiological information, lies in the prediction of the additional information itself. The advantage of using additional information in the forecasting method diminishes if the additional information cannot be forecast more accurately than mortality itself.

The inclusion of trends in other countries in the models is based on the observation that mortality evolution in most developed countries is similar because of similarities in socio-economic factors, life style, medical treatment, etc. Mortality levels will probably continue to develop in parallel. Again, several decisions have to be made, such as about how the information should be included: i.e., by a target approach or by coherent forecasting. Another crucial question is which countries determine the central tendency, or the basic mortality trend, that will be applied to the other countries. In addition, many different coherent forecasting methods exist, such as those of Li and Lee (2005); of Hyndman, Booth, and Yasmeen (2013); and of Li (2012). The methods for coherent forecasting are relatively new, and more research on this topic is needed.

All in all, the determination of which extrapolation method is optimal will depend on the amount of linearity in the past. If past trends in mortality have been largely linear, it is better to use the simple extrapolation methods, especially because the outcomes of different extrapolation methods using the same explicit assumptions do not differ greatly. If the past trends have been non-linear, including additional information is likely to result in a more robust forecast if the main effects of the non-linearity are successfully captured. Even though this involves more assumptions and more subjectivity, the right balance between added information and added subjectivity should be achieved.

2.7 Overall conclusion

In this paper, we have examined the various projection methods currently used in official mortality forecasts in Europe and mortality projections and forecasts in the Netherlands. The methods and the historical period used seem to reflect past mortality trends in the country. The Netherlands, along with other countries, use methods that take into account the non-linearity observed in the past trends. The different projections for a country lead to different results, which may have large implications for the insurance and pension industries.

For the Netherlands, the differences in the outcomes proved to be smaller if the same explicit assumptions were used, such as the same historical period and observed jump-off rates. The remaining sensitivity was shown to be small compared to other forms of uncertainty, and small compared to the increase in life expectancy over the long time horizon. The remaining differences in the outcomes mainly reflect differences between the methods which include the observed non-linearity, and the simple linear extrapolation methods which do not. For countries with more linear trends, smaller differences are likely to result.

Because the choice of explicit assumptions contributes more to the differences in outcome than the choice of the forecasting approach, the choice of the projection method should be based not only on different approaches, but, more importantly, on the explicit assumptions. The method which depends the least on the choice of the explicit assumptions might be the best option. Moreover, in order to improve the comparability of mortality projections from different institutions, the use of sensitivity analyses in which the range of different underlying explicit assumptions are applied would be an important step forward. Finally, it is important to realise that prediction intervals do not capture the uncertainty of the projections due to

the choice of the explicit assumptions. Caution is therefore warranted when judging the uncertainty of projections based on prediction intervals only.

References

Actuarieel Genootschap & Actuarieel Instituut (2010). *AG Prognosetafel 2010-2060* [electronic resource]. Utrecht: Actuarieel Genootschap & Actuarieel Instituut. [http://www.ag-ai.nl/download/9166-HR-binnenwerk+Prognosetafel.pdf].

Alders, M., Keilman, N., and Cruijsen, H. (2007). Assumptions for long-term stochastic population forecasts in 18 *European countries. European Journal of Population* 23(1): 33–69. doi:10.1007/s10680-006-9104-4.

Andreev, K. and Vaupel, J. (2005). *Patterns of mortality improvement over age and time in developed countries: Estimation, presentation and implications for mortality forecasting.* Paper presented at the Association of America 2005 Annual Meeting program, Philadelphia, Pennsylvania, March 31 – April 2 2005.

Bell, W.R. (1997). Comparing and assessing time series methods for forecasting age-specific fertility and mortality rates. *Journal of Official Statistics* 13(3): 279-303.

Blanpain, N. and Chardon, O. (2006). *Projections de population 2007-2060* pour la France métropolitaine: méthode et principaux résultats [electronic resource]. France: Institut national de la statistique et des études économiques. [http://www.insee.fr/fr/publications-et-services/docs_doc_travail/docf1008.pdf].

Bongaarts, J. (2006). How long will we live? *Population and Development Review* 34(4): 605-628. doi:10.1111/j.1728-4457.2006.00144.x.

Booth, H., Hyndman, R.J., Tickle, L., and De Jong, P. (2006). Lee-Carter mortality forecasting: A multi-country comparison of variants and extensions. *Demographic Research* 15(9): 289–310. doi:10.4054/DemRes.2006.15.9.

Booth, H., Maindonald, J., and Smith, L. (2002). Applying Lee-Carter under conditions of variable mortality decline. *Population Studies* 56(3): 325–336. doi:10.1080/ 00324720215935.

Booth, H. and Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. *Annals of Actuarial Science* 3(1-2): 3-43. doi:10.1017/S17484 99500000440.

Booth, H., Tickle, L., and Smith, L. (2005). Evaluation of the variants of the Lee-Carter method of forecasting mortality: A multi-country comparison. *New Zealand Population Review* 31(1): 13–34.

Brass, W. (1971). On the scale of mortality. In: Brass, W. (ed.). *Biological aspects of demography*. London: Taylor and Francis: 69-110.

Bravo, J.M. (2007). *Tábuas de mortalidade contemporâneas e prospectivas: Modelos estocásticos, aplicações actuariais e cobertura do risco de longevidade*. [PhD thesis]. Portugal: University of Évora, Economics.

Brouhns, N., Denuit, M., and Vermunt, J. (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics* 31(3): 373-393. doi:10.1016/S0167-6687(02)00185-3.

Brunborg, H. and Texmon, I. (2010). Befolkningsframskrivninger 2010-2060. *Økonomiske analyser* 4/2010: 28-39.

Bureau fédéral du Plan (2009). *Quotients de mortalité prospectifs par sexe et unisexes* [electronic resource]. Brussel: Bureau fédéral du Plan. [http://www.plan.be/admin/uploaded/201002040818570.wp200918.pdf].

Cairns, A.J.G. (2000). A discussion of parameter and model uncertainty in *insurance. Insurance: Mathematics and Economics* 27(3): 313-330. doi:10.1016/S0167-6687(00)00055-X.

Cairns, A.J.G., Blake, D., and Dowd, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance* 73(4): 687-718. doi:10.1111/j.1539-6975.2006.00195.x.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., and Khalaf-Allah, M. (2011). Mortality density forecasts: An analysis of six stochastic mortality models. *Insurance: Mathematics and Economics* 48(3): 355–367. doi:10.1016/j.insmatheco.2010.12.005.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., Ong, A., and Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from England & Wales and the United States. *North American Actuarial Journal* 13(1): 1-35. doi:10.1080/10920277.2009.10597538.

Carolina, N. and van Duin, C. (2010). Onzekerheidsmarges voor de sterfteprognose van het CBS. *Bevolkingstrends* 58(2): 32-37.

Central Statistics Office (2008). *Population and labour force projections 2011-2040* [electronic resource]. Ireland: Central Statistics Office. [http://www.cso.ie/en/media/csoie/releasespublications/documents/population/2008/poplabfor_2011-2041.pdf].

Continuous Mortality Investigation (CMI) (2005). *Projecting future mortality: Towards a proposal for a stochastic methodology*. (working paper 15).

Continuous Mortality Investigation (CMI) (2006). *Stochastic projection methodologies: Further progress and P-Spline model features, example results and implications*. (working paper 20).

Continuous Mortality Investigation (CMI) (2007). *Stochastic projection methodologies: Lee-Carter model features, example results and implications*. (working paper 25).

De Beer, J. (2012). Smoothing and projecting age-specific probabilities of deaths by TOPALS. *Demographic Research* 27(20): 543-592. doi:10.4054/DemRes. 2012.27.20.

De Jong, A. and Van Der Meulen, A. (2005). Prognose van sterfte naar doodsoorzaken: Model en veronderstellingen. *Bevolkingstrends* 53(2): 50-62.

De Jong, P. and Tickle, L. (2006). Extending Lee-Carter *mortality forecasting. Mathematical Population Studies* 13(1): 1-18. doi:10.1080/08898480 500452109.

Ediev, D.M. (2008). *Extrapolative projections of mortality: Towards a more consistent method. Part I: The central scenario.* Vienna Institute of Demography of Austrian Academy of Sciences (working paper; WP 03).

Ediev, D.M. (2011). Life expectancy in developed countries is higher than conventionally estimated. Implications from improved measurement of human longevity. *Journal of Population Ageing* 4(1-2): 5-32. doi:10.1007/s12062-011-9040-x.

Eurostat (2007). EUROPOP2007 *Convergence Scenario: Summary Note.* Paper presented at the Joint Eurostat-AWG Working Group on Population Projections, Luxembourg, 29-30 November 2007.

Eurostat (2012). *Population statistics* [electronic resource]. [http://epp.eurostat.ec.europa.eu/portal/page/portal/population/data/database].

Glówny Urzad Statystyczny (2009). *Prognoza ludnooeci na lata 2008–2035* [electronic resource]. Warsaw: Glówny Urzad Statystyczny. [http://www.stat.gov.pl/cps/rde/xbcr/gus/P_Population_projection_for_Poland_ 2008_2035.pdf].

Hanika, M.A. (2010). *Kleinräumige Bevölkerungsprognose für Österreich 2010-2030 mit Ausblick bis 2050* („ÖROK-Prognosen) [electronic resource]. Vienna: Statistics Austria. [http://www.statistik.at/web_de/static/kleinraeumige_bevoelkerungsprognose_fuer_oesterreich_2010-2030_mit_ausblic_051902.pdf].

Hansen, M.F. and Stephensen, P. (2010). *Danmarks fremtidige befolkning - Befolkningsfremskrivning 2010* [electronic resource]. Copenhagen: DREAM. [http://www.dreammodel.dk/pdf/Befolkning2010.pdf].

HMD. *Human Mortality Database* [electronic resource]. [http://www.mortality.org].

Hyndman, R.J., Booth, H., and Yasmeen, F. (2013). Coherent mortality forecasting: The product-ratio method with functional time series models. *Demography* 50(1): 261-283. doi:10.1007/s13524-012-0145-5.

Hyndman, R.J. and Ullah, M.S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics and Data Analysis* 51(10): 4942-4956. doi:10.1016/j.csda.2006.07.028.

Instituto Nacional de Estadistica (2009). *Proyección de la Población de España a Largo Plazo (2009-2049)* [electronic resource]. Madrid: Instituto Nacional de Estadistica. [http://www.ine.es/metodologia/t20/t2030251.pdf].

Instituto Nacional de Estatística (2009). *Projecções de população residente em Portugal* [electronic resource]. Lisboa: Instituto Nacional de Estatística. [http://www.ine.pt/ngt_server/attachfileu.jsp?look_parentBoui=65945988&att_display=n&att_download=y].

Janssen, F. and Kunst, A. (2007). The choice among past trends as a basis for the prediction of future trends in old-age mortality. *Population Studies* 61(3): 315-326. doi:10.1080/00324720701571632.

Janssen, F. and Kunst, A. (2010). De toekomstige levensverwachting. In: Luijben, A.H.P. and Kommer, G.J. (eds.). *Tijd en toekomst; deelrapport van de VTV 2010 Van gezond naar beter*. RIVM-rapport 270061008, Houten: Bohn Stafleu Van Loghum: 13-20.

Janssen, F., Kunst, A., and Mackenbach, J. (2007). Variations in the pace of old-age mortality decline in seven European countries, 1950–1999: The role of smoking and other factors earlier in life. *European Journal of Population* 23(2): 171-188. doi:10.1007/s10680-007-9119-5.

Janssen, F., Van Wissen, L.J.G., and Kunst, A.E. (2013). Including the smoking epidemic in internationally coherent mortality projections. *Demography* 50(4): 1341-1362. doi:10.1007/s13524-012-0185-x.

Keilman, N. and Pham, D.Q. (2004). *Empirical errors and predicted errors in fertility, mortality and migration forecasts in the European Economic Area*. Statistics Norway, Social and Demographic Research (Discussion Papers No. 386).

Keilman, N. and Pham, D.Q. (2005). Hvor lenge kommer vi til å leve? Levealder og aldersmønster for dødeligheten i Norge, 1900–2060. *Økonomiske analyser* 6/2005: 43-49.

King, G. and Soneji, S. (2011). The future of death in America. *Demographic Research* 25(1): 1-38. doi:10.4054/DemRes.2011.25.1.

Lee, R.D. and Carter, L.R. (1992). Modelling and forecasting US mortality. *Journal of the American Statistical Association* 87(419): 659-671.

Lee, R.D. and Miller, T. (2001). Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography* 38(4): 537-549. doi:10.1353/dem.2001.0036.

Li, J. (2012). A Poisson common factor model for projecting mortality and life expectancy jointly for females and males. *Population Studies* 67(1): 111-126. doi:10.1080/00324728.2012.689316.

Li, N.R. and Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography* 42(3): 575-594. doi:10.1353/dem.2005.0021.

OECD (2007). Pension reform: *The unfinished agenda*. Policy brief, September 2007. doi:10.1787/growth-2007-en.

Oeppen, J. and Vaupel, J.W. (2002). Broken limits to life expectancy. *Science* 296(5570): 1029-1031. doi:10.1126/science.1069675.

Office for National Statistics (2009). *2008-based National Population Projections* [electronic resource]. Newport: Office for National Statistics. [http://www.ons.gov.uk/ons/rel/npp/national-population-projections/2008-based-projections/national-population-projections.pdf]

Olshansky, S.J., Passaro, D., Hershow, R., Layden, J., Carnes, B., Brody, J., Hayflick, L., Butler,R., Allison, D., and Ludwig, D. (2005). A potential decline in life expectancy in the United States in the 21st century. *The New England Journal of Medicine* 352(11): 1138–1145. doi:10.1056/NEJMsr043743.

Pampel, F.C. (2002). Cigarette use and the narrowing sex differential in mortality. *Population and Development Review* 28(1): 77-104. doi:10.1111/j.1728-4457.2002.00077.x.

Pampel, F.C. (2005). Forecasting sex differences in mortality in high income nations: The contribution of smoking. *Demographic Research* 13(18): 455-484. doi:10.4054/DemRes.2005.13.18.

Pollard, J.H. (1987). Projection of age-specific mortality rates. *Population Bulletin of the United Nations* 21-22: 55-69.

Renshaw, A.E. and Haberman, S. (2006). A cohort-based extension to the Lee–Carter model for mortality reduction factors. Insurance: *Mathematics and Economics* 38(3): 556-570. doi:10.1016/j.insmatheco.2005.12.001.

Salvini, S., Santini, A., and Vignoli, D. (2006). *Le Previsiono della Popolazione. Teoria ed applicazioni* [electronic resource]. Firenze: Università degli Studi di Firenze. [http://www.ds.unifi.it/ricerca/pubblicazioni/altre/didattica/didattica2006_1.pdf].

Shang, H.L., Booth, H., and Hyndman, R.J. (2011). Point and interval forecasts of mortality rates and life expectancy: A comparison of ten principal component methods. *Demographic Research* 25(5): 173-214. doi:10.4054/DemRes. 2011.25.5.

STATEC (2005). *Potentiel de croissance économique et Démographie Projections 2005–2055* [electronic resource]. Luxembourg: Conférence de presse du 7 octobre 2005. [http://www.statistiques.public.lu/fr/actualites/population/population/2005/10/20051007/PDF-Present-ProjectDemograph-2005.pdf].

Statistics Netherlands (2012). *Statline database* [electronic resource]. Den Haag/Heerlen, Centraal Bureau voor de Statistiek. [http://statline.cbs.nl/StatWeb/default.aspx].

Statistiska centralbyrån (2005). *SCB:s modell för befolkningsprognoser. En dokumentation* [electronic resource]. Stockholm: Statistics Sweden. [http://www.scb.se/statistik/_publikationer/BE0401_2005A01_BR_BE52OP0501.pdf].

Stewart, S.T., Cutler, D.M., and Rosen, A.B. (2009). Forecasting the effects of obesity and smoking on U.S. life expectancy. *The New England Journal of Medicine* 361(23): 2252-2260. doi:10.1056/NEJMsa0900459.

Tabeau, E. (2001). A review of demographic forecasting models for mortality. In: Tabeau, E., Van Den Berg Jeths, A., and Heathcote, C. (eds.). *Forecasting mortality in developed countries: Insights from a statistical, demographic and epidemiological perspective.* Dordrecht: Kluwer Academic Publishers: 1-32.

Van Duin, C., De Jong, G., Stoeldraijer, L., and Garssen, J. (2011). Bevolkingsprognose 2010-2060: Model en veronderstellingen betreffende de sterfte. *Bevolkingstrends* 59(2): 28-41.

VID/IIASA/PRB (2012). [electronic resource] [http://www.oeaw.ac.at/vid/datasheet/download_2012.shtml].

Wang, C.-W. and Liu, Y.-L. (2010). Comparisons of mortality and forecasting – empirical evidence from Taiwan. *International Research Journal of Finance and Economics* 37: 46-55.

Wang H. and Preston, S.H. (2009). Forecasting United States mortality using cohort smoking histories. *PNAS* 106(2): 393-398. doi:10.1073/pnas.0811809106.

White, K.M. (2002). Longevity advances in high-income countries, 1955-96. *Population and Development Review* 28(1): 59-76. doi:10.1111/j.1728-4457.2002.00059.x.

WHO. Global Health Observatory Data Repository [electronic resource]. [http://apps.who.int/ghodata//].

Wong-Fupuy, C. and Haberman, S. (2004). Projecting mortality trends: Recent developments in the United Kingdom and the United States. *North American Actuarial Journal* 8(2): 56–83. doi:10.1080/10920277.2004.10596137.

3.

The future of smoking-attributable mortality: the case of England & Wales, Denmark and the Netherlands

Abstract

AIMS

We formally estimate future smoking-attributable mortality up to 2050 for the total national populations of England & Wales, Denmark and the Netherlands, providing an update and extension of the descriptive smokingepidemic model.

METHODS

We used smoking prevalence and population-level lung cancer mortality data for England & Wales, Denmark and the Netherlands, covering the period 1950–2009. To estimate the future smoking-attributable mortality fraction (SAF) we: (i) project lung cancer mortality by extrapolating age–period–cohort trends, using the observed convergence of smoking prevalence and similarities in past lung cancer mortality between men and women as input; and (ii) add other causes of death attributable to smoking by applying a simplified version of the indirect Peto–Lopez method to the projected lung cancer mortality.

FINDINGS

The SAF for men in 2009 was 19% (44 872 deaths) in England & Wales, 22% (5861 deaths) in Denmark and 25% (16 385 deaths) in the Netherlands. In our projections, these fractions decline to 6, 12 and 14%, respectively, in 2050. The SAF for women peaked at 14% (38 883 deaths) in 2008 in England & Wales, and is expected to peak in 2028 in Denmark (22%) and in 2033 in the Netherlands (23%). By 2050, a decline to 9, 17 and 19%, respectively, is foreseen. Different indirect estimation methods of the SAF in 2050 yield a range of 1–8% (England & Wales), 8–13% (Denmark) and 11–16% (the Netherlands) for men, and 7–16, 12–26 and 13–31% for women.

CONCLUSIONS

From northern European data we project that smoking-attributable mortality will remain important for the future, especially for women. Whereas substantial differences between countries remain, the age-specific evolution of smoking-attributable mortality remains similar across countries and between sexes.

Keywords: Age-period-cohort, Europe, lung cancer mortality, Peto-Lopez method, projection, smokingattributable mortality, smoking-epidemic.

3.1 Introduction

Smoking is a life-style with a considerable effect on health, mortality and trends therein over time. Within Europe, smoking is the leading risk factor of premature mortality (Lin et al. 2012). However, smoking behaviour and, consequently, smoking-attributable mortality (i.e. the number of all deaths in a population caused by smoking) differ strongly by country and cause a major gender gap in mortality (McCartney et al. 2011; Lopez et al. 1994).

The smoking-epidemic model by Lopez et al. in 1994 (Lopez et al. 1994) described that, in general, men in Anglo-Saxon countries were the first to take up smoking in the early 20th century. After a rapid rise lasting two to three decades, male smoking prevalence started to decline. Smoking-attributable mortality followed the increase and subsequent decline in smoking prevalence some 30–40 years later. For women, the increase in smoking started about 20 years later than men but, depending on the country, this period may be shorter or longer (Thun et al. 2013). The maximum levels in female smoking prevalence would be considerably lower than for men and, consequently, female smoking-attributable mortality would be lower than that for men.

In the last stage of the original smoking-epidemic model, similar (declining) levels of smoking prevalence for men and women were put forward, suggesting that smoking-attributable mortality for men and women will converge in the future (McCartney et al. 2011; Lopez et al. 1994). Smoking-attributable mortality for women, however, still increased during this last stage. Currently, some countries have already experienced the peak in smoking-attributable mortality for women, e.g. England &Wales (Thun et al. 2013). In other countries in northern and western Europe, such as Denmark and the Netherlands, this peak is also approaching, due to the past peak in smoking prevalence for women. An update of the smoking-epidemic model is therefore warranted.

A previous update of the smoking-epidemic model by Thun et al. (Thun et al. 2013) in which the experience of developing countries was added, and previous projections of smoking-attributable mortality (Pampel 2005; Wen et al. 2005), however, only included the short-term future. Whereas Thun et al. (Thun et al. 2013) qualitatively suggested a parallel future decrease in smoking-attributable mortality for men and women, and Pampel (2005) also revealed the equalization of smoking mortality rates for men and women, the long-term future evolution of the gap between the sexes in smoking-attributable mortality has not formally been studied previously. Furthermore, in the original smoking-epidemic model and its

Iupdate (Thun et al. 2013), little information is provided concerning differences by age groups.

Our objective is to update and extend the smoking-epidemic model by estimating future levels of smoking-attributable mortality up to 2050 for England & Wales, Denmark and the Netherlands, three countries that are ahead in the smoking-epidemic. We shall formally estimate the peak and subsequent decline in smoking-attributable mortality for women, and will provide information on the differences by sex and age groups for the long-term future. Our results will aid policymakers and public health professionals in setting goals for tobacco control programmes and can provide important input to all-cause mortality projections.

3.2 Estimation methodology

We studied past trends in age- and sex-specific smoking prevalence, lung cancer mortality rates and smoking-attributable mortality for England &Wales, Denmark and the Netherlands during the period 1950–2009.

Data on smoking prevalence by sex and age group were obtained from Cancer Research UK (2013) for England &Wales for 1950–2009 and The Dutch Expert Centre on Tobacco Control (STIVORO) (2013) for the Netherlands for 1958–2009. For Denmark, data on smoking prevalence among adults by sex was obtained from International Smoking Statistics WEB Edition (2013), Organization for Economic Co-operation and Development (OECD) Health Data (2013) and the World Health Organization (WHO) (2013) for 1950–69, 1970–93 and 1994–2009, respectively.

Annual lung cancer mortality deaths [International Classification of Diseases (ICD)-9: 162; ICD-10: C33–C34] by age (40–44, 45–49, . . . , 80+) and sex were obtained through the WHO Statistical Information System (2012) for England & Wales (1950–2009), Denmark (1951–2006) and the Netherlands (1950–2009). For Denmark, additional death numbers for 2007–09 were obtained through the Nordic Cancer Statistics Database NORDCAN (2013). Rates were calculated by dividing the deaths by population exposure data from the Human Mortality Database (2012).

To estimate the smoking-attributable mortality fraction (SAF), i.e. the proportion of all deaths due to smoking, an adapted and simplified version of the indirect Peto–Lopez method (Janssen et al. 2013) was used. Our method, like Peto et al. (Peto et al. 1992), uses observed lung cancer mortality—controlled for background

lung cancer mortality—as an indicator of the accumulated damage from smoking. That is, the observed national lung cancer mortality rates are compared with the rates of smokers and never-smokers of the American Cancer Society's Cancer Prevention Study II (ACS CPS-II) study to obtain the proportion of the population that is exposed to smoking (p) (Peto et al. 1992). We combined this indicator with relative risks (RR) for all-cause mortality for smokers versus nonsmokers from the ACS CPS-II study to obtain the age- and sex-specific SAF: SAF = p(RR-1)/(p(RR-1) + 1) (Mackenbach et al. 2004). The RRs were smoothed by applying a second-level polynomial and the excess risk was reduced by 30% to allow for confounding (Ezzati and Lopez 2003).

Lung cancer mortality and the SAF for all ages combined were directly age-standardized using sex- and country-specific population and death numbers, respectively, in 2009 as the standard.

To summarize the past trends more formally, age–period–cohort (APC) analysis was applied to lung cancer mortality. We chose an APC model with drift (Clayton and Schifflers 1987), defined as:

$$y_{a,p} = N_{a,p} \exp\left(\delta * p + \alpha_a + \beta_p + \gamma_{p-a}\right) + \varepsilon_{a,p} \qquad (1)$$

where $y_{a,p}$ is the number of deaths in age group a in period p which follows a Poisson distribution, $N_{a,p}$ is the number of person-years at risk in age group a in period p, and $\varepsilon_{a,p}$ is the error term. δ, α_a, β_p, γ_{p-a} are the drift, age (nonlinear) period and (non-linear) cohort effect, respectively. The model is applied to data by 5-year age groups (45–49,. . . , 80+) and 5-year calendar periods (1950–2009).We set the first and last cohort and first and last period to 0 to ensure identifiability (Clayton and Schifflers 1987). The model is fitted in R version 2.10 using the function glm.

3.3 Past trends

For men, smoking prevalence in the 1950s was very high: 60% in England & Wales, 80% in Denmark and 90% in the Netherlands (Fig. 1). During the period 1950–2009, smoking prevalence for men declined in all three countries, reaching a level of 30% in the Netherlands and approximately 20% in the other two countries. For women, smoking prevalence in the 1950s was between 30 and 40%. After reaching a maximum of approximately 45% between 1970 and 1980, smoking prevalence for

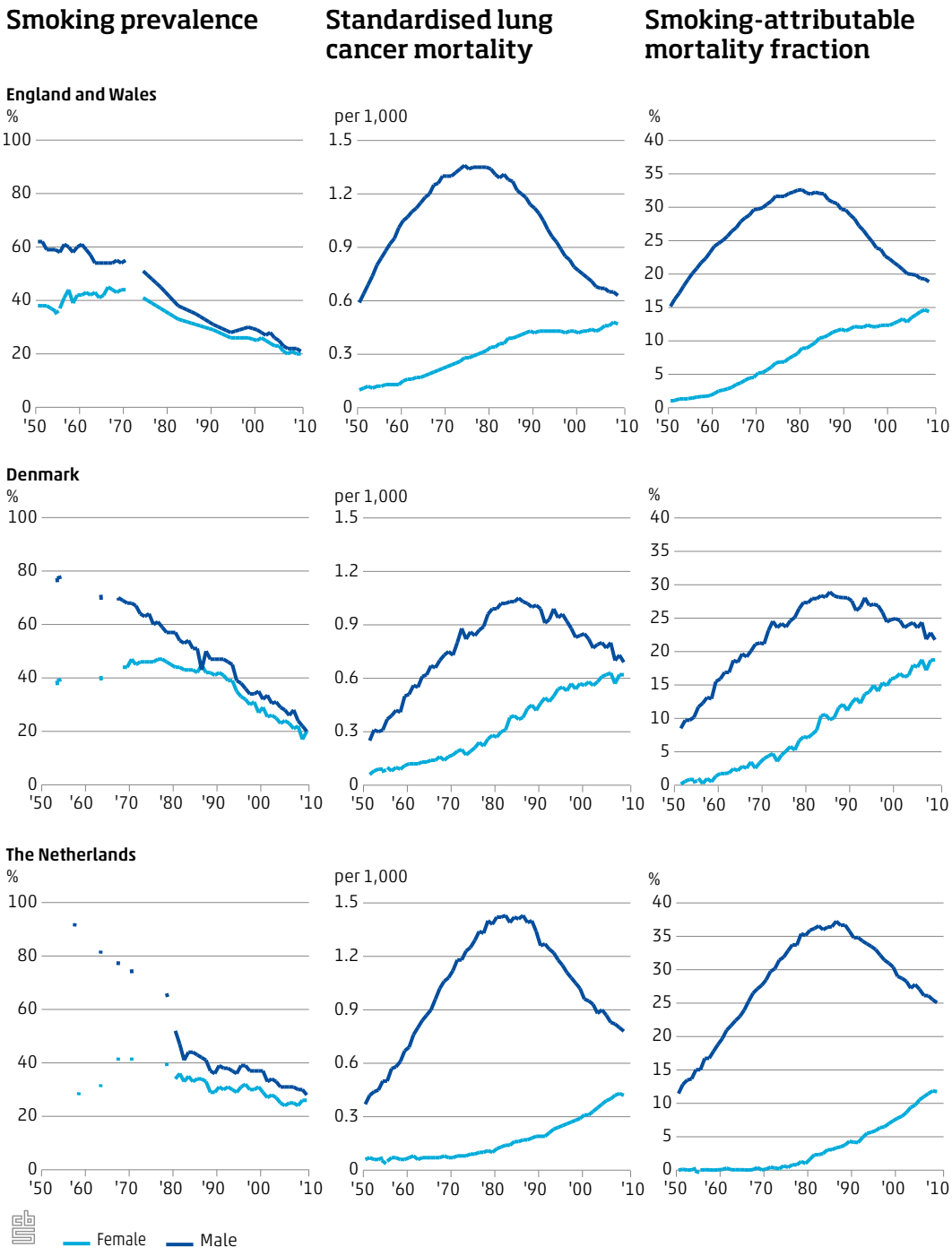
women also started to decline. From 1990 onwards the decline in smoking prevalence for women was parallel with the decline for men in all three countries. In 2009, smoking prevalence for women was almost 20% in England & Wales and Denmark and approximately 24% in the Netherlands.

For men, lung cancer mortality and the corresponding age-standardized SAF reached its maximum around 1975 in England & Wales and almost 10 years later in Denmark and the Netherlands (Fig. 3.3.1). The SAF was 33% (90,087 deaths), 29% (9,167 deaths) and 37% (25,578 deaths), respectively. Thereafter, the SAF showed a steady decline in all three countries, leading to a level of SAF in 2009 of 19% (44,872 deaths), 22% (5,861 deaths) and 25% (16,385 deaths), respectively. For women, lung cancer mortality and the SAF increased during the whole period in all three countries and converged to the level of men. The female SAF in 2009 was 14% (36,479 deaths) in England & Wales, 19% (5,249 deaths) in Denmark and 12% (8,099 deaths) in the Netherlands.

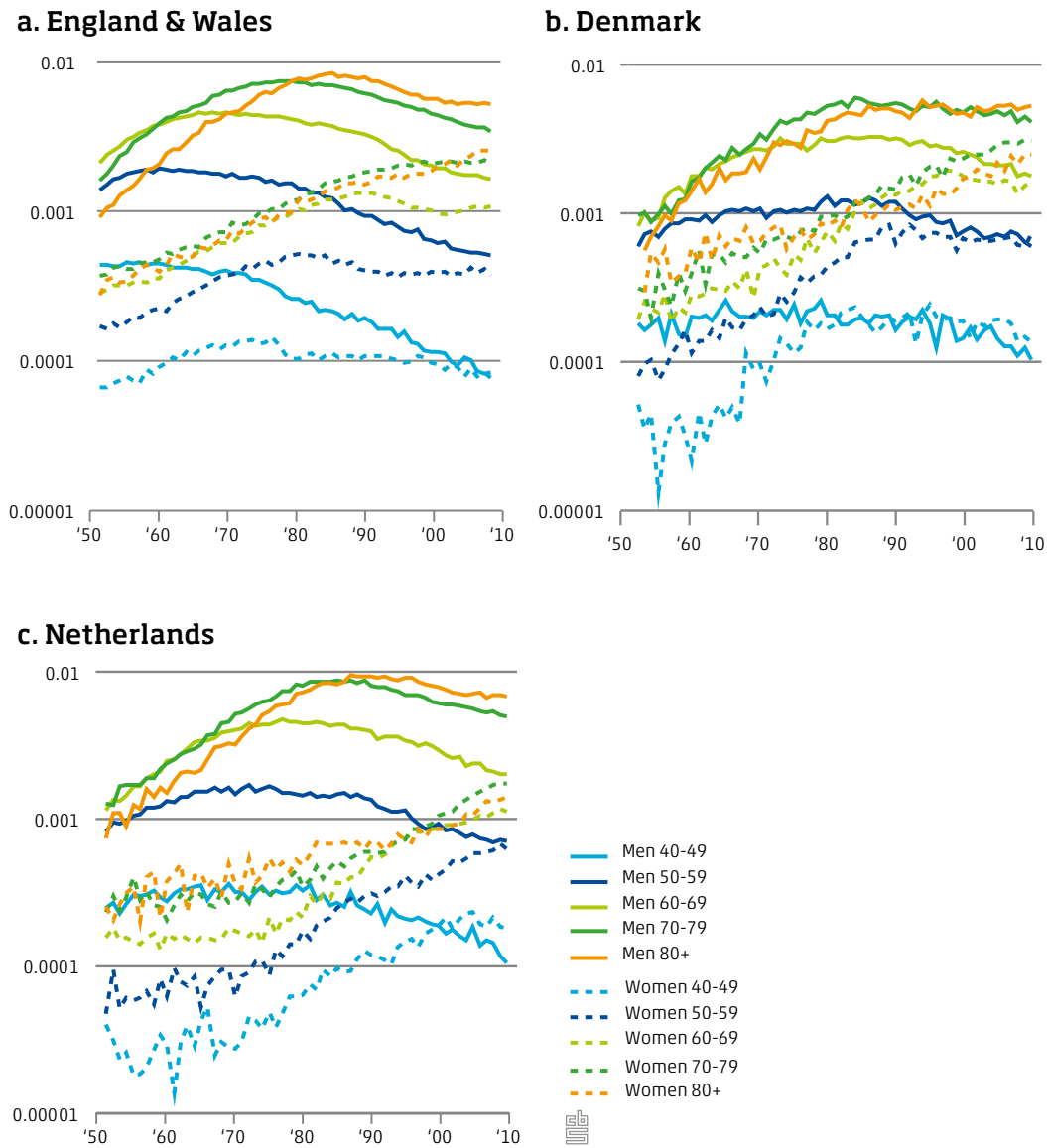
For men, the age-specific lung cancer mortality rates (Fig. 3.3.2) show a clear cohort pattern in the timing of the maximum, reflecting the uptake of smoking. The maximum is followed by a more period pattern after the peak, reflecting the quitting of smoking as a result of, for instance, tobacco control or changes in life-style when there is a decline in the lung cancer mortality rates at the same time for different age groups. The declines after the peak show parallel trends at the log scale for the different age groups, indicating that the age-specific patterns converge. For women, the cohort pattern in lung cancer mortality is less clear, but visible at the moment the lung cancer mortality starts to rise for each successive age group, and at the moment the increase for the youngest age groups ceases. For the youngest age groups we can observe that the moment the rates for women cross the rates for men, the rates start to decline at the same pace. The rates for women at higher ages show a steady increase over time. These observations also hold for the age-specific SAF's (results not shown).

Our APC analysis shows that men with the highest lung cancer mortality are born around 1900 in England & Wales, around 1925 in Denmark and around 1910 in the Netherlands (see Supporting information, Online Resource 1). The increase in lung cancer mortality among the oldest cohorts is very similar for the three countries, as well as the decline after the maximum. For women, differences in the timing of the increase in lung cancer mortality show England & Wales to be the forerunner. Women in England & Wales born around 1930 experienced the highest lung cancer mortality. For Denmark and the Netherlands no such maximum occurred.

3.3.1 Smoking prevalence (%), age-standardized lung cancer mortality rate (per 1,000) and age-standardized smoking-attributable mortality fraction (SAF) for Denmark, England & Wales and the Netherlands between 1950 and 2009, by sex



3.3.2 Lung cancer mortality rate (log-scale) for Denmark, England & Wales and the Netherlands between 1950 and 2009, by sex and age group (the 10-year age groups are weighted averages of two 5-year age groups)



3.4 Projection methodology

Based on our study of past trends (see previous section), we were able to formulate the basic assumptions behind our projection methodology:

- convergence of smoking prevalence and lung cancer mortality between men and women;
- a similar decline in age-specific lung cancer mortality rates for women as for men after the age-specific rates for women reached the age-specific rates for men;
- a cohort approach for the increase in lung cancer mortality and a period approach for its decrease.

We projected lung cancer mortality up to 2050, making qualitative use of the predictive value that current smoking prevalence has on mortality for the next 30–40 years. We then apply indirect estimation techniques to estimate the future SAF. For our main results we use the same simplified Peto–Lopez estimation technique. In addition, we performed a sensitivity analysis including four additional indirect estimation techniques (see Supporting information, Online Resource 3).

For men, the observed decline in lung cancer mortality for different age groups is projected to continue into the future. That is, we first estimated the maximum cohort exposed to smoking using an APC model applied to the lung cancer mortality data, and then projected the drift from the APC model applied to the lung cancer mortality data after this estimated maximum cohort (see Supporting information, Online Resource 1).

For women, we needed to estimate the year and level of the maximum in lung cancer mortality as well as the trend up to and after this maximum. We extrapolated the age-specific increase through an APC model with drift using the drift and non-linear cohort component. The peak years for the separate age groups were obtained by estimating the year in which the age-specific trends for women would reach the age-specific trends for men. The long-term decline after the maximum for women has been set equal to the drift from the model of men.

The limited reliability of historical smoking prevalence—due mainly to changed definitions and samples (Forey et al. 2002)—and the fact that smoking prevalence is a poor proxy of smoking intensity—mainly because it does not include dosage and age at onset (Ezzati and Lopez 2003)—are important restrictions of incorporating smoking prevalence directly in any projection methodology. Smoking prevalence is thus used merely to generate assumptions, i.e. the similarities in current smoking prevalence for men and women and its main effect on mortality 30–40 years later (Lopez et al 1994).

We project lung cancer mortality and not smoking-attributable mortality, because of the different indirect estimation techniques that exist to estimate smoking-attributable mortality, and the probable impact on the projection.

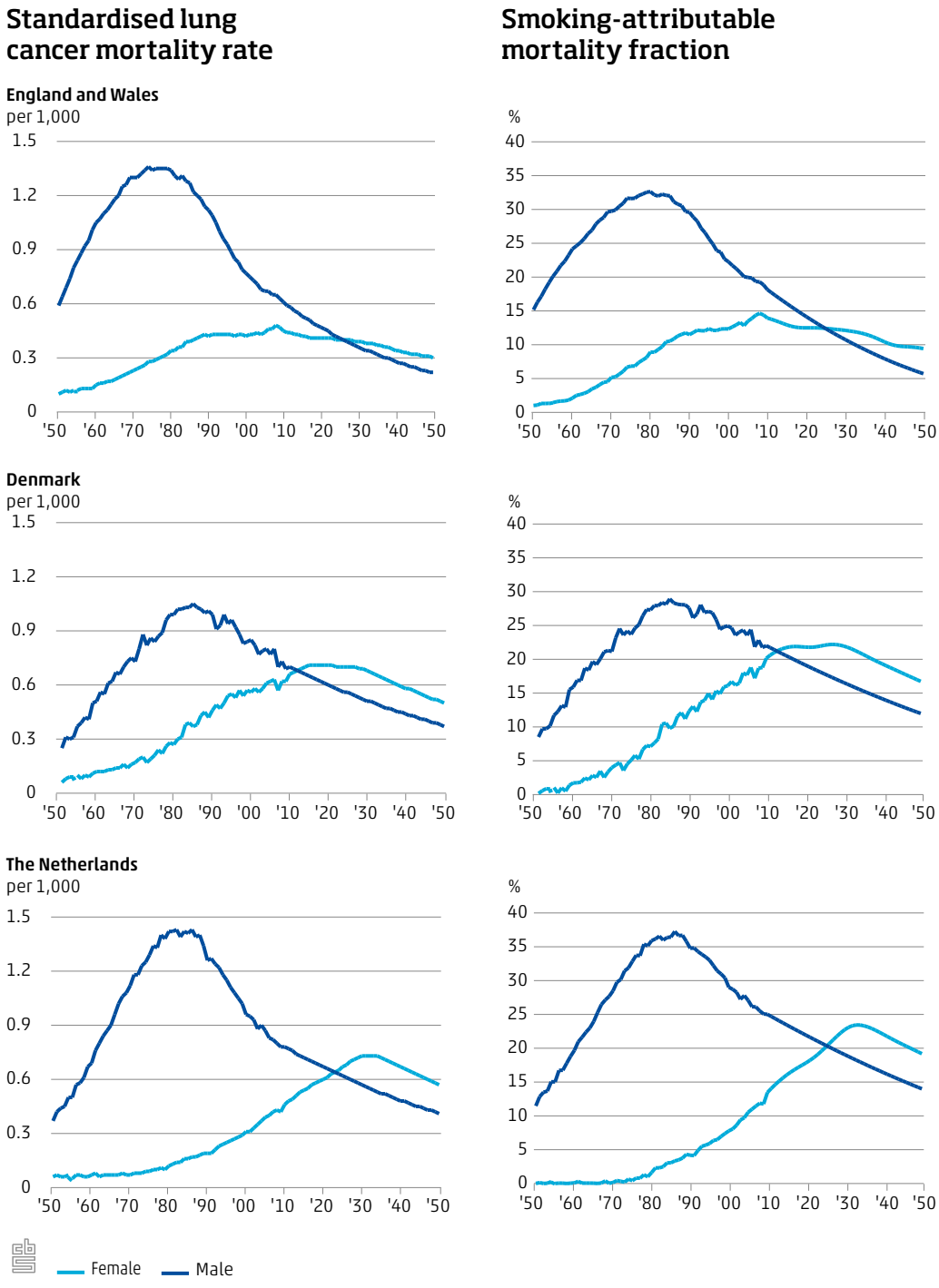
3.5 Future levels of smoking-attributable mortality

Figure 3.5.1 shows the projected age-standardized lung cancer mortality and SAF. For men in England & Wales, the SAF is estimated to decline from 19% in 2009 to 6% in 2050. The maximum SAF for women in England & Wales was already reached in 2008 (14%, 38 883 deaths), and the SAF is estimated to decline from 14% in 2009 to 9% in 2050. The SAF for men in Denmark is estimated to drop from 22% in 2009 to 12% in 2050. The level for Danish women is estimated to first increase from 19% in 2009 to 22% in 2028 and then decline to 17% in 2050. For men in the Netherlands, the SAF is estimated to decline from 25% in 2009 to 14% in 2050. For Dutch women, the SAF is estimated to increase from 12% in 2009 to 23% in 2033 and then decline to 19%.

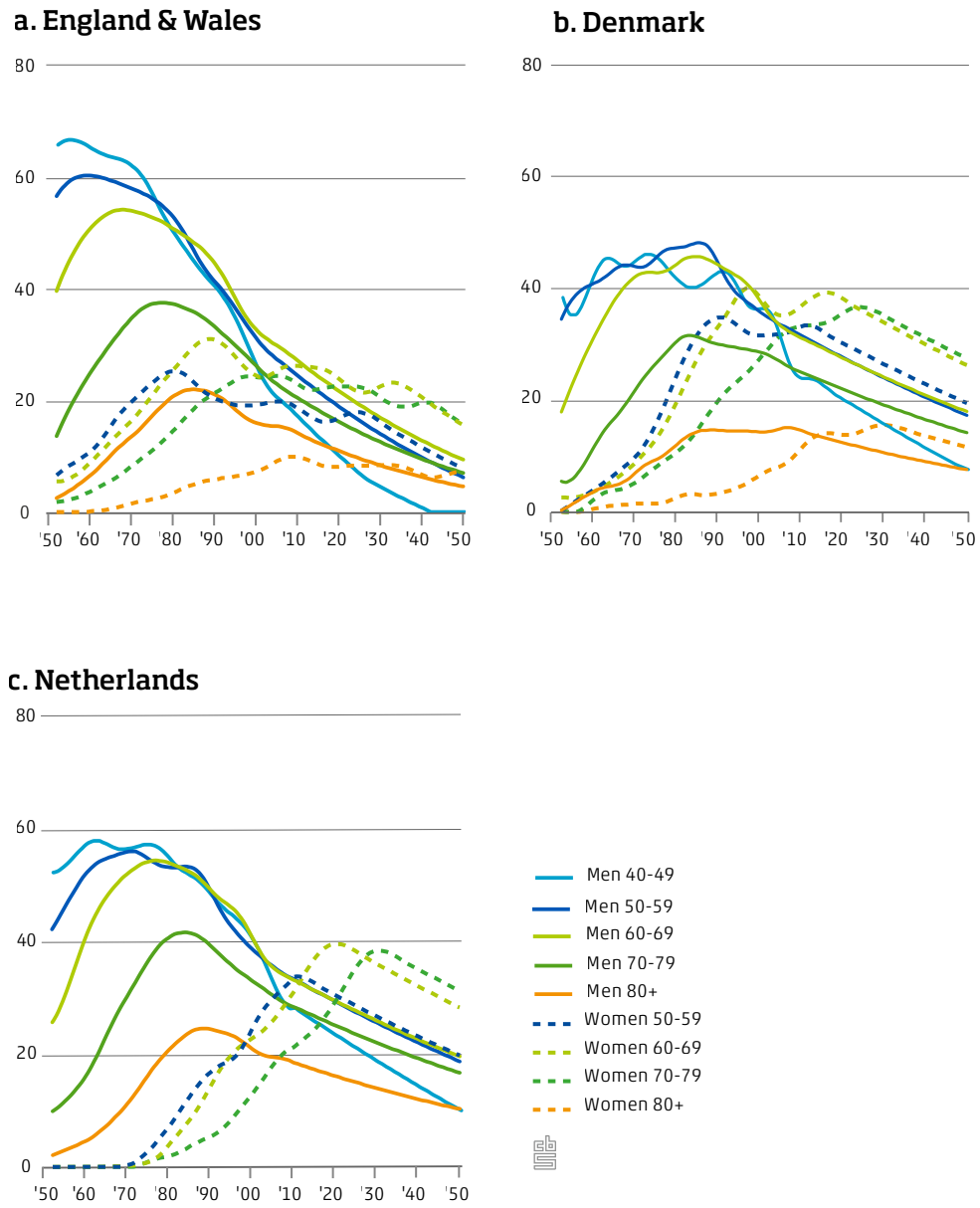
Figure 3.5.2 presents the future SAF by age (see Supporting information, Online Resource 2 for the projected lung cancer mortality by age).The results show the continuing convergence between the age groups and the (more pronounced) cohort pattern in the trend up to the maximum for women. For each country and each age group, it is expected that the SAF in 2050 for women is higher than the SAF for men.

When we apply five different indirect estimation methods (see Supporting information, Online Resource 3)—including using the National Health Interview Survey–Linked Mortality Files (NHIS-LMF) cohort study and the recent regression methods (Preston et al. 2010; Rostron 2010; Fenelon and Preston 2012)—the SAF's for men in 2050 range from 1 to 8% (England & Wales), from 8 to 13% (Denmark) and from 11 to 16% (the Netherlands). For women the ranges are 7–16%, 12–26% and 13–31%, respectively. Note that without the outliers, method 2 (NHIS-LMF cohort study) for men and the regression method 3 (Preston et al 2010) for women, the ranges were much lower.

3.5.1 Age-standardized lung cancer rate (per 1,000) and age-standardized smoking-attributable mortality fraction for Denmark, England & Wales and the Netherlands for 1950-2009 (observations) and 2010-2050 (projections), by sex



3.5.2 Smoking-attributable mortality fractions for Denmark, England & Wales and the Netherlands for 1950-2009 [age-period-cohort (APC)-estimates] and 2010-2050 (projected), by sex and age group (the 10-year age groups are weighted averages of two 5-year age groups)



3.6 Discussion

3.6.1 Summary of the results

The SAF for men in 2009 was 19% in England & Wales, 22% in Denmark and 25% in the Netherlands. In our projections, these fractions decline to 6, 12 and 14%, respectively, in 2050. The SAF for women peaked at 14% in 2008 in England & Wales, and is expected to peak in 2028 in Denmark (22%) and in 2033 in the Netherlands (23%). By 2050, a decline to 9, 17 and 19%, respectively, is foreseen.

3.6.2 Update and extension of the smoking-epidemic model

The original smoking-epidemic model assumes that, after a rapid rise, the SAF among women could be expected to peak at approximately 20–25% of all deaths, significantly lower than experienced by men (33%) and occurring approximately 20 years later. Thereafter, smoking-attributable mortality for both sexes would decline progressively (Lopez et al. 1994).

Our projected maximum levels of SAF for women in Denmark (22%) and the Netherlands (23%) correspond with the expected peak of SAF for women in the smoking-epidemic model. However, the observed maximum level for women in England &Wales (14%) is clearly lower. The difference in the timing of the maximum level between men and women, which amounts to 20 years in the smoking-epidemic model, is much greater in England & Wales (35 years), Denmark (43 years) and the Netherlands (48 years), and supports earlier findings of differential results for different countries (Thun et al. 2013).

Our observed differences between countries in the future level of smoking-attributable mortality and in sex differences in the (timing of the) smoking-epidemic are related clearly to differences in historical smoking prevalence, especially for women. These differences in smoking prevalence can be related to differences in cultural, political and economic determinants that led to differences in tobacco control and life-style (Thun et al. 2013). For instance, in England & Wales tobacco companies began the pursuit of female smokers after World War I (Action on Smoking and Health 2014). In other countries the government promoted traditional social roles for women that, among other things, discouraged tobacco use (Gomez 1999).

Our analyses also highlighted some important differences and commonalities between the different age groups. The SAF by age is characterized by a clear cohort pattern, and starts to rise, to peak and subsequently to decline first at younger ages. The SAF for younger age groups is higher than for older age groups, but after the maximum there is convergence between the age groups. These age patterns are similar across countries and sexes.

In spite of the observed convergence in smoking prevalence and lung cancer mortality rates between men and women, the SAFs in 2050 for women are higher than for men. This is because of lower relative risks and lower all-cause mortality rates for women compared to men in each age group.

3.6.3 Reflection on the projection methodology

Previous projections of smoking-attributable or smoking-related mortality consisted mainly of methods incorporating lagged smoking prevalence or different smoking scenarios (Pampel 2005; Wen et al. 2005). Probably because of the limited historical data on smoking prevalence, these projections were limited to a short projection period. Our methodology can be used for a longer projection period.

Previous projections of lung cancer mortality all used APC methodologies, although in different ways (e.g. Bashir and Estéve 2001; Kaneko et al. 2003; Shibuya et al. 2005; Olsen et al. 2008). Most of these methods do not perform well in a situation where the past trend in lung cancer mortality does not continue in the future, as we expect to happen with the trend for women. An exception is Shibuya et al. (2005), who replaced the period variable by lagged information on smoking. Their method might project changes in the trend in lung cancer mortality due to changes in smoking habits, although to obtain projections for the long term the smoking habits themselves need to be projected. Thus, previous projection methods of lung cancer mortality were relevant only for short-term projections.

Our methodology—differently from earlier studies— takes into account the expectation that future smoking-attributable mortality will first increase and then decline among women. Our assumption, and subsequent estimation, of the maximum level in lung cancer mortality for women resulted from the observed similar smoking prevalences for men and women and our assumption that this would result in similar lung cancer mortality rates 30–40 years later (as described by the smoking epidemic model by Lopez et al. (1994) and already observed for the youngest age groups). Applying our methodology to part of the data for England &Wales (1950–99), our assumption and methodology proved able to

predict the observed maximum in 2008, justifying the use of the trend and level in lung cancer mortality of men to determine the maximum for women.

3.6.4 Reflection on the indirect estimation method

The adapted and simplified Peto–Lopez method that we used to estimate SAF has the advantage of a low demand of data, is easy to use and is widely used (Pérez-Rios and Montes 2008). Furthermore, potential benefits of smoking cessation and probable effects of second-hand smoking are taken into account indirectly, because of the use of lung cancer mortality. The results of the simplified method are comparable to the results of the original Peto–Lopez method (Mackenbach et al. 2004).

A limitation of the (adapted and simplified) Peto–Lopez method is the use of the ACS CPS-II study, which may not be representative for the population under consideration due, for instance, to generally lower lung cancer mortality rates for female smokers. Furthermore, Mehta and Preston (2012) show a continuing increase over time in the relative risk of death for current and former smokers. Finally, the Peto–Lopez method assumes that the temporal relationship between accumulated exposure (including cessation) and risk will be similar between lung cancer and other smoking-determined risks (e.g. vascular disease, chronic respiratory disease).

In recent years additional indirect estimation methods have been developed, making use of regression analysis (Preston et al. 2010; Rostron 2010; Fenelon and Preston 2012). These methods rely only on observed lung cancer mortality and all-cause death rates. The two most recent methods (Rostron 2010; Fenelon and Preston 2012) showed a large similarity with the method we used, showing the validity of the three methods. Because differences at higher age groups had the largest effect on the SAF of all ages combined, its estimation should receive special attention.

3.7 Overall conclusion and implications

Our results for England &Wales, Denmark and the Netherlands illustrate clearly that smoking-attributable mortality will remain important for the future, especially for women. Substantial differences between countries are expected, both in the future level of smoking-attributable mortality and in the sex difference in the (timing of

the) smoking-epidemic. However, because of similarities in smoking prevalence, the similar age-specific evolution of smoking-attributable mortality across countries and between sexes, with convergence between the age groups, is also likely to occur for other countries currently at the fourth stage of the smoking epidemic.

Because our projection methodology requires a limited amount of data, it can be applied easily to other countries where lung cancer is dominated by smoking. The methodology would be suitable especially for countries where (i) the maximum level of lung cancer mortality for men was reached quite some time ago (e.g. Finland, Ireland, Italy, Sweden and Switzerland) and (ii) recent smoking prevalence is similar for men and women. In countries where the maximum for men was reached only recently (e.g. France, Norway, Portugal and Spain), an APC model would be more difficult to estimate and information from forerunners would also be needed. In addition, for countries at an earlier stage of the smoking epidemic, detailed information on smoking prevalence would be necessary.

Our formal quantification of future health effects of past smoking behaviour and differences therein by age and sex can aid policymakers and public health professionals in setting goals for tobacco control programmes. The effect of recent control measures, such as the WHO Framework Convention on Tobacco Control (2005), is expected to have its main effect on mortality after 2050. Moreover, it is essential to take into account the nonlinear development of the smoking-epidemic to project all-cause mortality correctly for the future.

References

Action on Smoking and Health. *Key dates in the history of anti-tobacco campaigning*. 2014. Available at: http://www.ash.org.uk/files/documents/ASH_741.pdf (accessed 25 July 2014) (Archived at http://www.webcitation.org/6TssMZ9SE on 6 November 2014).

Bashir, S. and Estéve, J. (2001). Projecting cancer incidence and mortality using Bayesian age–period–cohort models. *Journal of Epidemiology Biostatistics* 6: 287–96.

Cancer Research UK. *Lung cancer and smoking statistics*. 2013. Available at: http://www.cancerresearchuk.org/cancer-info/cancerstats/types/lung/smoking/lung-cancer-and-smoking-statistics#cancer (accessed 3 December 2013) (Archived at http://www.webcitation.org/6Tsrac4xz on 6 November 2014).

Clayton, D. and Schifflers, E. (1987). Models for temporal variation in cancer rates II: age–period–cohort models. *Statistics in Medicine* 6: 469–81.

Engholm, G., Ferlay, J., Christensen, N., Johannesen, T.B., Khan, S., Køtlum J.E. et al. NORDCAN: *Cancer Incidence, Mortality, Prevalence and Survival in the Nordic Countries*, Version 6.0 (04.12.2013). Association of the Nordic Cancer Registries. Danish Cancer Society. 2013. Available at: http://www.ancr.nu (accessed 3 December 2013) (Archived at http://www.webcitation.org/6TssBegoW on 6 November 2014).

Ezzati, M., Lopez, A. (2003). Measuring the accumulated hazards of smoking: global and regional estimates for 2000. *Tobacco Control* 12: 79–85.

Fenelon, A., Preston, S. (2012). Estimating smoking-attributable mortality in the United States. *Demography* 49: 797–818.

Forey, B., Hamling, J., Lee, P. and Wald, N. (eds.) (2002). *International Smoking Statistics: A Collection of Historical Data from 30 Economically Developed Countries*, 2nd edn. London, Oxford: Wolfson Institute of Preventive Medicine and Oxford University Press.

Forey, B., Hamling, J., Hamling, J., Thornton, A. and Lee, P. (2011). *International Smoking Statistics WEB Edition*. SupplementIndex. Available at: http://www.pnlee.co.uk/ISS3.htm#SupplementIndex (accessed 3 December 2013) (Archived at http://www.webcitation.org/6TsrqAUX1 on 6 November 2014).

Gomez, A.M. (1999). Shaping true Catholic womanhood: francoist educational discourse on women. In: Radcliff P. B., editor. *Constructing Spanish Womanhood: Female Identity in Modern Spain*. Albany, NY: State University of New York Press; pp. 51–68.

Human Mortality Database. *The Human Mortality Database*. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). 2012. Available at: http://www.mortality.org/ (accessed 15 March 2012) (Archived at http://www.webcitation.org/6TssIJVWX on 6 November 2014).

Janssen, F., Van Wissen, L., and Kunst, A. (2013). Including the smoking epidemic in internationally coherent mortality projections. *Demography* 50: 1341–62.

Kaneko, S., Ishikawa, K., Yoshimi, I., Narugame, T., Hamashima, C., Kamo, K. et al. (2003). Projection of lung cancer mortality in Japan. *Cancer Science* 94: 919–23.

Lim, S.S., Vos, T., Flaxman, A.D., Danaei, G., Shibuya, K., Adair-Rohani, H. et al. (2012). A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 380: 2224–60.

Lopez, A., Collishaw, A. and Piha, T. (1994). A descriptive model of the cigarette epidemic in developed countries. *Tobacco Control* 3: 242–7.

Mackenbach, J.P., Huisman, M., Andersen, O., Bopp, M., Borgan, J.K., Borrell, C. et al. (2004). Inequalities in lung cancer mortality by the educational level in 10 *European populations. European Journal of Cancer* 40: 126–35.

McCartney, G., Mahmood, L., Leyland, A.H., Batty, G.D., Hunt, K. (2011). Contribution of smoking-related and alcoholrelated deaths to the gender gap in mortality: evidence from 30 European countries. *Tobacco Control* 20: 166–8. doi:10.1136/tc.2010.037929.

Mehta, N. and Preston, S. (2012). Continued increases in the relative risk of death from smoking. American *Journal of Public Health* 3: 242–7.

Olsen, A., Parkin, D. and Sasieni, P. (2008). Cancer mortality in the United Kingdom: projections to the year 2025. *British Journal of Cancer* 99: 1549–54.

Organization for Economic Co-operation and Development (OECD) Health Data. *OECD Health Data 2013—frequently requested data*. 2013. Available at: http://www.oecd.org/els/health-systems/oecdhealthdata2013-frequentlyrequesteddata.htm (accessed 3 December 2013) (Archived at http://www.webcitation.org/6Tsrukvvl on 6 November 2014).

Pampel, F. (2005). Forecasting sex differences in mortality in high income nations: the contribution of smoking. *Demographic Research* 13: 455–84.

Pérez-Ríos, M. and Montes, A. (2008). Methodologies used to estimate tobacco-attributable mortality: a review. *BMC Public Health* 8: 1–11.

Peto, R., Lopez, A., Boreham, J., Thun, M. and Heath Jr, C. (1992). Mortality from tobacco in developed countries: indirect estimation from national statistics. Lancet 339: 1268–78.

Preston, S., Glei, D. and Wilmoth, J. (2010). A new method for estimating smoking-attributable mortality in high-income countries. International Journal Epidemiology 39: 430–8.

Rostron, B. (2010). A modified new method for estimating smoking-attributable mortality in high-income countries. *Demographic Research* 23: 399–420.

Shibuya, K., Inoue, M. and Lopez, A. (2005). Statistical modeling and projections of lung cancer mortality in 4 industrialized countries. *International Journal of Cancer* 117: 476–85.

The Dutch Expert Centre on Tobacco Control (STIVORO). *Trend Publication Percentage Smoking 2011: Smoking rates in the Dutch population from 1958 to 2011* [in Dutch]. 2012. Available at: http://stivoro.nl/wp-content/uploads/2012/docs/rapporten/diversen/Trendpublicatie%20Percentage%20Rokers%202011%20(2).pdf (accessed 18 December 2013) (Archived at http://www.webcitation.org/6Tsrlxq1f on 6 November 2014).

Thun, M., Peto, R., Boreham, J. and Lopez, A.D. (2013). Stages of the cigarette epidemic on entering its second century. *Tobacco Control* 21: 96–101.

Wen, S., Tsai, S., Chen, C.J., Cheng, T.Y., Tsai, M.-C. and Levy, D.T. (2005). Smoking attributable mortality for Taiwan and its projection to 2020 under different smoking scenarios. *Tobacco Control* 14: i76–80.

World Health Organization (WHO) (2005). *WHO Framework Convention on Tobacco Control. Geneva, Switzerland*: WHO Document Production Services.

World Health Organization Statistical Information System (WHOSIS). *Health statistics and health information systems. 2013*. Available at: http://www.who.int/healthinfo/statistics/mortality_rawdata/en/ (accessed 9 July 2012) (Archived at http://www.webcitation.org/6Tss5e9cj on 6 November 2014).

World Health Organization. *European health for all database (HFA-DB)*. 2013. Available at: http://data.euro.who.int/hfadb/ (accessed 3 December 2013) (Archived at http://www.webcitation.org/6Tss08hcf on 6 November 2014).

Supporting information

Additional Supporting information may be found in the online version of this article at the publisher's web-site (https://doi.org/10.1111/add.12775):

- Online Resource 1 Age–period–cohort (APC) estimation.
- Online Resource 2 Projected lung cancer mortality.
- Online Resource 3 Sensitivity of (future) smoking-attributable mortality to different indirect estimation techniques.

4.

An evaluation of methods to coherently forecast mortality based on both quantitative and qualitative criteria

Abstract

BACKGROUND

Methods to forecast mortality coherently are valuable as they can better identify the most likely long-term mortality trend and produce non-divergent outcomes. An evaluation of both quantitative and qualitative aspects of the different coherent forecasting methods is lacking, however.

OBJECTIVE

We evaluate different coherent forecasting methods in terms of accuracy (fit to historical data), robustness (stability across different fitting periods), subjectivity (sensitivity to the choice of the group of countries) and plausible outcomes (smooth continuation of trends from the fitting period).

METHODS

Mortality data from the Human Mortality Database (1970-2011) are used to produce both individual Lee-Carter (LC) and coherent mortality forecasts for France, Italy, the Netherlands, Norway, Spain, Sweden and Switzerland up to 2050. We compare a co-integrated Lee-Carter (CLC) method, the Li-Lee (LL) method, and the Coherent Functional Data (CFD) method.

RESULTS

The CFD method performed best on the accuracy measures. Both the CLC and LL method were robust. The CLC method (for women) and the LL method (for men) were least sensitive to the choice of the group of countries. The LL method generated the most plausible results, with convergence of future life expectancy similar to the fitting period and a smooth pattern of age-specific improvements.

CONTRIBUTION

To assess the suitability of coherent forecasting methods for particular forecasting applications it is essential to include both quantitative and qualitative evaluation criteria. This could imply the use of the LL method – which performed best on robustness, subjectivity and plausibility – over the CFD method – whose accuracy (model fit) was better.

Keywords: coherent mortality forecasting, accuracy, robustness, sensitivity, countries, fitting period

4.1 Introduction

Against a background of rapid population aging, mortality forecasting is becoming ever more important. Mortality forecasts are valuable for social security programmes and are often used to predict the sustainability of pension schemes (Currie et al., 2004). Forecasts of future mortality levels, especially among the elderly, are important for governments to be able to provide for health and other needs in their societies (Bengtsson and Christensen (Eds.), 2006).

The growing importance of mortality forecasts has resulted in the development of numerous models for mortality modelling and forecasting (for reviews see Pollard, 1987; Tabeau, 2001; Wong-Fupuy and Haberman, 2004; and Booth and Tickle, 2008). The majority of these methods can be classified as extrapolative, i.e. they make use of the regularity typically found in both age patterns and trends over time, with the Lee-Carter method (Lee and Carter, 1992) currently the most widely used one (Booth and Tickle, 2008). The Lee-Carter method summarises mortality by age and period for one single population into a time-varying index, an age component, and the extent of change over time by age (Lee and Carter, 1992). It forecasts probability distributions of age-specific death rates using standard time series procedures.

One of the strengths of the Lee-Carter method, and extrapolation methods in general, is its robustness in situations where age-specific log mortality rates have linear trends (Booth et al., 2006). However, herein also lies a drawback of the Lee-Carter method: there are examples of countries which have less linear trends, such as the Netherlands, Denmark and Norway. If the trend is not linear, the forecasted mortality could be very different, depending on the fitting period (Stoeldraijer et al., 2013).

Another important issue with the Lee-Carter method is that mortality forecasts using extrapolation methods based on information of each country separately might result in divergence, contrary to historic trends. In western Europe, convergence has been observed in mortality levels (White, 2002; Wilson, 2001) and in old-age mortality (Janssen, Mackenbach and Kunst, 2004). Continued convergence between countries is likely because of common socio-economic policies, similar progress in medical technology, and shared importance of certain lifestyle factors over time (Janssen, van Wissen and Kunst, 2013). Furthermore, it is likely that mortality levels of countries with similar mortality evolutions will continue to evolve similarly.

To avoid divergence, coherent forecasting methods are introduced, where "coherent" refers to non-divergent forecasts for sub-populations within a larger population (Li and Lee, 2005). The idea behind coherent forecasting is that mortality forecasts for populations with similar mortality developments will not diverge radically, but also that structural differences will remain (for instance, consistently higher mortality for men than for women (Hyndman, Booth and Yasmeen, 2013)).

The coherent forecasting methods are an important asset to obtain coherent forecasts either between the sexes or between countries (Li and Lee, 2005). Till now, coherent forecasting methods are applied more often to take into account male-female differentials than to obtain coherent forecasts between countries (Stoeldraijer et al., 2013). For instance, insurers and annuity providers need to model both sexes properly in a joint fashion because of EU rules on gender-neutral pricing in the insurance industry (European Commission, 2012). Li et al. (2016) found significant financial implications in allowing for the comovement of mortality of females and males properly. Obtaining coherent forecasts between countries is important as well, especially when past trends have been non-linear, as different fitting periods could lead to different forecasted mortality in individual forecasting. In coherent forecasting, the more linear trends for a group or an average of countries is likely to provide better information about the future direction of mortality trends in other countries with less linear trends. Experiences in other countries can thus be used to create a broader empirical basis for the identification of the most likely long-term trend, as has been suggested previously (Janssen and Kunst, 2007).

In coherent forecasting methods, non-divergence is derived by applying constraints to the parameters of individual forecasts of multiple populations. Most existing coherent forecasting methods are based on the Lee-Carter structure (Carter and Lee, 1992; Li and Lee, 2005; Li and Hardy, 2011; Zhou et al., 2012; Zhou, Li and Tan, 2013; Yang and Wang, 2013; Wan, Bertschi and Yang, 2013; Kleinow, 2015), but there are also examples in the age-period-cohort structure (Dowd et al., 2011; Cairns et al., 2011a; Jarner and Kryger, 2011; Börger and Aleksic, 2014) and the functional data paradigm (Hyndman, Booth and Yasmeen, 2013; Shang and Hyndman, 2016). Other structures are usually more complex. Even within one structure, these coherent forecasting methods are very different from each other. So far, few methods have been compared in terms of forecast accuracy (Shang, 2016; Enchev, Kleinow and Cairns, 2016; Shair, Purcal and Parr, 2017). Because a good fit to historical data does not guarantee sensible forecasts (Cairns et al., 2011b), a comparison on more qualitative aspects is important as well.

The purpose of this study is to evaluate different coherent forecasting methods in terms of accuracy (i.e. how well the model fits to historical data), robustness (i.e. stability across different fitting periods), subjectivity (i.e. sensitivity to the choice of the group of countries) and plausible outcomes (i.e. smooth continuation of trends from the fitting period). We compare the outcomes of the individual Lee-Carter method and three well-known (often cited) coherent forecasting methods that are all extensions of the individual Lee-Carter method: (i) the co-integrated Lee-Carter method (Li and Hardy, 2011; Cairns et al., 2011a); (ii) the Li-Lee method (Li and Lee, 2005); and (iii) the Coherent functional data method (Hyndman, Booth and Yasmeen, 2013).

4.2 Data and methodology

4.2.1 Data

Unsmoothed data on all-cause mortality numbers and exposures by sex, age (0, 1-4, 5-9, ..., 90-94, 95+), and year (1970-2011) were obtained from the Human Mortality Database (www.mortality.org, accessed February 9, 2016). The results are presented for France, Italy, the Netherlands, Norway, Spain, Sweden and Switzerland; seven low-mortality countries in Western Europe. Age and sex-specific death rates were calculated by dividing the mortality numbers by the exposures.

4.2.2 Analysis

Based on data for the period 1970-2011, we produced out-of-sample mortality forecasts to 2050 for the individual Lee-Carter (LC) method and three coherent forecasting methods (see section 4.3). We compared the models from two perspectives: quantitative, i.e. how well the models fit to historical data (accuracy), and qualitative, i.e. whether or not the forecasts are credible given historical data (robustness, subjectivity and plausibility).

To assess the accuracy of the method we examined the explanation ratio (ER), the Root Mean Squared Error (RMSE) and the Mean Absolute Percent Error (MAPE) of the log death rates averaged over ages and years. The explanation ratio can be interpreted as the proportion of variance in historic mortality rates explained by the method (Li and Lee, 2005). The higher the ER, the lower the RMSE and the

lower the MAPE, the better the fit to the data. Because a method with more parameters normally gives a higher ER, a lower RMSE and a lower MAPE, we also performed a Diebold Mariano test (Diebold and Mariano, 1995) to test if a method is more accurate than another method. For this, the errors of fitted values are used, but also errors outside the fitting period (using an in-sample forecast based on 1970-2001 for the period 2002-2011).

To evaluate the robustness of the coherent forecasting methods, we assessed the stability of the out-of-sample forecast outcomes across different fitting periods. For this purpose, we not only used 1970-2011 as the fitting period, but also 1970-2001 and 1970-2006. For each method we calculated the standard deviation of the life expectancy at birth (e0) in 2050 resulting from the use of the three fitting periods, averaged over the seven countries and the three selected main country groups (see below). In formula,

$$\frac{1}{c}\sum_{c}\sqrt{\frac{1}{p}\sum_{p}\left(\left(\frac{1}{g}\sum_{g}e_0^{p,g,c}\right)-\mu\right.}$$

where c denotes the seven countries, p the three fitting periods and g the three main country groups. μ denotes the (unweigthed) average of the life expectancy at birth. A lower average standard deviation implies that the method is less sensitive to the fitting period, and consequently more robust.

To compare the subjectivity of the different coherent forecasting methods, we assessed the sensitivity of the method to the choice of the included group of countries for their mortality experience (hereafter referred to as main country groups). For this purpose, we produced forecasts with three different main country groups who differ especially in their e0 values in 2011, in the amount of increase in e0 over the period 1970-2011, and as well in the linearity of the past trend (table 4.4.1.1):

- Group 'All HMD': all countries in the HMD with sufficient data (France, Italy, Netherlands, Norway, Spain, Sweden, Switzerland, Australia, Austria, Belarus, Belgium, Canada, Czech Republic, Denmark, East Germany, Estonia, Finland, Iceland, Ireland, Japan, Latvia, Lithuania, Luxembourg, New Zealand, Portugal, Slovakia, Ukraine, United Kingdom, U.S.A., West Germany).
- Group 'Top 10': the ten countries with the highest life expectancy at birth in 2011 (men and women combined; France, Italy, Netherlands, Norway, Spain, Sweden, Switzerland, Australia, Canada, Japan)

– Group ‘Western Europe’: western Europe (France, Italy, Netherlands, Norway, Spain, Sweden, Switzerland, Austria, Belgium, Denmark, East Germany, Finland, Ireland, Portugal, United Kingdom, West Germany).

For each method we calculated the standard deviation of e0 in 2050 resulting from the selection of the three main country groups, averaged over the seven countries and the three fitting periods. In formula,

$$\frac{1}{c}\sum_{c}\sqrt{\frac{1}{g}\sum_{g}\left(\left(\frac{1}{p}\sum_{p}e_0^{p,g,c}\right)-\mu\right)^2}$$

where *c* denotes the seven countries, p the three fitting periods and g the three main country groups. μ denotes the (unweigthed) average of the life expectancy at birth. A lower average standard deviation implies that the method is less sensitive to the choice of the group of countries, and consequently less subjective.

‘Plausibility’ is a rather subjective concept that is difficult to define. To assess if forecasts are plausible, we judged to what extent future patterns are in line with historical patterns or in line across age groups. That is, for each method we compared the amount of convergence in the projection period relative to the fitting period. The amount of convergence is calculated using the standard deviation of e0 in 2050 resulting from the mortality forecasts for the seven countries, averaged (unweigthed) over the three main country groups and the three different fitting periods. A smaller value of the standard deviation in 2050 compared to the observation period means that the forecast is convergent while a higher value means there is divergence. Furthermore, we compared the methods based on the improvement of the mortality rates by age between the last year of the fitting period and 2050. The forecasts are plausible if the age pattern of age-specific mortality improvements is smooth.

All forecasts were made with the program R. For the Coherent Functional Data method (Hyndman, Booth and Yasmeen, 2013) we used the Demography package for R (Hyndman, 2010).

For all methods we used the observed values in the last year of the fitting period as the jump-off rates. The forecasts are made for each sex separately without any assumption of gender coherence. Also the main country group in each coherent forecast does not include the other sex. In other words, the sexes are treated separately. We treated the sexes separately because there is no unified method to incorporate county and gender coherence at the same time.

To estimate the mortality rates of the three main country groups, the death rates for the individual countries were weighted by the population numbers. This means that the largest population dominates the mortality rates of the group. We choose to do so because the population of the whole region is relevant, not the countries separately. By creating the largest population possible of comparable countries, the most likely long-term trend of each country within the group could be determined.

4.3 The mortality forecasting methods

We compared the accuracy, robustness, subjectivity and plausibility for the individual Lee-Carter method and three coherent mortality forecasting methods: (i) the co-integrated Lee-Carter method (Li and Hardy, 2011; Cairns et al., 2011a); (ii) the Li-Lee method (Li and Lee, 2005); and (iii) the Coherent functional data method (Hyndman, Booth and Yasmeen, 2013). The three coherent forecasting methods are well known (often cited) and are all extensions of the individual Lee-Carter method.

4.3.1 The Lee-Carter method (LC method)

A well-known mortality forecasting method for individual populations was developed by Lee and Carter in 1992:

$$\ln\left(m_{x,t,i}\right) = a_{x,i} + b_{x,i}k_{t,i} + \varepsilon_{x,t} \qquad (1)$$

where $m_{x,t,i}$ denotes the death rate of population i, $a_{x,i}$ equals the average over time of $\ln\left(m_{x,t,i}\right)$, $b_{x,i}$ is the set of age-specific constants that describe relative rate of change at any age, $k_{t,i}$ denotes the underlying time development and $\varepsilon_{x,t,i}$ the residual error. Singular Value Decomposition is used to estimate $b_{x,i}$ and $k_{t,i}$ under the assumptions $\sum_x b_{x,i} = 1$ and $\sum_t k_{t,i} = 0$. After estimation, $k_{t,i}$ is extrapolated using a random walk with drift.

For more detailed information about the Lee-Carter method, see Lee and Carter (1992).

4.3.2 The Co-integrated Lee-Carter method (CLC method)

A simple extension of the Lee-Carter method is to assume a relationship between the mortality rates of two populations, by modelling the underlying time development of both processes together (Li and Hardy, 2011; Cairns et al., 2011a). Essentially, we have two Lee-Carter models: one for all populations combined and one for population :

$$\ln\left(M_{x,t}\right) = A_x + B_x K_t + E_{x,t} \quad (2a)$$

$$\ln\left(m_{x,t,i}\right) = a_{x,i} + b_{x,i} k_{t,i} + \varepsilon_{x,t,i} \quad (2b)$$

$M_{x,t}$ denotes the death rate at age x and year t of all populations combined, A_x equals the average over time of $\ln\left(M_{x,t}\right)$, B_x is the set of age-specific constants that describe the relative rate of change at any age, K_t denotes the underlying time development and $E_{x,t}$ the residual error. K_t and B_x are found using Singular Value Decomposition under the assumptions $\sum_x B_x = 1$ and $\sum_t K_t = 0$.

Because we have the situation where the main group is much larger than population i, we model the parameter of the time development for all populations combined (K_t) as a random walk, similar to a one-population model, while the spread between population i and the group ($K_t - k_{t,i}$) is modelled as an AR(1) time series, i.e. $\left(K_t - k_{t,i}\right) = c_{1,i} + c_{2,i} * \left(K_{t-1} - k_{t-1,i}\right) + \varepsilon_{t,i}$ in such a way that it will tend toward a certain constant level over time. $k_{t,i}$ is then calculated using the extrapolated values for K_t and $K_t - k_{t,i}$.

The difference in the parameters B_x and $b_{x,i}$ may still lead to diverging mortality forecasts, thus the co-integrated Lee-Carter method is only partly coherent.

For more detailed information about co-integration within the Lee-Carter method, see Li and Hardy (2011) and Cairns et al. (2011a).

4.3.3 The Li-Lee method (LL method)

Li and Lee (2005) extended the (co-integrated) Lee-Carter method so that the forecasted mortality rates will not diverge. In essence, the Lee-Carter method is applied twice: first to all populations combined, and then to the residuals.

Again, the model for all populations combined is given by

$$\ln\left(M_{x,t}\right) = A_x + B_x K_t + E_{x,t} \tag{3}$$

K_t is extrapolated using a random walk with drift.

The model for the residuals is given by

$$\ln\left(m_{x,t,i}\right) - a_{x,i} - \hat{B}_x \hat{K}_t = b_{x,i}^{res} k_{t,i}^{res} + \varepsilon_{x,t,i}^{res} \tag{4}$$

where $m_{x,t,i}$ denotes the death rate of population i , $a_{x,i}$ equals the average over time of $\ln\left(m_{x,t,i}\right)$ and $\hat{B}_x \hat{K}_t$ are the estimates from the first equation. $b_{x,i}^{res}$ is the set of age-specific constants that describe relative rate of change at any age, $k_{t,i}^{res}$ denotes the underlying time development and $\varepsilon_{x,t,i}^{res}$ the residual error. Again, Singular Value Decomposition is used to estimate $b_{x,i}^{res}$ and $k_{t,i}^{res}$. $k_{t,i}^{res}$ is extrapolated using an autoregressive model (AR(1) or a higher order model if $k_{t,i}^{res}$ does not converge to a constant when AR(1) is used).

The estimates are combined into one model for the population concerned:

$$\ln\left(m_{x,t,i}\right) = a_{x,i} + B_x K_t + b_{x,i}^{res} k_{t,i}^{res} + \varepsilon_{x,t,i} \tag{5}$$

For more detailed information about the Li-Lee method, see Li and Lee (2005).

4.3.4 The Coherent Functional Data method (CFD method)

The coherent functional data (CFD) method (Hyndman, Booth and Yasmeen, 2013) can be viewed as a generalisation of the Li-Lee method, with the difference that the CFD method uses up to six principal components (B_x and $b_{x,i}^{res}$ are the first principal components of model (3) and (4)), more general extrapolation models and smoothing. It involves forecasting interpretable product and ratio functions of rates using functional time series models introduced in Hyndman and Ullah (2007).

First, the death rates $m_{x,t,i}$ for population i at age x and year t are smoothed using weighted penalised regression splines (Wood 1994) so that each curve is monotonically increasing above age 65. The weights take care of the heterogeneity in death rates across ages. Let $\tilde{m}_{x,t,i}$ be the smoothed death rates. Then the products ($product_{x,t}$) and ratios ($ratio_{x,t,i}$) of the smoothed rates for each population i are defined:

$$product_{x,t} = \left[\prod_{i=1}^{I} \widetilde{m}_{x,t,i}\right]^{1/I} \quad \text{and} \quad ratio_{x,t,i} = \widetilde{m}_{x,t,i} \,/\, product_{x,t} \qquad (6)$$

These products and ratios - which behave roughly independently of each other and, on the log scale, are approximately uncorrelated - are then modelled using functional time series models, which are estimated using the weighted principal components algorithm of Hyndman and Shang (2009):

$$\ln\left(product_{x,t}\right) = a_x^{product} + \sum_{k=1}^{K} \beta_{t,k}\phi_{x,k} + e_{x,t} \qquad (7a)$$

$$\ln\left(ratio_{x,t,i}\right) = a_{x,i}^{ratio} + \sum_{l=1}^{L} \gamma_{t,l,i}\psi_{x,l,i} + w_{x,t,i} \qquad (7b)$$

where $a_x^{product}$ and $a_{x,i}^{ratio}$ are the means of $product_{x,t}$ and $ratio_{x,t,i}$, respectively, $\phi_{x,k}$ and $\psi_{x,l,i}$ are the principal components obtained from decomposing $product_{x,t}$ and $ratio_{x,t,i}$, respectively, and $\beta_{t,k}$ and $\gamma_{t,l,i}$ are the corresponding principal component scores. $e_{x,t}$ and $w_{x,t,i}$ are the error terms.

Forecasts are obtained by forecasting each coefficient $\beta_{t,1}$, ..., $\beta_{t,K}$ and $\gamma_{t,1,i}$, ..., $\gamma_{t,L,i}$ independently. $\beta_{t,1}$, ..., $\beta_{t,K}$ are forecasted using autoregressive integrated moving average (ARIMA) models. $\gamma_{t,1,i}$, ..., $\gamma_{t,L,i}$ are forecasted using any stationary autoregressive moving average (ARMA) or autoregressive fractionally integrated moving-average (ARFIMA) process.

The implied model for each population is given by

$$\begin{aligned}\ln\left(\widetilde{m}_{x,t,i}\right) &= \ln\left(product_{x,t}\, ratio_{x,t,i}\right) \\ &= a_x^{product} + a_{x,i}^{ratio} + \sum_{k=1}^{K} \beta_{t,k}\phi_{x,k} + \sum_{l=1}^{L} \gamma_{t,l,i}\psi_{x,l,i} + e_{x,t} + w_{x,t,i}\end{aligned} \qquad (8)$$

For more detailed information about the Coherent Functional Data method, see Hyndman, Booth and Yasmeen (2013).

4.4 Results

4.4.1 Past trends

The life expectancy at birth (e0) in France, Italy, the Netherlands, Norway, Spain, Sweden and Switzerland, increased in the period 1970-2011 with strong fluctuations from year to year. For France and Italy, the trend in e0 is almost a straight line, whereas for the Netherlands and men in Norway there are periods with a strong increase and periods with a weak increase. For men, the increase was higher than for women. Group 'Top 10' has the highest e0 of the three groups and was for women less linear than group 'All HMD' and 'Western Europe'. All three groups have a strongly linear evolution of e0 over the fitting period, comparable with Italy and France.

4.4.1.1 Life expectancy at birth (e0) in 2011 and past trends since 1970, for the seven countries under study, and the three groups used in the coherent forecasts, by sex

Country	Life expectancy (e0) in 2011		Slope of e0 1970–2011		Formal test of linearity (unexplained variance, 1-R^2) of e0 1970–2011	
	Women	Men	Women	Men	Women	Men
France	85.0	78.5	0.23	0.25	0.010	0.005
Italy	84.5	79.6	0.25	0.28	0.009	0.005
The Netherlands	82.8	79.2	0.13	0.19	0.071	0.036
Norway	83.4	79.0	0.14	0.20	0.019	0.044
Spain	85.1	79.3	0.24	0.22	0.022	0.025
Sweden	83.6	79.8	0.15	0.21	0.008	0.019
Switzerland	84.7	80.3	0.19	0.24	0.014	0.016
Average (unweigthed)	84.2	79.4	0.19	0.23	0.022	0.021
Group 'All HMD'	82.7	77.1	0.18	0.21	0.010	0.007
Group 'Top 10'	85.0	79.3	0.24	0.24	0.011	0.003
Group 'Western Europe'	83.8	78.7	0.22	0.25	0.005	0.005

4.4.2 Future trends

Averaged over all seven countries and the three fitting periods, e0 in 2050 of the LC forecasts is equal to 89.3 years for women and equal to 84.3 years for men (Table 4.4.2.1). The LC forecasts show a clear divergence in e0 for women between the Netherlands, Norway and Sweden on one side and France, Italy, Spain and

Switzerland on the other side (Table 4.4.2.1). For men, a cross-over in e0 between France and the Netherlands, Norway and Sweden occurs. In the fitting period e0 for men was always higher in the Netherlands, Norway and Sweden than in France. For both women and men, the increase in e0 in the Netherlands, Norway and Sweden is less than the increase in the other four countries.

Averaged over all seven countries, three fitting periods and three groups of countries, e0 in 2050 for women is 89.6 years using the CLC method, 89.9 years using the LL method and 88.8 years using the CFD method (Table 4.4.2.1, see Table A1 for all outcomes). For men, e0 in 2050 is equal to 84.3 years (CLC method), 85.0 years (LL method), and 84.5 years (CFD method). The outcomes of the coherent forecasts are generally closer together than the outcomes of the individual forecasts. The coherent forecasts for France, Italy, Spain (women) and Switzerland are on average lower than the individual forecasts, the coherent forecasts for The Netherlands, Norway, Spain (men), Sweden are on average higher than the individual forecasts (Table 4.4.2.1).

When applying coherent forecasts, divergence or crossover between countries occurs less often than for individual forecasts.

4.4.2.1 Period life expectancy in 2050 for the seven countries under study, by forecasting method and sex(unweighted averages over the three fitting periods and the three main country groups)

	Women				Men			
	LC	CLC	LL	CFD	LC	CLC	LL	CFD
France	91.1	91.1	90.9	89.4	84.7	84.7	84.7	84.0
Italy	91.3	90.2	90.5	88.8	85.8	85.0	85.5	84.5
The Netherlands	86.4	87.5	88.2	88.1	82.4	82.6	84.1	84.2
Norway	87.2	88.9	89.2	88.5	82.8	83.3	84.6	84.5
Spain	91.0	90.5	90.9	89.3	84.4	84.9	84.9	84.6
Sweden	87.7	88.6	89.6	88.8	84.3	84.3	85.3	84.8
Switzerland	90.2	90.2	90.2	89.1	85.9	85.2	85.5	84.9
Average (unweigthed)	89.3	89.6	89.9	88.8	84.3	84.3	85.0	84.5

4.4.3 Accuracy

By calculating the ER, RMSE and MAPE, using the historical data for 1970-2011, we determined the accuracy of the method, i.e. how well the models fit to historical data. The CFD method outperforms the (C)LC method and LL method for all

countries and sexes (Table 4.4.3.1) and is thus the most accurate. The ER, RMSE and MAPE values for the LL method is higher for some countries than values of the (C)LC method and lower for other countries. On average, the LL method performs equally or better than the (C)LC method. For all methods and countries, the MAPE for men is much higher than for women, indicating that all methods fit the data for women better than for men. The ER and RMSE are more equal for men and women.

4.4.3.1 Explanation Ratio (ER), Root Mean Squared Error (RMSE) and Mean Absolute Percent Error (MAPE) in log death rates (averaged over the three fitting periods and the three main country groups)

	Women				Men			
	LC	CLC	LL	CFD	LC	CLC	LL	CFD
ER								
France	0.96	0.96	0.97	0.98	0.94	0.94	0.95	0.98
Italy	0.95	0.95	0.96	0.98	0.93	0.93	0.94	0.98
The Netherlands	0.90	0.90	0.90	0.96	0.93	0.93	0.93	0.97
Norway	0.74	0.74	0.72	0.90	0.87	0.87	0.86	0.95
Spain	0.94	0.94	0.94	0.97	0.84	0.84	0.90	0.96
Sweden	0.86	0.86	0.86	0.95	0.88	0.88	0.88	0.97
Switzerland	0.85	0.85	0.86	0.95	0.82	0.82	0.87	0.94
Average (unweigthed)	0.89	0.89	0.89	0.96	0.89	0.89	0.90	0.96
RMSE								
France	0.049	0.049	0.047	0.036	0.060	0.060	0.054	0.036
Italy	0.060	0.060	0.058	0.044	0.077	0.077	0.071	0.045
The Netherlands	0.063	0.063	0.065	0.042	0.059	0.059	0.061	0.041
Norway	0.112	0.112	0.115	0.068	0.088	0.088	0.092	0.054
Spain	0.072	0.072	0.068	0.046	0.106	0.106	0.082	0.050
Sweden	0.087	0.087	0.089	0.051	0.088	0.088	0.090	0.048
Switzerland	0.099	0.099	0.095	0.058	0.117	0.117	0.101	0.066
Average (unweigthed)	0.078	0.078	0.077	0.049	0.085	0.085	0.079	0.048
MAPE								
France	0.92	0.92	0.97	0.72	1.49	1.49	1.64	1.04
Italy	1.23	1.23	1.20	0.90	1.80	1.80	1.76	1.22
The Netherlands	1.49	1.49	1.35	0.99	1.99	1.99	1.89	1.29
Norway	1.98	1.98	1.96	1.48	2.13	2.13	2.27	1.57
Spain	1.26	1.26	1.36	0.99	2.13	2.13	2.06	1.25
Sweden	1.42	1.42	1.52	1.15	2.19	2.19	2.32	1.60
Switzerland	2.06	2.06	1.99	1.50	3.35	3.35	2.65	2.15
Average (unweigthed)	1.48	1.48	1.48	1.10	2.15	2.15	2.08	1.44

To take into account the different number of model parameters, also a Diebold-Mariano test is performed to examine the accuracy of the methods. Based on the errors of the fitted values in the fitting period 1970-2011, the CFD method is more accurate than the (C)LC and LL method in all countries and both sexes

(Table 4.4.3.2), although, especially for women, the difference is not always statistically significant. Based on the errors of the forecasted values over the period 2002-2011 using the fitting period 1970-2001, the accuracy of the CFD method is statistically higher compared to the other methods for only a few countries/sexes (6 out of 28). Only for men in Sweden the CFD method is statistically more accurate then both the CLC and LL method. However, for France and women in Sweden the CLC method is statistically more accurate than both the LL and CFD method.

4.4.3.2 Results of the Diebold-Mariano test, both when applied to the fitting period 1970-2011 and when 2002-2011 is forecasted based on 1970-2001

	Fitting period 1970–2011[1]			Forecast 2002–2011 based on 1970–2001					
	(C)LC - LL	(C)LC - CFD	LL - CFD	LC - CLC	LC - LL	LC - CFD	CLC - LL	CLC - CFD	LL - CFD
France									
Women	−1.95	1.31	2.25	1.00	−2.24[2]	-3.56[2]	−2.75[2]	-3.55[2]	−2.69
Men	−1.50	2.52[2]	2.85[2]	1.08	−1.38	-3.30[2]	−1.98[2]	-3.33[2]	−1.62
Italy									
Women	−0.36	2.41	2.37	2.28[2]	0.71	−0.16	−1.98[2]	−1.47	−0.58
Men	0.07	2.83[2]	2.63[2]	0.50	−1.01	−2.68[2]	−1.17	-2.75[2]	-3.19[2]
The Netherlands									
Women	2.28	2.41[2]	1.73	4.05[2]	1.00	3.39[2]	0.11	2.64[2]	1.90
Men	1.70	3.23[2]	2.73[2]	−0.72	2.79[2]	2.61[2]	2.77[2]	2.58[2]	0.63
Norway									
Women	0.61	2.80[2]	2.82[2]	4.81[2]	1.66	0.16	0.96	−0.05	0.19
Men	−1.49	1.58	1.98[2]	0.42	2.29	2.31[2]	2.11	2.28	0.88
Spain									
Women	−1.14	1.44	2.03	−0.04	−1.68	−1.00	−1.80	−1.05	0.65
Men	−0.98	3.79[2]	3.55[2]	−2.51[2]	2.42[2]	2.22[2]	2.43[2]	2.28[2]	−1.11
Sweden									
Women	−0.09	1.57	1.60	−1.23	−2.36[2]	−3.03[2]	−2.41[2]	−3.08[2]	−1.88
Men	−0.63	1.81[2]	2.27[2]	-5.85[2]	1.99	3.38[2]	2.15	3.44[2]	2.55[2]
Switzerland									
Women	0.69	2.58[2]	1.93	−1.70	−0.39	−1.34	0.54	−0.86	−0.59
Men	1.14	2.33[2]	1.00	0.55	0.57	2.24[2]	0.56	2.27[2]	0.44

[1] The fitted values for the LC and CLC method are equal in the fitting period 1970-2001, and the values for the DB tests are equal as well.

[2] significance at the five percent level. A negative value of the DB-test indicates the first mentioned method is more accurate, a positive values that the second mentioned method is more accurate.

4.4.4 Robustness

By calculating the standard deviation of the mean e0 in 2050, averaged over groups, from Table A.1, we determined the robustness of each method, i.e. stability across different fitting periods. The coherent forecasting methods are sensitive to the fitting period, just as the individual forecasting method. For women the dependence on the choice of the fitting period of the coherent forecasts is lower than of the individual forecast (Table 4.4.5.1), i.e. the coherent forecasting methods are more robust than the LC method. The LL method depends the least on the choice of the fitting period and is therefore the most robust. For men, the CFD method depends more on the fitting period than the individual forecast (i.e. less robust), and the dependence for the CLC and LL method are close to each other and less than for the individual forecast (i.e. more robust).

4.4.5 Subjectivity

By calculating the standard deviation of the mean e0 in 2050, averaged over the three fitting periods, from Table A.1, we determined the subjectivity of the method, i.e. the sensitivity to the choice of the group of countries. The coherent forecasting methods are sensitive to the choice of the group of countries. The higher the group dependence, the more subjective the method is. The group dependence for women is the highest for the LL method, but close to the group dependence of the CFD method, and the lowest for the CLC method; for men the CFD method results in the highest dependence and the LL method in the lowest dependence (Table 4.4.5.1).

4.4.5.1 Sensitivity of the different methods to the use of the three different fitting periods and the three different selections of the main country group, by sex

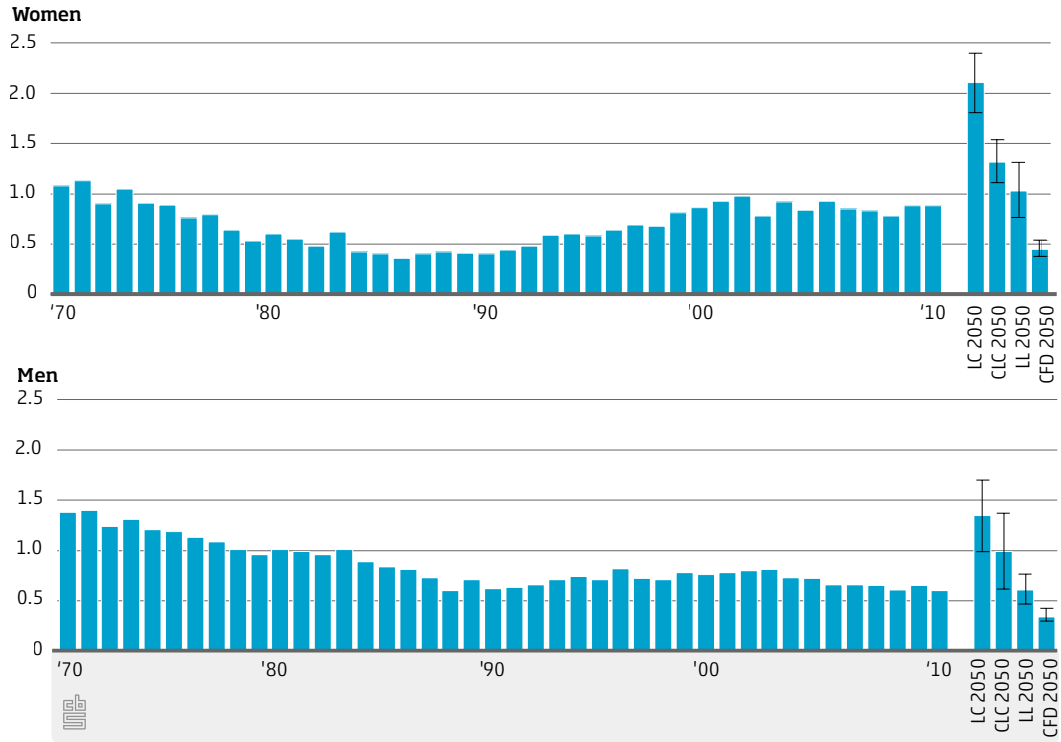
	Fitting period dependence Standard deviation of e0 in 2050, averaged over all seven countries and the three selected main country groups		**Main country group dependence** Standard deviation of e0 in 2050, averaged over all seven countries and the three fitting periods	
	Women	**Men**	**Women**	**Men**
LC method	0.33	1.10	0.00	0.00
CLC method	0.22	0.89	0.69	0.68
LL method	0.16	0.91	0.89	0.54
CFD method	0.32	2.02	0.85	0.87

4.4.6 Plausibility

By comparing the amount of convergence in the projection period related to the fitting period and the improvement of mortality rates by age, we determined the plausibility of the forecasts.

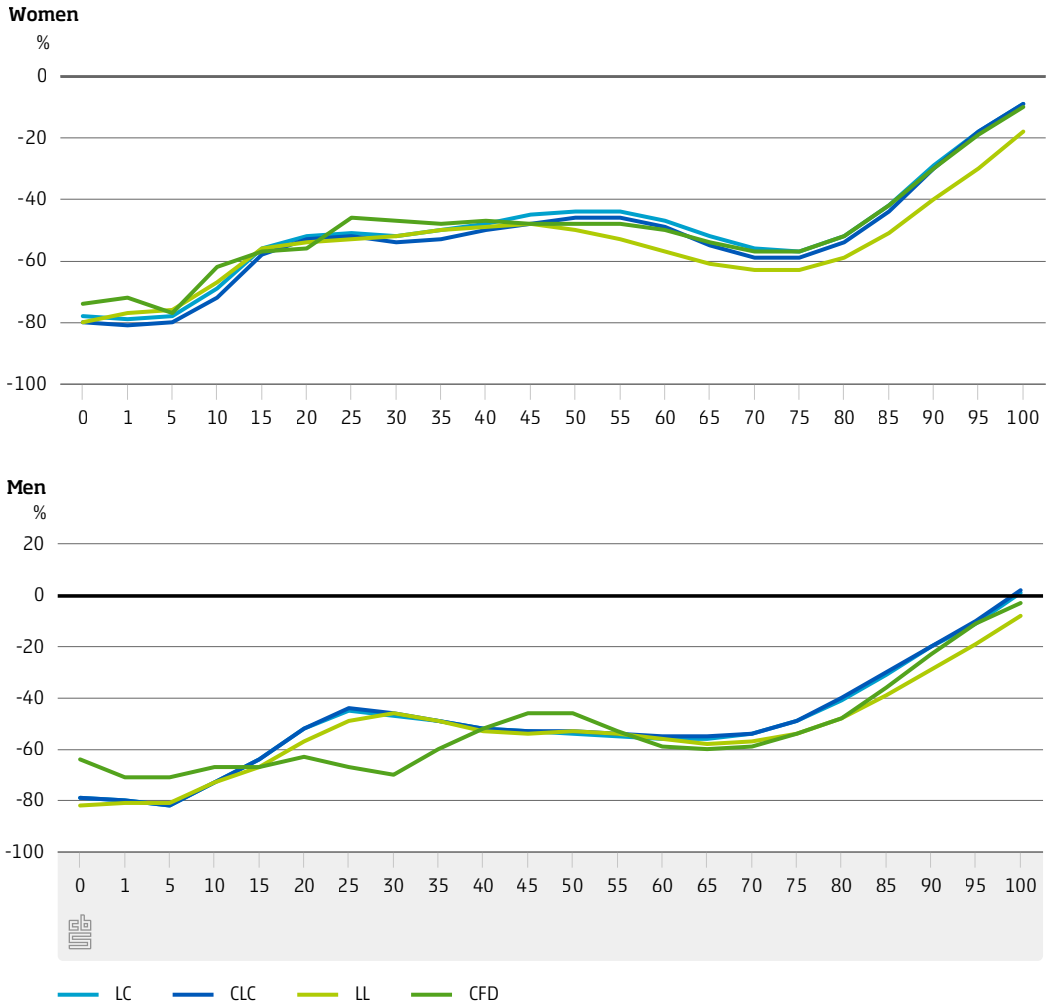
The (unweigthed) average standard deviation of e0 across the seven countries, representing the amount of convergence, in the last ten years of observation is equal to 0.87 for women and 0.69 for men (Figure 4.4.6.1). For the CLC method the average standard deviation of e0 in 2050 is equal to 1.32 for women and 0.99 for men, which means that the CLC method still shows divergence compared to the fitting period. The LL method shows some divergence for women (1.03 in 2050) and some convergence for men (0.61 in 2050) compared to the fitting period. The CFD method results in a clear convergence (0.45 for women and 0.34 for men), much stronger than in the fitting period.

4.4.6.1 Standard deviation of e0 across the seven countries: observations 1970-2011 and projections for 2050 for each method by sex (unweigthed averages over the three main country groups and the three fitting periods)



The improvement of the mortality rates of the LC, CLC and LL method changes gradually by age. The shape of the mortality rates by age of the CLC method is similar to the shape of the LC method, only higher or lower, depending on the fitting period and group of countries used. The shape of the LL method is somewhat different than the LC and CLC method. For the CFD method the improvement of the mortality rates change less gradually by age: there is much difference between adjacent ages, most often around ages 25-30 and 35-40.

4.4.6.2 Mortality improvement between 2011 and 2050 by age group, compared for the four forecasting methods, by sex (fitting period: 1970-2011, main country group: Top10, unweighted averages over the seven countries)



4.5 Discussion

4.5.1 Coherent versus individual forecasts

The outcomes of the coherent forecasts were not only less divergent than the outcomes of the individual forecasts, but also, on average, more accurate and robust. This clearly shows the added value of using coherent forecasts instead of individual forecasts. Shang (2016) also demonstrated that the coherent methods performed better than the individual methods, especially for populations with large variability over age and year. Also Shair, Purcal and Parr (2017) showed that, in terms of overall accuracy, the forecasts of the coherent models are consistently more accurate than those of the independent models.

Our finding that the coherent forecasts resulted, depending on the country, in either higher or lower outcomes than the individual forecast, can be linked to the initial position of the country relative to the group. For countries with lower initial positions, the coherent forecast generally leads to higher outcomes than the individual forecasts, and vice versa.

4.5.2 The different coherent forecasting methods evaluated

The coherent forecasting methods CLC, LL and CFD were evaluated in terms of accuracy, robustness, subjectivity and plausible outcomes.

Accuracy To assess the accuracy of the forecasting methods, the explanation ratio (ER), root mean squared error (RMSE) and mean absolute percent error (MAPE) for the different methods were compared. Furthermore, a Diebold-Mariano test was performed.

The CFD method performed better than the CLC and the LL method, both on the different accuracy measures and on the Diebold-Mariano test applied to the fitting period 1970-2011. The LL method ranked second regarding accuracy. This was expected because models with more parameters normally perform better on these measures. The Diebold-Mariano test applied to the forecast 2002-2011 showed that none of the methods outperformed the other methods on forecast accuracy for all countries.

An earlier comparison of the LL and CFD method, based on a dataset of 16 countries, (Shang, 2016) showed that the point forecast errors for the CFD method are almost always lower than the point forecast errors for the LL method. This is supplementary to our analysis and based on this we can cautiously conclude that the CFD is more accurate than the LL method.

However, for certain forecast applications, examining purely the accuracy of the forecasting method is not sufficient. Therefore we evaluated the three coherent forecasting methods as well on the more qualitative indicators robustness, subjectivity and plausible outcomes.

Robustness To assess the robustness of the forecasting methods, the stability across different fitting periods was evaluated based on out of sample forecasting.

Both the CLC method and the LL method proved robust, i.e. with stable outcomes across different fitting periods. The CFD method proved less robust then the other two coherent forecasting methods and for men even less robust than the individual LC method. This can be related to the CFD method using the weighted principal components algorithm of Hyndman and Shang (2009), which places more weight on recent data. If new recent data, with a different trend than the older data, is added to the fitting period, the out of sample forecast will be different as well. As a result, the CFD method is less stable across different fitting periods.

The weighting on recent data in the CFD method is an advantage in situations where the rates of decrease were not constant for each age in the fitting period, such as the past acceleration in the increase in e0 for men and some improvement at older ages in recent years, whereas in past decades most of the mortality improvement occurred within the younger age groups. Because weighting is at the expense of robustness (i.e. stability across different fitting periods), consideration should be made between weighting (to better fit the recent data) and robustness (to keep stability across different fitting periods).

The LL method in our analysis is estimated using Singular Value Decomposition (SVD). Earlier research (Enchev, Kleinow and Cairns, 2016) showed that when the LL method is calibrated using maximum likelihood estimation (MLE), the model potentially suffers from robustness problems. Therefore, the use of SVD is recommended for a robust forecast.

Looking at the results by main group (table A.2), it stands out that for men the groups with the more linear trend and higher e0 in 2011 (group 'Top 10'), give a lower standard deviation of e0 in 2050, irrespective of the method. For women the

group 'Western Europe' is more linear than group 'Top 10', but the standard deviation of e0 is about equal for both groups, irrespective of the method. From this we may infer that a higher life expectancy in the recent period combined with a more linear trend (in the future) of the group of countries contributes to a more robust method. This means that also coherent methods, like individual methods (Booth et al., 2006), are more robust in situations where age-specific log mortality rates have linear trends.

Subjectivity To assess the subjectivity of the methods, the sensitivity to the choice of the group of countries was examined.

Where the CLC method proved less sensitive to the choice of the group of countries than the CFD method and LL method for women, the LL method proved least subjective for men. The CFD method was most dependent on the choice of the group of countries, which can be related to the strong convergence that seems embedded in this method. It puts more weight on the trend of the group of countries than on the individual country in comparison to the other methods. If other groups with different trends are used this will consequently have a larger effect.

Coherent forecasting methods are sensitive to the choice of the group of countries. Kjaergaard, Canudas-Romo and Vaupel (2015) showed – with preliminary results – that the selection of the main group of countries in coherent forecasting methods has a large effect on the forecasted life expectancy for some Danish women, but not so much for Spanish women. Based on our results we recommend a group of countries with a linear trend in the past to improve robustness of the coherent forecasting method.

Plausible outcomes To assess if the outcomes are plausible, the continuation of trends from the fitting period are examined in terms of convergence/divergence of e0 between the seven countries and consistent age patterns.

In terms of convergence it was observed that the LL outcomes seems most plausible, with convergence level similar to the fitting period. The CFD outcomes revealed a strong convergence, which is a continuation of the trend in recent years, and, therefore, also plausible. The CLC method, however, showed divergence relative to the fitting period.

The CLC method assumes convergence through the parameter for the underlying time development. This method may still lead to diverging mortality forecasts, however, if the relative rate of change differs between the group and the country of interest.

As regards the consistent age patterns, both the CLC method and the LL outcomes looked plausible: the CLC method resulted in improvements in the mortality rates by age that are similar to the LC method; the LL outcomes showed a smooth pattern of age-specific mortality improvements. The CFD method, however, showed strong differences between adjacent ages in the age-specific mortality improvements.

Like most methods, the CLC and LL method assume that the rates of decrease are constant for each age, which not only results in plausible age patterns, but also in a slowdown in the increase in e0. This was evident from our results: the average annual increase in future e0 diminished slightly over the forecast horizon. For a constant (or increasing) annual increase, the rate of decrease of death rates must be nonlinear, and in particular must accelerate for at least some ages (White, 2002). With the extra parameters in the CFD method (up to six principal components) it is possible to produce a variable age pattern of change over time. Furthermore, the weighting ensures that the future age pattern and change in age pattern is more in line with recent data. This is for instance relevant in situations where mortality decline shifted from lower to higher ages. However, there are examples where the changing age pattern is reversed in the projection period (Hyndman, Booth and Yasmeen, 2013). Our results also showed that the CFD method has a deviating pattern of the improvement in the death rate for the younger age groups, while other results with respect to the age distribution seemed plausible.

Overall The CLC method was robust and, for women, least sensitive to the choice of the group of countries, but showed less plausible results: divergence of e0 in the future of the seven countries relative to the fitting period. Also its results were least accurate. The LL method was also robust, least sensitive to the choice of the group of countries for men, and its outcomes seemed plausible, with convergence of future e0 of the seven countries similar to the fitting period and a smooth pattern of age-specific mortality improvements. In terms of accuracy the LL method ranked second. The CFD method performed best on the accuracy measures in the fitting period, but was less robust and most dependent on the choice of the group of countries. Its outcomes revealed a strong convergence of future e0 and – less plausible – difference between adjacent ages in the age-specific mortality improvements.

Based on the above, we deduce that, overall, the CFD method performed best on accuracy (model fit), while the LL method performed best on the qualitative evaluation criteria (robustness, plausible outcomes and subjectivity. The choice of the best method can therefore differ depending on the forecasting application, and the value attached to quantitative versus qualitative criteria. For instance, for forecasts that are updated regularly, robustness should be given higher priority. Given that the

outcomes of future updates are uncertain, a robust forecasting method gives a little bit more certainty for users of the forecasts. Therefore, when robust forecasts are the aim, we would recommend the LL method over the CFD method.

4.5.3 Additional recommendations for coherent forecasting

In this paper we focused on coherence between countries, and not between genders. As a result, in our forecasts, coherence between genders is not guaranteed. Coherence between countries is often neglected in national forecasts, but especially important for countries with a less linear trend in the past. Coherence between genders is important as well to assess the future long-term trend. An approach to ensure coherence between countries and genders that has been used before is to also use the other gender in the group of countries (Janssen, van Wissen and Kunst, 2013). Also other approaches exist (e.g. Hyndman et al., 2011; Shang and Hyndman, 2016; Shang, 2016; Li et al. 2016). These different approaches will likely result in different outcomes. Both gender coherence and country coherence should ideally be incorporated in coherent forecasting.

The coherent mortality projection can be improved by taking into account smoking and other (lifestyle) factors affecting mortality. The non-linear pattern in mortality of most lifestyle factors affects the long-term trend for the country concerned, but also for the total group of countries. Coherent forecasting is most helpful in case of structural improvements in life expectancy because of medical improvements and socio-economic improvements. (Temporary) deviations from the general improvement, caused by lifestyle factors, should be projected separately (Janssen and Kunst, 2007). For smoking, this has been done for example by distinguishing smoking attributable mortality from non-smoking attributable mortality, and by performing the coherent forecast on non-smoking attributable mortality (see e.g. Janssen, van Wissen and Kunst, 2013).

This paper focused on point forecasts, but because future mortality is difficult to predict, measures of uncertainty are also important to users of mortality projections. With all methods analysed here or extensions to the methods, it is possible to produce prediction intervals by using a (Bayesian) stochastic model (see for example Cairns et al. 2011a and Antonio et al. 2015). It should be noted however that prediction intervals do not provide all uncertainty. In stochastic forecasts ideally also the uncertainty due to different selections of groups of countries and different fitting periods should be incorporated.

4.6 Conclusion

In this article, we evaluated three different coherent forecasting methods in terms of accuracy (i.e. fit to historical data), robustness (i.e. stability across different fitting periods), subjectivity (i.e. sensitivity to the choice of the group of countries) and plausible outcomes (i.e. smooth continuation of trends from the fitting period). Out of the three examined methods (the co-integrated Lee-Carter method (CLC); the Li-Lee method (LL); and the Coherent functional data method (CFD)), the CFD method performed the best on the accuracy measures (model fit), whereas the LL method performed best on the qualitative criteria (robustness, subjectivity and plausible outcomes).

Performing better on one quantitative evaluation criteria (e.g. accuracy) clearly does not mean performing better as well on more qualitative evaluation criteria (e.g. robustness, subjectivity and plausibility). To assess the suitability of (coherent) forecasting methods for particular forecasting applications it is essential to include both quantitative and qualitative evaluation criteria. Based on our results, and when the aim is to obtain robustness, subjectivity and plausibility, this would imply the use of the LL method over the CFD method.

Appendix A

A.1 e0 in 2050, given fitting period and group of countries, for each sex, method and country

	LC	CLC	LL	CFD	LC	CLC	LL	CFD	LC	CLC	LL	CFD
a. Women	**1970–2001**				**1970–2006**				**1970–2011**			
Group 'All HMD'												
FRATNP	90.8	90.7	90.0	87.9	91.3	90.7	90.7	88.6	91.2	90.6	90.6	88.9
ITA	91.5	89.6	89.8	87.4	91.7	89.4	89.4	88.0	90.9	89.9	90.2	88.4
NLD	85.7	86.8	86.0	86.6	86.3	86.8	87.5	87.1	87.2	86.7	88.0	87.7
NOR	86.5	87.5	88.2	87.1	87.5	87.6	88.3	87.5	87.5	88.1	88.4	87.9
ESP	91.0	90.8	90.4	87.9	91.0	89.4	90.9	88.3	91.0	90.6	90.9	88.9
SWE	87.6	87.4	88.2	87.2	87.8	87.4	88.4	87.6	87.8	87.8	88.6	88.2
CHE	90.3	89.2	89.0	87.7	90.2	89.0	89.0	88.0	90.1	88.7	89.2	88.5
mean	89.0	88.9	88.8	87.4	89.4	88.6	89.2	87.9	89.4	88.9	89.4	88.4
sd	2.4	1.6	1.5	0.5	2.1	1.4	1.2	0.5	1.8	1.5	1.2	0.5
Group 'Top 10'												
FRATNP	90.8	91.9	91.6	89.8	91.3	91.7	91.8	89.9	91.2	90.9	91.5	90.0
ITA	91.5	91.2	91.5	89.4	91.7	90.9	91.8	89.4	90.9	89.9	90.9	89.4
NLD	85.7	88.2	90.0	88.3	86.3	88.2	88.9	88.9	87.2	87.9	89.2	88.8
NOR	86.5	89.6	90.0	89.0	87.5	89.7	90.5	89.2	87.5	89.4	90.0	89.0
ESP	91.0	91.1	91.6	89.8	91.0	90.8	91.6	89.8	91.0	90.3	91.3	90.0
SWE	87.6	89.4	91.1	89.6	87.8	89.3	91.0	89.5	87.8	89.0	90.2	89.4
CHE	90.3	91.7	91.4	89.8	90.2	91.3	91.4	89.8	90.1	90.1	90.8	89.8
mean	89.0	90.4	91.0	89.4	89.4	90.3	91.0	89.5	89.4	89.6	90.6	89.5
sd	2.4	1.4	0.7	0.6	2.1	1.3	1.0	0.4	1.8	1.0	0.8	0.5
Group 'Western Europe'												
FRATNP	90.8	91.0	90.3	89.1	91.3	91.3	90.8	89.7	91.2	91.1	90.8	90.2
ITA	91.5	90.3	90.2	88.7	91.7	90.5	90.6	89.2	90.9	90.1	90.2	89.6
NLD	85.7	87.2	88.6	88.3	86.3	87.4	87.3	88.7	87.2	88.0	88.3	88.8
NOR	86.5	88.8	88.9	88.4	87.5	89.3	89.4	89.0	87.5	89.6	89.4	89.1
ESP	91.0	90.3	90.3	89.0	91.0	90.4	90.6	89.6	91.0	90.4	90.6	90.1
SWE	87.6	88.7	89.2	88.8	87.8	89.0	89.9	89.3	87.8	89.1	89.8	89.5
CHE	90.3	90.8	90.1	89.0	90.2	90.9	90.3	89.5	90.1	90.3	90.3	89.9
mean	89.0	89.6	89.6	88.8	89.4	89.8	89.8	89.3	89.4	89.8	89.9	89.6
sd	2.4	1.4	0.7	0.3	2.1	1.3	1.2	0.4	1.8	1.0	0.8	0.5

A.1 e0 in 2050, given fitting period and group of countries, for each sex, method and country (continued)

	LC	CLC	LL	CFD	LC	CLC	LL	CFD	LC	CLC	LL	CFD
b. Men	**1970–2001**				**1970–2006**				**1970–2011**			
Group 'All HMD'												
FRATNP	83.6	83.2	82.6	80.4	85.1	84.6	85.3	82.4	85.5	84.3	85.2	86.1
ITA	84.8	83.7	84.1	81.3	86.2	84.1	86.0	83.1	86.3	85.7	86.3	86.6
NLD	80.5	80.6	82.6	81.1	82.7	81.8	83.5	82.6	84.2	83.3	84.8	86.1
NOR	81.2	80.9	83.1	81.2	83.3	82.6	84.2	82.8	84.0	83.4	84.6	86.5
ESP	83.2	83.4	83.2	81.4	84.3	83.6	84.1	82.8	85.5	85.5	85.2	86.6
SWE	83.7	82.4	84.2	81.8	84.2	83.3	84.5	83.0	84.9	84.1	85.4	86.7
CHE	85.0	83.3	83.8	81.6	86.1	84.3	84.3	83.2	86.6	85.1	85.7	87.2
mean	83.1	82.5	83.4	81.3	84.6	83.5	84.6	82.9	85.3	84.5	85.3	86.5
sd	1.7	1.3	0.7	0.5	1.3	1.0	0.8	0.3	1.0	1.0	0.6	0.4
Group 'Top 10'												
FRATNP	83.6	84.7	84.5	82.9	85.1	85.2	85.4	84.1	85.5	85.2	85.5	85.5
ITA	84.8	84.9	85.2	83.4	86.2	85.5	86.1	84.5	86.3	85.5	86.0	85.8
NLD	80.5	81.2	83.8	83.1	82.7	83.3	85.0	84.2	84.2	84.5	85.4	85.7
NOR	81.2	82.4	84.2	83.5	83.3	84.4	85.6	84.6	84.0	84.7	85.6	85.7
ESP	83.2	85.3	84.8	83.6	84.3	85.8	85.4	84.7	85.5	85.1	86.2	86.0
SWE	83.7	83.9	85.4	83.9	84.2	85.1	86.1	84.9	84.9	85.4	86.1	86.1
CHE	85.0	85.4	85.5	83.9	86.1	86.0	86.1	85.0	86.6	85.8	86.9	86.3
mean	83.1	84.0	84.8	83.5	84.6	85.0	85.7	84.6	85.3	85.2	86.0	85.9
sd	1.7	1.6	0.6	0.4	1.3	0.9	0.4	0.3	1.0	0.4	0.5	0.3
Group 'Western Europe'												
FRATNP	83.6	84.1	83.7	82.2	85.1	85.0	85.0	85.3	85.5	85.7	85.6	87.0
ITA	84.8	84.3	84.3	82.8	86.2	85.3	85.6	85.6	86.3	85.7	86.0	87.3
NLD	80.5	81.1	82.3	82.6	82.7	83.2	84.1	85.2	84.2	84.7	85.5	87.0
NOR	81.2	82.0	83.3	83.0	83.3	84.2	85.3	85.8	84.0	85.0	85.6	87.1
ESP	83.2	84.6	84.0	83.0	84.3	85.4	85.1	85.7	85.5	85.5	86.0	87.5
SWE	83.7	83.5	84.5	83.4	84.2	84.9	85.6	85.8	84.9	85.7	86.1	87.4
CHE	85.0	84.8	84.7	83.3	86.1	85.8	85.7	86.0	86.6	86.1	86.7	87.7
mean	83.1	83.5	83.8	82.9	84.6	84.8	85.2	85.6	85.3	85.5	85.9	87.3
sd	1.7	1.4	0.8	0.4	1.3	0.9	0.6	0.3	1.0	0.5	0.4	0.3

A.2 Mean and standard deviation of period life expectancy in 2050 (averaged over all seven countries) for each group of countries, by sex

	Women				Men			
	LC	CLC	LL	CFD	LC	CLC	LL	CFD
Mean								
'All HMD'	89.3	88.8	89.1	87.9	84.3	83.5	84.4	83.6
'Top 10'	89.3	90.1	90.9	89.5	84.3	84.7	85.5	84.6
'Western Europe'	89.3	89.7	89.8	89.2	84.3	84.6	85.0	85.3
St.dev.								
'All HMD'	2.0	1.4	1.3	0.6	1.6	1.3	1.1	2.3
'Top 10'	2.0	1.2	0.9	0.5	1.6	1.2	0.7	1.1
'Western Europe'	2.0	1.2	0.9	0.5	1.6	1.3	1.1	1.9

References

Antonio, K., Bardoutsos, A. and Ouburg, W. (2015). Bayesian Poisson log-bilinear models for mortality projections with multiple populations. *European Actuarial Journal* 5: 245-281.

Bengtsson, T. and Christensen, K. (Eds.) (2006). Perspectives on Mortality Forecasting: IV. The Causes of Death. Stockholm, Swedish *Social Insurance Agency*. Social Insurance Studies 4: 1–73.

Booth, H., Hyndman, R.J., Tickle, L. and De Jong, P. (2006). Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions. *Demographic Research* 15(9): 289–310.

Booth, H. and Tickle, L. (2008). Mortality modelling and forecasting: a review of methods. *Annals of Actuarial Science* 3(1&2): 3–43.

Börger, M. and Aleksic, M-C. (2014). *Coherent Projections of Age, Period, and Cohort Dependent Mortality Improvements*. Presented at the Living to 100 Symposium, Orlando, Fla., January 8–10, 2014.

Cairns, A.J.G, Blake, D., Dowd, L., Coughlan, G.D. and Khalaf-Allah, M. (2011a). Bayesian Stochastic Mortality Modelling for Two Populations. *Astin Bulletin* 41(1): 29–59.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D. and Khalaf-Allah, M. (2011b). Mortality Density Forecasts: An Analysis of Six Stochastic Mortality Models. *Insurance: Mathematics and Economics* 48: 355–367.

Carter, L.R. and Lee, R.D. (1992). Modeling and forecasting US sex differentials in mortality. *International Journal of Forecasting* 8(3): 393–411.

Currie, I.D., Durban, M. and Eilers, P.H.C. (2004). Smoothing and forecasting mortality rates. *Statistical Modellin*g 4: 279–298.

Dowd, K., Blake, D., Cairns, A.J.G., Coughlan, G.D. and Khalaf-Allah, M. (2011). A gravity model of mortality rates for two related populations. *North American Actuarial Journal* 15: 334–356.

Diebold, F.X. and Mariano, R.S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13: 253-263.

Enchev, V., Kleinow, T. and Cairns, A.J.G. (2016). Multi-population mortality models: Fitting, Forecasting and Comparisons. *Scandinavian Actuarial Journal* (forthcoming).

European Commission (2012). *EU rules on gender-neutral pricing in insurance industry enter into force, News, 20-12-2012*. http://ec.europa.eu/justice/newsroom/gender-equality/news/121220_en.htm.

Girosi, F., and King, G. (2006). *Demographic Forecasting*. Cambridge: Cambridge University Press.

Hyndman, R.J. (2010). *Demography: Forecasting mortality, fertility, migration and population data*. R package version 1.07. With contributions from Heather Booth and Leonie Tickle and John Maindonald. Retrieved from http://robjhyndman.com/software/demography.

Hyndman, R.J., Ahmed, R.A., Athanasopoulos, G. and Shang, H.L. (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis* 55: 2579–2589.

Hyndman, R.J., Booth, H. and Yasmeen, F. (2013). Coherent Mortality Forecasting: The Product-Ratio Method With Functional Time Series Models. *Demography* 50: 261–283.

Hyndman, R. J. and Shang, H.L. (2009). Forecasting functional time series (with discussion). *Journal of the Korean Statistical Society* 38: 199–221.

Hyndman, R.J. and Ullah, M.S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis* 51: 4942–4956.

Janssen, F., Mackenbach, J.P. and Kunst, A.E. (2004). Trends in old-age mortality in seven European countries, 1950-1999. *Journal of Clinical Epidemiology* 57(2): 203–216.

Janssen, F. and Kunst, A. (2007). The choice among past trends as a basis for the prediction of future trends in old-age mortality. *Population studies* 61(3): 315–326.

Janssen, F., van Wissen, L.J.G. and Kunst, A.E. (2013). Including the smoking epidemic in internationally coherent mortality projections. *Demography* 50(4): 1341–1362.

Jarner, S.F. and Kryger, E.M. (2011). Modelling Adult Mortality in Small Populations: The SAINT Model. *Astin Bulletin* 41(2): 377–418.

Kjaergaard, S., Canudas-Romo, V. and Vaupel, J.W. (2015). *The importance of the reference population for coherent mortality forecasting models*. Extended abstract for the European Population Conference 2016, Germany.

Kleinow, T. (2015). A common age effect model for the mortality of multiple populations. *Insurance: Mathematics and Economics* 63: 147–152.

Lee, R.D. and Carter, L.R. (1992). Modelling and forecasting US mortality. *Journal of the American Statistical Association* 87(419): 659–671.

Li, N. and Lee, R. (2005). Coherent mortality forecasts for a group of populations: an extension of the Lee-Carter method. *Demography* 42(3): 575–94.

Li, J.S-H. and Hardy, M.R. (2011). Measuring Basis Risk in Longevity Hedges. *North American Actuarial Journal* 15(2): 177–200.

Li, J., Tickle, L. and Parr, N. (2016). A multi-population evaluation of the Poisson common factor model for projecting mortality jointly for both sexes. *Journal of Population Research* 33: 333–360.

Pascariu, M., Canudas-Romo, V. and Vaupel, W.J. (2016). *The double-gap life expectancy forecasting model*. Conference: Population Association of America, 31 mei 2016, Washington D.C.

Pollard, J.H. (1987). Projection of age-specific mortality rates. *Population Bulletin of the United Nations* 21-22: 55–69.

Shair, S., Purcal, S. and Parr, N. (2017). Evaluating Extensions to Coherent Mortality Forecasting Models. *Risks* 5(16): 1-20.

Shang, H.L. (2016). Mortality and life expectancy forecasting for a group of populations in developed countries: a multilevel functional data method. *The Annals of Applied Statistics* 10(3): 1639-1672.

Shang, H.L. and Hyndman, R.J. (2016). Grouped functional time series forecasting: An application to age-specific mortality rates. *Journal of Computational and Graphical Statistics* (to appear).

Stoeldraijer, L., van Duin, C., van Wissen, L. and Janssen, F. (2013). Impact of different mortality forecasting methods and explicit assumptions on projected future life expectancy: The case of the Netherlands. *Demographic Research* 29(13): 323–354.

Tabeau, E. (2001). A review of demographic forecasting models for mortality. In E. Tabeau, A. Van Den Berg Jeths & C. Heathcote (Eds.) *Forecasting mortality in developed countries: insights from a statistical, demographic and epidemiological perspective* (1–32). Kluwer Academic Publishers, Dordrecht.

Wan, C., Bertschi, L. and Yang, Y. (2013). *Coherent mortality forecasting for small populations: an application to Swiss mortality data*. Paper for the AFIR/ERM Colloqium, Lyon, France, June 2014.

White, K.M. (2002). Longevity advances in high-income countries, 1955-96. *Population and Development Review* 28(1): 59–76.

Wilson, C. (2001). On the Scale of Global Demographic Convergence 1950–2000. *Population and Development Review* 27(1): 155–172.

Wong-Fupuy, C. and Haberman, S. (2004). Projecting Mortality Trends: Recent Developments in the United Kingdom and the United States. *North American Actuarial Journal* 8(1): 56–83.

Yang, S.S. and Wang, C.W. (2013). Pricing and Securitization of Multi-Country Longevity Risk with Mortality Dependence. *Insurance: Mathematics and Economics* 52: 157–169.

Zhou, R., Wang, Y., Kaufhold, K., Li, J.S-H. and Tan, K.S. (2012). *Modeling Mortality of Multiple Populations with Vector Error Correction Models: Applications to Solvency II*. Paper for the AFIR/ERM Colloqium, Lyon, France, June 2013.

Zhou, R., Li, J.S-H. and Tan, K.S. (2013). Pricing Standardized Mortality Securitizations: A Two-Population Model with Transitory Jump Effects. *Journal of Risk and Insurance* 80: 733–774.

5.

Comparing strategies for matching mortality forecasts to the most recently observed data: exploring the trade-off between accuracy and robustness

Abstract

BACKGROUND

Given the increased link between retirement age and payments to the development in life expectancy, a precise and regular forecast of life expectancy is of utmost importance. The choice of the jump-off rates, i.e. the rates in the last year of the fitting period, is essential for matching mortality forecasts to the most recently observed data. A general approach to the choice of the jump-off rates is currently lacking.

OBJECTIVE

We evaluate six different options for the jump-off rates and examine their effects on the robustness and accuracy of the mortality forecast.

DATA AND METHODS

Death and exposure numbers by age for eight European countries over the years 1960-2014 were obtained from the Human Mortality Database. We examined the use of model values as jump-off rates versus observed values in the last year or averaged over the last couple of years. The future life expectancy at age 65 is calculated for different fitting periods and jump-off rates using the Lee-Carter model and examined on accuracy (mean absolute forecast error) and robustness (standard deviation of the change in projected e65).

RESULTS

The choice for the jump-off rates clearly influences the accuracy and robustness of the mortality forecast, albeit in different ways. For most countries using the last observed values as jump-off rates resulted in the most accurate method, which relates to the relatively high estimation error of the model in recent years. The most robust method is obtained by using an average of observed years as jump-off rates. The more years that are averaged, the better the robustness, but accuracy decreases with more years averaged.

CONCLUSION

Carefully considering the best choice for the jump-off rates is essential when forecasting mortality. The best strategy for matching mortality forecasts to the most recently observed data depends on the goal of the forecast, the country-specific past mortality trends observed, and the model fit.

Keywords: mortality forecasting, robustness, accuracy, jump-off rates

5.1 Introduction

The growth in public expenditure, such as expenditure on state pension, due to an ageing population is one of the key challenges in European countries (Lanzieri, 2011). To ensure the sustainability of the pension system expenditures, pension reforms in several countries in Western Europe have been carried out, linking the retirement age and/or retirement payments to the rapidly increasing life expectancy (Carone et al., 2016). In some countries, such as Finland and the Netherlands, the link is made with a forecasted remaining life expectancy (OECD, 2015). Given the increased link between retirement age and/or retirement payments to the development in life expectancy, a precise and regular forecast of life expectancy is of utmost importance.

The growing relevance of life expectancy forecasts has resulted in a lot of attention regarding the quality of mortality forecasts. There has been a growing range of models for forecasting mortality and studies performing quantitative and qualitative comparisons of these models (Booth and Tickle, 2008; Cairns et al., 2011; etc.). Also, in recent literature, there has been a lot of attention for the elements that influence the quality of mortality forecasts, i.e. the fitting period (Booth et al. 2002) or additional information, such as smoking (Jansen et al., 2013) or trends in other countries (Li and Lee, 2005). Less attention has been given, however, to the choice of the jump-off rates, i.e. the rates in the last year of the fitting period or jump-off year (Booth et al., 2006). The choice of the jump-off rates is leading when matching the mortality forecast to the most recently observed data. The matching is in turn important for a precise and regular forecast of the life expectancy and thus for the determination of the retirement age and payments.

The choice of the jump-off rates is essential for matching mortality forecasts to the most recently observed data (Lee and Miller, 2001; Booth et al., 2006) and is a practical consideration in every mortality forecast, regardless of the method chosen. A different choice of the jump-off rates may improve the accuracy of a single forecast and/or reduce the discontinuity between the last observed death rate and the first forecasted death rate (Lee and Miller, 2001; Booth et al., 2006). A forecast is called accurate if the out-of-sample forecast errors, examined using historical data, are small (Booth et al., 2008). An accurate method produces precise forecasts which are relevant to determine the retirement age in a future year based on the forecasted life expectancy. However, the choice of the jump-off rates can also influence how much successive forecasts differ, thereby affecting the robustness of the forecast (Cairns et al., 2011). A forecast is called robust if only modest changes in the forecasts occur after a small change to the sample period

(for example, adding the latest mortality data). For instance, if a retirement age in a future year is set based on the forecasted life expectancy, it is undesirable if a forecast based on one more year of data gave a different outcome. Both accuracy and robustness are important for a mortality forecast (Cairns et al., 2011), but can be differently affected by the choice of the jump-off rates.

The choice for the jump-off rates being more a practical problem than a theoretical one, is also highlighted by the fact that there are only four papers about the choice for the jump-off rates. Lee and Carter (1992) used model values (i.e. fitted rates in the jump-off year) as jump-off rates and accepted the discontinuity in observed to forecasted death rates. They stated that the jump-off bias affects only death rates which are absolutely very low and have little impact on the forecasted life expectancy. However, Bell (1997) as well as Lee and Miller (2001) concluded that a correction for the jump-off bias improves the accuracy of the forecast of life expectancy, especially in the early years of the forecast. They used the last observed values (i.e. actual rates) as jump-off rates. Finally, Booth et al. (2006) evaluated as well a two year average of the last observed values as jump-off rates as part of the evaluations of Lee-Carter models and variants. The literature thus gives us only three options to choose from: model values, last observed values, and a two year average of last observed values.

In practice, statistical and actuarial offices use different options for the jump-off rates (mostly last observed values) and, with a new update of the forecast, the choice of the jump-off rates might differ as well. Often, however, it is not explained how they reached these jump-off rates. There are some examples where there are more extensive adjustments of the jump-off bias, but they are relevant for the practical problem at hand and not for universal use (for instance, the statistical office of New Zealand adjusts the rates in the first few years to give plausible life expectancy at birth and death numbers (Woods and Dunstan, 2014)). In fact, a general approach on how to choose between different options for the jump-off rates seems to be lacking.

In the literature (Lee and Carter, 1992; Bell, 1997; Lee and Miller, 2001) and in practice, the jump-off rates are adjusted to improve the accuracy of the forecast. Also quantitative and qualitative comparisons of different models are mainly focused on improving accuracy. However, in light of regular forecasts for the determination of retirement age and payments, it is also of interest to take into account the robustness of the method.

This article examines the effects of different options for the jump-off rates on the accuracy and the robustness of the mortality forecast. This information can be used to determine the optimal choice for a given forecast, which will depend on the relative importance of accuracy and robustness for the applications for which the forecast is used.

We will do so by forecasting future life expectancy at age 65 for eight Western European countries using different fitting periods and six different options for the jump-off rates. A accurate and robust forecast of the life expectancy at age 65, with the mortality forecast matched as optimally as possible to the most recently observed data, is important for the pension reforms in Western Europe.

5.2 Data and methods

5.2.1 Data

For the analysis, deaths and exposures by calendar year and single year of age from the Human Mortality Database (2018) are used, from 1960 to 2014. In our calculations, we aggregated the data for ages 95 and over (Wunsch and Termote, 1978).

To contribute to the debate about the retirement age in Western Europe, and to observe commonalities and differences in the effect of the choice of the jump-off rates for the mortality forecast, data from eight Western European countries is used: the Netherlands ('NLD'), France ('FRA'), Belgium ('BEL'), Spain ('ESP'), Finland ('FIN'), United Kingdom ('UK'), Norway ('NOR') and Sweden ('SWE').

These countries experienced foremost fairly regular mortality trends in the chosen period, for which extrapolative forecasting methods would be suitable. Differences exist however in the extent of mortality decline between the individual countries.

We selected data from 1960 up until 2014, which gave us the opportunity to compute forecasts for the more recent years in the period in order to test the accuracy of the forecasting method.

5.2.2 Model

Many statistical offices are currently using extrapolation methods to forecast mortality (Stoeldraijer et al., 2013). To evaluate the effect of different choices for the jump-off rates, we will apply the most used extrapolation method: the Lee-Carter method (Lee and Carter, 1992; Booth and Tickle, 2008).

The Lee-Carter model (Lee and Carter, 1992) is given by:

$$\ln(m_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t}$$

where $m_{x,t}$ denotes the death rate at age x and year t, a_x equals the average over time of $\ln(m_{x,t})$, b_x is the set of age-s pecific constants that describe the relative rate of change at any age, k_t denotes the underlying time development and $\varepsilon_{x,t}$ the residual error (Lee and Carter, 1992). Singular Value Decomposition is used to estimate b_x and k_t under the assumptions $\sum_x b_x = 1$ and $\sum_t k_t = 0$ (Lee and Carter, 1992). After estimation, k_t is extrapolated using a random walk with drift (as also found by Lee and Carter, 1992, after carrying out the standard model specifications (see Box and Jenkins, 1970)).

5.2.3 Jump-off rates

For the analysis, three options for the jump-off rates are compared:

- Jump-off rates equal to the model values in the last year of the fitting period (Lee and Carter, 1992);
- Jump-off rates equal to the last observed death rates (Lee and Miller, 2001); this corresponds to taking a_x equal to the last observed values of $\ln(m_{x,t})$ and k_t equal to zero in the last observed year;
- Jump-off rates equal to an average of multiple years of the observed death rates; this corresponds to taking a_x equal to the average of a multiple observed years of $\ln(m_{x,t})$ and $k_{\sum t_i / n}$, the midpoint of the years on which is averaged, equal to zero.

By distinguishing four alternatives for the last option (average over two, three, four or five years) we end up with in total six different alternatives.

5.2.4 Analysis

For the analysis, we made for each country ten forecasts of life expectancy at age 65, men and women combined, using data for ten different fitting periods: from

1960-2005, 1960-2006, ..., and 1960-2014. The forecasts are calculated using the six different alternatives of the jump-off rates: the model values, the observed values and an average of two/three/four/five observed values.

Subsequently we compared between the different choices for the jump-off rates the accuracy (fit of the model) and the robustness of the forecast, as these are the most important evaluation criteria for mortality forecasts (e.g. Dowd et al., 2010a,b; Cairns et al., 2011; Booth and Tickle, 2008).

A model is accurate if the out-of-sample forecast errors, examined using historical data, are small (Booth et al. 2008). For evaluating the accuracy we used the mean absolute forecast error (MAFE) (Booth et al. 2006). The MAFE measures how close forecasts are to the eventual outcomes. The smaller the error, the more accurate the forecast, given the option for the jump-off rates. For each country and choice of the jump-off rates, we calculated the MAFE by comparing the forecasted values with the actual values of the life expectancy at age 65. The MAFE for the first year of the forecasting period (i.e. the first year after the fitting period) was calculated using fitting periods 1960–2005, ..., 1960–2013. The forecast of 2006 (with fitting period 1960–2005) was compared with the actual value in 2006, the forecast of 2007 (with fitting period 1960–2006) was compared to the actual value in 2007, and in a similar way for the subsequent forecasts until the forecast of 2014 (with fitting period 1960–2013) which is compared to the actual value of 2014. The errors are then averaged across the nine different forecasts (2006–2014). The MAFE for the second year of the forecasting period was calculated using fitting periods 1960–2005, ..., 1960–2012 and then averaged across the eight different forecasts. In the results (see Table 5.3.2.1), only the MAFE for the first and fifth year of the forecasting period are presented, because the results for the intervening years did not provide useful additional information.

Furthermore, to explain the results regarding the accuracy of the forecast (fit of the model), we calculated the mean absolute (percent) error over the period 1960-2014 and the mean error over the period 2005-2014 of the log death rates, limited to age 65 and above, of the Lee-Carter forecast (estimated over the period 1960-2014).

With a robust forecast, only modest changes would occur to the forecasted life expectancy after a small change to the sample period (e.g. adding one more year) (Cairns et al. 2011). It is important here to look at the stability of each incremental change to the sample period. This is relevant in the case a forecast is regularly updated, i.e. when a new forecast is made each time new data becomes available. Normally, robustness is measured by looking at the changes in model parameters

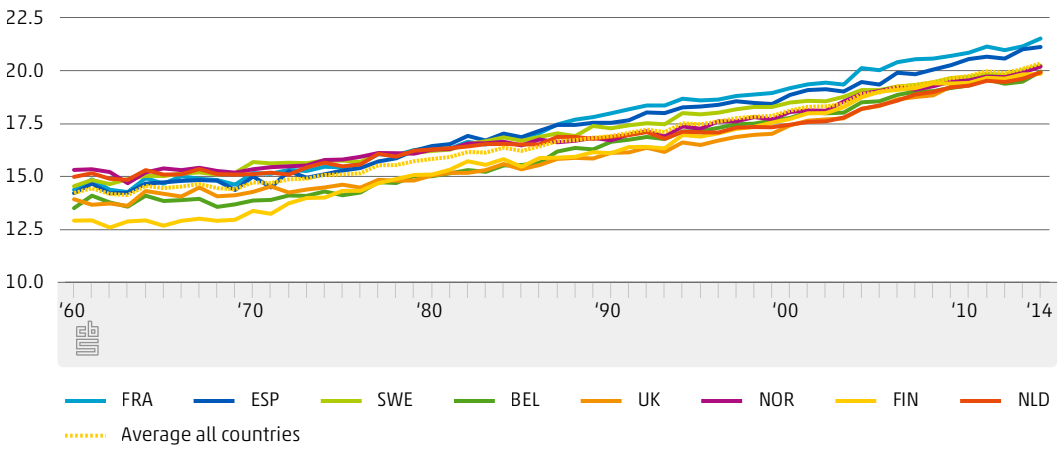
(Cairns et al. 2008). However, these parameters do not depend on the option that is used for the jump-off rates. Therefore, to evaluate the robustness of the forecast given the different options for the jump-off rates, we calculated the standard deviation (SD) of the increase/decrease of the (out-of-sample) life expectancy at age 65 in 2020 obtained for ten successive forecasts using fitting periods 1960-2005, ..., 1960-2014. A lower SD means that the forecast is more robust.

5.3 Results

5.3.1 Past trends in remaining life expectancy at age 65

Over the period 1960 to 2014, the remaining life expectancy at age 65 (e65) increased in the eight selected European countries, for men and women combined (Figure 5.3.1.1). On average, from 14.2 years in 1960 to 20.3 years in 2014. France has seen the largest increase over the whole period, while Norway has seen the lowest increase. Especially for the Netherlands and the United Kingdom there was a higher increase in e65 in the last decade of the observation period than in the decades before. In 2014, the highest e65 was observed for France (21.5 years) and the lowest for Finland (19.8 years), United Kingdom (19.9 years) and the Netherlands (19.9 years).

5.3.1.1 Life expectancy at age 65, 1960-2014 for eight countries, men and women combined



5.3.2 Effect of choice of jump-off rate on accuracy of mortality forecast

The six different options for the jump-off rates resulted in clear differences in the outcome for the accuracy of the forecast (see Table 5.3.2.1): on average there was a difference of 0.18 between the option which gave the minimum accuracy and the option which gave the maximum accuracy for the first year of the forecasting period. The minimum difference was found for France (0.03) and the maximum difference was found for United Kingdom (0.43). With a large difference, such as for United Kingdom, it makes a clear difference which option is chosen for the jump-off rates. The larger the difference between the options for the jump-off rates, the more important it is to choose the correct jump-off rates so that the accuracy of the forecast can be improved.

Using the last observed values as jump-off rates or an average of two years resulted in the most accurate forecast in the first year of the forecasting period for most countries (looking at the minimum MAFE by country). Only for Belgium the most accurate forecast was achieved by using the model values as jump-off rates. The minimum MAFE in the first year of the forecast ranged from 0.07 (Sweden) to 0.15 (Spain). The most accurate forecast for the fifth year of the forecasting period was achieved by using the last observed values as jump-off rates, except for Spain, where the most accurate forecast was achieved by using the model values as jump-off rates. The minimum MAFE in the fifth year of the forecast ranged from 0.10 (Sweden) to 0.38 (United Kingdom).

Except for France, Spain and Belgium using the model values as jump-off rates resulted in the least accurate forecast in the first year. For France, Spain and Belgium using the average of five observed years resulted in the least accurate forecast in the first year. The least accurate for the fifth year of the forecasting period showed the same pattern as the accuracy for the first year of the forecasting period. For all countries the accuracy decreases distinctly with the averaging of more years.

Generally, the MAFE in the fifth year is higher than in the first year (using the same option for the jump-off rates), reflecting that uncertainty further in the future is greater.

The most optimal choice for the jump-off rates for an accurate forecast is related to the error the model makes in the recent estimation period (fitting errors, Table 5.3.2.2). The mean error over the estimation period 2005-2014 is close to zero for Belgium.

5.3.2.1 Mean absolute forecast error (MAFE) of remaining life expectancy at age 65 for the first and fifth year in the forecasting period, for six different choices of the jump-off rates applied to a Lee-Carter model, for eight Western European countries, men and women combined, fitting periods 1960-2005, 1960-2006, ..., 1960-2014. The lowest MAFE is marked in bold, the highest MAFE in italic

Jump-off rates	FRA	ESP	SWE	BEL	GBR	NOR	FIN	NLD	Av[1)]
	Mean absolute forecast error in the first year of the forecasting period								
Model values	0.16	0.16	*0.24*	**0.10**	*0.56*	*0.24*	*0.41*	*0.39*	*0.28*
Last observed values	**0.13**	0.17	0.08	0.14	**0.13**	0.12	0.10	**0.13**	**0.12**
Average two years observed	0.13	**0.15**	**0.07**	0.15	0.16	**0.11**	**0.08**	0.16	0.13
Average three years observed	0.13	0.16	0.09	0.14	0.19	0.12	0.09	0.21	0.14
Average four years observed	0.14	0.17	0.10	0.15	0.24	0.13	0.12	0.26	0.16
Average five years observed	*0.16*	*0.19*	0.11	*0.17*	0.28	0.16	0.16	0.32	0.19
	Mean absolute forecast error in the fifth year of the forecasting period								
Model values	0.20	**0.17**	*0.31*	0.18	*0.76*	*0.45*	*0.41*	*0.55*	*0.38*
Last observed values	**0.16**	0.30	**0.10**	**0.17**	**0.38**	**0.20**	**0.13**	**0.31**	**0.22**
Average two years observed	0.17	0.32	0.11	0.18	0.41	0.21	0.14	0.35	0.24
Average three years observed	0.18	0.35	0.13	0.19	0.44	0.24	0.15	0.41	0.26
Average four years observed	0.20	0.38	0.16	0.20	0.46	0.29	0.18	0.47	0.29
Average five years observed	*0.21*	*0.40*	0.18	*0.21*	0.49	0.34	0.19	0.55	0.32

[1)] Unweigthed average of all eight countries

5.3.2.2 Mean absolute (percent) error over the period 1960-2014 and mean error over the period 2005-2014 of the log death rates limited to age 65 and above of the Lee-Carter forecast estimated over the period 1960-2014, for eight Western European countries, men and women combined

	FRA	ESP	SWE	BEL	GBR	NOR	FIN	NLD
Mean abs error (1960–2014)	0.028	0.040	0.031	0.032	0.036	0.037	0.046	0.036
Mean abs perc error (1960–2014)	1.20%	1.72%	1.49%	1.61%	1.51%	1.61%	2.34%	1.65%
Mean error (2005–2014)	0.017	0.019	–0.016	0.001	–0.035	–0.016	–0.028	–0.019

This was the only country for which the model values as jump-off rates gave the most accurate results. For Sweden, United Kingdom, Norway, Finland and the Netherlands the mean error is negative, i.e. the recent period was underestimated by the Lee-Carter model. For these countries the model values as jump-off rates were the worst option for an accurate forecast. These countries had a stronger increase in e65 in the recent decade compared to earlier decades. Hence using observed values as jump-off rates would mean the forecast is already closer to the observed future values than using the model values. For France and Spain the mean error was positive and differences in accuracy between the options for the jump-off rates were small.

5.3.3 Effect of choice of jump-off rate on robustness of mortality forecast

The choice of the jump-off rates clearly affected the robustness of the forecast: on average there was a difference of 0.18 between the minimum value of the SD and the maximum value of the SD. The minimum difference was found for the Netherlands (0.10) and maximum for Finland (0.36) (Table 5.3.3.1).

Using an average of recent observed years as jump-off rates results in a lower standard deviation of the increase/decrease of the life expectancy at age 65 in 2020 and thus a more robust method. This holds for all countries under study. The minimum standard deviation (per country) ranges from 0.05 (Spain and Sweden) to 0.15 (Finland). For most countries an average of at least four years gives the minimum standard deviation. The difference between a two year average and a five year average (maximum 0.5) is small compared to the differences with last observed or model values.

The worst options for the jump-off rates in terms of the robustness are either the model values (Sweden, Norway, Finland and the Netherlands) or the observed values (France, Spain, Belgium and United Kingdom). The maximum standard deviation (per country) ranges from 0.17 (United Kingdom and the Netherlands) to 0.51 (Finland).

5.3.3.1 Standard deviation (SD) of the increase/decrease of the life expectancy at age 65 in 2020 between ten successive forecasts (fitting periods 1960-2005, 1960-2006, ..., 1960-2014) for six different choices of the jump-off rates applied to a Lee-Carter model, for eight Western European countries, men and women combined. The lowest SD is marked in bold, the highest SD in italic

Jump-off rates	FRA	ESP	SWE	BEL	GBR	NOR	FIN	NLD	Av[1)]
	Standard deviation in 2020								
Model values	0.11	0.11	*0.23*	0.17	0.15	*0.26*	*0.51*	*0.17*	*0.22*
Last observed values	*0.19*	*0.24*	0.09	*0.20*	*0.17*	0.12	0.16	0.13	0.16
Average two years observed	0.11	0.10	0.06	0.10	0.11	0.09	**0.15**	0.08	0.10
Average three years observed	0.10	0.09	0.06	0.08	0.09	0.09	0.18	0.08	0.10
Average four years observed	0.08	**0.05**	**0.05**	0.06	0.07	0.10	0.18	0.08	0.08
Average five years observed	**0.07**	0.05	0.06	**0.06**	**0.06**	**0.09**	0.16	**0.07**	**0.08**

[1)] Unweigthed average of all eight countries

The fact that the last observed values as jump-off rates are not performing well on robustness is related to the nature of the data: the observed life expectancy fluctuates greatly from year to year. By using the last observed values as jump-off rates in the forecasting model, also the future values will fluctuate when recent data is added. Taking an average of multiple years makes sure there are fewer fluctuations. The model values are similar to taking an average, but over the whole period in that case. Because the relative decline of the model will influence the forecast more when using the model values as jump-off rates than an average of recent observed years, the robustness of the forecast is better using the average as jump-off rates.

Another feature of the results is also apparent: the countries in the south of Western Europe (France and Spain) have the last observed values as the worst option for the jump-off rates, but for the countries in the north of Western Europe (Sweden, Norway and Finland) have the model values as the worst option. For France and Spain the model values are not much different with the average as jump-off rates, while for Sweden, Norway and Finland the difference between the last observed values and the average are small. Belgium, United Kingdom and the Netherlands are more in between (in location and in the results). For these three countries it also holds that the difference between using the model values and last observed values as jump-off rates does not differ as much as for the other five countries.

5.4 Discussion

We evaluated the accuracy and robustness of the forecast of life expectancy at age 65 in Western Europe for six different options for the jump-off rates. We observed that the options for the jump-off rates clearly influence the accuracy and robustness of the mortality forecast, albeit in different ways. For most countries, the most accurate forecast resulted from taking the last observed values as jump-off rates, which relates to the relatively high estimation error of the model in recent years. The most robust forecast was obtained by using an average of the most recent observed years as jump-off rates. The more years that are averaged, the better the robustness, but accuracy decreases with more years averaged. The best choice for the jump-off rates, thus, seems to depend on whether you are interested mainly in accuracy or robustness, on the country-specific past mortality trends, and the model fit.

The influence of the choice of the jump-off rates on the accuracy and robustness of the forecast can be substantial. Figure A.1 in Appendix A gives an example for the Netherlands of a forecast with the model values as jump-off rates, a forecast with the last observed values as jump-off rates and a forecast with an average of five observed years as jump-off rates, with different fitting periods. The forecasts with the model values as jump-off rates are not accurate, i.e. there are large gaps between the observed values and the forecasts in the first year. The forecasts with the model values are also not robust: the successive forecasts, using different fitting periods, show large differences (i.e. increases and decreases between successive forecasts) between the successive forecasted e65 for a particular year. For the forecasts with the last observed values as jump-off rates the accuracy is improved, and, from the analysis, the most accurate from the six options for the jump-off rates. However, the successive forecasts are also showing large differences between the successive forecasted e65 in a particular year. Lastly, the successive forecasts with an average of five observed years as jump-off rates are slowly increasing with each new year of data added to the fitting period. This option for the jump-off rates was the most robust for the Netherlands.

5.4.1 Evaluation of analysis

We assessed the effect of the choice of the jump-off rates by means of two important evaluation criteria for a mortality forecasting method: robustness and accuracy (Dowd et al., 2010a,b; Cairns et al., 2011). A third evaluation criterion for evaluating a mortality forecast is plausibility (Cairns et al., 2011): is the outcome of the forecast reasonable given what we know? This is rather a subjective issue for which there are no objective measures and for that reason we did not include it in the analysis. Nonetheless, plausibility is important to consider when performing a mortality forecast. A plausible future age pattern is an important issue related to the plausibility of the results. Different characteristics of the jump-off rates, such as a rough age pattern of the last observed values, have an effect on the plausibility of the future age pattern of mortality. To limit the effect of the choice of the jump-off rates on the plausibility of the future age pattern, smoothing the observed mortality rates by age is recommended.

We performed the different mortality forecasts using the Lee-Carter method, which is frequently used for mortality forecasting in practice (Stoeldraijer et al., 2013) , as benchmark method (Booth and Tickle, 2008), and as the basis for more recent mortality forecasting models (Booth and Tickle, 2008; Lee and Carter, 1992). The Lee-Carter method, however, is known to be biased and tends to underpredict future mortality (Bell, 1997; Lee, 2000; Lee and Miller, 2001; Booth et al., 2002;

Girosi and King, 2007; Liu and Yu, 2011), as we have also seen in Table 5.3.2.2. where the mean error in the last ten years of the fitting period was negative for most countries. Therefore, differences between the last observed values and the model values tend to be relatively large. For this reason we performed a sensitivity analysis using two additional models: (i) a Lee-Carter model using three principal components (Appendix B), because based on earlier research it is unnecessary to adjust the jump-off rates when several principal components are used (Hyndman et al., 2013), and (ii) the Cairns-Blake-Dowd model (Cairns et al., 2006; Appendix C), which is considered a different stochastic model compared to the Lee-Carter model and widely used in actuarial sciences. The results show smaller differences in outcomes compared to differences we observed earlier with the Lee-Carter model, but, especially for accuracy, the importance of the jump-off rates remains. This highlights the importance of the model for the best choice of the jump-off rates.

We showed the results of our analysis for men and women combined. Similar results are observed however for men and women separately (see Tables D.1 and D.2 in the Appendix D). Also for men (with the exception of Finland) and women separately, an average of multiple years as jump-off rates was preferred for the most robust forecast. For the most accurate forecast there was some more variation in the results for men and women separately compared to men and women combined. For men in France and Spain the forecast is most accurate when using model values as jump-off rates, although accuracy is only slightly higher compared to the last observed values. For women, the most accurate forecast in the fifth year of the forecasting period is obtained by using the last observed values. The accuracy of the forecast for the first year of the forecasting period shows for women mostly small differences between choices for the jump-off rates, but resulted in model values (France, Sweden), last observed values (Belgium, United Kingdom), and an average (Spain, Norway, Finland, the Netherlands).

We deliberately computed the accuracy and the robustness measures directly for life expectancy at age 65, because of the use of this indicator in the pension reforms. For different contexts, e.g. life insurance and pension valuation, an evaluation of other outcomes (e.g. death rates or probabilities) would be relevant and could lead to different outcomes. That is for different age groups the model fit, and subsequently the choice of the jump off rates, might be different. Booth et al. (2006) compared both errors in life expectancy and log death rates when analysing the accuracy for different choices of the jump-off rate. They concluded that accuracy in log death rates does not necessarily translate into accuracy in life expectancy. Analysis based on forecasted log death rates might therefore lead to different conclusions, but in general last observed values as jump-off rates would give the most accurate forecast (Booth et al. 2006). The above indicates that the context of

the forecast determines the outcome measure used in the analysis of the jump-of rates and, hence, the final choice for the best jump-off rates.

5.4.2 Generalizability of our outcomes

We evaluated the results based on the life expectancy at age 65 in relation to pension reforms. Results based on the life expectancy at birth (e0) are very similar to the results based on e65 (Appendix E). The differences between the six options for the jump-off rates for both accuracy and robustness are slightly larger for e0 than for e65. This means that our conclusions can be generalized to other ages of life expectancy.

We focused our analysis on Western Europe, because of the prevalence of the pension reforms. Our findings can be generalized to countries which have seen similar trends in the past. For example, the results for the Netherlands are expected to be close to the results for Denmark, since both experienced a stagnation of the increase in life expectancy at approximately the same time (Janssen et al., 2004). Similarly our results for the remaining Western European countries can be generalized to other countries exhibiting fairly regular increases in life expectancy, like Japan since 1970 (Leon, 2011). Generalising our results to Eastern Europe however will be more daunting because these countries experienced very different past mortality trends due to the health crisis from 1975 onwards (McKee and Shkolnikov, 2001; Vallin and Meslé, 2004; Leon, 2011). The Lee-Carter method is most likely not suited to account for these specific past mortality trends (Bohk and Rau, 2015). Before evaluating different choices for the jump-off rates in the context of Eastern Europe, first the forecasting method needs to be improved.

5.4.3 Recommendations

Following our findings, we recommend the goal of the forecast, and the related emphasis on accuracy, robustness or both, to be leading for determining the best choice of the jump-off rates.

If the goal of the mortality forecast is focused on accuracy, it is relevant to examine the error of the estimates of the model over the period it is applied to, following its importance in explaining our results for accuracy. We recommend the model values as most suitable as jump-off rates for an accurate forecast when the errors are small. We recommend the last observed values as most suitable jump-off rates when the model errors are large and there is underestimation of the model in the

most recent period. With large errors and an overestimation of the model in the most recent period, we recommend to use the model values as jump-off rates, following our results of men and women separately.

If the goal of the mortality forecast is focused on robustness, we recommend using an average of multiple years as jump-off rates, as was the most suitable for a robust forecast for all countries in our analysis. There was little difference in the outcomes between a two year average and a five year average, thus the number of years used in the averaging is less important. Robustness becomes more important in situations where the forecast is made regularly, for instance when the future retirement age based on the forecasted life expectancy needs to be determined every year.

Because often the goal of the forecast is focused both on accuracy and robustness, the most optimal choice for the jump-off rates must give the most accurate as well as the most robust forecast. For each country in our analysis, there was no option of the jump-off rates that guaranteed accuracy and robustness at the same time. Thus, there always has to be a trade-off between accuracy and robustness. Therefore, we recommend looking into developing a choice for the jump-off rates that is both accurate and robust. Our four recommendations for determining the best choice for the jump-off rates that give both an accurate and robust forecast are: (1) Because the accuracy of the forecast decreases distinctly with the averaging of more observed years as jump-off rates, whereas the robustness of the forecast stayed approximately the same, it is preferable to use an average using as few years as possible to improve the accuracy with a robust forecast. (2) Using the observed values instead of the model values in case the model fits the data well does not improve accuracy and deteriorates the robustness. Thus, in the case the model fits well, it is best to use the model values as jump-off rates and not the observed values as is often done by force of habit. (3) The further ahead, the less accurate the forecast gets. This means that the relative price you pay for more robustness is lower for a forecast further in the future. If the forecast further in the future is of more importance than the short-term forecast, there should be a greater value attached to the robustness of the forecast, and thus the best option for the most robust forecast can be selected. (4) In line with the previous recommendations: to best unite the results for robustness and accuracy we would recommend interpolation (see Appendix F for an example). Robustness is more important for the long-term forecast (for instance, from five years in the future) as a result of the increasing uncertainty with duration. For the first few years accuracy would be more relevant because data for these years will be available quickly. Our recommendation would be to start with a forecast using a jump-off rate that is the most accurate in the first year. Subsequently, make a forecast that is most robust in, say, the fifth year of the forecast period. Between the two forecast, each year more

weight should be given to the most robust forecast, i.e. we recommend interpolating from the most accurate forecast to the most robust forecast. By interpolating between the two forecasts both accuracy in the first year of the forecast and robustness of the forecast five years ahead is obtained.

An additional issue to consider: To match the forecast to recent data it is important that it is of good quality. Preliminary data might underestimate or overestimate the life expectancy. Using jump-off rates based on this data might not work well for the accuracy (to final data) of the forecast. It might also turn out to be disadvantageous for the robustness if the preliminary data is replaced by final data. The use of preliminary data is therefore not recommended when matching the forecast to recent data.

5.5 Overall conclusion

The choice of the jump-off rates clearly influences, in different ways, the accuracy and robustness of the mortality forecast. It is therefore important to carefully consider the best choice for the jump-off rate when forecasting mortality. This is especially relevant when a forecast is regularly updated, as is the case for the pension reforms.

The best choice depends on the goal of the forecast, the country-specific past mortality trends observed, and the model fit. Because the best option of the jump-off rates for accuracy (most often last observed values) and the best option for robustness (average of observed years) are not equal, there will always have to be a trade-off between the two. The recommendations presented, of which interpolation between the jump-off rates with optimal accuracy and optimal robustness combines accuracy and robustness, give guidelines to make a just trade-off between accuracy and robustness of the forecast.

References

Bell, W.R. (1997). Comparing and Assessing Time Series Methods for Forecasting Age-Specific Fertility and Mortality Rates. *Journal of Official Statistics* 13(3): 279–303.

Bohk, C. and Rau, R. (2015). Impact of Economic Conditions and Crises on Mortality and its Predictability. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 67: 271–294.

Booth, H., Maindonald, J. and Smith, L. (2002). Applying Lee-Carter under conditions of variable mortality decline. *Population Studies* 56(3): 325–336. doi:10.1080/00324720215935.

Booth, H., Hyndman, R., Tickle, L. and De Jong, P. (2006). Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions. *Demographic Research* 15(9): 289–310.

Booth, H. and Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. *Annals of Actuarial Science* 3: 3–43. DOI: 10.1017/S1748499500000440.

Box, G. & Jenkins, G. (1970). Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day.

Cairns A.J.G., Blake, D.P. and Dowd, K. (2006). A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration. *Journal of Risk and Insurance* 73(4): 687–718.

Cairns, A.J.G., Blake, D.P. and Dowd, K. (2008). Modelling and Management of Mortality Risk: A Review. *Scandinavian Actuarial Journal* 2(3): 79-113.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D. and Khalaf-Allah, M. (2011b). Mortality Density Forecasts: An Analysis of Six Stochastic Mortality Models. *Insurance: Mathematics and Economics* 48: 355–367.

Carone, G., Eckefeldt, P., Giamboni, L., Laine, V. and Pamies Sumner, S. (2016). *Pension Reforms in the EU since the Early 2000's: Achievements and Challenges Ahead*. European Economy, Discussion Papers 42. December 2016. Brussels. 64pp. DOI: 10.2765/620267.

Dowd, K., Blake, D., Cairns, A.J.G., Coughlan, G.D., Epstein, D. and Khalaf-Allah, M. (2010a). Evaluating the goodness of fit of stochastic mortality models. *Insurance: Mathematics and Economics* 47: 255–265.

Dowd, K., Blake, D., Cairns, A.J.G., Coughlan, G.D., Epstein, D. and Khalaf-Allah, M. (2010b). Backtesting stochastic mortality models: An ex-post evaluation of multi-period-ahead density forecasts. *North American Actuarial Journal* 14: 281–298.

Girosi, F. and King, G. (2007). *Understanding the Lee-Carter Mortality Forecasting Method.* Harvard University Working Paper, Cambridge, MA

Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on 22-05-2018).

Hyndman, R., Booth, H. and Yasmeen, F. (2013). Coherent mortality forecasting: the product-ratio method with functional time series model. *Demography* 50: 261-283.

Janssen, F., Mackenbach, J.P. and A.E. Kunst (2004). Trends in old-age mortality in seven European countries, 1950-1999. *Journal of Clinical Epidemiology* 57(2): 203-216.

Janssen, F., Van Wissen, L.J.G. and Kunst, A.E. (2013). Including the smoking epidemic in internationally coherent mortality projections. *Demography* 50(4): 1341–1362. doi:10.1007/s13524-012-0185-x.

Lanzieri, G. (2011). The greying of the baby boomers, A century-long view of ageing in European populations. *Population and social conditions* Statistics in focus 23/2011.

Lee, R.D. (2000). The Lee-Carter Method for Forecasting Mortality, with Various Extensions and Applications. *North American Actuarial Journal* 4(1): 80–93.

Lee, R.D. and Carter, L.R. (1992). Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association* 87(419): 659–671.

Lee, R.D. and Miller, T. (2001). Evaluating the Performance of the Lee-Carter Method for Forecasting Mortality. *Demography* 38(4): 537–549.

Leon, D.A. (2011). Trends in European life expectancy: a salutary view. *International Journal of Epidemiology* 40(2): 271–277.

Li, N.R. and Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography* 42(3): 575–594. doi:10.1353/dem.2005.0021.

Liu, X. and Yu, H. (2011). *Assessing and extending the lee-carter model for long-term mortality prediction*. Orlando, Fla., January 2011. Living to 100 Symposium.

McKee, M. and Shkolnikov, V. (2001). Understanding the toll of premature death among men in eastern Europe. *BMJ* 323:1051.

OECD (2015). *Pensions at a Glance 2015: OECD and G20 indicators*. OECD Publishing, Paris. http://dx.doi.org/10.1787/pension_glance-2015-en

Stoeldraijer, L., Van Duin, C., Van Wissen, L. and Janssen, F. (2013). Impact of different mortality forecasting methods and explicit assumptions on projected future life expectancy: The case of the Netherlands. *Demographic Research* 29(13): 323–354. DOI: 10.4054/DemRes.2013.29.13.

Vallin, J. and Meslé, F. (2004). Convergences and divergences in mortality: A new approach of health transition. *Demographic Research* S2(2): 11-44.

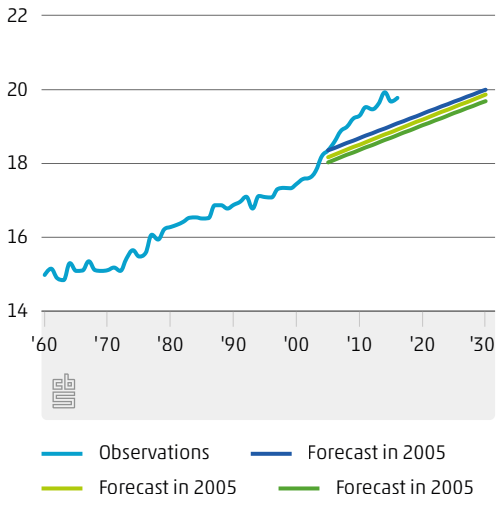
Woods, C. and Dunstan, K. (2014). *Forecasting mortality in New Zealand: A new approach for population projections using a coherent functional demographic model.* Statistics New Zealand Working Paper No 14–01. Available from www.stats.govt.nz.

Wunsch, G.J. and Termote, M.G. (1978). *Introduction to demographic analysis; principles and methods*. Plenum Press, New York.

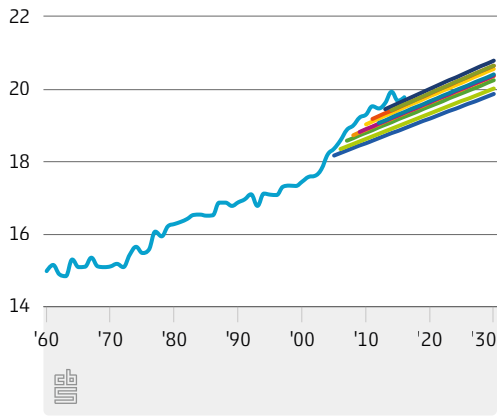
Appendix A

A.1 Example of forecasts with different choices for the jump-off rates, the Netherlands, men and women combined, Lee-Carter model, fitting periods 1960-2005, 1960-2006, ..., 1960-2014

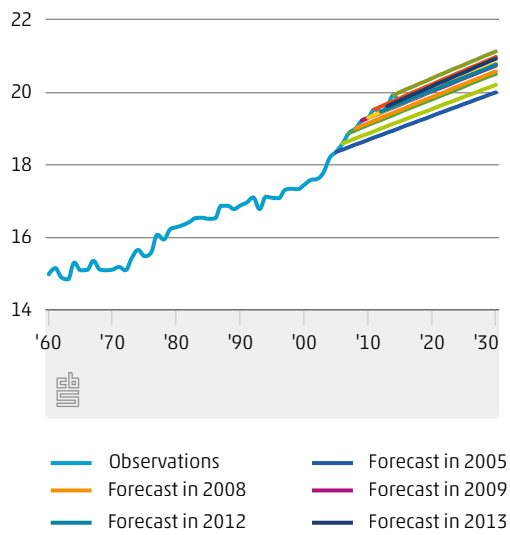
a. Forecast in 2005 with three options of the jump-off rates



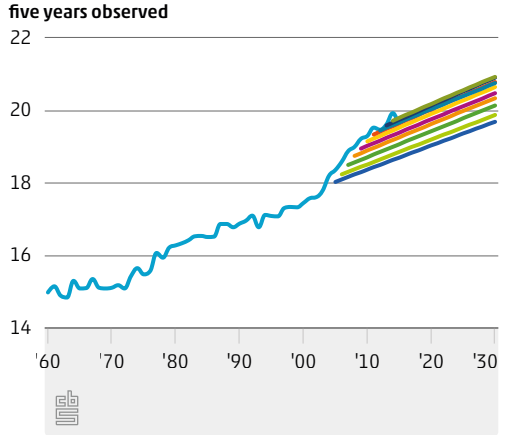
b. Forecasts with jump-off rates equal to model values



c. Forecasts with jump-off rates equal to last observed values



d. Forecasts with jump-off rates equal to the average five years observed



Appendix B

In Tables B.1 and B.2 the results of the different choices for the jump-off rates with a Lee-Carter model using three principle components are presented. The different model mostly influences the robustness results, with for France, Spain and the Netherlands the model values as jump-off rates as the most optimal. Differences in outcome between the choices of the jump-off rates are much smaller than the difference in outcome with the Lee-Carter model with only one principal component. The results for the accuracy of the model are similar as with the Lee-Carter model with only one principal component (mostly observed values as jump-off rates are most optimal). Again we see less differences in outcomes between the different choices of the jump-off rates.

B.1 Mean absolute forecast error (MAFE) of remaining life expectancy at age 65 for the first and fifth year in the forecasting period and standard deviation (SD) of the increase/decrease of the life expectancy at age 65 in 2020 between ten successive forecasts, for six different choices of the jump-off rates applied to a Lee-Carter model with three principal components, for eight Western European countries, men and women combined, fitting periods 1960-2005, 1960-2006, ..., 1960-2014. The lowest MAFE is marked in bold, the highest MAFE in italic

	FRA	ESP	SWE	BEL	GBR	NOR	FIN	NLD	Av[1)]
Jump-off rates	**Standard deviation in 2020**								
Model values	**0.17**	**0.23**	*0.25*	0.14	0.16	*0.28*	0.50	**0.20**	*0.24*
Last observed values	*0.21*	*0.30*	**0.17**	*0.20*	0.14	0.19	**0.45**	*0.25*	0.24
Average two years observed	0.20	0.25	0.18	0.15	**0.13**	**0.17**	0.51	0.20	**0.22**
Average three years observed	0.18	0.24	0.19	0.14	0.16	0.22	*0.54*	0.21	0.23
Average four years observed	0.18	0.23	0.20	0.13	0.17	0.23	0.53	0.21	0.23
Average five years observed	0.18	0.24	0.21	**0.13**	*0.17*	0.24	0.52	0.21	0.24
	Mean absolute forecast error in the first year of the forecasting period								
Model values	*0.24*	*0.24*	*0.22*	0.12	*0.21*	*0.22*	*0.36*	*0.21*	*0.23*
Last observed values	**0.14**	0.18	0.13	*0.14*	**0.14**	0.13	**0.30**	0.18	0.17
Average two years observed	0.15	0.15	0.12	0.12	0.15	**0.13**	0.31	**0.16**	**0.16**
Average three years observed	0.15	**0.15**	0.12	**0.12**	0.17	0.14	0.31	0.17	0.17
Average four years observed	0.16	0.16	**0.12**	0.12	0.18	0.15	0.31	0.18	0.17
Average five years observed	0.17	0.16	0.13	0.12	0.19	0.15	0.32	0.19	0.18
	Mean absolute forecast error in the fifth year of the forecasting period								
Model values	*0.31*	*0.40*	*0.31*	0.18	0.51	*0.43*	*0.38*	*0.41*	*0.37*
Last observed values	**0.18**	**0.23**	**0.15**	**0.17**	**0.43**	0.28	**0.24**	0.36	**0.26**
Average two years observed	0.20	0.24	0.20	0.20	0.46	**0.28**	0.25	**0.36**	0.27
Average three years observed	0.21	0.25	0.19	0.21	0.48	0.29	0.25	0.37	0.28
Average four years observed	0.23	0.27	0.20	*0.22*	0.50	0.30	0.25	0.37	0.29
Average five years observed	0.24	0.28	0.22	0.21	*0.51*	0.32	0.28	0.39	0.31

[1)] Unweigthed average of all eight countries

B.2 Mean absolute (percent) error over the period 1960-2014 and mean error over the period 2005-2014 of the log death rates limited to age 65 and above of the Lee-Carter forecast with three principal components estimated over the period 1960-2014, for eight Western European countries, men and women combined

	FRA	ESP	SWE	BEL	GBR	NOR	FIN	NLD
Mean abs error (1960–2014)	0.020	0.025	0.028	0.030	0.022	0.035	0.042	0.024
Mean abs perc error (1960–2014)	0.91%	1.17%	1.35%	1.50%	1.04%	1.54%	2.19%	1.17%
Mean error (2005–2014)	−0.007	−0.006	−0.012	−0.004	−0.004	−0.011	−0.019	−0.001

Appendix C

In Tables C.1 and C.2 the results of the different choices for the jump-off rates with the Cairns-Black-Dowd (CBD) model (Cairns et al. 2006), a variant of the Lee-Carter model which relies on the linearity of the logit of one-year death probabilities at older ages, are presented. With this model the differences between the outcome of the different choices of the jump-off rates are much more smaller compared to the differences in outcomes for the Lee-Carter model. The fit of this model over the whole fitting period is similar to the fit of the Lee-Carter method, but the mean error in recent years of the fitting period is now positive, which might influence the outcomes as well.

C.1 Mean absolute forecast error (MAFE) of remaining life expectancy at age 65 for the first and fifth year in the forecasting period and standard deviation (SD) of the increase/decrease of the life expectancy at age 65 in 2020 between ten successive forecasts, for six different choices of the jump-off rates applied to the Cairns-Black-Dowd method (ages 65 to 95), for eight Western European countries, men and women combined, fitting periods 1960-2005, 1960-2006, ..., 1960-2014. The lowest SD/MAFE is marked in bold, the highest SD/MAFE in italic

Jump-off rates	FRA	ESP	SWE	BEL	GBR	NOR	FIN	NLD	Av[1)]
	Standard deviation in 2020								
Model values	*0.24*	*0.29*	*0.12*	*0.24*	*0.20*	*0.15*	*0.13*	*0.16*	*0.19*
Last observed values	0.22	0.28	0.11	0.23	0.19	0.15	**0.12**	0.15	0.18
Average two years observed	0.22	0.28	0.11	0.23	0.19	0.15	0.13	**0.15**	0.18
Average three years observed	0.22	0.28	**0.11**	**0.22**	0.19	0.15	0.13	0.15	0.18
Average four years observed	0.22	**0.28**	0.11	0.22	0.19	0.15	0.13	0.15	**0.18**
Average five years observed	**0.22**	0.28	0.11	0.22	**0.19**	**0.15**	0.13	0.15	0.18
	Mean absolute forecast error in the first year of the forecasting period								
Model values	*0.15*	0.17	**0.08**	*0.15*	*0.12*	**0.11**	*0.12*	**0.11**	*0.13*
Last observed values	**0.13**	*0.17*	0.08	0.14	0.12	0.12	**0.10**	0.13	**0.12**
Average two years observed	0.13	0.17	0.08	0.13	0.12	0.12	0.10	0.13	0.12
Average three years observed	0.13	0.17	0.08	0.13	0.12	0.12	0.10	0.13	0.12
Average four years observed	0.13	0.17	0.09	0.13	0.12	0.12	0.10	0.13	0.12
Average five years observed	0.13	**0.17**	*0.09*	**0.13**	**0.12**	*0.12*	0.10	*0.13*	0.12
	Mean absolute forecast error in the fifth year of the forecasting period								
Model values	*0.19*	**0.25**	**0.04**	*0.17*	**0.34**	**0.11**	*0.26*	**0.27**	**0.20**
Last observed values	**0.16**	*0.32*	0.06	0.17	*0.35*	0.18	0.19	*0.28*	*0.21*
Average two years observed	0.16	0.32	0.06	0.17	0.35	0.18	**0.19**	0.28	0.21
Average three years observed	0.16	0.32	*0.06*	**0.17**	0.35	0.18	0.19	0.28	0.21
Average four years observed	0.16	0.31	0.06	0.17	0.35	0.19	0.19	0.28	0.21
Average five years observed	0.17	0.31	0.06	0.17	0.35	*0.19*	0.19	0.28	0.21

[1)] Unweigthed average of all eight countries

C.2 Mean absolute (percent) error over the period 1960-2014 and mean error over the period 2005-2014 of the log death rates limited to age 65 and above of the Cairns-Black-Dowd method (ages 65 and above) estimated over the period 1960-2014, men and women combined

	FRA	ESP	SWE	BEL	GBR	NOR	FIN	NLD
Mean abs error (1960–2014)	0.055	0.039	0.041	0.042	0.033	0.042	0.047	0.038
Mean abs perc error (1960–2014)	2.18%	1.71%	1.84%	2.03%	1.65%	1.94%	2.44%	1.89%
Mean error (2005–2014)	0.025	0.019	0.016	0.022	0.009	0.013	0.019	0.012

Appendix D

D.1 Mean absolute forecast error (MAFE) of remaining life expectancy at age 65 for the first and fifth year in the forecasting period, and standard deviation (SD) of the increase/decrease of the life expectancy at age 65 in 2020 between ten successive forecasts, for six different choices of the jump-off rates applied to a Lee-Carter model, for eight countries, men (a) and women (b), fitting periods 1960-2005, 1960-2006, ..., 1960-2014. The lowest SD/MAFE is marked in bold, the highest SD/MAFE in italic

Jump-off rates	FRA	ESP	SWE	BEL	GBR	NOR	FIN	NLD	AV[1)]
a. Men	**Standard deviation in 2020**								
Model values	0.11	0.11	*0.36*	0.16	*0.16*	*0.26*	*0.52*	0.13	*0.23*
Last observed values	*0.15*	*0.23*	0.10	*0.21*	0.13	0.16	**0.13**	*0.14*	0.16
Average two years observed	0.09	0.08	0.09	0.11	0.09	0.11	0.14	0.09	0.10
Average three years observed	0.08	0.08	**0.07**	0.08	0.08	**0.09**	0.15	0.08	0.09
Average four years observed	0.06	**0.05**	0.08	0.07	0.07	0.11	0.15	0.07	0.08
Average five years observed	**0.05**	0.05	0.09	**0.06**	**0.06**	0.10	0.14	**0.06**	**0.08**
	Mean absolute forecast error in the first year of the forecasting period								
Model values	**0.11**	**0.14**	*0.69*	*0.52*	*0.75*	*0.80*	*0.58*	*1.15*	*0.59*
Last observed values	0.12	0.18	**0.10**	**0.16**	**0.12**	**0.16**	0.10	**0.20**	**0.14**
Average two years observed	0.13	0.17	0.13	0.17	0.16	0.17	0.10	0.25	0.16
Average three years observed	0.15	0.19	0.17	0.19	0.21	0.20	**0.09**	0.34	0.19
Average four years observed	0.19	0.22	0.21	0.22	0.27	0.26	0.12	0.44	0.24
Average five years observed	*0.23*	*0.24*	0.25	0.26	0.34	0.32	0.15	0.53	0.29
	Mean absolute forecast error in the fifth year of the forecasting period								
Model values	**0.21**	**0.35**	*0.93*	*0.69*	*1.04*	*1.25*	*0.84*	*1.60*	*0.86*
Last observed values	0.25	0.42	**0.35**	**0.24**	**0.42**	**0.48**	**0.16**	**0.65**	**0.37**
Average two years observed	0.28	0.45	0.38	0.28	0.45	0.53	0.17	0.74	0.41
Average three years observed	0.33	0.50	0.42	0.33	0.50	0.58	0.20	0.84	0.46
Average four years observed	0.38	0.54	0.47	0.39	0.56	0.65	0.26	0.95	0.52
Average five years observed	*0.43*	*0.57*	0.52	0.44	0.62	0.73	0.31	1.05	0.58
b. Women	**Standard deviation in 2020**								
Model values	0.13	0.14	*0.15*	*0.23*	0.18	*0.40*	*0.62*	*0.26*	*0.26*
Last observed values	*0.21*	*0.24*	0.14	0.19	*0.20*	0.19	0.22	0.15	0.19
Average two years observed	0.13	0.11	0.08	0.11	0.12	0.12	**0.18**	**0.09**	0.12
Average three years observed	0.11	0.09	0.04	0.09	0.11	0.10	0.23	0.09	0.11
Average four years observed	0.10	**0.05**	0.03	**0.06**	0.08	0.09	0.21	0.09	0.09
Average five years observed	**0.08**	0.06	**0.03**	0.07	**0.06**	**0.07**	0.20	0.09	**0.08**

D.1 Mean absolute forecast error (MAFE) of remaining life expectancy at age 65 for the first and fifth year in the forecasting period, and standard deviation (SD) of the increase/decrease of the life expectancy at age 65 in 2020 between ten successive forecasts, for six different choices of the jump-off rates applied to a Lee-Carter model, for eight countries, men (a) and women (b), fitting periods 1960-2005, 1960-2006, ..., 1960-2014. The lowest SD/MAFE is marked in bold, the highest SD/MAFE in italic (continued)

Jump-off rates	FRA	ESP	SWE	BEL	GBR	NOR	FIN	NLD	AV[1)]
	Mean absolute forecast error in the first year of the forecasting period								
Model values	**0.13**	*0.20*	**0.08**	*0.14*	*0.47*	*0.14*	*0.54*	0.15	*0.23*
Last observed values	0.14	0.17	*0.11*	**0.12**	**0.14**	0.13	0.11	0.12	0.13
Average two years observed	0.15	**0.15**	0.10	0.13	0.17	0.11	**0.09**	**0.10**	**0.13**
Average three years observed	0.15	0.15	0.09	0.13	0.19	0.10	0.13	0.14	0.14
Average four years observed	*0.17*	0.16	0.10	0.13	0.21	0.10	0.17	0.17	0.15
Average five years observed	0.16	0.17	0.08	0.13	0.24	**0.09**	0.20	*0.20*	0.16
	Mean absolute forecast error in the fifth year of the forecasting period								
Model values	**0.26**	*0.37*	**0.12**	*0.25*	*0.61*	*0.14*	*0.46*	0.24	*0.31*
Last observed values	0.18	0.26	*0.09*	**0.17**	**0.36**	0.11	0.25	0.21	0.20
Average two years observed	0.21	**0.28**	0.10	0.18	0.38	0.12	**0.26**	**0.23**	**0.22**
Average three years observed	0.19	0.31	0.11	0.18	0.40	0.11	0.27	0.27	0.23
Average four years observed	*0.22*	0.33	0.10	0.20	0.41	0.15	0.28	0.31	0.25
Average five years observed	0.22	0.34	0.10	0.19	0.44	**0.16**	0.29	*0.34*	0.26

[1)] Unweighted average of all countries

D.2 Mean absolute (percent) error over the period 1960-2014 and mean error over the period 2005-2014 of the log death rates limited to age 65 and above of the Lee-Carter forecast estimated over the period 1960-2014, men (a) and women (b)

	FRA	ESP	SWE	BEL	GBR	NOR	FIN	NLD
a. Men								
Mean abs error (1960–2014)	0.027	0.039	0.046	0.049	0.043	0.061	0.064	0.067
Mean abs perc error (1960–2014)	1.39%	1.91%	2.25%	2.61%	2.06%	2.82%	3.66%	3.06%
Mean error (2005–2014)	0.005	0.019	–0.046	–0.028	–0.047	–0.052	–0.045	–0.063
b. Women								
Mean abs error (1960–2014)	0.029	0.042	0.036	0.039	0.035	0.046	0.062	0.040
Mean abs perc error (1960–2014)	1.09%	1.79%	1.57%	1.73%	1.36%	1.89%	2.82%	1.73%
Mean error (2005–2014)	0.018	–0.007	0.004	0.015	–0.034	0.000	–0.055	0.010

Appendix E

In Tables E.1 and E.2 the results of the different choices for the jump-off rates with the Lee-Carter model are presented for the life expectancy at birth (e0). For the most accurate forecast the last observed values as jump-off rates are the best choice, similar as with the life expectancy at age 65 (e65). For Spain the best option of the model values as jump-off rates is more clear than in the results with e65. For France an average of multiple years as jump-off rates is better than the last observed values, but the difference is small. For the most robust forecast an average of multiple years is the best choice for the jump-off rates based on the results of e0, which was also the best choice based on the results of e65.

E.1 Mean absolute forecast error (MAFE) of remaining life expectancy at birth for the first and fifth year in the forecasting period, and standard deviation (SD) of the increase/decrease of the life expectancy at birth in 2020 between ten successive forecasts, for six different choices of the jump-off rates applied to a Lee-Carter model, for eight countries, men and women combined, fitting periods 1960-2005, 1960-2006, ..., 1960-2014. The lowest SD/MAFE is marked in bold, the highest SD/MAFE in italic

Jump-off rates	FRA	ESP	SWE	BEL	GBR	NOR	FIN	NLD	Av[1)]
	Standard deviation in 2020								
Model values	0.16	0.15	*0.36*	*0.25*	*0.22*	*0.37*	*0.71*	*0.25*	*0.31*
Last observed values	*0.20*	*0.25*	0.09	0.22	0.18	0.11	0.25	0.16	0.18
Average two years observed	0.12	0.11	0.08	0.11	0.12	**0.10**	0.19	0.11	0.12
Average three years observed	0.11	0.10	0.07	0.10	0.10	0.11	0.20	0.11	0.11
Average four years observed	0.10	**0.05**	**0.07**	0.07	0.08	0.13	0.20	0.11	0.10
Average five years observed	**0.08**	0.07	0.09	**0.07**	**0.06**	0.12	**0.19**	**0.10**	**0.10**
	Mean absolute forecast error in the first year of the forecasting period								
Model values	*0.21*	**0.17**	*0.38*	**0.08**	*0.46*	*0.38*	*0.42*	*0.47*	*0.32*
Last observed values	0.15	0.19	**0.08**	0.15	**0.13**	**0.11**	0.13	**0.16**	**0.14**
Average two years observed	0.14	0.19	0.08	0.15	0.15	0.12	**0.09**	0.20	0.14
Average three years observed	**0.14**	0.21	0.10	0.15	0.19	0.16	0.10	0.26	0.16
Average four years observed	0.15	0.24	0.11	0.16	0.24	0.22	0.12	0.33	0.20
Average five years observed	0.19	*0.27*	0.13	*0.17*	0.29	0.27	0.13	0.42	0.23
	Mean absolute forecast error in the fifth year of the forecasting period								
Model values	0.24	**0.37**	*0.49*	0.18	*0.78*	*0.74*	*0.62*	0.70	*0.51*
Last observed values	0.17	0.48	**0.16**	**0.16**	**0.50**	**0.38**	**0.25**	**0.39**	**0.31**
Average two years observed	**0.16**	0.52	0.16	0.17	0.51	0.40	0.25	0.44	0.33
Average three years observed	0.18	0.58	0.19	0.19	0.52	0.45	0.26	0.52	0.36
Average four years observed	0.22	0.61	0.22	0.21	0.54	0.52	0.28	0.62	0.40
Average five years observed	*0.25*	*0.65*	0.25	*0.24*	0.57	0.60	0.31	*0.71*	0.45

[1)] Unweighted average of all countries

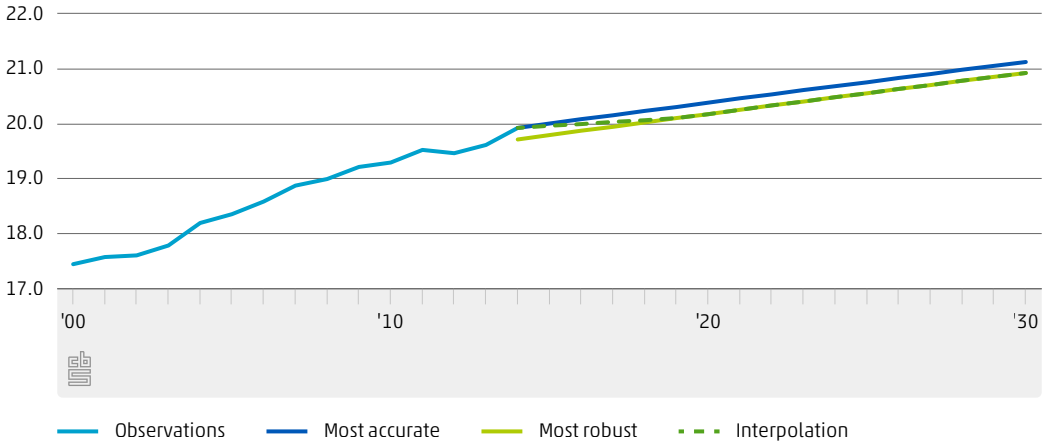
E.2 Mean absolute (percent) error over the period 1960-2014 and mean error over the period 2005-2014 of the log death rates of the Lee-Carter forecast estimated over the period 1960-2014

	FRA	ESP	SWE	BEL	GBR	NOR	FIN	NLD
Mean abs error (1960–2014)	0.054	0.076	0.087	0.071	0.051	0.097	0.094	0.060
Mean abs perc error (1960–2014)	1.07%	1.49%	1.53%	1.39%	1.08%	1.70%	1.88%	1.22%
Mean error (2005–2014)	−0.005	−0.010	−0.010	−0.010	−0.007	−0.016	−0.010	−0.015

Appendix F

In Figure F.1 an example of a forecast using interpolation is shown for the Netherlands. It starts with a forecast using a jump-off rate that gives the most accurate forecast in the first year (red line). Subsequently, a forecast using jump-off rates that is most robust in the fifth year of the forecast period is made (green line). Between the two forecast each year more weight is given to the most robust forecast until the weight is equal to 1 from the fifth year of the forecast onwards (dashed purple line).

F.1 Example of forecasts with interpolation from the forecast with the option for the jump-off rates which is considered the most accurate to the forecast with the option for the jump-off rates which is considered the most robust, the Netherlands, men and women combined, Lee-Carter model, fitting period 1960-2014



6.

Conclusion and discussion

6.1 Introduction and research questions

This study aimed to evaluate mortality forecasting in the context of non-linear mortality trends. Particularly in populations among whom the past trends have been non-linear (like in the Netherlands), the use of an objective extrapolative mortality forecasting method will be more problematic: the level of forecasted mortality could differ greatly depending on the fitting period, and mortality forecasts for individual countries might result in unrealistically divergent outcomes between countries.

Among the potential approaches for improving mortality forecasts when the trends are non-linear trends are making explicit adjustments for the distorting effects of smoking on mortality trends, and using the more linear trends of other countries as the underlying long-term mortality trend. However, both of these approaches require the inclusion of more subjective information in the mortality forecast. Thus, there is an important debate about whether only "objective" extrapolation methods should be employed even in cases of non-linearity, or whether it is preferable to include additional information, even if doing so introduces additional subjectivity. To address this question, mortality forecasting approaches must be evaluated in the context of non-linear past mortality trends. Most previous studies on this topic were purely quantitative evaluations of mortality forecasting models that focused solely on their accuracy, or they evaluated purely objective forecasting approaches that are less relevant for non-linear trends. Moreover, most of these studies did not evaluate the sensitivity of future mortality to explicit assumptions; i.e., to the specific choices that are explicitly stated in a method, such as the choices of the length of the fitting period and of the jump-off rates. This PhD thesis included these important elements. Furthermore, the findings of this research can be used to evaluate, validate, and further improve the mortality forecasts of Statistics Netherlands, which take into account the mortality trends in other Western European countries, and which systematically include in the calculation information about developments in smoking, following the approach by Janssen et al. (2013).

This study was guided by the following research questions:

1) In a context in which mortality trends are non-linear, how does the choice of the mortality forecasting method and the explicit assumptions affect future forecasted mortality
2) How can future levels of smoking-attributable mortality be formally estimated?

3) Which model should be used when the goal is to forecast mortality coherently , namely by taking into account the mortality experiences of other countries?
4) How can mortality forecasts be adjusted to take into account more recently observed data?

In the remainder of this concluding chapter, summaries of both the overall results and the results by chapter are provided (6.2). Reflections on the main findings are then offered (6.3). Next, the implications of the results for mortality forecasting and for the official mortality forecasts of Statistics Netherlands are discussed (6.4), and reflections on the approach are provided (6.5). Finally, recommendations for further research on mortality forecasting and for users of mortality forecasts are made (6.6).

6.2 Summary of the findings

6.2.1 Summary of the results by chapter

Chapter 2 reviewed the different mortality forecasting methods and their assumptions in Europe, and assessed their impact on projections of future life expectancy for the Netherlands. More specifically, (i) the current methods used in official mortality forecasts in Europe were reviewed; (ii) the outcomes and the assumptions of different projection methods within the Netherlands were compared; and (iii) the outcomes of different types of methods for the Netherlands using similar explicit assumptions, including the same historical period, were compared. The findings of a review of the current methods indicated that most statistical offices in Europe use simple linear extrapolation methods, but that countries with less linear trends employ other approaches or different assumptions. The approaches employed in the Netherlands include the use of explanatory models, the separate projection of smoking- and non-smoking-related mortality, and the projection of the age profile of mortality. There are, however, clear differences in the explicit assumptions used in these approaches, and the resulting e0 in 2050 varies by approximately six years. Using the same historical period (1970-2009) and the observed jump-off rates, the findings generated by different methods result in a range of 2.1 years for women and of 1.8 years for men. For e65, the range is 1.4 years for men and 1.9 years for women. These findings suggest that the choice of explicit assumptions is more important than the choice of the forecasting method.

In Chapter 3, a formal estimation of future levels of smoking-attributable mortality up to 2050 was proposed for the total national populations of England and Wales, Denmark, and the Netherlands. An update and an extension of the descriptive smoking epidemic model were provided in the estimation. A two-step method for estimating the future smoking-attributable mortality fraction was presented: (i) lung cancer mortality was projected by extrapolating age-period-cohort trends (1950-2009), while using the observed convergence among men and women of smoking prevalence and past lung cancer mortality levels as input; and (ii) other causes of death attributable to smoking were added by applying a simplified version of the indirect Peto–Lopez method to the projected levels of lung cancer mortality. The smoking-attributable mortality fractions (SAF) for men in 2009 were found to be 19% (44,872 deaths) in England and Wales, 22% (5,861 deaths) in Denmark, and 25% (16,385 deaths) in the Netherlands. In the projections, these fractions declined to 6%, 12%, and 14%, respectively, in 2050. The SAF for women peaked at 14% (38,883 deaths) in 2008 in England and Wales, and is expected to peak in 2028 in Denmark (22%) and in 2033 in the Netherlands (23%). By 2050, declines to 9%, 17%, and 19%, respectively, are foreseen. The use of different indirect methods for estimating the SAF in 2050 yielded ranges of 1–8% in England and Wales, 8–13% in Denmark, and 11–16% in the Netherlands for men; and of 7–16%, 12–26%, and 13–31%, respectively, for women.

In Chapter 4, different coherent forecasting methods were evaluated in terms of their accuracy (fit to historical data), robustness (stability across different fitting periods), subjectivity (sensitivity to the choice of the group of countries), and plausible outcomes (smooth continuation of trends from the fitting period). The coherent forecasting methods investigated in this chapter were as follows: the co-integrated Lee-Carter (CLC) method, the Li-Lee (LL) method, and the coherent functional data (CFD) method. The methods were applied to data from France, Italy, the Netherlands, Norway, Spain, Sweden, and Switzerland in order to generate forecasts up to 2050; and the results were compared to those of the individual Lee-Carter (LC) method. Of the three coherent forecasting methods evaluated, the CFD method was found to perform best on the accuracy measures. However, after the CFD method's higher number of parameters was controlled for, the differences disappeared. Both the CLC and the LL methods were found to be robust. The CLC method (for women) and the LL method (for men) were shown to be the least sensitive to the choice of the group of countries. The LL method generated the most plausible results, as it showed a convergence of future life expectancy levels that was in line with the fitting period and the smooth pattern of age-specific improvements. This finding could imply that the LL method, which performed best in terms of robustness, subjectivity, and plausibility, provided a better fit than the CFD method, which had better accuracy (model fit).

Finally, in Chapter 5, six different options for the jump-off rates were evaluated and their effects on the robustness and the accuracy of the mortality forecasts were examined. As the jump-off rates, the use of the model values, the observed values in the last year, and the averaged over the last couple of years are examined for data from eight European countries (Belgium, Finland, France, the Netherlands, Norway, Spain, Sweden, and United Kingdom, 1960-2014 period). The future life expectancy at age 65 was calculated for different fitting periods and jump-off rates using the Lee-Carter model, and the accuracy (mean absolute error) and the robustness (standard deviation of the change in projected e65) of the results were examined. The findings of the analysis showed that which jump-off rates were chosen clearly influenced the accuracy and robustness of the mortality forecast, albeit in different ways. For most of the countries, using the last observed values as the jump-off rates resulted in the most accurate method, due in part to the estimation error of the model in recent years. The most robust method was obtained when using an average of observed years as jump-off rates. The more years that were averaged, the higher the degree of robustness; but the level of accuracy decreased with more years averaged. These results imply that the best strategy for matching mortality forecasts to the most recently observed data depends on the goal of the forecast, the country-specific past mortality trends, and the model fit.

6.2.2 Overall summary of results

For countries with non-linear mortality trends, like the Netherlands, approaches and assumptions were used that differ from the simple linear extrapolation methods that are commonly used by national statistical offices. It was found that the choice of explicit assumptions (i.e., the assumptions that had to be explicitly stated in a method, such as the length of the fitting period and the jump-off rates) proved more important than the choice of the forecasting approach for the mortality forecast. Because the inclusion of additional information on the smoking epidemic or on the mortality experiences of other countries is generally known to diminish the effect of the length of the historical period, doing so is expected to result in a more robust forecast.

One way that additional information on the smoking epidemic could be included was by separately forecasting smoking-attributable mortality. The age-period-cohort methodology – informed by assumptions derived from the smoking epidemic model and a careful study of past trends – proved valid for this purpose.

When the mortality experiences of other countries by means of coherent mortality forecasting were included, it was found that the Li-Lee method (Li and Lee 2005) outperformed the co-integrated Lee-Carter method (Li and Hardy 2011; Cairns et al. 2011a) and the coherent functional data method (Hyndman et al. 2013) in terms of robustness, subjectivity, and plausibility.

Another important explicit assumption was the choice of the jump-off rates; i.e., how mortality forecasts should be matched to the most recently observed data. It was found that which jump-off rates were chosen clearly influenced the accuracy and the robustness of the mortality forecast, albeit in different ways. It was therefore concluded that which strategy was best depended on the goal of the forecast, the country-specific past mortality trends, and the model fit.

All in all, it was found that forecasting mortality when the trends were non-linear involved more than the direct (linear) extrapolation of past mortality trends. Even though including additional information (like data on the smoking epidemic and/or on the mortality experiences of other countries) made the method more subjective, it also made the method less dependent on an important explicit assumption: namely, the historical period. This insight is important, because this PhD thesis has also demonstrated that explicit assumptions play an essential role in mortality forecasts.

6.3 Reflections on the main findings

6.3.1 Importance of explicit assumptions

This PhD thesis found that in the Netherlands, where the past mortality trends are non-linear, the choice of explicit assumptions contributed more to the differences in the estimates of different mortality forecasts than the choice of the forecasting method/approache. Thus, the findings showed that when the same historical period and the same jump-off rates were used in different mortality forecasts, the differences in the life expectancy levels at birth projected for 2050 declined by approximately two-thirds.

This is a novel and important finding. Most of the previous studies that evaluated mortality forecasts focused primarily on the method itself, and only very rarely on which explicit assumptions were chosen (examples are Booth et al. 2002; Bell

1997; Lee and Miller 2001; Janssen and Kunst 2007). This finding is important because explicit assumptions are an essential part of mortality forecasting methods; that is, forecasting methods cannot generate outcomes unless specific assumptions are made. It is important that this key point is acknowledged.

The explicit assumptions also play an important role in the comparability of different mortality forecasting approaches and the related mortality forecasting methods, and of the outcomes from different institutions. Different forecasting approaches/methods are more comparable when the same explicit assumptions are used. Thus, the differences in outcomes reflect the different methods used in forecasts.

Furthermore, the explicit assumptions can have a large effect on the prediction intervals of mortality forecasts. Prediction intervals do not capture the level of uncertainty of the forecasts depending on which explicit assumptions are chosen; i.e., they are actually conditional intervals based on the assumptions. Because the explicit assumptions clearly contribute to the differences in mortality forecasts, they also contribute to the level of uncertainty of mortality forecasts. Ideally, in order to paint a more complete picture of the role of uncertainty, the prediction intervals would also include the level of uncertainty due to the explicit assumptions.

6.3.2 Inclusion of additional information

There is an important debate in the mortality forecasting literature about whether only "objective" extrapolation methods should be employed even in cases of non-linearity, or whether it is preferable to include additional information – e.g., data on trends in other countries and/or epidemiological information on smoking – even if doing so introduces additional subjectivity. The results of the analyses presented in this PhD thesis contribute to this debate. First, the literature review in Chapter 2 showed that the statistical offices in countries with non-linear past mortality trends often use more subjective methods that take into account the non-linearity observed in the past mortality trends, rather than the simple linear extrapolation methods typically used by national statistical offices in countries with more linear trends. These more subjective forecasting methods usually rely on a very short fitting period, a quadratic age effect, or epidemiological information. If, however, the past mortality trends have been largely linear, simple extrapolation methods will suffice, especially given that the outcomes of different extrapolation methods using the same explicit assumptions do not differ greatly.

Chapter 2 also revealed that once the effect of explicit assumptions was controlled for, the remaining differences in the outcomes mainly reflect differences between

the methods that include additional information to account for the observed non-linearity, and the extrapolation methods that do not. When applied to the Netherlands based on the fitting period 1970-2009, the methods that included additional information – through either age-period-cohort modelling or the inclusion of data on smoking and/or other countries – resulted in higher future life expectancy estimates and less linear future trends than the Lee-Carter method and linear extrapolation.

Both observations (higher future life expectancy estimates and less linear future trends) can be linked to the non-linearity observed in the past trends in the Netherlands, and to the main determinant of this non-linearity: the smoking epidemic. In the Netherlands, the impact of the smoking epidemic resulted in an overall mortality trend that was less optimistic than the trend in non-smoking-attributable mortality. But when the continuing decline in smoking prevalence (and, hence, in smoking-attributable mortality) was taken into account, the estimates of future life expectancy were higher (see also "6.3.3 Methodology for forecasting smoking-attributable mortality" below). The less linear future trend found among women was the result of a projected increase in smoking-attributable mortality, followed by a decline. Such a non-linear future pattern does not arise when using the Lee-Carter and the linear extrapolation methods, because these methods extrapolate the average increase in all-cause mortality over the whole period into the future, which results in a straight-line projection.

Because the recent mortality trends in the Netherlands (1970-2009) have been less positive than the average trends in certain other countries, a method that includes these other countries will also result in a higher future life expectancy than a method that does not include these other countries. This was also shown in Chapter 4: the countries that had less positive past mortality trends than those of the main group of countries used in the coherent forecasting model had higher future life expectancy levels than when the Lee-Carter method was applied, and vice versa (see also "6.3.4 Inclusion of the mortality experiences in other countries" below).

6.3.3 Methodology for forecasting smoking-attributable mortality

The inclusion of information on the smoking epidemic can add non-linearity to the trend; and, depending on the phase in the smoking epidemic, can lead not only to higher estimated life expectancy outcomes, but to a more robust forecast. Thus, the inclusion of additional information on the smoking epidemic may be expected to

diminish the effect of the length of the historical period (Janssen and Kunst 2007). Janssen et al. (2013) clearly demonstrated for the Netherlands that past trends in non-smoking-attributable mortality were more linear over time than past trends in all-cause mortality. Furthermore, Janssen and Kunst (2007) demonstrated that because past declines in non-smoking-attributable mortality were more similar across countries and between the sexes than declines in all-cause mortality, including the former information in a forecast can lead to more comparable outcomes between countries and between men and women than relying on the latter data alone. Therefore, providing a separate forecast of smoking-attributable mortality clearly has value when forecasting mortality in a context in which mortality trends are non-linear, such as the Netherlands.

In Chapter 3, a method for forecasting smoking-attributable mortality was introduced and validated that, unlike previous methods, is capable of forecasting the long-term future. In contrast to the methodologies used in earlier studies, the methodology takes into account the expectation that among women, future smoking-attributable mortality will increase, and then decrease. This expectation is based on the smoking epidemic model described by Lopez et al. (1994), in which the wave pattern in smoking prevalence was followed 30-40 years later by a similar wave pattern in smoking-attributable mortality, first for men and then for women. In addition, the trends in smoking prevalence and smoking-attributable mortality for the three examined countries reported in Chapter 3 clearly show that smoking-attributable mortality is already declining for women for the youngest age groups.

The age-period-cohort methodology developed in this PhD thesis was guided by the smoking epidemic model and by past trends in both smoking prevalence and smoking-attributable mortality. This methodology was shown to be valid for forecasting lung cancer mortality and, subsequently, smoking-attributable mortality. For example, when the methodology was applied to some of the data for England and Wales (1950–99), where smoking-attributable mortality among women peaked in 2008, it was found that the assumptions and methodology were able to predict the observed maximum in 2008 for women. This finding justifies the use of the trends in and the levels of lung cancer mortality for men to determine the maximum for women.

It is projected that the peak in smoking-attributable mortality will be reached in 2033 for Dutch women and in 2028 for Danish women, and that smoking-attributable mortality for these groups will decrease thereafter. Including these irregular trends in the forecast of total mortality will add non-linearity to the projected trend in mortality and, consequently, in life expectancy.

This methodology can be applied to other countries as well. For countries that are well into the fourth stage of the smoking epidemic (in which male mortality from smoking has peaked a few decades ago and smoking prevalence has been slowly declining for both men and women), the method can be easily applied. For countries where male mortality from smoking has peaked more recently, information about the forerunners would be needed to complement the methodology. For countries at an even earlier stage of the smoking epidemic, detailed information on smoking prevalence would be necessary as well.

6.3.4 Inclusion of the mortality experiences in other countries

Another way of introducing additional information into the mortality forecast is to include the mortality experiences of other countries. An approach that is often used to take into account the experiences of other countries is coherent forecasting, of which the best-known methods are the co-integrated Lee-Carter method (Li and Hardy 2011; Cairns et al. 2011a), the Li-Lee method (Li and Lee 2005), and the coherent functional data method (Hyndman et al. 2013).

The results in this PhD thesis showed that the Li-Lee (LL) coherent forecasting method performed better than the co-integrated Lee-Carter (CLC) method and the coherent functional data (CFD) method in terms of robustness, subjectivity, and plausibility (Chapter 4). Specifically, it was found that the LL method – when estimated using singular value decomposition – generated stable outcomes across different fitting periods; that the LL method (for men) was the least sensitivity to the choice of the group of countries; and that the LL method resulted in a convergence of future life expectancy trends that was in line with the fitting period and the smooth pattern of age-specific improvements. The high degree of stability observed across fitting periods can be explained by the equal weight the LL method placed on all data in the sample. This aspect of the LL method tends to diminish the dependence on new data being added, which can be higher when more weight is placed on recent data (such as in the CFD method). Because the LL method is less sensitive to the choice of the group of countries, it is less likely to result in convergence, particularly in comparison to the CFD method. Although the LL method scored lower than the CFD method on some accuracy measures, this difference in the degree of accuracy proved negligible when the number of model parameters was accounted for in the comparison.

Whereas projections for separate countries based on the past mortality trends in each individual country will lead almost inevitably to divergence (Lundström 2003; Giannakouris 2004; Li and Lee 2005; Janssen and Kunst 2007), including information from other countries in the forecasting method/approach will prevent this from happening. Furthermore, by using the experiences of other countries, a broader empirical basis can be created for the identification of the most likely long-term trends, which can improve the robustness of the forecast.

In coherent mortality forecasting, an important explicit assumption that should be taken into account is the choice of the main group of countries that will be included in the model. Which main group is chosen determines the long-term trends for a specific country in the coherent mortality forecast. In Chapter 4, three different main groups were compared. The results showed that the coherent forecasting methods were sensitive to the choice of the group of countries. Among the important criteria for the selection of the main group were the linearity of the mortality trend of the total group (extrapolative methods perform better if the trends are linear) and the similarities (political, economic, health care) between the countries in the group and the country for which the forecast is being made.

6.3.5 Choice of the jump-off rates

In addition to the choice of the historical period (the effect of which is diminished by including additional information on the smoking epidemic and/or the mortality experiences of other countries) and of the main group of countries, another important explicit assumption in coherent mortality forecasting is the choice of the jump-off rates. The aim of the jump-off rates is matching the mortality forecast to the most recently observed data. The main problem encountered in the majority of forecasting methods is the appearance of a discontinuity between the observed and the predicted trends, which leads to a jump-off that is usually considered implausible. This is a practical problem more than it is a theoretical one, as it has a large impact on the outcomes of the forecasting methods. This issue has not been the main topic of any previous research article, and has not been adequately addressed in the scientific literature.

The results of the analysis (Chapter 5) showed that which jump-off rates were chosen clearly influenced the accuracy and the robustness of the mortality forecast. For most countries, using the last observed values as the jump-off rates resulted in the most accurate method, which was related to the estimation error of the model in recent years. That is, if the model had been underestimated, the last observed values would have already been closer to the future values than the model values,

which would have automatically favoured the last observed values as jump-off rates. The most robust method is generally obtained by using an average of the observed years as jump-off rates, as this approach can smooth out fluctuations in life expectancy. That is, if the last observed values are used as jump-off rates, life expectancy will fluctuate when recent data were added; whereas if an average of these values is used, these fluctuations will diminish.

However, the use of different strategies can affect the robustness and the accuracy of the forecasts differently. For example, the more years with observed values that are averaged, the greater the robustness, but the lower the degree of accuracy. Thus, in determining which strategy is best, it is important to take into account not just the model fit and the country-specific past mortality trends, but the goal of the forecast. If the goal of the forecast is robustness, using an average of the observed years as jump-off rates (the most robust approach for most countries) may be preferable to using the last observed values as jump-off rates (the most accurate approach for most countries).

6.4 Implications

6.4.1 Implications of the results for mortality forecasting

The results of this PhD thesis have a number of implications for mortality forecasting.

When past trends in mortality are non-linear, adding more information could have value; as the added information could reveal the true underlying trend in mortality, and could thus provide a solid basis for the mortality forecast. However, before any information is added to mortality forecasting models, a careful examination of past trends should be undertaken, and a careful assessment of the pros and cons of its inclusion should be performed. If more information is included in a model, more assumptions need to be made, which increases the subjectivity of the forecast. Therefore, the decision to add information should not be taken lightly. A key challenge that can arise when using methods that include additional information, such as a cohort effect or epidemiological information, is that the additional information can be hard to predict. The advantage of using additional information

in the forecasting method diminishes if the additional information cannot be forecast more accurately than mortality itself. It should, however, be possible to strike the right balance between the risks associated with including additional information in a forecasting method and the risks associated with increased subjectivity.

This PhD thesis showed the important role explicit assumptions can play in mortality forecasting. It is therefore essential that the explicit assumptions used in the forecasting method are selected carefully when making a forecast or when choosing a new forecasting method or approach. Choosing the right explicit assumptions can improve the accuracy and the robustness of the mortality forecast, but which assumptions are selected is likely to differ depending on the forecasting method/approach and the forecasting goal. It should be noted that currently, prediction intervals do not provide information about the levels of uncertainty associated with these explicit assumptions. The results presented in this PhD thesis strongly suggest that stochastic forecasts should also incorporate the levels of uncertainty associated with different explicit assumptions in order to provide a fuller picture of the degree of uncertainty. Currently, the methodology that would allow us to do so is not yet well developed.

New forecasting methods should be evaluated based not only on their accuracy, but on other more qualitative criteria. This PhD thesis showed, for example, that which coherent forecasting method is chosen can differ depending on whether the methods are evaluated solely on their accuracy, or also on the robustness, subjectivity, and plausibility of their outcomes. By adopting different evaluation criteria – both more quantitative (accuracy) and qualitative (the robustness, subjectivity, and plausibility of the results) – this PhD thesis was able to demonstrate that the best method might not be the most accurate method. Judging an approach or model using one type of criteria only will clearly not provide the full story.

The most appropriate method can differ depending on the forecasting application/goal, and the value assigned to quantitative versus qualitative criteria. For instance, for forecasts that are updated regularly, robustness should be given higher priority. It is therefore advisable to keep the forecasting application/goal in mind when choosing the method, and to explicitly mention the forecasting application/goal when reporting on it.

In addition, it is essential to remain flexible when forecasting mortality. Both mortality trends and their determinants are constantly changing, as is our

knowledge of them. Moreover, new forecasting methodologies are constantly being developed. Mortality forecasting can thus be described as "a work in progress"; and remind ourselves of the need to keep learning from new developments.

6.4.2 Implications for the official mortality forecasts in the Netherlands

This PhD thesis examined in detail the different components of the new mortality forecasting approach adopted by Statistics Netherlands in 2012 (Stoeldraijer et al. 2012, see the Appendix). This new mortality forecasting approach was developed by Janssen and Kunst (2010) and Janssen et al. (2013). The approach made use of extrapolation, but included additional information on trends in other countries in Western Europe, and took into account the clear non-linear pattern in smoking-attributable mortality. It combined the separate forecast of smoking-attributable mortality with the coherent forecast of non-smoking-attributable mortality.

The method used by Statistics Netherlands differs from the method presented in Janssen and Kunst (2010) and Janssen et al. (2013) in two main ways: the methods use different jump-off rates and different approaches to forecasting smoking-attributable mortality based on lung cancer mortality.

Janssen and Kunst (2010) and Janssen et al. (2013) used the projected rates rather than the observed rates as the jump-off rates. The focus in Janssen et al. (2013) was on differences between the gains in life expectancy and the projected life expectancy in the jump-off year, which are not affected by the choice of jump-off rates. For Statistics Netherlands, it was important that the forecast was perfectly aligned with the last observation (i.e., had no jump-off bias). Thus, Statistics Netherlands used the observed rates in the last year as the jump-off rates.

The most important difference in the forecasting methods used for smoking-attributable mortality is that Janssen and Kunst (2010) and Janssen et al. (2013) used an age-period-cohort method applied to lung cancer mortality to estimate the year in which the smoking-attributable mortality fraction will reach its maximum for females (by adding the average age at dying from lung cancer to the cohort with the highest lung cancer mortality), but subsequently used the smoking-attributable mortality fractions to forecast smoking-attributable mortality; whereas Statistics Netherlands projected levels of lung cancer mortality directly via the age-period-cohort method, and used the projected lung cancer mortality rates to

calculate projected smoking-attributable mortality by applying an indirect method to estimate smoking-attributable mortality.

Another difference in the approaches used for estimating future smoking-attributable mortality lies in the indirect method used to estimate smoking-attributable mortality from lung cancer mortality. Janssen and Kunst (2010) and Janssen et al. (2013) used the simplified Peto-Lopez method (Bonneux et al. 2003; Peto et al. 1992) for this purpose; whereas Statistics Netherlands, after comparing different indirect estimation methods, chose to use the indirect estimation method of Rostron (2010) instead. Both methods use lung cancer death rates as an indicator of the damage caused by smoking. Whereas the Peto-Lopez method merely uses epidemiological information from the American Cancer Study (lung cancer death rates among smokers and non-smokers, relative risks of dying from smoking); the method more recently developed by Rostron (2010) uses – instead of the relative risks – a regression model to predict mortality from causes other than lung cancer as a function of lung cancer mortality and other variables (dummy variables for age, year and country, and interaction terms), using data from several low-mortality countries (many of which are in Western Europe, see also Preston et al. 2010).

A third and smaller difference between the methods employed by Statistics Netherlands and Janssen and Kunst (2010) and Janssen et al. (2013) is that Statistics Netherlands uses the total population of Germany instead of the population of West Germany in the group of countries used in the coherent forecasting for the non-smoking-attributable mortality.

As a result of the research within this PhD thesis, several components of the mortality forecasting approach of Statistics Netherlands were closely evaluated, validated, and – if necessary – improved.

The projection of smoking-attributable mortality by means of the age-period-cohort model applied to lung cancer mortality was validated (Chapter 2) using in-sample forecasting, as well as data for Denmark and England and Wales. That is, the observed maximum of smoking-attributable mortality for women in 2008 was correctly estimated by using a portion of the data for England and Wales.

The indirect estimation method used to estimate smoking-attributable mortality was validated by comparing five different methods (Chapter 2). It was found that the regression-based method of Rostron (2010) was very similar to the simplified Peto-Lopez method (Bonneux et al. 2003; Peto et al. 1992), and thus concluded that both methods are valid. Because the regression-based method by Rostron

(2010) has a stronger empirical base (compared to the simplified Peto-Lopez method, it has data from more countries and data that are more recent) and uses recent new estimation techniques that were introduced in the field, Statistics Netherlands continues to use this method.

In the original approach by Janssen and Kunst (2010), Janssen et al. (2013), and Statistics Netherlands, the Li-Lee method (Li and Lee 2005) was used as the coherent forecasting method for the non-smoking-attributable mortality projection because it was at that point in time (2009/2010) the most known coherent mortality forecasting technique (Hyndman et al. 2013). The method was easy to understand and easy to apply. In this PhD research, the use of this method rather other more recently developed coherent mortality forecasting methods was assessed (Chapter 4, based on all-cause mortality). It was found that compared to two other coherent forecasting methods (CFD, Hyndman et al. 2013; CLC, Li and Hardy 2011, Cairns et al. 2011a), the Li-Lee method performed just as well in terms of accuracy and better in terms of the robustness, subjectivity, and plausibility of the outcomes. These findings validated the use of the Li-Lee method over the other methods, and provided a stronger empirical basis for the use of the Li-Lee method by Statistics Netherlands.

The results presented in Chapter 4 on the group of countries that is used in the coherent forecasting method did not lead to a modification of the group of countries used by Statistics Netherlands in their forecasting method. The group of countries used in the coherent forecasting method of Statistics Netherlands consists of countries in Western Europe that had similar trends in the past. Moreover, the mortality trend of the group is relatively linear, which is in line with previous recommendation in this PhD thesis (see 6.3, "Reflections on the main findings").

In the original application of the method by Statistics Netherlands, the last observed years were used as the jump-off rates. However, in this PhD research, the strong effect of explicit assumptions, like the jump-off rate, led to a more detailed appraisal of the choice of the jump-off rates. As a result of this finding, the average of the mortality rates in three recent observed years are used as the jump-of rates instead of the rates in the last observed year (Van Duin and Stoeldraijer 2014). To ensure there was no jump-off bias, the first five years of the forecast were also adjusted: i.e., there was an interpolation between a forecast with jump-off rates equal to the last observed rates to a forecast with jump-off rates equal to the average of the three recent observed years. The interpolation was suggested in Chapter 5. Both the accuracy and the robustness of the mortality forecast was improved by this modification.

More generally, the findings of this PhD research demonstrate how important it is that the mortality forecasts of Statistics Netherlands are adjusted in response to scientific developments and recent mortality trends, not only in the Netherlands, but in surrounding countries as well. Therefore, it is critical that Statistics Netherlands communicates with other statistical offices in Europe (through Eurostat) and other bureaus in the Netherlands that make mortality forecasts about their problems and the potential solutions to these problems, as well as about recent developments in research and methods. Furthermore, over the course of working on this PhD thesis, the need to better explain what forecasts are became clear, and to publish prediction intervals to inform users of the uncertainty surrounding forecasts.

All in all, the mortality forecasts of Statistics Netherlands have become more evidence-based.

6.5 Reflections on the approach

The approach used in this research was data-driven and had a strong empirical basis that relied heavily on the careful study of past trends. By investigating how different mortality trends (especially linear versus non-linear trends) were affecting the performance of different mortality forecasting methods, both quantitatively and qualitatively, important new insights on mortality forecasting in the context of non-linear mortality trends were obtained.

The focus of the PhD thesis was on Western Europe, and, more specifically, on the Netherlands. It is therefore possible that the results of the thesis might not apply to other countries with non-linear trends, such as countries in Eastern Europe and other high-mortality countries. These countries have very different past mortality trends than the Western European countries. For instance, a key reason why the Eastern European countries have very different past mortality trends is that they experienced a health crisis from 1975 onwards (McKee and Shkolnikov 2001; Vallin and Meslé 2004; Leon 2011). When forecasting mortality for these countries, extrapolation methods are not suitable because of the clear non-linear trends and the breaks in the trends, and because the non-linearity might be caused by factors other than smoking (alone). However, the approach used for these countries can be very similar: namely, the past mortality trends should be studied carefully; and the non-linear trends should be filtered out from the general trend, which can be

captured using extrapolation; attention to the explicit assumptions; and evaluation based on both quantitative and qualitative criteria.

In this PhD thesis, it was assumed that the data were of good quality, because on average, developed countries have the resources to collect and maintain extensive records of mortality and population data (Mathers et al. 2005). However, a different assessment of the forecasting approach might be made if this was not the case, or if the data did not satisfy the needs of the mortality forecast, or was preliminary in nature. Finally, forecasting might be improved by weighting or smoothing the data.

As the outcome measure of the predictive ability of the mortality forecast, the focus in this PhD thesis was primarily on life expectancy, both at birth and at age 65. These parameters were useful for the criteria that were evaluated. Looking at future life expectancy provided information about the forecasted expected mean age at death. Some of the more novel outcome measures used in the field of mortality are the modal age at death – i.e., the age at which most of the deaths are occurring – and the variability of the age at death around the modal age (Canudas-Romo 2008). The performance of mortality methods can be evaluated more comprehensively by analysing not only the mean age at death (life expectancy), but the modal age and the variability of the age at death (Bohk-Ewald et al. 2017).

The evaluation approach of this PhD thesis was extensive, and comprised (i) an evaluation of not just different mortality forecasting methods, but different forecasting approaches; (ii) an evaluation of both quantitative and qualitative criteria; (iii) the assessment of the sensitivity of future mortality to different explicit assumptions (e.g., historical period, jump-off rates); and (iv) the evaluation of different elements of a mortality forecasting approach that deals with non-linear past mortality trends (e.g., the forecasting of mortality attributed to smoking, a model for coherently forecasting mortality). The use of this approach has led to important new insights, as was discussed in the previous sections.

Although this PhD thesis covered many different aspects of mortality forecasting, much more research on this topic is possible, as the list of approaches, methods, evaluation criteria, and explicit assumptions which were compared is by no means exhaustive. This PhD focused on models based on extrapolation, which are the most frequently used, and which are more objective than models based on expectation or explanation. Furthermore, the focus was limited to models based on death rates. More recently, other models that can be used to study mortality have been developed, such as Bayesian models (in which prior knowledge and various sources of uncertainty can be included, Czado et al. 2002; Pedroza 2006) or models using mortality delay (with a shift in the age-at-death distribution towards older

ages, Janssen and de Beer 2016; Basellini et al. 2016; de Beer et al. 2017). In addition, while the forecasts were evaluated using different criteria, such as the accuracy, robustness, and plausibility of the results; it should be emphasised that these criteria represent only a selection of all the criteria that might be applied (Cairns et al. 2011). Among the explicit assumptions that could be added are whether and, if so, how mortality can be projected up to higher ages (de Beer et al. 2017).

Despite its limitations, the evaluation in this PhD thesis resulted not only in the evaluation, validation, and further improvement of the mortality forecasts of Statistics Netherlands, but contributed to the scientific literature and to research on mortality forecasting in general.

6.6 Recommendations

6.6.1 Recommendations for further research on mortality forecasting

In connection with the evaluation of the approach (6.5), the following recommendations for further research on mortality forecasting are offered.

To obtain a fuller picture of the evaluation of mortality forecasting in contexts with non-linear past mortality trends, the patterns in Eastern European countries should be evaluated as well. Most countries in this region have past mortality trends that differ from those of Western European countries. The causes of the non-linearity might be different for these countries than for their Western counterparts. For instance, after the fall of the Berlin Wall, the morality trends in these countries changed suddenly, and the high levels of alcohol consumption among large parts of the population have had a clear impact on mortality in the Eastern European countries (Trias-Llimos at al. 2017). These causes and other potential sources of non-linearity should be investigated, and, if possible, incorporated into the forecasting method. If the cause of the non-linearity is purely a period effect or a break in the trend, the consequences for the method are different from those for the approach used here for smoking.

In Western Europe, the main cause of the (measurable) non-linearity in past trends has been smoking (Janssen et al. 2007; Janssen et al. 2013; Lindahl-Jacobsen et al.

2016). At the moment, large shares of the population in Western Europe are obese, which might influence future mortality trends (Vidra et al. 2018). It may be necessary to revise mortality forecasts in response to these changes in trends. Thus, it is important that past mortality trends are studied continuously, not just in the country of interest, but in other countries as well.

The findings presented in this PhD thesis call for future studies to focus on a wider range of mortality forecast outcome measures. The focus on life expectancy in this PhD thesis was sufficient to address the research questions, but further research might explore other measures (such as the variability of the age at death, Bohk-Ewald et al. 2017), not only in order to evaluate the mortality forecasts more comprehensively, but to improve upon the methods themselves.

While the focus in this PhD thesis was on national populations, there are also differences within these populations that are associated with mortality differences, such as differences in educational attainment, migration background, income, and type of employment. While important advances in mortality forecasting have been made (Janssen, forthcoming: GENUS thematic issue), mortality forecasts that are disaggregated beyond age, sex, and region are almost non-existent (Samir et al. 2010; van Baal et al. 2016, Villegas and Haberman 2014). Developing such forecasts would be an important way forward in mortality forecasting, as policies can be better targeted to specific groups if the differences between them are known.

In the course of meeting the two main goals of this PhD thesis (i.e., contributing to the debate on objective versus subjective mortality forecasting methods and further improving the mortality forecasts of Statistics Netherlands), the importance of developing a closer relationship between the professional and the academic worlds became apparent. The approaches to mortality forecasting used in academia differ greatly from the approaches used in practice, and the two worlds could learn from each other. For instance, in practice it is important that a method/approach is understandable and reproducible, and the academic world can do more to support these aims. A collaboration between the various institutes and disciplines involved in mortality forecasting is also recommended, as fields such as demography and actuarial sciences employ different approaches, but have similar goals.

6.6.2 Recommendations for the users of mortality forecasts

Mortality forecasts have many users, as mortality rates affect many aspects of society (e.g., Currie et al. 2004). The most widespread users of mortality forecasts are the government (health care and public retirement), planning bureaus (population projections), and actuarial companies (life insurance and annuities). Examples of aspects of society that are affected by mortality rates (Bengtsson and Christensen (Eds.) 2006) are the cost of old-age income support in social security systems, public retirement policies, the financial position of defined benefit pension funds, the solvency requirements of life insurers, the pricing and reserves of other mortality-linked products, the planning and resources of social welfare programs, industries like care services for the elderly, and life course planning for individuals. Planning in all these areas requires institutions and individuals to understand and be knowledgeable about the present and the forecasted rates of mortality.

The new mortality forecasting methodology that was implemented by Statistics Netherlands for the first time in 2012 resulted in higher long-term life expectancy, added non-linearity, and more robust outcomes (fewer changes between consecutive forecasts). It is essential that users are aware of the implications of the impact this new methodology has. For example, if long-term life expectancy is projected to be higher than it was in previous forecasts, users might conclude that the reserves for mortality-linked products or payments should be higher for a longer period of time, or be delayed to a later date. For instance, retirement benefits would have to be paid over a longer period of time if people are expected to live longer. If mortality forecasts become more robust, users will have to make fewer adjustments to the estimates they rely on. For instance, plans to increase the state pension age in the Netherlands, which is linked by law to the forecasted life expectancy, will develop more evenly if the forecast is more robust.

The outcomes of the new mortality forecasts of Statistics Netherlands also affect the official population forecasts for the Netherlands issued by Statistics Netherlands, for which the mortality forecasts represent an important input. Together with assumptions about future migration and fertility, this information (based on the cohort component method) will contribute to a comprehensive forecast of the future population in the Netherlands. From this forecast, other measures can be derived, such as the extent of ageing. If the forecasted life expectancy is higher, the extent of ageing will also be higher than previously expected. The greater robustness of the mortality forecasts will lead to more robust population forecasts

(with respect to mortality) as well. Users should also be aware of how the changes in the population forecasts are related to the new methodology for the mortality forecasts, because once again this association affects various aspects of life (for instance, how many people will receive a pension).

When applying the outcomes of the mortality forecasts (and, subsequently, the population forecasts), users should keep in mind that these measures (like life expectancy at birth) are averages of the population, and will not apply to all segments of the population, as there are very large differences in life expectancy based on, for instance, socio-economic status (Mackenbach et al. 2008; Van Kippersluis et al. 2010). On average, people with fewer years of education have much shorter lives than their better educated counterparts, and some studies have even reported a widening of inequalities in life expectancy between different socio-economic groups. Thus, the overall life expectancy numbers become less informative over time. Users should be aware of this diversity within the population.

A flexible attitude towards the outcomes of mortality forecasts is required of users, as the results of a given mortality forecast will change in response to new mortality developments, new underlying factors, new knowledge about mortality developments, and new methodologies. Moreover, users should be cognisant that a degree of uncertainty is inevitable in every forecast.

References

van Baal, P., Peters, F., Mackenbach, J. and Nusselder, W. (2016). Forecasting differences in life expectancy by education. *Population Studies* 70(2): 201-216.

Basellini, U., Camarda, C.G. and Canudas-Romo, V. (2016). *Modeling and Forecasting Age at Death Distributions: A Nonparametric Approach*. Abstract for IUSSP conference 2017. https://iussp.confex.com/iussp/ipc2017/mediafile/ExtendedAbstract/Paper6042/ModelAndForecastDx_IPC2017.pdf

de Beer, J., Bardoutsos, A. and Janssen, F. (2017). Maximum human lifespan may increase to 125 years. *Nature* 546: E16-E17.

Bengtsson, T. and Christensen, K. (eds.) (2006). *Perspectives on Mortality Forecasting. IV. Causes of Death*. vol. IV, Swedish Social Insurance Agency, Stockholm.

Bohk-Ewald, C., Ebeling, M. and Rau, R. (2017). Life disparity as an additional indicator for evaluating mortality forecasting. *Demography* 54(4): 1559-1577.

Bonneux, L., Looman, C.W.N. and Coebergh, J.W. (2003). Sterfte door roken in Nederland: Meer dan een miljoen doden tussen 1950 en 2015 [Mortality due to smoking in the Netherlands: 1.2 million tobaccorelated deaths between 1950 and 2015]. *Nederlands Tijdschrift voor Geneeskunde 147, 917–921.*

Booth, H., Maindonald, J. and Smith, L.(2002). Applying Lee-Carter under conditions of variable mortality decline. *Population Studies* 56 (3): 325–336.

Cairns, A.J.G, Blake, D., Dowd, L., Coughlan, G.D. and Khalaf-Allah, M. (2011a). Bayesian Stochastic Mortality Modelling for Two Populations. *Astin Bulletin* 41(1): 29–59.

Canudas-Romo, V. (2008). The modal age at death and the shifting mortality hypothesis. *Demographic Research* 19(30): 1179–1204. doi:10.4054/DemRes.2008.19.30.

Currie, I.D., Durban, M. and Eilers, P.H.C. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling* 4: 279–298.

Czado, C., Delwarde, A. and Denuit, M. (2002). Bayesian Poisson log-bilinear mortality projections. *Insurance Mathematics and Economics* 36: 260-284.

Van Duin, C. and Stoeldraijer, L. (2014). Kernprognose 2013–2060: tijdelijk minder geboorten. *Bevolkingstrends* januari 2014.

Giannakouris, K. (2004). *EUROPOP2004: methodology for drafting mortality assumptions.* Working Paper for the Ageing Working Group of the Economic Policy Com mittee. Luxembourg: European Commission.

Hyndman, R.J., Booth, H., and Yasmeen, F. (2013). Coherent mortality forecasting: The product-ratio method with functional time series models. *Demography* 50(1): 261-283. doi:10.1007/s13524-012-0145-5.

Janssen, F. and Kunst, A. (2007). The choice among past trends as a basis for the prediction of future trends in old-age mortality. *Population Studies 61: 315–326.*

Janssen, F. and Kunst, A. (2010). De toekomstige levensverwachting. In: Luijben, A.H.P. and Kommer, G.J. (eds.). *Tijd en toekomst; deelrapport van de VTV 2010 Van gezond naar beter.* RIVM-rapport 270061008, Houten: Bohn Stafleu Van Loghum: 13-20.

Janssen, F., van Wissen, L. and Kunst, A. (2013). Including the smoking epidemic in internationally coherent mortality projections. *Demography* 50: 1341–1362.

Janssen, F. and de Beer, J. (2016). *Projecting future mortality in the Netherlands taking into account mortality delay and smoking.* Joint Eurostat/UNECE Work Session on Demographic Projections. Working Paper 18.

Van Kippersluis, H., O'Donnell, O., Van Doorslaer, E., and Van Ourti, T. (2010). Socioeconomic differences in health over the life cycle in an egalitarian country. *Social Science and Medicine* 70(3): 428-438.

Lee, R. and Miller, T. (2001). Evaluating the Performance of the Lee-Carter Approach to Modeling and Forecasting Mortality. *Demography* 38(4): 537-549.

Leon, D.A. (2011). Trends in European life expectancy: a salutary view. *International Journal of Epidemiology* 40(2): 271-277. https://doi.org/10.1093/ije/dyr061

Li, N.R. and Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography* 42(3): 575-594. doi:10.1353/dem.2005.0021.

Li, J.S-H. and Hardy, M.R. (2011). Measuring Basis Risk in Longevity Hedges. *North American Actuarial Journal* 15(2): 177–200.

Lindahl-Jacobsen, R., Oeppen, J., Rizzi, S., Moller, S., Zarulli, V., Christensen, K., et al. (2016). Why did Danish women's life expectancy stagnate? The influence of interwar generations' smoking behaviour. *European Journal of Epidemiology* 31(12):1207-1211.

Lundström, H. (2003). Mortality assumptions for Sweden. The 2000-2050 Population Projection. In Bengtsson, T. and Keilman, N. (eds.), *Perspectives on Mortality Forecasting I. Current Practice*. Stockholm: Swedish National Social Insurance Board, pp. 59-74.

Mackenbach, J.P., Stirbu, I., Roskam, A. ., Schaap, M.M., Menvielle, G., Leinsalu, M., and Kunst, A.E. (2008). Socioeconomic inequalities in health in 22 european countries. *New England Journal of Medicine* 358(23): 2468-2481.

Mathers, C.D., Ma Fat, D., Inoue, M., Rao, C. and Lopez, A.D. (2005). Counting the dead and what they died from: an assessment of the global status of cause of death data. *Bulletin of the WHO* 2005 83:171-177.

McKee, M. and Shkolnikov, V. (2001). Understanding the toll of premature death among men in eastern Europe. *British Medical Journal* 323. doi: https://doi.org/10.1136/bmj.323.7320.1051

Oeppen J, Vaupel, J.W. (2002). Demography. Broken limits to life expectancy. *Science* 296(5570): 1029-31.

Pedroza, C. (2006). A Bayesian forecasting model: predicting U.S. male mortality. *Biostatistics* 7(4): 530-550.

Peto R., Lopez A., Boreham J., Thun M. and Heath Jr C. (1992). Mortality from tobacco in developed countries: indirect estimation from national statistics. *Lancet* 339: 1268–78.

Rostron, B. 2010. A modified new method for estimating smoking-attributable mortality in high-income countries. *Demographic Research* 23: 399–420.

Samir, K.C., Skirbekk, V., Barakat, B., Sanderson, W., Goujon, A. and Lutz, W. (2010). Projection of populations by level of educational attainment, age, and sex for 120 countries for 2005-2050. *Demographic Research 22(15): 383-472. DOI: 10.4054/DemRes.2010.22.15*

Stoeldraijer, L., van Duin, C. and Janssen, F. (2012). Bevolkingsprognose 2012-2060: model en veronderstellingen betreffende de sterfte. *Bevolkingstrends* 27-6-2013.

Trias-Llimós, S. and Janssen, F. (2017). Country differences in past trends in alcohol-attributable mortality in Europe. *European Journal of Public Health* 27(S3): 359. https://doi.org/(...)93/eurpub/ckx189.147 (abstract)

Vallin, J., and Meslé, F. (2004). Convergences and divergences in mortality: A new approach of health transition. *Demographic Research* 2:11-44.

Vidra, N., Bijlsma, M.J., Trias-Llimós, S. and Janssen, F. (2018). Past trends in obesity-attributable mortality in eight European countries: an application of age-period-cohort analysis. *International Journal of Public Health*. https://doi.org/(...)07/s00038-018-1126-2

Villegas, A. M., and Haberman, S. (2014). On the modeling and forecasting of socioeconomic mortality differentials: An application to deprivation and mortality in England. *North American Actuarial Journal* 18(1): 168-193.

Annex

Bevolkingsprognose 2012-2060: model en veronderstellingen betreffende de sterfte

Als onderdeel van de bevolkingsprognose publiceert het CBS om het jaar een langetermijnprognose voor de sterftekansen en de levensverwachting in Nederland. De nieuwste update van deze prognose is op 13 december 2012 verschenen. Het model voor de sterfteprognose is daarin aangepast. De afgelopen jaren waren regelmatig grote bijstellingen in de prognose noodzakelijk. Om tot een betrouwbaardere en meer robuuste prognose te komen, wordt in de nieuwe methode ook rekening gehouden met de sterfteontwikkelingen in andere West-Europese landen. Bovendien wordt informatie over ontwikkelingen in het rookgedrag op een systematische wijze in de prognose betrokken, wat in het oude model op beperktere schaal gebeurde. Waar de oude prognose een sterke remming voorzag van het stijgingstempo van de levensverwachting, daalt dit volgens de nieuwe prognose geleidelijker. Volgens de nieuwe prognose stijgt de periode-levensverwachting bij geboorte tussen 2012 en 2060 voor mannen met 7,8 en voor vrouwen met 7,0 jaar, om uit te komen op respectievelijk 87,1 en 89,9 jaar. Ten opzichte van de vorige prognose is dit een verhoging van 2,6 jaar voor mannen en van 2,5 jaar voor vrouwen. Op de korte termijn zijn er slechts kleine verschillen tussen de prognoses van 2012 en 2010.

A.1. Inleiding

Het CBS publiceert om het jaar een nieuwe bevolkingsprognose voor de lange termijn. Een onderdeel van de bevolkingsprognose is de prognose van de leeftijdsspecifieke sterftekansen. Deze liggen ten grondslag aan de berekening van het verwachte aantal sterfgevallen per jaar. Op basis van de sterftekansen worden verder sterftematen zoals de periode- en cohort-levensverwachtingen berekend. Deze worden gebruikt door onder meer pensioenverstrekkers en bij doorberekeningen van toekomstige overheidsuitgaven.

De prognoses van 2004 tot en met 2010 gebruikten een model dat uitging van uitsluitend Nederlandse data en maakten onderscheid naar tien groepen van doodsoorzaken (De Jong en Van der Meulen, 2005, Van Duin et al., 2011). Het model voor de prognose van 2012 gebruikt in plaats daarvan een onderverdeling naar rookgerelateerde en niet-rookgerelateerde sterfte en houdt ook rekening met sterfteontwikkelingen in andere West-Europese landen. Deze benadering is gebaseerd op de belangrijke rol van roken in de sterftetrends en plaatst de geobserveerde sterftefluctuaties voor Nederland in een internationale context. De methodologie leidt tot een stabielere trend die vervolgens als basis dient voor de projectie. De verwachting is dat hierdoor de bijstellingen tussen opeenvolgende

prognoses kleiner zullen worden. De nieuwe methode voor de sterfteprognose is een verfijning van de methode gebruikt als onderdeel van de Volksgezondheid Toekomst Verkenning 2010 van het RIVM (Janssen en Kunst, 2010, Janssen et al., te verschijnen).

Dit artikel beschrijft, in hoofdlijnen, de inhoudelijke argumentatie achter het gekozen sterfteprognosemodel (paragraaf A.1.1–A.1.4) en de schattingen op basis van dit model (paragraaf A.2). Paragraaf A.3 beschrijft de uitkomsten van de huidige prognose en maakt een vergelijking met de uitkomsten van de prognose in 2010. Paragraaf A.4 geeft een korte samenvatting.

A.1.1 Ontwikkelingen in de sterfte

Sinds het midden van de 19e eeuw dalen de sterftekansen in Nederland en andere geïndustrialiseerde landen. Vooruitgang in medische kennis en technologie en betere hygiëne, voeding en leefomstandigheden leidden ertoe dat het risico om voortijdig te overlijden sterk terugliep. Op basis van de sterftekansen in hun geboortejaar zouden een Nederlands jongetje en meisje uit 1870 gemiddeld 36 en 39 jaar oud zijn geworden. Een eeuw later was de periode-levensverwachting bij geboorte opgelopen tot respectievelijk 71 en 77 jaar. Het grootste deel van deze stijging kwam door een daling van de sterfte op jonge leeftijden. Meer recent wordt de ontwikkeling van de levensverwachting vooral bepaald door sterftedalingen bij de hogere leeftijden, aangezien de sterfte op jonge leeftijden al zo laag is dat een verdere daling de levensverwachting nog maar weinig beïnvloedt.

De daling van de sterftekansen en de toename in de levensverwachting verlopen niet gelijkmatig over de jaren. Periodes van relatieve stagnatie worden gevolgd door periodes van versnelde stijging, die later vaak tijdelijk blijken te zijn. Zo is tussen 1950 en 1970 de levensverwachting van mannen toegenomen met maar 0,3 jaar per decennium, terwijl de levensverwachting van vrouwen in die periode toenam met 2,0 jaar per decennium (grafiek A.1.1.1). In de periode 1970-2002 steeg de levensverwachting voor mannen vervolgens met 1,6 jaar per decennium. Bij de vrouwen was er in de jaren '70 nog een sterke toename van 2,7 jaar, maar daarna vertraagde de groei tot 0,7 jaar per decennium over de periode 1980-2002. Vanaf 2002 versnelde de stijgende trend bij zowel mannen als vrouwen sterk. De gemiddelde toename per decennium in de periode 2002-2011 bedroeg 3,5 jaar voor mannen en 2,4 jaar voor vrouwen.

Doordat de toename in de levensverwachting voor mannen anders verliep dan voor vrouwen, varieert het verschil in levensverwachting tussen mannen en vrouwen over de jaren. In 1950 bedroeg het verschil 2,3 jaar. Door de stagnatie in de toename van de levensverwachting van mannen groeide dit naar 6,7 jaar begin

jaren '80. Daarna stagneerde de groei in de levensverwachting van vrouwen, waardoor het verschil weer afnam. Ook in de periode vanaf 2002, waarin de levensverwachting van zowel mannen als vrouwen sterk toenam, daalde het verschil. In 2011 was het teruggelopen tot 3,7 jaar.

De (onverwacht) versnelde stijging van de levensverwachting sinds 2002 betekende dat de kortetermijnontwikkeling niet aansloot bij de langetermijntrend. Dit leidde tot onderschattingen in opeenvolgende CBS-prognoses (grafiek A.1.1.1). De levensverwachting die voor 2050 werd voorzien in de prognose van 2002 en 2004, werd in 2011 al vrijwel bereikt. In de prognoses van 2006 en 2008 werden de vooruitzichten voor de levensverwachting naar boven bijgesteld, maar bleek er nog steeds een onderschatting te zijn. In de prognose van 2010 werden de vooruitzichten voor de levensverwachting verder verhoogd. In 2012 bleek dat de prognose van 2010 de levensverwachting voor mannen correct heeft voorspeld, maar dat deze voor vrouwen 0,2 jaar te hoog is geschat[1)].

Het herhaaldelijk onderschatten van de toekomstige levensverwachting door prognoses gebeurt niet alleen in Nederland, maar is een wereldwijd fenomeen (Oeppen en Vaupel, 2002). Vaak waren onderschattingen een gevolg van de aanname dat de levensverwachting een biologisch vastliggende maximum waarde naderde en dus niet veel meer kon stijgen. Hoewel in de CBS-prognose uit 2010 niet zo'n aanname is gedaan, voorzag ook deze prognose een sterke afremming van de toename van de levensverwachting.

A.1.2 De rol van roken

Voor Nederland kunnen de historische periodes van stagnatie (vooral voor mannen) en de schommelingen in het verschil in levensverwachting tussen mannen en vrouwen voor een groot deel verklaard worden door roken. Roken is een belangrijke factor voor de levensverwachting (Peto et al., 1996). Met de methode die beschreven wordt in paragraaf 2.3 kan de levensverwachting gecorrigeerd worden voor het effect van roken. Na deze correctie blijken de stagnatie onder mannen in de periode 1950–1970 en de verschillen in de ontwikkeling tussen de geslachten grotendeels verdwenen. Over de periode 1970–2011 bleef het verschil in levensverwachting zonder roken tussen mannen en vrouwen rond de 3 jaar (grafiek A.1.2.1).

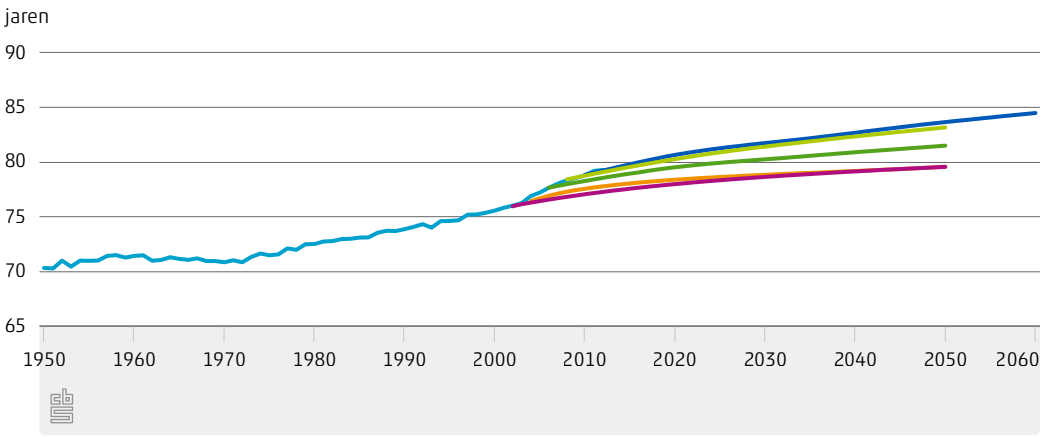
Het verschil in de totale levensverwachting en de levensverwachting zonder roken is verschillend voor mannen en vrouwen, doordat mannen eerder en massaler zijn

1) We vergelijken met de geraamde levensverwachting voor 2012, op basis van de voorlopige sterftecijfers tot en met week 44. Hierop zit nog een onzekerheidsmarge van ongeveer 0,2 jaar

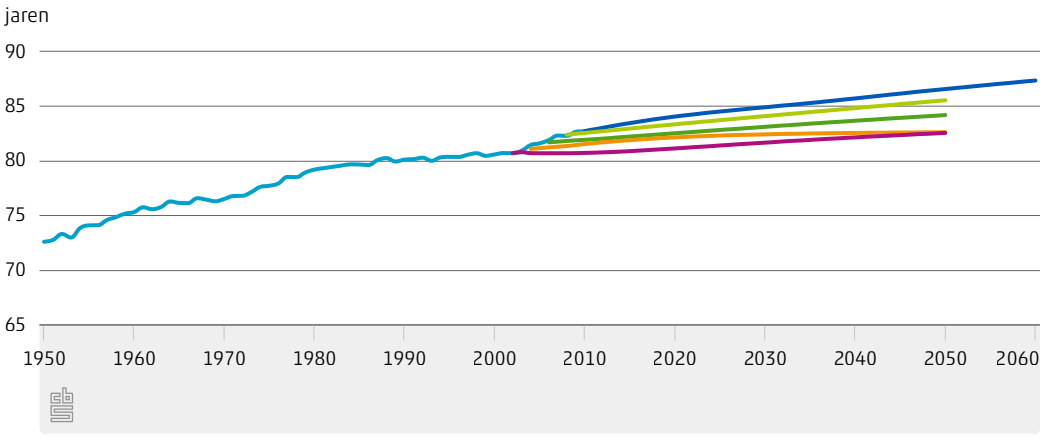
begonnen met roken (zie ook Lopez et al., 1994). Tot begin jaren '70 nam de totale levensverwachting voor mannen nauwelijks toe, terwijl de levensverwachting zonder roken wel toenam (grafiek A.1.2.1). Door de sterke afname van het aandeel rokers onder mannen vanaf de jaren '50 liep de totale levensverwachting voor mannen sinds begin jaren '80 in op de levensverwachting zonder roken.

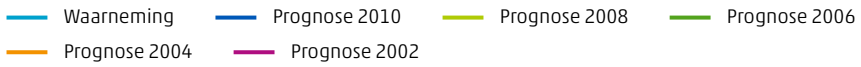
A.1.1.1 Levensverwachting bij geboorte, waarneming en prognose

a. Mannen



a. Vrouwen

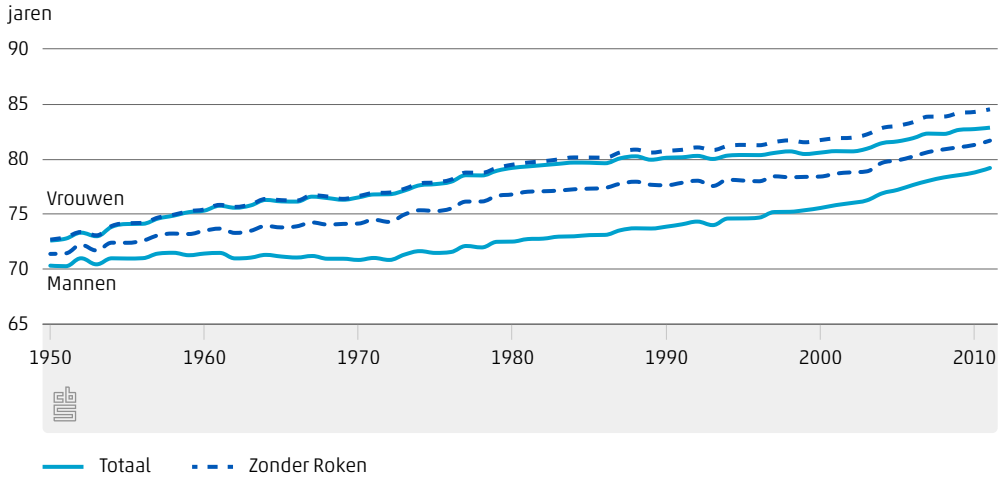




Ook vrouwen roken momenteel minder dan in de jaren '70, maar doordat het effect van roken vertraagd doorwerkt op de sterfte, is voor vrouwen de totale levensverwachting sinds 1980 minder toegenomen dan de levensverwachting zonder roken. Veranderend rookgedrag zorgt dus voor een duidelijk niet-lineair patroon in de levensverwachting.

Doordat veranderingen in rookgedrag vertraagd doorwerken in de sterfte, geven de huidige rooktrends informatie over wat er in de (nabije) toekomst verwacht mag worden. Als gevolg van de daling van het aandeel rokende vrouwen in de afgelopen decennia zal de totale levensverwachting van vrouwen weer inlopen op de levensverwachting zonder roken, net als bij mannen is gebeurd. Het doortrekken van de huidige trend bij vrouwen in de totale sterfte zal daarom leiden tot een onderschatting van de levensverwachting op de lange termijn, omdat dan geen rekening wordt gehouden met een afname van de sterfte gerelateerd aan roken als gevolg van het gedaalde aandeel rokende vrouwen. Daarnaast valt te verwachten dat, wanneer de afname van het percentage rokers doorzet, het geslachtsverschil uiteindelijk terugloopt tot rond de 3 jaar, overeenkomstig het verschil in levensverwachting voor niet-rokende mannen en vrouwen.

A.1.2.1 Totale levensverwachting en levensverwachting zonder roken, bij geboorte



A.1.3 Resterende fluctuaties in de levensverwachting

De periodes van stagnatie en sterkere toename van de levensverwachting die overblijven na correctie voor roken zijn voor mannen en vrouwen bijna gelijk (grafiek A.1.2.1). Eind jaren '60 was er een stagnatie, waarna er vanaf begin jaren '70 een sterke stijging inzette. De toename van de levensverwachting stagneerde vervolgens weer in de jaren '80 en '90, gevolgd door een sterke toename vanaf 2002. Onderzoek naar de omslag van stagnatie naar sterke groei in 2002 suggereert dat de causale factoren onmiddellijk optraden in plaats van met een langdurige vertraging: de sterfte nam in bijna alle leeftijdsgroepen gelijktijdig af

(Mackenbach en Garssen, 2011). Van de vier onderzochte categorieën lijken veranderingen in de gezondheidszorg de beste kandidaat om de verandering van de sterftetrends te verklaren. Voorbeelden van deze veranderingen zijn een toename van de uitgaven aan gezondheidszorg en het aantal ziekenhuisopnames, beide gevolgen van een versoepeling van budgettaire beperkingen door de overheid. Het gepresenteerde bewijs is echter slechts indirect. Er waren geen plotselinge veranderingen in de gezondheidstoestand van ouderen, in hun fysieke of sociale omgeving of in hun levensstijl die deze omkering zouden kunnen verklaren.

Zoals eerder aangegeven blijken periodes van relatieve stagnatie of versnelde stijging later vaak tijdelijk. Dit komt ook naar voren als de ontwikkeling van de levensverwachting in Nederland in een internationale context wordt geplaatst.

A.1.4 Nederland in vergelijking met andere West-Europese landen

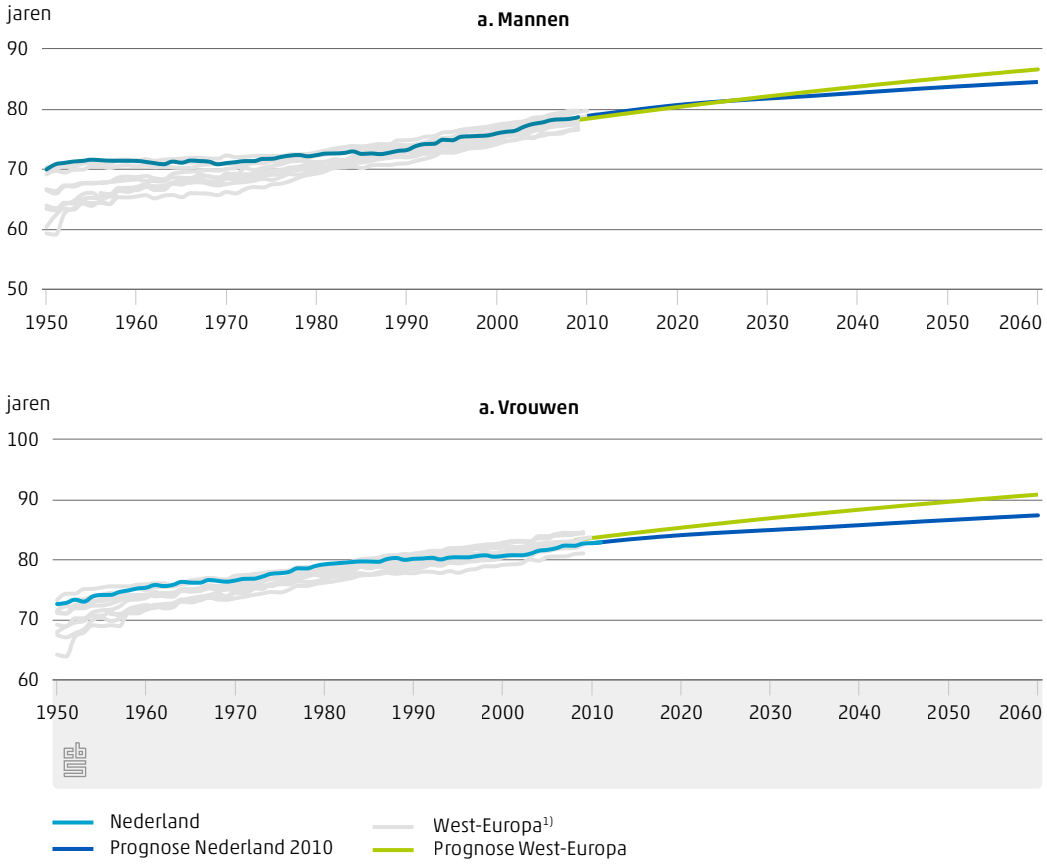
Als gevolg van de schommelingen in de sterftetrends heeft het niveau van de levensverwachting in Nederland door de jaren heen een andere positie ingenomen dan die in andere geselecteerde landen in West-Europa[2]. Tot 1970 was de levensverwachting in Nederland voor zowel mannen als vrouwen bovengemiddeld vergeleken met andere West-Europese landen (grafiek A.1.4.1). Doordat voor mannen de jaarlijkse toename na 1970 in deze andere landen doorgaans hoger lag, was de levensverwachting in Nederland rond 2000 afgezakt naar het gemiddelde. Door de sterke stijging van de levensverwachting voor Nederlandse mannen sinds 2002 ligt hun levensverwachting sinds 2005 weer boven het gemiddelde niveau voor West-Europese mannen. De levensverwachting van Nederlandse vrouwen stagneerde vanaf 1980 in vergelijking met die van West-Europese vrouwen. De levensverwachting van Nederlandse vrouwen lag daardoor rond de eeuwwisseling onder het West-Europees gemiddelde. Sinds 2002 loopt de levensverwachting van Nederlandse vrouwen ook iets in, maar de levensverwachting blijft nog ruim onder het West-Europese gemiddelde.

Ook andere landen in West-Europa hebben te maken gehad met tijdelijke versnellingen en vertragingen in de ontwikkelingen van de levensverwachting (grafiek A.1.4.1). De omslagpunten en de duur van periodes met stagnatie of herstel verschillen echter per land. Opvallend aan de gemiddelde

2) Denemarken, Finland, Frankrijk, Duitsland, Italië, Noorwegen, Spanje, Zweden, Zwitserland en Engeland&Wales, bron HMD (2012). In het vervolg wordt verwezen naar deze groep als West-Europese landen.

levensverwachting van de geselecteerde West-Europese landen is dat de jaarlijkse toename over de periode 1970-2008 nagenoeg constant is en de levensverwachting voor mannen en vrouwen een bijna rechte lijn vormt. Bovendien neemt de spreiding in levensverwachting tussen de landen niet toe, wat betekent dat over de langere termijn de individuele landen de trend van de groep volgden. Dit is een logisch gevolg van vooral universele sociaaleconomische ontwikkelingen en een over het algemeen gelijke ontwikkeling van medische zorg en leefstijl in de West-Europese landen. Daarnaast is er sprake van benchmarking: preventie- of behandelmethoden uit beter presterende landen kunnen worden overgenomen wanneer blijkt dat de ontwikkelingen op het gebied van

A.1.4.1 Levensverwachting bij geboorte, waarnemingen en prognose West-Europese landen



1) Denemarken, Finland, Frankrijk, Duitsland, Italië, Noorwegen, Spanje, Zweden, Zwitserland en Engeland&Wales, bron Human Mortality Database.

sterftereductie in een bepaald land achterblijft. Een voorbeeld is de toegenomen aandacht voor zuigelingensterfte in Nederland toen bleek dat de daling hiervan achterbleef bij die in andere Europese landen.

In grafiek A.1.4.1 is een prognose opgenomen voor de levensverwachting in de West-Europese landen met behulp van het Lee-Carter extrapolatiemodel (Lee en Carter, 1992). Hierbij is geen rekening gehouden met de invloed van roken op de sterfte. Het Lee-Carter extrapolatiemodel is een in andere landen veelgebruikte methode waarbij de historische sterftequotiënten worden geëxtrapoleerd. De toename in de levensverwachting van West-Europese mannen daalt volgens dit model van de huidige 3,0 jaar per decennium naar 1,4 jaar per decennium na 2050. Voor West-Europese vrouwen daalt de toename van 2,0 jaar per decennium naar 1,2 jaar per decennium.

Het valt op dat de CBS-prognose in 2010, die zich baseert op de relatief minder gunstige ontwikkelingen in Nederland, een veel sterkere afname laat zien en daardoor divergeert van deze toekomstige trend voor de West-Europese landen. Als aangenomen wordt dat de stabiele trend in de sterfteontwikkelingen van de West-Europese landen zich in de toekomst doorzet, valt te verwachten dat de CBS-prognose de levensverwachting op de lange termijn onderschat.

A.2. Methode

A.2.1 Methode sterfteprognose vóór 2012

Sinds 2004 onderscheidt het CBS verscheidene doodsoorzaken in het prognosemodel (De Jong en Van der Meulen, 2005, Van Duin et al., 2011). Dit biedt de mogelijkheid om inhoudelijke informatie mee te wegen, bijvoorbeeld over determinanten van de sterfte door een bepaalde doodsoorzaak. Ook kunnen niet-lineariteiten in het verloop van de sterfte worden gemodelleerd die ontstaan doordat bij verschillende doodsoorzaken op verschillende momenten trendbreuken optreden. Voor een prognose die regelmatig wordt geactualiseerd, zoals de CBS-prognose, heeft deze aanpak als voordeel dat per doodsoorzaak kan worden nagegaan in hoeverre de prognose afwijkt van de realisatie, en of er nieuwe inzichten zijn over te verwachten ontwikkelingen, waarna gefundeerde bijstellingen kunnen worden doorgevoerd. Het nadeel van deze aanpak is dat het aantal benodigde veronderstellingen zeer groot is, omdat per doodsoorzaak en per

leeftijdscategorie veronderstellingen moeten worden opgesteld over het toekomstig verloop van de sterftekans. Dit maakt de methode in de praktijk weinig transparant. Daarnaast bleek het in de praktijk moeilijk om gefundeerde expertverwachtingen over toekomstige ontwikkelingen per doodsoorzaak te verkrijgen, waardoor grotendeels met extrapolatie moest worden gewerkt. Bovendien maakte het detailniveau van de prognose het problematisch om internationale trends mee te nemen.

Naast deze praktische problemen zijn er meer fundamentele bezwaren tegen het oude model. In de literatuur wordt als algemeen nadeel van het maken van een sterfteprognose naar doodsoorzaken genoemd dat de doodsoorzaak met de minst gunstige ontwikkeling de algehele sterfteontwikkeling gaat domineren (Wilmoth, 1995). Impliciet veronderstelt een model dat sterfte per doodsoorzaak extrapoleert dat toekomstige doorbraken op medisch gebied plaatsvinden bij dezelfde ziekten waar in het verleden veel vooruitgang is geboekt, met als gevolg dat die doorbraken steeds minder effect hebben. Te verwachten valt echter dat medisch onderzoek zich juist meer zal richten op de ziekten die een toenemend aandeel in de sterfte hebben. Extrapoleren van trends per doodsoorzaak geeft dan een te pessimistisch beeld. Wanneer over een langere periode wordt gekeken, is ook zichtbaar dat de stijging van de levensverwachting in het verleden in verschillende perioden met vooruitgang in de bestrijding van verschillende doodsoorzaken samenhing (Wilmoth, 2000).

Dat de doodsoorzaak met de minst gunstige ontwikkeling de algehele sterfteontwikkeling gaat domineren in het prognosemodel naar doodsoorzaken verklaart gedeeltelijk de sterke afremming van de stijging van de geprojecteerde levensverwachting op de lange termijn. Wat verder meespeelt is dat de sterftekansen bij ouderen in Nederland sinds 1970 relatief weinig gedaald zijn ten opzichte van die in andere West-Europese landen. Met name in de jaren '80 en '90 stagneerde de daling in de sterftekansen van ouderen (Janssen et al., 2004). De West-Europese trend in de sterftekansen bij ouderen is gunstiger. Doordat Nederland de laatste tien jaar bij deze trend heeft aangehaakt, lijkt de stagnatie in Nederland in de jaren '80 en '90 tijdelijk te zijn geweest. In de prognose van 2010 werkten de ongunstige Nederlandse ontwikkelingen van de jaren '80 en '90 echter sterk door in de sterftetrend bij ouderen, doordat het model werkt met de veronderstelling dat het dalingstempo van de sterftekansen bij de verschillende leeftijden in de toekomst hetzelfde zal zijn als in het verleden. Dit heeft tot gevolg dat op de lange duur de tragere daling bij de hoogste leeftijden de sterftetrend gaan domineren, waardoor de stijging van de levensverwachting wordt afgeremd.

Door enkel rekening te houden met de ontwikkeling in Nederland leidde dit dus tot een extra afremming in de stijging van de geprojecteerde levensverwachting (grafiek A.1.4.1).

Samengevat was het voordeel van het oude model dat het een inhoudelijke toelichting gaf op de prognose van de levensverwachting. De toekomstverwachtingen werden geformuleerd op basis van duiding van recente ontwikkelingen en, waar mogelijk, op basis van beargumenteerde verwachtingen voor de toekomst. Voor een prognose die kort vooruit kijkt is dat mogelijk een goede aanpak, maar voor een langetermijnprognose is het essentieel om tijdelijke afwijkingen goed te onderscheiden van structurele ontwikkelingen. Hiervoor is het juist van belang om op een hoger niveau te kijken. Het onderscheid naar doodsoorzaken had meerwaarde waar het de mogelijkheid gaf om het effect van veranderend rookgedrag op de sterfte aan longkanker en COPD te modelleren. Dat aspect is in het huidige model gehandhaafd, maar uitgebreid naar een breder onderscheid tussen longkankersterfte en overige rookgerelateerde sterfte enerzijds en niet-rookgerelateerde sterfte anderzijds.

A.2.2 Methode sterfteprognose 2012

Ten opzichte van de voorgaande sterfteprognoses is er een grote verandering aangebracht in de methode die de sterftekansen projecteert. De methode is een verfijning van de methode uit de Volksgezondheid Toekomst Verkenning 2010 van het RIVM (Janssen en Kunst, 2010, Janssen et al., te verschijnen). Hierin werd een nieuwe methodologie gebruikt die rekening houdt met de sterfteontwikkelingen in andere Europese landen en die bovendien informatie over ontwikkelingen in roken op een systematische wijze in de berekening betrekt. De nieuwe methodologie sluit aan bij de inhoudelijke observaties uit paragraaf A.1 over de belangrijke rol van roken in de sterftetrends en de plaatsing van de geobserveerde sterftefluctuaties voor Nederland in een internationale context.

Roken verklaarde voor een groot deel de historische stagnaties en de verschillen in de ontwikkeling in de sterfte tussen mannen en vrouwen. Daarnaast liet roken een duidelijk niet-lineair patroon zien als gevolg van de toename en afname van het percentage rokers in Nederland, eerst voor mannen en vervolgens voor vrouwen (zie ook paragraaf A.1.2). Het is belangrijk om deze niet-lineaire patronen te onderscheiden van de algehele sterfteontwikkelingen. De resterende niet-rookgerelateerde sterfte is stabieler en kan hierdoor beter gebruikt worden voor de projectie (Janssen en Kunst, 2007). De niet-lineaire ontwikkelingen in de rookgerelateerde sterfte dienen daarnaast afzonderlijk te worden voorspeld.

Na correctie voor roken blijven er periodes van stagnatie en sterkere toename aanwezig in de sterfteontwikkelingen in Nederland. Een eenduidige verklaring ontbreekt echter en dergelijke fluctuaties – die ook in andere West-Europese landen voorkomen – blijken achteraf vaak tijdelijk. Over het algemeen is een duidelijke toename van de levensverwachting in alle West-Europese landen te zien, door overeenkomstige sociaaleconomische ontwikkelingen, medische vooruitgang en leefstijl. De stabiele ontwikkelingen van een groep met landen geven daarom een betere indicatie van de toekomstige ontwikkelingen in de niet-rookgerelateerde sterfte. Daarnaast valt te verwachten dat Nederland ook in de toekomst niet langdurig uit de pas zal lopen met andere West-Europese landen. Voor de projectie van de niet-rookgerelateerde sterfte wordt daarom aangenomen dat de ontwikkeling voor de afzonderlijke geslachten in Nederland op de lange termijn parallel zal lopen aan de sterfteontwikkeling in vergelijkbare landen voor de afzonderlijke geslachten.

Voor de projectie van de niet-rookgerelateerde sterfte wordt aangenomen dat de daling van de leeftijdsspecifieke sterftekansen in West Europa in hetzelfde tempo doorzet als in de periode sinds 1970. Er wordt dus niet verondersteld dat deze kansen al dicht bij een ondergrens liggen en om die reden niet veel verder zouden kunnen dalen. Vroegere prognoses die wel een dergelijke ondergrens veronderstelden, met uiteenlopende inhoudelijke argumentaties, zijn steeds te conservatief gebleken (Wilmoth, 2000). Omdat sterftekansen op hoge leeftijden in het verleden echter minder snel zijn gedaald dan op jonge en middelbare leeftijden, en dit ook voor de toekomst is te veronderstellen, wordt de toekomstige groei van de levensverwachting in de sterfteprognose van 2012 alsnog afgeremd. Naarmate meer mensen de hoge leeftijden bereiken, gaat het tragere dalingstempo van de sterfte op hoge leeftijden domineren, waardoor de stijging van de levensverwachting afzwakt. Dit wordt rectangularisatie genoemd. Dit is een duidelijk verschil ten opzichte van extrapolatiemodellen die de stijging in de levensverwachting lineair doortrekken, want deze modellen nemen impliciet aan dat de sterftedalingen bij de hogere leeftijden zullen versnellen. De methode voor de sterfteprognose 2012 neemt in dat opzicht een meer conservatief uitgangspunt: dat het waargenomen gemiddelde dalingstempo uit het verleden in de toekomst doorzet. Dit is in lijn met de huidige praktijk bij andere statistische bureaus (Stoeldraijer et al., ingediend). Omdat er geen consensus is dat deze zogenoemde rectangularisatie in de toekomst blijft optreden (Oeppen en Vaupel, 2002), is het mogelijk dat de gekozen methode voor de sterfteprognose 2012 alsnog te pessimistisch is.

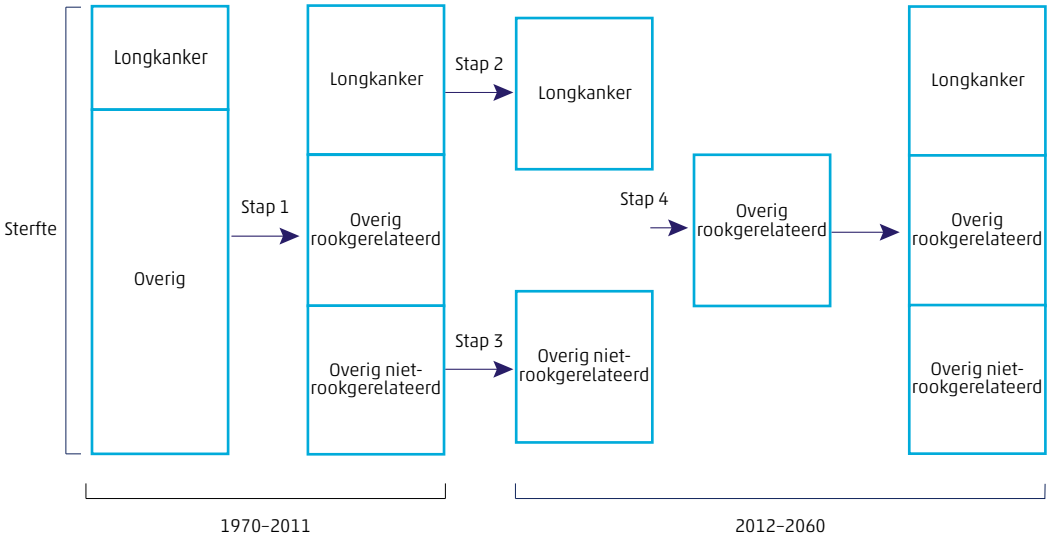
De nieuwe methode voor de CBS-sterfteprognose 2012 bestaat uit drie belangrijke onderdelen: 1) het opsplitsen van de totale sterfte in niet-rookgerelateerde en

rook-gerelateerde sterfte met behulp van een indirecte schattingsmethode voor de sterfte gerelateerd aan roken, 2) de projectie van de niet-lineaire ontwikkelingen in rookgerelateerde sterfte door middel van een projectie van de longkankersterfte, en 3) de toevoeging van sterfteontwikkeling in andere landen aan de projectie van niet-rookgerelateerde sterfte waarin de trends uit het verleden worden doorgetrokken.

A.2.3 Beschrijving methode sterfteprognose 2012

De methode voor de sterfteprognose 2012 is opgedeeld in vier stappen. Deze worden grafisch weergegeven in grafiek A.2.3.1 en worden hieronder toegelicht (zie ook Bijlage A voor de formules). De methode start met het opdelen van de totale sterfte uit het verleden in twee groepen: de doodsoorzaak longkanker en alle andere doodsoorzaken ('overig'). In stap 1 wordt de groep overig opgesplitst in rookgerelateerd en niet-rookgerelateerd aan de hand van een indirecte schattingsmethode op basis van de waargenomen longkankersterfte. Stap 2 is de projectie van de longkankersterfte en stap 3 een projectie van de overig niet-rookgerelateerde sterfte. Hieruit volgen dus toekomstige waarden voor deze twee groepen. Met behulp van de toekomstige waarden uit stap 2 en 3 en opnieuw de indirecte schattingsmethode die ook in stap 1 is gebruikt, kunnen vervolgens in stap 4 de toekomstige waarden van de overig rookgerelateerde sterfte worden uitgerekend. Door optellen volgt dan de totale sterfte in de prognoseperiode.

A.2.3.1 Stappen CBS-sterfteprognose 2012



Voor de berekeningen is globaal gebruik gemaakt van de leeftijdsintervallen 0, 1–4, 5-9, ..., 90–94, 95+ en de jaren 1970–2011. Welke data en leeftijdsintervallen precies per stap zijn gebruikt, staat in bijlage A. De leeftijdsspecifieke sterftekansen worden uit de sterftekansen per leeftijdsinterval afgeleid en gefit aan de raming van 2012.

Stap 1: Schatting overig rookgerelateerde en overig niet-rookgerelateerde sterfte

De eerste stap in het maken van de sterfteprognose is het uitrekenen van de sterfte die toe te kennen is aan roken. Idealiter zouden hiervoor data over rookgedrag en sterfte worden gebruikt zodat het verhoogde risico van rokers kan worden uitgerekend, maar uitgebreide gegevens zijn hierover niet beschikbaar. In plaats daarvan wordt een indirecte methode (Rostron, 2010) gebruikt om een schatting te maken van het aantal sterfgevallen dat niet zou plaatsvinden als rokers dezelfde sterftecijfers zouden hebben als nietrokers. Hierbij wordt aangenomen dat rookgedrag de enige factor is die het risico van longkankersterfte van rokers ten opzichte van niet-rokers verhoogt. De methode gebruikt een statistisch regressiemodel om de relatie tussen longkankersterfte en sterfte aan andere doodsoorzaken in ontwikkelde landen tussen 1950 en 2003 te schatten. Aan de hand van deze relatie kan worden uitgerekend hoeveel sterfgevallen er zouden zijn gegeven de waargenomen longkankersterfte in de populatie en hoeveel sterfgevallen er zouden zijn als de longkankersterfte het niveau van een niet-rokende populatie zou hebben. Het verschil tussen deze aantallen, gedeeld door het geschatte aantal sterfgevallen gegeven de waargenomen longkankersterfte in de populatie, geeft het aandeel rookgerelateerde sterfte. De methode is toepasbaar op verschillende populaties. Naast de totale sterftecijfers zijn alleen gegevens over longkankersterfte nodig.

De schatting wordt uitgevoerd voor Nederland en voor de West-Europese landen Denemarken, Duitsland, Engeland en Wales, Finland, Frankrijk, Italië, Noorwegen, Spanje, Zweden en Zwitserland. De in deze stap verkregen 'overig niet-rookgrelateerde sterfte' wordt vervolgens geprojecteerd in stap 3.

Stap 2: Projectie longkankersterfte

De longkankersterfte wordt geprojecteerd met behulp van het zogenaamde Age-Period-Cohort-model (APC-model). Dit model is populair in de epidemiologie met verscheidene toepassingen in de sterfte naar doodsoorzaken en ziekte-incidentie en wordt ook binnen de demografie toegepast op historische sterftedata (Bonneux et al., 2003, Barendregt et al., 2002).

Het APC-model beschrijft (de log van) de sterftecijfers als een som van effecten van leeftijd (Age), periode (Period) en geboortejaar (Cohort). Zo wordt onderscheid gemaakt tussen effecten met een duidelijke periode-signatuur, zoals de invoering van maatregelen die stoppen met roken stimuleren, of het beschikbaar komen van nieuwe behandelmethoden, en historische factoren die gedurende de rest van het leven doorwerken, zoals structurele stijgingen of dalingen van het aantal jongeren dat begint met roken.

Voor de projectie wordt gebruik gemaakt van de aanname dat de huidige trends in de toekomst zullen voortzetten. Daarbij wordt aangenomen dat de toename van longkankersterfte vooral veroorzaakt is – en bij vrouwen nog zal worden veroorzaakt – door een cohortpatroon en dat de afname van de longkankersterfte een periodepatroon volgt. Aangenomen wordt dat de longkankersterfte van mannen in de projectieperiode verder zal dalen, terwijl de longkankersterfte van vrouwen voor de meeste leeftijden zal stijgen tot het niveau van mannen is bereikt en vervolgens gelijk met de mannen zal dalen. Voor deze aanname is gebruik gemaakt van het algemene rookepidemiemodel van Lopez et al. (1994) en ervaringen in Engeland en Denemarken. De aanname spoort bovendien met de ontwikkelingen in de longkankersterfte bij vrouwen onder de 50, waar de omslag van stijgende naar dalende sterfte al heeft plaatsgevonden. In bijlage B is een grafiek met de resultaten opgenomen.

Stap 3: Projectie overig niet-rookgerelateerde sterfte

De niet-rookgerelateerde sterfte in Nederland wordt geprojecteerd door rekening te houden met de sterfte ontwikkelingen in andere landen in West-Europa (Janssen en Kunst, 2010). Hiertoe wordt het Li-Lee-model gebruikt (Li en Lee, 2005). Dit model is een uitbreiding van de Lee-Carter-methodologie (Lee en Carter, 1992), die ervan uitgaat dat de dynamiek van de sterftecijfers in de tijd wordt aangedreven door een enkele tijdsafhankelijke parameter. Het Lee-Carter-model veronderstelt een constant dalingstempo van de leeftijdsspecifieke sterftecijfers, waarbij het tempo wel tussen leeftijden kan verschillen. Door het verwijderen van de niet-lineaire rookgerelateerde sterfte uit de totale sterfte, wordt de fit van het model aan de waarnemingen verbeterd, wat tot betere parameterschattingen en een stabielere extrapolatie leidt.

De Li-Lee-methodologie is ontwikkeld met het idee dat de sterftetrends van verschillende landen in de toekomst niet langdurig zullen divergeren. Hierbij worden verschillen in de sterfte op korte termijn behouden, maar op de lange termijn zijn de leeftijdspecifieke sterftecijfers beperkt tot een constante verhouding met elkaar. In de praktijk komt het erop neer dat de Lee-Carter-methodologie twee

keer wordt toegepast: eerst op de sterfte in de totale populatie om de gemeenschappelijke trend in leeftijdsspecifieke sterftecijfers te bepalen, en vervolgens op de residuen voor ieder land afzonderlijk om de afwijking ten opzichte van de gemeenschappelijke trend te bepalen.

Als totale populatie wordt gekozen voor de eerder genoemde West-Europese landen (inclusief Nederland) per geslacht. De gekozen landen hebben een vergelijkbaar (relatief hoog) sociaaleconomisch ontwikkelingsniveau. De overige niet-rookgerelateerde sterfte uit stap 1 wordt hiertoe gewogen naar de bevolking van deze landen. De geslachten worden afzonderlijk bekeken omdat de levensverwachtingen voor mannen en vrouwen verschillen. Het samenvoegen van de geslachten in een model leidt daardoor bijna automatisch tot een convergentie. Deze aanname willen wij niet doen. Bovendien is de trend in de overige niet-rookgerelateerde sterfte voor mannen en vrouwen afzonderlijk robuust genoeg.

Deze stap geeft vervolgens als uitkomst de projectie van de overige niet-rookgerelateerde sterftecijfers per geslacht voor de totale groep West-Europese landen en voor Nederland.

Stap 4: Berekenen projectie totale sterfte

In de laatste stap worden de geprojecteerde longkankersterfte en de geprojecteerde overige nietrookgerelateerde sterfte gebruikt om de overige rookgerelateerde sterfte uit te rekenen. Hiervoor wordt weer de methode van Rostron (2010) uit stap 1 gebruikt, maar dan in omgekeerde volgorde. Door optelling van de longkankersterfte, niet-rookgerelateerde sterfte en rookgerelateerde sterfte ontstaat de totale sterfte.

A.2.4 Aannames

In het voorgaande is een aantal aannames langsgekomen die aan de prognose ten grondslag liggen. Deze paragraaf vat deze voor de helderheid samen.

Verondersteld wordt dat de stabiele dalende trend in de leeftijdsspecifieke sterftekansen in West-Europese landen in hetzelfde tempo doorzet, en dat de tempoverschillen tussen de verschillende leeftijdsgroepen blijven bestaan. Daarbij lopen de niet-rookgerelateerde sterfteontwikkelingen voor Nederlandse mannen en vrouwen op de lange termijn parallel aan de niet-rookgerelateerde sterfteontwikkelingen voor mannen en vrouwen in de geselecteerde West-Europese landen.

De periode 1970–2011 wordt gebruikt als basisperiode voor de projectie. Dit is een subjectief element in de prognose, maar omdat de sterfte wordt opgesplitst in rookgerelateerd en niet-rookgerelateerd en er rekening wordt gehouden met internationale trends, is dit element minder van belang vanwege de stabielere historische trend.

Er wordt aangenomen dat de huidige daling in de longkankersterfte van mannen zal doorzetten. Voor de vrouwen wordt verondersteld dat de stijging van de longkankersterfte in de verschillende leeftijdscategorieën doorzet tot het niveau bij mannen is bereikt, waarna een daling zal inzetten, zoals ook bij de mannen is gebeurd. Dit vertaalt zich in soortgelijke toekomstige trends voor mannen en vrouwen voor de rookgerelateerde sterfte. Omdat de rookgerelateerde sterfte vertraagd reageert op veranderingen in rookgedrag, rechtvaardigen het dalende aandeel rokende mannen en de toename van het aandeel rokende vrouwen dat gevolgd werd door een daling in de afgelopen decennia deze aannames kwalitatief voor ongeveer de komende twintig jaar. Voor de veronderstelde langeretermijndaling van de longkankeren rookgerelateerde sterfte moet de dalende trend in het percentage rokers ook in de toekomst doorzetten, of moet roken minder schadelijk worden.

De indirecte schattingsmethode voor de rookgerelateerde sterfte berust op de aanname dat de relatie tussen veranderingen in de longkankersterfte en de rookgerelateerde sterfte aan andere doodsoorzaken dan longkanker over de tijd stabiel is. Daarnaast is aangenomen dat roken geen invloed heeft op de sterfte voor mannen onder de 40 en voor vrouwen onder de 45 jaar.

A.3. Resultaten

A.3.1 Levensverwachting bij geboorte en op leeftijd 65

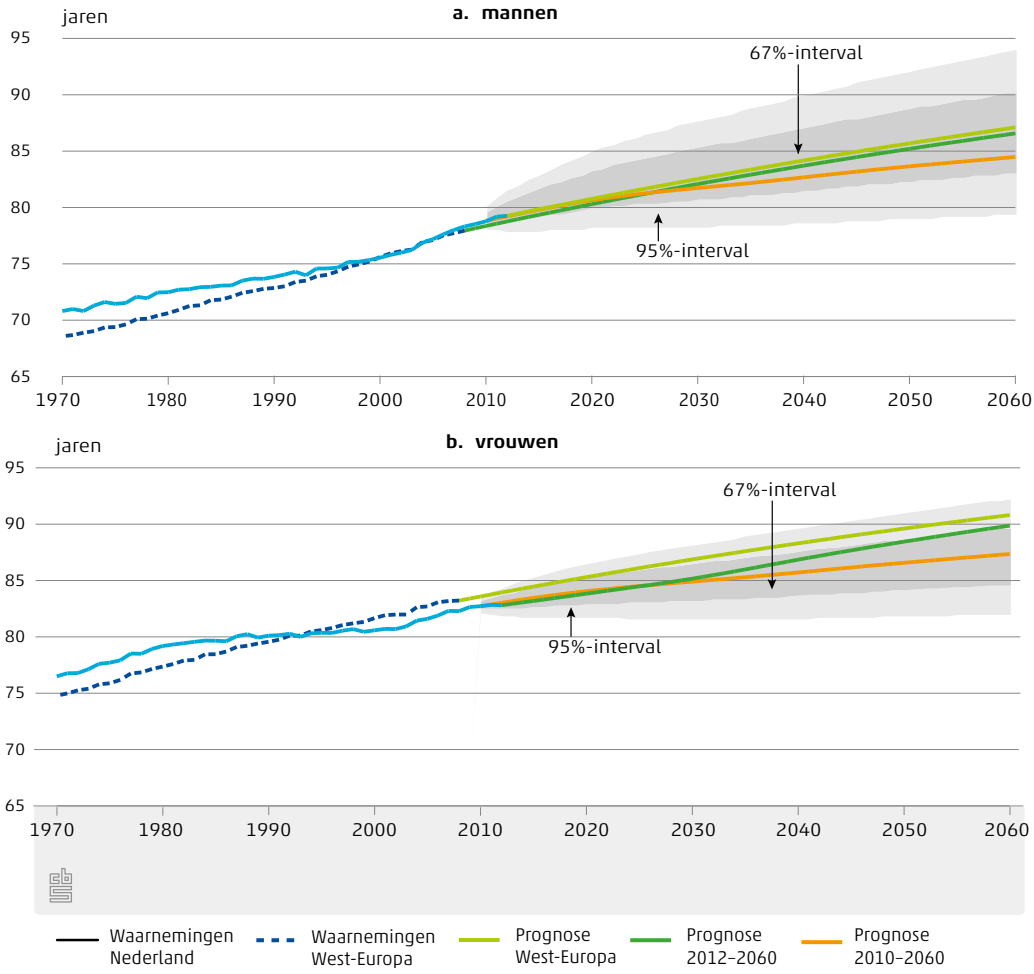
De levensverwachting bij geboorte stijgt volgens de nieuwe prognose tot 87,1 jaar voor mannen en 89,9 jaar voor vrouwen in 2060 (grafiek A.3.1.2). Ten opzichte van de raming van 2012 betekent dit een toename van 7,8 jaar bij mannen en 7,1 jaar bij vrouwen. Het verschil in levensverwachting tussen mannen en vrouwen loopt volgens de prognose terug van 3,6 jaar op dit moment tot 2,8 jaar in 2060.

A.3.1.1 Toename in de levensverwachting bij geboorte per decennium

	Mannen		Vrouwen	
	Nederland	West-Europa[1)]	Nederland	West-Europa[1)]
1970–1980	1,7	2,0	2,7	2,6
1980–1990	1,4	2,2	0,9	2,2
1990–2000	1,7	2,7	0,5	2,1
2000–2010	3,2	2,8	2,1	1,9
2010–2020	2,0	1,9	1,1	1,7
2020–2030	1,8	1,8	1,3	1,6
2030–2040	1,6	1,6	1,7	1,4
2040–2050	1,5	1,5	1,6	1,3
2050–2060	1,4	1,4	1,4	1,2

[1)] Denemarken, Finland, Frankrijk, Duitsland, Italië, Noorwegen, Spanje, Zweden, Zwitserland en Engeland&Wales, bron Human Mortality Database. De trend van de West-Europese landen is doorgetrokken met een Lee-Carter-model.

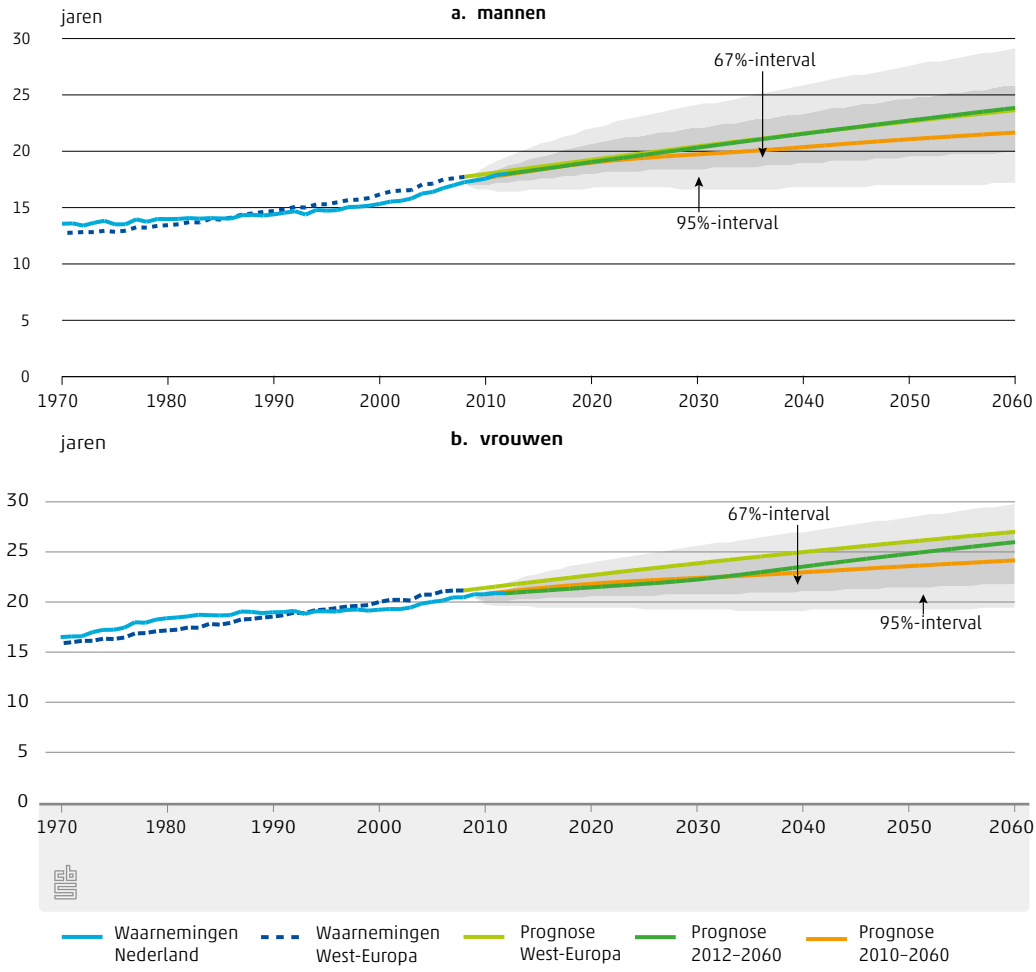
A.3.1.2 Levensverwachting bij geboorte



Voor vrouwen is een duidelijk niet-lineair patroon te zien in de toekomstige levensverwachting bij geboorte als gevolg van de toenemende sterfte door roken.

Vergeleken met het West-Europese gemiddelde zal de levensverwachting voor Nederlandse vrouwen de komende periode minder snel stijgen (tabel A.3.1.1). Wanneer de sterfte door roken gaat dalen, neemt de levensverwachting weer sterker toe. Daarnaast remt bij beide geslachten de stijging van de levensverwachting op de lange termijn af, doordat de minder gunstige sterftetrends bij de hoge leeftijden het beeld steeds meer gaan domineren naarmate mensen gemiddeld ouder worden.

A.3.1.3 Levensverwachting op leeftijd 65



Verwacht wordt dat de levensverwachting op 65-jarige leeftijd over de periode 2012 – 2060 voor mannen zal toenemen met 5,9 jaar en voor vrouwen met 5,1

jaar (grafiek A.3.1.3). Het verschil in levensverwachting tussen mannen en vrouwen zal naar verwachting afnemen van 2,9 jaar nu tot 2,1 jaar in 2060. Ook op oudere leeftijd is het niet-lineaire patroon in de levensverwachting van vrouwen overduidelijk te zien.

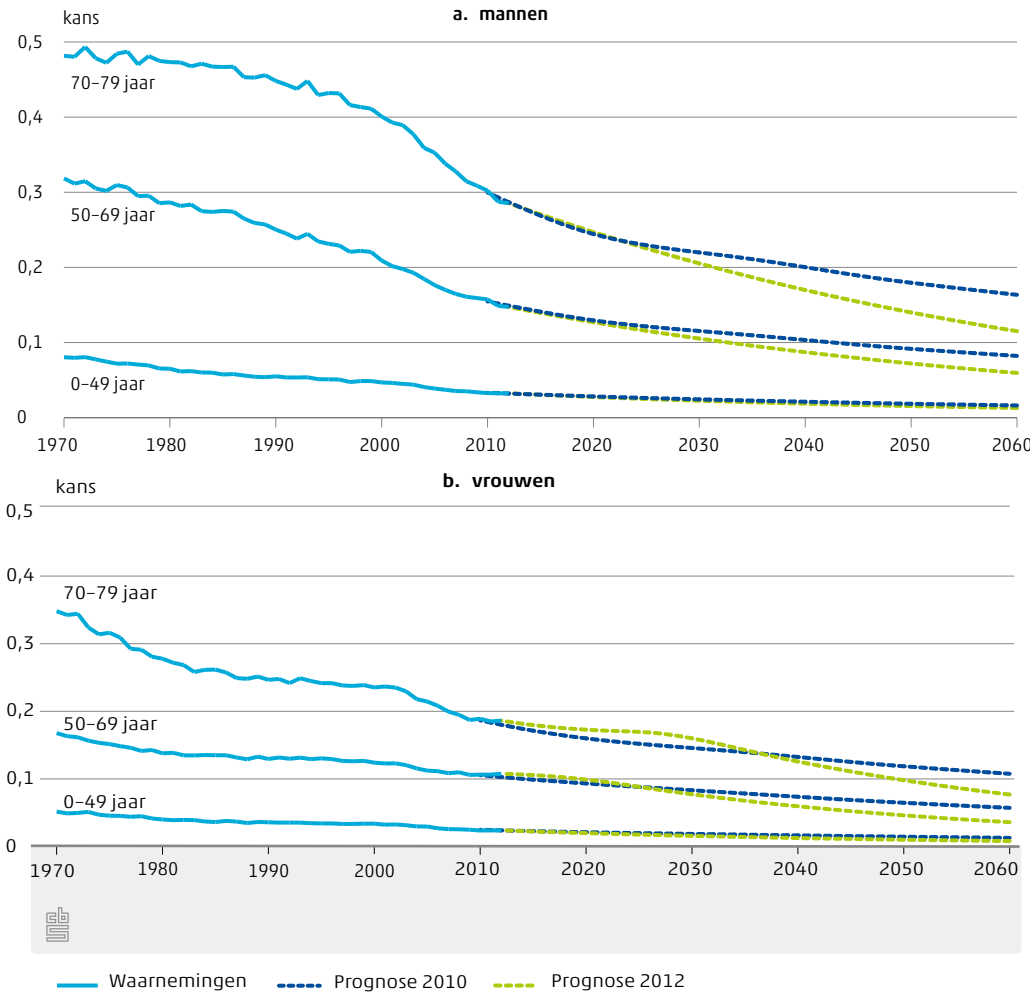
A.3.2 Bijstellingen ten opzichte van de vorige prognose

Vergeleken met de CBS-prognose 2010 komt de levensverwachting bij geboorte in 2060 voor mannen 2,6 jaar hoger uit en voor vrouwen 2,5 jaar hoger, als gevolg van de aangepaste methodologie en daaraan gerelateerde aannames (grafiek A.3.1.2). De nieuwe prognose ligt daarbij op de bovenste rand van het 67-procent-interval uit de prognose van 2010. Op de korte termijn zijn er voor mannen echter nauwelijks verschillen tussen de prognoses en voor vrouwen ligt de levensverwachting bij geboorte zelfs iets lager.

De levensverwachting op 65-jarige leeftijd komt voor mannen 2,2 jaar en voor vrouwen 1,8 jaar hoger uit vergeleken met de vorige prognose (grafiek A.3.1.3). Net als de levensverwachting bij geboorte, komt de levensverwachting op 65-jarige leeftijd uit op de bovenste rand van het 67-procent-interval van de prognose uit 2010.

Grafieken A.3.2.1 en A.3.2.2 tonen de sterftekansen voor mannen en vrouwen in verschillende leeftijdscategorieën volgens de huidige prognose en die uit 2010. De bijstellingen hebben vooral betrekking op de leeftijden boven de 50 jaar. De sterftekansen voor mannen liggen in deze leeftijdscategorieën onder het niveau uit de prognose in 2010. De sterftekans voor vrouwen in de leeftijd van 70-79 jaar ligt de eerste 20 jaar boven het niveau van de vorige prognose en is op oudere leeftijd ongeveer gelijk. Daarna treedt een sterkere afname op.

A.3.2.1 Sterftekansen van mannen en vrouwen 0-79 jaar per leeftijdsgroep

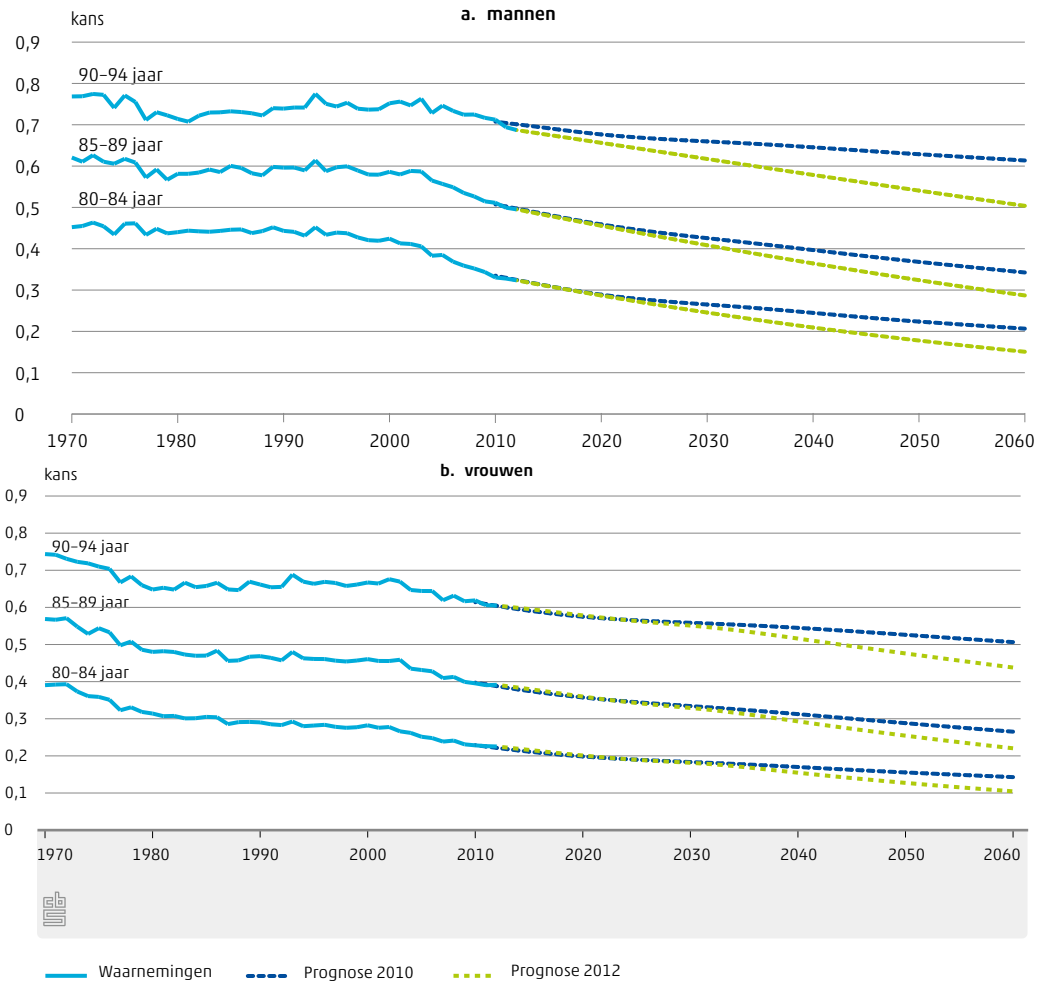


A.3.3 Evaluatie van de effecten van de verschillende componenten

Grafiek A.3.3.1 toont een vergelijking van de levensverwachting bij geboorte volgens de prognose en drie varianten. De eerste variant is een model waarin het effect van roken op de sterfte niet wordt meegenomen. Voor vrouwen krijgt de geprojecteerde levensverwachting in deze variant een meer lineair verloop doordat geen rekening wordt gehouden met de aanstaande omslag van stijgende naar dalende rookgerelateerde sterfte door veranderingen in het rookgedrag. Rond

2030 ligt de levensverwachting bij geboorte van vrouwen volgens deze variant ongeveer een half jaar lager dan volgens de prognose en in 2060 een kwart jaar lager. De levensverwachting voor mannen komt met deze variant over de hele prognoseperiode lager uit (tot een half jaar lager in 2060). Dit komt doordat de sterftetrend van de totale sterfte minder goed was dan voor de niet-rookgerelateerde sterfte ten opzichte van de Europese landen.

A.3.2.2 Sterftekansen van mannen en vrouwen 80-plus per leeftijdsgroep

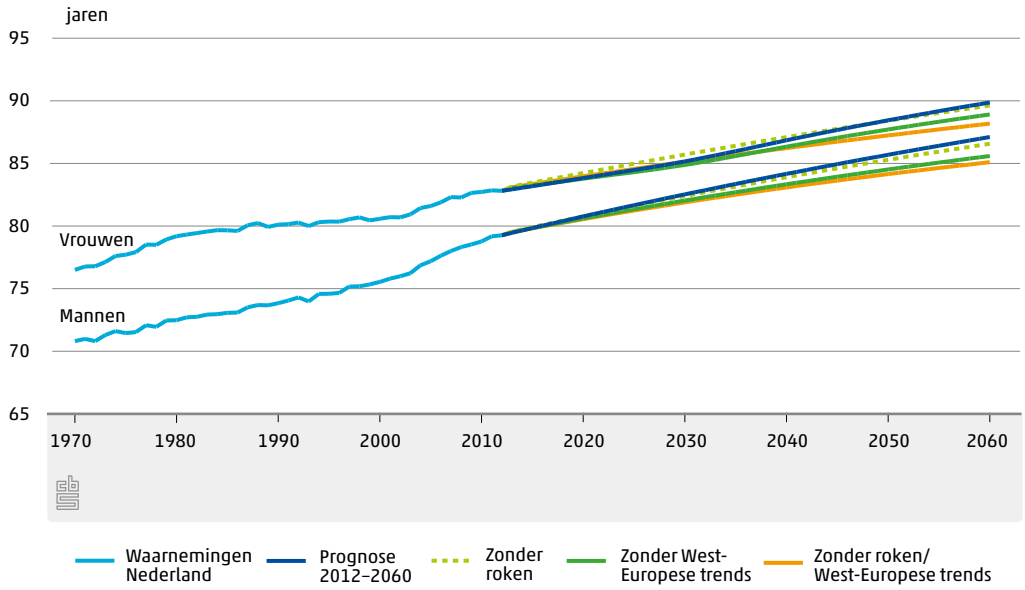


De tweede variant neemt wel het effect van roken mee, maar houdt geen rekening met de trends in andere West-Europese landen. Voor zowel mannen als vrouwen komt de levensverwachting met deze variant lager uit. In 2060 bedraagt het verschil 1,5 jaar voor mannen en 1,0 jaar voor vrouwen. De laatste variant is een model waarin noch roken noch de West-Europese trends zijn meegenomen. Dit

komt neer op een Lee-Carter-schatting op basis van alleen de Nederlandse sterftecijfers sinds 1970. Deze variant geeft de laagste prognose van de levensverwachting: in 2060 ligt deze 2,0 jaar lager voor mannen en 1,7 jaar lager voor vrouwen. De uitkomsten van deze variant liggen dicht bij de CBS-prognose uit 2010.

Bovenstaande laat zien dat voor het berekenen van de levensverwachting bij geboorte het meenemen van Europese landen belangrijker is dan het onderscheid tussen roken en niet-roken. Het meenemen van de invloed van rookgedrag heeft vooral invloed op de kortetermijnontwikkeling bij vrouwen. Het relatieve belang van de componenten verschilt echter per leeftijd.

A.3.3.1 Levensverwachting bij geboorte, alternatieve methoden

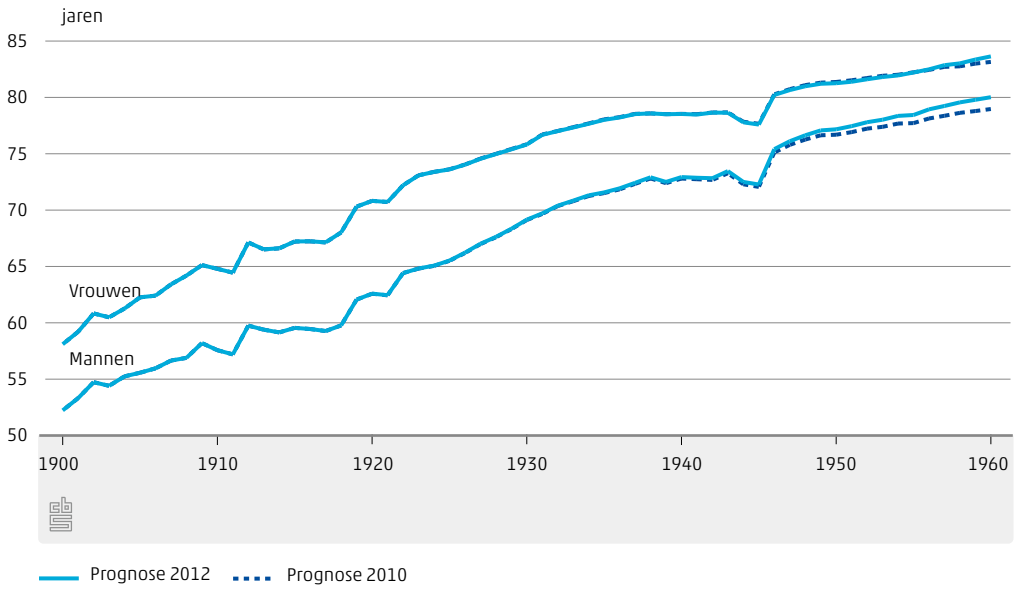


A.3.4 Cohort-levensverwachting

De periode-levensverwachting bij geboorte geeft een sterke onderschatting van de werkelijke levensduur, omdat in de berekening wordt verondersteld dat de sterftekansen in het geboortejaar gedurende het hele leven gelden. Profijt van bijvoorbeeld medische vooruitgang wordt dus buiten beschouwing gelaten. Uit de geprognosticeerde sterftekansen tot en met 2060 kan echter de cohort-levensverwachting worden berekend voor generaties met een geboortejaar tot en met 1960. Van de jongere generaties zal een belangrijk deel na 2060 nog in leven zijn, zodat zonder extra aannames op basis van de prognose niets kan worden gezegd over de levensverwachting van deze groepen.

Grafiek 3.4.1 toont de cohort-levensverwachting bij geboorte voor de geboortecohorten vanaf 1900. De generatie uit 1900 had een werkelijke levensverwachting van ongeveer 52 jaar voor mannen en 58 jaar voor vrouwen. Mannen die in 1960 zijn geboren leven naar verwachting gemiddeld 80 jaar en vrouwen bijna 84 jaar. Dit betekent dat de levensverwachting ten opzichte van de generatie uit 1900 voor mannen 28 jaar hoger ligt en voor vrouwen 26 jaar; ongeveer een half jaar winst per cohort. De toename is voor een groot deel het gevolg van dalende sterftekansen op jonge leeftijden.

A.3.4.1 Cohort-levensverwachting bij geboorte



De periode-levensverwachting in 1960 was 71,4 jaar voor mannen en 75,3 jaar voor vrouwen. Het verschil met de cohort-levensverwachting is daarmee voor zowel mannen als vrouwen ruim 8,5 jaar. Dit betekent dat de personen geboren in 1960 naar verwachting 8,5 jaar langer zullen leven dan de periodelevensverwachting aangeeft.

Vergeleken met de prognose uit 2010 is de geschatte cohort-levensverwachting bij geboorte in 1960 voor mannen 1,0 jaar hoger en voor vrouwen 0,8 jaar hoger. Bij oudere geboortecohorten zijn de verschillen kleiner. Omdat de bijstelling van de sterfteprognose vooral de lange termijn betreft, werkt het effect daarvan met name door in de levensverwachting van jongere geboortecohorten.

A.4. Conclusie

De nieuwe methode van de CBS-sterfteprognose 2012 houdt rekening met de sterfteontwikkelingen in andere West-Europese landen. Bovendien wordt informatie over ontwikkelingen in roken er op een systematische wijze in betrokken. Volgens de nieuwe prognose stijgt de periode-levensverwachting bij geboorte tussen 2012 en 2060 met 7,8 jaar voor mannen en met 7,0 jaar voor vrouwen, om dan uit te komen op respectievelijk 87,1 en 89,9 jaar. Ten opzichte van de vorige prognose is dit een verhoging van 2,6 jaar voor mannen en een verhoging van 2,5 jaar voor vrouwen. Op de korte termijn zijn er echter nauwelijks verschillen tussen de twee prognoses.

Bijlage A Methode

Deze bijlage beschrijft de methode van de CBS-sterfteprognose 2012. Alle berekeningen zijn uitgevoerd in het programma R (www.r-project.org/).

Naast gegevens over Nederland zijn gegevens opgenomen over Denemarken, Duitsland, Engeland en Wales, Finland, Frankrijk, Italië, Noorwegen, Spanje, Zweden en Zwitserland. Voor deze landen zijn de gegevens over longkanker verkregen van de World Health Organization (WHOSIS, 2012). De gegevens over de totale sterfte en populatie zijn overgenomen uit de Human Mortality Database (HMD, 2012). De gegevens voor Nederland zijn verkregen via Statline. De data zijn uitgesplitst naar jaar, geslacht en 5-jaars leeftijdsgroepen. Voor het omzetten van de sterftekansen van 5-jaars leeftijdsgroepen naar 1-jaarsgroepen is de 1-jaars-sterftekans voor de jaren 2010 en 2011 van Statline gebruikt.

Notatie

$m_{OTH}(x,t)$ sterftecijfer overige doodsoorzaken, leeftijd x en jaar t

$m^0_{OTH}(x,t)$ niet- rookgerelateerde sterftecijfer overige doodsoorzaken, leeftijd x en jaar t

$m_{LC}(x,t)$ sterftecijfer longkanker, leeftijd x en jaar t

Stap 1: Schatting overig rookgerelateerde en overig niet-rookgerelateerde sterfte

In stap 1 wordt de niet-rookgerelateerde sterfte geschat met behulp van de methode van Rostron (2010). Hiervoor wordt een fractie berekend met behulp van de longkankersterfte en de coëfficiënten uit het artikel (tabel A.1).

A.1 Coëfficiënten voor longkankersterfte (Rostron, 2010) en longkankersterfte voor niet-rokers (Thun et al., 1997) per leeftijdsgroep

	Coëfficiënten		Longkankersterfte niet-rokers	
	Mannen	Vrouwen	Mannen	Vrouwen
	per 1 000		per 100 000	
50–54	0,348	0,707	5,5	5,8
55–59	0,174	0,510	5,3	7,2
60–64	0,113	0,382	11,6	12,3
65–69	0,079	0,218	21,5	16,7
70–74	0,060	0,137	34,9	30,5
75–79	0,046	0,061	52,0	32,5
80+	0,028	0,013	89,2	57,6

De coëfficiënten uit Rostron (2010) zijn het resultaat van een regressie van de (log) overige sterfte op de longkankersterfte en enkele sets van dummyvariabelen voor leeftijd, jaar en het land. De regressie ziet er als volgt uit (let op: de notatie uit het artikel van Rostron is hier overgenomen):

$$ln(M_O) = \beta_x X_x + \beta_t X_t + \beta_c X_c + \beta_{ct}(t \times X_c) + \beta_{xt}(t \times X_x) + \beta_L M_L + \beta_{Lt}(t \times M_L) + \beta_{Lx}(X_x \times M_L) \quad (1)$$

waarin

M_O : sterftecijfer overige doodsoorzaken
M_L : sterftecijfer longkanker
X_x : set van dummyvariabelen voor iedere leeftijdsgroep
X_t : set van dummyvariabelen voor ieder jaar
X_c : set van dummyvariabelen voor ieder land
t : jaar (als een lineaire variabele)

Aan de hand van deze relatie kan worden uitgerekend hoeveel sterfgevallen er zouden zijn gegeven de waargenomen longkankersterfte in de populatie en hoeveel sterfgevallen er zouden zijn als de longkankersterfte het niveau van een niet-rokende populatie zou hebben. Dit verschil, gedeeld door het geschatte aantal sterfgevallen gegeven de waargenomen longkankersterfte in de populatie, geeft

het aandeel rookgerelateerde sterfte. Deze procedure is gelijk aan het toepassen van de volgende formule:

$$A_{OTH}(x,t) = 1 - e^{-\beta'_{LC}(x)\left(m_{LC}(x,t) - \lambda^{O}_{LC}(x)\right)}$$

waarbij β'_{LC} de som van β_L, β_{Lt}, en β_{Lx} (naar leeftijdsgroep) uit vergelijking (1) is en λ^{O}_{LC} de longkankersterfte voor niet-rokers (Thun et al., 1997, zie tabel A.1). Het aandeel $A_{OTH}(x,t)$ geeft dus per leeftijdsgroep en per jaar het gedeelte van de overige sterfte dat gerelateerd is aan roken.

De fractie wordt vervolgens gebruikt om de overige doodsoorzaken op te splitsen in niet-rookgerelateerd en rookgerelateerd met behulp van de volgende formule:

$$m^{O}_{OTH}(x,t) = \left(1 - A_{OTH}(x,t)\right) m_{OTH}(x,t)$$

De berekening van de fracties en de opsplitsing in rookgerelateerd en niet-rookgerelateerd worden voor ieder land in de prognose berekend voor de jaren 1970-2011 (of het meest recent beschikbare jaar). Doordat de coëfficiënten uit Rostron (2010) beschikbaar zijn voor de leeftijden 50–54, 55–59,...,80+, wordt de fractie voor de leeftijdsgroep 80+ toegepast op de leeftijdsgroepen 80–85, 85–90, 90–95 en 95+, en de coëfficiënt voor de leeftijdsgroep 50–54 op de leeftijden 40-45 en 45-50 voor mannen en 45-50 voor vrouwen. Alle sterfte onder leeftijd 40 voor mannen en leeftijd 45 voor vrouwen wordt gerekend als niet-rookgerelateerd (fractie is hier dus gelijk aan nul), omdat op de jongere leeftijden rookgerelateerde sterfte nauwelijks voorkomt. De waargenomen fracties worden weergegeven in grafiek C.1 in bijlage C.

Stap 2: Projectie longkankersterfte

Stap 2 is de projectie van de longkankersterfte met behulp van een Age-Period-Cohort-model (APC-model). De longkankerdata uit de periode 1950-2009 wordt hiervoor omgezet naar vijfjaarsperioden. De leeftijden voor mannen zijn 40–44, 45–49,.., 80+ voor mannen en 45–49,..., 80+ voor vrouwen.

Het APC-model heeft de volgende vorm:

$$m_{LC}(x,c) = exp(a_x + b_t + c_{t-x} + dt)$$

De parameters a_x, b_t en c_{t-x} stellen de leeftijd, periode en cohort-effecten voor en is een driftterm voor de periode. Deze driftterm geeft het lineaire effect weer van de periode. Om het model identificeerbaar te maken, worden de eerste en

laatste parameters van de periode en cohort-effecten op nul gezet. Het model wordt geschat met behulp van een poisson-regressie.

Voor mannen is puur de projectie van de afname nodig. Daarom wordt het APC-model eerst geschat om de data vanaf het maximale cohort te kunnen bepalen en vervolgens nogmaals met deze nieuwe dataset (data vanaf cohort 1910) om de afnemende trend te berekenen. De projectie voor mannen heeft de volgende vorm:

$$\widetilde{m}_{LC}(x,t) = exp\left(ln\left(m_{LC}(x, 2005/2009)\right) + \hat{d}\left(t - code(2005/2009)\right)\right)$$

Voor vrouwen is de methode iets ingewikkelder, omdat eerst nog een stijging in de longkankersterfte wordt verwacht als gevolg van de toename van het aandeel rokers in het verleden, en daarna pas een daling. Het model voor de stijging heeft de volgende vorm:

$$\widetilde{m}_{LC}(x,t) = exp\left(\hat{a}_x + \hat{b}_t + \hat{c}_{t-x} + \hat{d}t\right)$$

Vervolgens wordt het omslagpunt vanwaar de daling inzet uitgerekend met behulp van de projectie van mannen. Het jaar waarin $\widetilde{m}_{LC}(x,t)$ voor vrouwen per leeftijd hoger is dan de waarde van mannen is het omslagpunt. Wanneer de waarde voor vrouwen hoger is dan de waarde van mannen in het jaar ervoor, wordt de waarde voor vrouwen aangepast aan die van mannen in het jaar ervoor. De daling verloopt vervolgens met dezelfde trend als bij mannen vanaf het omslagpunt. De geschatte driftterm $\hat{d}$ uit het model van mannen wordt dus toegepast vanaf het omslagpunt bij de vrouwen.

De keuze voor het omslagpunt volgt uit de observaties bij de jongere leeftijden waar de longkankersterfte van mannen en vrouwen elkaar al hebben gekruist, op basis van het epidemiemodel van Lopez et al. (1994) en op basis van een vergelijking met Engeland en Denemarken. Het maximum aan de toename van vrouwen wordt net na de kruising met mannen ingesteld omdat er door het gebruik van vijfjaarsdata anders een enorme toename ten opzichte van mannen kan optreden als het punt net voor de kruising al bijna gelijk is.

Als laatste wordt de 5-jaarsprojectie omgezet naar een 1-jaarsprojectie met behulp van een spline en gefit aan de laatste waarneming (2011). In principe kan de projectiemethode op alle landen worden toegepast , maar voor de sterfteprognose van Nederland is dit alleen noodzakelijke voor de longkankersterfte in Nederland.

Stap 3: Projectie overig niet-rookgerelateerde sterfte

Stap 3 is de projectie van de niet-rookgerelateerde sterfte met behulp van een Li-Lee-methode (Li en Lee 2005). Hiervoor wordt meerdere keren een Lee-Carter-model (Lee en Carter 1992) geschat met behulp van Singular Value Decomposition. Alle schattingen worden gedaan voor mannen en vrouwen afzonderlijk.

De gebruikte periode in deze stap is 1970-2011. Doordat voor enkele landen de meest recente gegevens (nog) niet beschikbaar zijn, worden de sterftecijfers voor de totale groep over 2009–2011 geschat met het Lee-Carter-model. Daarna wordt de Li-Lee-methode toegepast op deze aangevulde dataset.

Als eerste wordt het model geschat met de Nederlandse data:

$$ln\left(m^{O}_{OTH}(x,t)\right) = a(x) + b(x)k(t) + e(x,t)$$

Dit geeft $a(x)$, de leeftijdsverdeling van de sterftecijfers die constant is over de tijd. $b(x)$ is een leeftijdsspecifieke constante die de verandering over de tijd weergeeft en $k(t)$ de onderliggende tijdsverandering. $e(x,t)$ is de fout. Uit deze schatting is alleen $a(x)$ nodig.

Vervolgens wordt het Lee-Carter-model geschat voor de totale groep landen:

$$ln\left(m^{O}_{OTH}(x,t)\right) = A(x) + B(x)K(t) + E(x,t)$$

Dit geeft $B(x)$, de verbeteringen in de sterftecijfers per leeftijdsgroep, en $K(t)$, de algemene trend. $K(t)$ wordt vervolgens geëxtrapoleerd met een random walk met drift-model.

Als laatste wordt een schatting gemaakt voor de afwijking van Nederland ten opzichte van de groep. Hiervoor wordt eerst de schatting met de coëfficiënten van hiervoor uitgerekend en een Lee-Carter-model opgezet voor het verschil ten opzichte van de Nederlandse observaties:

$$a(x) + B(x)K(t) - ln\left(m^{O}_{OTH}(x,t)\right) = b^{res}(x)k^{res}(t) + \varepsilon^{res}(x,t)$$

Hierin wordt het leeftijdspatroon van Nederland genomen ($a(x)$) en niet dat van de totale groep, omdat dit element niet tot divergentie leidt en het dus niet voor alle landen hetzelfde hoeft te zijn. De schatting geeft $b^{res}(x)$ en $k^{res}(t)$. $k^{res}(t)$ wordt geëxtrapoleerd met een AR(1)-model. Indien de coëfficiënt hiervan groter is dan 1 (oftewel, er is divergentie op de lange termijn), dan wordt $k^{res}(t)$ voor de toekomst constant gehouden aan de schatting van het laatste waarneemjaar.

De toekomstige waarden voor de niet-rookgerelateerde sterfte overige doodsoorzaken worden als volgt berekend:

$$\tilde{m}^{O}_{OTH}(x,t) = exp\Big(ln\Big(\tilde{m}^{O}_{OTH}(x,2011)\Big) + B(x)K(t) - K(2011) + b^{res}(x)\big(k^{res}(t) - k^{res}(2011)\big)\Big)$$

Stap 4: Berekenen projectie totale sterfte

Stap 4 is het berekenen van de projectie van de totale sterfte. Hiervoor wordt eerst de fractie nietrookgerelateerde sterfte van de overige doodsoorzaken uitgerekend voor de projectieperiode met behulp van de geprojecteerde longkankersterfte:

$$\tilde{A}_{OTH}(x,t) = 1 - e^{-\beta'_{LC}(x)\left(m_{LC}(x,t) - \lambda^{O}_{LC}(x)\right)}$$

De geprojecteerde fracties worden weergegeven in grafiek C.1 in bijlage C.

Vervolgens kan de totale sterfte overige doodsoorzaken uitgerekend worden met de geprojecteerde rookgerelateerde fractie en de geprojecteerde niet-rookgerelateerde sterfte overige doodsoorzaken:

$$\tilde{m}_{OTH}(x,t) = \tilde{m}^{O}_{OTH}(x,t) / \Big(1 - \tilde{A}_{OTH}(x,t)\Big)$$

De laatste stap is het optellen van de sterfte overige doodsoorzaken en de longkankersterfte:

$$\tilde{m}(x,t) = \tilde{m}_{OTH}(x,t) + \tilde{m}_{LC}(x,t)$$

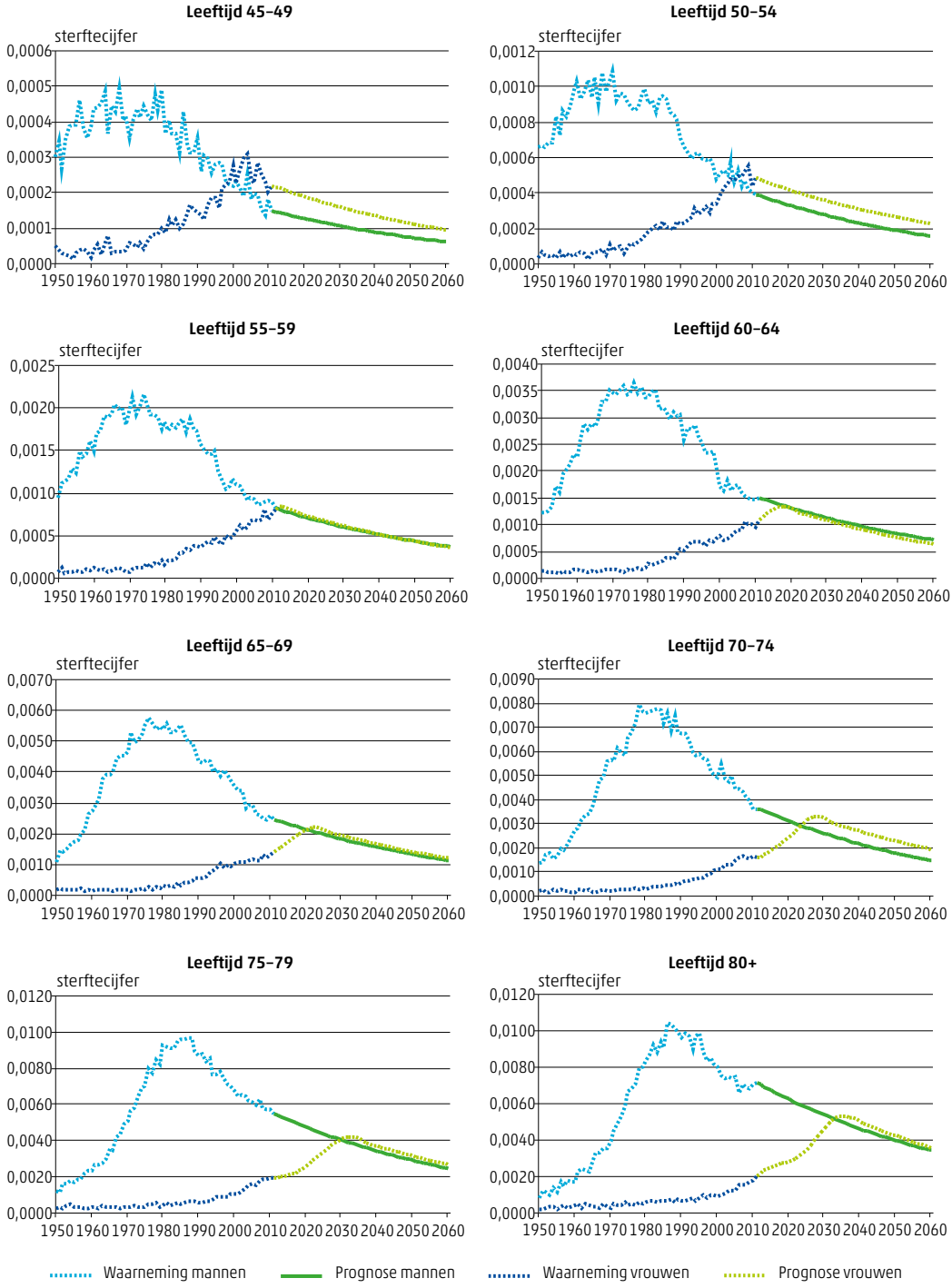
Omzetten naar 1-jaars-sterftekansen

Het patroon van de sterftekansen in 2010–2011 wordt toegepast op het leeftijdsverloop binnen de vijfjaarscategorieën. Vervolgens worden de log-sterftekansen glad gemaakt met een 5-punts lopend gemiddelde. De kansen voor de twee laagste leeftijdscategorieën worden daar niet bij betrokken.

Bijlage B
Projectie van de longkanker

Grafiek B.1 geeft de longkankerprojectie weer. Hierin is te zien dat de toekomstige longkankersterfte voor mannen zal afnemen en voor vrouwen eerst zal toenemen, maar daarna ook zal afnemen nadat de trend van mannen is gekruist.

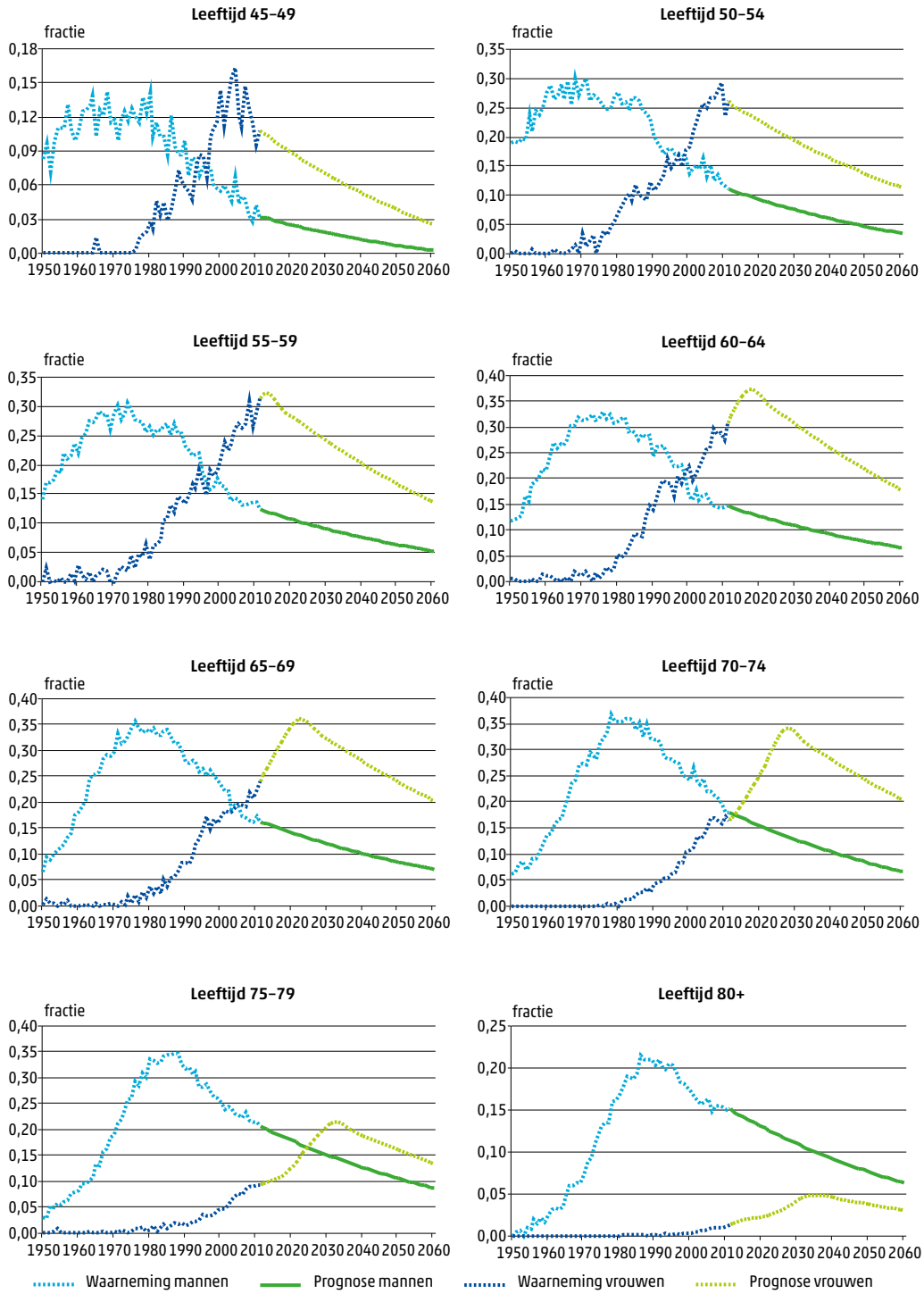
B.1 Projectie longkankersterfte per leeftijdsgroep



Bijlage C
Fractie rookgerelateerde sterfte

Grafiek C.1 geeft de fractie rookgerelateerde sterfte weer van de overige doodsoorzaken per leeftijdsgroep voor mannen en vrouwen in de periode 1950–2060. Deze fractie wordt berekend op basis van de waargenomen en geprojecteerde longkankersterfte en de methode van Rostron (2010).

C.1 Projectie fractie rookgerelateerde sterfte van de overige doodsoorzaken per leeftijdsgroep



Literatuur

Barendregt J.J., Looman, C.W.M. en Bronnum-Hansen, H. (2002). Comparison of cohort smoking intensities in Denmark and the Netherlands. *Bulletin of the World Health Organization* 80: 26–32.

Bonneux, L.G.A., Looman, C.W.N. en Coebergh, J.W. (2003). Sterfte door roken in Nederland: 1,2 miljoen tabaksdoden tussen 1950 en 2015. *Nederlands Tijdschrift voor Geneeskunde* 147: 917–921.

Van Duin, C., de Jong, G., Stoeldraijer, L. en Garssen, J. (2011). Bevolkingsprognose 2010-2060: model en veronderstellingen betreffende de sterfte. *Bevolkingstrends* 59(2): 28–41.

HMD (2012). *The Human Mortality Database.* Berkeley: University of California and Rostock: Max Planck Institute for Demographic Research (http://www.mortality.org/).

Janssen, F., Mackenbach, J.O. en Kunst, A.E. (2004). Sterftetrends onder ouderen in zeven Europese landen van 1950-1999. *Bevolkingstrends* 52(3): 40–51.

Janssen, F. en Kunst, A. (2007). The choice among past trends as a basis for the prediction of future trends in old-age mortality. *Population Studies* 61(3): 315–326.

Janssen, F. en Kunst, A. (2010). De toekomstige levensverwachting. In: A.H.P. Luijben & G.J. Kommer (eds.), *Tijd en toekomst; deelrapport van de VTV 2010 Van gezond naar beter.* RIVM-rapport 270061008, Houten: Bohn Stafleu Van Loghum, 13–20.

Janssen, F., van Wissen, L. en Kunst, A. (te verschijnen): A new mortality projection methodology-evaluation and application to the Netherlands, *Demography*.

Jong, A. de, en van der Meulen, A. (2005). Prognose van sterfte naar doodsoorzaken: model en veronderstellingen. *Bevolkingstrends* 53(2): 50–62.

Lee, R. D. en Carter, L. R. (1992). Modelling and forecasting US mortality. *Journal of the American Statistical Association* 87(419): 659–671.

Li, N. en Lee, R. (2005). Coherent mortality forecasts for a group of populations: an extension of the Lee-Carter method. *Demography* 42(3): 575–94.

Lopez, A.D., Collishaw, N.E. en Piha, T. (1994). A descriptive model of the cigarette epidemic in developed countries. *Tobacco Control* 3(3): 242–247.

Mackenbach, J. en Garssen, J. (2011). Renewed progress in life expectancy: The case of the Netherlands. In: Crimmins, E.M., S.H. Preston en B. Cohen (eds.). *International differences in mortality at older ages: Dimensions and sources.* National Academies Press, Washington DC, 369– 384.

Meulen, A. van der, en Janssen, F. (2007). Achtergronden en berekeningswijzen van CBS-overlevingstafels. *Bevolkingstrends* 55(3): 66–77.

Oeppen, J. en Vaupel, J.W. (2002). Broken limits to life expectancy. *Science* 296(5570): 1029–1031.

Peto, R., Lopez, A.D., Boreham, J., Thun, M., Heath, C. Jr. en Doll, R. (1996). Mortality from smoking worldwide. *British Medical Bulletin* 52(1): 12–21.

Rostron, B. (2010). A modified new method for estimating smoking-attributable mortality in high-income countries. *Demographic Research* 23: 399–420.

Stoeldraijer, L, van Duin, C., van Wissen, L. en Janssen, F. (ingediend): Sensitivity of projected future life expectancy at birth and at age 65 due to different mortality forecasting methods. *Demographic Research.*

Thun, M.J., Day-Lally, C., Myers, D.G., Calle, E.E., Flanders, W.D., Zhu, B.P., Namboodiri, M.M. en Heath, C.W. (1997). Trends in tobacco smoking mortality from cigarette use in Cancer Prevention Studies I (1959 through 1965) and II (1982 through 1988). In: *Changes in cigarette-related disease risks and their implications for prevention and control.* Smoking and Tobacco Control Monograph no. 8. Bethesda, MD: National Cancer Institute, 305–382.

WHOSIS (2012). *World Health Organization Statistical Information System:* http://www.who.int/whosis/mort/download/en/ (update 9 juli 2012).

Wilmoth, J.R. (1995). Are mortality projections always more pessimistic when disaggregated by cause of death? *Mathematical Population Studies* 5(4): 293–319.

Wilmoth, J. R. (2000). Demography of longevity: past, present, and future trends. *Experimental Gerontology* 35: 1111–1129.

English summary

Mortality forecasting in the context of non-linear past mortality trends: an evaluation

The aim of the current PhD research was to evaluate mortality forecasting in the context of non-linear past mortality trends. Having accurate and high-quality mortality forecasts has become increasingly important due to the general increase in life expectancy and the social consequences of this (ageing, health care, housing, social security, pensions).

The majority of current methods of mortality forecasting are extrapolative in nature; that is, they extend a past mortality trend by assuming that both age patterns and trends remain regular over time. Compared with other forecasting approaches, the extrapolative methods are highly objective; i.e., they reduce the role of subjective judgment involved in mortality forecasting. However, particularly in situations in which past trends have been non-linear, like in the Netherlands, the use of an objective extrapolative method will be more problematic. Among the potential approaches for improving mortality forecasts when the trends are non-linear trends are making explicit adjustments for the distorting effects of smoking on mortality trends, and using the more linear trends of other countries as the underlying long-term mortality trend. However, both of these approaches require the inclusion of more subjective information in the mortality forecast. Whether only "objective" extrapolation methods should be employed even in cases of non-linearity, or whether it is preferable to include additional information, even if doing so introduces additional subjectivity, is an important topic of debate. To address this question, it is essential to evaluate mortality forecasting approaches in the context of non-linear past mortality trends.

Most previous studies employed purely quantitative evaluations of mortality forecasting models that focused solely on their accuracy, or they evaluated purely objective forecasting approaches that are less relevant for non-linear trends. Moreover, most of these studies did not evaluate the sensitivity of future mortality to explicit assumptions; i.e., to the specific choices that are explicitly stated in a method, such as the choices of the length of the fitting period and of the jump-off rates.

This PhD research evaluates mortality forecasting methods and forecasting approaches, both from a quantitative and qualitative perspective. Furthermore, the sensitivity of future mortality based on different explicit assumptions is assessed. Moreover, different elements of a mortality forecasting approach that deals with non-linear past mortality trends are evaluated (e.g., the forecasting of smoking-attributable mortality, a model that forecasts mortality coherently).

This PhD thesis includes a careful study of past mortality trends. Although the focus of the thesis is mainly on the Netherlands, mortality trends in other Northwest

European countries are also studied to create a broader empirical basis. The emphasis here is on how different mortality trends (especially linear versus non-linear trends) were affecting the performance of different mortality forecasting methods, both quantitatively and qualitatively.

This PhD thesis not only contributes to the debate on the degree of subjectivity in mortality forecasting, but the findings of this research are used to evaluate, validate, and further improve the mortality forecasts of Statistics Netherlands.

The study is guided by the following research questions:
1) In a context in which mortality trends are non-linear, how does the choice of the mortality forecasting method and the explicit assumptions affect future forecasted mortality?
2) How can future levels of smoking-attributable mortality be formally estimated?
3) Which model should be used when the goal is to forecast mortality coherently , namely by taking into account the mortality experiences of other countries?
4) How can mortality forecasts be adjusted to take into account more recently observed data?

After the introductory chapter, the empirical chapters 2 through 5 address the research questions above. In Chapter 6, the main findings of the PhD thesis as a whole are summarised and discussed.

Chapter 2 reviewed the different mortality forecasting methods and their assumptions in Europe, and assessed their impact on projections of future life expectancy for the Netherlands. More specifically, (i) the current methods used in official mortality forecasts in Europe were reviewed; (ii) the outcomes and the assumptions of different projection methods within the Netherlands were compared; and (iii) the outcomes of different types of methods for the Netherlands using similar explicit assumptions, including the same historical period, were compared. The findings of a review of the current methods indicated that most statistical offices in Europe use simple linear extrapolation methods, but that countries with less linear trends employ other approaches or different assumptions. The approaches employed in the Netherlands include the use of explanatory models, the separate projection of smoking- and non-smoking-related mortality, and the projection of the age profile of mortality. There are, however, clear differences in the explicit assumptions used in these approaches, and the resulting e0 in 2050 varies by approximately six years. Using the same historical period (1970-2009) and the observed jump-off rates, the findings generated by different methods result in a range of 2.1 years for women and of 1.8 years for men. For e65, the range is 1.4 years for men and 1.9 years for women. These findings

suggest that the choice of explicit assumptions is more important than the choice of the forecasting method.

In Chapter 3, a formal estimation of future levels of smoking-attributable mortality up to 2050 was proposed for the total national populations of England and Wales, Denmark, and the Netherlands. An update and an extension of the descriptive smoking epidemic model were provided in the estimation. A two-step method for estimating the future smoking-attributable mortality fraction was presented: (i) lung cancer mortality was projected by extrapolating age-period-cohort trends (1950-2009), while using the observed convergence among men and women of smoking prevalence and past lung cancer mortality levels as input; and (ii) other causes of death attributable to smoking were added by applying a simplified version of the indirect Peto–Lopez method to the projected levels of lung cancer mortality. The smoking-attributable mortality fractions (SAF) for men in 2009 were found to be 19% (44,872 deaths) in England and Wales, 22% (5,861 deaths) in Denmark, and 25% (16,385 deaths) in the Netherlands. In our projections, these fractions declined to 6%, 12%, and 14%, respectively, in 2050. The SAF for women peaked at 14% (38,883 deaths) in 2008 in England and Wales, and is expected to peak in 2028 in Denmark (22%) and in 2033 in the Netherlands (23%). By 2050, declines to 9%, 17%, and 19%, respectively, are foreseen. The use of different indirect methods for estimating the SAF in 2050 yielded ranges of 1–8% in England and Wales, 8–13% in Denmark, and 11–16% in the Netherlands for men; and of 7–16%, 12–26%, and 13–31%, respectively, for women.

In Chapter 4, different coherent forecasting methods were evaluated in terms of their accuracy (fit to historical data), robustness (stability across different fitting periods), subjectivity (sensitivity to the choice of the group of countries), and plausible outcomes (smooth continuation of trends from the fitting period). The coherent forecasting methods we investigated were as follows: the co-integrated Lee-Carter (CLC) method, the Li-Lee (LL) method, and the coherent functional data (CFD) method. The methods were applied to data from France, Italy, the Netherlands, Norway, Spain, Sweden, and Switzerland in order to generate forecasts up to 2050; and the results were compared to those of the individual Lee-Carter (LC) method. Of the three coherent forecasting methods evaluated, the CFD method was found to perform best on the accuracy measures. However, after the CFD method's higher number of parameters was controlled for, the differences disappeared. Both the CLC and the LL methods were found to be robust. The CLC method (for women) and the LL method (for men) were shown to be the least sensitive to the choice of the group of countries. The LL method generated the most plausible results, as it showed a convergence of future life expectancy levels that was in line with the fitting period and the smooth pattern of age-specific

improvements. This finding could imply that the LL method, which performed best in terms of robustness, subjectivity, and plausibility, provided a better fit than the CFD method, which had better accuracy (model fit).

Finally, in Chapter 5, six different options for the jump-off rates were evaluated and their effects on the robustness and the accuracy of the mortality forecasts were examined. As the jump-off rates, we examined the use of the model values, the observed values in the last year, and the averaged over the last couple of years for data from eight European countries (Belgium, Finland, France, the Netherlands, Norway, Spain, Sweden, and United Kingdom, 1960-2014 period). The future life expectancy at age 65 was calculated for different fitting periods and jump-off rates using the Lee-Carter model, and the accuracy (mean absolute error) and the robustness (standard deviation of the change in projected e65) of the results were examined. The findings of the analysis showed that which jump-off rates were chosen clearly influenced the accuracy and robustness of the mortality forecast, albeit in different ways. For most of the countries, using the last observed values as the jump-off rates resulted in the most accurate method, due in part to the estimation error of the model in recent years. The most robust method was obtained when using an average of observed years as jump-off rates. The more years that were averaged, the higher the degree of robustness; but the level of accuracy decreased with more years averaged. These results imply that the best strategy for matching mortality forecasts to the most recently observed data depends on the goal of the forecast, the country-specific past mortality trends, and the model fit.

The results of the empirical chapters of this thesis show that for countries with non-linear mortality trends, like the Netherlands, mortality forecasting approaches and assumptions were used that differ from the simple linear extrapolation methods that are commonly used by national statistical offices. The choice of the explicit assumptions proved more important than the choice of the forecasting approach. Because the inclusion of additional information on the smoking epidemic or on the mortality experiences of other countries is generally known to diminish the effect of the length of the historical period, doing so is expected to result in a more robust forecast. One way that additional information on the smoking epidemic could be included was by separately forecasting smoking-attributable mortality. The age-period-cohort methodology developed in this thesis – informed by assumptions derived from the smoking epidemic model and a careful study of past trends – proved valid for this purpose. When the mortality experiences of other countries by means of coherent mortality forecasting is included, it was found that the Li-Lee method outperformed the co-integrated Lee-Carter method and the coherent functional data method in terms of robustness,

subjectivity, and plausibility. Another important explicit assumption was the choice of the jump-off rates. It was found that the choice of the jump-off rates clearly influenced the accuracy and the robustness of the mortality forecast, albeit in different ways. Therefore, it was concluded that which strategy was best depended on the goal of the forecast, the country-specific past mortality trends, and the model fit.

All in all, this PhD thesis found that forecasting mortality when the trends were non-linear involved more than the direct (linear) extrapolation of past mortality trends. Even though including additional information (like data on the smoking epidemic and/or on the mortality experiences of other countries) made the method more subjective, it also made the method less dependent on an important explicit assumption: namely, the historical period. This insight is important, because this PhD thesis has also demonstrated that explicit assumptions play an essential role in mortality forecasts. However, before any information is added to mortality forecasting models, a careful examination of past trends should be undertaken, and a careful assessment of the pros and cons of its inclusion should be performed.

The results of this PhD thesis have a number of implications for mortality forecasting in general. First, the strong effect of explicit assumptions (including the main group of countries that will be included in coherent mortality forecasting) should be underlined. A more important role must be assigned to explicit assumptions than is currently the case. Ideally, stochastic forecasts should also incorporate the levels of uncertainty associated with different explicit assumptions. Furthermore, new forecasting methods should be evaluated based not only on their accuracy, but on other more qualitative criteria, such as the robustness, subjectivity, and plausibility of their outcomes. It should be noted that the most appropriate method can differ depending on the forecasting application/goal. For example, a long-term forecast requires a different approach than a forecast for the short term. It is therefore advisable to explicitly mention the forecasting application/goal. In addition, it is essential to remain flexible when forecasting mortality. Both mortality trends and their determinants are constantly changing, as is our knowledge of them. Moreover, new forecasting methodologies are constantly being developed. These developments are important to take into account when forecasting mortality.

As a result of the research within this PhD thesis, several components of the mortality forecasting approach of Statistics Netherlands were closely evaluated, validated, and – if necessary – improved. Based on this thesis, the following components were validated: (i) the projection of smoking-attributable mortality by means of the age-period-cohort model applied to lung cancer mortality; (ii) the

use of the Li-Lee method over the other coherent forecasting methods. This validation does not apply exclusively to the mortality forecast of Statistics Netherlands, but also more generally. The forecasting method used for smoking-attributable mortality can be applied as well to other countries in the final stage of the smoking epidemic. As a result of the findings in this PhD thesis, the jump-off rates are modified in the mortality forecast by Statistics Netherlands since 2014 to improve both the accuracy and the robustness of the mortality forecast. More generally, the findings of this PhD research demonstrate how important it is that the mortality forecasts of Statistics Netherlands are adjusted in response to scientific developments and recent mortality trends, not only in the Netherlands, but in surrounding countries as well.

Including data on the smoking epidemic and on the mortality experiences of other countries in the mortality forecasts by Statistics Netherlands resulted in higher future life expectancy values, and – especially for women – added non-linearity in the future mortality trends. The first observation can be linked to the impact of the smoking epidemic on the historical increase in life expectancy and because the recent non-smoking-attributable mortality trends in the Netherlands have been less positive than the average trends in certain other countries. The latter is the result of a projected increase in smoking-attributable mortality, followed by a decline. In addition, the mortality forecasting methodology by Statistics Netherlands is more robust resulting in fewer changes between the outcomes of the yearly published forecasts.

Based on the above findings, this PhD thesis offers the following recommendations for the various users of mortality forecasts, including the government, planning bureaus, and actuarial companies. First, it is essential that users are aware of the implications of the new mortality forecasting methodology by Statistics Netherlands. For example, if long-term life expectancy is projected to be higher than it was in previous forecasts, users might conclude that the reserves for mortality-linked products or payments should be higher for a longer period of time, or be delayed to a later date. The outcomes of the new mortality forecasts of Statistics Netherlands also affect the official population forecasts for the Netherlands issued by Statistics Netherlands. For example, if the forecasted life expectancy is higher, the extent of ageing will also be higher than previously expected. When applying the outcomes of the mortality forecasts (and, subsequently, the population forecasts), users should keep in mind that these measures (like life expectancy at birth) are averages of the population, and will not apply to all segments of the population, as there are very large differences in life expectancy based on, for instance, socio-economic status. Users should be aware of this diversity within the population. A flexible attitude towards the outcomes of

mortality forecasts is required of users, as the results of a given mortality forecast will change in response to new mortality developments, new underlying factors, new knowledge about mortality developments, and new methodologies.

The data-driven approach of this PhD thesis, as well as the extensive evaluation, have led to important new insights on mortality forecasting. For future research on mortality forecasting in the context of non-linear mortality trends, the evaluation of other countries with non-linear mortality trends, such as Eastern European countries, would be important. Furthermore, attention is needed for other potential sources of non-linearity in addition to the smoking epidemic, which might influence current and future mortality trends. Examples include excessive alcohol consumption (Eastern Europe) or obesity. In addition, future research might explore a wider range of mortality forecast outcome measures (such as the variability of the age at death) not only in order to evaluate the mortality forecasts more comprehensively, but to improve upon the methods themselves. Moreover, it would be an important way forward in mortality forecasting if more attention is paid to heterogeneity within populations. While important advances in mortality forecasting have been made, mortality forecasts that are disaggregated beyond age, sex, and region are almost non-existent. Finally, closer collaboration between the academic and practical world, but also between different disciplines (such as demographic and actuarial sciences), is important to further develop the field of mortality forecasting.

Nederlandse samenvatting

Het voorspellen van toekomstige sterfte bij niet-lineaire sterfteontwikkelingen: een evaluatie

Het doel van dit proefschrift is om de voorspelling van de toekomstige sterfte te evalueren in de context van niet-lineaire sterfteontwikkelingen in het verleden. Een goede en nauwkeurige voorspelling van de toekomstige levensverwachting en sterfte wordt steeds belangrijker door de algehele toename van de levensverwachting en de maatschappelijke consequenties hiervan (vergrijzing, gezondheidzorg, woningbouw, sociale zekerheid, pensioenen).

De meeste huidige methoden voor het voorspellen van de toekomstige sterfte zijn gebaseerd op directe (lineaire) extrapolatie, het doortrekken van bestaande sterfteontwikkelingen. Deze methoden worden zo min mogelijk beïnvloed door persoonlijke meningen. Wanneer de sterfteontwikkelingen niet-lineair zijn, zoals in Nederland, zijn deze meer objectieve methoden minder geschikt. De sterfteprognose bij niet-lineaire sterfteontwikkelingen kan worden verbeterd door rekening te houden met het verstorende effect van roken op de sterfteontwikkelingen en gebruik te maken van sterfteontwikkelingen in andere landen. Hiermee wordt echter meer subjectieve informatie meegenomen in de sterfteprognose. De keuze voor objectieve of subjectieve prognosemethoden is daarom een belangrijk onderwerp van debat. Moeten alleen objectieve methoden worden toegepast, hoewel ze minder geschikt zijn bij niet-lineaire sterfteontwikkelingen? Of heeft het toevoegen van aanvullende informatie de voorkeur, ondanks de extra subjectiviteit? Deze vragen vereisten een evaluatie van de sterfteprognose in de context van niet-lineaire sterfteontwikkelingen in het verleden.

In de meeste eerdere evaluaties van sterfteprognosemethoden werd de nauwkeurigheid van prognoses vergeleken, een puur kwantitatieve benadering. Bovendien waren deze meestal gericht op puur objectieve methoden die minder relevant zijn voor niet-lineaire sterfteontwikkelingen. Hierbij ontbrak een evaluatie van de sensitiviteit van de toekomstige sterfte als gevolg van specifieke in de methodes benoemde keuzes (expliciete veronderstellingen), bijvoorbeeld voor de lengte van de schattingsperiode[3)] en de jump-off rates[4)].

3) De historische periode die gebruikt wordt om het model op te schatten.
4) De sterftekansen die gebruikt worden als startwaarden van de sterfteprognose.

In dit proefschrift worden verschillende prognosemodellen en –methodes geëvalueerd, zowel vanuit een kwantitatief als kwalitatief perspectief. Ook is beoordeeld hoe sensitief de toekomstige sterfte is voor het gebruik van verschillende expliciete veronderstellingen. Daarnaast zijn verschillende elementen van een sterfteprognose geëvalueerd, die rekening houdt met niet-lineaire sterfteontwikkelingen in het verleden, bijvoorbeeld door aan roken gerelateerde sterfte te voorspellen en een model te gebruiken waarbij sterfteontwikkelingen van verschillende landen worden meegenomen.

Dit proefschrift omvat een zorgvuldige studie van sterfteontwikkelingen in het verleden. Het onderzoek is vooral op Nederland gericht. Maar om een bredere empirische basis te genereren, zijn ook de sterfteontwikkelingen in andere Noordwest Europese landen bestudeerd. Hierbij ligt de nadruk op hoe verschillende sterfteontwikkelingen (met name lineair versus niet-lineair) de prestaties van verschillende sterfteprognosemodellen beïnvloeden, zowel kwantitatief als kwalitatief.

Dit proefschrift draagt niet alleen bij aan het debat over de mate van subjectiviteit in een sterfteprognose, maar wordt ook gebruikt om de sterfteprognose van het Centraal Bureau voor de Statistiek (CBS) te evalueren, valideren en – uiteindelijk – te verbeteren.

In deze studie worden de volgende onderzoeksvragen beantwoord:
1) Wat is, in de context van niet-lineaire sterfteontwikkelingen, de impact van de keuze van het sterfteprognosemodel versus expliciete veronderstellingen op de sterfteprognose?
2) Hoe kan de toekomstige aan roken gerelateerde sterfte op een formele manier worden geschat?
3) Welk model kan het beste worden gebruikt om de sterfte op een coherente manier te voorspellen, d.w.z. door rekening te houden met de sterfteontwikkelingen in andere landen?
4) Hoe kan de sterfteprognose het beste aansluiten op recent waargenomen data?

Na een eerste inleidend hoofdstuk behandelen de empirische hoofdstukken 2 tot en met 5 onderwerpen die helpen bij het beantwoorden van de hierboven genoemde onderzoeksvragen. Hoofdstuk 6 sluit het proefschrift af met een samenvatting en algemene discussie van de bevindingen.

In hoofdstuk 2 worden verschillende Europese sterfteprognosemethoden en hun veronderstellingen besproken, alsook de impact hiervan op prognoses van de toekomstige levensverwachting in Nederland. In het bijzonder zijn hierbij (i) de

huidige methoden die werden gebruikt voor de officiële sterfteprognoses in Europa onderzocht, (ii) de uitkomsten en veronderstellingen van verschillende prognosemethoden voor Nederland vergeleken en (iii) de uitkomsten voor Nederland van verschillende soorten methodes met dezelfde expliciete veronderstellingen - waaronder dezelfde historische periode - vergeleken. Uit de evaluatie van de huidige methoden blijkt dat statistische bureaus in Europa meestal eenvoudige lineaire extrapolatiemodellen gebruiken. Landen met minder lineaire sterfteontwikkelingen hanteren echter andere benaderingen of veronderstellingen. Binnen Nederland worden verklarende modellen, de afzonderlijke projectie van rook- en niet-rook-gerelateerde sterfte, en de projectie van het leeftijdsprofiel van de sterfte gebruikt. Voor Nederland verschilt de levensverwachting bij geboorte in 2050 gegeven deze methoden – met ook duidelijke verschillen in de gebruikte expliciete veronderstellingen – met ongeveer zes jaar. Door gebruik te maken van dezelfde expliciete veronderstellingen (historische periode 1970-2009 en de laatst waargenomen sterftekansen als jump-off rates) zijn de verschillen tussen de methoden slechts 1,8 jaar voor vrouwen en 1,8 jaar voor mannen. Voor de levensverwachting op 65-jarige leeftijd is het verschil respectievelijk 1,4 en 1,9 jaar. De resultaten laten zien dat de keuze voor de expliciete veronderstellingen belangrijker is dan de keuze van de prognosemethode.

In hoofdstuk 3 wordt een voorstel gedaan voor een projectie tot 2050 van de toekomstige aan roken gerelateerde sterfte voor Engeland en Wales, Denemarken en Nederland. Het beschrijvende model voor de rookepidemie is hiervoor bijgewerkt en uitgebreid. Om de toekomstige aan roken gerelateerde sterfte te schatten behandelt het hoofdstuk een methode in twee stappen: (i) de longkankersterfte wordt geprojecteerd met behulp van de leeftijd-periode-cohort-methodologie (periode 1950-2009). Hierbij wordt gebruikgemaakt van de waargenomen convergentie van rookprevalenties en overeenkomsten in longkankersterfte in het verleden tussen mannen en vrouwen; (ii) andere aan roken gerelateerde doodsoorzaken worden toegevoegd aan de geprojecteerde longkankersterfte door een versimpelde versie van de indirecte Peto-Lopez methode toe te passen. Het aandeel van sterfte dat is toe te schrijven aan roken (de rook-gerelateerde attributieve fractie (RAF)) was voor mannen in 2009 gelijk aan 19% (44 872 overledenen) in Engeland en Wales, 22% (5 861 overledenen) in Denemarken en 25% (16 385 overledenen) in Nederland. In onze projecties namen deze fracties af tot, respectievelijk, 6, 12 en 14% in 2050. De RAF voor vrouwen bereikte in Engeland en Wales in 2008 een piek van 14% (38 883 overledenen) en zal naar verwachting in 2028 een piek bereiken in Denemarken (22%) en in 2033 in Nederland (23%). Tegen 2050 is een daling naar respectievelijk 9, 17 en 19% voorzien. Verschillende indirecte schattingsmethodes van de RAF in 2050 leveren

een bereik op van 1–8% (Engeland en Wales), 8–13% (Denemarken) en 11–16% (Nederland) voor mannen, en 7–16%, 12–26% en 13–31% voor vrouwen.

In hoofdstuk 4 worden verschillende coherente prognosemethoden geëvalueerd in termen van nauwkeurigheid (fit op historische data), robuustheid (stabiliteit over verschillende schattingsperiodes), subjectiviteit (gevoeligheid voor de keuze van de groep van landen) en plausibele resultaten (goede aansluiting met de ontwikkelingen in de schattingsperiode). De onderzochte coherente prognosemethoden zijn een co-geïntegreerde Lee-Carter-methode (CLC), de Li-Lee-methode (LL) en de Coherent-Functional-Data-methode (CFD). De methoden zijn toegepast op gegevens uit Frankrijk, Italië, Nederland, Noorwegen, Spanje, Zweden en Zwitserland, met prognoses tot 2050. De resultaten zijn ook vergeleken met de individuele Lee-Carter-methode (LC). Van de drie geëvalueerde coherente prognosemethoden blijkt de CFD-methode het best te presteren op nauwkeurigheid. De verschillen verdwijnen echter als er wordt gecontroleerd voor het aantal parameters. Zowel de CLC- als LL-methode blijken robuust. De CLC-methode (voor vrouwen) en de LL-methode (voor mannen) zijn het minst gevoelig voor de keuze van de groep landen. De LL-methode levert de meest plausibele resultaten op, met een convergentie van de toekomstige levensverwachting die vergelijkbaar is met de schattingsperiode en een regelmatig patroon van leeftijdsspecifieke verbeteringen. Dit kan betekenen dat de LL-methode, die het best presteert op robuustheid, subjectiviteit en plausibiliteit, gebruikt kan worden in plaats van de CFD-methode, waarvan de nauwkeurigheid beter is.

In hoofdstuk 5 worden zes verschillende opties voor de jump-off rates geëvalueerd en de effecten ervan op de robuustheid en nauwkeurigheid van de sterfteprognose onderzocht. We hebben de volgende jump-off rates onderzocht: de modelwaarden, de waarden in het laatste waargenomen jaar en een gemiddelde over de laatste paar waargenomen jaren. Op basis van gegevens uit acht Europese landen (België, Finland, Frankrijk, Nederland, Noorwegen, Spanje, Zweden en Verenigd Koninkrijk, periode 1960-2014) is de toekomstige levensverwachting op 65-jarige leeftijd (e65) berekend voor verschillende schattingsperioden en jump-off rates met behulp van het Lee-Carter-model. De e65 is onderzocht op nauwkeurigheid (gemiddelde absolute fout ten opzichte van de waarnemingen) en robuustheid (standaardafwijking van de verandering in de toekomstige e65). De analyse toont aan dat de keuze voor de jump-off rates duidelijk de nauwkeurigheid en de robuustheid van de sterfteprognose beïnvloedt, op verschillende manieren. Voor de meeste landen geldt dat de laatste geobserveerde waarden als jump-off rates tot de meest nauwkeurige methode leidt. Dat is gerelateerd aan de fout dat het model in de afgelopen jaren maakt. Het blijkt de meest robuuste methode om een gemiddelde van geobserveerde jaren als jump-

off rates te gebruiken. Hoe meer jaren gemiddeld worden, hoe beter de robuustheid. Aan de andere kant neemt de nauwkeurigheid af als meer jaren worden gemiddeld. De beste strategie voor het aansluiten van de sterfteprognose op de meest recent waargenomen gegevens is afhankelijk van het doel van de prognose, de waargenomen land specifieke sterfteontwikkelingen in het verleden en de fit van het model, zo impliceren de resultaten.

De resultaten van de empirische hoofdstukken van dit proefschrift laten ziet dat in prognoses voor niet-lineaire sterfteontwikkelingen, zoals die van Nederland, verschillende benaderingen en veronderstellingen worden gebruikt en zich daarin onderscheiden van eenvoudige lineaire extrapolatiemethoden bij de meeste nationale bureaus voor de statistiek. De keuze voor de expliciete veronderstellingen blijkt belangrijker dan de keuze van de methode voor de sterfteprognose. Omdat bekend is dat het opnemen van aanvullende informatie over de rookepidemie of de sterfteontwikkelingen uit andere landen – op zijn minst – het effect van de lengte van de historische periode vermindert, zal dit waarschijnlijk resulteren in een robuustere prognose. Het afzonderlijk voorspellen van de rook-gerelateerde sterfte is een manier om aanvullende informatie over de rookepidemie op te nemen. De in dit proefschrift ontwikkelde leeftijd-periode-cohort-extrapolatiemethode – gebaseerd op aannames afgeleid van het rookepidemie-model en een zorgvuldige studie van historische ontwikkelingen – blijken geschikt voor dit doel. De Li-Lee-methode blijkt de voorkeursmethode om de sterfteontwikkelingen uit andere landen mee te nemen door middel van coherente prognosemethoden. Deze methode presteerde namelijk beter dan de co-geïntegreerde Lee-Carter-methode en de Coherent Functional Data-methode in termen van robuustheid, subjectiviteit en plausibiliteit. Een andere belangrijke expliciete veronderstelling is de keuze van de jump-off rates. Deze keuze blijkt duidelijk de nauwkeurigheid en de robuustheid van de sterfteprognose te beïnvloeden, zij het op verschillende manieren. De beste strategie bij het kiezen van de jump-off rates hangt af van het doel van de prognose, de waargenomen nationale sterfteontwikkelingen in het verleden en de model fit.

Al met al houdt de prognose van sterfte in de context van niet-lineaire sterfteontwikkelingen meer in dan directe (lineaire) extrapolatie van waargenomen sterfteontwikkelingen. Hoewel het opnemen van aanvullende informatie (zoals de rookepidemie en/of de sterfteontwikkelingen in andere landen) automatisch resulteert in een meer subjectieve methode, resulteert dit ook in een methode die minder afhankelijk is van een belangrijke expliciete veronderstelling: de historische periode. Dit is belangrijk, omdat expliciete veronderstellingen een essentiële rol blijken te spelen in de prognose van de sterfte. De aanvullende informatie die bij de sterfteprognose wordt gebruikt, moet

wel gebaseerd zijn op zorgvuldig onderzoek van historische sterfteontwikkelingen en accuraat voorspeld kunnen worden. Een zorgvuldige afweging van de voor- en nadelen van de aanvullende informatie is dan ook essentieel.

Dit proefschrift heeft ook implicaties voor de sterfteprognose in het algemeen. Allereerst dient het enorme belang van de expliciete veronderstellingen (waaronder ook de keuze van de landengroep bij een coherente sterfteprognose) onderstreept te worden. Aan expliciete veronderstellingen dient een belangrijkere rol toegekend te worden dan nu gebruikelijk is. Idealiter moeten deze ook meegenomen worden bij de bepaling van de onzekerheid van de sterfteprognose. Daarnaast dienen nieuwe prognosemethoden niet alleen te worden beoordeeld op nauwkeurigheid, maar ook op meer kwalitatieve criteria zoals robuustheid, subjectiviteit en plausibele resultaten. Hierbij dient opgemerkt te worden dat het doel van de prognose bepalend is voor de keuze van de beste methode. Een prognose voor de lange termijn vergt bijvoorbeeld een andere benadering dan een prognose voor de korte termijn. Het is dan ook goed dit altijd te expliciteren. Daarnaast is het belangrijk om flexibiliteit te hanteren bij de sterfteprognose. Zowel de sterfteontwikkelingen zelf, als de achterliggende factoren, onze kennis hierover en de methodologie om sterfteprognoses te maken, zijn voortdurend in beweging. Dit is belangrijk om mee te nemen in toekomstige sterfteprognoses.

In dit proefschrift zijn verschillende componenten van de CBS-sterfteprognose nauwgezet geëvalueerd, gevalideerd en – indien nodig – verbeterd. Op basis van het proefschrift, blijken de onderstaande componenten van de CBS prognose gevalideerd: (i) de projectie van de aan roken gerelateerde sterfte aan de hand van de leeftijd-periode-cohort-extrapolatiemethode toegepast op longkankersterfte; (ii) het gebruik van de Li-Lee-methode ten opzichte van andere coherente sterfteprognosemethoden. Deze validatie geldt overigens niet uitsluitend voor de CBS-prognose, maar ook meer in het algemeen. Zo kan de extrapolatiemethode voor aan roken gerelateerde sterfte ook makkelijk toegepast worden op andere landen die in het finale stadium van de rookepidemie zitten. Op basis van de resultaten uit dit proefschrift zijn de jump-off rates in de CBS-prognoses vanaf 2014 aangepast om de robuustheid en nauwkeurigheid van de prognose te verbeteren. Meer algemeen toont het onderzoek in dit proefschrift aan hoe belangrijk het voor de CBS-prognose is om wetenschappelijke ontwikkelingen en recente sterfteontwikkelingen nauw in de gaten te houden, niet alleen in Nederland, maar ook in andere West-Europese landen.

Het tegelijkertijd meenemen van het effect van de rookepidemie en de sterfteontwikkelingen in andere landen heeft in de CBS-sterfteprognose geleid tot een hogere toekomstige levensverwachting en – vooral voor vrouwen – meer

niet-lineariteit in de toekomstige sterfte. Het eerste wordt verklaard door de negatieve invloed van de rookepidemie op de historische toename van de levensverwachting en doordat de afname in niet-rook-gerelateerde sterfte in andere landen hoger is dan in Nederland. Het laatste komt vooral doordat rook-gerelateerde sterfte bij vrouwen naar verwachting eerst nog zal toenemen, alvorens het zal afnemen. Daarnaast is de methode robuuster geworden en zijn er daardoor minder veranderingen tussen de prognoses die jaarlijks worden gepubliceerd.

Op basis van bovenstaande bevindingen doen we de volgende aanbevelingen voor de verschillende gebruikers van sterfteprognoses, waaronder de overheid, planbureaus en actuariële instanties. Allereerst moeten gebruikers zich realiseren dat de verhoging van de levensverwachting van de nieuwe CBS-sterfteprognose verschillende aspecten beïnvloedt. Producten gekoppeld aan de sterfteprognose vereisten bijvoorbeeld een hogere reserve en pensioenen moeten langer worden uitbetaald wanneer verwacht wordt dat mensen langer leven. De nieuwe uitkomsten van de CBS-sterfteprognose werken ook door in de nationale bevolkingsprognose van het CBS. De mate van veroudering wordt bijvoorbeeld hoger geprognosticeerd dan eerst werd verwacht. Het is daarnaast goed voor gebruikers om zich te realiseren dat de uitkomsten van de sterfteprognose (en de bevolkingsprognose; zoals de levensverwachting bij geboorte) betrekking hebben op de gemiddelde bevolking en niet van toepassing zijn op alle segmenten van de bevolking. Zo leven bijvoorbeeld personen met minder onderwijsjaren gemiddeld korter. Het wordt daarom aanbevolen om aandacht te hebben voor deze diversiteit binnen de bevolking. Bij het plannen voor de toekomst is het belangrijk om te beseffen dat zowel de uitkomst als de methode van de sterfteprognose kan veranderen. Flexibiliteit is dan ook geboden.

De data-gedreven benadering van dit proefschrift, alsook de omvangrijke evaluatie, hebben geleid tot belangrijke nieuwe inzichten wat betreft de sterfteprognose. Voor toekomstig onderzoek op het gebied van sterfteprognoses bij niet-lineaire sterfteontwikkelingen is het belangrijk om andere landen met niet-lineaire sterfteontwikkelingen mee te nemen in de evaluatie, zoals landen in Oost-Europa. Ook is aandacht gewenst voor andere oorzaken van niet-lineariteit naast de rookepidemie, die nu of mogelijk in de toekomst van invloed kunnen zijn. Voorbeelden hiervan zijn overmatige alcoholconsumptie (Oost-Europa) of obesitas. Daarnaast is het belangrijk om een breder scala aan uitkomstenmaten (bijvoorbeeld de variatie in de leeftijd bij overlijden) te onderzoeken om de sterfteprognoses vollediger te evalueren en wellicht te verbeteren. Bovendien zou het een belangrijke stap vooruit zijn in onderzoek naar sterfteprognoses als er

meer aandacht komt voor de heterogeniteit binnen populaties. Hoewel er vooruitgang is geboekt in de voorspelling van de sterfte, worden sterfteprognoses vaak nog puur uitgesplitst naar geslacht, leeftijd en regio. Als laatste is een nadere samenwerking tussen de academische en praktische wereld, maar ook tussen verschillende disciplines (zoals demografische en actuariële wetenschappen), belangrijk om het onderzoek naar en de praktijk van de sterfteprognose nog verder te ontwikkelen.

Acknowledgements

This PhD project was initiated out of a wish to collaborate more intensively between those working on the mortality forecast at Statistics Netherlands and academics doing research on mortality forecasting.

Without the opportunity given by Statistics Netherlands to spend time (of the author, but also of others) on this PhD project, the author could never have achieved this result.

The supervision by Fanny Janssen is financed by the Netherlands Organisation for Scientific Research (NWO) (grant no. 452-13-001) in line with Janssen's VIDI research project "Smoking, alcohol and obesity – ingredients for improved and robust mortality projections" (see www.futuremortality.com).

About the author

Lenny Stoeldraijer was born on March 20th, 1985 in Terheijden, the Netherlands. After completing secondary school in 2003, she studied Econometrics at Tilburg University. She received her bachelor degree in 2007. In 2008 she obtained her master degree in Mathematical Economics and Econometric Methods. Her dissertation concerned forecasting the sales of 'MCB Nederland', a metal wholesaler in Valkenswaard supplying metal for more than 75 years. After graduating, she started as a research assistant at Tilburg University for Professor Jan C. van Ours, doing research in the area of age, wage and productivity. For this research, data from Statistics Netherlands was used. Therefore, when the research at Tilburg University ended, she applied for a position at Statistics Netherlands and started working at the Demography Department in June 2010, working mainly for the population forecast. In 2012 she officially started her PhD research at the University of Groningen under the guidance of Fanny Janssen, Leo van Wissen and her colleague at Statistics Netherlands, Coen van Duin. Her PhD research was combined with her work at Statistics Netherlands. Currently, she is the main person responsible for the mortality forecast by Statistics Netherlands.

Mortality forecasting in the context of non-linear past mortality trends: an evaluation

Having accurate and high-quality mortality forecasts has become increasingly important due to the general increase in life expectancy and the social consequences of this.

The aim of this dissertation was to evaluate mortality forecasting in the context of non-linear past mortality trends. In this way, it contributes to the debate on the degree of subjectivity in mortality forecasting, but also to the evaluation, validation, and further improvement of the mortality forecasts of Statistics Netherlands (CBS).

When making mortality forecasts, different approaches and assumptions are used, in which additional, often subjective, information is included to deal with non-linear mortality trends. A careful examination of past trends, and a careful assessment of the pros and cons of including additional information, is therefore important. The mortality forecast in the context of non-linear past mortality trends may be improved by making explicit adjustments for the distorting effects of smoking on mortality trends and the use of mortality developments in other countries. The specific choices that are explicitly stated in a method proved more important than the choice of the forecasting approach.

The methods used by CBS to include mortality trends from other countries in the forecast and to project smoking-attributable mortality, proved valid. The matching of the forecast to recent observations has been adjusted.

Both the mortality developments themselves and the underlying factors, our knowledge hereof and the mortality forecasting methodology are constantly changing. Flexibility in both making mortality forecasts and interpreting his outcomes is therefore required.

Uitgave 2019