

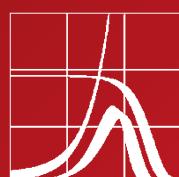
**Demographic
Research
Monographs**

Tommy Bengtsson
Nico Keilman
Editors



Old and New Perspectives on Mortality Forecasting

 Springer Open



Demographic Research Monographs

A Series of the Max Planck Institute
for Demographic Research

Editor-in-chief

Mikko Myrskylä
Max Planck Institute for Demographic Research
Rostock, Germany

More information about this series at <http://www.springer.com/series/5521>

Tommy Bengtsson • Nico Keilman
Editors

Old and New Perspectives on Mortality Forecasting



Springer Open

Editors

Tommy Bengtsson
Centre for Economic Demography
Lund University
Lund, Sweden

Nico Keilman
Department of Economics
University of Oslo
Oslo, Norway



ISSN 1613-5520

ISSN 2197-9286 (electronic)

Demographic Research Monographs

ISBN 978-3-030-05074-0

ISBN 978-3-030-05075-7 (eBook)

<https://doi.org/10.1007/978-3-030-05075-7>

Library of Congress Control Number: 2018968401

© The Editor(s) (if applicable) and The Author(s) 2019. This book is an open access publication.

Open Access This book is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this book are included in the book's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the book's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

The *Stockholm Committee on Mortality Forecasting* was created in 2000 by the National Social Insurance Board, which is the national agency in Sweden responsible for providing a financial picture of Sweden's public pension system. The aim of the Committee was to survey the state of the art and to provide an impetus for the advancement of knowledge and better practice in forecasting mortality. To do so they brought together scholars from different disciplines working on issues in projecting mortality resulting in five booklets, which now are reprinted together with a new introduction.

Members of the Stockholm Committee on Mortality Forecasting 2000–2007:
Professor Edward Palmer, Uppsala University and National Social Insurance Board,
Sweden
Professor Tommy Bengtsson, Committee Secretary, Lund University, Sweden
Professor Juha M. Alho, University of Joensuu, Finland
Professor Kaare Christensen, University of Southern Denmark, Odense
Professor Nico Keilman, University of Oslo, Norway
Professor James W. Vaupel, Max Planck Institute for Demographic Research,
Rostock, Germany

Preface

More than 10 years have gone since the Swedish Social Insurance Agency published a series of booklets entitled *Perspectives on Mortality Forecasting*. Five volumes appeared in print, on different topics that all are relevant for anyone faced with the task of computing a forecast of mortality in future years. Each volume contained the papers presented in a series of workshops organized by the Stockholm Committee on Mortality Forecasting during the years 2002–2007. The current volume contains reprints of the contributions to the five booklets. Each part of this book corresponds to one original booklet. We gratefully acknowledge the financial support from the Max Planck Institute for Demographic Research for the publication of this book and from the Swedish Social Insurance Agency for organizing the Committee's workshops.

The field of mortality forecasting is in continuous change. The purpose of the current volume is to track this development by showing the reader what the main issues were some 10–15 years ago, together with an update. Therefore, the book starts with an introductory chapter, which summarizes recent new insights on the following topics:

- The need for accurate mortality forecasting today for systems for pension, health care, and elderly care
- Determinants and dynamics of life expectancy
- Causes of death and lifestyle factors, such as smoking and obesity
- Cohort and period perspectives
- Compression of mortality
- Joint forecasting of mortality in similar populations
- From scenarios to stochastic modelling
- The way conditions in early life affect mortality in later life
- The increasing gap in life expectancy with respect to position in the income distribution

Some of the material in the introductory chapter relates to only one of the five parts, while other items in the list above cut across themes. The last item brings up a new topic of mortality forecasting, which was not dealt with in any of the workshops.

Lund, Sweden
Oslo, Norway

Tommy Bengtsson
Nico Keilman

Contents

1	Introduction	1
	Tommy Bengtsson, Nico Keilman, Juha M. Alho, Kaare Christensen, Edward Palmer, and James W. Vaupel	
Part I Current Practice		
2	Life Expectancy Is Taking Center Place in Modern National Pension Schemes – A New Challenge for the Art of Projecting Mortality	23
	Edward Palmer	
3	Experiences from Forecasting Mortality in Finland	33
	Juha M. Alho	
4	Mortality Projections in Norway	43
	Helge Brunborg	
5	Mortality Assumptions for Sweden. The 2000–2050 Population Projection	59
	Hans Lundström	
6	Forecasting Life Expectancy: The SCOPE Approach	73
	James W. Vaupel	
7	Mortality Forecasts. Comments on How to Improve Existing Models – An Epidemiologist’s Perspective	79
	Kaare Christensen	
8	The Need for Looking Far Back in Time When Predicting Future Mortality Trends	85
	Tommy Bengtsson	

Part II Probabilistic Models

- 9 Erroneous Population Forecasts** 95
Nico Keilman
- 10 Remarks on the Use of Probabilities in Demography and Forecasting** 113
Juha M. Alho
- 11 An Expert Knowledge Approach to Stochastic Mortality Forecasting in the Netherlands** 123
Maarten Alders and Joop de Beer
- 12 Stochastic Forecasts of Mortality, Population and Pension Systems** 145
Shripad Tuljapurkar

Part III The Linear Rise in Life Expectancy: History and Prospects

- 13 The Linear Rise in the Number of Our Days** 159
Jim Oeppen and James W. Vaupel
- 14 Mortality Forecasts and Linear Life Expectancy Trends** 167
Ronald Lee
- 15 Forecasting Life Expectancy: A Statistical Look at Model Choice and Use of Auxiliary Series** 185
Juha M. Alho
- 16 Life Expectancy Convergence Among Nations Since 1820: Separating the Effects of Technology and Income** 197
Jim Oeppen
- 17 Linear Increase in Life Expectancy: Past and Present** 221
Tommy Bengtsson

Part IV Causes of Death

- 18 How Useful Are the Causes of Death When Extrapolating Mortality Trends. An Update** 237
Graziella Caselli, Jacques Vallin, and Marco Marsili
- 19 Forecasting Life Expectancy and Mortality in Sweden – Some Comments on Methodological Problems and Potential Approaches** 261
Måns Rosén
- 20 How Analysis of Mortality by Cause of Death Is Currently Influencing UK Forecasts** 271
Richard Willets

Part V Cohort Factors: How Conditions in Early Life Influence Mortality Later in Life

- 21 A Life Course Perspective to the Modern Secular Mortality Decline and Socioeconomic Differences in Morbidity and Mortality in Sweden** 295
Martin Lindström and George Davey Smith
- 22 Early Life Events and Later Life Health: Twin and Famine Studies** 311
Kaare Christensen
- 23 The Month of Birth: Evidence for Declining but Persistent Cohort Effects in Lifespan** 319
Gabriele Doblhammer
- 24 Early-Life Conditions and Old-Age Mortality in a Comparative Perspective: Nineteenth Century Sweden and Belgium** 335
Tommy Bengtsson and George Alter

Chapter 1

Introduction



**Tommy Bengtsson, Nico Keilman, Juha M. Alho, Kaare Christensen,
Edward Palmer, and James W. Vaupel**

1.1 The Need for Accurate Mortality Forecasts Is Greater Than Ever Before

Globally, the twenty-first century will witness rapid population ageing. Already in 2050, one out of five persons in the world, and one out of three in Europe, is expected to be 60 or over (UN 2015). Moreover, we have entered into a new stage of population ageing in terms of its causes, which have altered its consequences. In the first stage, lasting until the middle of the twentieth century in developed countries, population ageing was entirely due to the decline in fertility, with Sweden being commonly used as an example (Coale 1957; Bengtsson and Scott 2010; Lee and Zhou 2017). During this stage, the increase in life expectancy was primarily driven by declines in infant and child mortality. It worked in the opposite direction to

T. Bengtsson (✉)

Centre for Economic Demography, Lund University, Lund, Sweden

e-mail: tommy.bengtsson@ekh.lu.se

N. Keilman (✉)

Department of Economics, University of Oslo, Oslo, Norway

e-mail: nico.keilman@econ.uio.no

J. M. Alho

University of Joensuu, Joensuu, Finland

K. Christensen

Institute of Public Health, University of Southern Denmark, Odense, Denmark

E. Palmer

Social Insurance Economics, Uppsala University, Uppsala, Sweden

Research Division, Swedish National Social Insurance Board, Stockholm, Sweden

J. W. Vaupel

Max Planck Institute for Demographic Research, Rostock, Germany

the fertility decline, making the population younger since it added more years before, than after retirement (Coale 1957; Lee 1994). In the second stage of population ageing, which is the current situation, population ageing is primarily driven by the increase in life expectancy, which is now due to declining old-age mortality. As a result, more years are added after retirement than in working ages (Lee 1994). Could immigration or an upswing in fertility stop population ageing? The short answer is *most likely not*. The effect of migration on population aging is generally regarded as minor (Murphy 2017), and since population ageing is a global phenomenon, it will be of no general help anyway. A rapid increase in fertility is improbable and, in any case, an increase would take some 25 years before adding to the labor force. Instead, attention has been focused on how to adapt our social systems to the increasing number of elderly per worker – more so since the increase in the elderly-per-worker ratio came in parallel with a rise in per capita costs for the institutional care, home care, and general health care for the elderly.

The increase in per capita costs for the elderly is most clearly visible in the United States and Japan, but also in a selection of European countries, among them Sweden (Lee and Mason 2011; Qi 2016). Attempts to adapt social care, health care, and pension systems to this new reality, among other things by increasing the minimum pension age and even indexing them to life expectancy, are not sufficient. Accurate life expectancy projections for planning the supply of health care and other services to the elderly forms of care become then crucial. This said, the public pension system is still a main driver of costs in a large number of countries, which is why we need to learn how to design national pension systems that adapt to the ageing process. Specifically, the re-design of public pension systems has increasingly been directed to creating public life-cycle savings-transfer schemes that are financially sustainable.

An example of a financially stable pension scheme that by design is demographically robust is the Non-financial Defined Contribution (NDC) public pension scheme, introduced in a number of European countries since the mid-1990s (Latvia, Italy, Poland, Norway, and Sweden). What's more, the demographic stability features of such NDC systems are being emulated by other countries, where Germany is a prime example. This is not the end of the story, however. In fact, it is more where the story of this publication begins. What has happened is that life expectancy these days is an important parameter in calculating newly granted pensions, not only in public NDC schemes, but also increasingly in occupational and or employer-based financially defined contribution schemes. Indeed, more and more countries are converting from defined benefit to defined contribution schemes (e.g. Australia, Denmark, the Netherlands, Sweden, the United Kingdom and the United States).

What all systems for public and private pension, health care, and elderly care have in common is that accurate life expectancy and mortality projections are important for both planning and managing expenditures. This is why the topic of the endeavor presented in this publication is how to project future mortality accurately. Ideally, projected values should be very close to outcomes observed later. The problems arises when there is a systematic tendency to over- or underestimate the remaining number of years a group of persons of a certain age may be expected to be alive.

In practice, during recent decades, demographers, actuaries, economists, and other social scientists have done just this. They have systematically underestimated improvements in future life expectancy for people 60 and older. This means that current pension schemes are under-financed. Similarly, to calculate future costs for social and health care systems, accurate mortality forecasts are equally important.

The demographic models used in projecting mortality are usually based on statistical modelling of historical data. The question is, is it possible to bring the results of mortality modelling closer to the ideal, and if so, what do demographers need to do to achieve this result? This reasoning provided the motivation for forming the Stockholm Committee on Mortality Forecasting in 2000, and the question is still a high priority, which is an important message of this publication. Here, we attempt to identify what we have learned since this collaboration began, and discuss what remains to be done. Doing so, we go beyond the five themes addressed in our previous publications. We discuss whether information on socioeconomic status can be used to improve mortality forecasting – in particular the fact that persons with low socio-economic status (income, education, occupation) tend to be worse off in terms of life expectancy than those with higher status.

1.2 Determinants and Dynamics of Life Expectancy – Pensions Are Upping the Ante for the Challenge Facing the Art of Projecting Mortality

In 2003, the first chapter of the first volume of the Stockholm Committee on Mortality Forecasting was titled “*Life Expectancy Is Taking Center Place in Modern National Pension Schemes – A New Challenge for the Art of Projecting Mortality*” (see Chap. 2 in the current volume). Now, 15 years later, we are beginning to get a clear perspective of the challenge of life expectancy dynamics, and its repercussions for public policy in general and for pensions in particular. Since 2003, we have become much more aware of the importance of understanding both the determinants and dynamics of life expectancy. Below we will argue that accurate projections of life expectancy are more necessary than before. Pension scheme design is nowadays increasingly proactive rather than reactive. Life expectancy now has the leading role in the design of pension schemes – in setting the minimum age at which benefits can be claimed and how they are calculated.

As we moved into the twenty-first century, the “philosophy” for projecting life expectancy employed by many official statistical agencies, embraced a general belief that improvements in mortality could not continue into the twenty-first century at the same rate as they did in the preceding half century. This belief is exemplified in the approach adopted in the three country-practice contributions in the first volume of *Perspectives on Mortality Forecasting* – Finland (Chap. 3 by Alho), Norway (Chap. 4 by Brunborg), and Sweden (Chap. 5 by Lundström).

In projecting life expectancy in the 1990s, Statistics Finland adjusted future age-specific mortality rates so that the implied increase in life expectancy gradually slowed down until the line reached a certain target value. Norway employed a similar procedure, i.e. a trend extrapolation of past rate(s) of change in mortality with a gradual decrease in the rate of change towards an “end target”. In Sweden, the extrapolation included a downward adjustment of the historical rate of change, for the period 2000–2050 as time passed, first by 25 percent and then by 50 percent.

For many, the eye-opening event of the time was Oeppen and Vaupel’s article “Broken Limits to Life Expectancy”, published in *Science*. In this article, they concluded that there is no evidence that the upward linear trend in best-performance life expectancy is coming to an end:

Continuing belief in imminent limits (to improvements in mortality) is distorting public and private decision-making. [...] If life expectancy were close to a maximum, then the increase in the record expectation of life should be slowing. It is not. For 160 years, best-performance life expectancy has steadily increased by a quarter of a year per year, an extraordinary constancy of human achievement (Oeppen and Vaupel 2002).

In Chap. 13 of the present volume, Oeppen and Vaupel write: “If the trend of the past 160 years continues for the next half century, (best performance) life expectancy in 2050 will reach a record of 97.5.” At the time, the message of these two papers conflicted with the cautious note “Overly optimistic forecasts of life expectancy have already influenced important areas of public policy” (Olshansky et al. 2001).

Waldron (2005), commissioned by the USA Office of the Government Actuary, examined the life expectancy projections underlying the US’s official projections for Social Security, and even projection models employed in several European countries. She concluded that, regardless of the choice of procedure, including use of the “state of the art” Lee-Carter model, ex-post evaluations show that the models systematically underestimate expected lifespans, frequently owing to the widespread use of “expert opinion”. Keilman and Pham (2004) evaluated official mortality projections in 14 European countries, published after World War II, against actual outcomes. They showed slow increases in projected life expectancies. On average, the under-prediction of life expectancy amounted to 1.0–1.3 years of life for projections 10 years ahead, and 3.2–3.4 years of life 20 years ahead (see Chap. 9 by Keilman).

Mortality rates of the older population have declined faster and faster since at least the 1960s. At a joint World Bank – Swedish conference held in Stockholm in December 2009 on NDC pension schemes, a paper examined patterns of mortality decline in the Scandinavian countries – Denmark, Finland, Norway and Sweden, and four other countries – Japan, Portugal, the United Kingdom and the United States (Alho et al. 2013). At that time, Japan (# 3) and Sweden (# 10) were in the top 10 group of performers and Portugal (# 51) and the US (# 47) at the bounds of the top quartile, with the remaining countries in-between these four. The examination covered birth cohorts in these eight countries beginning earliest in 1908 and ending in 2007.

The clearest pictures are those for Japan and Sweden. In war-free Sweden, age groups 65–74 and 75–84 show an unmistakable acceleration in the rate of decline in mortality around 1945–1950, when death rates started to decline faster than before

1945. This new trend is still visible in 2014. The same process is observable for Japan, but it is much stronger and shorter in duration. It begins in the mid-1950s and continues up to 1990, but even as it decreases in strength, the rate of acceleration in these age groups *is still on a par with all the other countries* but the US. The latter country goes between phases of accelerating and decelerating increases in the rate of change in mortality. Most noteworthy is that in Japan, from around 1990, acceleration in the decline in mortality also came to the age group 85–99 and is still going on.

The success of procedures or lack of it, for projecting life expectancy is reflected directly in the fairness and financial sustainability of national pension schemes. To begin with, more and more life expectancy is entering directly into determination of individual benefits in national pension schemes. This means that non-biased cohort life expectancy projections over a succession of birth cohorts are essential for fulfilling the social criterion of fairness of individual outcomes within and between generations, and the fiscal criterion of maintaining financial balance. Both require that the procedure employed to project life expectancy does not lead to systematically biased estimates of life expectancy – i.e. systematic under or overestimation of the life expectancy of the pension insurance pool’s birth cohorts. The other side of the same coin is that systematic underestimation of life expectancy creates financial deficits that have to be financed by younger generations, which goes against the notion of intergenerational fairness.

The challenge is to design statistical models that capture the *dynamics of the rate of decline* in mortality. It has been demonstrated that, with data for 2400 individual birth cohorts for the period 1907–2014 for eight countries,¹ both classical period models and the Lee-Carter model systematically underestimate life expectancy, and that the scale of the errors is increasing (Palmer et al. 2018). As an alternative, a new model for projecting cohort life expectancy from the changing relationship between period and cohort mortalities has been used (Palmer et al. 2018). Applying the new model to the entire sample, the most important finding is that it delivers systematically unbiased projections for the individual cohort observations for the eight countries separately, and in the aggregate. Other promising alternative models are also being developed that are not systematically biased and that have improved accuracy (Bergeron-Boucher et al. 2017; Pascariu et al. 2018).

1.3 Cause of Death Forecasts

Is it useful to include causes of death (CoD) in mortality forecasting models? Willett (Chap. 20) lists some of the problems. First, interactions between different causes are difficult to model. Second, CoD reporting is unreliable at older ages where most deaths occur. Third, there is the danger of misclassification of deaths by cause (e.g.,

¹Denmark, France, Italy, the Netherlands, Norway, Sweden, the United Kingdom and the United States.

primary or secondary cause) particularly for the very oldest. Fourth, there are important major causes that change over time with medical advances. Finally, cause classification methods change over time and may distort observed trends (Tuljapurkar and Boe 1998; Booth 2006). Wilmoth (2005) showed that when trends are extrapolated linearly, cause-specific mortality forecasts result in lower future values for life expectancy than do all-cause forecasts, because the CoD that declines most slowly is the one that dominates in the end. Caselli, Vallin, and Marsili (Chap. 18) illustrate this point empirically for England and Wales. Alho (1991) found that using CoD did not improve the accuracy of his mortality forecasts for the US. A major review of mortality forecasting methods for the UK population concluded that such forecasts should not be carried out by cause of death (GAD 2001). However, cause-elimination projections may be useful as reference scenarios; see Rosén (Chap. 19) for an application to Swedish data.

Some causes of death are related to life style factors, such as smoking, obesity, exercise etc. Examples are lung cancer, cardiovascular diseases, and problems with the respiratory system. Chetty et al. (2016) analyzed the correlates in local variation in life expectancy in the United States. They showed that life expectancy was negatively correlated with rates of smoking and obesity and positively correlated with exercise rates among individuals in the bottom income quartile. Janssen, De Beer, and their colleagues have done innovative work and developed mortality projection models that explicitly account for smoking-related mortality (Janssen et al. 2013; Stoeldraijer et al. 2013; Janssen and de Beer 2016). They separated smoking-related mortality from non-smoking-related mortality, estimated age- and sex-specific mortality rates for both types of mortality, parameterized the age schedules, and extrapolated the parameters of these schedules. The model contains separate parameters for the delay of mortality and for the compression of the age at death distribution; hence the name CoDE-model. To include smoking-related mortality of men and women may potentially result in more accurate mortality projections, because the onset and the level of the smoking epidemics (and subsequent fall in prevalence) differ between the sexes. Smoking behavior explains a large part of the divergence in life expectancies between the sexes after World War II, and their convergence in recent decades.²

1.4 Period and Cohort Perspectives

Virtually all life expectancies discussed in this volume are computed from period life tables. A period life expectancy may be interpreted as the expected life length of a newborn child if all age-specific period mortality rates would remain constant over

²See, for instance, Lindahl-Jacobsen et al. (2016) for the case of Denmark, and Janssen and Van Poppel (2015) for The Netherlands. See also Sect. 1.5 on joint forecasting of mortality in similar populations in this Chapter.

time. It relates to the behavior of many different birth cohorts during just one calendar year. However, this is a hypothetical situation, and real people do not live in this way. They are members of only one birth cohort, and they live their lives during many years.

Survival in real birth cohorts is different from survival in the hypothetical situation of constant period mortality rates, for various reasons (Borgan and Keilman 2016). First, in times of falling mortality, the life expectancy of a certain birth cohort is comparable to that based on age-specific mortality rates for a particular year many decades later. This is what demographers call a “tempo effect”. Second, there are cohort effects. Mortality decreased regularly for subsequent cohorts, but some cohorts may display a typical pattern. An example is the case of Denmark, discussed earlier, where women born in the interwar period adopted unhealthy life styles.

Because of their high mortality, Danish female period life expectancy stagnated in the 1970s and 1980s (Lindahl-Jacobsen et al. 2016). Then, Danish female period life expectancy started to rise again as this cohort died out. Third, there is mortality selection. The frail members of a birth cohort tend to die earlier than members that are more robust. That leaves a selected population at the higher ages. When longevity improves, more of the frail cohort members will live longer. Thus for a country where the living conditions improved early (examples are Norway and Sweden) the elderly population in our days may contain more frail persons than it does in countries where the living conditions improved later (for example, in Italy). On the other hand, factors that enabled more members of a cohort to survive may have also led to improved health for the members of the cohort that were not at high risk of death.

Because of these complications, cohort life expectancy may increase faster than what period life tables suggest. For instance, “best practice” *cohort* life expectancies for women born between 1870 and 1920 increased by 0.43 years of age per calendar year (Shkolnikov et al. 2011) – almost twice as fast as the improvement in best practice *period* life expectancy for women since 1840 (0.24 years of age per calendar year). Here, best practice life expectancy may refer to the maximum life expectancy observed among national populations in a given year or for a given birth cohort (see also Oeppen and Vaupel in Chap. 13). Wilmoth (2005) shows empirical evidence for Sweden for the period 1751–2002 and birth cohorts 1751–1911. Goldstein and Wachter (2006) use official estimates and projections of female life expectancy from Sweden and the United States. In addition, they derive analytical expressions for gaps and lags between the two types of life expectancies. The life expectancy lag is the number of years it takes for a period life expectancy to reach the current level of the cohort life expectancy. The gap is the difference between the life expectancy for a cohort born in a particular year, and the period life expectancy for that year. Goldstein and Wachter (2006) find that period life expectancies are approximately equal to cohort life expectancies for cohorts born about 40–50 years earlier. They note also that the lag lengthens as mortality improves. The life expectancy gap has risen and then fallen over time. Canudas-Romo and Schoen (2005) analyze the Siler model of age-specific mortality combined with constant rates of mortality decline, and find qualitatively similar effects for the lag. Gaps were about one-ninth to

one-tenth of the lags. Missov and Lenart (2011) assume a Gompertz model for age-specific mortality and the same yearly improvement in mortality at all ages. Under these conditions, the temporal change in period life expectancy is approximately proportional to but less than the change in cohort life expectancy. If period life expectancy improves by 2 years of age per decade, cohort life expectancy would improve by 2.5 years per decade, or by 25 per cent more. Wilmoth (2005) assumes that distributions of deaths by age change over time in accordance with a linear shift model. Under this model, he also finds that cohort life expectancies increase faster than period life expectancies. Like Missov and Lennart, he establishes a simple linear relationship between period and cohort life expectancies.

Another consequence of the distorted view that period life tables give in times of changing mortality is the compression around the mean or the modal age of the age-at-death distribution. Ouellette and Bourbeau (2011) show an on-going process of mortality compression in Canada, France, Japan, and the United States. Keilman et al. (2018) find the same for Norway. Deaths tend to be progressively concentrated near the mean or the modal age at death. The findings for these five countries are based on period data. Observed and projected cohort patterns for the age at death distribution (AADD) for Norway suggest that compression in cohorts born 1900–1950, measured as a decrease in the standard deviation of the AADD beyond 30 years, is twice as fast compared to compression in period AADDs during the years 1900–2060. Canudas-Romo (2008) studies the period standard deviation from the modal age at death, assuming the Siler mortality change model as in Canudas-Romo and Schoen (2005) and a Gompertz mortality change model. The period standard deviation is constant under the Gompertz model, but falls regularly under the Siler model. Ediev (2013) discusses the theoretical link between period and cohort compression measures. He concludes that compression in periods may very well go together with decompression in cohorts. Tuljapurkar and Edwards (2011) show that the variance in the age at death is inversely related to the Gompertz slope of log mortality.

Given the problems inherent to a period approach, it will not come as a surprise that methods that project cohort life expectancies perform rather well. Zhao de Gossen de Varennes et al. (2016) demonstrate that compared to the classical Lee-Carter method for extrapolating age-specific mortality rates period-wise, projecting cohort life expectancy based on its rate of change in mortality works well in eight countries with long data series. See also Sect. 1.2 on determinants and dynamics of life expectancy in this chapter.

1.5 Joint Forecasting of Mortality in Similar Populations

Since the early days of cohort-component forecasting, official forecasters have been keen to assure forecast users of the external validity of their results. For example, the forecast of a given country should not markedly deviate from the developments in

countries with similar culture, economics, or social life. Secondly, male and female mortalities should not diverge in an unreasonable manner.

During the past two decades, extrapolation methods have regained prominence in mortality forecasting, primarily because assessments of earlier judgmental forecasts have been shown to grossly underestimate future life expectancy. However, separate extrapolations of the mortality of genders, or of countries, will eventually lead to divergence. Happily, the time series tools used in extrapolation offer multiple ways of constraining forecasts to behave in a coherent manner. A rather large literature has evolved that links varying forms of substantive reasoning regarding trends in smoking, obesity etc. ([Staetsky 2009](#); [King and Soneji 2011](#)) to extrapolation. Notable statistical contributions include [Li \(2013\)](#), [Kleinow \(2015\)](#), and [Raftery et al. \(2014\)](#).

However, the empirical evidence regarding the convergence of male and female mortality, or across countries, is equivocal. [Thorslund et al. \(2013\)](#) provide evidence of convergence between male and female life expectancy at 65, since the 1990s, in, for instance, several Northern European countries. However, an analysis of 23 European countries in [Alho \(2016\)](#) shows that both the creation of the male-female difference occurred differently in different countries and considerable differences in trends continue to exist to this date. Notably, European countries that have suffered large war casualties, and the former socialist States, have histories that differ from those observed in northern Europe, and show little evidence of convergence.

1.6 From Scenarios to Stochastic Modelling

The scenario-based approach as a way to express forecast uncertainty has been criticized from a statistical point of view ([Alho and Spencer 1985](#); [Lee 1998](#)). Obviously, uncertainty is not quantified as no probability is attached to the high-low interval. A second drawback is that one assumes perfect correlation over time: when mortality is low one year, it is also low in all other years. In reality, mortality shows random fluctuations. Third, the scenario-based approach is inconsistent when variant assumptions for mortality are combined with variant assumptions for fertility (or migration). With three variants for each of the two components, nine different combinations for population results can be made. However, a variant pair that is extreme for one variable is not necessarily extreme for another variable.

Since the future is inherently uncertain, one should compute a forecast by stochastic methods. However, official forecasters of mortality have been slow to adopt such methods. A notable exception is the implementation of a stochastic approach by the U.N. ([Raftery et al. 2012](#)). This work builds on ideas similar to those of [Grosi and King \(2008\)](#). In both works, data from similar countries is used in joint forecasting, a feature that is particularly important for countries with poor data.

[Raftery et al. \(2012\)](#) adopted a Bayesian approach to population forecasting. This view is particularly useful when one combines expert opinions with empirical data. The application of the Bayesian approach in demography, “Bayesian demography”,

started to gain popularity about 10 years ago.³ The Bayesian approach to demographic forecasting is a welcome development in addition to frequentist methods that have been used more widely. For instance, see Alho and Spencer (2005, p. 235) for stochastic modelling of future mortality for 18 European countries in the framework of the Uncertain Population of Europe (UPE) Project.

In contrast to population forecasting by statistical agencies, in actuarial science stochastic modelling has become, during the past decade, the standard. The difficulty in using safety margins and other scenarios in the pricing of annuities and other mortality related insurance products has proven to be difficult, to the point that liquid markets for such products are few. Notable examples of stochastic modelling are the formulations by Hahn and Christiansen (2016) and van Berkum et al. (2017), who both develop full probability models for age-specific mortality of one or more populations, and compute the posterior distributions of the relevant risk measures using Markov Chain Monte Carlo techniques.

Macroeconomics is another field where the inadequacy of the scenario-based assessment of forecast uncertainty is becoming evident. Generational accounting and the sustainability of pension systems can now be analyzed in stochastic demographic settings (Alho et al. 2008). Recent analyses (Lassila et al. 2014) allow for adaptive responses to stochastic shocks. Extensions of stochastic demographic modelling to migration (Bijak 2011) and household composition (Keilman 2017) are now available.

1.7 How Conditions in Early Life Affect Mortality in Later Life

The idea that childhood conditions are important for health in later life, which can be traced far back in time, came into focus in the 1920–1930s, when demographers and epidemiologists became aware of a long-term decline in mortality (Derrick 1927; Kermack et al. 1934; Bengtsson and Mineau 2009). They found that as each birth cohort got healthier, and infant mortality declined, health improved in the remaining part of their lives as well. The focus was on infant mortality, regardless of the determinants of infant mortality, as a predictor of mortality throughout the life course. The role of fetal development, which also can be traced far back in time, came into focus much later, in the 1970s (Forsdahl 1977; Barker and Osmond 1986). It has since then dominated the research, in both medical and social sciences.

Adverse conditions in early life influence cardiovascular diseases, respiratory and allergic diseases, diabetes, hypertension and obesity, breast and testicular cancers, neuropsychiatric and some other disorders later in life (Kuh and Ben-Shlomo 2004). Three specific diseases, respiratory tuberculosis, hemorrhagic stroke, and bronchitis, which have accounted for two thirds of the total decline in mortality in ages

³See Bijak and Bryant (2016) for a review and Alho and Spencer (1985), for early discussions.

15–64 years from the mid-nineteenth century to the first decade of the twentieth century, reflect demonstrable responses from conditions in infancy and childhood (see Chap. 21 by Lindström and Davey Smith). The question is then, if children early in life, during fetal stage or in first years of life are “programmed” for a certain health status, can this information be used to improve mortality forecasting?

Birth weight has become a common indicator of fetal development, both in medical and economic research. Low birth weight is associated with high blood pressure later in life. Despite the effects not being large (Huxley et al. 2002),⁴ this measure is still correlated with increased risks of heart disease (Barker 1994) and diabetes (Hales et al. 1991). Many diseases have since then been added to the list of outcomes associated with low birth weight, including schizophrenia (Brown and Derkets 2010). Recently, *high* birth weight, somewhat counterintuitively, has been associated with increased incidences of breast and prostate cancer (Risnes et al. 2011). Thus, the relationship between birth weight and diseases later in life appears to be U-shaped.

Low birth weight has also been shown to be linked to educational and labor market outcomes (Currie and Moretti 2007; Johnson and Schoeni 2011; Royer 2009). Children with low birth weight perform worse in school, and earn less, than children with normal birth weight do (Currie and Hyson 1999; Case et al. 2005). However, some studies have only found small effects of birth weight. A study of Norway using register data, for example, found that a 10% lower birth weight, a sizable difference from the norm is associated with only 1.2% lower IQ for males, 0.3% shorter height, and 0.9% less earnings (Black et al. 2007).

Birth weight does not change much in the long term, only during extreme situations, such as during wars and famines, if at all (Abolins 1962). For example, since 1860 in Norway, adult height, weight and lifespan have increased substantially, and still the average birth weight has only varied within a range of 200 g (Rosenberg 1988). Since the overall effects on both health and earnings are small, the birth-weight measure is unlikely to be an important indicator predicting mortality trends.

Instead, recent research has shown that infections due to certain bacteria, viruses, and parasites during pregnancy lead to health problems later in life (Adams Waldorf and McAdams 2013). Children exposed to the 1918 pandemic during the fetal stage, for example, show elevated risks for cancers and coronary heart disease at older ages (Mazumder et al. 2010; Myrskylä et al. 2013). Furthermore, children are vulnerable to maternal and environmental factors not only during the fetal stage but also in the first years of life. In fact, the development of cells and organs are gradually slowing down and completed only when the individual is about 20 years of age. Taking the lungs as an example, they not only keep growing in size after birth, but new alveolar sacs develop for several years, with this structures being also very sensitive to infections (Broman 1923).

⁴See also Chap. 22 by Christiansen.

Streptococcal infections in early childhood are associated with rheumatic heart disease in later life (Cunningham 2012), whereas respiratory infections have been suggested to impair lung function (Bengtsson and Lindström 2003).⁵ Recent research, studying effects of exposure to specific diseases, like whooping cough, in first years of life on mortality over the full life course shows that scarring overpowers selection at ages around 25 years, where after the net negative health effects increase with age (Quaranta 2014). Moreover, infections in early life are associated with cancer and diabetes at older ages (Finch and Crimmins 2004). They may also cause inflammation in atherosclerosis, which is a risk factor for a variety of diseases including cardiovascular diseases, diabetes and some forms of cancer later in life (Libby et al. 2002; Finch 2010).

Getting back to the question of whether information on conditions early in life can be used to improve mortality forecasting, we conclude that information on birth weights cannot be relied upon to predict life spans. Instead, infections in the first year of life, manifested in infant mortality, seems far more promising, as emphasized by the development of research in the last decade. However, since infant mortality has declined to very low levels during the twentieth century, this predictor is likely to lose importance over time, at least for developed countries. The same is likely the case for information on season of birth, which is another predictor of life expectancy.⁶ Instead, early life factors at the family level, which in the past were not associated with mortality, are likely to become more and more important (Bengtsson and Van Poppel 2011). Whether they can be used for mortality forecasting, is, however doubtful, since the effect is not strong enough to have any practical importance for the projections themselves. Alternatively, it seems more promising to use information on life-style factors such as obesity or smoking, which have become increasingly relevant predictors for health and mortality (Janssen et al. 2013; Neovius et al. 2010).

1.8 The Increasing Gap in Life Expectancy with Respect to Position in the Income Distribution

In this section, we discuss some issues that were not addressed in any of the five booklets. Recently, socioeconomic differences in length of life have received a lot of attention, in particular the fact that persons with low socio-economic status (income, education, occupation) tend to be worse off in terms of adult mortality than those with higher status. The gradient is stronger among men than among women, and has grown stronger in the recent past (Elo 2009; Mackenbach et al. 1997; Marmot et al. 1991; Torssander and Erikson 2010; Cutler et al. 2012; de Gelder et al. 2017; Chetty et al. 2016). While some argue that the socioeconomic differences in adult mortality

⁵ See also Chap. 24 by Bengtsson and Alter.

⁶ See Chap. 23 by Doblhammer.

have always existed, others are skeptical. Most of this research is based on cross-sectional data for cities. Studies based on longitudinal data for Sweden, the Netherlands, and Canada find barely any social differences in adult mortality in the past (Bengtsson and Dribe 2011; Bengtsson and Van Poppel 2011).

What characterizes individuals in the lower income deciles is that their *relative* income status remains largely unchanged throughout their working lives and follows them into retirement. Consequently, and not surprisingly, the result is a low pension in retirement, since all pensions are related one way or another to individual working life incomes. In addition, many studies for Western European and Anglo-Saxon countries show population differences in life expectancy with respect to other socioeconomic characteristics than income, such as education and occupation (National Academies of Sciences 2015; Chetty et al. 2016). This research has also shown that income gaps we see at retirement, originate already at younger ages, sometimes even earlier than age 40. What's more, it is likely that differences in life expectancy by income group are partially at least indirectly related to level of education and choice of occupation, as the empirical evidence reveals that persons with low education and low-skilled occupations have lower life expectancy than highly educated persons with occupations requiring high intellectual skills.

What these studies make clear is that the gap in life expectancy by income has several important characteristics. The first is that seen over a longer period of time, life expectancy in some populations is largely stagnant in the low deciles. Persons with higher income and those who live in more economically dynamic local cultures are experiencing growth in life expectancy. Geography, i.e., the social environment of peoples' home communities, also turns out to be important. For example, in New York City, Chetty et al. (2016) find that the life expectancy of low-income individuals is 5 years more than in Detroit, which is attributed to differences in "social" infrastructure. They also find that the strongest determinant of life expectancy differentials across income deciles is the prevalence of unhealthy life-style factors (smoking, drinking, exercise, dietary, etc.) among persons in the lowest deciles.⁷ One has to bear in mind, however, that causality runs both ways: not only from low income (through bad health) to low life expectancy, but also the other way around.

What are the ramifications in the present context? The average total working-life income of women is generally less than that of men, which means that women are likely to be overrepresented as low-income individuals in *all* income deciles. The reasons are well-known: this is in part due to an uneven division of child care and care of elderly relatives, in part due to lower earnings for equal work and lower wages in female dominated occupations (such as the care sector and sales personnel), partly due to individual decisions to work part time even as children become older. It

⁷On the other hand, they found that access to health care services was not a significant factor, indicating that provision is generally sufficient on a regional level in the United States. They also report that, as a group, first generation immigrants have a longer life expectancy than persons born in the United States.

comes as no surprise then that the gender difference in the scale of pensions at retirement is generally a reflection of women's working careers.

Despite their lower income, however, the life expectancy of women exceeds that of men. This gender gap in life expectancy leaves its imprint on the redistribution of income within a pension insurance pool. As a group, since women live several years longer than men, money is ultimately transferred from men to women. This has generally been viewed as an acceptable transfer of resources, partially compensating within the insurance pool for women's lower share of the income of male-female partners prior to retirement.

The general picture emerging from the empirical literature is nevertheless that the poor are increasingly transferring income to the rich in the universal unisex old-age insurance pool, if we focus singularly on how annuities are created. And, as a consequence, the literature is beginning to produce proposals of how to even out the distributional outcomes within the insurance pool, in the context of a public pension scheme.

There are two proposals – in addition to maintaining the status quo in insurance pool distribution. One is a technical tax-transfer solution that can be designed to redistribute money *within the annuity pension pool* (Holzmann et al. in preparation). Alternatively, information on the socio-economic determinants of life expectancy can be used to create separate insurance pools. For instance, a simple procedure is to stratify new retirees into separate, say, birth-cohort decile-based insurance pools by income, as measured by Defined Contributions (DC) account balances at the commencement of retirement.

However, stratifying by income is already what is happening as the default in the overall DC format in countries that have NDC and/or FDC public pension schemes. In addition, status quo is that persons with the lowest working-life income in all countries already receive some form of guarantee or social pension, often means-tested – financed by general tax revenues. This may nevertheless be a first-best solution, provided the amount transferred through the guarantee contributes sufficient income for the least well-off 20–30 percent of the pensioners' population.

What's more, in a broader perspective, the amount of money on pension account balances is by far not the only component of overall personal resources, which include capital income (also with a high-income profile) and other sources that one would normally take into account in granting a means-tested income transfer for retired persons. Therefore, in practice the status quo approach might not be so bad after all. In addition to this, the economists will remind us to bring into the discussion individuals' supply of labor, and its determinants, which explains individual lifetime earnings – and ultimately the magnitude of pensions granted – through individual preferences for work and leisure.

Let us return now to the discussion from a previous section – that of the development of mortality and forecasting life expectancy. The point of departure was the ongoing increase in life expectancy of the older population and developing projection models that produce unbiased cohort-based forecasts over time. Of course, the emerging empirical evidence regarding socio-economic heterogeneity in life expectancy has implications for projections, even only if implicitly.

To begin with, we are gaining a better understanding of the underlying mechanisms. The knowledge emanating from the empirical research (Chetty et al. 2016), together with the many “bits and pieces” from other studies, suggests that there remains considerable potential for further dramatic improvements in life expectancy. This would occur through improved community social-infrastructure and public health policy directed towards creating awareness of individuals’ “self-destructive” life styles. If this is the direction in which we are moving, the result will be that a larger share of birth cohorts will reach the pension age and beyond, a share that will enjoy healthier years in old age. This would create then even more decades of considerable improvements in life expectancy, adding to the effect of continued advancements in medical technology.

References

- Abolins, J. (1962). Weight and length of the newborn during the Years 1918–1922 and 1938–1942. *Acta Paediatrica*, 51(S135), 8–13.
- Adams Waldorf, K., & McAdams, R. (2013). Influence of infection during pregnancy on fetal development. *Reproduction*, 146(5), R151–R162.
- Alho, J. M. (1991). Effect of aggregation on the estimation of trend in mortality. *Mathematical Population Studies*, 3(1), 53–67.
- Alho, J. (2016). *Descriptive findings on the convergence of female and male mortality in Europe* (Technical report, ETLA Working Papers 40).
- Alho, J. M., & Spencer, B. D. (1985). Uncertain population forecasting. *Journal of the American Statistical Association*, 80(390), 306–314.
- Alho, J., & Spencer, B. (2005). *Statistical demography and forecasting* (Springer Series in Statistics). New York: Springer.
- Alho, J. M., Jensen, S. E. H., & Lassila, J. (2008). *Uncertain demographics and fiscal sustainability*. Cambridge: Cambridge University Press.
- Alho, J., Bravo, J., & Palmer, E. (2013). Annuities and life expectancy in NDC. In R. Holzmann, E. Palmer, & D. Robalino (Eds.), *Nonfinancial defined contribution pension schemes in a changing pension world* (Vol. 2, pp. 395–436). Washington, DC: World Bank.
- Barker, D. J. P. (1994). *Mothers, babies and disease in later life*. London: BMJ Publishing Group.
- Barker, D. J. P., & Osmond, C. (1986). Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales. *The Lancet*, 327(8489), 1077–1081.
- Bengtsson, T., & Dribe, M. (2011). The late emergence of socioeconomic mortality differentials: A micro-level study of adult mortality in southern Sweden 1815–1968. *Explorations in Economic History*, 48, 389–400.
- Bengtsson, T., & Lindström, M. (2003). Airborne infectious diseases during infancy and mortality in later life in southern Sweden, 1766–1894. *International Journal of Epidemiology*, 32(2), 286–294.
- Bengtsson, T., & Mineau, G. P. (2009). Early-life effects on socio-economic performance and mortality in later life: A full life-course approach using contemporary and historical sources. *Social Science & Medicine*, 68(9), 1561–1564.
- Bengtsson, T., & Scott, K. (2010). The ageing population. In *Population ageing – a threat to the welfare state? The case of Sweden* (pp. 7–22). Heidelberg: Springer.
- Bengtsson, T., & Van Poppel, F. (2011). Socioeconomic inequalities in death from past to present: An introduction. *Explorations in Economic History*, 48(3), 343–356.

- Bergeron-Boucher, M. P., Canudas-Romo, V., Oeppen, J., & Vaupel, J. W. (2017). Coherent forecasts of mortality with compositional data analysis. *Demographic Research*, 37, 527–566.
- Bijak, J. (2011). *Forecasting international migration in Europe: A Bayesian view* (Vol. 24). Dordrecht: Springer Science & Business Media.
- Bijak, J., & Bryant, J. (2016). Bayesian demography 250 years after Bayes. *Population Studies*, 70 (1), 1–19.
- Black, S. E., Devereux, P. J., & Salvanes, K. G. (2007). From the cradle to the labor market? The effect of birth weight on adult outcomes. *The Quarterly Journal of Economics*, 122(1), 409–439.
- Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting*, 22(3), 547–581.
- Borgan, O., & Keilman, N. (2016). Do Japanese and Italian women live longer than women in Scandinavia? *European Journal of Population*, 1–13.
- Broman, I. (1923). Zur Kenntnis der Lungenentwicklung. *Anatomischer Anzeiger*, 57, 83.
- Brown, A. S., & Derkets, E. J. (2010). Prenatal infection and schizophrenia: A review of epidemiologic and translational studies. *American Journal of Psychiatry*, 167(3), 261–280.
- Canudas-Romo, V. (2008). The modal age at death and the shifting mortality hypothesis. *Demographic Research*, 19, 1179–1204.
- Canudas-Romo, V., & Schoen, R. (2005). Age-specific contributions to changes in the period and cohort life expectancy. *Demographic Research*, 13(3), 63–82.
- Case, A., Fertig, A., & Paxson, C. (2005). The lasting impact of childhood health and circumstance. *Journal of Health Economics*, 24(2), 365–389.
- Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A., & Cutler, D. (2016). The association between income and life expectancy in the United States, 2001–2014. *Journal of the American Medical Association*, 315(16), 1750–1766.
- Coale, A. J. (1957). How the age distribution of a human population is determined. In *Cold spring harbor symposia on quantitative biology* (Vol. 22, pp. 83–89). New York: Cold Spring Harbor Laboratory Press.
- Cunningham, M. W. (2012). Streptococcus and rheumatic fever. *Current Opinion in Rheumatology*, 24(4), 408.
- Currie, J., & Hyson, R. (1999). Is the impact of health shocks cushioned by socioeconomic status? The case of low birthweight. *American Economic Review*, 89(2), 245–250.
- Currie, J., & Moretti, E. (2007). Biology as destiny? Short-and long-run determinants of intergenerational transmission of birth weight. *Journal of Labor Economics*, 25(2), 231–264.
- Cutler, D. M., Lleras-Muney, A., & Vogl, T. (2012). Socioeconomic Status and health: Dimensions and mechanisms. In M. Gleid & P. Smith (Eds.), *The Oxford handbook of health economics* (pp. 124–163). Oxford: Oxford University Press.
- de Gelder, R., Menvielle, G., Costa, G., Kovács, K., Martikainen, P., Strand, B. H., & Mackenbach, J. P. (2017). Long-term trends of inequalities in mortality in 6 European countries. *International Journal of Public Health*, 62(1), 127–141.
- Derrick, V. (1927). Observations on (1) errors of age in the population statistics of England and Wales, and (2) the changes in mortality indicated by the national records. *Journal of the Institute of Actuaries*, 58(2), 117–159.
- Ediev, D. M. (2013). Mortality compression in period life tables hides decompression in birth cohorts in low-mortality countries. *Genus*, 69(2), 53–84.
- Elo, I. T. (2009). Social class differentials in health and mortality: Patterns and explanations in comparative perspective. *Annual Review of Sociology*, 35, 553–572.
- Finch, C. E. (2010). Evolution of the human lifespan and diseases of aging: Roles of infection, inflammation, and nutrition. *Proceedings of the National Academy of Sciences*, 107(suppl 1), 1718–1724.
- Finch, C. E., & Crimmins, E. M. (2004). Inflammatory exposure and historical changes in human life spans. *Science*, 305(5691), 1736–1739.

- Forsdahl, A. (1977). Are poor living conditions in childhood and adolescence an important risk factor for arteriosclerotic heart disease? *Journal of Epidemiology & Community Health*, 31(2), 91–95.
- GAD. (2001). *National population projections: Review of the methodology for projecting mortality* (Technical report). London: Government Actuary's Department.
- Girosi, F., & King, G. (2008). *Demographic forecasting*. Princeton: Princeton University Press.
- Goldstein, J. R., & Wachter, K. W. (2006). Relationships between period and cohort life expectancy: Gaps and lags. *Population Studies*, 60(3), 257–269.
- Hahn, L. J., & Christiansen, M. C. (2016). Mortality projections for non-converging groups of populations. *SSRN Papers*.
- Hales, C. N., Barker, D. J. P., Clark, P. M., Cox, L. J., Fall, C., Osmond, C., & Winter, P. (1991). Fetal and infant growth and impaired glucose tolerance at age 64. *BMJ*, 303(6809), 1019–1022.
- Holzmann, R., Palmer, E., Palacios, R., & Sacchi, S. (in preparation). *Non-financial defined contribution pension schemes: Facing the challenges of marginalization and polarization in the economy and society*.
- Huxley, R., Neil, A., & Collins, R. (2002). Unravelling the fetal origins hypothesis: Is there really an inverse association between birthweight and subsequent blood pressure? *The Lancet*, 360 (9334), 659–665.
- Janssen, F., & 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.
- Janssen, F., & van Poppel, F. (2015). The adoption of smoking and its effect on the mortality gender gap in Netherlands: A historical perspective. *BioMed Research International*, 2015. Article ID 370274.
- Janssen, F., van Wissen, L. J., & Kunst, A. E. (2013). Including the smoking epidemic in internationally coherent mortality projections. *Demography*, 50(4), 1341–1362.
- Johnson, R. C., & Schoeni, R. F. (2011). The influence of early-life events on human capital, health status, and labor market outcomes over the life course. *The BE Journal of Economic Analysis & Policy*, 11(3).
- Keilman, N. (2017). A combined Brass-random walk approach to probabilistic household forecasting: Denmark, Finland, and the Netherlands, 2011–2041. *Journal of Population Research*, 34 (1), 17–43.
- Keilman, N., & Pham, D. (2004). *Empirical errors and predicted errors in fertility, mortality and migration forecasts in the European Economic Area* (Discussion Paper No. 386). Oslo: Statistics Norway.
- Keilman, N., Syse, A., & Pham, D. (2018). *Mortality shifts and mortality compression: The case of Norway, 1900–2060*. Submitted.
- Kermack, W. O., McKendrick, A. G., & McKinlay, P. L. (1934). Death rates in Great Britain and Sweden some general regularities and their significance. *The Lancet*, 223(5770), 698–703.
- King, G., & Soneji, S. (2011). The future of death in America. *Demographic Research*, 25, 1.
- Kleinow, T. (2015). A common age effect model for the mortality of multiple populations. *Insurance: Mathematics and Economics*, 63, 147–152.
- Kuh, D., & Ben-Shlomo, Y. (2004). *A life course approach to chronic disease epidemiology* (Number 2). Oxford: Oxford University Press.
- Lassila, J., Valkonen, T., & Alho, J. M. (2014). Demographic forecasts and fiscal policy rules. *International Journal of Forecasting*, 30(4), 1098–1109.
- Lee, R. D. (1994). The formal demography of population aging, transfers, and the economic life cycle. In L. Martin & S. Preston (Eds.), *Demography of aging* (pp. 8–49). Washington, DC: National Academy Press.
- Lee, R. D. (1998). Probabilistic approaches to population forecasting. *Population and Development Review*, 24, 156–190.
- Lee, R. D., & Mason, A. (Eds.). (2011). *Population aging and the generational economy: A global perspective*. Cheltenham: Edward Elgar Publishing.

- Lee, R. D., & Zhou, Y. (2017). Does fertility or mortality drive contemporary population aging? The revisionist view revisited. *Population and Development Review*, 43(2), 285–301.
- Li, J. (2013). A Poisson common factor model for projecting mortality and life expectancy jointly for females and males. *Population Studies*, 67(1), 111–126.
- Libby, P., Ridker, P. M., & Maseri, A. (2002). Inflammation and atherosclerosis. *Circulation*, 105(9), 1135–1143.
- Lindahl-Jacobsen, R., Rau, R., Jeune, B., Canudas-Romo, V., Lenart, A., Christensen, K., & Vaupel, J. W. (2016). Rise, stagnation, and rise of Danish women's life expectancy. *Proceedings of the National Academy of Sciences*, 113(15), 4015–4020.
- Mackenbach, J. P., Kunst, A. E., Cavelaars, A. E., Groenhof, F., Geurts, J. J., Health, E. W. G., & o. S.I. i., and others. (1997). Socioeconomic inequalities in morbidity and mortality in Western Europe. *The Lancet*, 349(9066), 1655–1659.
- Marmot, M. G., Stansfeld, S., Patel, C., North, F., Head, J., White, I., Brunner, E., Feeney, A., & Smith, G. D. (1991). Health inequalities among British civil servants: The Whitehall II study. *The Lancet*, 337(8754), 1387–1393.
- Mazumder, B., Almond, D., Park, K., Crimmins, E. M., & Finch, C. E. (2010). Lingering prenatal effects of the 1918 influenza pandemic on cardiovascular disease. *Journal of Developmental Origins of Health and Disease*, 1(1), 26–34.
- Missov, T. I., & Lenart, A. (2011). Linking period and cohort life-expectancy linear increases in Gompertz proportional hazards models. *Demographic Research*, 24, 455–468.
- Murphy, M. (2017). Demographic determinants of population aging in Europe since 1850. *Population and Development Review*, 43(2), 257–283.
- Myrskylä, M., Mehta, N. K., & Chang, V. W. (2013). Early life exposure to the 1918 influenza pandemic and old age mortality by cause of death. *American Journal of Public Health*, 103(7), e83–e90.
- National Academies of Sciences, Engineering, a. M. (2015). *The growing gap in life expectancy by income: Implications for federal programs and policy responses*. Washington, DC: The National Academies Press.
- Neovius, K., Rasmussen, F., Sundström, J., & Neovius, M. (2010). Forecast of future premature mortality as a result of trends in obesity and smoking: Nationwide cohort simulation study. *European Journal of Epidemiology*, 25(10), 703–709.
- Oeppen, J., & Vaupel, J. W. (2002). Broken limits to life expectancy. *Science*, 296(5570), 1029–1031.
- Olshansky, S. J., Carnes, B. A., & Désesquelles, A. (2001). Prospects for human longevity. *Science*, 291(5508), 1491–1492.
- Ouellette, N., & Bourbeau, R. (2011). Changes in the age-at-death distribution in four low mortality countries: A nonparametric approach. *Demographic Research*, 25, 595–628.
- Palmer, E., Alho, J. M., & Zhao de Gosselin de Varennes, Y. (2018). *Projecting cohort life expectancy from the changing relationship between period and cohort mortalities*. Submitted.
- Pascariu, M. D., Canudas-Romo, V., & Vaupel, J. W. (2018). The double-gap life expectancy forecasting model. *Insurance: Mathematics and Economics*, 78, 339–350.
- Qi, H. (2016). Live longer, work longer? Evidence from Sweden's ageing population. In *Lund studies in economic history* (Vol. 75). Lund: Lund University.
- Quaranta, L. (2014). Early life effects across the life course: The impact of individually defined exogenous measures of disease exposure on mortality by sex in 19th- and 20th-century Southern Sweden. *Social Science and Medicine*, 119(Online May 15, 2014), 266–273.
- Raftery, A. E., Li, N., Ševčíková, H., Gerland, P., & Heilig, G. K. (2012). Bayesian probabilistic population projections for all countries. *Proceedings of the National Academy of Sciences*, 109(35), 13915–13921.
- Raftery, A. E., Lalic, N., & Gerland, P. (2014). Joint probabilistic projection of female and male life expectancy. *Demographic Research*, 30, 795.

- Risnes, K. R., Vatten, L. J., Baker, J. L., Jameson, K., Sovio, U., Kajantie, E., Osler, M., Morley, R., Jokela, M., & Painter, R. C. (2011). Birthweight and mortality in adulthood: A systematic review and meta-analysis. *International Journal of Epidemiology*, 40(3), 647–661.
- Rosenberg, M. (1988). Birth weights in three Norwegian cities, 1860–1984. Secular trends and influencing factors. *Annals of Human Biology*, 15(4), 275–288.
- Royer, H. (2009). Separated at birth: US twin estimates of the effects of birth weight. *American Economic Journal: Applied Economics*, 1(1), 49–85.
- Shkolnikov, V. M., Jdanov, D. A., Andreev, E. M., & Vaupel, J. W. (2011). Steep increase in Best-practice Cohort life expectancy. *Population and Development Review*, 37(3), 419–434.
- Staetsky, L. (2009). Diverging trends in female old-age mortality: A reappraisal. *Demographic Research*, 21, 885.
- Stoeldraijer, L., van Duin, C., van Wissen, L., & 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, 323.
- Thorslund, M., Wastesson, J. W., Agahi, N., Lagergren, M. a., & Parker, M. G. (2013). The rise and fall of women's advantage: A comparison of national trends in life expectancy at age 65 years. *European Journal of Ageing*, 10(4), 271–277.
- Torssander, J., & Erikson, R. (2010). Stratification and mortality—A comparison of education, class, status, and income. *European Sociological Review*, 26(4), 465–474.
- Tuljapurkar, S., & Boe, C. (1998). Mortality change and forecasting: How much and how little do we know? *North American Actuarial Journal*, 2(4), 13–47.
- Tuljapurkar, S., & Edwards, R. D. (2011). Variance in death and its implications for modeling and forecasting mortality. *Demographic Research*, 24, 497.
- UN. (2015). *World population ageing, 1950–2050*. New York: United Nations Publications.
- van Berkum, F., Antonio, K., & Vellekoop, M. (2017). A Bayesian joint model for population and portfolio-specific mortality. *ASTIN Bulletin: The Journal of the IAA*, 47(3), 681–713.
- Waldron, H. (2005). Literature review of long-term mortality projections. *Social Security Bulletin*, 66(1), 16.
- Wilmoth, J. R. (2005). On the relationship between period and cohort mortality. *Demographic Research*, 13, 231–280.
- Zhao de Gossen de Varennes, Y., Palmer, E., & Alho, J. M. (2016). Projecting cohort life expectancy based on its rate of change in mortality. In *Benefit design, retirement decisions and welfare within and across generations in defined contribution pension schemes* (Economic studies, Vol. 157). Uppsala: Uppsala University.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Part I

Current Practice

Tommy Bengtsson and Nico Keilman

This part focuses on current practices of mortality forecasting in Sweden, Norway, and Finland with some reflections on what can be done to improve them. The first paper, written by Edward Palmer, introduces the question of why we are interested in improving mortality projections. To ease the stress of population ageing on public pension systems, several countries have changed from defined benefit to defined contribution systems, including Sweden. For these countries, life expectancy projections are used to calculate individual pensions. To get financially robust systems, that also are fair across generations, it is very important to calculate life expectancy accurately, since it determines the size of the pensions.

Juha M. Alho, Helge Brunborg, and Hans Lundström discuss current projection models in Finland, Norway and Sweden, respectively. Although the methods used consider cohort trends, and projections are updated regularly, they underestimate the progress of life expectancy, as has been the case for forecasts by national agencies in other countries. This means that current contribution based pension systems are too generous since the benefits are based on an underestimation of remaining life expectancy. It also means that the pension systems are under-financed, which creates a burden on future generations.

James Vaupel presents the SCOPE approach to forecasting life expectancy. It is a kind of scenario method, with scenarios structured conditionally, and with the possibility of stochastic scenarios. Kaare Christensen discusses how to improve existing models from the perspective of an epidemiologist, focussing on cohort approaches. If we know the “risk profile” of the current cohorts compared to the previous cohorts, then our forecasts may improve. Self-rated health, as well as

T. Bengtsson

Centre for Economic Demography, Lund University, Lund, Sweden
e-mail: tommy.bengtsson@ekh.lu.se

N. Keilman

Department of Economics, University of Oslo, Oslo, Norway
e-mail: nico.keilman@econ.uio.no

physical and cognitive abilities, are possible indicators of remaining life expectancy, which may provide an opportunity to improve forecasting. Finally, Tommy Bengtsson stresses the role of historical experience in getting forecasting right. He argues that since the cause of the great mortality decline is clearly multi-factorial, and the importance of the various factors changes over time, both period and cohort factors must be taken into account when forecasting mortality. The use of demographic cohort information in the 1990s was a considerable advancement. The question now is whether we are ready to take a great step forward by using multivariate causal models that combine information at the individual and societal levels.

Chapter 2

Life Expectancy Is Taking Center Place in Modern National Pension Schemes – A New Challenge for the Art of Projecting Mortality



Edward Palmer

2.1 Introduction

Mortality scenarios are a standard tool in projecting costs of defined benefits (DB) pay-as-you-go pension systems. Government actuaries and statisticians have used mortality projections to estimate the costs of national pay-as-you go defined benefit pension systems since their conception, although the development of computer technology in the 1970s and 1980s was a prerequisite for more sophisticated analysis. With the maturation of computer technology, the possibilities to model and examine various assumptions are nowadays more-or-less unbounded. We have become used to official publications from, for example, the Office of the Government Actuary in the US or, in Sweden, the National Social Insurance Board, that provide a picture of future contributions and payments given various demographic and economic scenarios. These reports provide a picture of the financial development of national pension systems.

Financial projections illuminating the financial status of pension systems are indisputably important, not the least because pay-as-you-go systems transfer considerable resources from workers to pensioners. As life expectancy increases, and assuming current retirement patterns, 25–30% of the total population in the OECD will be pensioners within the coming three decades. There is now considerable debate in countries all around the world about how to cope with the aging population. The World Bank's publication *Averting the Old Age Crisis* (1994) constituted a milestone in the debate. The World Bank recommended that its client countries adopt multi-pillar schemes, with a funded “second pillar” playing an important role.

E. Palmer (✉)

Social Insurance Economics, Uppsala University, Uppsala, Sweden

Research Division, Swedish National Social Insurance Board, Stockholm, Sweden
e-mail: edward.e.palmer@gmail.com

The Swedish pension reform, first published as a proposal in Swedish in 1992, and then legislated by parliament as a framework law in early 1994, introduced the notional defined contribution scheme to the international political arena. Since the mid-1990s, there has been a large-scale conversion of national pension schemes from defined benefit (DB) to defined contribution (DC) – both notional defined contribution (NDC) and financial defined contribution (FDC). As countries convert to DC schemes, life expectancy projections are used to *calculate individual pensions* in national systems.

The purpose of this paper is to summarize the dramatic change that has occurred in the design of national pension schemes in the past 10 years and to discuss the need for accurate mortality projections in this context.

2.2 Basic Pension Economics – The Role of Mortality

The financial costs of a pay-as-you-go pension system are determined by a few key parameters. In the aggregate these are the size of an average benefit relative to the average wage, the number of workers and the number of pensioners. The ratio of the latter two is the system dependency ratio. The system dependency ratio can also be computed as the number of years people work related to the number of years they are retired with a benefit. These are two sides of the same coin. The policy parameters of a country are the size of a pension and the level of the contribution rate. The demographic and economic determinants of pensions are largely exogenous to policy, although these can respond to system design.

These ideas follow from the two fundamental equations of pension economics, which provide the point of departure for an overview of the links between mortality and pay-as-you-go pension systems. Where bars denote averages, and p and w depict pensions and wages, these equations are:

$$\text{Contribution rate} = \frac{\bar{p}}{w} * \frac{\text{pensioners}}{\text{contributors}}$$

or for the average participant in the scheme (see e.g. Palmer 1999)

$$\text{Contribution rate} = \frac{\bar{p}}{w} * \frac{\text{years in retirement}}{\text{years of work}}.$$

The first equation tells us that pension costs, measured in terms of the contribution rate on wages, is determined by the average benefit times the number of recipients, and the contribution base, expressed in the first equation as the product of the average wage and the number of contributors. If pension payments grow faster than contributions, the contribution rate needed to finance them increases, and vice versa. The second equation expresses the same relationship for the average scheme participant in terms of years worked and years of retirement. It tells us that more years of retirement in relation to years of work will result in higher pension costs.

Mortality affects both the numerator and the denominator of the first equation. Declining mortality among working age persons increases the size of the working

age population and the (potential) labor force. Decreasing mortality of persons above the pension age increases the population of retirees. In other words, the age distribution of mortality improvements matters.

In terms of the second equation, increasing mortality from the age of retirement implies that people will have to work more years – at the same wage – to receive a benefit of a given level and with a fixed contribution rate. Conversely, if the pay-as-you-go system is designed so that a benefit is defined, for example, in terms of the number of covered years of work and contributions required to qualify for a benefit of a fixed amount, then declining mortality, *i.e.* more years in retirement, increases the contribution rate that the younger generation of workers has to pay in order to sustain the fixed level benefit over more years.

This is a static picture, however. Assume now that the labor force grows at the rate λ and productivity at the rate g , and that the average wage grows at the rate of growth of productivity. Then the first equation tells us a number of things. First, if benefits are inflation-indexed, which is a common form of indexation for pay-as-you-go systems, then positive values of productivity and labor force growth will counteract negative effects on costs of increasing longevity after retirement. During the past two decades some countries have taken advantage of this mechanism to reduce costs, by replacing wage indexation of benefits with price indexation.¹

The first equation also tells us that a country can afford *wage* indexation as long as the labor force grows faster than the number of pensioners. With wage indexation, the real-valued (inflation adjusted) pension grows with the rate of increase in productivity, and productivity growth is shared between workers and pensioners.² Many countries have seen this as a desirable redistributive aspect of a pay-as-you-go pension system. Whether or not this level of ambition can be attained depends on the connection between the construction of the pension system and the demography behind changes in the work force and the number of pensioners.

Assume now net migration of zero and a birth rate that reproduces the population. Maintaining a fixed ratio between the average pension and the average wage through wage indexation of benefits over time requires that the positive effect on the work force of declining mortality surpass the increase in pension payments resulting from decreased mortality among the retired population.

Of course, the effect on the working population of decreasing mortality will have to be associated with sufficiently healthy years of life to make a difference.

¹The UK did this in the early 1980s. This has also been one of the mechanisms used by transition country governments during the 1990s to cut back on burdensome pension costs – where the immediate problem was a dependency ratio around 1.5 workers per pensioner, due to very generous pension ages inherited from the old communist regimes.

²Denoting the rate of inflation by p , the *nominal* wage rate grows at $(1 + p)(1 + g)$, as does the *nominal* wage base. For a fixed ratio of pensioners to contributors (the first equation) it is possible to index benefits with this factor, and maintain a constant contribution rate, since the numerator increases at the same rate as the denominator.

In addition, the economy must be able to employ the extra labor created in this way.³ Employability is usually viewed as a short-run problem, however, as the growth of the working age population and labor supply are usually regarded as the long-run determinants of a country's economic growth, together with the rate of growth of productivity.

The second equation provides technical insight into how a pay-as-you-go system can be designed to adjust to changes in the life expectancy of pensioners. Given that people continue to work with the average wage, increasing life expectancy after retirement can be dealt with by increasing the minimum benefit age at the rate of change in the life expectancy of retirees. This also requires that people work and contribute during the additional time prior to a retirement age⁴ that is sufficient to maintain a fixed ratio of years of retirement to years of work. Alternatively, the pay-as-you-go scheme can be constructed as a notional defined contribution scheme as discussed below.

2.3 NDC and FDC Schemes – And Life Expectancy

A notional defined contribution – NDC – scheme is the pay-as-you-go equivalent of the financial defined contribution – FDC – scheme. The difference is in the rate of return, which in the NDC scheme is based on economic growth, whereas it is a financial rate of return in the FDC scheme. In the neo-classical Golden Rule, these are equivalent in the long run. However, many current writers make the claim that FDC schemes should always be expected to yield a higher return (*e.g.*, Feldstein and Samwick 1997, 1998). This claim is based on the observation that the financial rate of return has surpassed the rate of economic growth over the past half century in the US. Data for Sweden (*e.g.*, Frennberg and Hansson 1992) yield a similar result, at least up to the fall of the stock market in 2001–2002.

How does an NDC scheme work in practice? In the NDC public scheme, just as in any public or privately managed FDC scheme, wage earners pay contributions based on a fixed contribution rate. The value of these are accredited their accounts – this is the defined-contribution feature of the system. Contributions are paid on earnings as long as people work, and if people combine work with a pension then they continue to pay contributions on earned income and increase their pension capital accordingly. The previous year's account value is indexed annually with a nominal per

³In addition, the years added to the older population need to be (relatively) healthy years in order not to create other social costs, for example, increasing costs for health and home care of the elderly. In a broader model of the social “costs” of caring for the elderly, *e.g.*, health and home care, could be added to the numerator of the first equation by adding the cost per capita.

⁴In a DB scheme where the right to a full benefit is based on a certain number of years, *e.g.*, 40, it might be necessary to increase this number to assure that the period of working and contributing also become longer as the minimum pension age is increased.

capita wage index in Sweden,⁵ where the system was conceived in 1992–1994. The wage sum is used for indexation in Latvia and GDP in Italy, both of which were legislated in 1995.

The NDC annuity is calculated by dividing the value on the account at the chosen age of retirement with a factor based on unisex life expectancy at the age of retirement. In addition to this, in Sweden a real rate of return of 1.6% and in Italy 1.5% is calculated into the annuity. This form of front-loading is an alternative to possible wage indexation (from a lower initial level) over the lifetime. Annuities are also indexed to annual changes in prices in both countries. Annual indexation in the Swedish scheme also includes an adjustment for digressions of actual real growth from 1.6%, and if estimated system liabilities exceed assets a balancing adjustment to bring the system back into financial equilibrium. This keeps the aggregate contribution rate in line with the individual contribution rate of 16%.⁶

Góra and Palmer (2002) have recently made the claim that the main difference between NDC and FDC is in the nature of the “fund,” which in the NDC case can be seen as a fund of bonds bearing the rate of return of the wage sum (tax base) in the economy, where it is a needless exercise to sell the bonds on the market fund, since this simply creates transaction costs. These authors also note that both NDC and FDC funds are illiquid until retirement, and that from retirement both are paid out as yearly annuities. Both FDC and NDC have the advantage that they eliminate negative externalities by creating a direct link between contributions and benefits. On the other hand, an FDC scheme is associated with the positive externality of creating financial funds. If these are placed in non-government debt instruments they contribute to financing private investment and – if they do not offset private saving – provide additional financing for economic growth. The major difference between NDC and FDC is, thus, that NDC does not create this opportunity. On the other hand, an FDC scheme that invests solely in government bonds can be viewed as a cost-inefficient NDC scheme, since the transaction costs of marketing the debt do not create offsetting revenues.

Proponents of NDC claim that it represents a paradigm shift in social security thinking (*e.g.*, Palmer 2000). By creating a direct link between contributions and the annuity and by basing the size of the annuity on life expectancy at retirement, NDC

⁵In principle, a system must follow the development of contributions, i.e. the wage sum on which contributions are based, in order to maintain financial equilibrium. This means that if the per capita wage is used to index notional capital and benefits, then the system must be equipped with a brake that keeps it in financial equilibrium when labor force growth (λ) is negative. In the Swedish system a financial balance is kept that relates estimated system assets to liabilities. *Ceteris paribus*, if liabilities exceed assets because labor force growth is negative, both benefits and notional capital will be indexed downwards to bring the system back into equilibrium (See Palmer 2000, 2002; Settergren 2001).

⁶In principle, the Italian system should achieve a long-term equilibrium around the weighted contribution rates of the employed and self-employed if contributions and accrual factors are brought more in line. In practice, the absence of a mechanism to offset chronic divergence from the imputed return of 1.5% may lead to financial difficulties. Palmer (1999) examines the stability conditions of the NDC PAYGO system, as does Valdés-Prieto (2001).

systems reduce the impact on system costs of individual behavioral choice and of unexpected changes in longevity. In comparison, in the DB framework the burden of the risk is unclear. It may fall on future generations of workers or on present workers before they retire.

NDC pension schemes are subject to the same “political risks” as DB pay-as-you-go – “political” management of public funds causing low rates of return, special interest lobbying, etc. For example, in Italy, the rate credited into the notional account is actually higher than the payroll tax earmarked for pensions, while in Poland it is lower, which is synonymous to taxing accounts. The NDC provides a framework for monitoring the costs of these interventions, as Fox and Palmer (1999) have argued in discussing the Latvian NDC scheme.

An NDC scheme with demographic reserves and indexation of notional capital and benefits that follows the growth of the contribution base, and with an annuity based on life expectancy projections that on average do not deviate from the outcome for birth cohorts, yields approximate long-run financial stability (Palmer 1999). If as in Sweden indexation follows the per capita wage rate and life expectancy is based on current outcomes – rather than a cohort projection – then there is a built in risk that the assets of the system will fall below the liabilities. Financial balance is secured in Sweden through a balance mechanism, based on current estimates of system assets and liabilities.

Briefly, the Swedish balance mechanism works as follows. If assets, measured as the estimated future stream of revenues from contributions and the current value of the system’s buffer fund(s), are less than liabilities, *i.e.*, claims on future payments of pensions of pensioners and the notional account values of workers, the balance index falls below unity. In this case, both notional account values and benefits are adjusted to bring the system back into balance.

The Swedish NDC scheme was started with a large buffer fund inherited from the old system – these funds will help to cover the demographic pressure associated with the large birth cohorts of the 1940s. Various scenarios using relatively extreme demographic and economic assumptions show that it is likely that if the balance index has to be used in the future, the negative effect on future benefits during the whole retirement period of a pensioner is not likely to be more than 10 per cent in total, given some of the worse scenarios (Settergren et al. 2000).

Most importantly for the present topic, both FDC and NDC schemes work more efficiently with good expectancy projections. Poor projections give rise to a need for adjustment, and in general, do not provide the information needed by scheme participants to plan their “economic lives” given information about the development of their own cohort’s life expectancy and the expected value of their own stream of pension benefits.

2.4 It's More Important Than Ever to Project Life Expectancy Accurately

Public NDC and FDC schemes have been introduced in a number of countries since the mid-1990s.⁷ In addition, also since 1995, public FDC schemes have been introduced in a large number of Latin American countries, and have become popular in especially the transition countries of Central and Eastern Europe.⁸ These countries join, then, some of the forerunners of funded schemes, among them, Chile, Australia, Denmark, the Netherlands, Switzerland and the UK. The newer countries in the league differ from many of the forerunners, however, in their explicit DC construction. For example, the mandated employer schemes in Australia and the “opting out” schemes in the UK have been largely financial defined benefit (FDB) schemes, although there is a recent tendency for these also to convert to FDC. The difference between an FDB and an FDC scheme is small, however, since financial solidarity requires that life expectancy projections be on target in both.

In principle, there are two approaches that can be applied in estimating the life expectancy factor to be used in the calculation of annuities for NDC, and FDC, schemes. One is to base the projection on the current period tables – perhaps with some form of smoothed moving average. This is the approach applied by Sweden.⁹ The major alternative is to produce a cohort projection.

This is the approach applied by Latvia. The Swedish projection is revised yearly with new information on mortality, as is the Latvian projection.

The revision process differs in Latvia, however. Latvia bases changes in the projection of life expectancy on a professional judgment,¹⁰ with a demographic analysis as the point of departure. In the revision process, it is asked whether new information provides sufficient evidence to revise the existing long-term projection.¹¹ One could view this approach as a sort of error-correction mechanism. In principle, the methods available for projecting life expectancy for FDC schemes

⁷In addition to Sweden, Italy, Latvia and Poland, which have already been mentioned, versions of NDC schemes have been introduced in Kyrgyzstan, and in Russia.

⁸See Fox and Palmer (2001) for a discussion of the driving forces behind this movement.

⁹Although the life expectancy factor is continuously updated in Sweden, since people continue to live longer after receiving their pension, the procedure used to calculate it will underestimate the actual outcome. This is counterbalanced either through other, positive factors contributing to the financial balance, or through triggering the balance index (see above).

¹⁰The judgment is the product of a white paper written by a leading demographer and a discussion in an official committee of demographers and actuaries set up for this purpose. This is discussed in Krumins et al. (2001).

¹¹The procedure was initiated in 1999, and in the first 3 years thereafter, no revision of the long-term projection had been made.

are exactly those available for NDC. In the FDC context an effort is made to project a life expectancy factor that is expected to give long-run system solvency.¹²

In sum, as countries convert to DC schemes, life expectancy projections are used to *calculate individual pensions* in national systems. The question is whether the state of the art in projecting mortality can meet new demand created by large public schemes that are turning towards annuities based on life expectancy.

2.5 Final Comments

With conversion to public NDC and FDC schemes in Sweden,¹³ as well as elsewhere, individuals have been given greater responsibility to plan their own working careers and saving with respect to a desired level of resources during retirement. In the DC framework resources during retirement are linked to contributions during the whole working career, and the level of a benefit is also determined by life expectancy at the age chosen for retirement. In this framework, knowledge about the development of cohort life expectancy becomes an important informational input into the economic plans of individuals.

One of the goals of policy makers is to loosen up the idea of “a” pension age at which everyone is expected to exit the labor force, that is the concept of the “statutory” or “mandatory” pension age that implanted itself so deeply in the minds of employees, unions and employers since the 1960s. The focus is to shift from the “right” to leave the labor force at age, for example 60, with a defined lifelong benefit, to the “right” to work as an older worker – but in a work environment that is friendly to older workers.

The transition from national DB to DC schemes presupposes a future where people can freely choose between work and retirement – combining a partial or full benefit with partial or full retirement. It also presupposes that individuals invest enough in their human capital (personal health, education and training) to be able to remain in the labor force longer. Today the *de facto* age of exit from the labor force for men and women together is below 60 in the OECD countries. There is evidence that this has been influenced by national benefit schemes (Gruber and Wise 1999).

¹²In FDC schemes there is a trade off between the rate of return and life expectancy in the sense that inaccuracies in the projection of the life expectancy factor can be counterbalanced by a better rate of financial return on funded capital. In the Swedish NDC construction, since positive labor force gains are undistributed, i.e. indexation is with $(1 + g)(1 + p)$, these together with good returns on the buffer fund(s) can counterbalance the clear inaccuracy in the projection of life expectancy at retirement resulting from not attempting to account for some additional increase in life expectancy of cohorts after retirement.

¹³Note that Sweden, as other countries that have introduced NDC and FDC schemes, has a minimum guarantee, which has not been discussed in the present context.

Countries are now aiming to raise this age. National DC (NDC or FDC) schemes are a tool that can help promote this goal, and this is an important reason why countries are introducing them.¹⁴

To conclude, in order to plan for retirement in a DC environment, individuals need good information about their cohort's life expectancy. Work career and saving plans will then be formulated in accordance with this information, thereby determining lifetime resources and their distribution over the life cycle.

References

- Feldstein, M., & Samwick, A. (1997). *The economics of prefunding social security and medical care benefits* (NBER Working Paper, No. 6055). Cambridge, MA: NBER.
- Feldstein, M., & Samwick, A. (1998). *Two percent personal retirement accounts: Their potential effects on social security taxes and national saving* (NBER Working Paper No. 6540). Cambridge, MA: NBER.
- Fox, L., & Palmer, E. (1999). *Latvian pension reform* (Social Protection Discussion Paper No. 9922). Washington, DC: The World Bank.
- Fox, L., & Palmer, E. (2001). New approaches to multi-pillar pension systems: What in the world is going on? In J. Stiglitz et al. (Eds.), *New ideas about old age security*. Chicago: World Bank/University of Chicago Press.
- Frennberg, P., & Hansson, B. (1992). Swedish stocks, bonds, bills and inflation. *Applied Financial Economics*, 2, 79.
- Góra, M., & Palmer, E. (2002). *Shifting perspectives in pensions* (Working Paper). Warsaw: Center for Economic and Social Research (CASE).
- Gruber, J., & Wise, D. A. (1999). *Social security programs and retirement around the world*. Chicago: University of Chicago Press.
- Krumins, J., Palmer, E., Svensson, I., & Vanovska, I. (2001, June 7–9). *Life expectancy trends and calculations of projected life expectancy at retirement – From transition to post-transition – The case of Latvia*. Paper presented at the European Population Conference, Helsinki, Finland.
- Palmer, E. (1999). *Individual decisions and aggregate stability in the NDC system*. Mimeograph.
- Palmer, E. (2000). *The Swedish pension reform: Framework and issue* (World Bank Pension Primer). Washington, DC: The World Bank.
- Palmer, E. (2002). Swedish pension reform – How did it evolve and what does it mean for the future? In M. Feldstein & H. Siebert (Eds.), *Coping with the pension crisis: Where does Europe stand?* Chicago: University of Chicago Press.
- Settergren, O. (2001). The automatic balance mechanism of the Swedish pension system – A non-technical introduction. *Wirtschaftspolitische Blätter*, 4, 339.
- Settergren, O., Holmgren, N., Leander, J., & Mikula, B. (2000). *The automatic balance mechanism in the Swedish pension system* (RFV Analyser 2000:1). Stockholm: The National Social Insurance Board.

¹⁴Another is that the value of pension rights equals the account value at any given time. This makes it easy to move between jobs, occupations and countries without losing rights, eliminates the potential locking in effect of some formulations of DB schemes, for example, DB final-salary schemes.

- Valdés-Prieto, S. (2001). The financial stability of notional account pensions. *Scandinavian Journal of Economics*, 102(3), 395–417.
- World Bank. (1994). *Averting the old age crisis: Policies to protect the old and promote growth*. Washington, DC: World Bank.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 3

Experiences from Forecasting Mortality in Finland



Juha M. Alho

3.1 Modeen and Törnqvist

The first official cohort-component projection of the population of Finland was prepared by Gunnar Modeen (1934a), an actuary with the Central Statistical Office of Finland at the time. Modeen's work had elements of genuine forecasting in that he commented on past trends in fertility, mortality, and migration, and discussed their possible long-term implications (Modeen 1934b). On the other hand, the work was rather schematic in nature. In particular, age-specific mortality was assumed not to change during the projection period, although Modeen was aware of its declining trend since the late nineteenth century (Modeen 1934a, 38). Unable to pinpoint the future rate of decline exactly, Modeen rejected any alternative assumption as speculative (cf., Modeen and Fougstedt 1938).

Analyses of mortality trends in presumably more advanced countries were used as *leading indicators* in the United States by Whelpton et al. (1947), for example. In Finland, Leo Törnqvist (1949) proposed similar methods. In particular, he used Swedish mortality as a target towards which he assumed Finnish mortality to converge. Both series were first transformed via a logistic type transformation. Then, the curves were aligned, and the Finnish curve was prolonged in accordance with the Swedish development.

A problem with Modeen's projection was that it soon became outdated. Fertility started to increase at the time the projection was published, and mortality continued to decline. Modeen's calculations suggested that the Finnish population would never exceed four million, but this mark had already been crossed by 1950. Törnqvist's collected works (Viren et al. 1981) do not mention the error of past forecasts as a motivation for his own early work. Nevertheless, the future statistics professor, a

J. M. Alho (✉)
University of Joensuu, Joensuu, Finland
e-mail: juhala.helsinki.fi

specialist in time-series analysis (among other fields, cf. Nordberg 1999), was well aware that forecasts cannot be made without error. He appears to have been the first to formulate the problem of uncertainty in population forecasting in probabilistic terms in Törnqvist (1949). Later, Törnqvist also conducted what must be one of the earliest assessments of the empirical accuracy of Finnish forecasts (including his own!) in Törnqvist (1967).

In this note, we will outline current developments in Finnish mortality forecasting. In Sect. 3.2, we describe the methods used by official forecasters. These derive mainly from the tradition of early cohort-component forecasters (cf. DeGans 1999). In Sect. 3.3, we discuss how uncertainty can be taken into account using probabilistic models and present-day computing facilities. We conclude in Sect. 3.4 by commenting on some applications for which mortality forecasts are particularly relevant.

3.2 Official Forecasts¹

The arithmetic underlying cohort-component forecasts was understood a hundred years ago (DeGans 1999). Since the method relies on detailed assumptions concerning future age-specific rates, the real key to forecast accuracy lies with those assumptions. One would think that major improvements would have occurred during the past century, judging from the way the assumptions are formulated. Yet, the methods were essentially perfected by Whelpton back in the 1940s.

The two producers of official population forecasts in Finland are Statistics Finland and the Social Insurance Institution of Finland (or KELA, an abbreviation of the Finnish name). Since the forecasters of the two institutions cooperate on an informal basis, the forecasts have many similarities.

Both institutions produce forecasts approximately every 3 years. More frequent updates are made if unexpected developments occur. Both disaggregate the population by sex and single years of age (0, 1, 2, ..., 99, 100+). Currently both organizations forecast until 2050.

KELA produces a national forecast only, whereas Statistics Finland forecasts the population of every one of the 448 municipalities of Finland. In the case of mortality, the country is divided into three relatively homogeneous areas: Northern and Eastern Finland, which have a high level of mortality (due in particular to cardio-vascular diseases among males); the Swedish-speaking coastal areas, with low mortality; the rest of the country, with intermediate mortality. The reason for the low mortality among the Swedish speakers has not been established, but both socio-economic and lifestyle factors apparently play a part (Koskinen and Martelin 1995).

¹The author would like to thank Mr. Matti Saari, Statistics Finland, and Mr. Markku Ryyynänen, KELA, for information on the practice of forecasting. Any misunderstandings are the sole responsibility of the author.

Neither organization uses cause-specific mortality data in the preparation of their assumptions. This is contrast with the U.S. Office of the Actuary, for example (*e.g.*, Wade 1987). However, we have argued elsewhere that cause-specific information cannot be expected to increase forecast accuracy unless one of two conditions are met: either leading indicators can be identified in the preparation of forecasts, or structural changes can be anticipated based on other available information (as in the case of AIDS, for example) (Alho 1991).

Both organizations use trend extrapolation as a basis for their mortality forecast. Starting from a target value for life expectancy at birth, e_0 , Statistics Finland adjusts future age-specific mortality rates so that the implied increase in life expectancy gradually slows down until the target of e_0 is reached. Age-dependent proportional adjustment is used to modify the jump-off rates. In KELA the starting point is a classification of individual ages into aggregates with similar mortality levels. Regression analysis is used in the log-scale to estimate rates of decline that gradually decelerate. The assumption, made by both organizations, that the rate of decline eventually falls off, is far from self-evident. In fact, we have used U.S. data to show that such an assumption has historically made the U.S. mortality forecasts *worse* than simpler trend extrapolations (Alho 1990).

Neither organization formulates their targets on a cohort basis although both occasionally examine cohort trends to see whether there are any irregularities. A current example of such an irregularity was reported by KELA: the female cohorts born in the 1950s appear to have higher mortality than cohorts born earlier, during WWII.

The methods of trend extrapolation used by the organizations blend judgment and empirical analysis. Neither organization has experimented with the method proposed by Lee and Carter (1992). Its performance in regard to ages 65+ in Finland, was investigated in a University of Joensuu *pro gradu* thesis by Eklund (1995), who found that a one-dimensional singular-value decomposition produced a good fit to the data. Because of random variation, however, the resulting forecast was not always an increasing function of age.

In addition to the trend forecast, KELA produces another mortality variant in which it is assumed (as Modeen did) that mortality will remain at the jump-off level. Statistics Finland limits itself to a single variant even though high and low variants have previously been used in national forecasts.

3.3 Predictive Distribution of Mortality

A major contribution by Törnqvist (1949) was that he was apparently the first to maintain that since the future values of a vital rate cannot be totally known, they must be treated as random variables. The actual future values are then “samples” from their distributions. In modern terminology, the uncertainty of the future value is

expressed in terms of a predictive distribution that represents both our best guess and its uncertainty. The distribution is conditioned on all information available at the jump-off time of the forecast (*e.g.*, Gelman et al. 1995, 9).

Törnqvist's contribution may have been ahead of its time. In particular, correct formal treatment of the predictive distribution would have been difficult before the availability of high-speed computing. In recent years, the potential usefulness of a probabilistic approach to uncertainty has been noted on several occasions.² At the University of Joensuu, we have written a computer program, PEP (Program for Error Propagation), which is capable of simulation samples from a wide range of predictive distributions.

The main concept of PEP is that it allows us to describe the uncertainty connected with a forecast at the time it is being made. All sources of uncertainty – age-specific fertility and age and sex-specific mortality and migration – are taken into account and propagated throughout to derive the predictive distribution of the population. In this sense, PEP is merely a stochastic version of the cohort-component bookkeeping system. The usefulness of the results depends on the assumptions underlying the calculations. The user of PEP must specify a point forecast for each of the vital rates for all future years, just as in ordinary cohort-component forecasting. An additional step is required in the form of specifying the uncertainty surrounding the forecast.

Suppose $R(j,t)$ is the mortality rate for age $j = 0, 1, \dots, \omega$ in a future year $t = 1, 2, \dots, T$. PEP assumes that

$$R(j, t) = \exp(\hat{r}(j, t) + X(j, t)),$$

where $\hat{r}(j, t)$ is the point forecast of the log-rate, and $X(j, t)$ is a random error with a mean of $E[X(j, t)] = 0$. The random error can always be written in the form

$$X(j, t) = \varepsilon(1, t) + \dots + \varepsilon(j, t).$$

In PEP, the *error increments* $\varepsilon(j, t)$ are *assumed* to be of the form

$$\varepsilon(j, t) = S(j, t) (\eta_j + \delta(j, t)),$$

where the $S(j, t)$'s are known scale factors that can be chosen to match any sequence of error variances $\text{Var}(X(j, t))$ that increases with t . Fixing j , we can think of the terms η_j as representing errors in forecasted trends. In the case of mortality, the trend corresponds to the rate of decline, for example. Since the terms $\delta(j, t)$ are independent for any fixed j , they represent unpredictable random variation. The relative roles of the two types of uncertainties derive from the assumption $\eta_j \sim N(0, \kappa_j)$, and $\delta(j, t) \sim N$

²Review of Land (1986); “Special Section on Statistical Analysis of Errors in Population Forecasting and Its Implications on Policy,” *Journal of Official Statistics*, September 1997; “Frontiers of Population Forecasting,” *Population and Development Review*, 1998 Supplement; review of U.N. forecasts by the National Research Council (2000).

$(0, 1 - \kappa_j)$, where $0 \leq \kappa_j \leq 1$. The terms η_j are assumed to be independent of the terms $\delta(j,t)$. Finally, the terms η_j can either have a constant correlation across j , or an AR(1) type correlation. The same is true for the $\delta(j,t)$'s, when t is fixed. This *scaled model* for error was introduced in Alho and Spencer (1997).

In Alho (1998) we provide details of the application of PEP to the population of Finland for 1999–2050. The point forecasts for each vital rate were as specified by Statistics Finland. We now present some details on the treatment of uncertainty in the mortality forecast.

Age-specific mortality data in 5-year age-groups 0–4, 5–9, . . . , 75–79, and 80+ were available for the years 1900–1994. After a preliminary analysis, the data were aggregated into the broader age groups 0–4, 5–34, 35–59, 60–79, and 80+ by adding the age-specific rates together. This increased the stability of the trends. The analysis was carried out in terms of the logarithm of the sum (cf. Alho 1998, Figures 5a–e, pp. 19–21). The unusual values produced by the civil war in 1918 and WWII in 1939–1944 were smoothed using values from the previous year. For each of the five broad age groups, we produced baseline forecasts as follows:

- Starting from year $y = 1915$, we used the data for the previous 15 years ($y, y - 1, \dots, y - 15$) to calculate a trend forecast for all future years until 1994.
- A linear trend was estimated from the first and the last observation of the 15-year data period.
- In case the linear trend was positive, it was replaced by a constant value (*i.e.*, slope = 0).

For each $y = 1915, 1916$, we calculated the empirical forecast error for lead times $t = 1, 2, \dots, 50$. For each lead time t , we could then estimate the standard deviation of the error around zero (*i.e.*, assuming that the forecasts are unbiased). This would give us estimates of $\text{Var}(X(j,t))$ directly, from which the scales $S(j,t)$ could be deduced. However, it turned out that especially for younger ages the estimates were somewhat erratic because of the large random (Poisson-type) variation in the counts. Therefore, final estimates were produced by averaging the estimates from the six time series corresponding to the three broad age groups of 35–59, 60–79, and 80+ for males and females. The resulting estimate of the standard deviation of the relative error starts from approximately 0.06 at $t = 1$ and increases in a linear fashion to about 0.6 at $t = 50$. Otherwise expressed, the relative error one might expect for a single age group increases from 6% to roughly 60% in 50 years (cf., Alho 1998, Fig. 6, p. 22). These estimates were used for all ages.

The results were checked by fitting an ARIMA(1,1,0) model to the data series, and similar results were obtained (Alho 1998, Fig. 6, p. 22).

The parameter κ was estimated by the least-squares method. The single value $\kappa = 0.149$ was applied for all ages.

An AR(1) process was used to model the autocorrelation of the error terms η_j and $\delta(j,t)$ across age j . Otherwise expressed, the correlation was assumed to be $\rho^{|i-j|}$ for any two single years of age i and j , where the empirical estimate $\rho = .945$ was used for η_j 's and $\rho = .977$ was used for $\delta(j,t)$'s. Finally, a parameter for contemporaneous

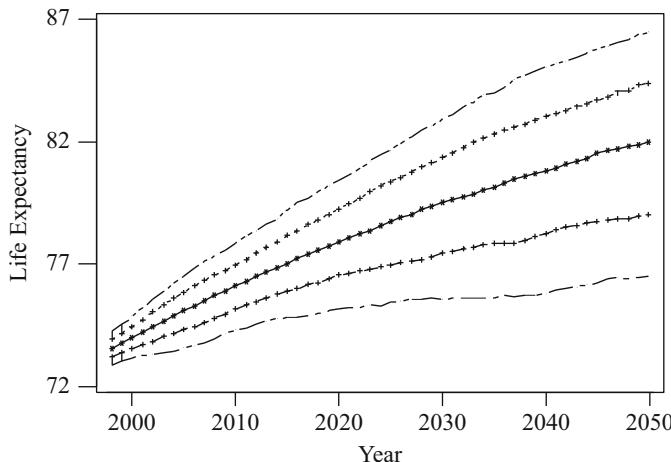


Fig. 3.1 Predictive distribution of male life expectancy in Finland in 1998–2050

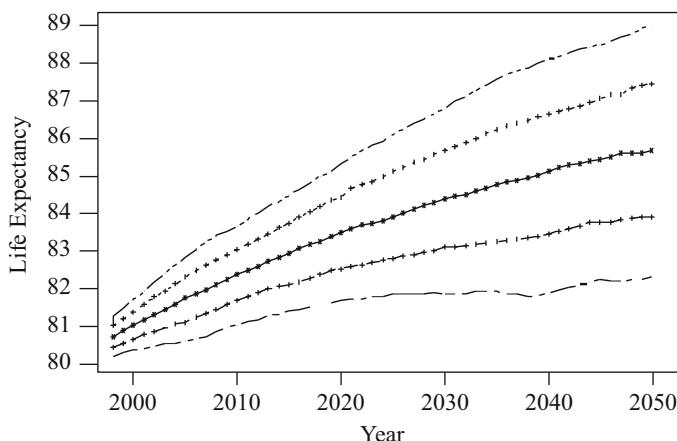


Fig. 3.2 Predictive distribution of female life expectancy in Finland in 1998–2050

crosscorrelation between the error of male mortality and the female mortality was estimated as .795.

The details are fairly complex. One way to assess the reasonableness of the procedures is to consider their implications for life expectancy. Figure 3.1 has a predictive distribution for male life expectancy at birth, and Fig. 3.2 has a plot for female life expectancy. The median of the predictive distribution is 82.0 for males and 85.6 for females in 2050. A 50% prediction interval (or interval between the first and third quartile) is [79.0, 84.4] for males and [83.9, 87.5] for females. An 80% prediction interval for males is [76.5, 86.5] and [82.3, 89.0] for females. The narrower spread for females is probably due to their lower level of mortality.

Two concerns can be raised concerning the intervals. First, the long-term point forecasts are based on an eventual slowdown of the decline in mortality; this may make the Finnish forecast too conservative, as it did in the U.S. earlier. However, we may note that the life expectancy implied by the current Swedish forecasts for males is 82.6 and for females 86.5 years in 2050. In the intermediate variant of the Norwegian forecast, the corresponding ages assumed are 80.0 and 84.5 years. We see that despite the assumption of a slow-down in the mortality decline in Finland, the Finnish forecast is the most optimistic of the three in terms of improvement, since the current life expectancy in Finland is the lowest. Even though the Finnish point forecast may be too low at the end of the forecast period, from this perspective the Finnish forecast appears less conservative.

Second, in view of the vast potential for new medical advances, one could argue that the range of uncertainty expressed by the widths of the intervals might be overly narrow. Two arguments seem relevant here. For the U.S. (both sexes combined), Lee and Carter (1992, p. 660, Fig. 3.1) calculated model-based 95% intervals for life expectancy 50 years ahead. The width of these intervals was approximately 8.4 years. In a normal model, the corresponding width for an 80% interval would be approximately 5.5 years. Thus, our intervals are clearly wider. In a discussion of the paper by Lee and Carter, we noted that by including all sources of variation, the Lee and Carter intervals would have been approximately one half wider (Alho 1992). This would have resulted in estimates close to ours. (One could also argue that in a large country with heterogeneous sub-populations there might be some offsetting variation, resulting in a national average more stable than in a small homogeneous country. While conceivable, this possibility does not seem to be an adequate reason for inflating the Finnish intervals, since it does not show up in the Finnish time series).

A related criticism suggests that future advances in medical knowledge may be so unprecedented that intervals based on past outcomes are too narrow. However, in 1915, at the start of our observation period, both life expectancy for both males and females was substantially lower – 43.2 and 49.2 years (Kannisto and Nieminen 1996); the improvements by the end of the twentieth century were 27.5 and 29.5 years, respectively. Our estimates reflect the variation in this turbulent period of major improvement. The future can be even more volatile, but the advantage of our intervals is that they correspond to actual past variation, rather than to a subjective assessment. Of course, a subjective assessment may be used as a basis for other calculations that would complement our own.

3.4 Applications

Now that we have the analytical capability to produce predictive distributions for future vital rates and future population, it is of some interest to consider how they might be applied. There are two aspects to this.

First, it is critical that we understand how the predictive distribution can be understood. As noted, *e.g.*, in Alho (1998), predictive distributions can be based on (1) formal statistical models, (2) errors of past forecasts can be used to estimate the error of future forecasts, (3) errors of baseline forecasts can be used to estimate future error, and (4) error specification can be purely subjective. Of course, any mixture of the four is also a possibility. The results we have shown rely primarily on (3), although they have elements of (1) and (4), as well. The aim was to provide an empirical assessment of the difficulty of forecasting (or “forecastability”) of the vital processes for different times. As such, the *results for mortality correspond to the uncertainty in mortality forecasting during the twentieth century*. One can reasonably question whether forecasting is now easier, or more difficult, than in the past, but at least we now have a quantitative empirical assessment of how things were before.

Second, the predictive distribution can be used to address numerous social-policy issues that depend on future population and its age structure. For example, in Alho et al. (2001) we review an example in which output from PEP was used in combination with the Finnish overlapping-generations model to devise alternative pension-funding rules, and another example in which output from PEP was used to assess the stability of the current rules for state aid to municipalities. In a University of Joensuu *pro gradu* thesis, Polvinen (2001) used PEP to form a predictive distribution of the so-called generational accounts. All these questions are of fundamental concern for the long-term planning of the social-support systems in Finland. In no case has the effect of uncertain population age-distribution previously been recognized. Other applications have been presented by Lee and Tuljapurkar (1998) for the Social Security system of the United States, for example. Further research opportunities are discussed in Auerbach and Lee (2001).

References

- Alho, J. (1990). Stochastic methods in population forecasting. *International Journal of Forecasting*, 6, 521–530.
- Alho, J. (1991). Effect of aggregation on the estimation of trend in mortality. *Mathematical Population Studies*, 3, 53–67.
- Alho, J. (1992). Comment. *Journal of the American Statistical Association*, 87, 673–675.
- Alho, J. (1998). *A stochastic forecast of the population of Finland* (Reviews 1998/4). Helsinki: Statistics Finland.
- Alho, J., & Spencer, B. D. (1997). The practical specification of expected error of population forecasts. *Journal of Official Statistics*, 13, 203–225.
- Alho, J., Lassila, J., Salo, M., Spencer, B. D., & Valkonen, T. (2001). *Predictive distributions for future population with applications to aging*. Manuscript.
- Auerbach, A. J., & Lee, R. D. (Eds.). (2001). *Demographic change and fiscal policy*. Cambridge: Cambridge University Press.
- DeGans, H. A. (1999). *Population forecasting 1895–1945*. Dordrecht: Kluwer.
- Eklund, K. (1995). *Kuolevuuden mallittaminen ja ennustaminen Suomessa vanhusväestön keskuudessa*. Pro Gradu thesis, University of Joensuu.

- Gelman, A., Carlin, J. B., Rubin, D. B., & Stern, H. S. (1995). *Bayesian data analysis*. London: Chapman-Hall.
- Kannisto, V., & Nieminen, M. (1996). *Revised life tables for Finland. Population 1996:2*. Helsinki: Statistics Finland.
- Koskinen, S., & Martelin, T. (1995). *Why is mortality low among the Swedish speaking minority in Finland?* Paper presented at the 11. Nordic Demographic Symposium, June 1995, in Helsinki, Finland.
- Land, K. C. (1986). Methods for national population forecasts: A review. *Journal of the American Statistical Association*, 81, 888–901.
- Lee, R. D., & Carter, L. (1992). Modeling and forecasting the time series of U.S. mortality. *Journal of the American Statistical Association*, 87, 659–671.
- Lee, R. D., & Tuljapurkar, S. (1998). Uncertain economic futures and social security finances. *American Economic Review*, May, 237–241.
- Modeen, G. (1934a). *Finlands framtida befolkningsutveckling* (Tilastokatsauksia No: 10). Helsinki: Tilastokeskus.
- Modeen, G. (1934b). Suomen väkiluvun tuleva kehitys ja sen taloudelliset seuraukset. *Kansantaloudellinen aikakauskirja*, VI, 351–378.
- Modeen, G., & Fougestadt, G. (1938). *Laskelmia Suomen vastaisesta väestönkehityksestä* (Sarja VI, Väestötilastoa 89). Helsinki: Tilastokeskus.
- National Research Council. (2000). *Beyond six billion: Forecasting the world's population*. Washington, DC: National Academies Press.
- Nordberg, L. (1999). Leo Törnqvist – The “Grandfather” of Finnish Statistics. In J. Alho (Ed.), *Statistics, registries and science. Experiences from Finland*. Helsinki: Statistics Finland.
- Polvinen, A. (2001). *Stokastisen ennusteen käyttö julkisen talouden kehityksen arvioinnissa* (Eläketurvakeskuksen raportteja 2001:23). Helsinki: Eläketurvakeskus.
- Törnqvist, L. (1949). Nämäkohdat, jotka ovat määritteet primääriesten prognosiolettamusten valinnan. In J. Hyppölä, A. Tunkelo, & L. Törnqvist (Eds.), *Suomen väestöä, sen uusiutumista ja tulevaa kehitystä koskevia laskelmia* (Tilastollisia tiedonantoja 38). Helsinki: Tilastokeskus.
- Törnqvist, L. (1967). The post-war development of Finland compared with predictions made after the war. In *World population conference, 1965, Vol. III, Projections, measurement of population trends*. New York: United Nations.
- Viren, M., Alho, K., Teräsvirta, T., Vartia, P., & Vartia, Y. (1981). *Collected scientific papers of Leo Törnqvist* (Series A7). Helsinki: The Research Institute of the Finnish Economy.
- Wade, A. (1987). *Social security area population projections 1987*. Baltimore: U.S. Office of the Actuary.
- Whelpton, P. K., Eldridge, H. T., & Siegel, J. S. (1947). *Forecasts of the population of the United States 1945–1975*. Washington, DC: U.S. Bureau of the Census.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 4

Mortality Projections in Norway



Helge Brunborg

4.1 A Brief Description of the Norwegian Population Projection Model

The official population projections for Norway are produced and published by Statistics Norway. As in Finland, though not in Denmark and Sweden, the national statistical agency makes regional projections as well. The smallest geographical units projected are the 435 municipalities (*kommuner*), which range in size from about 250 (Utsira) to about $\frac{1}{2}$ million (Oslo) people.

Using the projection model BEFREG, the population by age and sex is projected 1 year at a time by the cohort-component method. The method employs a migrant pool approach for migration between the 90 economic regions of Norway (NUTS), with 1–19 municipalities in each region. The regional projection results are subsequently broken down into results for individual municipalities according to the size and historical growth rate of broad population age groups in each municipality. Thus, the cohort component method is generally not applied at the municipal level. The national population figures are found by aggregating the regional projections, *i.e.* the bottom-up principle.

The model has been virtually unchanged during the last 15–20 years; see Rideng et al. (1985) for a general description and Hetland (1998) for a technical description of the computer system.

After this paper was written, a new set of population projections for Norway for the period 2002–2050 was published in December 2002, see <http://www.ssb.no/folkfram/>. One important change compared to the 1999–2050 projections is that the assumed life expectancies up to 2050 are considerably higher than previously. The method for projecting the age-specific death probabilities is the same as that for the 1999–2050 projections.

H. Brunborg (✉)
Statistics Norway, Oslo, Norway
e-mail: helge.brunborg@gmail.com

In the most recent projections, for the period 1999–2050 and with the registered population as of 1 January 1999 as the initial population, we assumed three variants for each of the following demographic components: fertility, mortality, net immigration and the degree of centralisation for internal migration (plus a variant with zero migration). This would yield 144 different population projections; however, only a few of these combinations have been computed and published. The populations of municipalities were projected until 2020, those of the counties were projected until 2050 (but published only for the years up to 2030), and the population for the entire country was computed until 2050.

All data for the projections come from registers through the population-statistics system BESYS, which is used to build up a structure of aggregate data for population, births, deaths, domestic and international migration, where the smallest unit is age*sex*municipality.

As mentioned above, BEFREG projects the population only by age, sex and region. Previously, Statistics Norway has also made national projections by age, sex and marital status (Brunborg et al. 1981; Kravdal 1986) and by age, sex and household status (Keilman and Brunborg 1995). The last two of these publications used mortality rates by formal marital status, whereas the first did not. Moreover, the stochastic microsimulation model MOSART projects a sample (from 1% to 10%) of Norway's population by age, sex, marital status, household status, educational activity and level, labour-market earnings and public-pension status/benefits (including disability and old-age). In this model the mortality probabilities have been estimated from 1993 data with sex, age, marital status, educational attainment and disability status as covariates (Fredriksen 1998). MOSART has a complicated data structure that is infrequently updated.

4.2 A Short History of Mortality Projections in Norway

Since 1969, thirteen sets of regional and national projections have been prepared and published, usually every 3 years.¹ In the first projections the mortality rates were kept constant for the entire projection period and set equal to the most recently observed rates, usually for a period ranging from two to five calendar years for national rates and 10-year periods or more for regional rates and time trends. The use of observations for several years is done to reduce random variations due to Norway's small population (about four million).

Gradually the projections of the age- and sex-specific mortality rates have become more sophisticated²:

¹See Texmon (1992) for a survey of Norwegian population projections for the period 1969–1990.

²For more detail, see the publication for each set of projections, with text in both Norwegian and English, the most recent being Statistics Norway (2002).

- In the projections of 1977 and 1979, the observed rates were reduced by 2% and 3%, respectively, and kept constant for the entire projection period (until 2010).
- Regional mortality rates (specific for each of the 19 counties) were introduced in the projections of 1979. This affects overall mortality if, for example, net migration flows go primarily from high to low-mortality areas.
- A declining trend in future mortality rates was assumed for the first time in the 1982 projection, when all rates were reduced by 1% per year for the first 10 years but kept constant for the rest of the projection period.
- The Brass logit life-table model was used for the projections of 1987 (but not later).
- A model with time- and age-specific death rates was used for the period 1990–2050 (but not later); see Statistics Norway (1991) and Goméz de Leon and Texmon (1992).
- Alternative mortality projections were introduced in the 1993 projections, when sets of rates yielding low, medium and high life expectancies were assumed (Statistics Norway 1994).
- Target life expectancies for the final projection year were also introduced in the 1993 projections (Statistics Norway 1994).
- Age- and sex-specific reductions in future death rates were used in the 1999–2050 projections.

Most of these changes were made because it was discovered that mortality had been persistently overestimated. It is perhaps surprising that it took so long to realise this, since Norway had experienced an almost uninterrupted mortality decline since the 1820s (though with some stagnation for men in the 1950s). The decline has been particularly rapid since the 1960s – see Fig. 4.1.

4.3 Current Methodology of Mortality Projections

In this section, we will describe the methodology and thinking behind the current mortality projections, focusing on a number of separate issues that need to be considered.

4.3.1 Target Life Expectancies

The assumed target life expectancies at birth for the final projection year, 2050, vary from 77 to 83 years for men and from 81.5 to 87.5 years for women (*i.e.* the same as for the previous projections, those of 1996) – see Fig. 4.1. There is no theory or profound thinking behind these assumptions, which have been based on the development in Norway and elsewhere:

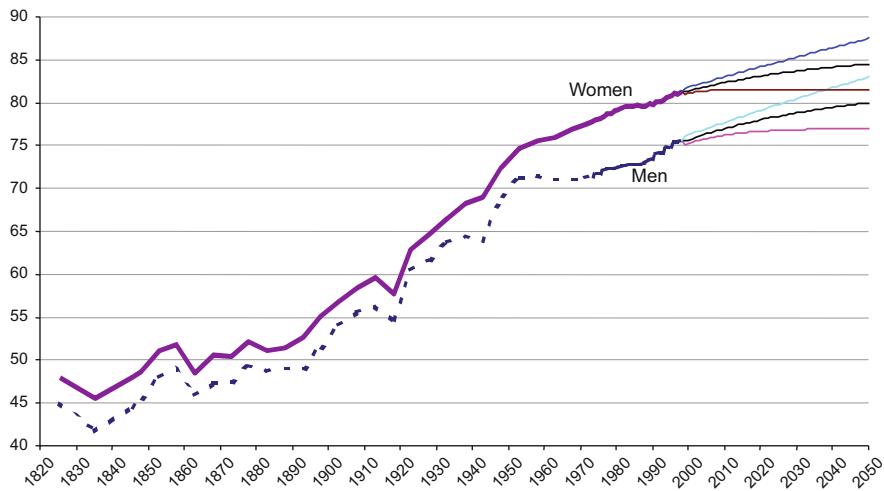


Fig. 4.1 Life expectancy at birth for women and men. Registered 1825–1998 and projected 1999–2050: Low, medium and high assumptions of life expectancy

- The assumed future increase of e_0 –7.5 years for males and 6.3 years for women – in the most optimistic alternative from 1998 to 2050 is of the same magnitude as the *actual* increase during the previous 50-year period from about 1950 to 1998 (5.5 and 7.7 years, respectively).
- In the medium UN series, it is assumed that e_0 for Norway will increase to 80.5 years for men and 86.4 years for women in 2040–2050 (United Nations 1999), *i.e.* well within our range. The long-term UN projections to 2150 assume that e_0 will increase to 85.2 for men and 91.3 for women in Europe and to 87.1 and 92.5 in North America; these constitute the upper limits of the projections.
- The most recent Swedish projections assume life expectancies for 2050 that are slightly lower than in our high alternative, 82.6 for men and 86.5 for women (Statistics Sweden 2000).

We conclude that our assumptions about target life expectancies in 2050 are not extreme and perhaps even slightly conservative. In our next projections, we will consider assuming a wider range of life expectancies. It would be useful to base this range on probabilistic considerations, for example, the work done by Juha Alho and Nico Keilman – see *e.g.*, Keilman et al. 2001.

4.3.2 Difference in Target e_0 for Males and Females

We have assumed a continued gradual decline in the difference between female and male life expectancies, to 4½ years in 2050. This is consistent with historical trends, where the difference increased from about 3½ years around 1950 to almost 7 years in

the 1980s, declining later to 5.7 years in 1998. This decrease, which has been observed in most European countries, is usually explained as due to a narrowing of the gender differences in life style. A similar assumption has been made in Sweden, where the differential is reduced to 3.9 years for 2050.

The most specific manifestation, as well as the principal reason, for this tendency is probably the change in smoking behaviour. The proportion of daily smokers among men 16–76 in Norway has decreased from about $\frac{1}{2}$ in 1973 to $\frac{1}{3}$ in 2000, while it has remained constant at $\frac{1}{3}$ for women. In 2000, smoking had become more common for the first time among women than among men (32% vs. 31%) – see Fig. 4.2.

4.3.3 Life Expectancies in the First Projection Year

If we had assumed a smooth development of e_0 from the most recent observations to the target, there would have been practically no difference between the alternatives for the first years of the projection. We would like, however, to have a reasonably wide range of e_0 to cover the possibility of random year-to-year fluctuations in mortality. For this purpose, we chose, as described in Sect. 5.5, a parameter that yields a change in e_0 for the first projection year (*i.e.* from 1998 to 1999) that is similar to the largest observed annual change in e_0 during the period 1965–1998. In other words, we chose the largest annual *decrease* for the low alternative – about 0.4 years for men 0.2 years for women – and the largest annual *increase* for the high alternative – about 0.6 years for each sex (see Table 4.1). In 1996 a similar approach was taken to widen the mortality range in the first projection year, but then the standard deviation of the difference between observed and *projected* life expectancies was subtracted (or added) to yield the low (or high) values for the first projection year. Actual and assumed life expectancies for the first projection years are shown in Figs. 4.3 and 4.4. It can be seen that the observed values are close to the medium variant in both rounds of projections, with a slightly faster improvement for men than for women.

4.3.4 Path of e_0 from the Initial Until the Target Year

In Sect. 5.5, we describe the method used to project the mortality rates in the 1999–2050 population projections for Norway. The e_0 trajectory is found by interpolating between the life expectancies assumed for the initial and target years. This is done by assuming a dampening of the annual age- and sex specific rates of change in death rates. Figures 4.5 and 4.6 show e_0 for males and females for each of the three mortality alternatives in the two most recently completed projections, *i.e.* for 1996–2050 and 1999–2050.

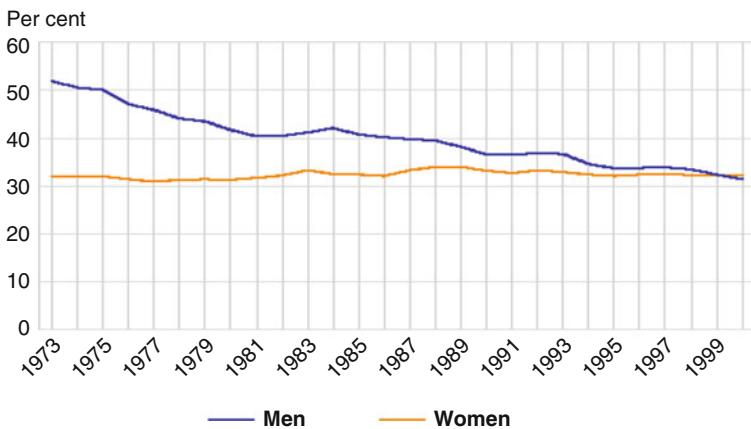


Fig. 4.2 Proportion of daily smokers among men and women 16–74 Years of Age, 1973–2000.
(Source: <http://www.ssb.no/emner/03/01/roy/>)

Table 4.1 Change factors and target life expectancies, population projections 1999–2050^a

Life expectancy variant	First projection year (1999)		Subsequent projection years (2000 ... 2050)	
	Parameter α_s	Assumed life expectancy	Parameter β_s	Target life expectancy
Males				
L (low)	3.315	75.1	0.927	77.0
M (medium)	0.741	75.5	0.982	80.0
H (high)	-3.026	76.1	1.005	83.0
Females				
L (low)	1.417	81.0	0.791	81.5
M (medium)	-0.212	81.2	0.978	84.5
H (high)	-4.195	81.7	1.006	87.5

^aThe observed life expectancy at birth in 1998 was 75.5 years for males and 81.2 years for females

We notice that the 1999 projections generally assume a more linear development of e_0 than the 1996 projections, except for the low alternative. The main reason is that we did not impose the restriction of no mortality change in the target year, as discussed below. It is, however, not yet possible to determine whether the 1999 paths are more realistic than the 1996 paths – only time will tell!

4.3.5 Slope of e_0 in the Target Year

In the previous round of projections, for 1996–2050, it was assumed that life expectancy would cease to decline in 2050 due to high uncertainty about mortality

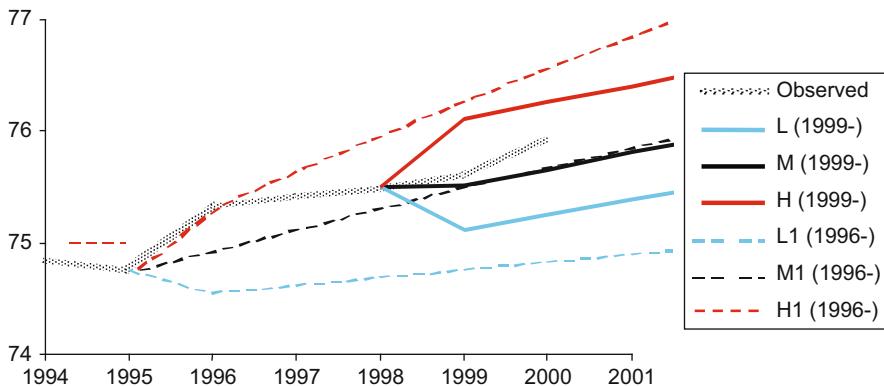


Fig. 4.3 Life expectancy at birth for men, registered 1994–2000 and projected to 2001

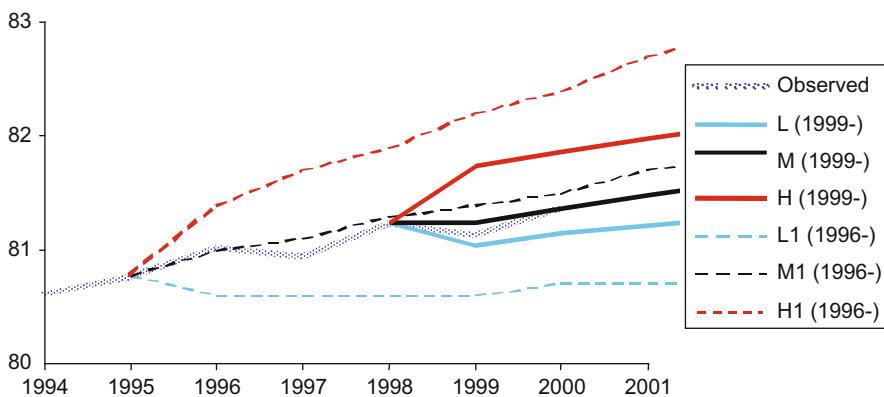


Fig. 4.4 Life expectancy at birth for women, registered 1994–2000 and projected to 2001

in such a distant future. In the most recent projections, however, we have not been concerned about the slope of e_0 in the target year. This is due partly to the high uncertainty, but also to the likelihood – in our opinion – that mortality will continue to decline, even after 2050.

4.3.6 Alternative Mortality Assumptions

As mentioned in the introduction, Statistics Norway did not begin to introduce alternative mortality assumptions before 1993. Although future fertility (and migration) may seem more uncertain than future mortality, there are a number of uncertainties related to the development of mortality. To mention briefly just a few:

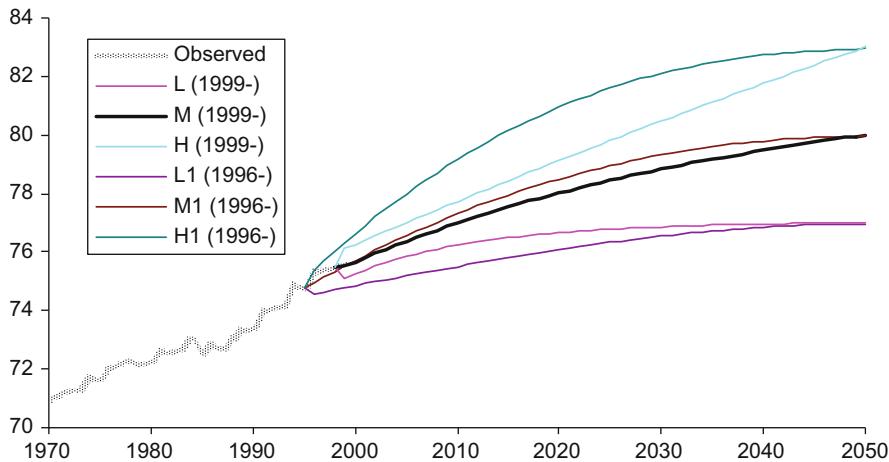


Fig. 4.5 Life expectancy at birth for males, registered 1970–1999 and projected to 2050

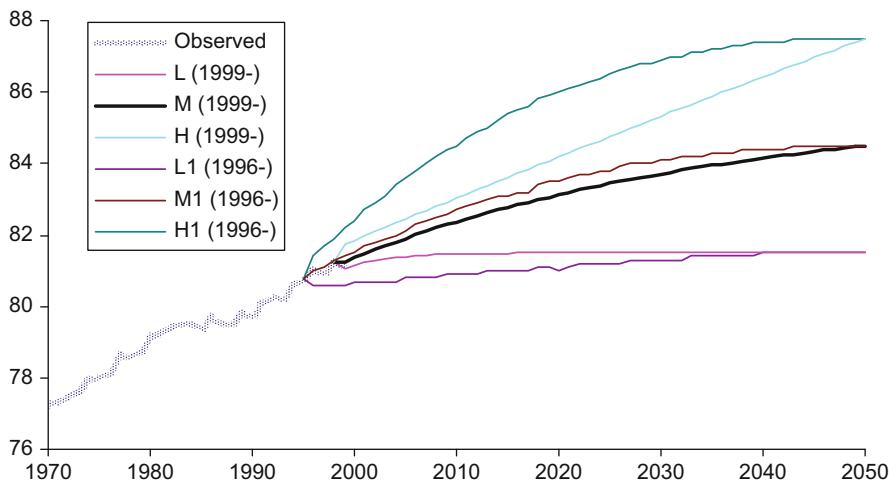


Fig. 4.6 Life expectancy at birth for females, registered 1970–1999 and projected to 2050

technological breakthroughs in the diagnosis and treatment of diseases, new epidemics and other diseases, pollution and other environmental problems, increasingly unhealthy life styles, etc.

Thus, it seems wise to base population projections on several mortality alternatives. The alternatives should cover a realistic future range of variation. If the range is too small, there are in reality no alternatives; if it is too large, the projections may cease to be interesting. Ideally, the range should be based on a probability distribution.

The number of alternatives is another choice that needs to be considered. The number should be greater than one but not so large that it becomes confusing and

complicated to present and use the population projections. Thus, three seems to be a sensible number. In the last three rounds of projections, we have assumed one low, one medium and one high alternative for life expectancy. Note that “low”, “medium” and “high” do not refer to mortality levels as such but to life expectancies, for consistency with the assumptions about other demographic components; for example, “low” implies low population growth.

4.3.7 Age Groups

The current version of BEFREG has 100 age groups, 0,1,2,...,98 and 99+. We are considering including more single-year age groups, at least age 99, in the next version of the model for two reasons: First, there is an increasing number of (and interest in) centenarians. Second, the lack of a decline in mortality for the oldest of the old, as described in Sect. 4.4, may necessitate a more differentiated approach for this particular group.

4.3.8 Cohort Mortality

Cohort mortality is not directly considered in our mortality assumptions. However, we have calculated the implications of our assumptions for cohort mortality – see Fig. 4.7.³ We may note that projected cohort mortality approaches projected period mortality; this tendency may be expected in view of the assumptions made. We also see that the difference between cohort and period mortality is particularly high for the 1960s–1980s.

4.4 Age-Specific Trends in Mortality Rates

The future age pattern of mortality has been given little attention in previous Norwegian population projections. The reason may be that the past mortality decline was implicitly assumed to be about the same for all ages and perhaps also that the age pattern of mortality was not considered to matter so much for the population projections – the only important factor being the level of mortality as measured by life expectancy at birth.

It is clear, however, that the mortality decline has varied greatly by age and sex. It is well known that historically the decline was largest for infants and children. To

³The life expectancy for cohorts born in 1950 and later and having members still living in 2050 has been estimated from extrapolated death probabilities. For example, $q(101, c1950) = q(101, c1949)$.

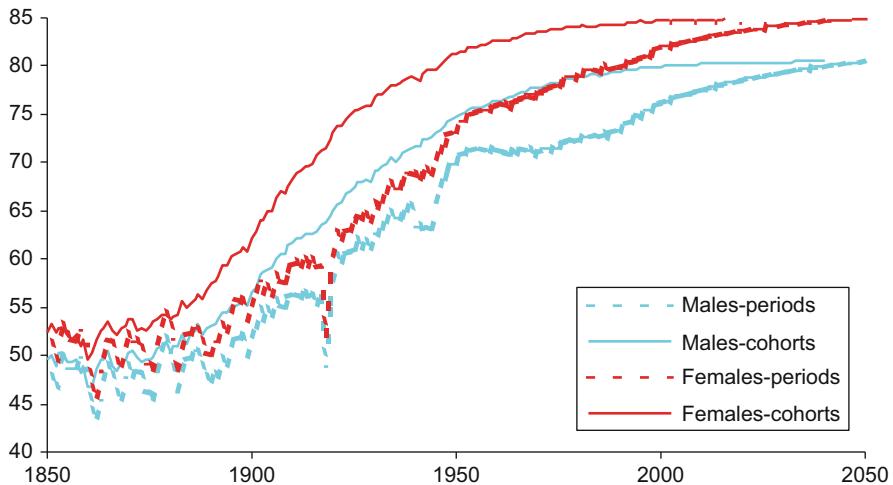


Fig. 4.7 Life expectancy at birth for periods and cohorts, registered and projected

learn more about recent trends, we have studied the development of age-and sex-specific mortality rates for the past 30–40 years.

Let $q(x,t,s)$ be the probability of death at age x , time t and sex s . The annual relative rate of change during the period from t_1 to t_2 is

$$\check{r}(x, s) = \ln(q(x, s, t_2)/q(x, s, t_1))/(t_2 - t_1). \quad (4.1)$$

We estimated the rate of change for $x = 0, 1, \dots, 99$; $t_1 = 1965–1968$, $t_2 = 1998$; and $s = \text{male, female}$.⁴ As these rates of change exhibit a rather ragged appearance we applied a smoothing procedure. The rate of change $r(x,s)$ is graduated using a 21-term weighted moving average, with coefficients from Hoem (1995). The smoothed rates of change for males and females are shown in Fig. 4.8.

As expected, the decline has been the most rapid for children, but it has also been very impressive for adolescents and for men and women between 40 and 80. On the other hand, it has been nearly zero for persons over 90. There is even a tendency of *increasing* mortality for the oldest individuals, especially men over age 95, see Figs. 4.9 and 4.10. Thus, mortality rates for the oldest in Norway may have stagnated or started to increase in recent years. This is contrary to the trends in most other countries, but similar to those in the Netherlands (Nusselder and Mackenbach 2000). There are also indications of stagnating old-age mortality trends in Denmark, especially for men (Danmarks Statistik 2000), and in the USA (Kranczer 1997).

⁴We experimented with different periods, finding similar patterns, and chose the oldest single-year age-specific death rates that were available to us when we performed this analysis, i.e. for 1965.

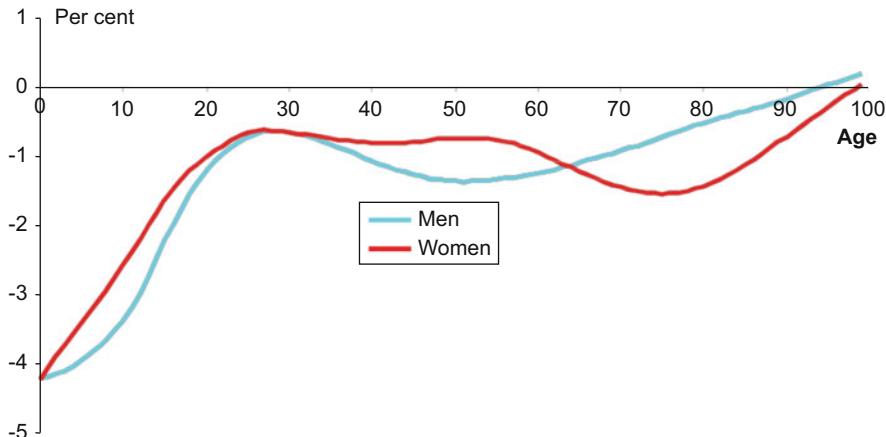


Fig. 4.8 Annual change in age-specific death probabilities, 1965–1998. Per cent. Men and women by age. Graduated by 21-term weighted moving average

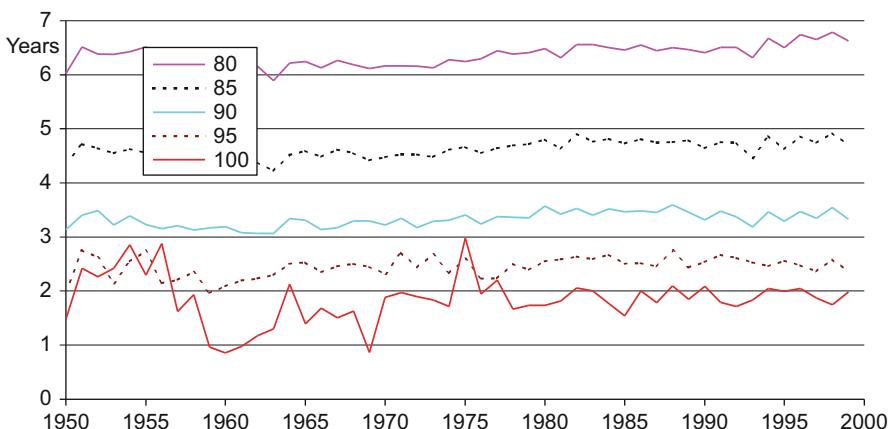


Fig. 4.9 Expected remaining years of life for elderly men at selected ages

4.5 Projections of Age-Specific Mortality Rates

In the most recent projections, we assume that the age-specific mortality declines are consistent with the patterns shown in Fig. 4.8 and as estimated in (1). To project the mortality probabilities $q(x, t, s)$ by age x , time t and sex s , we multiply them by a factor depending on the rates of change $r(x, t, s)$ estimated for the period 1965–1998, as shown in Fig. 4.8. These rates of change are also changed, however, 1 year at the time, through multiplication by the parameters α_s or β_s for each age.

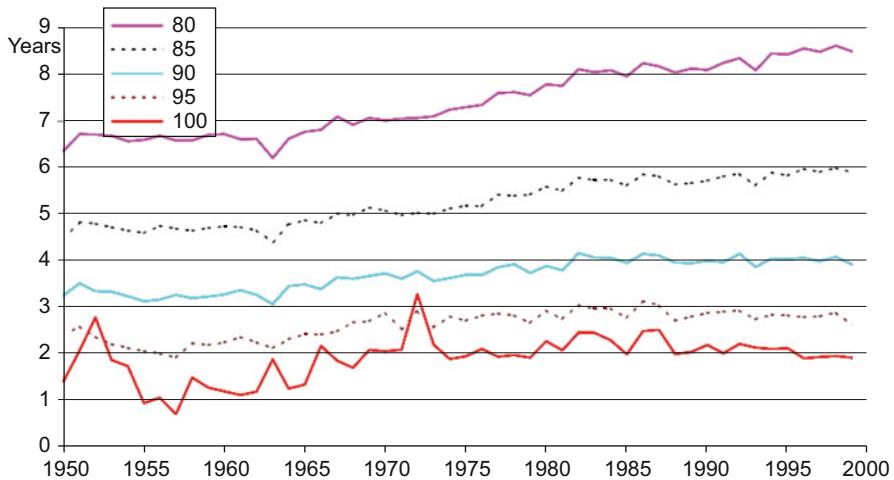


Fig. 4.10 Expected remaining years of life for elderly women at selected ages

Thus, for the first projection year, 1999, we find

$$q(x, s, 1999) = q(x, s, 1997.5) * (1 + r(x, s, t)), \text{ where} \quad (4.2)$$

$$q(x, s, 1997.5) = (q(x, s, 1997) + q(x, s, 1998))/2 \quad (4.3)$$

and 1997.5 denotes the average of 1997 and 1998. For 1999 the change factors are

$$r(x, s, 1999) = \check{r}(x, s) * \alpha_s. \quad (4.4)$$

For the subsequent years we compute

$$q(x, s, t) = q(x, s, t - 1) * (1 + r(x, s, t)) \text{ for } t = 2000, \dots, 2050, \text{ where} \quad (4.5)$$

$$r(x, s, t) = \check{r}(x, s) * \beta_s \text{ for } t = 2000 \text{ and} \quad (4.6)$$

$$r(x, s, t) = r(x, s, t - 1) * \beta_s \text{ for } t = 2001, \dots, 2050. \quad (4.7)$$

Equation (4.7) determines the path of e_0 from the most recently observed values of q_x (the mean of 1997 and 1998) to the assumed target values for 2050 (see Table 4.1). The estimates of parameters α_s and β_s employed here imply a dampening of the annual rates of change of the age- and specific death probabilities, see Table 4.1. The parameters are determined in the following way, for each sex and mortality alternative (low, medium and high):

- The parameter for the first projection year, α_s , is chosen on the following assumptions about the *change* in life expectancy from 1998 to the first projection year (1999), to obtain a wide range of death probabilities rates for the first year:

- It is set equal to the largest *decline* in e_0 observed during 1965–1998 in alternative L
 - It is set equal to zero in alternative M
 - It is set equal to the largest *increase* in e_0 observed during 1965–1998 in alternative H
- The parameter for the rest of the projection years, β_s , is chosen such that the resulting life expectancy in 2050 is equal to our target life expectancy for that year – see Table 4.1. This is done through iteration. (The procedure is made easy by using an Excel spreadsheet function that lets the user set the target whereby the program estimates the value that yields the desired value of the target).

4.6 Projection Results

The two most recent sets of population projections for Norway, *i.e.* those for 1996–2050 and 1999–2050, are very similar with regard to most of the assumptions. The fertility and migration assumptions are almost the same, and the target life expectancies in 2050 are identical although the trajectories are slightly different. As mentioned above, the greatest difference is that we have assumed a different age-pattern of mortality change for the projection period. This difference, and in particular the slower mortality decline for the oldest, result in significantly lower numbers of old people than in the previous projections (Fig. 4.11).⁵ The results are very similar for the low alternatives, where only a very small mortality decline has been assumed. However, the *high* alternative of the 1999 projections yields about the same number of people 90+ as the *medium* alternative of the 1996 projections. For 2050, for example, the projected population 90+ varies from 50,000 to 150,000 in the previous projection to 50,000–100,000 in the most recent projections.

We conclude that the specification of the age-structure of mortality decline has significant effects on the projected number of old people. Thus, the analysis of mortality trends for the oldest members of the population is important for population projections.

⁵The three different projections from 1999 shown here differ only with regard to the mortality assumptions. The 1996-projections are also different with regard to fertility and migration, however, but for the age groups and period considered here, the effects of these differences are marginal. Fertility differences for 1996–2050 will obviously not affect the number of persons 90+ for this period. The youngest age in 1996 of persons to enter age group 90+ in the projection period 1996–2050 is 36 years. Since persons aged over 35 accounted for only 12% of net immigration in 1996–1999, differences in net immigration can only have marginal effects on the number of persons 90+ projected for the period 1996–2050.

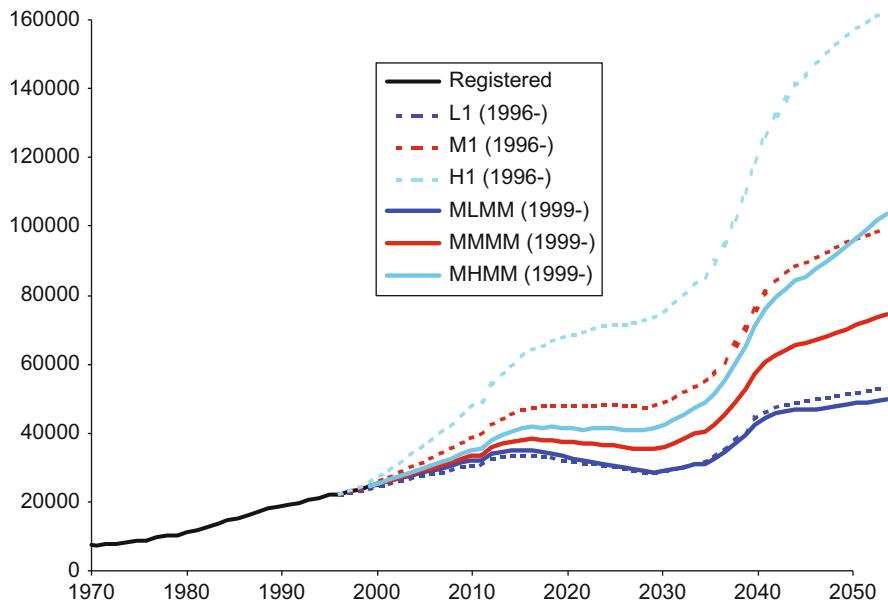


Fig. 4.11 Number of persons 90+. Registered 1970–2001 and projected 1996–2050 and 1999–2050

References

- Brunborg, H., Mønnesland, J., & Selmer, R. (1981). *Projections of the population by marital status 1979–2025* (Rapporter 81/12). Oslo/Kongsvinger: Statistics Norway.
- Danmarks Statistik. (2000). *Befolningens bevegelser 1999*. Copenhagen: Statistics Denmark.
- Fredriksen, D. (1998). *Projections of population, education, labour supply and public pension benefits: Analyses with the dynamic microsimulation model MOSART* (Social and Economic Studies 101). Oslo/Kongsvinger: Statistics Norway.
- Goméz, D. L. J., & Texmon, I. (1992). Methods of mortality projections and forecasts. In N. Keilman & H. Cruijsen (Eds.), *National Population forecasting in industrialized countries*. Amsterdam/Lisse: Swets & Zeitlinger.
- Hetland, A. (1998). *Systemdokumentasjon for BEFREG* (Interne dokumenter 98/4). Oslo/Kongsvinger: Statistics Norway.
- Hoem, J. M. (1995). Moving-average graduation of the timing of the historical demographic transition in Sweden. In C. Lundh (Ed.), *Demography, economy and welfare* (Scandinavian Population Studies 10). Lund: Lund University Press.
- Keilman, N., & Brunborg, H. (1995). *Household projections for Norway, 1990–20 Part I: Macrosimulations* (Report 95/21). Oslo/Kongsvinger: Statistics Norway.
- Keilman, N., Pham, D. Q., & Hetland, A. (2001). *Norway's uncertain demographic future* (Social and Economic Studies 105). Oslo: Statistics Norway. <http://www.ssb.no/emner/02/03/sos105/>.
- Kranczer, S. (1997). Record-high U.S. life expectancy. *Statistical Bulletin*, 78, 2–8.
- Kravdal, Ø. (1986). *Framskrivning av folkemengden etter kjønn, alder og ekteskapelig status 1985–2050* (Rapporter 86/22). Oslo/Kongsvinger: Statistics Norway.
- Nusselder, W. J., & Mackenbach, J. P. (2000). Lack of improvement of life expectancy. *The Netherlands. International Journal of Epidemiology*, 29, 140–148.

- Rideng, A., Sørensen, K. Ø., & Sørlie, K. (1985). *Modell for regionale befolkningsframkrivinger* (Rapporter 85/7). Oslo/Kongsvinger: Statistics Norway.
- Statistics Norway. (1991). *Population projections 1990–2050. National and regional figures* (Official Statistics of Norway, NOS B983). Oslo/Kongsvinger: Statistics Norway.
- Statistics Norway. (1994). *Population projections 1993–2050. National and regional figures* (Official Statistics of Norway, NOS C176). Oslo/Kongsvinger: Statistics Norway.
- Statistics Norway. (2002). *Population projections 1999–2050. National and regional figures* (Official Statistics of Norway, NOS C693). Oslo/Kongsvinger: Statistics Norway.
- Statistics Sweden. (2000). *Sveriges framtidia befolkning. Befolknings-framskriving för åren 2000–2050* (Demografiska Rapporter 2000:1). Stockholm: Statistics Sweden.
- Texmon, I. (1992). Norske befolkningsframskrivinger 1969–1990. In O. Ljones, B. Moen, & L. Østby (Eds.), *Mennesker og modeller. Livsløp og kryssløp* (Sosiale og økonomiske studier nr. 78). Oslo/Kongsvinger: Statistics Norway.
- United Nations. (1999). *World population prospects* (The 1998 revision. Volume I: Comprehensive Tables, ST/SER.A/177). New York: United Nations.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 5

Mortality Assumptions for Sweden.

The 2000–2050 Population Projection



Hans Lundström

5.1 Mortality Projection in Sweden

A new population projection for Sweden is prepared every third year with minor updates in between. Statistics Sweden is responsible for the projections.

Ideally, mortality projections should be made in a process-oriented manner. In practice, however, they are mainly based on trend extrapolations of period mortality rates for the last 50 years. The critical component in the assumptions concerns the ages of 50 and above. The problem is to handle both short-term and long-term developments, especially when there is a trend shift like the one today with a temporary slow-down in the mortality decline for middle-aged women. Does this change mark a new trend or not? Most users are mainly interested in the short-term development, *i.e.* in the next year or the next 5 years. Others have a time horizon of 50 years or longer. There are three alternative assumptions: no change in mortality, mortality decline according to the main alternative, and a more pronounced mortality decline.

The assumption of the future mortality decline is based not only on the observed trend but adjusted for other information such as smoking behavior.

For the coming projection we plan to base the analysis more on cohort mortality.

Mortality in Sweden has fallen ever since the mid-nineteenth century. In the beginning, the change was mainly due to a reduced risk of dying of infectious diseases and deficiency diseases. The factors underlying the greater chances of

To a large extent, this is a translation from Swedish of Chapter 3.3 in the publication Sveriges framtida befolkning. Befolkningsframskrivning för åren 2000–2050. Demografiska rapporter 2000:1 SCB. (Sweden's future population. Population forecast for the years 2000–2050. Demographic reports 2000:1, Statistics Sweden). ISBN 91-618-1068-1.

H. Lundström (✉)
Statistics Sweden, Stockholm, Sweden

survival were economic, social and sanitary improvements, and, not least, medical advances such as the introduction of vaccines and antibiotics.

More recent developments – say, since 1950 – have brought a continued decline in mortality. Changes in the last few decades have largely concerned chronic illnesses, including cardiovascular diseases and cancer, which are the major causes of death. The reasons for the changes in what are sometimes called the diseases of wealthy societies are a transition to a healthier lifestyle and improved medical care, leading to a considerable increase in survival at advanced ages. The decline in mortality at ages above 65 began considerably earlier for women than for men.

5.2 Sharply Lower Mortality in 1950–1999

In the second half of the twentieth century, mortality fell sharply. The risk of death has been reduced by more than half among men below the age of 50 and women below the age of 80. The drop in infant mortality has been particularly dramatic. In 1950, 21 of 1000 children born died before their first birthday. In 1999, just three per thousand live-born children died in their first year.

Among adult men, however, the risk of death changed relatively little in the period 1950–1975. For part of this period, there was even an observable increase in mortality, largely among middle-aged men. Since 1975, however, male mortality has declined substantially, by an average of about 2% per year among men between young middle age and the age of about 80, (see Fig. 5.1).

Among women, we see a decline in mortality throughout the post-1950 period. The annual rate of mortality reduction has averaged 1–2%. At ages in excess of

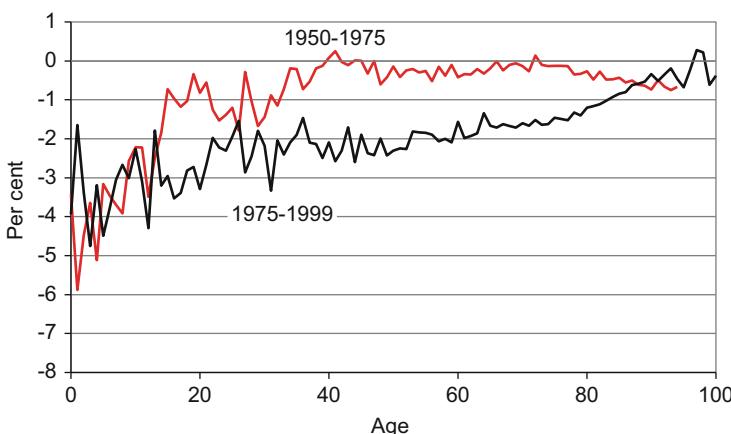


Fig. 5.1 Annual reduction of risk of death at different ages for two periods, 1950–1999. Men

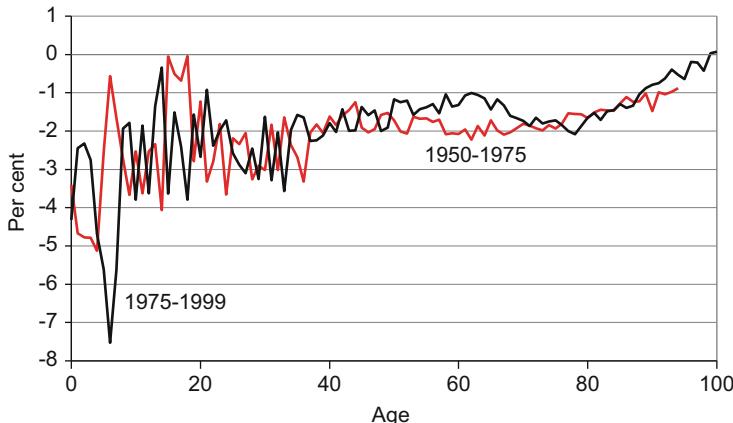


Fig. 5.2 Annual reduction of risk of death at different ages for two periods, 1950–1999. Women

85, the reduction of mortality has been somewhat lower. However, over the last 10 years the rate of mortality reduction among uppermiddle-aged women has slowed slightly, (see Fig. 5.2).

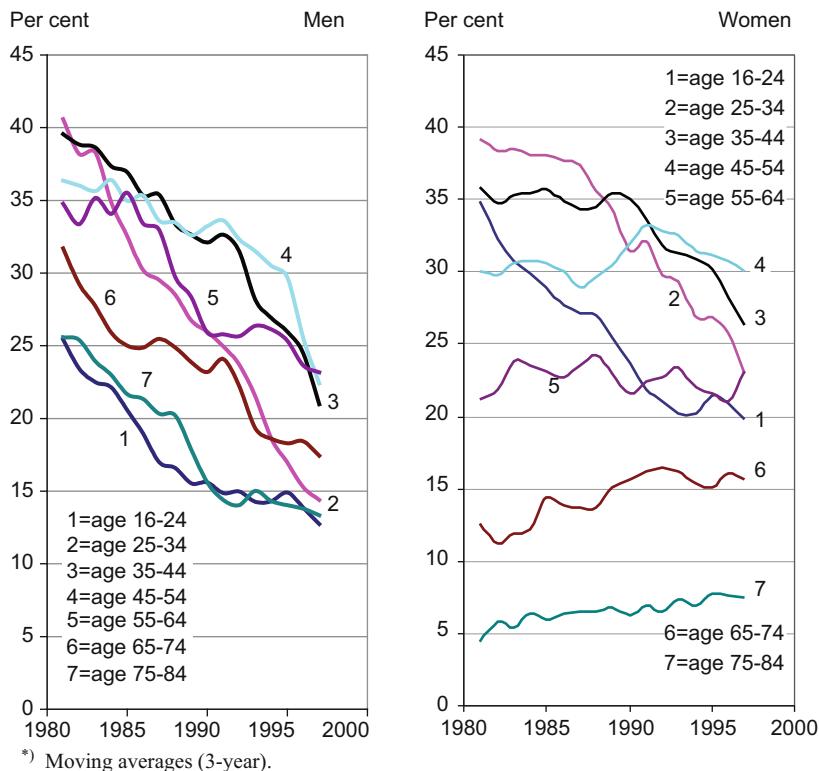
We may add that in recent years severe influenza epidemics have occurred more frequently. This has had a serious effect on many elderly people. The long-term rise in women's annual average life expectancy has therefore slowed temporarily after each influenza epidemic. One reason why women's average life expectancy is affected most is that a larger proportion of elderly women than men reach advanced ages.

5.3 Reasons for the Decline in Mortality in 1980–1999

Assumptions about future mortality trends are based on changes in mortality over the past two decades. During this period, the decline in male mortality has strongly resembled the decline in female mortality. In spite of this close correspondence, around 1980 the situation differed for men and women aged over 50 – and these are critical ages in the assumptions used in the forecast.¹

The rising mortality observable for men until the end of the 1970s was caused by an increase in deaths due to cardiovascular diseases and cancer. Changes in lifestyles helped to break this trend in the 1980s, when mortality began to fall. The proportion of people smoking every day has declined since the end of the 1970s, a tendency with a significant impact on the trends for cardiovascular diseases and cancer. Less

¹Except in the very long term, assumptions on relative changes in younger age groups have a limited impact on the numerical strength of the population, since mortality is so low at these ages.



*) Moving averages (3-year).

Fig. 5.3 Proportion of daily smokers by sex, 1981–1997. (Moving averages (3 year). Source: living conditions survey. (ULF, Statistics Sweden))

fatty food and increased exercise have probably been other important factors behind the decline in cardiovascular diseases. Alcohol consumption has also fallen during this period.

The evident rise in mortality for men until the end of the 1970s was not observable among women. However, a slight tendency towards higher mortality due to ischemic diseases has been discernible, along with a relatively sharp rise in mortality from lung cancer. Nevertheless, total mortality fell, though at a lower rate. The reasons for the reduction in female mortality over the last 20 years have probably been about the same as those affecting male mortality, but with one major exception: the proportion of smokers has continued to increase gradually at more advanced ages. This is due to a generational change involving smokers. The proportion of women who have smoked at some time increases in each successive generation over the age of about 30, (see Fig. 5.3). In spite of this, mortality from cardiovascular diseases has diminished, whereas mortality due to lung cancer, which is linked more closely to smoking habits, has continued to rise.

The medical treatment of cardiovascular diseases in particular has improved, and this has had a significant impact on the decline in the risk of death. A simple indicator of the changes is that mortality due to cardiovascular diseases (heart attacks) has fallen considerably more rapidly than the risk of falling ill (incidence).

5.4 Higher Average Life Expectancy in 1950–1999

Mortality trends since 1950 have resulted in an increase in the average life expectancy of men from 69 to 77 years, an average increase of 0.16 years per calendar year. For women average life expectancy went up from 72 to 82 years, an average increase of 0.20 years per calendar year.

Table 5.1 below shows changes in life expectancy between different periods in the second half of the twentieth century. Table 5.1 also reveals the positive impact on average life expectancy (at birth) of the improvements in mortality at different ages. The gains for men in recent years consist mainly of mortality improvements among

Table 5.1 Change in average life expectancy at birth between different periods (in years)

Sex period	Change (in years)			
	Total	0–19	20–64	65+
Men				
1951–55 to 1956–60	0.7	0.3	0.3	0.1
1956–60 to 1961–65	0.4	0.3	0.1	0.0
1961–65 to 1966–70	0.3	0.3	-0.1	0.1
1966–70 to 1971–75	0.2	0.3	-0.1	0.1
1971–75 to 1976–80	0.4	0.3	-0.0	0.1
1976–80 to 1981–85	1.1	0.2	0.5	0.4
1981–85 to 1986–90	0.8	0.0	0.4	0.4
1986–90 to 1991–95	1.2	0.2	0.6	0.5
1991–95 to 1998 ^a	1.3	0.1	0.6	0.5
Women				
1951–55 to 1956–60	1.2	0.3	0.6	0.4
1956–60 to 1961–65	1.0	0.2	0.3	0.5
1961–65 to 1966–70	0.9	0.3	0.2	0.5
1966–70 to 1971–75	1.1	0.2	0.2	0.7
1971–75 to 1976–80	0.9	0.2	0.1	0.5
1976–80 to 1981–85	1.0	0.1	0.3	0.6
1981–85 to 1986–90	0.7	0.1	0.1	0.5
1986–90 to 1991–95	0.8	0.1	0.2	0.5
1991–95 to 1998 ^a	1.0	0.1	0.3	0.6

Total change and distribution by different age intervals

^a1998 is the last point in time (the distance from the middle year of the previous period is 5 years here, as in the other cases)

young and middle-aged men, but improvements among the elderly have also made a major contribution. For women, the trend among the elderly accounts for most of the increase in average life expectancy over the same period.

5.5 Future Mortality

As stated above, improved living conditions in a range of areas are significant factors underlying the decline in mortality in recent decades. Given present trends, there is reason to hope for continued improvements in living conditions and lifestyles. We know that fewer and fewer young people are taking up smoking and increasing numbers are exercising regularly in their spare time, factors that are important to health and life expectancy. It is worth noting that even if no major improvements were to occur in future, the long-term (longitudinal) impact on mortality at a given age would be similar to that observed to date (perhaps for several decades). In certain cohorts, people could enjoy a favourable life expectancy throughout their entire lives, assuming the levels attained in the 1980s and 1990s are sustained (for factors like consumption, exercise, and men's smoking habits).

Nevertheless, there are lifestyle factors that give rise to concern. Even if smoking is now becoming less common among young people, there is a considerable difference between the smoking habits of elderly and middle-aged women, (see Fig. 5.3). At present, relatively few elderly women are smokers or former smokers. The number will increase during the forecast period, as those who are middle-aged grow old, and this may put a brake on the decline in mortality. As a result, we have assumed that the long-term decline in mortality will be less marked for women than for men. The increasing proportion of people who are overweight, greater stress in professional life, and a possible rise in alcohol consumption in the future are some examples of trends that could slow the decline in mortality. Better information on health matters and improvements in workplace organisation, in the broad sense of this term, may moderate such effects.

Medical progress has had a positive impact on mortality trends. In all probability, the positive trend observed until now in the medical area will continue, and these medical advances may help to improve quality of life and increase life expectancy. The possible impact of potential breakthroughs in genetic engineering and biotechnology surpasses our present comprehension. However, as serious illnesses become more curable, a higher proportion of elderly people will have previously had such illnesses. Despite successful treatment at the time, this factor may have a negative impact on mortality among the very oldest.

Thus, numerous trends may potentially have a – positive or negative – impact on mortality. However, it is hardly possible to quantify the effect of these factors with any precision. We should bear in mind that until now, mortality has changed slowly. Accordingly, we assume that in the immediate future, mortality will continue to follow the trend prevailing up to this point. In the longer term, we assume that the

reduction in the risk of death will continue throughout the forecast period² but will be slowed somewhat by the negative risk factors indicated above. It is far from clear when this slowdown in the reduction of mortality will set in and how significant it will be. Our assumptions have been guided in part by the growing uncertainty of assessments as the time elapsed increases. Here we have assumed a reduced decline in mortality for women from 2010 onwards and for men from 2015 onwards. The difference between men and women is due in part to our assumption that longitudinal effects will cease to be felt sooner among women than among men, since the decline in mortality started earlier among women than among men. We also put a brake on the decline towards the end of the forecast period. The reason is that the overall picture of causes of death may change by then. We should bear in mind that most of the extrapolated reduction in mortality is connected with cardiovascular diseases. In 30–40 years time, this cause-of-death category may well be considerably reduced, even at relatively advanced ages. The other causes of death, which are declining more slowly, will thus acquire greater significance in relative terms and will then automatically entail a slower decline in total mortality.

5.6 Assumptions Used in the Forecast for the Immediate Future

We have based our assumptions regarding mortality in the immediate future on observed risks of death during the period 1995–1999, extrapolated until 2000, (see Fig. 5.4).

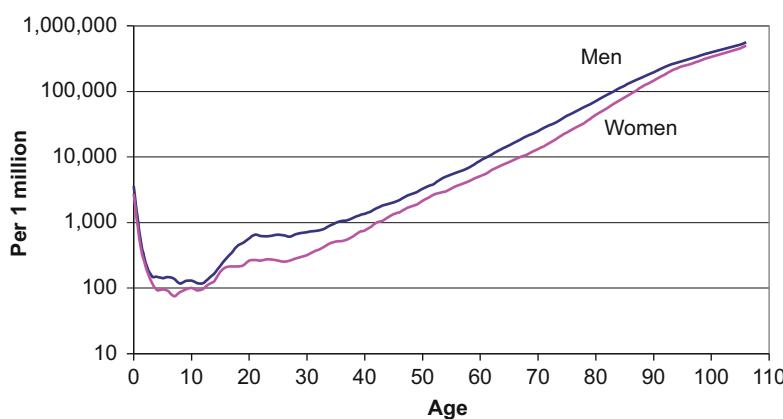


Fig. 5.4 Risks of death in 2000 by age and sex. Per million

²It may be noted that this assumption is very far-reaching in statistical terms, since in effect we are extrapolating 50 years forward in time from a 20 year trend (at least for men).

We assume that risk of death will subsequently be reduced according to the pattern shown in the figures below. Among men, we assume that the risk of death will decline by 1.5% per year at ages below 45, at a somewhat faster rate between 50 and 75, and at a gradually declining rate at more advanced ages. These reductions in the risk of death largely correspond to the trend observable in the 1990s among middle-aged and older men. We assume that this reduction in risk of death will continue unchanged until 2015.

For women, the risk of death has diminished over time in about the same way as for men. For the period until 2010, we have assumed an annual reduction in the risk of death of 1.5% up to the age of 80, in accordance with the trends observed in the 1990s.

It should be noted that the change in the rate of reduction in mortality at different dates proceeds in stages (linear progressive reduction). The transition to a new rate of reduction occurs over a 4-year period (for men in 2015–2018 and for women in 2010–2013).

5.7 Assumptions Used in the Forecast for the Longer Term

Among men, we assume that the annual rate of reduction during the period 2018–2039 will be 75% of its original level. After this, the rate of reduction will gradually decline over a 4-year period until it reaches 50% of the original level (due to the change in the overall composition of causes of death, see Fig. 5.5).

Fig. 5.5 Annual reduction of risk of death among men, by age. Per cent

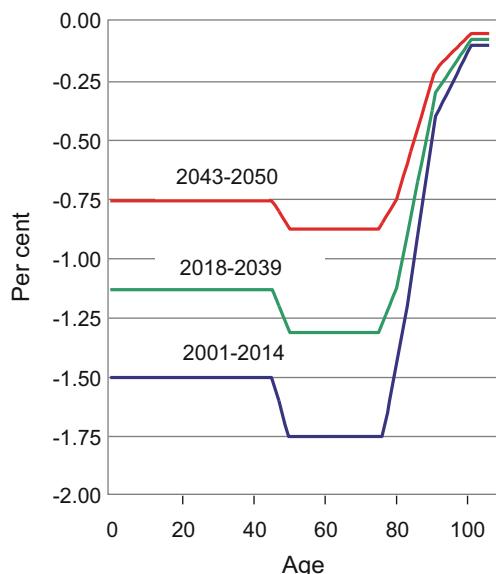
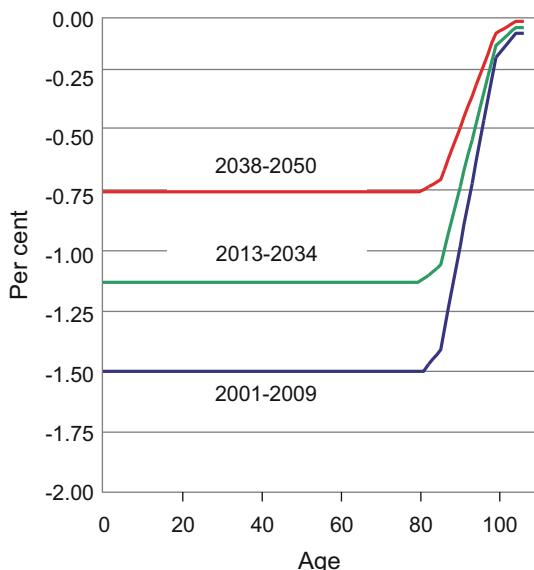


Fig. 5.6 Annual reduction of risk of death among women, by age. Per cent



Among women, we assume that the risk of death will be reduced at a slightly slower pace beginning in 2010. We set the rate of reduction at 75% of its original level over the period 2013–2034 and at half its original level in 2038–2050, (see Fig. 5.6).

Behind these assumptions, there is substantial uncertainty regarding the speed at which the chances of survival are capable of changing over so long a period of extrapolation. However, the future may bring both a more rapid slow-down in the decline in mortality and new medical advances resulting in sharply lower risks of death.

5.8 Mortality Trends over the Period 1950–2050

Figure 5.7 summarizes mortality trends between 1950 and 2050. A logarithmic scale has been used, thus making it possible to compare the mortality trends for different ages. The fact that the curves have the same slope shows that the percentage change in the risk of death has been the same.

5.9 Higher Average Life Expectancy

According to our estimates, average life expectancy for men will rise from 77.1 in 2000 to 82.6 in 2050, while the corresponding figures for women are 82.1 and 86.5. As shown in Table 5.2, we are forecasting a slower increase in average life

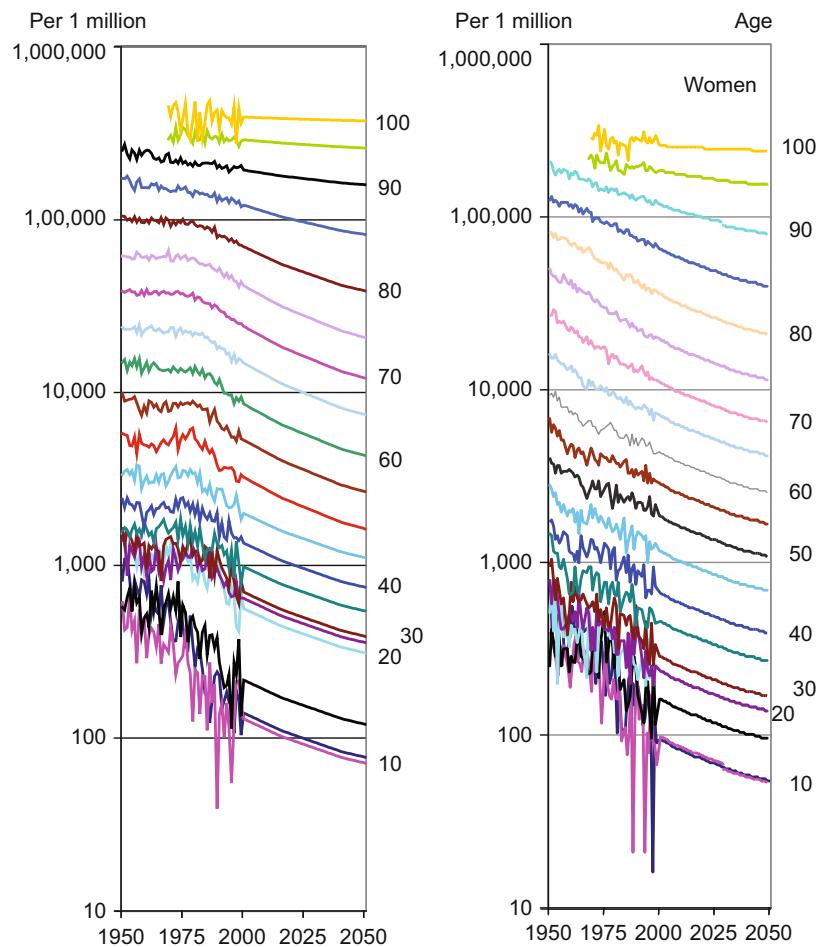


Fig. 5.7 Mortality trends by age 1950–1999 and assumed mortality 2000–2050. Men and women. Per 1 million

Table 5.2 Average life expectancy at birth and at age 65, 1951–2050

Year	At birth		At age 65	
	Men	Women	Men	Women
1951–60	70.9	74.1	13.9	15.0
1961–70	71.7	76.1	13.9	16.1
1971–80	72.3	78.1	14.1	17.5
1981–90	74.0	79.9	14.9	18.7
1991–95	75.6	81.0	15.7	19.4
2000	77.1	82.1	16.5	20.1
2010	78.7	83.4	17.6	21.1
2020	80.0	84.4	18.4	21.8
2030	81.0	85.2	19.1	22.5
2040	82.0	86.0	19.8	23.0
2050	82.6	86.5	20.3	23.5

expectancy over the coming 50-year period than we have observed over the past 50 years. We estimate that average life expectancy at 65 will rise by 3.8 years for men and 3.4 years for women over the next 50 years.

5.10 Assumptions Regarding Mortality Trends in Some Countries

For the sake of comparison with assumptions regarding future mortality trends in other countries, we have provided average life expectancies according to population forecasts for a number of countries in Europe and for the USA and Japan, (see Tables 5.3 and 5.4).

There is wide variation between the different countries. In France and Belgium, it is assumed that average life expectancy for men will increase by nearly 8 years over

Table 5.3 Average life expectancy for men, 2000–2050

Country	Average life expectancy at birth, in years					
	2000	2010	2020	2030	2040	2050
Sweden	77.1	78.7	80.0	81.0	82.0	82.6
France	74.6	76.4	78.0	79.5	80.9	82.2
Belgium	74.4	75.7	77.1	78.7	80.3	82.1
Austria	75.1	76.9	78.5	80.0	81.0	82.0
Switzerland	76.1	77.5	78.8	79.8	80.6	81.3
Finland	73.8	75.6	77.3	78.7	80.0	81.2
USA	74.2	75.5	76.8	78.3	79.8	81.2
Netherlands	75.3	76.6	77.8	78.8	79.5	80.0
Norway	75.8	77.3	78.5	79.4	79.8	80.0
Japan	77.4	78.1	78.6	79	79.2	79.4
UK	75.1	76.6	77.6	78.2	78.7	78.9
Italy	75.9	77.1	78.3	78.3	78.3	78.3
Denmark	73.4	73.5	73.5	73.5	73.5	73.5
Ireland	73.7	75.2	76.4			
Spain	74.1	75.3	76.0			
Germany	73.7	74.8	75.7	76.7	76.9	
Iceland	77.3	77.5	77.5	77.5		

Forecasts in different countries

Source: USA, US Bureau of Census; Japan, Ministry of Health and Welfare; Sweden, Forecast 2000–2050. Other countries: Eurostat, June 2000

Table 5.4 Average life expectancy for women, 2000–2050

Country	Average life expectancy at birth, in years					
	2000	2010	2020	2030	2040	2050
France	83.0	84.8	86.5	87.9	89.2	90.4
Belgium	81.1	82.3	83.6	85.0	86.5	88.1
Austria	81.3	82.8	84.2	85.5	86.2	87.0
Switzerland	83.0	84.4	85.6	86.0	86.4	86.9
USA	79.9	81.4	82.9	84.2	85.4	86.6
Sweden	82.1	83.4	84.4	85.3	86.0	86.5
Japan	84.1	85.1	85.6	86.0	86.3	86.5
Finland	80.9	82.0	83.1	84.0	84.8	85.5
Italy	82.3	83.5	84.7	84.7	84.7	84.7
Norway	81.5	82.7	83.5	84.1	84.4	84.5
UK	80.1	81.5	82.6	83.2	83.6	83.8
Netherlands	80.6	81.1	81.6	82.2	82.7	83.0
Denmark	78.5	78.6	78.6	78.6	78.6	78.6
Spain	81.9	83.0	83.7			
Iceland	81.9	82.1	82.1	82.1		
Ireland	79.4	81.0	82.4			
Germany	80.1	81.1	81.9	82.6	82.9	

Forecasts in different countries

Source: USA, US Bureau of Census; Japan, Ministry of Health and Welfare; Sweden, Forecast 2000–2050. Other countries: Eurostat, June 2000

the next 50 years, while in Japan it is expected to rise by just 2 years. In Sweden, the predicted increase is 5.5 years.

For women, too, there is a considerable range in assumed future mortality. France, Belgium and the USA are predicting that average life expectancy will go up by 7 years, whereas Japan and the Netherlands are anticipating a gain of barely more than 2 years. In Sweden, the rise is expected to be 4.4 years.

5.11 Alternative Assumptions

The purpose of alternative assumptions is to attempt to capture some of the uncertainty in the principal assumption that we have already presented.

Under an alternative assumption with lower mortality, the declining trend in mortality decline during the 1990s will continue uninterrupted throughout the

forecast period until 2050. We assume continuous improvements in lifestyle throughout the period. Moreover, further improvement in medical care and treatment is required (over and above the improvement in the principal assumption), particularly with regard to diseases other than cardiovascular diseases.

In an alternative with higher mortality, we assume no changes in mortality at all in future. Positive and negative lifestyle factors offset each other. This alternative provides a base level for the impact on the population of assumptions regarding mortality; *i.e.*, it functions as a form of sensitivity analysis.

In the first alternative, life expectancy rises from 77.1 in 2000 to 86.1 in 2050 for men and from 82.1 to 89.0 for women. In the second alternative, the figures remain at their initial level throughout the period.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 6

Forecasting Life Expectancy: The SCOPE Approach



James W. Vaupel

This note outlines a method for forecasting life expectancy. The method is based on the idea of structured conditional probabilistic estimation; it “scopes” out the range of possibilities the future may hold. I first described this SCOPE method at a workshop organized by Juha Alho several years ago in Finland.¹ It is a kind of scenario method – with probabilities attached to scenarios, with scenarios structured conditionally, and with the possibility of stochastic scenarios. It is a simple method, and it is by no means original; many other people have used a similar approach in various settings. This method might be helpful to those who want to forecast life expectancy. This note summarizes my presentation.

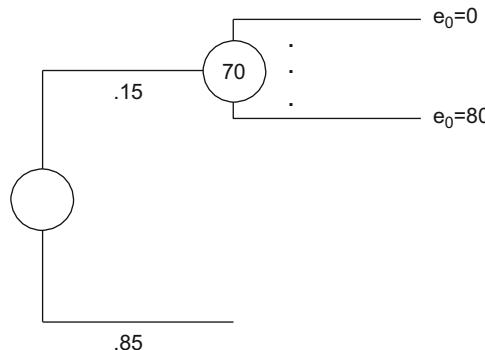
To be concrete, suppose the goal is to forecast female life expectancy at birth in Sweden in 2050. In the year 2000, the expected life span for Swedish women was just over 82 years. What will it be in the middle of the twenty-first century?

A central point of uncertainty is whether the kind of progress and development in Sweden that has marked the past couple of centuries will continue until 2050. Perhaps the future will be characterized by poverty, misery, and a shorter life expectancy. Coming decades could bring nuclear war, massive biochemical terrorism, epidemics more deadly than the AIDS epidemic, catastrophic environmental change, lasting economic depression, or some other disaster or combination of disasters that might cause female life expectancy in Sweden to plummet far below its current level of more than 82 years, perhaps even down to zero. These possibilities are indicated in the figure below. My “guesstimate,” given my current knowledge and the limited amount of time I have spent researching and thinking about this

¹In addition, I very briefly presented one simple version of it in the last paragraph on page 195 of Lutz, Vaupel and Alberg (eds.) (1998).

J. W. Vaupel (✉)
Max Planck Institute for Demographic Research, Rostock, Germany
e-mail: jwv@demogr.mpg.de

question, is that there is a 15% chance that life expectancy will decline in the future. If it does, then my best guess is that the mean value – of the range of possible life expectancies in 2050 – is 70 years, which is close to the current value of female life expectancy in the world as a whole. Discussions among a group of experts and systematic consideration of various scenarios would undoubtedly produce values different from 15% and 70 years, but these values illustrate the approach.



Suppose calamity is averted, with probability 0.85. Then the next major uncertainty would seem to be whether life expectancy is approaching a looming limit. This limit does not have to be an ultimate cap that will hold forever. It simply has to be some ceiling that Swedes will not be able to exceed by 2050. Perhaps there really is some biological limit to life expectancy of 85 or so. Current evidence suggests that this is unlikely, but it might be true. More plausibly, perhaps it will be impossible to make much progress in reducing death rates at very old ages. To achieve such reductions, new kinds of biomedical breakthroughs will be required, and these breakthroughs may not be forthcoming, at least over the next half century. Furthermore, there may be practical impediments to further reductions in mortality. For instance, taxpayers may not be willing (or able) to pay for the required interventions if there are too few workers to support increasing numbers of retired people. For illustrative purposes, suppose the probability of some such scenario is 20% and that, conditional on this, life expectancy in 2050 will be 85 for Swedish females. Of course, it might not be precisely 85, but suppose that 85 is the average value of the fairly narrow range of possibilities permitted by this line of thinking. Again, debate and systematic calculation would lead to values other than 20% and 85 years, but these values provide a suggestive example.

The final possibility, in my simple probability tree, is that the future will be roughly the same or perhaps even better than the past. The uncertainty here might be structured as follows:

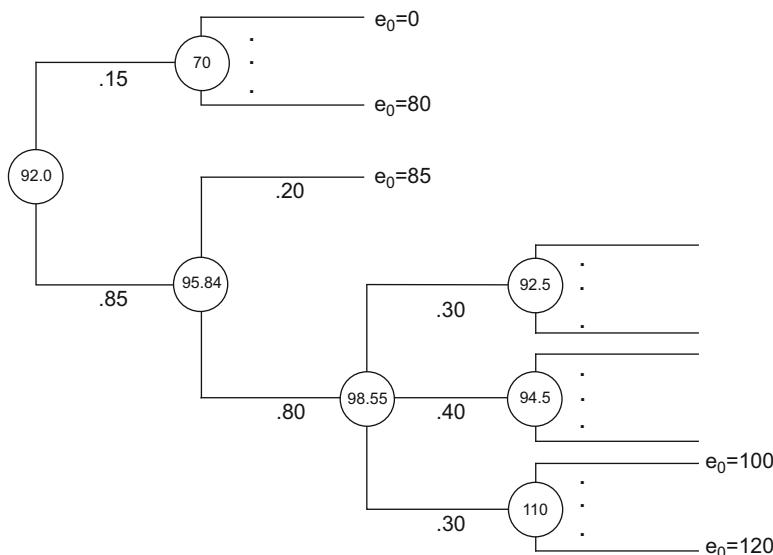
- One possibility is that the pace of age-specific mortality improvement over the next half century will be similar to the pace of improvement over the last 50 or 100 years. This is the general idea underlying the Lee-Carter-Tuljapurkar time-series method of forecasting life expectancy and Juha Alho has also done research along this line.² I am not sure what the latest results are if this approach is applied to Swedish females, but to be specific, suppose the mean value of predicted life expectancies in 2050 is 92.5.
- A second possibility is that the pace of life-expectancy increase over the next half century will be similar to the rate of increase over past decades. Jim Oeppen and I studied the rise in the expectation of life over the last 200 years. We discovered that “best-practice” life expectancy, *i.e.*, female life expectancy in the country that held the record, had risen linearly by 3 months per year from 1840 to 2000. Over this entire period, Sweden has been either at or near the top. Currently, the life expectancy of Japanese women is almost 85 years, whereas that of Swedish females is a bit more than 82 years. Thus, the current gap is between two and a half and three years. If the trend of the past 160 years continues for the next half century, then life expectancy in 2050 will reach a record 97.5. If time-series methods are applied to the data, then there will be some small band of uncertainty around this value; the band will be narrow because the trend over the past 160 years has been remarkably linear. I have not applied time-series methods to the data, however, and I am not sure exactly how wide the range of uncertainty will be. In any case, a greater source of uncertainty will involve the gap between Swedish and record life expectancy in 2050. In the figure below, I assume that the mean value of this gap is 3 years, yielding a life-expectancy estimate of 94.5.
- Finally, the third possibility is that mortality improvements will accelerate in the future. Biology and biomedicine may be on the verge of unprecedented breakthroughs in knowledge about specific diseases and about the aging process itself – many knowledgeable scientists are of this opinion. Specifically, instead of increasing by 2.5 years per decade, life expectancy may increase by 3, then 4, and then 5 years per decade over the next three decades and perhaps by 6, 8, or even 10 years per decade in the 2030s and 2040s. Since the sum of $3 + 4 + 5 + 7 + 9$ is 28, female life expectancy in Sweden in 2050 may be 110 years rather than the current 82 years or so. In the figure below, I suggest a range from 100 to 120 years if mortality improvements accelerate, with a mean value of 110. This is really just a guess, and some informed discussion and structured debate might yield quite a different picture. For illustrative purposes, however, a value of 110 may not be inappropriate.

What is the chance that the future will be like the past in terms of age-specific mortality change, that the future will be like the past in terms of life expectancy

²See Lee and Carter (1992: 659); Tuljapurkar et al. (2000); Alho (1998).

change, or that the future will bring an accelerated rate of increase in life expectancy? The trend in best-practice life expectancy is so regular that I assigned a probability of 40% to a continuation of this trend, and I gave each of the other two possibilities a 30% chance. Folding the tree back, these values lead to a mean life expectancy of 98.5 if the future is like the past or even better. If there is no disaster, then the mean is 95.84. All factors considered, the mean is 92.0. (The calculations just happened to produce a value close to this nice round number).

The future is enveloped in uncertainty, and there is a wide probability distribution around this value of 92.0, stretching from 0 to 120 in the tree below. This predictive distribution could be estimated. Some components of uncertainty could be assessed by expert judgement. Other components, as noted above, could be estimated by time-series methods. A structured conditional probability tree, of the kind shown below, could be used to organize the forecasting problem. And that is the concept underlying the SCOPE approach.



References

- Alho, J. (1998). A stochastic forecast of the population of Finland. *Reviews* 1998/4. Helsinki: Statistics Finland.
- Lee, R. D., & Carter, L. (1992). Modeling and Forecasting U.S. Mortality. *Journal of American Statistical Association*, 87(419), 659–671.
- Lutz, W., Vaupel, J. W., & Ahlburg, D. A. (Eds.). (1998). Ways to improve population forecasting: What should be done differently in the future? *Population and Development Review*, 24 (Supplement: Frontiers of Population Forecasting), 191–198.
- Tuljapurkar, S., Li, N., & Boe, C. (2000). A universal pattern of the mortality decline in the G7 countries. *Nature*, 405, 789–792.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 7

Mortality Forecasts. Comments on How to Improve Existing Models – An Epidemiologist’s Perspective



Kaare Christensen

From an epidemiologist’s perspective, one way to improve mortality forecasts is to gain insight into the causes and predictors of mortality. If we know the “risk profile” of the current cohorts compared to the previous cohorts, then our forecasts may improve.

If we also include genetics in our perspective, the initial task will be to quantify the contribution of genes and environment to lifespan. Therefore, the first question will be, “Are the lifespans of relatives correlated?” and if so, “Is the correlation due to a shared environment or to shared genes?” It is important to pose these questions initially because the answers determine whether it is worthwhile to seek the causes and predictors of mortality in the environment as well as in the genetic make-up.

7.1 Are the Lifespans of Relatives Correlated?

Traditional family studies suggest a correlation in lifespan within families. However, studies have generally found only small correlations in lifespan between parents and offspring (0.01–0.15) (Pearl 1931; Cohen 1964; Wyshak 1978), whereas correlations between siblings tend to be higher (0.15–0.35) (Cohen 1964; Wyshak 1978). Heritability estimates based on regression analysis were in the range of 0.10–0.33 for parents-offspring and 0.33–0.41 for siblings, constantly over a period of 300 years (Meyer 1991), but these estimates include both genetic factors and shared environmental factors. Some family studies have found a stronger maternal than paternal effect (Abbott et al. 1974), but not all (Wyshak 1978). The lower correlation found for parents and offspring than for siblings, suggests that genetic non-additivity

K. Christensen (✉)

Institute of Public Health, University of Southern Denmark, Odense, Denmark
e-mail: KChristensen@health.sdu.dk

(genetic effects due to gene interaction which are not passed from one generation to the next) is present. However, it may also reflect a higher degree of shared environment among siblings than among parents and offspring; the latter constitute two generations living under very different conditions.

7.2 The Relative Effects of Genetic and Environmental Factors on Lifespan

Twin studies are designed to separate the effects of additive and non-additive genetic factors, as well as shared and non-shared environmental factors. However, most of the early twin studies had methodological problems due to left-truncation of the cohorts included, selection bias, lack of zygosity diagnosis, or heavy right-censoring. Carmelli and Andersen (1981) included a sample of 2242 Mormon pairs of twins born 1800–1899 in which both twins had died, a criterion met by 60% of the original sample. Wyshak (1978) followed 972 Mormon twin-pairs (possibly included in the study of Carmelli and Andersen) until death. Unfortunately, since both studies lacked zygosity diagnosis, heritability estimates could not be provided. However, similarity in length of life was found. The similarity was more pronounced for twins of the same sex (including both MZ and DZ twins) than for twins of the opposite sex (always DZ twins), suggesting genetic influences on lifespan. Jarvik et al. (1960) followed a sample of 853 pairs of twins for 12 years; the sample included only pairs with at least one twin surviving to age 60. At the end of the follow-up period, both twins had died in only 35% of the pairs. The mean intra-pair difference in lifespan was found to be higher for DZ than in MZ twins, suggesting genetic influences on lifespan. Hrubec and Neel (1981) followed a sample of 31,848 male twin veterans born 1917–27 for 30 years to ages 51–61. Around 10% were deceased at the time of analysis. To avoid censoring problems, longevity was analyzed as a categorical variable (dead/alive). In this study, the heritability of “liability” to die was estimated to be 0.5.

The first non-censored and population-based twin study that could provide an estimate of the magnitude of genetic influences on lifespan was conducted by McGue et al. (1993). It covered 600 Danish pairs of twins born 1870–1880. Using path analysis, a heritability of 0.22 was found, with genetic influences being mainly non-additive. Later this study was expanded by Herskind et al. (1996) to include more than 2800 twin-pairs with known zygosity born 1870–1900. These cohorts were followed from age 15 to death. The study confirmed that approximately a quarter of the variation in lifespan in this population could be attributed to non-additive genetic factors, while the remaining three-quarters were due to non-shared environmental factors.

Ljungquist et al. (1998) studied the 1886–1900 Swedish cohorts of twins and concluded that around a third at most of the variance in longevity is attributable to genetic factors.

Hence, it seems to be a rather consistent finding in the Nordic countries that approximately 25% of the variation in lifespan is caused by genetic differences. It is interesting that animal studies have revealed similar estimates for a number of species not living in the wild (Curtsinger et al. 1995; Finch and Tanzi 1997).

Hence, the conclusion from these studies is that it is worthwhile to seek the causes and predictors of mortality in the environment as well as in the genetic make-up. However, the results from family studies with low correlations between family members suggest an absence of common genes with a substantial impact on lifespan.

7.3 Prediction of Mortality

From a forecast perspective, which focuses on reduction of mortality (rather than on sudden increases in mortality due to new diseases, war, etc.), there is little interest in estimating survival at younger ages, when the room for improvement is very limited. What is important for future mortality trajectories is mortality among elderly people (Vaupel et al. 1998).

A number of risk factors seem to lose their importance with age, probably because of heterogeneity and selection, *e.g.*, smoking, obesity, diseases and SES (socio-economic status). However, this does not mean that the survival rate of the elderly, including the very oldest, cannot be raised. In the latter category, the fraction dying every year is high (for example, 1/4 – 1/3 among nonagenarians), and there is both practical and theoretical evidence that intervention can have substantial positive effects on both quality of life and survival.

In relation to forecasting, it is important to note that certain predictors remain valid even at the highest ages, *e.g.*, self-rated health as well as physical and cognitive abilities (Nybo et al. 2001). This may provide an opportunity to improve forecasting if for example the physical abilities of new cohorts of elderly persons are assessed and compared to previous cohorts; *i.e.* “Are the new cohorts of the elderly healthier than the previous cohorts (and therefore expected to live longer)?” The first reports of this kind have been published based on US studies (Manton and Gu 2001; Manton et al. 1997). They indicate that the new cohorts of elderly are increasingly healthy.

7.4 Conclusion

The cohort differences in physical abilities among the elderly and the correlation between physical abilities and mortality may be the basis for improving the forecasting of mortality.

References

- Abbott, M. H., Abbey, H., Bolling, D. R., & Murphy, E. A. (1974). The familial component in longevity. A study of offspring of nonagenarians II. Preliminary analysis of the completed study. *Johns Hopkins Medical Journal*, 134(1), 1–16.
- Carmelli, D., & Andersen, S. (1981). A longevity study of twins in the mormon genealogy. *Progress in Clinical and Biological Research*, 69(Pt C), 187–200.
- Cohen, B. H. (1964). family pattern of mortality and life-span. *The Quarterly Review of Biology*, 39, 130–181.
- Curtsinger, J. W., Fukui, H. H., Khazaeli, A. A., Kirscher, A., Pletcher, S. D., Promislow, D. E., & Tatar, M. (1995). Genetic variation and aging. *Annual Review of Genetics*, 29, 553–575.
- Finch, C. E., & Tanzi, R. E. (1997). Genetics of aging. *Science*, 278(5337), 407–411.
- Herskind, A. M., Harvald, B., Holm, N. V., McGue, M., Sorensen, T. I., & Vaupel, J. W. (1996). The Heritability of human longevity: A population-based study of 2872 Danish Twin Pairs Born 1870–1900. *Human Genetics*, 97(3), 319–323.
- Hrubec, Z., & Neel, J. V. (1981). Familial factors in early deaths: Twins followed 30 years to ages 51–61 in 1978. *Human Genetics*, 59(1), 39–46.
- Jarvik, L., Falek, A., Kallmann, F. J., & Lorge, I. (1960). Survival trends in a senescent twin population. *American Journal of Human Genetics*, 12, 170–179.
- Ljungquist, B., Berg, S., Lanke, J., McClearn, G. E., & Pedersen, N. L. (1998). The Effect of genetic factors for longevity: A comparison of identical and fraternal twins in the Swedish twin registry. *The Journals of Gerontology. A Biological Science and Medical Science*, 53(6), M441–M446.
- Manton, K. G., & Gu, X. (2001). Changes in the prevalence of chronic disability in the United States black and nonblack population above age 65 from 1982 to 1999. *Proceedings of the National Academy of Sciences of the U S A*, 98(11), 6354–6359.
- Manton, K. G., Corder, L., & Stallard, E. (1997). Chronic disability trends in elderly United States populations: 1982–1994. *Proceedings of the National Academy of Sciences of the U S A*, 94(6), 2593–2598.
- McGue, M., Harvald, B., Holm, N., & Vaupel, J. W. (1993). Longevity is moderately heritable in a sample of Danish Twins Born 1870–1880. *Journal of Gerontology*, 48(6), B237–B244.
- Meyer, P. J. (1991). Inheritance of longevity evinces no secular trend among members of six New England families born 1650–1874. *American Journal of Human Biology: The Official Journal of the Human Biology Council*, 3, 49–58.
- Nybo, H., Andersen, K., Christensen, K., Gaist, D., Jeune, B., McGue, M., Petersen, H. C., & Vaupel, J. W. (2001). *Predictors of mortality among 2,249 nonagenarians the Danish 1905-Cohort Survey*. Manuscript.
- Pearl, R. (1931). Studies on human longevity IV. The inheritance of longevity. *Human Biology*, 3, 245–269.

- Vaupel, J. W., Carey, J. R., Christensen, K., Curtsinger, J. W., Holm, N. V., Iachine, I. A., Johnson, T. E., Kannisto, V., Khazaeli, A. A., Liedo, P., Longo, V. D., Manton, K. G., Yashin, A. I., & Zeng, Y. (1998). Biodemographic trajectories of longevity. *Science*, 280(5365), 855–860.
- Wyshak, G. (1978). Fertility and longevity in twins, sibs, and parents of twins. *Social Biology*, 25 (4), 315–330.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 8

The Need for Looking Far Back in Time When Predicting Future Mortality Trends



Tommy Bengtsson

There are three reasons why we must look far back in time when predicting future mortality trends. Firstly, the mortality decline that we can observe today has its roots in improvements achieved long ago in living standards and diet, public health institutions, medicine, and other areas relevant to the physical well-being of the population. Speaking in general terms, living conditions improve from one period to the next. Such improvements are called *period* factors, since they relate to living conditions for the entire population during one specific period.

Secondly, the health and remaining life span for people living today is determined not only by contemporary period factors but also by living conditions earlier in life. Since even conditions in the foetal stage have an influence on longevity, improvements during early childhood could have an effect on mortality trends today. These are called *cohort* factors, since they relate to the conditions for a certain cohort, often birth cohort, while previous cohorts are unaffected. Since the oldest people living today were born a hundred or more years ago, we have to consider cohort factors far back in time in predicting future mortality.

Thirdly, predicting future mortality trends also requires a multivariate point of departure. No single factor, but a variety of factors, determines health and remaining life span, and we do not know *a priori* which one is the most important. Thus, predicting future mortality trends calls for a long-term multivariate, causal approach, in which both period and cohort factors are taken into account. This kind of holistic view has not always been in favour. Early in the twentieth century, cohort factors were considered most important, while multivariate period factors became more popular later on, followed in turn by a preference for single-period factors as the main determinant of the great mortality decline since the mid-eighteenth century. Now we are once again discussing cohort factors.

T. Bengtsson (✉)

Centre for Economic Demography, Lund University, Lund, Sweden
e-mail: tommy.bengtsson@ekh.lu.se

Thus, there has been a shift over time in our explanations for the mortality decline (Bengtsson 1998). Below I shall develop the reasons why I think we need to take a multivariate, period, and cohort approach. I will start with an overview of the great mortality decline, commencing several hundred years ago, and then move on to discuss how the explanations for the historical decline have changed over time.

Over the past 300 years, human physiology has undergone profound changes. These were made possible by numerous advances whereby humans have gained an unprecedented degree of control over their environment. The changes include increases in body size by over 50%, in length of life by over 100%, and in time lived in retirement by several hundred per cent, with obvious implications for expenditures on pensions and health care. The extension of the human life span has been gradual, starting a long time ago with what now is called the great mortality decline.

The timing of the great mortality decline was strikingly similar in the countries of Western and Northern Europe despite differences in economic structure and development. It started in the mid-eighteenth century, levelled off for a couple of decades in the mid-nineteenth century, and then continued. Over this period, life expectancy at birth rose from some 35 years to more than 70 years. The development in North America was similar.

Few scholars today will argue that the great mortality decline was due solely or primarily to a single factor. Economic growth, for example, was probably not a major determinant before or during the initial stages of the great mortality decline, and its effect may be far less than expected during later stages of the decline. Instead, the causes are multi-factorial and vary from the start to the end of the decline.

In countries for which long historical series of age-specific mortality rates (England, France, and some others, including the Nordic countries) are available, we know that the increase in life expectancy began with a fall in the rates of infant and child mortality. This was mainly due to a reduction in deaths from smallpox, which was a very common childhood disease in the eighteenth century.

While infant mortality continued to decline in Sweden throughout the nineteenth century, it levelled off in England and remained stable until the end of the nineteenth century, when it again dropped rapidly, as was the case in all other Western countries. Childhood mortality in both countries actually increased in the mid-nineteenth century before commencing a persistent decline. This exemplifies two patterns of the historical decline in infant and childhood mortality: a northern pattern with declines essentially throughout the nineteenth century, and a western one with a levelling off during the nineteenth century and a further decline after the 1880s (Perrenoud 1984).

Adult and old-age mortality started to decline gradually at the beginning of the nineteenth century, possibly earlier for England. The decrease was generally sharper in the later years of the nineteenth century and even more so after the First World War, as was the case with mortality at all other ages. The decline slowed for adults and the elderly around 1950 but accelerated again in the 1970s.

With the great mortality decline followed a change in the leading causes of death, from pestilence, to receding pandemics, and then to man-made diseases (Omran

1971). In his theory of epidemiological transition, Omran identified three different development patterns: the classical or western pattern, described above, the accelerated pattern, and the delayed pattern. They are distinguished by differences in timing and speed. The decline that took 200 years in the West started 150–200 years later in the Third World but then took less than 50 years to complete. In many countries, the great mortality decline has not yet occurred, and the gains from low mortality are still to be reaped.

The great mortality decline can be viewed in light of two approaches, one based on *period factors*, the other on *cohort factors*. Most studies focus largely on period factors. In addition, factors based almost entirely on human actions should be distinguished from those beyond deliberate human control. One widely accepted multi-factorial explanation (United Nations 1953) is based on *period factors* that depend upon human activity: these include public health reforms, advances in medical knowledge, improved personal hygiene, and rising income and standards of living. This explanation is highly similar to the demographic-transition theory (Davis 1945; Notestein 1953; Bengtsson and Ohlsson 1994).

On a broader front, McKeown (1976) questioned the multi-factorial explanation of the great mortality decline. He argued that a single factor – better nutrition – could almost entirely explain the great mortality decline. His criticism was based on a study of cause-specific mortality in England and Wales from 1838 to 1947, where he observed that two-thirds of the mortality decline was due to a reduction in infectious diseases. In later work, he also analysed mortality rates and economic development for other countries and further back in time, though not in such detail as for England and Wales.

McKeown argued that medical advances had little influence on mortality trends before the breakthrough of sulphonamides and antibiotics in the 1930s and 1940s. Previously, the only curable disease had been diphtheria. Very few deaths, however, were due to diphtheria, and its incidence was already receding when the antitoxin came into use around 1900. For periods before 1838, McKeown held that inoculation and vaccination for smallpox had little or no impact on the general mortality decline. Vaccination started in England at the end of the eighteenth century, but it did not become widespread until after 1840, when it was made available at public expense. The history of vaccination in the Nordic countries is similar. There, smallpox mortality was declining even before vaccination started in the first years of the nineteenth century and well before it became common in the 1820s. However, not all scholars agree on this interpretation. Easterlin (1999) downplays the role of economic development and better diet, contending that advancement in medical knowledge is the principal reason for the mortality decline from the mid-nineteenth century onwards.

The improvement of personal hygiene may have had some effect on mortality in England and Wales after about 1880, when the incidence of intestinal infections declined, probably because of substantial improvements in water supply and sewage control at that time. Prior to 1870, the decrease in intestinal infections accounted for only a small part of the general decline in mortality, and according to McKeown, neither better personal hygiene nor measures to improve public health had any

significant impact on the overall decline before 1870. Other public-health measures, such as the breast-feeding campaigns in Sweden in the 1830s and thereafter, were not discussed by McKeown. Fridlizus (1984) has argued that while these campaigns might well have had an impact on infant mortality in Sweden, they began much later than the decline in mortality. In fact, childhood mortality went up somewhat at the time of the first breast-feeding campaigns.

In 1957, Helleiner argued that the population of Western Europe must have increased from the mid-eleventh century to the late thirteenth century and from the mid-fifteenth century to the end of the sixteenth century (Helleiner 1957). The population increase observed in the eighteenth century was therefore unique only in that the mortality decline started from a higher level and went on longer than before. Helleiner and other scholars have contended that these changes in mortality were spontaneous (natural), and in 1973 the UN added natural factors as a fifth determinant of the great mortality decline (United Nations 1973). Later, others like Fridlizus (1984), Perrenoud (1984) and Schofield (1984) asserted that a change in the virulence of pathogens initiated the great mortality decline. Their basic assumption was that the virulence of pathogens changes spontaneously over time. Virulence in a pathogenic organism is generally understood as its ability to overcome host defences. At this point, it is important to note that some pathogens develop more quickly in a malnourished host, whereas others do not depend on a weakening of the host to produce very high mortality.

McKeown's main argument on this subject, though not proved, is that the initial development of mortality in the late eighteenth century was an integral part of the great mortality decline. Since the decline continued for the next two centuries, it cannot be due to a spontaneous reduction in the virulence of pathogens; thus, according to McKeown, the only explanation left is better nutrition. Improved nutrition would explain not only the decline in infectious diseases from the mid-nineteenth century onwards, but also the initial decline. This reasoning is part of McKeown's attempt to find one single explanation for the entire mortality decline.

Fogel (1994) criticized McKeown's reasoning for considering only nutritional intake, or diet, while ignoring the needs of the body to maintain itself and build up cells. Thus, McKeown only took into account gross nutrition, rather than net nutrition; the latter must be more closely related to health and mortality. I will come back to this issue shortly.

Cohort explanations for the mortality decline refer to factors that initially affect only certain young age groups but may have a long-lasting impact on these groups. Such factors would consist mainly of improvements in childhood conditions, or even conditions during the foetal stage, that have lasting effects on health and on the life span. Net nutrition is seen as the principal determinant of cellular development, which is most rapid during the foetal stage and gradually diminishes until the body is fully developed around the age of 20. Net nutrition is what is left for the development of cells after the nutritional requirements of other life-sustaining functions and work have been met. Thus, low net nutrition could be due either to low nutritional intake or to additional, disease-related needs of the body for nutrition. Moreover, many diseases not only claim nutrients but also make it more difficult for the body to

absorb nutrients in general, as is the case with infectious diseases. If cells and organs consequently fail to develop properly, a child's growth and development may be inhibited, and the child may be less healthy in general. Thus, we can differentiate between two basic types of cohort explanations for the mortality decline, namely (i) increased nutritional intake during the foetal stage and/or early years of life, and (ii) decreased needs for nutrition during the foetal stage or early years of life owing to less disease in the mother or the child.

The importance of early childhood conditions for later life has probably been well known since time immemorial. It is often assumed that each generation shows the same relative mortality from childhood to old age. Kermack, McKendrick and McKinley adopted this assumption in 1934 when scholars were starting to become aware of the great mortality decline (Kermack et al. 1934). Strangely enough, cohort factors were out of fashion when the UN in 1953 and 1973 made its synthesis of the causes of the great mortality decline. Over the last couple of decades, both medical and historical research on this matter has expanded rapidly (for an overview, see Elo and Preston 1992).

In this connection, the work by Barker (1994 and elsewhere) has been of major importance. He has summarised the medical evidence showing the importance of foetal and neonatal nutrition for adult health. In historical research, Fogel (1996) is probably the leading advocate of these ideas. In addition, Preston and van de Walle (1978) for urban France, and Fridlizius (1989) for Sweden, emphasised the importance of cohort factors for the mortality decline. Bengtsson and Lindström found that exposure to disease from airborne infections has a strong effect on mortality among the elderly (Bengtsson and Lindström 2000, 2001).

Steckel (1983) and Fogel (1996 and previous work) use final body height as a measure of net nutrition and health during childhood. Individuals who have had well-nourished and healthy mothers, and thus have been well nourished themselves during the foetal stage, have a lower risk of death during infancy. If they are well nourished and healthy, their cells and organs develop better, they grow taller, and they live longer. Since health is determined by net nutrition and not gross nutrition, there is no direct link between gross nutrition and height attained. Improvements in health and height may be due either to better nutrition (better diet) or to more limited claims on nutrition from disease. Thus, a decline in the prevalence of smallpox, for example, has a positive effect on height and extends the life span, everything else being equal. The problem is to evaluate how much of the improvement in health is due to diet and how much is due to less disease. Calculating diets for pre-modern populations is a difficult task (Fogel 1996), and it is even harder to calculate disease-related claims on nutrition. Still, historical records show similarities between trends in height and gross consumption of nutrition (Fogel 1994, 1996), indicating that trend in disease-related claims has been of minor importance. If that is the case, then McKeown's focus on gross nutrition may be justified.

Whether due to a low or badly composed nutritional intake, or to greater claims on nutrition from diseases, undernourishment may stunt growth in height or weight and lead to illness, disease and mortality later in life more than in the immediate future. The immediate relationship, or period link, between the economy and

mortality, therefore, is much weaker than Malthus believed, according to Fogel. The rather tenuous short-term relationship often found between prices and deaths for many European countries, as shown by Lee (1981, 1993), Galloway (1988) and others, supports this interpretation (Fogel 1994). Thus, cohort factors matter more to the mortality decline than period factors, according to Fogel.

To summarise, few scholars today will argue that any single factor is the primary determinant of the great mortality decline. Of course, it is no coincidence that the vast growth in resources resulting from the transformation of our economies in the eighteenth and nineteenth centuries was concurrent with the great mortality decline and with the fertility transition as well. In a millennial perspective, these events took place at about the same time. This is not to say, however, that there is a close relationship between the economy and the great mortality decline. On the contrary, economic growth has probably not been a major determinant, either before the great mortality decline or in its initial stages, and its impact may have been far less than expected during later stages of the decline. Instead, the causes are multi-factorial and vary from the start to the end of the decline.

Initially, the decline may well have been partly due to pure luck, for example spontaneously less aggressive smallpox pathogens as part of an old demographic pattern rather than the result of a modernisation process. Later development and compulsory use of vaccine surely prevented the re-emergence of the more aggressive virus. Improvements in nourishment and in the care of mothers and children had long-lasting effects on life span. Advancements in water supply and sanitation as well as better housing contributed to the decline from the second part of the nineteenth century onwards. Medical progress in the twentieth century prolonged life. The fact that health is determined by net nutrition – intake of nutrition minus claims on nutrition – and that claims are partly disease-related, makes it difficult to evaluate the determinants of the mortality decline. The influence of conditions in childhood – including temporary ones – on mortality at older ages adds further complexity. The impact of each individual factor is therefore very difficult to measure; to do so would require a long series of high-quality data. Several variables, such as the virulence of pathogens and the claims on nutrients due to disease can at best be estimated indirectly, if at all. The analyses based on highly aggregated longitudinal data have identified the problems and directed us toward the solutions. It is more doubtful, however, whether solutions will be found at that level of analysis. Perhaps the use of longitudinal micro data will serve as valuable complement, as it has in other areas.

In my opinion, the development of theories about the great mortality decline can be summarised in Fig. 8.1, which also could give us some guidance for the future.

Thus, the cause of the great mortality decline is clearly multi-factorial, and the importance of the various factors changes over time. Both period and cohort factors must be taken into account in analysing the decline. A longterm perspective is essential in predicting future mortality trends. Combining longitudinal information for individuals with information at the societal level is likely to provide important information about mortality determinants in the past and may be useful for predicting

	Period Factors	Cohort Factors
Univariate Models	1970s McKeown <i>1940s forecasts</i>	1930s Kermack et al <i>1990s forecasts</i>
Multivariate Models	1950s UN	1990s Barker and Fogel <i>2000s mortality forecasts?</i>

Fig. 8.1 Development of models for explaining the great mortality decline and population forecasts

future trends as well. Previously, both population and mortality projections were based largely on periodic information about demographic trends. The use of demographic cohort information in the 1990s was a considerable advancement. The question now is whether we are ready to take a great step forward by using multivariate causal models that combine information at the individual and societal levels.

References

- Barker, D. J. P. (1994). *Mothers, babies and disease in later life*. London: BMJ.
- Bengtsson, T. (1998). The great mortality decline: Its causes and consequences. In C.-E. Núñez (Ed.), *Debates and controversies in economic history*. Madrid: Fundación Ramón Areces.
- Bengtsson, T., & Lindström, M. (2000). Childhood misery and disease in later life: The effects on mortality in old-age of hazards experienced in early life, southern Sweden, 1760–1894. *Population Studies*, 54(2), 263–277.
- Bengtsson, T., & Lindström, M. (2001, November). *Airborne infectious diseases in infancy and mortality in later life, southern Sweden 1766–1894*. Paper presented at the Social Science History Association conference in Chicago.
- Bengtsson, T., & Ohlsson, R. (1994). The demographic transition revised. In T. Bengtsson (Ed.), *Population, economy, and welfare in Sweden*. Berlin: Springer.
- Davis, K. (1945). The world demographic transition. *Annals of the American Academy of Political and Social Sciences*, 237, 1–11.
- Easterlin, R. A. (1999). How beneficent is the market? A look at the modern history of mortality. *European Review of Economic History*, 3(3), 257–294.
- Elo, I., & Preston, S. (1992). Effects of early-life conditions on adult mortality: A review. *Population Index*, 59(2), 186–212.
- Fogel, R. W. (1994). The relevance of Malthus for the study of mortality today: Long-run influences on health, mortality, labour force participation, and population growth. In K. Lindahl-Kiessling & H. Landberg (Eds.), *Population, economic development, and the environment*. Oxford: Oxford University Press.

- Fogel, R. W. (1996). *The Escape from hunger and premature death 1700–2100: Europe, America and the Third World*. The 1996 Ellen McArthur Lectures, Cambridge.
- Fridlizius, G. (1989). The deformation of cohorts: Nineteenth century mortality decline in a generational perspective. *Scandinavian Economic History Review*, 37(3), 3–17.
- Fridlizius, G. (1984). The mortality decline in the first phase of the demographic transition: Swedish experiences. In T. Bengtsson, G. Fridlizius, & R. Ohlsson (Eds.), *Pre-industrial population change: The mortality decline and short-term population movements* (pp. 71–114). Lund: Almqvist & Wiksell International.
- Galloway, P. R. (1988). Basic patterns in annual variations in fertility, nuptiality, mortality and prices in pre-industrial Europe. *Population Studies*, 42, 275–302.
- Helleiner, K. F. (1957). The vital revolution reconsidered. *Canadian Journal of Economics and Political Science*, 23, 1–9.
- Kermack, W. O., McKendrick, A. G., & McKinley, P. L. (1934, March). Death rates in Great Britain and Sweden: Some regularities and their significance. *The Lancet*, 223, 698–703.
- Lee, R. D. (1981). Short-term variation: Vital rates, prices and weather. In E. A. Wrigley & R. S. Schofield (Eds.), *The population history of England, 1541–1871: A reconstruction*. London: Edward Arnold.
- Lee, R. D. (1993). Inverse projections and demographic fluctuations: A critical assessment of new methods. In D. Reher & R. S. Schofield (Eds.), *Old and new methods in historical demography*. Oxford: Oxford University Press.
- McKeown, T. (1976). *The modern rise of population*. London: Edward Arnold.
- Notestein, F. W. (1953). *Economic problems of economic change*. In Proceedings of the eight international conference of agricultural economics, London.
- Omran, A. R. (1971). The epidemiological transition. A theory of the epidemiology of population change. *Millbank Memorial Fund Quarterly*, 49(4), 509–538.
- Perrenoud, A. (1984). The mortality decline in a long term perspective. In T. Bengtsson, G. Fridlizius, & R. Ohlsson (Eds.), *Pre-industrial population change: The mortality decline and short-term population movements*. Almqvist & Wiksell International: Lund.
- Preston, S. H., & van de Walle, E. (1978). Urban French history in the nineteenth century. *Population Studies*, 32, 275–297.
- Schofield, R. S. (1984). Population growth in the century after 1750: The role of mortality decline. In T. Bengtsson, G. Fridlizius, & R. Ohlsson (Eds.), *Pre-industrial population change*. Almqvist & Wiksell International: Lund.
- Steckel, R. (1983). Height and per capita income. *Historical Methods*, 16, 1–7.
- United Nations. (1953). *The determinants and consequences of population trends* (Population Studies, No 17). New York.
- United Nations. (1973). *The determinants and causes of population trends* (Population Studies, No 50). New York.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Part II

Probabilistic Models

Nico Keilman

The chapters in this part focus on probabilistic (also labelled as stochastic) forecasts, in other words forecasts in which uncertainty has been quantified. Given a history of sizable forecasting errors, the first paper, by Nico Keilman, addresses the question of why demographic forecasts are uncertain. He summarizes main patterns in *ex-post* observed errors in historical population forecasts, and discusses probabilistic forecasts. Illustrations come from a probabilistic forecast for the population of Norway.

In the second paper, Juha Alho outlines the statistical background of uncertain events and forecasts of these. The purpose of his remarks is to review some probability models, and to show how people with different backgrounds may interpret probabilities differently. The basic cause of the difficulties, and disagreements, is that there are several layers of probabilities to be considered. Consequently, it is essential to be explicit about the details of the model.

In the third paper, Maarten Alders and Joop de Beer examine the approach taken by Statistics Netherlands in their stochastic forecast of mortality. They discuss the use of expert knowledge for the specification of the uncertainty of future mortality. They also describe the methodology underlying the Dutch stochastic population forecasts.

Finally, in the fourth paper, Shripad Tuljapurkar presents a model for mortality analysis and forecasting that has proven to be feasible for probabilistic forecasts. He gives illustrations of US and Swedish mortality, and discusses also possible implications of uncertain mortality for future pension expenditures.

N. Keilman

Department of Economics, University of Oslo, Oslo, Norway

e-mail: nico.keilman@econ.uio.no

Chapter 9

Erroneous Population Forecasts



Nico Keilman

9.1 Forecast Accuracy

World population in the year 2000 was 6.09 billion, according to recent estimates by the United Nations (UN 2005). This number is almost 410 million lower than the year 2000-estimate that the UN expected in 1973. The UN has computed forecasts for the population of the world since the 1950s. Figure 9.1 shows that the calculations made in the 1980s were much closer to the current estimate than those published around 1990. Subsequent forecasts for the world population in 2000 show an irregular pattern: apparently, in 1973 and around 1990 it was rather difficult to predict world population size and much less so in the mid-1980s.

At first sight, the relative differences in Fig. 9.1 appear small. The highest forecast came out in 1973. That forecast numbered 6.49 billion, only 6% higher than the current estimate of 6.09 billion. However, the difference is much larger in terms of population *growth*. The 1973 forecast covered the period 1965–2000. During those 35 years, a growth in world population by 3.20 billion was foreseen. According to the current estimate, the growth was 16% lower: only 2.7 billion persons.

An important reason for lower population growth is that the world's birth rates fell stronger than previously thought. Thirty years ago, the UN expected a drop in total fertility by 1.4 children between the periods 1965–1970 and 1995–2000: from 4.7 to 3.3 children per woman on average. Recent estimates indicate that fertility initially was higher than previously thought, and that it fell steeper than expected in that 30-year period, from 4.9 to 2.8.

Accuracy statistics of the type given here are important indicators when judging the quality of population forecasts. Other aspects, such as the information content (for instance, does the forecast predict only total population, or also age groups?) and

N. Keilman (✉)

Department of Economics, University of Oslo, Oslo, Norway
e-mail: nico.keilman@econ.uio.no

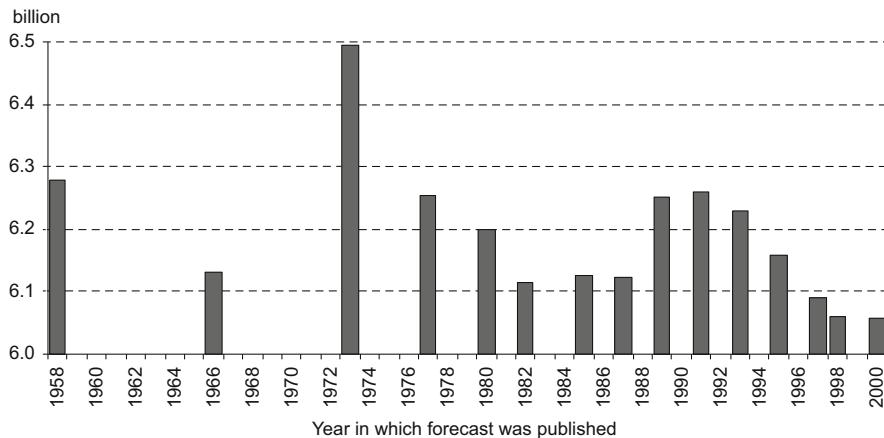


Fig. 9.1 Zooming in on the year 2000 world – population at the end of the twentieth century

the usefulness for policy purposes (for instance, does the predicted trend imply immediate policy measures?) are relevant as well. Nevertheless, the degree to which the forecast reflects real trends is a key factor in assessing its quality, in particular when the forecast is used for planning purposes. For example, imagine a forecast, for which the odds are one against two that it will cover actual trends. This forecast should be handled much more cautiously than one that can be expected to be in error only one out of five times.

The purpose of this chapter is to give a broad review of the notions of population forecast errors and forecast accuracy. Why are population forecasts inaccurate? How large are the errors involved, when we analyse historical forecasts of fertility, mortality, and the age structure? Moreover, how can we compute expected errors in recent forecasts? We shall see that *probabilistic* population forecasts are necessary to assess the expected accuracy of a forecast, and that such probabilistic forecasts quantify expected accuracy and expected forecast errors much better than traditional deterministic forecasts do. The chapter concludes with some challenges in the field of probabilistic population forecasting.

The focus in this chapter is on population forecasts at the *national* level, computed by means of the *cohort component method*. I have largely restricted myself to national forecasts, because most of the empirical literature on forecast errors and forecast accuracy deals with forecasts at that level. Notable exceptions, to be discussed below, are analyses for major world regions by Lutz et al. (1996, 2001), and for all countries in the world by the US National Research Council (NRC 2000). The empirical accuracy of subnational population forecasts has been evaluated since the 1950s (Smith et al. 2001), but the *expected* accuracy of such forecasts is largely uncharted terrain, *cf.* the concluding section. I focus on the cohort component method of population forecasting, because this method is the standard approach for population forecasting at the national level (Keilman and Cruijsen 1992). Most of the empirical evidence stems from industrialized countries, although findings for less-developed countries will be mentioned occasionally.

Various terms are in use to express accuracy, and lack thereof. I shall use inaccuracy and uncertainty as equivalent notions. When a forecast is accurate, its errors are small. Forecast errors are a means of quantifying forecast accuracy and forecast uncertainty. *Empirical* errors may be computed based on a historical forecast, when its results are compared with actual population data observed some years after the forecast was computed. For a recent forecast, this is not possible. In that case, one may compute *expected* errors, by means of a statistical model.

9.2 Why Population Forecasts Are Inaccurate

Population forecasts are inaccurate because our understanding of demographic behaviour is imperfect. Keyfitz (1982) assessed various established and rudimentary demographic theories: demographic transition, effects of development, Caldwell's theory concerning education and fertility, urbanization, income distribution, Malthus' writings on population, human capital, the Easterlin effect, opportunity costs, prosperity and fertility, and childbearing intentions. He tried to discover whether these theories had improved demographic forecasting, but his conclusion was negative. Although many of the theories are extensively tested, they have limited predictive validity in space and time, are strongly conditional, or cannot be applied without the difficult prediction of non-demographic factors. Keyfitz' conclusion agrees with Ernest Nagel's opinion from 1961, that "... (un)like the laws of physics and chemistry, generalizations in the social sciences ... have at best only a severely restricted scope, limited to social phenomena occurring during a relatively brief historical epoch with special institutional settings." Similarly, Raymond Boudon (1986) concluded that general social science theories do not exist – they are all partial and local, and Louis Henry (1987) supports that view for the case of demography. Applied to demographic forecasting, this view implies that uncertainty is inherent, and not merely the result of our ignorance. Individuals make unpredictable choices regarding partnership and childbearing, health behaviour, and migration. Note that the views expressed by Nagel and Boudon are radically different from Laplace's view on chance and uncertainty: "Imagine ... an intelligence which could comprehend all the forces by which nature is animated ... To it nothing would be uncertain, and the future, as the past, would be present to its eyes." (Laplace 1812–1829). This view suggests that our ignorance is temporary, and good research into human behaviour will increase our understanding and help formulating accurate forecasts.

Whichever view is correct, demographic behaviour is not well explained as of today. When explanation is problematic, forecasting is even more difficult. Therefore, in addition to whatever fragmentary insight demographers obtain from behavioural sciences, they rely heavily on current real trends in vital processes, and they extrapolate those trends into the future. Hence, they face a problem when the indicators show unexpected changes in level or slope. It will not be clear whether these are caused by random fluctuations, or whether there is a structural change in the

underlying trends. A trend shift that is perceived as random will first lead to large forecast errors. This effect is known in forecasting literature as assumption drag (Ascher 1978). Later, when the new trend is acknowledged, it will be included in the forecast updates and the errors will diminish. On the other hand, random fluctuations that are perceived as a trend shift will cause forecast errors, which will have a fluctuating effect on subsequent forecasts.

9.3 Empirical Evidence from Historical Forecasts

There is a large literature, in which historical population forecasts are evaluated against observed statistics (Preston 1974; Calot and Chesnais 1978; Inoue and Yu 1979; Keyfitz 1981; Stoto 1983; Pflaumer 1988; Keilman 1997, 1998, 2000, 2001; Keilman and Pham 2004; National Research Council 2000). These studies have shown, among others, that forecast accuracy is better for short than for long forecast durations, and that it is better for large than for small populations. They also learned us that forecasts of the old and the young tend to be less accurate than those of intermediate age groups, and that there are considerable differences in accuracy between regions and components. Finally, poor data quality tends to go together with poor forecast performance. This relationship is stronger for mortality than for fertility, and stronger for short-term than for long-term forecasts. Selected examples of these general findings will be given below.

9.3.1 *Forecasts Are More Accurate for Short Than for Long Forecast Durations*

Duration dependence of forecast accuracy is explained by the fact that the more years a forecast covers, the greater is the chance that unforeseen developments will produce unexpected changes in fertility, mortality, or migration.

The US National Research Council (NRC) evaluated the accuracy of nine total population size forecasts for countries of the world. Four of these were published by the United Nations (between 1973 and 1994), four by the World Bank (between 1972 and 1990), and one by the US Census Bureau (1987). The absolute percentage error, that is the forecast error irrespective of sign, increased from 5% on average for 5-year ahead forecasts, to 9% 15 years ahead, and to 14% 25 years ahead (NRC 2000). The average was computed over all countries and all forecasts. Developed countries had errors that were lower, and increased slower by forecast duration: from 2 (5 years ahead) to 4–5 (25 years ahead) %. A striking feature of these errors is that, even at duration zero, *i.e.*, in the forecast's base year, the errors are not negligible. Hence, forecasts start off with an incorrect base line population. For countries in

Africa and the Middle East this base line error was highest: 5%. Base line errors reflect poor data quality: when the forecasts were made, demographers worked with the best data that were available, but in retrospect, those data were revised.

Total fertility showed average errors from 0.4 children per woman after 5 years, to 0.6 and 0.8 children per woman after 15 and 25 years, with higher than average errors for European countries. In an evaluation of ten TFRforecasts made by the UN since 1965, I found that for Europe as a whole, TFR errors were lower, and increased slower: from 0.2 children per woman after 5 years, to 0.5 after 15 years (Keilman 2001). An analysis of the errors observed in TFR forecasts in 14 European countries made since the 1960s shows that TFR-predictions have been wrong by 0.3 children per woman for forecasts 15 years ahead, and 0.4 children per woman 25 years ahead (Keilman and Pham 2004). Life expectancy was wrong by 2.3 (5 years ahead), 3.5 (15 years ahead) and 4.3 (25 years ahead) years on average in the NRC evaluation. In 14 European countries, life expectancy forecasts tended to be too low by 1.0–1.3 and 3.2–3.4 years at forecast horizons of 10 and 20 years ahead, respectively.

9.3.2 *Forecasts Are More Accurate for Large Than for Small Populations*

A size effect in empirical errors at the sub national level was established already 50 years ago (White 1954), and reconfirmed repeatedly (see Smith et al. 2001 for an overview). Schéele (1981) found that the absolute error in small area forecasts within the Stockholm area was approximately proportional to the square root of population size, *i.e.*, a power of 0.5 (see also Bandel Bäckman and Schéele 1995). Later, Tayman et al. (1998) confirmed such a power law for small area forecasts in San Diego County, California, when they found that the mean absolute percentage forecast error was proportional to population size raised to the power 0.4.

At the international level, the NRC analysis referred to earlier showed that the absolute percentage error in forecasts of total population size was 5.5% on average, the average being taken over all countries and all nine forecast rounds. However, for countries with less than one million inhabitants, the average was 3 percentage points *higher*; for countries with a population of at least one million, the error was 0.7 percentage points *lower* (controlling, among others, for forecast length, year forecasted, forecast round, and whether or not the country had had a recent census; see NRC 2000, Appendix Table B7).

There are three reasons for the size effect in forecast accuracy. First, at the international scale, forecasters tend to pay less attention to the smallest countries, and take special care with the largest ones (NRC 2000). Second, both at the international and the local scale, small countries and areas are stronger affected by random fluctuations than large ones. In fact, many errors at the lower regional level cancel after aggregation. This explains irregular patterns and randomness in

historical series of vital statistics at the lower level, leading to unexpected real developments after the forecast was produced. Third, for small areas the impact of migration on total population is strong compared to fertility and mortality, while, at the same time, migration is the least predictable of the three components.

9.3.3 Forecasts of the Old and the Young Tend to Be Less Accurate Than Those of Intermediate Age Groups

In medium sized and large countries and regions, international migration has much less effect on the age structure than fertility or mortality. Therefore, a typical age pattern is often observed for accuracy. For many developed countries, a plot of relative forecast errors against age reveals large and positive errors (*i.e.*, too high forecasts) for young age groups, and large negative errors (too low forecasts) for the elderly. Errors for intermediate age groups are small. This age effect in forecast accuracy has been established for Europe, Northern America, and Latin America, and for countries such as Canada, Denmark, the Netherlands, Norway, and the United Kingdom (Keilman 1997, 1998). The fall in birth rates in the 1970s came fully unexpected for many demographers, which led to too high forecasts for young age groups. At the same time, mortality forecasts were often too pessimistic, in particular for women – hence the forecasts predicted too few elderly. The relative errors for the oldest old are often of the same order of magnitude as those for the youngest age groups: plus or minus 15% or more for forecasts 15 years into the future.

9.3.4 Accuracy Differs Between Components and Regions

In an analysis of the accuracy of 16 sets of population projections that the UN published between 1951 and 1998, I found considerable variation among ten large countries and seven major regions (Keilman 2001). Problems are largest in pre-transition countries, in particular in Asia. The quality of UN data for total fertility and the life expectancy has been problematic in the past for China, Pakistan, and Bangladesh. The poor data quality for these countries went together with large errors in projected total fertility and life expectancy. For Africa as a whole, data on total population and age structure have been revised substantially in the past, and this is a likely reason for the poor performance of the projections in that region. Nigeria, the only African country in my analysis, underwent major revisions in its data in connection with the Census of 1991. In turn, historical estimates of fertility and mortality indicators had to be adjusted, and this explains large projection errors in the age structure, in total fertility and in the life expectancy for this country. The problematic data situation for the former USSR is well known, in particular that

for mortality data. The result was that, on average, life expectancy projections were too high by 2.9 years, which in turn caused large errors in projected age structures for the elderly. For Europe and Northern America, data quality is generally good. Yet, as noted in Sect. 9.3.3, the two regions have large errors in long-range projections of their age structures, caused by unforeseen trend shifts in fertility and mortality in the 1960s and 1970s.

The analysis of the statistical distribution of observed forecast errors for 14 European countries showed that a normal distribution fitted well for errors in life expectancies (Keilman and Pham 2004): TFR-errors, on the other hand, were exponentially distributed. This indicates that the probability for extremely large error values was greater for the TFR than for the life expectancy. Extreme errors for net migration are even more likely.

9.4 The Expected Accuracy of Current Forecasts

Forecast users should be informed about the expected accuracy of the numbers they work with. It focuses their attention on alternative population futures that may have different implications, and it requires them to decide what forecast horizon to take seriously. Just because a forecast covers 100 years does not mean that one should necessarily use that long a forecast (NRC 2000). In that sense, empirical errors observed in a series of historical forecasts for a certain country can give strong indications of the accuracy of the nation's current forecast. However, these historical errors are just one realization of a statistical process, which applied to the past. Expected errors for the *current* forecast can only be assessed when the population forecast is couched in probabilistic form.

A probabilistic population forecast of the cohort component type requires the joint statistical distribution of all of its input parameters. Because there are hundreds of input parameters, one simplifies the probabilistic model in two ways. First, one focuses on just a few key parameters (for instance, total fertility, life expectancy, net immigration).¹ Second, one ignores certain correlations, for instance those between components, and sometimes also those in the age patterns of fertility, mortality, or migration.²

¹A cohort component forecast that has 1-year age groups requires 35 fertility rates, 200 death rates, and some 140 parameters for net migration *for each forecast year*. With age groups and time intervals equal to 50 years, a forecast for a period of 50 years, say, still requires that one specify the joint statistical distribution of $(7 + 40 + 28) \times 10 = 750$ parameters.

²For Western countries, there is little or no reason to assume correlation between the components of fertility, mortality, and migration. Nor is there any empirical evidence of such correlation (Lee and Tuljapurkar 1994; Keilman 1997). In developing countries, disasters and catastrophes may have an impact both on mortality, fertility, and migration, and a correlation between the three components cannot be excluded. There may also be a positive correlation between the levels of immigration and childbearing in Western countries with extremely high immigration from developing countries.

In probabilistic forecasts, an important type of correlation is that across time (serial correlation). Levels of fertility and mortality change only slowly over time. Thus, when fertility or mortality is high one year, a high level the next year is also likely, but not 100% certain. This implies a strong, but not perfect serial correlation for these two components. International migration is much more volatile, but economic, legal, political, and social conditions stretching over several years affect migration flows to a certain extent, and some degree of serial correlation should be expected. In the probabilistic forecasts for the United States (Lee and Tuljapurkar 1994), Finland (Alho 1998), the Netherlands (De Beer and Alders 1999), and Norway (Keilman et al. 2001, 2002) these correlation patterns were estimated based on time series models. For Austria (Hanika et al. 1997) and for large world regions (Lutz and Scherbov 1998a, b) perfect autocorrelation was assumed for the summary parameters (total fertility, life expectancy, and net migration). This assumption underestimates uncertainty (Lee 1999). In recent work for world regions, Lutz, Sanderson, and Scherbov relaxed the assumption of perfect autocorrelation (Lutz et al. 2001).

Three main methods are in use for computing probabilistic forecasts of the summary indicators: time series extrapolation, expert judgement, and extrapolation of historical forecast errors (Lee 1999; NRC 2000). The three approaches are complementary, and elements of all three are often combined. Time series methods and expert judgement result in the distribution of the parameter in question around its expected value. In contrast, an extrapolation of empirical errors gives the distribution centred around zero (assuming an expected error equal to zero), and the expected value of the population variable is taken from a deterministic forecast computed in the traditional manner.

Time series methods are based on the assumption that historical values of the variable of interest have been generated by means of a statistical model, which also holds for the future. A widely used method is that of Autoregressive Integrated Moving Average (ARIMA)-models. These time series models were developed for short horizons. When applied to long-run population forecasting, the point forecast and the prediction intervals may become unrealistic (Sanderson 1995). Judgmental methods (see below) can be applied to correct or constrain such unreasonable predictions (Lee 1993; Tuljapurkar 1996).

Expert judgement can be used when expected values and corresponding prediction intervals are hard to obtain by formal methods. In demographic forecasting, the method has been pioneered by Lutz and colleagues (Lutz et al. 1996; Hanika et al. 1997; Lutz and Scherbov 1998a, b). A group of experts is asked to indicate the probability that a summary parameter, such as the TFR, falls within a certain pre-specified range for some target year, for instance the range determined by the high and the low variant of an independently prepared population forecast. The subjective probability distributions obtained this way from a number of experts are combined in order to reduce individual bias. A major weakness of this approach, at

least based upon the experiences from other disciplines, is that experts often are too confident, *i.e.*, that they tend to attach a too high probability to a given interval (Armstrong 1985). A second problem is that an expert would have problems with sensibly guessing whether a certain interval corresponds to probability bounds with 90% coverage versus 95% or 99% (Lee 1999).

Extrapolation of empirical errors requires observed errors from historical forecasts. Formal or informal methods may be used to predict the errors for the current forecast. Keyfitz (1981) and Stoto (1983) were among the first to use this approach in demographic forecasting. They assessed the accuracy of historical forecasts for population growth rates. The Panel on Population Projections of the US National Research Council (NRC 2000) elaborated further on this idea and developed a statistical model for the uncertainty around total population in UN-forecasts for all countries of the world. Others have investigated and modelled the accuracy of predicted TFR, life expectancy, immigration levels, and age structures (Keilman 1997; De Beer 1997). There are two important problems. First, time series of historical errors are usually rather short, as forecasts prepared in the 1960s or earlier generally were poorly documented. Second, extrapolation is often difficult because errors may have diminished over successive forecast rounds as a result of better forecasting methods.

Irrespective of the method that is used to determine the prediction intervals for all future fertility, mortality and migration parameters, the next step is to apply these to the base population in order to compute prediction intervals for future population size and age pyramids. This can be done in two ways: analytically, and by means of simulation.

The *analytical approach* is based on a stochastic cohort component model, in which the statistical distributions for the fertility, mortality, and migration parameters are transformed into statistical distributions for the size of the population and its age-sex structure. Alho and Spencer (1985) and Cohen (1986) employ such an analytical approach, but they need strong assumptions. Lee and Tuljapurkar (1994) give approximate expressions for the second moments of the distributions.

The *simulation approach* avoids the simplifying assumptions and the approximations of the analytical approach. The idea is to compute several hundreds or thousands of forecast variants (“sample paths”) based on input parameter values for fertility, mortality, and migration that are randomly drawn from their respective distributions, and store the results in a database. Early contributions based on the idea of simulation are those by Keyfitz (1985), Pflaumer (1986, 1988), and Kuijsten (1988).

In order to illustrate that probabilistic forecasts are useful when uncertainty has to be quantified, I shall give an example for the population of Norway. I shall compare the results from a probabilistic forecast with those from a traditional deterministic one, prepared by Statistics Norway.

9.5 Probabilistic Forecasts: An Alternative to Forecast Variants

Technical details of the methods used to construct the probabilistic forecast are presented elsewhere (Keilman et al. 2001, 2002). Here I shall give a brief summary.

ARIMA time series models were estimated for observed annual values of the TFR, the life expectancy for men and women, and total immigration and immigration in Norway since the 1950s. Based on these ARIMA models, repeated stochastic simulation starting in 1996 yielded 5,000 sample paths for each of these summary parameters to the year 2050. The predictive distributions for the TFR and the life expectancy at birth were checked against corresponding empirical distributions based on historical forecasts published by Statistics Norway in the period 1969–1996. The predicted TFR, life expectancy, and gross migration flows were broken down into age specific rates and numbers by applying various model schedules: a Gamma model for age specific fertility, a Heligman-Pollard model for mortality, and a RogersCastro model for migration. Next, the results of the 5000 runs of the cohort component model for the period up to 2050 were assembled in a data base containing the future population of Norway broken down by 1-year age group, sex, forecast year (1997–2050), and forecast run. For each variable of interest, for example the total population in 2030, or the old age dependency ratio (OADR) in 2050, one can construct a histogram based on 5000 simulated values, and read off prediction intervals with any chosen coverage probability.

The results showed odds equal to four against one (80% chance) that Norway's population, now 4.5 million, will number between 4.3 and 5.4 million in the year 2025, and 3.7–6.4 million in 2050. Uncertainty was largest for the youngest and the oldest age groups, because fertility and mortality are hard to predict. As a result, prediction intervals in 2030 for the population younger than 20 years of age were so wide, that the forecast was not very informative. International migration showed large prediction intervals around expected levels, but its impact on the age structure was modest. In 2050, uncertainty had cumulated so strongly, that intervals were very large for virtually all age groups, in particular when the intervals are judged in a relative sense (compared to the median forecast).

Figure 9.2 shows the high and the low bound of the various prediction intervals for the old age dependency ratio, defined as the population 67 and over relative to that aged 20–66.³ The prediction intervals are those with 95%, 80%, and 67% coverage. The median of the predictive distributions is also plotted. The intervals widen rapidly, reflecting that uncertainty increases with time. We see that ageing is certain in Norway, at least until 2040. In that year, the odds are two against one (67% interval) that the OADR will be between 0.33 and 0.43, *i.e.*, at least 10 points higher than today's value of 0.23. The probability of a ratio in 2040 that is lower than today's is close to zero.

³The legal retirement age in Norway is 67.

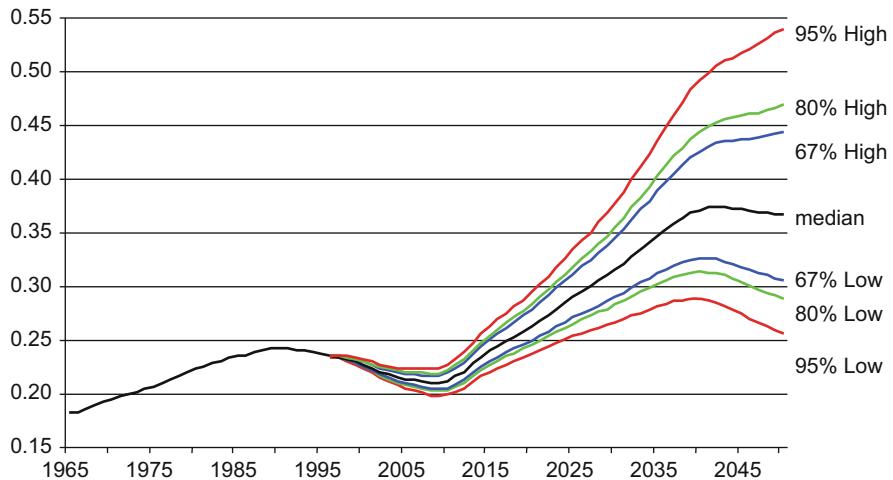


Fig. 9.2 Old age dependency ratio, Norway

How do these probabilistic forecast results compare with those obtained by a traditional deterministic forecast? Statistics Norway's most recent population forecast contains variants for high population growth and low population growth, among others (Statistics Norway 2005). The high population growth forecast results from combining a high fertility assumption with a high life expectancy assumption (*i.e.*, low mortality) and a high net immigration assumption. Likewise, the low growth variant combines low fertility with low life expectancy and low immigration. The forecast predicts a population aged 67 and over in 2050 between 1,095,000 (low growth) and 1,406,000 (high growth). However, the corresponding OADR-values are 0.409 for low population growth, and 0.392 for high population growth. Therefore, while there is a considerable gap between the absolute numbers of elderly in the two variants, the relative numbers, as a proportion of the population aged 20–66, are almost indistinguishable. The interval for the absolute number thus reflects uncertainty in some sense, but the OADR-interval for the same variant pair suggests almost no uncertainty. On the other hand, the probabilistic forecast results in Fig. 9.2 show a two-thirds OADR-prediction interval in 2050 that stretches from 0.31 to 0.44.⁴

This example illustrates that it is problematic to use forecast variants from traditional deterministic forecast methods to express forecast uncertainty. First, uncertainty is not quantified. Second, the use of high and low variants is inconsistent from a statistical point of view (Lee 1999, Alho 1998). In the high variant, fertility is assumed to be high in every year of the forecast period. Similarly, when fertility is

⁴The median OADR-value of the stochastic forecast in 2050 (0.37) is lower than the medium value of Statistics Norway's forecast for that year (0.395). Life expectancy in 2050 rises to 86 years in Statistics Norway's forecast, but only to 82.3 years in the median of the stochastic forecast. The latter forecast was prepared 4 years earlier than Statistics Norway's forecast.

Table 9.1 Prediction intervals for retirement age, Norway

	Median	67% interval	80% interval	95% interval
OADR = 0.24				
2010	64.8	64.5–65.2	64.4–65.3	64.1–65.5
2030	69.2	68.2–70.3	67.9–70.6	67.2–71.3
2050	71.9	69.4–74.4	68.7–75.1	67.0–76.8
OADR = 0.18				
2010	67.6	67.2–68.1	67.1–68.2	66.8–68.5
2030	72.0	71.0–73.0	70.7–73.3	70.1–74.0
2050	75.1	72.8–77.2	72.2–77.9	70.7–79.4

low in one year, it is 100% certain that it will be low in the following years, too. Things are even worse when two or more mortality variants are formulated, in addition to the fertility variants, so that high/low growth variants result from combining high fertility with high life expectancy/low fertility with low life expectancy. In that case, *any* year in which fertility is high, life expectancy is high as well. In other words, one assumes perfect correlation between fertility and mortality, in addition to perfect serial correlation for each of the two components. Assumptions of this kind are unrealistic, and, moreover, they cause inconsistencies: two variants that are extreme for one variable need not be extreme for another variable.

As a further illustration of the use of stochastic population forecasts when analyzing pension systems, let me consider the possibility of a flexible retirement age. When workers postpone retirement, they contribute longer to the pension fund, and the years they benefit from it become shorter (other factors remaining the same). Therefore I analyse the following question: which retirement age is necessary in Norway in the future in order to achieve a constant OADR (see also Sect. 12.4 of the chapter by Tuljapurkar in this volume for a similar analysis for the United States)? I will investigate two cases. First I assume a constant OADR equal to 0.24, which is the highest value observed in the past (around 1990, see Fig. 9.2). Second, I assume an OADR equal to 0.18. This is the value in 1967, the year when the Norwegian pension system in its current form was introduced. Since the future age structure is uncertain, the retirement age necessary to obtain a constant OADR becomes a stochastic variable. Table 9.1 gives the results.

The table shows that the retirement age in Norway must increase strongly from its current value of 67 years, if the OADR were to remain constant at 0.24. The median of 71.9 years in 2050 indicates that the rise is almost 5 years. Yet the uncertainty is large here. In four out of five cases would the retirement age in 2050 be between 69 and 75 years. In the short run the situation is completely different. The age structure of the population of Norway is such that the retirement age can *decrease* to 2010, and yet the ratio of elderly to the population in labour force ages could remain constant. This finding is almost completely certain. Even the upper bound of the 95% interval (65.5) is much lower than today's retirement age.

If one would require an OADR as low as the one in 1967, the median age at retirement has to increase to no less than 75.1 years in 2050. A higher retirement age

is necessary even in the short run: the median in 2010 is 67.6 years, and the lower bound to the 80% prediction interval indicates that the probability that we may can an increase is about 10% or lower, given the assumptions made.

9.6 Challenges in Probabilistic Population Forecasting

A probabilistic forecast extrapolates observed variability in demographic data to the future. For a proper assessment of the variability, one needs long series with annual data of good quality. The minimum is about 50 years, but a longer series is preferable. At the same time, one would ideally have a long series of historical forecasts, and estimate empirical distributions of observed forecast errors based on the old forecasts. There are very few countries that have so good data. Therefore, a major challenge in probabilistic forecasting is to prepare such forecasts for countries with poorer data. Two research directions seem promising. First, when time series analysis cannot be used to compute predictive distributions, one has to rely strongly on expert opinion. Lutz et al. (1996, 2001) have indicated how this can be done in practice. An important task here is a systematic elicitation of the experts' opinions, in order to avoid too narrow prediction intervals. Second, in case the data from historical forecasts are lacking, one could replace actual forecasts by naïve or baseline forecasts (Keyfitz 1981; Alho 1998). Historical forecasts often assumed constant (or nearly constant) levels or growth rates for summary indicators such as the TFR, the life expectancy, or the level of immigration. Thus we can study how accurate past fertility forecasts would have been if they had assumed that the base value had persisted. Similarly, we can compute mortality errors based on an assumption of a linear increase in life expectancy. Such naïve error estimates would be expected to lead to conservative, that is, too large variability estimates, in some cases only slightly so but in others substantially.

Most applications of probabilistic forecasting so far focus on one country. Very few have a regional or an international perspective. One important exception is the work by Lutz et al. (1996, 2001), who used a probabilistic cohort component approach for 13 regions of the world.⁵ For fertility and mortality, they combined the three methods mentioned in Sect. 9.4 to obtain predictive distributions for summary indicators. An important challenge was the probabilistic modelling of interregional migration, because migration data show large volatility in the trends, are unreliable, not consistent between countries, or often simply lacking. In their 1996 study, Lutz and colleagues assumed a matrix of constant annual interregional migration flows, with the 90% prediction bounds corresponding to certain high and

⁵Probabilistic forecasts of total population size for all countries of the world have been prepared by the Panel on Population Projections (NRC 2000), but these forecasts do not give age detail.

low migration gains in each region. In the recent study, net migration into the regions was modelled as a stochastic vector with a certain autocorrelation structure. A second challenge was the treatment of interregional correlations for fertility, mortality, and migration. Due to the paucity of the necessary data, these correlations are difficult to estimate. Therefore, the authors combined qualitative considerations with sensitivity analysis, and investigated alternative regional correlation levels.

Because of these data problems, the development of a sound method for probabilistic multiregional cohort component forecasting is an important research challenge. For sub-national forecasts, the problems are probably easier to overcome than for international forecasts, because the data situation is better in the former case, at least in a number of developed countries. The way ahead would thus be to collect better migration data, and to invest efforts in estimating cross-regional correlation patterns for fertility, mortality, and migration. An alternative strategy could be to start from a probabilistic cohort component forecast for the larger region, and to compute such forecasts at the lower regional level (by age and sex) by means of an appropriate multivariate distribution with expected values corresponding to the regional shares from an independently prepared deterministic forecast.

Not only regional forecasts, but also other types of population forecasts should be couched in probabilistic terms, such as labour market forecasts, educational forecasts, and household forecasts, to name a few. Very few of such probabilistic forecasts have been prepared. Lee and Tuljapurkar (2001) have investigated the expected accuracy of old age security funds forecasts in the United States. A major topic of research here is to analyse the relative contribution to uncertainty of demographic factors (fertility, mortality, migration) and non-demographic factors (labour market participation, educational attainment, residential choices).

References

- Alho, J. M. (1998). *A stochastic forecast of the population of Finland. Reviews 1998/4*. Helsinki: Statistics Finland.
- Alho, J., & Spencer, B. (1985). Uncertain population forecasting. *Journal of the American Statistical Association*, 80, 306–314.
- Armstrong, J. (1985). *Long-range forecasting: From crystal ball to computer* (2nd ed.). New York: Wiley.
- Ascher, W. (1978). *Forecasting: An appraisal for policy-makers and planner*. Baltimore: The Johns Hopkins University Press.
- Bandel Bäckman, J., & Schéele, S. (1995). Uncertainty in population forecasts for small areas. In C. Lundh (Ed.), *Demography, economy and welfare. Lund studies in economic history* (Vol. 1, pp. 419–434). Lund: Lund University Press.
- Boudon, R. (1986). *Theories of social change: A critical appraisal*. Cambridge: Polity Press.
- Calot, G., & Chesnais, J. C. (1978). La prévision démographique. In A.-C. Dans (Ed.), *Découflé Traité élémentaire de prévision et de prospective*. Paris: PUF.

- Cohen, J. (1986). Population forecasts and confidence intervals for Sweden: A comparison of model-based and empirical approaches. *Demography*, 23, 105–126.
- De Beer, J. (1997). The effect of uncertainty of migration on national population forecasts: The case of the Netherlands. *Journal of Official Statistics*, 13, 227–243.
- De Beer, J., & Alders, M. (1999). *Probabilistic population and household forecasts for the Netherlands Working Paper nr. 45*. Joint ECE-Eurostat Work Session on demographic projections, Perugia, Italy, 3–7 May 1999.
- Hanika, A., Lutz, W., & Scherböv, S. (1997). Ein probabilistischer Ansatz zur Bevölkerungsvorausschätzung für Österreich. *Statistische Nachrichten*, 12(1997), 984–988.
- Henry, L. (1987). Perspectives et prévision. In Les projections démographiques. *Actes du VIII^e Colloque National de Démographie* (Tome 1). Paris: Presses Universitaires de France (Traveaux et Documents Cahier no. 116).
- Inoue, S., & Yu, Y. C. (1979). *United Nations new population projections and analysis of ex post facto errors*. Paper annual meeting Population Association of America, Philadelphia, April 1979.
- Keilman, N. (1997). Ex-post errors in official population forecasts in industrialized countries. *Journal of Official Statistics*, 13, 245–277.
- Keilman, N. (1998). How accurate are the United Nations world population projections? *Population and Development Review*, 24(Supplement), 15–41.
- Keilman, N. (2000). Précision et incertitudes des prévisions nationales de population In C. Wattelar et J. Duchêne (red.). *Le défi de l'incertitude: Nouvelles approches en perspectives et prospective démographiques. Actes de la Chaire Quetelet 1995* (pp. 33–67). Louvain-la-Neuve: Academia-Bruylants/L'Harmattan.
- Keilman, N. (2001). Data quality and accuracy of United Nations population projections, 1950–95. *Population Studies*, 55(2), 149–164.
- Keilman, N., & Cruisjen, H. (Eds.). (1992). *National population forecasting in industrialized countries*. Amsterdam etc.: Swets & Zeitlinger.
- Keilman, N., & Pham, D. Q. (2004). *Empirical errors and predicted errors in fertility, mortality and migration forecasts in the European economic area* (Discussion Paper 386). Statistics Norway.
- Keilman, N., Pham, D. Q., & Hetland, A. (2001). *Norway's uncertain demographic future* (Social and Economic Studies 105). Oslo: Statistics Norway. Internet www.ssb.no/english/subjects/02/03/sos105_en
- Keilman, N., Pham, D. Q., & Hetland, A. (2002, May 6–15). Why population forecasts should be probabilistic – Illustrated by the case of Norway. *Demographic Research*, 409–453. Internet www.demographic-research.org/volumes/vol6/15/6-15.pdf
- Keyfitz, N. (1981). The limits of population forecasting. *Population and Development Review*, 7, 579–593.
- Keyfitz, N. (1982). Can knowledge improve forecasts? *Population and Development Review*, 8, 729–751.
- Keyfitz, N. (1985). A probability representation of future population. *Zeitschrift für Bevölkerungswissenschaft*, 11, 179–191.
- Kuijsten, A. (1988). Demografische toekomstbeelden van Nederland (Demographic scenarios for the Netherlands). *Bevolking en Gezin*, 1988(2), 97–130.
- Laplace, Pierre Simon de. (1812–1829). Œuvres Vol. VII, Théorie Analytique des Probabilités.
- Lee, R. (1993). Modeling and forecasting the time series of US fertility: Age distribution, range, and ultimate level. *International Journal of Forecasting*, 9, 187–202.

- Lee, R. (1999). Probabilistic approaches to population forecasting. In W. Lutz, J. Vaupel, & D. Ahlburg (Eds.), *Frontiers of population forecasting* (Supplement to Vol. 24 of Population and development review) (pp. 156–190). New York: Population Council.
- Lee, R., & Tuljapurkar, S. (1994). Stochastic population forecasts for the United States: Beyond high, medium, and low. *Journal of the American Statistical Association*, 89, 1175–1189.
- Lee, R. And S. Tuljapurkar (2001) Population forecasting for fiscal planning: Issues and innovation. pp. 7–57 in A. Auerbach and R. Lee (eds.) Demographic change and fiscal policy. Cambridge: Cambridge University Press.
- Lutz, W., & Scherbov, S. (1998a). An expert-based framework for probabilistic national population projections: The example of Austria. *European Journal of Population*, 14, 1–17.
- Lutz, W., & Scherbov, S. (1998b). Probabilistische Bevölkerungsprognosen für Deutschland (Probabilistic population projections for Germany). *Zeitschrift für Bevölkerungswissenschaft*, 23, 83–109.
- Lutz, W., Sanderson, W., & Scherbov, S. (1996). Probabilistic population projections based on expert opinion. In W. Lutz (Ed.), *The future population of the world: What can we assume today?* (rev. ed., pp. 397–428). London: Earthscan.
- Lutz, W., Sanderson, W., & Scherbov, S. (2001). The end of world population growth. *Nature*, 412, 543–545.
- Nagel, E. (1961). *The structure of science: Problems in the logic of scientific explanations* (Vol. 29, p. 716). New York: Harcourt, Brace and World.
- National Research Council. (2000). Beyond six billion: Forecasting the world's population. Panel on population projections. In J. Bongaarts & R. Bulatao (Eds.), *Committee on population, commission on behavioral and social sciences and education*. Washington, DC: National Academy Press. Internet <http://books.nap.edu/books/0309069904/html/index.html>.
- Pflaumer, P. (1986). Stochastische Bevölkerungsmodelle zur Analyse der Auswirkungen demographischer Prozesse auf die Systeme der sozialen Sicherung. *Allgemeine Statistics Archives*, 70, 52–74.
- Pflaumer, P. (1988). Confidence intervals for population projections based on Monte Carlo methods. *International Journal of Forecasting*, 4, 135–142.
- Preston, S. (1974). An evaluation of postwar mortality projections in Australia, Canada, Japan, New Zealand, and the United States. *WHO Statistical Report*, 27, 719–745.
- Sanderson, W. (1995). Predictability, complexity, and catastrophe in a collapsible model of population, development, and environmental interactions. *Mathematical Population Studies*, 5, 233–279.
- Schéele, S. (1981). *Osäkerhet i befolkningsprognosser*. Stockholm: Stockholms Kommune Utredningsoch Statistikkontor.
- Smith, S. K., Tayman, J., & Swanson, D. (2001). *State and local population projections: Methodology and analysis* (The Plenum series on demographic methods and population analysis). New York: Kluwer Academic/Plenum Publishers.
- Statistics Norway. (2005). *Framskrivning av folkemengden 2002/2050 (Population projections 2002/2050) NOS D 319*. Oslo: Statistics Norway.
- Stoto, M. (1983). The accuracy of population projections. *Journal of the American Statistical Association*, 78, 13–20.
- Tayman, J., Schafer, E., & Carter, L. (1998). The role of population size in the determination and prediction of forecast errors: An evaluation using confidence intervals for subcounty areas. *Population Research and Policy Review*, 17, 1–20.
- Tuljapurkar, S. (1996). *Uncertainty in demographic projections: Methods and meanings*. Paper sixth annual conference on applied and business demography. Bowling Green, 19–21 September 1996.

- United Nations. (2005). *World population prospects: The 2004 revision*. New York: United Nations.
- White, H. R. (1954). Empirical study of the accuracy of selected methods of projecting state populations. *Journal of the American Statistical Association*, 29, 480–498.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 10

Remarks on the Use of Probabilities in Demography and Forecasting



Juha M. Alho

10.1 Introduction

The concept of “probability” is used as a step in life table construction to get the expected number of survivors in a cohort. However, in traditional texts on demographic methods (e.g., Shryock and Siegel 1976), variance in the number of survivors plays no role. Similarly, concepts of estimation, estimation error, and bias are routinely used, but standard error and sampling distribution are not (except in connection with sample surveys). Although statistically satisfactory accounts of the life table theory have existed for a long time (e.g., Chiang 1968; Hoem 1970), a reason for neglecting population level random variability, and statistical estimation error, has been that the populations being studied are so large that random error must be so small as not to matter, in practice.

In contrast, when statistical methods started to become used in population forecasting in the 1970s, 1980s and 1990s, some of the resulting prediction intervals have been criticized as being implausibly wide. This view has not often been expressed in print, but Smith (2001, pp. 70–71) provides an example. Others, especially sociologically minded critics have gone further and argued that due to the nature of social phenomena, the application of probability concepts in general, is inappropriate. On the other hand, demographers coming with an economics background have tended to find probabilistic thinking more palatable.

The purpose of the following remarks is to review some probability models, and show how the apparent contradiction arises. We will see that the basic principles have been known for decades. The basic cause of the difficulties – and disagreements – is that there are several layers of probabilities that can be considered. Consequently, it is essential to be explicit about the details of the model.

J. M. Alho (✉)
University of Joensuu, Joensuu, Finland
e-mail: juha.alho@helsinki.fi

10.2 Binomial and Poisson Models

As emphasized by good introductory texts on statistics (*e.g.*, Freedman et al. 1978, p. 497), the concept of probability can only be made precise in the context of a mathematical model. To understand why one often might ignore other aspects of random variables besides expectation, let us construct a model for the survival of a cohort of size n for 1 year. For each individual $i = 1, \dots, n$, define an indicator variable such that $X_i = 1$, if i dies during the year, and $X_i = 0$ otherwise. The total number of deaths is then $X = X_1 + \dots + X_n$. We assume that the X_i 's are random variables (*i.e.*, their values are determined by a chance experiment). Suppose we make an assumption concerning their expectation

$$E[X_i] = q, i = 1, \dots, n, \quad (10.1)$$

and assume that

$$X_1, \dots, X_n \text{ are independent.} \quad (10.2)$$

It follows that X has a binomial distribution, $X \sim \text{Bin}(n, q)$. As is well known, we have the expectation $E[X] = nq$, and variance $\text{Var}(X) = n(q-q^2)$. Therefore, the coefficient of variation is $C = ((1-q)/nq)^{1/2}$.

Now, in industrialized countries the probability of death is about 1% and population size can be in the millions, so relative variation can, indeed, be small. For example, if $q = 0.01$ and $n = 1000,000$, we have that $C = 0.01$. Or, the relative random variability induced by the model defined in (10.1) and (10.2) is about 1%. Equivalent calculations have already been presented by Pollard (1968), for example.

One might object to the conclusion that relative variability is negligible on the grounds that (10.1) does not hold: surely people of different ages (and of different sex, socio-economic status etc.) have different probabilities of death. Therefore, suppose that

$$E[X_i] = q_i, \text{ with } q = (q_1 + \dots + q_n)/n. \quad (10.3)$$

In this case

$$\text{Var}(X) = nq - \sum_{i=1}^n q_i^2. \quad (10.4)$$

However, it follows from the Cauchy-Schwarz inequality that

$$n \sum_{i=1}^n q_i^2 \geq n^2 q^2. \quad (10.5)$$

Therefore, we have that the variance (10.4) is actually *less* than the binomial variance. The naive argument based on population heterogeneity simply does not hold.

Before proceeding further, let us note that, apart from substantive factors, heterogeneity of the type (10.3) is imposed on demographic data, because vital events are typically classified by age, so individuals contribute different times to the “rectangles” of the Lexis diagram. This is one reason why the basic data are typically collected in terms of rates, and a Poisson assumption is invoked. There is some comfort in the fact that if the assumptions (10.2) and (10.3) are complemented by the following assumptions: suppose the q_i ’s depend on n and as $n \rightarrow \infty$, (i) $nq = \lambda > 0$, and (ii) $\max\{q_1, \dots, q_n\} \rightarrow 0$, then the distribution of X converges to $\text{Po}(\lambda)$ (Feller 1968, p. 282). The Poisson model is of interest, because under that model $E[X] = \lambda$ as before, but $\text{Var}(X) = \lambda > n(q-q^2)$. In other words, the Poisson model has a larger variability than the corresponding binomial models. Quantitatively the difference is small, however, since now $C = \lambda^{-1/2}$. If $n = 1000,000$ and $q = 0.01$, then $\lambda = 10,000$, and $C = 0.01$, for example. Or, the relative variability is the same as that under the homogeneous binomial model, to the degree of accuracy used.

The usual demographic application of the Poisson model proceeds from the further identification $\lambda = \mu K$, where K is the person years lived in the population, and μ is the force of mortality. The validity of this model is not self-evident, since unlike n , K is also random. At least when λ is of a smaller order of magnitude than K , the approximation appears to be good, however (Breslow and Day 1987, pp. 132–133). As is well-known, the maximum likelihood estimator of the force of mortality is $\hat{\mu} = X/K$ with the estimated standard error of $X^{1/2}/K$. Extensions to log-linear models that allow for the incorporation of explanatory variables follow similarly.

Since the Poisson distribution has the variance maximizing property, and it provides a model for both the independent trials and occurrence/exposure rates, we will below restrict the discussion primarily to the Poisson case.

10.3 Random Rates

Since (10.1) is not the cause of the low level of variability in the number of deaths, we need to look more closely at (10.2). A simple (but unrealistic) example showing that there are many opportunities here is the following. Suppose we make a single random experiment with probability of success = q , and probability of failure = $1-q$. If the experiment succeeds, define $X_i = 1$ for all i . Otherwise define $X_i = 0$ for all i . In this case we have, for example, that $X = nX_1$, so $E[X] = nq$ as before, but $\text{Var}(X) = n^2(q-q^2)$, and $C = ((1-q)/q)^{1/2}$ independently of n . For $q = 0.01$ we have $C = 9.95$, for example, indicating a huge (nearly 1000%) level of variability.

More realistically, we may think that dependence across individuals arises because they may all be influenced by common factors to some extent, at least. For example, there may be year to year variation in mortality around a mean that is

due to irregular trends in economics, epidemics, weather etc. If the interest would center on a given year, the model might still be $X \sim Po(\mu K)$, but if several years are considered jointly, then the year-to-year variation due to such factors would have to be considered. In this case, we would entertain a hierarchical model of the type

$$X \sim Po(\mu K) \text{ with } E[\mu] = \mu_0, \text{Var}(\mu) = \sigma^2. \quad (10.6)$$

In other words, the rate μ itself is being considered random, with a mean μ_0 that reflects the average level of mortality over the (relatively short) period of interest, and variance σ^2 that describes the year to year variation.

In this case we have that

$$\begin{aligned} \text{Var}(X) &= E[\text{Var}(X|\mu)] + \text{Var}(E[X|\mu]) \\ &= \mu_0 K + \sigma^2 K^2. \end{aligned} \quad (10.7)$$

It follows that $\text{Var}(\hat{\mu}) = \sigma^2 + \mu_0/K$. This result is of fundamental interest in demography, because if K is large, then the dominant part of the error is due to the annual variability. If the interest centers (as in the production of official population statistics) on a given year, with no regard to other years, we would be left with the pure Poisson variance μ_0/K , which is often small. An exception is the oldest-old mortality, where Poisson variation is always an issue, because for ages high enough K will always be small and μ_0 large.

However, when the interest centers on the time trends of mortality, and eventually on forecasting its future values, then the year to year variation σ^2 must be considered. Under model (10.6) this is independent of population size K . This is a realistic first approximation, but we note that model (10.6) does not take into account the possibility that a population might consist of relatively independent subpopulations. In that case, populations having many such subpopulations would have a smaller variance than a population with no independent subpopulations.

10.4 Handling of Trends

Consider now two counts. Or assume that for $i = 1, 2$, we have that

$$X_i \sim Po(\mu_i K_i) \text{ with } E[\mu_i] = \mu_{0i}, \text{Var}(\mu_i) = \sigma_i^2, \text{Corr}(\mu_1, \mu_2) = \rho. \quad (10.8)$$

Repeating the argument leading to (10.7) for covariances yields the result

$$\text{Corr}(X_1, X_2) = \rho / \{(1 + \mu_{01}/\sigma_1 K_1)(1 + \mu_{02}/\sigma_2 K_2)\}^{1/2}. \quad (10.9)$$

Or, the effect of Poisson variability is to decrease the correlation between the observed rates. We note that if the K_i 's are large, the attenuation is small. However, for the oldest old the K_i 's are eventually small, and the μ_{0i} 's large, so attenuation is expected.

In concentrating on Poisson variability that is primarily of interest in the assessment of the accuracy of vital registration, demographers have viewed annual variation as something to be explained. Annual changes in mortality and fertility are analyzed by decomposing the population into ever finer subgroups in an effort to try to find out, which are the groups most responsible for the observed change. Often partial explanations can be found in this manner, but they rarely provide a basis for anticipating future changes (Keyfitz 1982). To be of value in forecasting, an explanation must have certain robustness against idiosyncratic conditions, and persistence over time. This leads to considering changes around a trend as random.

One cause for why some demographers find statistical analyses of demographic time-series irritating seems to lie here: what a demographer views as a phenomenon of considerable analytical interest, may seem to a statistician as a mere random phenomenon, sufficiently described once σ^2 is known. [This tension has counterparts in many parts of science. Linguists, for example, differ in whether they study the fine details of specific dialects, or whether they try to see general patterns underlying many languages.]

In forecasting, the situation is more complex than outlined so far. In mortality forecasting one would typically be interested taking account of the nearly universal decline in mortality, by making further assumptions about time trends. For example, suppose the count at time t is of the form $X_t \sim Po(\mu_t K_t)$, such that

$$\log(\mu_t) = \alpha + \beta t + \xi_t, \text{ where } E[\xi_t] = 0, \text{ Cov}(\xi_t, \xi_s) = \sigma^2 \min\{t, s\}. \quad (10.10)$$

If the ξ_t 's have normal distributions, under (10.10) we would have that $E[\mu_t] = \exp(\alpha + \beta t + \sigma^2 t/2) \equiv \mu_{0t}$ (This model is closely related to the so-called Lee-Carter model.)

One reason that makes (10.10) more complicated than (10.8), is that μ_{0t} involves parameters to be estimated, so standard errors become an issue. Especially, if $\text{Var}(\hat{\beta})$ is large, this source of error may have a considerable effect for large t , because it induces a quadratic term into the variance of error, whereas the effect of the random walk via σ^2 is only linear.

The way $\text{Var}(\hat{\beta})$ is usually estimated from past data assumes that the model specified in (10.10) is correct. Therefore, probabilistic analyses based on (10.10) are conditional on the chosen model. What these probabilities do not formally cover is the uncertainty in model choice itself (Chatfield 1996).

One should pay attention to model choice because it is typically based on iteration, in which lack of fit is balanced against parametric parsimony (*cf.*, Box and Jenkins 1976). One would expect error estimates calculated after a selection process to be too small, because of potential overfitting. Yet, a curious empirical fact seems to be that statistical time-series models identified and estimated in this

manner, for example demographic time-series, often produce prediction intervals that are rather wide, and even implausibly wide in the sense that in a matter of decades they may include values that are thought to be biologically implausible.

A possible explanation is that the standard time-series models belong to simple classes of models (e.g., (10.10) can be seen as belonging to models with polynomial trends with once integrated, or I (10.1), errors) and the identification procedures used are tilted in favor of simple models within those classes. This shows that although judgment is certainly exercised in model choice, it can be exercised in a relatively open manner that tends to produce models that are too simple rather than too complex. When such models are estimated from the data, part of the lack of fit is due to modeling error. Therefore, the estimated models can actually incorporate some aspects of modeling error.

Modeling error can sometimes be handled probabilistically by considering alternative models within a larger class of models, and by weighting the results according to the credibility of the alternatives (e.g., Draper 1995). Alho and Spencer (1985) discuss some minimax type alternatives in a demographic context. A simpler approach is to use models that are not optimized to provide the best possible fit obtainable. In that case the residual error may capture some of the modeling error, as well.

10.5 On Judgment and Subjectivity in Statistical Modeling

“One cannot be slightly pregnant”. In analogy, it is sometimes inferred from this dictum that if judgment is exercised in some part of a forecasting exercise, then all probabilistic aspects of the forecast are necessarily judgmental in nature. In addition, since judgment always involves subjective elements, then the probabilities are also purely subjective. I believe these analogies are misleading in that they fail to appreciate the many layers of probabilities one must consider.

First, the assumption of binomial or Poisson type randomness is the basis of grouped mortality analyses, and as such implicitly shared by essentially all demographers. It takes some talent to see how such randomness could be viewed as subjective.

Second, although models of random rates are not currently used in descriptive demography, they are implicit in all analyses of trends in mortality. Such analyses use smooth models for trends, and deviations from trends are viewed as random. The validity of alternative models can be tested against empirical data and subjective preferences have relatively little role.

On the other hand, models used in forecasting are different in that they are thought to hold in the future, as well as in the past. Yet, they can only be tested against the past. However, even here, there are different grades. In short term forecasting (say, 1–5 years ahead), we have plenty of empirical data on the

performance of the competing models in forecasting. Hence, there is an empirical and fairly formal basis for the choice of models. In medium term forecasting (say, 10–20 years ahead), empirical data are much more scant, and alternative data sets produce conflicting results of forecast performance. Judgment becomes an important ingredient in forecasting. In long-term forecasting (say, 30+ years ahead), the probabilities calculated based on any statistical model begin to be dominated by the possibility of modeling error and beliefs concerning new factors whose influence has not manifested itself in the past data. Judgment, and subjective elements that cannot be empirically checked, get an important role. Note that the binomial/Poisson variability, and the annual variability of the rates, still exist, but they have become dominated by other uncertainties.

In short, instead of viewing probabilities in forecasting as a black and white subjective/objective dichotomy suggested by the “pregnancy dictum”, we have a gradation of shades of gray.

10.6 On the Interpretation of Probabilities

A remaining issue is how one might interpret the probabilities of various types. Philosophically, the problem has been much studied (*e.g.*, Kyburg 1970). It is well-known that the so-called frequency interpretation of probabilities is not a logically sound basis for defining the concept of probability. (For example, laws of large numbers presume the existence of the concept of probability for their statement and proof.) However, it does provide a useful interpretation that serves as a basis of the empirical validation of statistical models we have discussed from binomial/Poisson variation to short and even medium term forecasting. For long term forecasting it is less useful, since we are not interested in what might happen if the history were to be repeated probabilistically again and again. We only experience one sample path.

It is equally well-known that there is a logically coherent theory of subjective probabilities that underlies much of the Bayesian approach to statistics. This theory is rather more subtle than is often appreciated. As discussed by Savage (1954), for example, the theory is prescriptive in the sense that a completely rational actor would behave according to its rules. Since mere mortals are rarely, if ever, completely rational, the representation of actual beliefs in terms of subjective probabilities is a non-trivial task.

For example, actual humans rarely contemplate all possible events that might logically occur. If a person is asked about three events, he might first say that A is three times as likely as B, and B is five times as likely as C; but later say that A is ten times as likely as C. Of course, when confronted with the intransitivity of the answers, he could correct them in any number of ways, but it is not clear that the likelihood of any given event would, after adjustment, be more trustworthy than before.

Actual humans are also much less precise as “computing machines” than the idealized rational actors. Suppose, for example, that a person P says that his uncertainty about the life expectancy in Sweden in the year 2050 can be represented by a normal distribution $N(100, 8^2)$. One can then imagine the following dialogue with a questioner Q:

- Q: So you think the probability that life expectancy exceeds 110 years is over 10%?
- P: I don't know if I think that. If you say so.
- Q: Why can't you say for sure?
- P: Because I can't recall the 0.9 quantile of the standard normal distribution right now.

(Upon learning that it is 1.2816, and calculating $100 + 1.2816*8 = 110.25$ P then agrees.)

Both difficulties suggest that any “subjective” probability statements need to be understood in an idealized sense. To be taken seriously, a person can hardly claim that he or she “feels” that some probabilities apply. Instead, careful argumentation is needed, if one were to want to persuade others to share the same probabilistic characterization (Why a mean of 100 years? Why a standard deviation of 8 years? Why a normal distributional form?).

10.7 Eliciting Expert Views on Uncertainty

Particular problems in the elicitation of probabilistic statements from “experts” are caused by the very fact that an expert is a person who should know how things are.

First, representing one's uncertainty truthfully may be tantamount to saying that one does not really know, if what he or she is saying is accurate. A client paying consulting fees may then deduce that the person is not really an expert! Thus, there is an incentive for the expert to downplay his or her uncertainty.

Second, experts typically share a common information basis, so presenting views that run counter to what other experts say, may label the expert as an eccentric, whose views cannot be trusted. This leads to expert flocking: an expert does not want to present a view that is far from what his or her colleagues say. An example (pointed out by Nico Keilman) is Henshels (1982, p. 71) assessment of the U.S. population forecasts of the 1930s and 1940s. The forecasts came too high because according to Henshel the experts talked too much to each other. Therefore, a consideration of the range of expert opinions may not give a reasonable estimate of the uncertainty of expert judgment.

Economic forecasting provides continuing evidence of both phenomena. First, one only has to think of stock market analysts making erroneous predictions with great confidence on prime time TV, week after week. Second, one can think of think-tanks making forecasts of the GDP. Often (as in the case of Finland in 2001), all tend to err on the same side.

Of course, an expert can also learn to exaggerate uncertainty, should it become professionally acceptable. However, although exaggeration is less serious than the underestimation of uncertainty, it is not harmless either, since it may discredit more reasonable views.

A third, and much more practical problem in the elicitation of probabilities from experts stems from the issues discussed in Sect. 10.7. It is difficult, even for a trained person, to express one's views with the mathematical precision that is needed. One approach that is commonly used is to translate probabilities into betting language. (These concepts are commonly used in the Anglo-Saxon world, but less so in the Nordic countries, for example.) If a player thinks that the probability is at least p that a certain event A happens, then it would be rational to accept a $p:(1-p)$ bet that A happens. (*I.e.*, if A does not happen, the player must pay p , but if it does happen, he or she will receive $1-p$. If the player thinks the true probability of A occurring is $\rho \geq p$, then the expected outcome of the game is $p(1-p) - (1-p)p = \rho - p \geq 0$.)

This approach has two problems, when applied in the elicitation of probabilities from experts. First, it is sometimes difficult to convince the experts to take the notion of a gamble seriously when they know that the “game” does not really take place. Even if the experts are given actual money with which to gamble, the amount may have an effect on the outcome. The second problem, in its standard form, the gambling argument assumes that the players are risk neutral. This may only be approximately true if the sums involved are small. If the sums are large, people tend to be more risk adverse (Arrow 1971). Moreover, experimental evidence suggests (*e.g.*, Kahneman et al. 1982) that people frequently violate the principle of maximizing expected utility.

The betting approach has been used in Finland to elicit expert views on migration (Alho 1998). In an effort to anchor the elicitation on something empirical, a preliminary time series model was estimated, and the experts were asked about the probability content of the model based prediction intervals for future migration. Experts had previously emphasized the essential unpredictability of migration, but seeing the intervals they felt that future migration is not as uncertain as indicated. The intervals were then narrowed down using the betting argument. In this case the use of an empirical bench mark may have lead to a higher level of uncertainty than would otherwise have been obtained.

References

- Alho, J. M. (1998). A stochastic forecast of the population of Finland. *Reviews 1998/4. Statistics Finland*.
- Alho, J. M., & Spencer, B. D. (1985). Uncertain population forecasting. *Journal of the American Statistical Association*, 80, 306–314.
- Arrow, K. J. (1971). *Essays in the theory of risk-bearing*. Chicago: Markham Publishing Co.
- Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis* (Rev ed.). San Francisco: Holden-Day.

- Breslow, N. E., & Day, N. E. (1987). *Statistical methods in cancer research* (Vol. II). Lyon: I.A.R.C.
- Chatfield, C. (1996). *The analysis of time series*. London: Chapman and Hall.
- Chiang, C. L. (1968). *Introduction to stochastic processes in biostatistics*. New York: Wiley.
- Draper, D. (1995). Assessment and propagation of model uncertainty (with discussion). *Journal of the Royal Statistical Society. Series B*, 57, 45–97.
- Feller, W. (1968). *An introduction to probability and its applications* (Vol. I, 3rd ed.). New York: Wiley.
- Freedman, D., Pisani, R., & Purves, R. (1978). *Statistics*. New York: Norton.
- Henshel, R. L. (1982). Sociology and social forecasting. *Annual Review of Sociology*, 8, 57–79.
- Hoem, J. (1970). *Grunnbegreper i formell befolkningslære*. Oslo: Universitetsforlaget.
- Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- Keyfitz, N. (1982). Can knowledge improve forecasts? *Population and Development Review*, 8, 719–751.
- Kyburg, H. E. (1970). *Probability and inductive logic*. London: Macmillan.
- Lee, R. D., & Auerbach, A. J. (2001). *Demographic change and fiscal policy*. Cambridge: Cambridge University Press.
- Pollard, J. H. (1968). A note on multi type Galton-Watson processes with random branching probabilities. *Biometrika*, 55, 589–590.
- Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley.
- Shryock, H. S., & Siegel, J. S. (1976). *The methods and materials of demography*. New York: Academic.
- Smith, J. P. (2001). Comment, in Lee and Auerbach, pp. 69–72.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 11

An Expert Knowledge Approach to Stochastic Mortality Forecasting in the Netherlands



Maarten Alders and Joop de Beer

11.1 Introduction

The Dutch population forecasts published by Statistics Netherlands every other year project the future size and age structure of the population of the Netherlands up to 2050. The forecasts are based on assumptions about future changes in fertility, mortality, and international migration. Obviously, the validity of assumptions on changes in the long run is uncertain, even if the assumptions are expected to describe the expected future according to the forecasters. It is important that users of forecasts are aware of the degree of uncertainty. In order to give accurate information about the degree of uncertainty of population forecasts Statistics Netherlands produces stochastic population forecasts. Instead of publishing two alternative deterministic (low and high) variants in addition to the medium variant, as was the practice up to a few years ago, forecast intervals are made. These intervals are calculated by means of Monte Carlo simulations. The simulations are based on assumptions about the probability distributions of future fertility, mortality, and international migration.

In the Dutch population forecasts the assumptions on the expected future changes in mortality primarily relate to life expectancy at birth. In the most recent Dutch forecasts assumptions underlying the medium variant¹ are based on a quantitative model projecting life expectancy at birth of men and women for the period

¹The description ‘medium variant’ originates from the former practice when several deterministic variants were published. Since no variants are published anymore it does not seem appropriate to speak of a medium variant anymore. However, abandoning this terminology would make users think that the medium variant is different from the expected value. For this reason we will still use ‘medium variant’ while we mean the expected value.

M. Alders · J. de Beer (✉)

Department of Statistical Analysis, Statistics Netherlands, The Hague, Netherlands
e-mail: beer@nidi.nl

2001–2050. The model describes the trend of life expectancy in the period 1900–2000 taking into account the effect of changes in smoking behaviour, the effect of the rectangularization of the survival curve and the effect of some other factors on changes in life expectancy at birth. Since the model is deterministic, it cannot be used directly for making stochastic forecasts. For that reason, the assumptions underlying the stochastic forecasts are based on expert judgement, taking into account the factors described by the model.

This paper examines how assumptions on the uncertainty of future changes in mortality in the long run can be specified. More precisely, it discusses the use of expert knowledge for the specification of the uncertainty of future mortality. Section 11.2 briefly describes the methodology underlying the Dutch stochastic population forecasts. Section 11.3 provides a general discussion on the use of expert knowledge in (stochastic) mortality forecasting. Section 11.4 applies the use of expert knowledge to the Dutch stochastic mortality forecasts. The paper ends with the main conclusions.

11.2 Stochastic Population Forecasts: Methodology

Population forecasts are based on assumptions about future changes in fertility, mortality, and migration. In the Dutch population forecasts assumptions on fertility refer to age-specific rates distinguished by parity, mortality assumptions refer to age- and sex-specific mortality rates, assumptions about immigration refer to absolute numbers, distinguished by age, sex and country of birth, and assumptions on emigration are based on a distinction of emigration rates by age, sex and country of birth.

Based on statistical models of fertility, mortality, and migration, statistical forecast intervals of population size and age structure can be derived, either analytically or by means of simulations. In order to obtain a forecast interval for the age structure of a population analytically a stochastic cohort-component model is needed. Application of such models, however, is very complicated. Analytical solutions require a large number of simplifying assumptions. Examples of applications of such models are given by Cohen (1986) and Alho and Spencer (1985). In both papers assumptions are specified of which the empirical basis is questionable.

Instead of an analytical solution, forecast intervals can be derived from simulations. On the basis of an assessment of the probability of the bandwidth of future values of fertility, mortality, and migration, the probability distribution of the future population size and age structure can be calculated by means of Monte Carlo simulations. For each year in the forecast period values of the total fertility rate, life expectancy at birth of men and women, numbers of immigrants and emigration rates are drawn from the probability distributions. Subsequently age-gender-specific fertility, mortality and emigration rates, and immigration numbers are specified. Each draw results in a population by age and gender at the end of each year. Thus the simulations provide a distribution of the population by age and gender in each forecast year.

To perform the simulations several assumptions have to be made. First, the type of probability distribution has to be specified. Subsequently, assumptions about the parameter values have to be made. The assumption about the mean or median value can be derived from the medium variant. Next, assumptions about the value of the standard deviation have to be assessed. In the case of asymmetric probability distributions additional parameters have to be specified. Finally, assumptions about the covariances between the forecast errors across age, between the forecast years, and between the components have to be specified (see *e.g.*, Lee 1998).

The main assumptions underlying the probability distribution of the future population relate to the variance of the distributions of future fertility, mortality, and migration. The values of the variance can be assessed in three ways:

- (a) an analysis of errors of past forecasts may provide an indication of the size of the variance of the errors of new forecasts;
- (b) estimates of the variance can be based on a statistical (time-series) model;
- (c) on the basis of expert judgement values of the variance can be chosen.

These methods do not exclude each other; rather they may complement each other. For example, even if the estimate of the variance is based on past errors or on a time-series model judgement plays an important role. However, in publications the role of judgement is not always made explicit.

11.2.1 An Analysis of Errors of Past Forecasts

The probability of a forecast interval can be assessed on the basis of a comparison with the errors of forecasts published in the past. On the assumption that the errors are approximately normally distributed – or can be modelled by some other distribution – and that the future distribution of the errors is the same as the past distribution, these errors can be used to calculate the probability of forecast intervals of new forecasts. Keilman (1990) examines the errors of forecasts of fertility, mortality, and migration of Dutch population forecasts published between 1950 and 1980. He finds considerable differences between the errors of the three components. For example, errors in life expectancy grow considerably more slowly than errors in the total fertility rate. Furthermore, he examines to what extent errors vary between periods and whether errors of recent forecasts are smaller than those of older forecasts, taking into account the effect of differences in the length of the forecast period.

The question whether forecast accuracy has been increasing is important for assessing to what extent errors of past forecasts give an indication of the degree of uncertainty of new forecasts. One problem in comparing old and new forecasts is that some periods are easier to forecast than others. Moreover, a method that performs well in a specific period may lead to poor results in another period. Thus one should be careful in drawing general conclusions on the size of forecast errors on the basis of errors in a given period. For example, since life expectancy at birth of

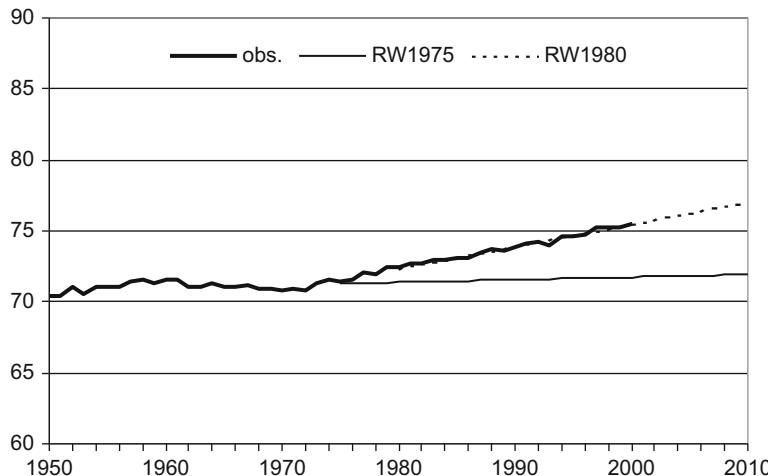


Fig. 11.1 Life expectancy at birth, men, projections of random walk (RW) with drift

men in the Netherlands has been increasing linearly since the early 1970s, a simple projection based on a random walk model with drift would have produced rather accurate forecasts. Figure 11.1 shows that a forecast that would have been made in 1980 on the basis of a random walk model in which the intercept is estimated by the average change in the preceding 10 years would have been very accurate for the period 1980–2000. However, this does not necessarily imply that forecasts of life expectancy are very certain in the long run. If the same method would have been used for a forecast starting in 1975 the forecast would have been rather poor (Fig. 11.1). Thus simply comparing the forecast errors of successive forecasts does not tell us whether recent forecasts are ‘really’ better than preceding forecasts.

The fact that a forecast of life expectancy of men made in 1980 is more accurate than a forecast made in 1975 does not imply that recent forecasts are more accurate than older forecasts. This is illustrated by Fig. 11.2 which shows projections of life expectancy of women. Projections of the life expectancy at birth of women based on a random walk model starting in 1980 are less accurate than projections starting in 1975.

In order to be able to assess whether new forecasts are ‘really’ better than older ones, we need to know the reasons why the forecaster chose a specific method for a certain forecast period. This information enables us to conclude whether a certain forecast was accurate, because the forecaster chose the right method for the right period, or whether the forecaster was just more lucky in one period than in another.

Thus, in order to assess whether errors of past forecasts provide useful information about the uncertainty of new forecasts, it is important not only to measure the size of the errors but also to take into account the explanation of the errors. One main explanation of the poor development of life expectancy of men in the 1960s is the increase in smoking in previous decades, whereas the increase in life expectancy in subsequent years can partly be explained by a decrease in smoking. As women

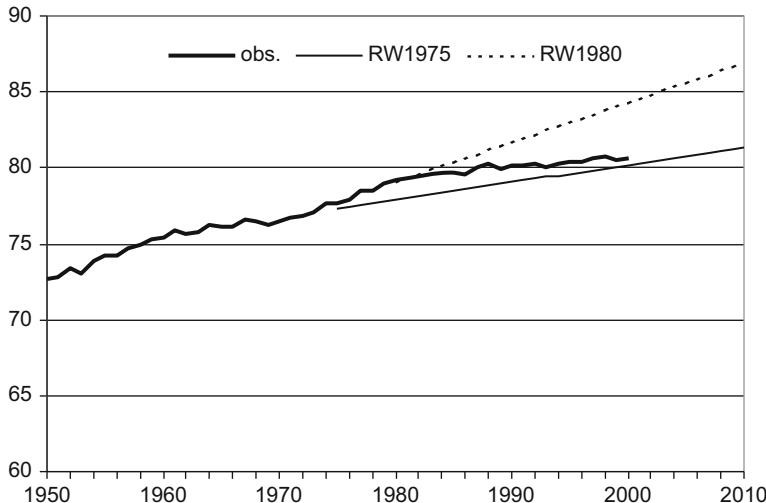


Fig. 11.2 Life expectancy at birth, women, projections of random walk (RW) with drift

started to smoke some decades later than men, the development of life expectancy of women was affected negatively not until the 1980s. This explains why a linear extrapolation of the trend in life expectancy of women starting in 1980 leads to overestimating the increase in the 1980s and 1990s. On the other hand a linear extrapolation of the trend in life expectancy of men starting in 1975 leads to underestimating the increase in subsequent years.

The question to what extent an analysis of past errors provides useful information about the degree of uncertainty of new forecasts depends on the question how likely it is that similar developments will occur again. The 1970- based forecasts were rather poor because forecasters did not recognize that the negative development was temporary (Fig. 11.3). If it is assumed that it is very unlikely that such developments will occur again, one may conclude that errors of new forecasts are likely to be smaller than the errors of the 1970- based forecasts. For that reason, the degree of uncertainty of new forecasts can be based on errors of forecasts that were made after 1970.

The decision which past forecasts to include is a matter of judgement. Thus, judgement plays a role in using errors of past forecasts for assessing the uncertainty of new forecasts. Obviously one may argue that an ‘objective’ method would be to include all forecasts made in the past. However, this implies that the results depend on the number of forecasts that were made in different periods. Since more forecasts were made after 1985 than in earlier periods, the errors of more recent forecasts weigh more heavily in calculating the average size of errors. On the other hand, for long-run forecasts one major problem in using errors of past forecasts for assessing the degree of uncertainty of new forecasts is that the sample of past forecasts tends to

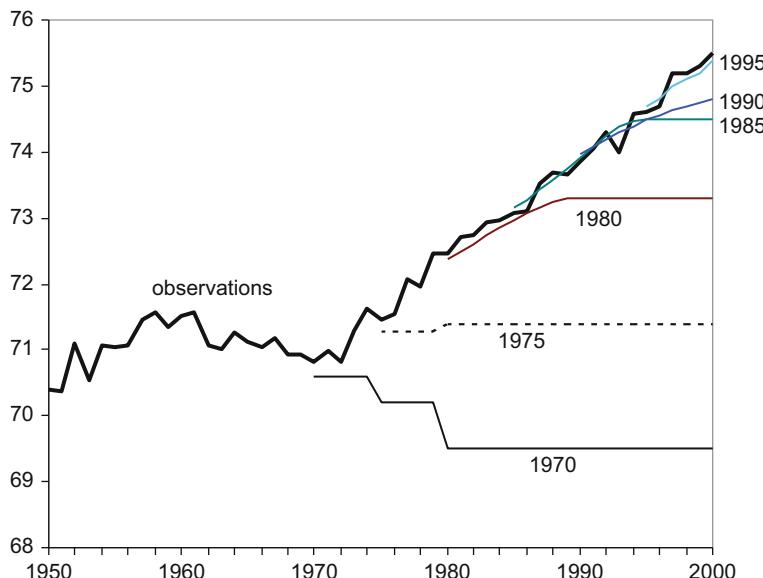


Fig. 11.3 Life expectancy at birth, men, observations and historic forecasts

be biased towards the older ones, as for recent forecasts the accuracy cannot yet be checked except for the short run (Lutz et al. 1996a). Forecast errors for the very long run result from forecasts made a long time ago. Figure 11.3 shows that the 30 and 25 years ahead forecasts made in 1970 and 1975 respectively are rather poor, so including these forecasts in assessing the uncertainty of new forecasts may lead to overestimating the uncertainty in the long run. Figure 11.3 suggests that the forecasts made in the 1980s and 1990s may well lead to smaller errors in the long run than the forecasts made in the 1970s, since the former forecasts are closer to the observations up to now than the latter forecasts were at the same forecast interval.

One way of assessing forecast errors in the long run is to extrapolate forecast errors by means of a time-series model (De Beer 1997). The size of forecast errors for the long run can be projected on the basis of forecast errors of recent forecasts for the short and medium run. Thus, estimates of ex ante forecast errors can be based on an extrapolation of ex post errors.

Rather than calculating errors in forecasts that were actually published, empirical forecast errors can be assessed by means of calculating the forecast errors of simple baseline projections. Alho (1998) notes that the point forecasts of the official Finnish population forecasts are similar to projections of simple baseline projections, such as assuming a constant rate of change of age-specific mortality rates. If these baseline projections are applied to past observations, forecast errors can be calculated. The relationship between these forecast errors and the length of the forecast period can be used to assess forecast intervals for new forecasts.

11.2.2 Model-Based Estimate of Forecast Errors

Instead of assuming that future forecast errors will be similar to errors of past forecasts, one may attempt to estimate the size of future forecast errors on the basis of the assumptions underlying the methods used in making new forecasts. If the forecasts are based on an extrapolation of observed trends, ex ante forecast uncertainty can be assessed on the basis of the time-series model used for producing the extrapolations. If the forecasts are based on a stochastic time-series model, the model produces not only the point forecast, but also the probability distribution. For example, ARIMA (Autoregressive Integrated Moving Average)-models are stochastic univariate time-series models that can be used for calculating the probability distribution of a forecast (Box and Jenkins 1970). Alternatively, a structural time-series model can be used for this purpose (Harvey 1989). The latter model is based on a Bayesian approach: the probability distribution may change as new observations become available. The Kalman filter is used for updating the estimates of the parameters.

One problem in using stochastic models for assessing the probability of a forecast is that the probability depends on the assumption that the model is correct. Obviously, the validity of this assumption is uncertain, particularly in the long run. If the point forecast of the time-series model does not correspond with the medium variant, the forecaster does apparently not regard the time-series model as correct. Moreover, time-series forecasting models were developed for short horizons, and they are not generally suitable for long run forecasts (Lee 1998). Usually, stochastic time-series models are identified on the basis of autocorrelations for short time intervals only. Alternatively, the form of the time-series model can be based on judgement to constraint the long-run behaviour of the point forecasts such that they are in line with the medium variant of the official forecast (Tuljapurkar 1996). However, one should be careful in using such a model for calculating the variance of ex ante forecast errors, because of the uncertainty of the validity of the constraint imposed on the model. In assessing the degree of uncertainty of the projections of the model one should take into account the uncertainty of the constraint, which is based on judgement.

11.2.3 Expert Judgement

In assessing the probability of forecast intervals on the basis of either an analysis of errors of past forecasts or an estimate of the size of model-based errors, it is assumed that the future will be like the past. Instead, the probability of forecasts can be assessed on the basis of experts' opinions about the possibility of events that have not yet occurred. For example, the uncertainty of long-term forecasts of mortality

depends on the probability of technological breakthroughs that may have a substantial impact on survival rates. Even though these developments may not be assumed to occur in the expected variant, an assessment of the probability of such events is needed to determine the uncertainty of the forecast. More generally, an assessment of ex ante uncertainty requires assumptions about the probability that the future will be different from the past. If a forecast is based on an extrapolation of past trends, the assessment of the probability of structural changes which may cause a reversal of trends cannot be derived directly from an analysis of historical data and therefore requires judgement of the forecaster. With regard to mortality the assessment of the probability of unprecedented events like medical breakthroughs cannot be derived directly from models. Lutz et al. (1996b) assess the probability of forecasts on the basis of opinions of a group of experts. The experts are asked to indicate the upper and lower boundaries of 90% forecast intervals for the total fertility rate, life expectancy, and net migration up to the year 2030. Subjective probability distributions of a number of experts are combined in order to diminish the danger of individual bias.

In the Dutch population forecasts the assessment of the degree of uncertainty of mortality forecasts is primarily based on expert judgement, taking into account errors of past forecasts and model-based estimates of the forecast errors.

11.3 Using Expert Knowledge

Expert knowledge or judgement usually plays a significant role in population forecasting. The choice of the model explaining or describing past developments cannot be made on purely objective, *e.g.*, statistical criteria. Moreover, the application of a model requires assumptions about the way parameters and explanatory variables may change. Thus, forecasts of the future cannot be derived unambiguously from observations of the past. Judgement plays a decisive role in both the choice of the method and the way it is applied. “There can never be a population projection without personal judgement. Even models largely based on past time-series are subject to a serious judgemental issue of whether to assume structural continuity or any alternative structure” (Lutz et al. 1996a).

Forecasts of mortality can be based on extrapolation of trends in mortality indicators or on an explanatory approach. In both cases forecasters have to make a number of choices. In projecting future changes in mortality on the basis of an extrapolation of trends, one important question is which indicator is to be projected. If age- and gender-specific mortality rates are projected one may choose to assume the same change for each age (ignoring changes in the age pattern of mortality) or one may project each age-specific rate separately (which may result in a rather irregular age pattern). Instead of projecting separate age-specific mortality rates one may project a limited number of parameters of a function describing the age pattern of mortality, *e.g.*, the Gompertz curve or the Heligman-Pollard model. One disadvantage of the Heligman-Pollard model is that it includes many parameters that

cannot be projected separately. This makes the projection process complex. On the other hand, the disadvantage of using a simple model with few parameters is that such models usually do not describe the complete age pattern accurately. Another possible forecasting procedure is based on a distinction of age, period and cohort effects. If cohort effects can be estimated accurately, such models may be appropriate for making forecasts for the long run. However, one main problem in using an APC model is that the distinction between cohort effects and the interaction of period and age effects for young cohorts is difficult. Finally the indicator most widely used in mortality forecasting is life expectancy at various ages, especially life expectancy at birth. Using life expectancy at birth additional assumptions have to be made about changes in the age pattern of mortality rates.

In addition to the choice of the indicator to be projected, other choices have to be made. One main question is which observation period should be the basis for the projections. An extrapolation of changes observed in the last 20 or 30 years may result in quite different projections than an extrapolation of changes in the last 50 or more years. Another important question is the choice of the extrapolation procedure: linear or non-linear. This question is difficult to be answered on empirical grounds: different mathematical functions may describe observed developments about equally well, but may lead to quite different projections in the long run. In summary, judgement plays an important role in extrapolations of mortality.

Instead of an extrapolation of trends forecasts of mortality may be based on an explanatory approach. In making population forecasts usually a qualitative approach is followed. On the basis of an overview of the main determinants of mortality (*e.g.*, changes in living conditions, life style, health care, safety measures, etc.) and of assumptions about both the impact of these determinants on the development of mortality and future changes in the determinants, it is concluded in which direction mortality may change. Clearly, if no quantitative model is specified, the assumptions about the future change in mortality are largely based on judgement. However, even if a quantitative model would be available, judgement would still play an important role, since assumptions would have to be made about the future development in explanatory variables.

In most developed countries life expectancy has been rising during a long period. Therefore, in assessing the uncertainty of forecasts of mortality the main question does not seem to be whether life expectancy will increase or decrease, but rather how strongly life expectancy will increase and how long the increase will continue. Basically, three types of change may be assumed. Firstly, one may assume a linear increase in life expectancy (which is not the same as a linear decrease of age-specific mortality rates) or a linear decline in the logarithm of the age-specific mortality rates. Such a trend may be explained by gradual improvements due to technological progress and increase in wealth. Secondly, one may assume that the rate of change is declining. For example, one may assume that the increase in life expectancy at birth will decline due to the fact that mortality rates for the youngest age groups are already so low that further improvements will be relatively small. More generally, the slowing down of the increase in life expectancy is related to the

rectangularisation of the survival curve. Thirdly, it may be assumed that due to future medical breakthroughs life expectancy may increase more strongly than at present.

The assumption about the type of trend is not only relevant for the specification of the medium variant but also for assessing the degree of uncertainty of the forecasts. Obviously, if one assumes that trends will continue and that the uncertainty only concerns the question whether or not the rate of change will be constant or will decline gradually, the uncertainty of the future value of life expectancy is much smaller than if one assumes that life expectancy may change in new, unprecedented directions due to medical breakthroughs.

As discussed above, one problem in determining the long-run trend in life expectancy is the choice of the base period. If one fits a mathematical function to the observed time series of life expectancy in a given period, the results may be quite different than if a model is fitted to another period. Either the estimated values of the parameters of the function may differ or even another function may be more appropriate. One cause of the sensitivity of the fitted function to the choice of the period is that part of the changes in life expectancy are temporary. For example, the increase in smoking by men in the Netherlands starting before the Second World War to a level of about 90% in the 1950s and the decline in the 1960s and 1970s to a level of about 40% had a significant effect on the trend in life expectancy: it caused a negative development in life expectancy in the 1960s and an upward trend in the 1980s and 1990s. It is estimated that smoking reduced life expectancy at birth around 1975 by some 4 years. If the percentage of smokers stabilizes at the present level, the negative effect of smoking can be expected to decline to 2 years. This pattern of change in mortality due to smoking is one explanation why the choice of the base period for projecting mortality has a strong effect on the extrapolation. It makes a lot of difference whether the starting year of the base period is chosen before the negative effect of smoking on life expectancy became visible or around the time that the negative effect reached its highest value. Another example of transitory changes is the decline in mortality at young ages. In the first half of the twentieth century the decline in mortality of newborn children was much stronger than at present. As a consequence, life expectancy at birth increased more strongly than at present. If these transitory changes are not taken into account in fitting a function to the time series of life expectancy, the long-run projections may be biased as temporary changes are erroneously projected in the long run.

In order to avoid these problems, Van Hoorn and De Beer (2001) developed a model in which the development of life expectancy over a longer period, 1900–2000, is described by a long-term trend together with the assumed effects of smoking, the rectangularization of the survival curve, the introduction of antibiotics after the Second World War, the increase and subsequent decline in traffic accidents in the 1970s and changes in the gender difference due to other causes than smoking. The long-run trend is described by a negative exponential curve. Figure 11.4 shows the assumed effects of selected determinants on the level of life expectancy. Figure 11.5 shows that the model fits the data very well (see the [appendix](#) for a more extensive description of the model). Figure 11.5 also shows projections up to 2050.

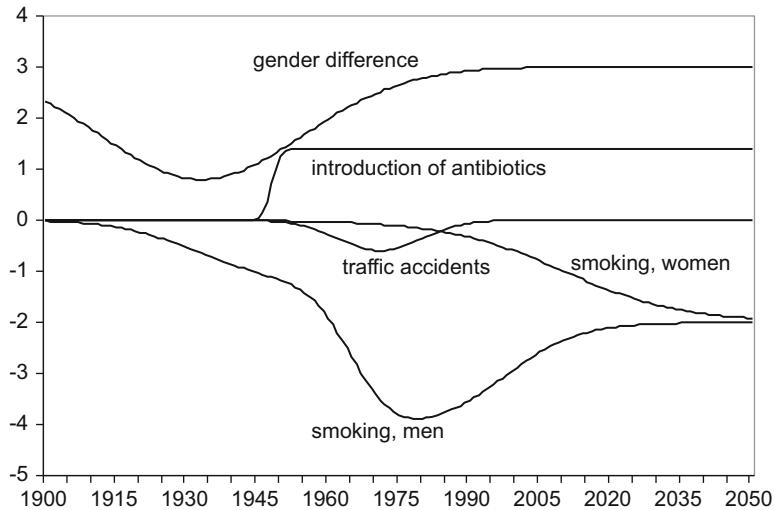


Fig. 11.4 Effects on life expectancy at birth



Fig. 11.5 Life expectancy at birth: observations and model

This model projects a smaller increase in life expectancy of men than, e.g., a linear extrapolation of the changes in the last 25 years or so would have done.

Since the model is deterministic, it cannot be used directly for making stochastic forecasts. The projections of the model are uncertain for at least two reasons. Firstly, it is not sure that the model is specified correctly. Several assumptions were made about effects on life expectancy which may be false. Secondly, in the future new unforeseen developments may occur that cannot be specified on the basis of

observations. For example, future medical breakthroughs may cause larger increases in life expectancy than what we have seen so far. For this reason expert knowledge is necessary to estimate the probability and the impact of future events that have not occurred in the past.

11.4 Expert Knowledge in the Dutch Stochastic Mortality Forecasts

For making stochastic forecasts of mortality it is assumed that the projections of the model described in Sect. 11.3 correspond with the expected values of future life expectancy. Assuming future life expectancy to be normally distributed, assumptions need to be made on the values of the standard deviations of future life expectancy.

As mentioned in Sect. 11.2 both an analysis of previous forecasts and model-based estimates of forecast variances can be combined with expert judgement. One problem in using information on historic forecasts to assess the uncertainty of new long-run forecasts is that there are hardly any data on forecast errors for the long run. Alternatively, forecast errors for the long run can be projected on the basis of forecast errors for the short and medium term. Time-series of historic forecast errors can be modelled as a random walk model (without drift). On the basis of this model the standard error of forecast errors 50 years ahead is estimated at 2 years. This implies that the 95% forecast interval for the year 2050 equals 8 years. Alternatively, the standard error of forecast errors can be projected on the basis of a time-series model describing the development of life expectancy. The development of life expectancy at birth for men and women in the Netherlands can be described by a random walk model with drift (Lee and Tuljapurkar 1994, model mortality in the United States as a random walk with drift too). The width of the 95% forecast interval produced by this model for the year 2050 equals 12 years. Thus, on the basis of the time-series models of life expectancy and models of forecast errors of life expectancy it can be expected that the 95% forecast interval of life expectancy in 2050 will be around 8–12 years. The decision which interval is to be used is based on judgement. Judgement ought to be based on an analysis of the processes underlying changes in life expectancy. The judgemental assumptions underlying the Dutch forecasts are based on four considerations.

- (1) It is regarded highly likely that the difference in mortality between men and women will continue to decrease. This difference has arisen in the past decades largely because of differences in smoking habits. As smoking habits of men and women have become more similar, the gender difference in life expectancy is assumed to decrease in the forecasts.
- (2) Changes in life expectancy at birth are the result of changes in mortality for different age groups. In assessing the degree of uncertainty of forecasts of life expectancy, it is important to make a distinction by age as the degree of

uncertainty of future changes in mortality differs between age categories. The effect of the uncertainty about the future development of mortality at young ages on life expectancy at birth is only small, because of the current, very low levels of mortality at young ages. On the basis of the current age specific mortality rates, 95.3% of live born men and 96.6% of women would reach the age of 50. Clearly, the upper limits are not far away. According to the medium variant of the 2000 Dutch population forecasts the percentage of men surviving to age 50 will rise to 97.0% in 2050 and the percentage of women to 97.4%. A much larger increase is not possible. A decrease does not seem very likely either. That would, *e.g.*, imply that infant mortality would increase, but there is no reason for such an assumption. The increase in the population with a foreign background could have a negative effect on mortality, since the infant mortality rates for this population group are considerably higher than those for the native population. However, it seems much more likely that infant mortality rates for the foreign population will decline rather than that they would increase. Furthermore, the effect on total mortality is limited. Another cause of negative developments at young ages could be new, deadly diseases. The experience with AIDS, however, has shown that the probability that such developments would have a significant impact on total mortality in the Netherlands (in contrast with, *e.g.*, African countries) does not seem very large. A third possible cause of negative developments at young ages would be a strong increase in accidents or suicides. However, there are no indications of such developments. Thus it can be concluded that the effect of the uncertainty about mortality at young ages on the uncertainty about the future development of life expectancy at birth is limited.

- (3) As regards older age groups one main assumption underlying the Dutch mortality forecasts is that the main cause of the increase in life expectancy at birth is that more people will become old rather than old people becoming still older. This implies the assumption that the survival curve will become more rectangular, an assumption based on an analysis of changes in age-specific mortality rates. The development of mortality rates for the eldest age groups in the 1980s and 1990s has been less favourable than for the middle ages. Another reason for assuming ‘rectangularisation’ of the survival curve is that expectations about a large increase in the maximum life span seem rather speculative, and even if they would become true, it is questionable whether their effect would be large during the next 50 years or so. A very strong progress of life expectancy can only be reached if life styles would change drastically or if medical technology would generate fundamental improvements (and health care would be available for everyone). Assuming a tendency towards rectangularisation of the survival curve implies that uncertainty about the future percentage of survivors around the median age of dying is relatively high. If the percentage of survivors around that age would be higher than in the medium variant (*i.e.*, if the median age would be higher), the decrease in the slope of the survival curve at the highest ages will be steeper than in the medium variant. Thus, the deviation from the medium variant at the highest ages will be smaller than around the median age. This implies that the degree of uncertainty associated with forecasts of life

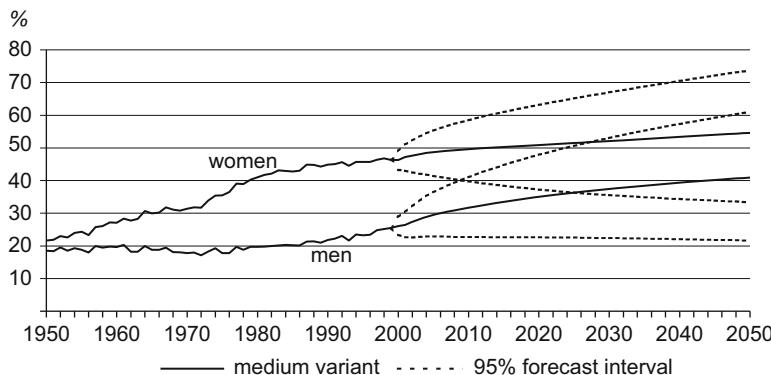


Fig. 11.6 Survivors at age 85; medium variant and 95% forecast interval

expectancy at birth mainly depends on changes in the median age of dying rather than on changes in the maximum life span. According to the medium variant of the 2000 population forecasts the percentage of survivors at age 85 in the year 2050 will be little under 50% for men and slightly over 50% for women. For that reason the degree of uncertainty of the mortality forecast is based on the assessment of a forecast interval at age 85.

- (4) The last consideration concerns the important point of discussion whether medical breakthroughs can lead to an unexpectedly strong increase of life expectancy. Even in case of a significant improvement of medical technology, it will be questionable to what extent this future improvement will lengthen the life span of present generations. It should be kept in mind that the mortality forecasts are made for the period up to 2050, and thus primarily concern persons already born. Experts who think that a life expectancy at birth could reach a level of 100 years or higher usually do not indicate when such a high level could be reached. It seems very unlikely that this will be the case in the period before 2050.

The four considerations discussed above are used to specify forecast intervals. According to the medium variant assumptions on the age-specific mortality rates for the year 2050, 41% of men will survive to age 85 (Fig. 11.6). According to the present mortality rates, little more than 25% of men would reach age 85. Because it is assumed to be unlikely that possible negative developments (*e.g.*, a strong increase in smoking or new diseases) will predominate positive effects of improvement in technology and living conditions during a very long period of time, the lower limit of the 95% forecast interval for the year 2050 is based on the assumption that it is very unlikely that the percentage of survivors in 2050 will be significantly lower than the current percentage. For the lower limit it is assumed that one out of five men will survive to age 85. This would imply that the median age of death is 77.5 years. The upper limit of the forecast interval is based on the assumption that it is very unlikely that about two thirds of men will survive past the age of 85. The median age of death

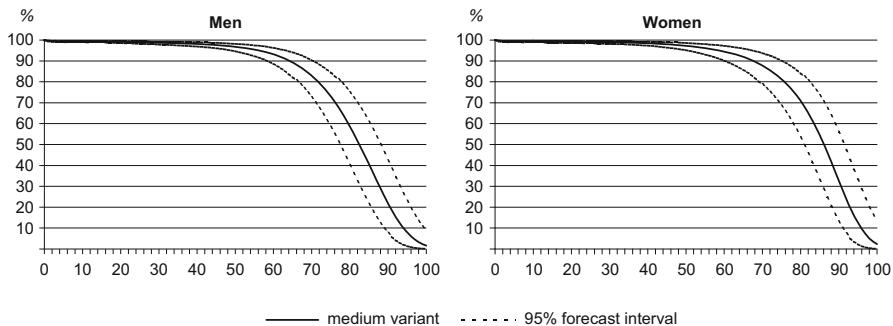


Fig. 11.7 Survival curves; medium variant and 95% forecast interval

would increase to 88 years. Currently, only 16% of men survive past the age of 88. A higher median age at dying than 88 seems thus very unlikely.

As discussed above, the medium variant assumes that the gender difference will become smaller. This implies that life expectancy of women will increase less strongly than that of men. This is in line with the observed development since the early 1980s. Consequently, the probability that future life expectancy of women will be lower than the current level is higher than the corresponding probability for men. The lower limit of the 95% forecast interval corresponds with a median age at dying of 81 years, which equals the level reached in the early 1970s. This could become true, *e.g.*, if there would be a strong increase in mortality by lung cancer and coronary heart diseases due to an increase in smoking. The upper limit of the forecast interval is based on the assumption that three quarters of women will reach age 85. This would imply that half of women would become older than 91 years. This is considerably higher than the current percentage of 21. It does not seem very likely that the median age would become still higher.

The intervals for the percentage of survivors at the age of 85 for the intermediate years are assessed on the basis of the random walk model (Fig. 11.6).

On the basis of these upper and lower limits of the 95% forecast interval for percentages of survivors at age 85, forecast intervals for percentages of survivors at the other ages are assessed, based on the judgemental assumption that for the youngest and eldest ages the intervals are relatively smaller than around the median age (Fig. 11.7). The age pattern of changes in mortality rates in the upper and lower limit are assumed to correspond with the age pattern in the medium variant.

The assumptions on the intervals of age-specific mortality rates are used to calculate life expectancy at birth. These assumptions result in a 95% forecast interval for life expectancy at birth in 2050 of almost 12 years. For men the interval ranges from 73.7 to 85.4 years and for women from 76.7 to 88.5 years (Fig. 11.8). The width of these intervals closely corresponds with that of the interval based on the random walk with drift model of life expectancy at birth mentioned before.

The intervals for the Netherlands are slightly narrower than the intervals for Germany specified by Lutz and Scherbov (1998). They assume that the width of

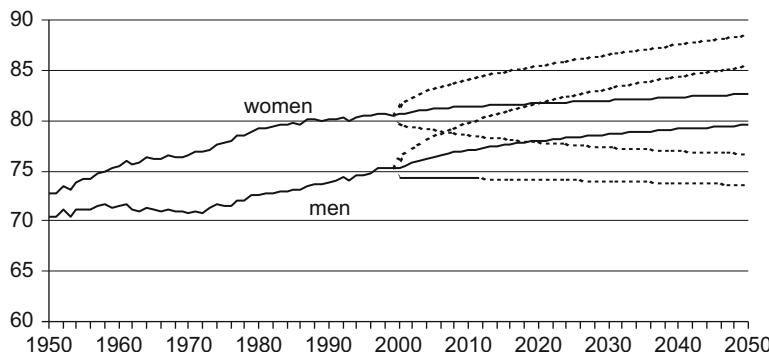


Fig. 11.8 Life expectancy at birth; medium variant and 95% forecast interval

the 90% interval equals 10 years in 2030. This is based on the assumption that the lower and upper limits of the 90% interval of the annual increase in life expectancy at birth equal 0 and 0.3 years respectively. This would imply that the width of the 90% interval in 2050 equals about 15 years.

11.5 Conclusions

Long-term developments in mortality are very uncertain. To assess the degree of uncertainty of future developments in mortality and other demographic events several methods may be used: an analysis of errors of past forecasts, a statistical (time-series) model and expert knowledge or judgement. These methods do not exclude each other; rather they may complement each other. For example, even if the assessment of the degree of uncertainty is based on past errors or on a time-series model judgement plays an important role. However, in publications the role of judgement is not always made explicit.

The most recent Dutch mortality forecasts are based on a model that forecasts life expectancy at birth. Implementation of the model is based on literature and expert knowledge. The model includes some important determinants of mortality, such as the effect of smoking and gender differences. Since the model is deterministic, it cannot be used for stochastic forecasting. Therefore, an expert knowledge approach is followed. This approach can be described as ‘argument-based forecasting’. Basically, four quantitative assumptions are made: (1) the difference in mortality between men and women will continue to decrease, (2) the effect of uncertainty about mortality is limited at young ages and is highest around the median age of dying, (3) the effect of medical breakthroughs on the life span will be limited up to 2050, and (4) more people will become old rather than old people will become still older (rectangularisation of the survival curve). Based on these assumptions target values for the boundaries of 95% forecast intervals are specified. It appears that the

width of the 95% interval of life expectancy at birth in 2050 is almost 12 years, both for men and women. This interval closely resembles the interval based on a random walk model with drift. It is about 4 years wider than the interval based on a time-series model of errors of historic forecasts.

Appendix: An Explanatory Model for Dutch Mortality

There are several ways to explain mortality. One approach is to assume a dichotomy of determinants of mortality – internal factors and external factors. For instance age, sex and constitutional factors are internal, whereas living and working conditions as well as socio-economic, cultural and environmental conditions are external. Other factors, such as life styles and education are partly internal, partly external.

An alternative approach takes the life course as a leading principle for a causal scheme. Determinants that act in early life are placed in the beginning of the causal scheme, those that have an impact later in life are put at the end. In this way heredity comes first and medical care comes last. Factors like life styles are in the middle. In the following scheme this approach is elaborated, though some elements of the first approach are used also. Eight categories of important determinants are distinguished:

- A. Heredity (including gender)
- B. Gained properties (education, social status)
- C. Life styles (risk factors like smoking, relationships)
- D. Environment (living and working conditions)
- E. Health policy (prevention of accidents, promotion of healthy life styles)
- F. Medical care (technological progress and accessibility of cure and care)
- G. Period effects (wars, epidemics)
- H. Rectangularity of survival curve

Interactions of gender with other factors should be taken in account in forecasting mortality because a lot of differences between men and women exist. Categories A, B, C and D reflect heterogeneity in mortality in the population, while groups E, F and G reflect more general influences. As life expectancy is the dependent variable in the explanatory model, a supplementary factor (H) is needed which is dependent on the age profile of the survival curve. When the survival curve becomes more rectangular, a constant increase in life expectancy can only be achieved through ever-larger reductions of mortality rates.

Most of the eight categories listed above contain many determinants. Of course it is not possible to trace and quantify all determinants. The selection of variables is based on the following criteria:

1. there is evidence about the magnitude of the effect and about changes over time;
2. independence of effects; and
3. there is a good possibility to formulate assumptions for the future.

In category C (life styles), smoking is a good example of a suitable explanatory factor since there is considerable evidence about the prevalence of smoking in the population and the effect on mortality. In category E the effect of safety measures on death from traffic accidents is an example of an independent and relatively easy factor to estimate. The same holds for category F for the introduction of antibiotics which caused a sharp drop in mortality by pneumonia.

On the contrary, general medical progress is not a very suitable factor, because there is much uncertainty and divergence of opinions about the impact on life expectancy. The effect is hard to separate from that of social progress, growth of prosperity, cohort-effects etc.

Part of the variation in mortality (life expectancy) can be modelled by separate effects, the rest is included in the trend. It must be stressed that the model does not quantify the effect of the determinants on causes of death (for instance smoking on death rates of lung cancer and heart diseases), but directly links them with overall mortality (life expectancy).

Six determinants that meet the three criteria were included in the model. Figure 11.4 in the paper shows the assumptions about the effect of these determinants on life expectancy at birth in the observation period (1900–2000) and the forecast period (2001–2050).

- A. *Heredity: gender difference.* Part of the gender difference in mortality can be attributed to differences in smoking behaviour (see C). Gender differences due to other factors were not constant through the entire twentieth century. However, since 1980 the difference seems to have stabilised. The change in the gender difference in life expectancy in the twentieth century can be described by a log normal curve.
- B. *Gained properties.* There are strong differences in mortality between social groups. However, these effects were not included in the model. One reason is that there is a strong correlation with, e.g., life styles, living and working conditions and access to medical care. Thus, the second criterion mentioned above is violated.
- C. *Life style: smoking.* On the basis of the literature the different effects for men and women are estimated. The effects can be quantified with a combination of a normal and a logistic curve. Smoking largely explains the quite different developments of life expectancy for men and women in the past decades. As smoking habits of men and women become more similar, in the forecasts the gender difference in life expectancy is assumed to decrease and finally become 3 years (in favour of women).
- D. *Environment.* These effects can be characterised as gradual long-term changes. It is hard to distinguish these effects from a general long-term trend. Hence these factors are not included in the model as separate effects but are included in the trend.

- E. *Health policy: traffic accidents.* In the early 1970s measures were taken to improve safety of traffic. As a result the number of deaths by traffic accidents declined. This effect is modelled as a deviation of the trend. Around 1970 the number of traffic accidents reached its highest level. A log normal curve appears to be appropriate for describing changes through time.
- F. *Medical care: introduction of antibiotics.* After the Second World War the introduction of antibiotics caused a sharp decline of mortality by pneumonia. A logistic function describes the rise in life expectancy and the flattening out to a constant level.
- G. *Period effects: outliers.* The Spanish flu and the Second World War caused sharp negative deviations of life expectancy, which are modelled by dummies.
- H. *Rectangularity.* The shape of the survival curve has an important impact on the pace of the increase of life expectancy. If the survival curve in year t is more rectangular than in year s , a given reduction in all age-specific death rates in both years will result in a smaller increase in life expectancy in year t than in year s . We define the rectangularity effect as the growth of life expectancy in year t divided by the growth in the year 1895 caused by the same percentage of decline of all age-specific mortality rates. A linear spline function is used to smooth the results. The increase in life expectancy in 1995 appears to be only 40% of that in 1895.

The new model contains a lot of parameters and simultaneous estimation can be problematic (unstable estimates). Therefore, the parameters were estimated in two steps. In the first step values of parameters of the functions describing the effect of smoking, traffic accidents, the introduction of antibiotics, and rectangularity were chosen in such a way that the individual functions describe patterns that correspond with available evidence. In the second step the values of the trend parameters and the outliers were estimated on the basis of non-linear least squares and some values of parameters fitted in the first step were ‘fine-tuned’.

Several functions were tested to fit the trend. A negative exponential curve appears to fit the development since 1900 best (see Fig. 11.5 in the paper). This function implies that there is a limit to life expectancy. However, according to the fitted model this is not reached before 2100.

The model is:

$$\begin{aligned}
 e_{0,g,t} &= T_t + S_{g,t} + V_t + A_t + G_{g,t} + \sum_{j=1917}^{1919} u_{j,g,t} + \sum_{j=1940}^{1945} u_{j,g,t} + \varepsilon_{g,t} \\
 T_t &= a_0 + H_t R_t \\
 H_t &= a_1 e^{-a_2(t-t_1)} \\
 R_t &= 1 - b_1(t-t_2) - b_2(t-t_3) D_{1,t}
 \end{aligned}$$

$$\begin{aligned}
D_{1,t} &= 0 \text{ if } t < t_3 \text{ and } D_{1,t} = 1 \text{ if } t \geq t_3 \\
S_{g,t} &= c_{g,1} e^{-c_{g,2}(t-t_4)^2} + \frac{c_{g,3}}{1 + c_{g,4} e^{-c_{g,5}(t-t_5)}} + \frac{c_{g,6}}{1 + c_{g,7} e^{-c_{g,8}(t-t_6)}} + c_{g,9} \\
V_t &= d_1 e^{-d_2 \ln(t/t_7)^2} \\
A_t &= \frac{f_1}{1 + f_2 e^{-f_3(t-t_8)}} + f_4 \\
G_{g,t} &= \left(h_1 e^{-h_2 \ln(t/t_9)^2} + h_3 \right) D_{2,g} \\
D_{2,g} &= 1 \text{ if } g = \text{female and } D_{2,g} = 0 \text{ if } g = \text{male} \\
u_{j,g,t} &= 1 \text{ if } t = j \text{ and } u_{j,g,t} = 0 \text{ in other years,}
\end{aligned}$$

where $e_{0,g,t}$ is life expectancy at birth for gender g in year t , T is trend, S is the effect of smoking, V is the effect of traffic accidents, A is the introduction of antibiotics, G is the unexplained gender difference, u are outliers, ϵ is error, H is slope of the trend and R is the effect of the rectangularity of the survival curve.

References

- Alho, J. M. (1998). *A stochastic forecast of the population of Finland* (Reviews 1998/4). Statistics Finland, Helsinki.
- Alho, J. M., & Spencer, B. D. (1985). Uncertain population forecasting. *Journal of the American Statistical Association*, 80, 306–314.
- Box, G. E. P., & Jenkins, J. M. (1970). *Time series analysis. Forecasting and control*. San Francisco: Holden-Day.
- Cohen, J. E. (1986). Population forecasts and confidence intervals for Sweden: a comparison of model-based and empirical approaches. *Demography*, 23, 105–126.
- De Beer, J. (1997). The effect of uncertainty of migration on national population forecasts: the case of the Netherlands. *Journal of Official Statistics*, 13, 227–243.
- Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge: Cambridge University Press.
- Keilman, N. W. (1990). *Uncertainty in population forecasting: issues, backgrounds, analyses, recommendations*. Amsterdam: Swets & Zeitlinger.
- Lee, R. (1998). Probability approaches to population forecasting. In W. Lutz, J. W. Vaupel, & D. A. Ahlburg (Eds.), *Frontiers of population forecasting* (A Supplement to Volume 24, Population And Development Review) (pp. 156–190). New York: Population Council.
- Lee, R., & Tuljapurkar, S. (1994). Stochastic population forecasts for the United States: beyond high, medium, and low. *Journal of the American Statistical Association*, 89, 1175–1189.
- Lutz, W., & Scherbov, S. (1998). Probabilistische Bevölkerungsprognosen für Deutschland. *Zeitschrift für Bevölkerungswissenschaft*, 23, 83–109.
- Lutz, W., Goldstein, J. R., & Prinz, C. (1996a). Alternative approaches to population projection. In W. Lutz (Ed.), *The future population of the world. What can we assume today?* (pp. 14–44). London: Earthscan.

- Lutz, W., Sanderson, W., Scherbov, S., & Goujon, A. (1996b). World population scenarios for the 21st century. In W. Lutz (Ed.), *The future population of the world. What can we assume today?* (pp. 361–396). London: Earthscan.
- Tuljapurkar, S. (1996, September 19–21). *Uncertainty in demographic projections: methods and meanings*. Paper presented at the sixth annual conference on applied and business demography, Bowling Green.
- Van Hoorn, W., & De Beer, J. (2001). Bevolkingsprognose 2000–2050: Prognosemodel voor de sterfte. Maandstatistiek van de bevolking. *Statistics Netherlands*, 49(7), 10–15.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 12

Stochastic Forecasts of Mortality, Population and Pension Systems



Shripad Tuljapurkar

12.1 Introduction

This paper discusses the construction of stochastic forecasts of human mortality and fertility rates and their use in making stochastic forecasts of pension funds. The method of mortality analysis was developed by Lee and Carter (1992), henceforth called the LC method. Lee and Tuljapurkar (1994) combined the LC method with a related fertility forecast to make stochastic population forecasts for the US. Tuljapurkar and Lee (1999) and Lee and Tuljapurkar (2000) combined these population forecasts with a number of other forecasts to generate stochastic forecasts of the US Social Security system.

My goal is to explain the distinctive features, strengths, and shortcomings of the stochastic method rather than to explain the method. I begin with a discussion of stochastic forecasts and their differences from scenario forecasts. Then I discuss mortality forecasts using Swedish mortality data, including a new forecast for Sweden. I go on to consider briefly how population forecasts are made and their use in modeling pension systems.

12.2 Stochastic Forecasts

A population forecast made in year T aims to predict population $P(t)$ for later years, where P includes numbers and composition. The information on which the forecast is based includes the history of the population and of environmental factors (economic, social, etc.). Every forecast maps history into a prediction. Scenario forecasts

S. Tuljapurkar (✉)
Stanford University, Stanford, CA, USA
e-mail: tulja@stanford.edu

rely on a subjective mapping made by an expert, whereas stochastic forecasts attempt to make an explicit model of the historical dynamics and project this dynamic into the future. Stochastic forecasts may rely partly on a subjective mapping as well. What are the pros and cons of the two approaches?

When historical data contain a strong “signal” that describes the dynamics of a process, it is essential to use the signal as a predictive mechanism. Equally, it is important to include information that is not contained in the signal – this residual information is an important element of uncertainty that should be reflected in the forecast. The LC method shows that there is such a signal in mortality history. When there is no strong signal in the historical data, a subjective prediction may be unavoidable. Fertility history tends to reveal relatively little predictive signal. Even here, uncertainty ought to be included because history does tell us about uncertainty, and we can estimate the variability around a subjective prediction.

The use of history to assess uncertainty certainly does make assumptions about persistence in the dynamic processes that drive the variables we study. This does not imply that we assume an absence of surprises or discontinuities in the future. Rather it assumes that all shocks pass through a complex filter (social, economic, and so on) into demographic behavior, and that future shocks will play out in the same statistical fashion as past shocks. I would not abandon this assumption without some demonstration that the filtering mechanisms have changed – witness for example the stock market bubble in the US markets in 1999–2000 and its subsequent decline. It may be useful to think about extreme scenarios that restructure aspects of how the world works – one example is the possibility that genomics may change the nature of both conception and mortality in fundamental ways – but I regard the exploration of such scenarios as educational rather than predictive.

I argue strongly for the systematic prediction of uncertainty in the form of probability distributions. This position does not argue against using subjective analysis where unavoidable. One way of doing a sound subjectively based analysis is to follow the work of Keilman (1997, 1998) and Alho and Spencer (1997) and use a historical analysis of errors in past subjective forecasts to generate error distributions and project them. The practice of using “high-low” scenarios should be avoided. Uncertainty accumulates, and must be assessed in that light. In my view, the best that a scenario can do is suggest extreme values that may apply at a given time point in the future – for example, demographers are often reluctant to believe that total fertility rate (TFR) will wander far from 2 over any long interval, so the scenario bounds are usually an acceptable window around 2, such as 1.5–2.2. Now this may be plausible as a period interval in the future but in fact tells us nothing useful about the dynamic consequences of TFR variation over the course of a projection horizon.

Uncertainty, when projected in a probabilistic manner, provides essential information that is as valuable as the central location of the forecast. To start with, probabilities tell us how rapidly the precision of the forecast degrades as we go out into the future. It can also be the case that our ability to predict different aspects of population may differ, and probability intervals tell us about this directly.

Probabilities also make it possible to use risk metrics to evaluate policy: these are widely used in insurance, finance, and other applications, and surely deserve a bigger place in population-related planning and analysis.

12.3 Mortality Forecasts

The LC method seeks a dominant temporal “signal” in historical mortality data in the form of the simplest model that captures trend and variation in death rates, and seeks it by a singular-value decomposition applied to the logarithms $\log m(x, t)$ of central death rates. For each age x subtract the sample average $a(x)$ of the logarithm, and obtain the decomposition

$$\log m(x, t) - a(x) = \sum_i s_i u_i(x) v_i(t).$$

On the right side above are the real singular values $s_1 \geq s_2 \dots \geq 0$. The ratio of s_1^2 to the sum of squares of all singular values is the proportion of the total temporal variance in the transformed death rates that is explained by just the first term in the singular-value decomposition.

In all the industrialized countries that we have examined, the first singular value explains well over 90% of the mortality variation. Therefore we have a dominant temporal pattern, and we write

$$\log m(x, t) = a(x) + b(x) k(t) + E(x, t).$$

The single factor $k(t)$ corresponds to the dominant first singular value and captures most of the change in mortality. The far smaller variability from other singular values is $E(x, t)$.

The dominant time-factor $k(t)$ displays interesting patterns. Tuljapurkar et al. (2000) analyzed mortality data in this way for the G 7 countries over the period from approximately 1955–1995. They found that the first singular value in the decomposition explained over 93% of the variation, and that the estimated $k(t)$ in all cases showed a surprisingly steady linear decline in $k(t)$. The mortality data for Sweden from 1861 to 1999 constitute one of the longest accurate series, and a similar analysis in this case reveals two regimes of change in $k(t)$. The estimated $k(t)$ for Sweden is shown in Fig. 12.1. There is steady decline from 1861 to about 1910 and after 1920 there is again steady decline but at a much faster rate. Note that the approximately linear long-term declines are accompanied by quite significant short-term fluctuations. It is possible that we can interpret period-specific fluctuations in terms of particular effects (*e.g.*, changes in particular causes of death) but it is difficult to project these forward. For example, the change in the pattern in the early 1900s is

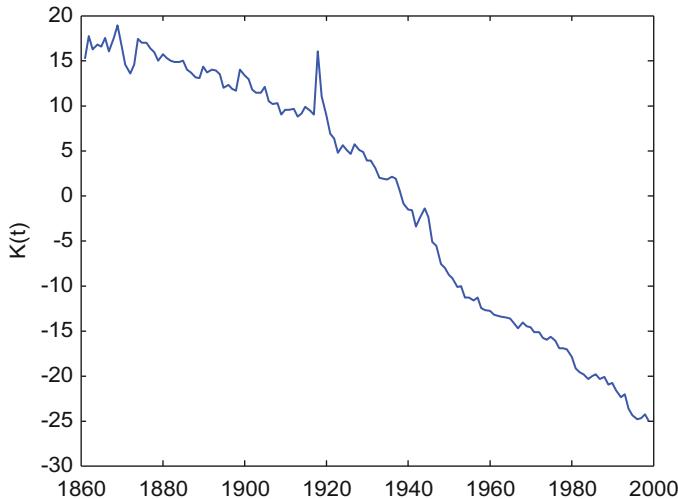


Fig. 12.1 Lee Carter $k(t)$ Sweden 1861 to 1999

consistent with our views of the epidemiological transition, but we do not know if the future will hold such a qualitative shift. Within the twentieth century we take the approach of using the dominant linearity coupled with superimposed stochastic variation.

Mortality decline at any particular age x is proportional to the signal $k(t)$ but its actual magnitude is scaled by the response profile value $b(x)$. Figure 12.2 shows the $b(x)$ profiles computed for Swedish data using 50 year spans preceding the dates 1925, 1945, 1965, and 1985. Note that there is a definite time evolution, in which the age schedules rotate (around an age in the range 40–50) and translate so that their weight shifts to later ages as time goes by. This shifting corresponds to the known sequence of historical mortality decline starting with declines initially at the youngest ages and then in later ages over time. An intriguing possibility is that temporal changes in the $b(x)$ schedules may be described by a combination of a scaling and translation – a sort of nonlinear mapping over time. An important matter for future work is to explore the time evolution of the $b(x)$, even though it appears (see below) that one can make useful forecasts over reasonable time spans of several decades by relying on a base span of several decades to estimate a relevant $b(x)$.

What accounts for the regular long-term decline in mortality that is observed over any period of a few decades? It is reasonable to assume that mortality decline in this century has resulted from a sustained application of resources and knowledge to public health and mortality reduction. Let us assume, as appears to be the case, that societies allocate attention and resources to mortality reduction in proportion to observed levels of mortality at different ages (*e.g.*, immunization programs against childhood disease, efforts to reduce cardiovascular disease at middle age). Such allocation would produce an exponential (proportional) change in mortality, though

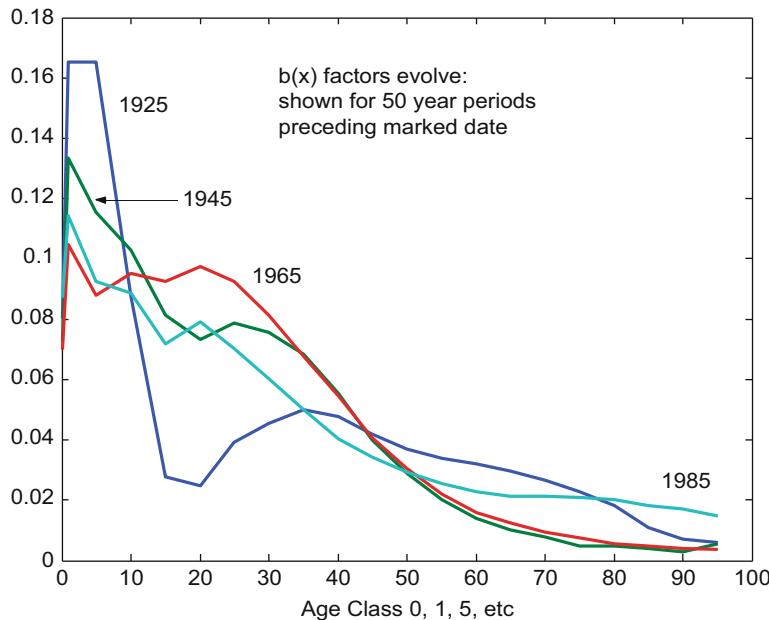


Fig. 12.2 b(x) Response factors Sweden

not necessarily at a constant rate over time. Over time, the rate of proportional decline depends on a balance between the level of resources focused on mortality reduction, and their marginal effectiveness. Historically, the level of resources has increased over time but their marginal effectiveness has decreased over time (because, for example, we are confronted with ever more complex causes of mortality that require substantial resources or new knowledge). The observation of linearly declining $k(t)$ – roughly constant long-run exponential rates of decline – implies that increasing level and decreasing effectiveness have balanced each other over long times. It is of course possible that the linear pattern of decline we report has some other basis. For the future, we expect a continued increase in resources spent on mortality reduction, and a growing complexity of causes of death. The balance between these could certainly shift if there were departures from history – for example, if new knowledge is discovered and translated into mortality reductions at an unprecedented rate. But this century has witnessed an amazing series of discoveries that have altered medicine and public health, and there is no compelling reason why the future should be qualitatively different. Therefore, I expect a continuation of the long-run historical pattern of mortality decline.

The LC method uses the long-term linear decline in $k(t)$ to forecast mortality. A naive forecast based on the long-run trend is not sensible because the short-term variation will accumulate over time, so it is essential to employ a stochastic forecast. In LC, the stochastic term $E(t)$ is modeled as a stationary noise term, and this

procedure leads to forecasts for Sweden as shown in Fig. 12.3, for life expectancy at birth, e_{00} , and in Fig. 12.4 for life expectancy at age 65, e_{65} . In both cases we use a 50-year span of historical data prior to a launch date of 1999. The intervals shown are 95% prediction intervals for each forecast year. Notice that there are separate forecasts for each sex, as well as a combined-sex forecast. The joint analysis of the two sexes in an LC framework has not been fully resolved, although Li et al. (2004) suggest one method for doing this.

Some previous comments on the LC method have asserted that it is simply equivalent to a linear extrapolation in the log scale of the individual rates at each age, but it is not. For one thing, the extrapolations would include elements of the $E(t)$ terms in each age, and these may be larger at some ages than at others. For another, I take the stochastic variation seriously as an integral part of the forecast, and the realized long run trend can be rather different depending on where in the prediction interval one ends up. Without this variability, the forecasts would not be terribly useful over long horizons.

To illustrate the robustness of the LC method, Lee and Miller (2001) have analyzed the performance of the method using internal validation. A more extensive analysis for Sweden echoes their finding that the method is surprisingly robust. To illustrate, I use different base periods to forecast e_0 in 1999. I first select a starting base year, say 1875, and then a launch year which is chosen from the set 1935, 1945, ..., 1995; this gives a total of seven forecasts starting in 1875. We expect that a forecast for 1999 using the 1875–1935 base period would be much less accurate than a forecast which uses the 1875–1995 base period. The object of the exercise is to see whether the projection intervals for e_0 in 1999 will decrease in some systematic way as we include more recent (relative to 1999) history and whether they speak to the accuracy of the method. Figure 12.5 plots the projection intervals obtained in this way, using each of three starting years (1875, 1900, or 1925) and the seven launch years indicated above, so for each starting year we have an upper and lower prediction “fan” for e_0 in 1999. The figure shows that as we use more recent histories, we close in on the true 1999 value of e_0 of 79.4 years – the 95% prediction interval brackets the true value most of the time which is impressive especially when compared with the historical performance of scenario forecasts. From a practical point of view, the prediction interval width is under 7 years for launch dates from 1960 to 1980 and any of the starting base years. This means that we may expect a reasonable performance from LC forecasts for as far as 40 years into the future.

12.4 From Population to Pension Systems and Policy

For a population forecast we must supplement mortality forecasts with similar forecasts for fertility and if necessary for immigration. These elements can then be combined in the usual cohort-component procedure to generate stochastic population forecasts. Fertility forecasts pose special challenges because there does not seem to be a strong temporal pattern to fertility dynamics. Lee and Tuljapurkar (1994) use

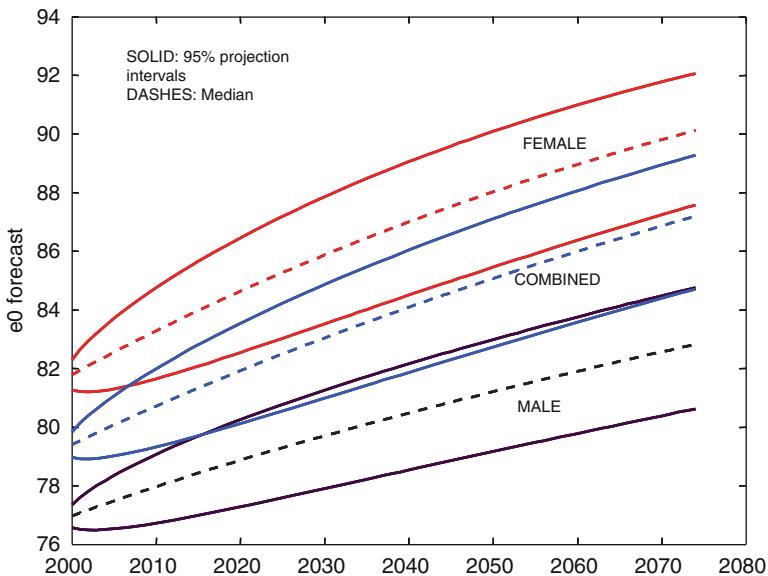


Fig. 12.3 e0 Stochastic forecast Sweden launch 1999

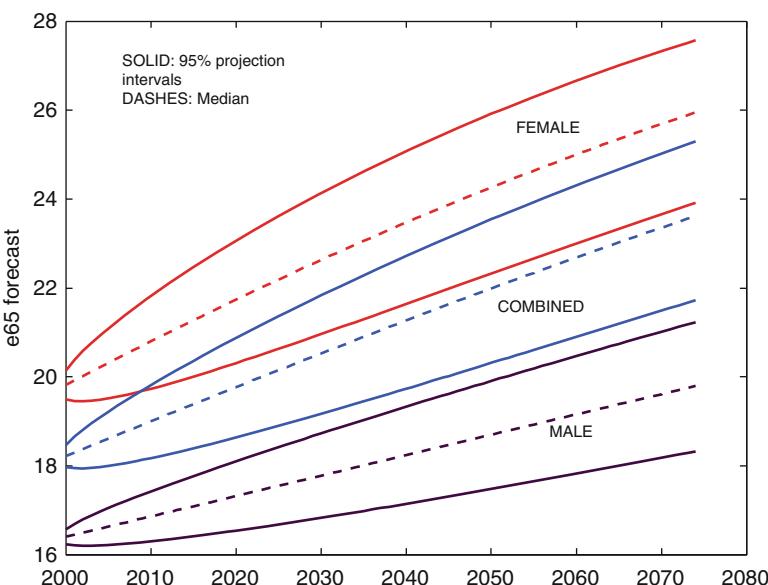


Fig. 12.4 e65 Stochastic forecast Sweden, launch data 1999

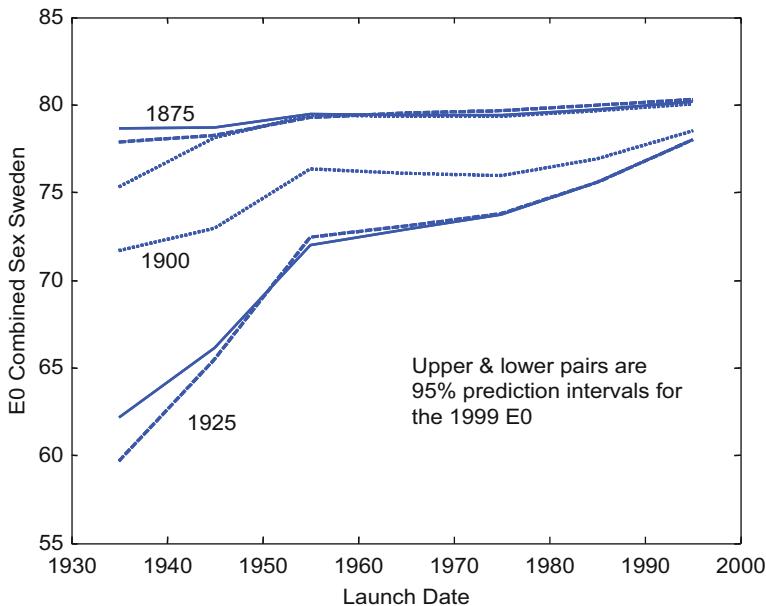


Fig. 12.5 Forecasts that use data going back to 1875, 1900, 1925

time series models for fertility to make stochastic forecasts for the US. Their simple models have been considerably extended by Keilman and Pham (2004) who suggest several ways of modeling and constraining the volatility of fertility forecasts.

How can stochastic forecasts be used in analyzing pension policy? At a purely demographic level, it is well known that the old-age dependency ratio is the key variable that underlies pension costs. As the old-age dependency ratio for a population increases, the more retirees-per-worker there are in the population, which implies greater stress on a pay-as-you-go pension system which relies on today's workers to pay the benefits of today's retirees. An interesting insight into the demographic impact of aging on the dependency ratio can be created by asking the following question. Suppose that the age at which people retire is, *e.g.*, 65. If this "normal retirement age" age cutoff could be changed arbitrarily, how high would we have to raise it in order to keep the dependency ratio constant? If we have a population trajectory forecast, then we can simply compute in each year the retirement age, say $R(t)$, at which the old-age dependency ratio would be the same ratio as in the launch year. When we have stochastic forecasts, there is, in each forecast year t , a set of random values $R(t)$; in our analysis we look for the integer value of $R(t)$ that comes closest to yielding the target dependency ratio. Figure 12.6 shows the results of computing these stochastic $R(t)$ for the US population. What is plotted is actually three percentiles of the distribution of $R(t)$ in each year, the median value, and the upper and lower values in a 95% projection interval. The plots show some long steps because the dependency ratio distribution changes fairly slowly over time. The

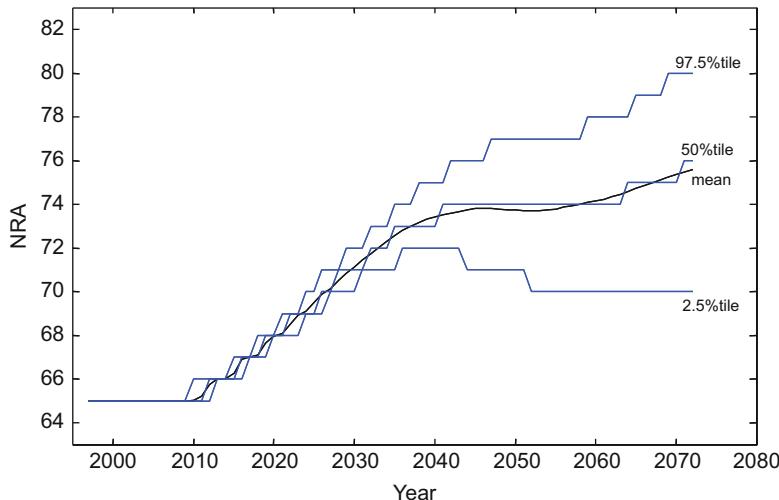


Fig. 12.6 Normal retirement age to maintain US old-age dependency ratio, 1997-2072

smooth line shows the average value of $R(t)$ for each forecast year, which is surprisingly close to the median. Observe, for example, that there is a 50% chance that the “normal retirement age” would have to be raised to 74 by 2060 in order to keep the dependency ratio constant at its 1997 value. There is only a 2½% chance that the “normal retirement age” of 69 years would suffice. Given that current US Social Security policy is only intended to raise the “normal retirement age” to 67 years, and that even the most draconian proposals would only raise it to 69 years, we conclude that changes in the “normal retirement age” are very unlikely to hold the dependency ratio constant. Anderson et al. (2001) present similar results for the G7 countries. In Tuljapurkar and Lee (1999) there are additional examples of how stochastic forecasts can be combined with objective functions to analyze fiscal questions.

To go beyond this type of analysis we need a full model of the structure of a pension system which may be “fully funded” or “pay-as-you-go” or some mixture. Many systems, in order to operate with a margin of security, are modified versions of pay-as-you-go systems that include a reserve fund. In the United States the OASDI (Old Age Survivors and Disability Insurance, or Social Security) Trust Fund, the holdings of the system are federal securities, and the “fund” consists of federally-held debt. A fund balance earns interest, is subject to withdrawals in the form of benefit payments, and receives deposits in the form of worker contributions (usually in the form of tax payments). Lee and Tuljapurkar (2000) discuss such models for the US Social Security system and also for other fiscal questions. The dynamics of such models proceed by a straightforward accounting method. Starting with a launch year (initial year) balance, we forecast contributions and benefit payments for each

subsequent year, as well as interest earned. This procedure yields a trajectory of fund balance over time. Future contributions depend both on how many workers contribute how much to the system. Future benefit payments depend on how many beneficiaries receive how much in the future. Our population forecasts do not directly yield a breakdown in terms of workers and retirees. Therefore, we estimate and forecast per-capita averages by age and sex, for both contributions and benefits. We combine these age and sex-specific “profiles” with age and sex-specific population forecasts to obtain total inflows and outflows for each forecast year.

Contribution profiles evolve over time according to two factors. First, increases in contributions depend in turn on increases in the real wage. We forecast real wage increases stochastically (as described below), and contributions increase in proportion to wages. Second, changes in the labor force participation rates also affect contributions; we forecast labor force participation rates deterministically. Benefit profiles evolve over time in response to several factors. In our model of the U.S. Social Security system, we disaggregate benefits into disability benefits and retirement benefits. Retirement benefit levels reflect past changes in real wages because they depend on a worker’s lifetime wages. Also, legislated or proposed changes in the Normal Retirement Age (the age at which beneficiaries become eligible to collect 100% of their benefits) will reduce benefits at the old NRA.

Demographic variables are obviously not the only source of uncertainty facing fiscal planners; there are sizable economic uncertainties as well. Taxes and future benefits usually depend on wage increases (economic productivity) and funds can accumulate interest or investment returns on tax surpluses. Our models combine uncertainty in productivity and investment returns by converting productivity to real 1999 dollars, subtracting out increases in the CPI. We then model productivity rates and investment returns stochastically.

There is substantial correlation between interest rates on government bonds and returns to equities, so it is important to model these two variables jointly. For our historical interest rate series we use the actual, effective real interest rate earned by the trust fund, and for historical stock market returns we use the real returns on the overall stock market as a proxy. These two series are modeled jointly as a vector auto-regressive process.

Our stochastic model allows us to simulate many (1000 or more, usually) trajectories of all variables and obtain time trajectories of the fund balance from which we estimate probabilities and other statistical measures of the system’s dynamics. This method may be used to explore the probability that particular policy outcomes are achieved, for example, that the “fund” stays above a zero balance for a specified period of years, or that the level of borrowing by the fund does not exceed some specified threshold.

Acknowledgements My work in this area involves a long-term collaboration with Ronald Lee and Nan Li. My work has been supported by the US National Institute of Aging.

References

- Alho, J., & Spencer, B. D. (1997). The practical specification of the expected error of population forecasts. *Journal of Official Statistics*, 13, 201–225.
- Anderson, M., Tuljapurkar, S., & Li, N. (2001). How accurate are demographic projections used in forecasting pension expenditure? In T. Boeri, A. Borsch-Supan, A. Brugiavini, R. Disney, A. Kapteyn, & F. Peracchi (Eds.), *Pensions: More information, less ideology* (pp. 9–27). Boston: Kluwer Academic Publishers.
- Keilman, N. (1997). Ex-post errors in official population forecasts in industrialized countries. *Journal of Official Statistics*, 13(3), 245–277.
- Keilman, N. (1998). How accurate are the United Nations world population projections? *Population and Development Review*, 24(Supplement), 15–41.
- Keilman, N., & Pham, D. Q. (2004). Time series errors and empirical errors in fertility forecasts in the Nordic countries. *International Statistical Review*, 72(1), 5–18.
- Lee, R., & Carter, L. (1992). Modeling and forecasting U.S. Mortality. *Journal of the American Statistical Association*, 87(419), 659–671.
- Lee, R., & Miller, T. (2001). Evaluating the performance of the Lee-Carter approach to modeling and forecasting mortality. *Demography*, 38(4), 537–549.
- Lee, R., & Tuljapurkar, S. (1994). Stochastic population projections for the U.S.: Beyond high, medium and low. *Journal of the American Statistical Association*, 89(428), 1175–1189.
- Lee, R., & Tuljapurkar, S. (2000). Population forecasting for fiscal planning: Issues and innovations. In A. Auerbach & R. Lee (Eds.), *Demography and fiscal policy*. Cambridge: Cambridge University Press.
- Li, N., Lee, R., & Tuljapurkar, S. (2004). Using the Lee-Carter method to forecast mortality for populations with limited data. *International Statistical Review*, 72, 19–36.
- Tuljapurkar, S., & Lee, R. (1999, August). Population forecasts, public policy, and risk, bulletin of the international statistical institute. In Proceedings, 52nd annual meeting, ISI, Helsinki.
- Tuljapurkar, S., Li, N., & Boe, C. (2000). A universal pattern of mortality decline in the G7 countries. *Nature*, 405(15), 789–792.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Part III

The Linear Rise in Life Expectancy: History and Prospects

Tommy Bengtsson

Jim Oeppen and James Vaupel's study "Broken limits to life expectancy", published in Science in 2002, provided the background for this part. This study showed that increases in life expectancy were initially due to reductions in death rates in the younger ages, later followed by a decrease in rates for older persons. This development was initially triggered by a decline in infectious diseases and at a later stage a downward trend in chronic diseases. The resulting substantial improvements in life expectancy led to what these authors identify as "best-practice populations." The authors show that the development of "best-practice populations" are best approximated by a linear trend, estimated over the past 160 years. Females have continuously gained almost 3 month per year, males slightly less. The amazing fact is not the increase as such, since countries entering the mortality transition at a later point of time will experience an even faster improvement in life expectancy than those that preceded them, but the regularity with which the world record has been broken. This leads at once to the questions: What are the causes of this linear increase and how much longer can it proceed? These are the questions discussed in this part.

The chapter by Jim Oeppen and James Vaupel brings their 2002 article up to the forefront once more, developing further the arguments put forward there and discuss the implications of their findings for mortality forecasting. They argue that the increase in life expectancy is not slowing down and that in the near future we should expect average life expectancy to continue to increase at the same rate as before. They also argue that countries lagging behind tend to catch up with the best-practice populations. Thus, best-practice life expectancy should be used when making national forecasts.

In the chapters by Ronald Lee and Juha M. Alho respectively, this standpoint is partly called into question and, from different perspectives, they argue that individual countries are unable to stay at the best-practice line for long time periods. Instead,

the trends for leading countries tend to “bend down” as time passes. When making forecasts, the issue is thus not only to capture the catch-up phase but also thereafter a phase when improvements no longer keep pace with newly emerging best-practice countries.

In his chapter, Jim Oeppen explores further the discussion of identifying the processes that have led to the linear increase in life expectancy over the past 160 years. By employing a causal model acknowledging the significance of factors such as per capital income and technical change for a large number of countries, Oeppen analyses convergence in national trends in life expectancy. The final chapter, by Tommy Bengtsson, is also devoted to the causes of the linear decline. Bengtsson argues that there is a variety of factors changing over time that determine trends in life expectancy, economic performance only being one of these and not always the most important one. A circumstance that has to be taken into consideration is that countries catching up and taking over the lead have had relatively small elderly populations. This would also imply that the elderly in these populations have gone through a process of selection. In addition, they may have access to more modern care resources per capita than their counterparts in countries that have experienced a slower mortality transition, albeit combined with a similar economic development. Since this advantage is not permanent, it disappears and the advantage of backwardness turns into a penalty for taking the lead.

Chapter 13

The Linear Rise in the Number of Our Days



Jim Oeppen and James W. Vaupel

If life expectancy¹ – also known as the expectation of life, is the mean life-span of a cohort of newborns if current age-specific death rates remain unchanged – in developed countries were close to an ultimate limit, then increases in record life expectancy – the average length of life in the best-practice population – should slow as the ceiling is asymptotically approached.

Best-practice national life expectancy has, contrary to what many believe² (Olshansky et al. 2001; Riley 2001; Dublin 1928; Dublin and Lotka 1936; Olshansky

We are grateful to the many people who have provided comments and information, including Kenneth Wachter and Yasuhik Saito. A version of this article that does not include some of the material here but that includes some additional material was published by Oeppen and Vaupel in 2002.

¹Most of the life-expectancy calculations in this article are based on data on death rates over age and time in the Human Mortality Database, see <http://www.demog.berkeley.edu/wilmoth/mortality>. Recent Japanese data can be found at <http://www.mhlw.go.jp/english/database/index.html>. Some data for the period before 1950 are from Keyfitz and Flieger (1968) and other sources.

²For reviews, see Preston (1974); Keilman (1997). For a critical account of the low mortality assumptions used by the U.S. Social Security Administration, see Lee (2000). A review of mortality forecasting in 13 European Union countries in the early- and mid-1990s found that all assumed that mortality improvements would decelerate and 10 constrained life expectancy to reach an ultimate limit by a target date (Cruijsen and Eding 2001). In a report notorious for missing the baby boom, Whelpton et al. (1947) focused their discussion on life-expectancy limits for U.S. native white males. They concluded that for this population a life expectancy in the year 2000 of 72.1 years was the upper limit of what could be achieved by the largest mortality “declines that seem reasonable” and close to what could be attained at the “biological minimum of mortality”. In two publications Frejka (1973, 1981) focus on population growth rather than life expectancy. In the first, Frejka writes that “within broad limits mortality can be fairly well predicted.” He believes that life expectancy will approach a limit and that 77.5 is the most likely limit. He notes, however, that “mortality might even take a course absolutely different from what has been assumed.”

J. Oeppen (✉) · J. W. Vaupel
Max Planck Institute for Demographic Research, Rostock, Germany
e-mail: joeppen@health.sdu.dk

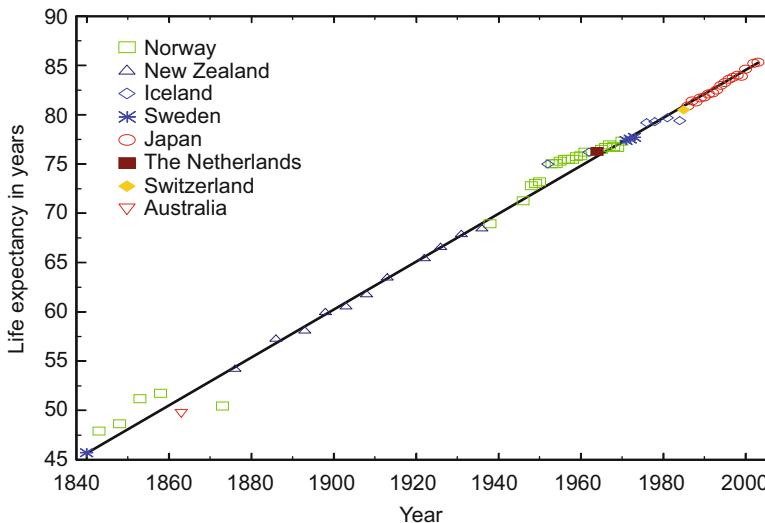


Fig. 13.1 Best-practice national life expectancy over the last 160 years

et al. 1990; Bourgeois-Pichat 1952, 1978; Fries 1980, 1990; Siegel 1980; Demeny 1984; United Nations 1973, 1985, 1989, 1999, 2001; NIPSSR 1997), risen for 160 years at a steady pace of 3 months per year as shown in Fig. 13.1.

Before 1950 most of the gain in life expectancy was due to large reductions in death rates at younger ages. The conventional view is that “future gains in life expectancy cannot possibly match those of the past, because they were achieved primarily by saving the lives of infants and children – something that happens only once for a population” (Olshansky et al. 2001). The sustained improvement in best-practice life expectancy belies this contention. In the second half of the twentieth century improvements in survival after age 65 propelled the rise in the length of people’s lives. For Japanese females, remaining life expectancy at age 65 grew from 13 years in 1950 to 22 years today, and the chance of surviving from 65 to 100 soared from less than 1 in 1000 to 1 in 20.³

The linear climb of record life expectancy suggests that reductions in mortality should not be seen as a disconnected sequence of unrepeatable revolutions but rather as a regular stream of continuing progress. Mortality improvements result from the intricate interplay of advances in income, salubrity, nutrition, education, sanitation, and medicine, with the mix varying over age, period, cohort, place and disease (Riley 2001). Reinforcing processes may help sustain the increase. For instance, reductions in premature deaths reduce bereavement, an important risk factor for mortality. The improvements also increase the number of people who survive to high ages, leading to greater attention to health at those ages. Increasingly prosperous, educated populations aided by armies of researchers, physicians, nurses and public-

³Recent Japanese data can be found at <http://www.mhlw.go.jp/english/database/index.html>

health workers incessantly seize opportunities to push death back. The details are complicated but the resultant – the straight line of life-expectancy increase – is simple.

For the world as a whole life expectancy has more than doubled over the past two centuries, from about 25 years to about 65 for men and 70 for women (Riley 2001). This transformation of the duration of life has greatly enhanced the quantity and quality of people's lives. It has fueled enormous increases in economic output and in population size, including an explosion in the number of the elderly (Fogel and Costa 1997; Martin and Preston 1994).

13.1 Better Forecasts

Although students of mortality eventually recognized the reality of improvements in survival, they blindly clung to the ancient notion that under favorable conditions the typical human has a characteristic lifespan, the Biblical three score and ten. As the expectation of life rose higher and higher, most experts were unable to imagine it rising much further. They envisioned various biological barriers and practical impediments. The notion of a fixed lifespan evolved into a belief in a looming limit to life expectancy.

Continuing belief in imminent limits is distorting public and private decision-making. Forecasts of the expectation of life are used to determine future pension, health-care and other social needs. Increases in life expectancy of a few years can produce large changes in the numbers of the old and very old, substantially augmenting these needs. The officials responsible for making projections – at the United Nations, the World Bank, and various national bureaus – recalcitrantly insist that life expectancy will increase slowly and not much further. The official forecasts distort people's decisions about how much to save and when to retire. They give politicians license to postpone painful adjustments to social-security and medical-care systems (Vaupel 2000).

Officials charged with forecasting trends in life expectancy over future decades should base their calculations on the empirical record of mortality improvements over a corresponding or even longer span of the past⁴ (Lee and Carter 1992; Alho 1998; Tuljapurkar et al. 2000; Wilmoth 1998; Olshansky et al. 2001; Lee 2001). Because best-practice life expectancy has increased linearly by two and a half years per decade for a century and a half, one reasonable scenario would be that this trend will continue in coming decades. If so, record life expectancy will reach 100 in about six decades. This is far from immortality: modest annual increments in life

⁴Olshansky et al. (2001) use changes in age-specific probabilities of death over the decade from 1985 to 1995 to make long-term projections, one out to the year 2577. It is more appropriate to base long-term projections on long-term historical data and to use changes in central death rates. See Wilmoth (1998); Notestein et al. (1944); Lee (2001).

expectancy will never lead to immortality. It is striking, however, that centenarians may become commonplace within the lifetimes of people alive today.

In all countries except the record holder, female life expectancy will be shorter than the best-practice level. Life expectancy could be estimated by forecasting the gap. The U.S. disadvantage varied from a decade in 1900 to less than a year in 1950 and about 5 years in 2000. If the trend in record life expectancy continues and if the U.S. disadvantage is between a year and a decade in 2070, then female life expectancy would be between 92.5 and 101.5, considerably higher than the U.S. Social Security Administration's forecast of 83.9.

An alternative method for forecasting life expectancy is to compute the average rapidity of improvement in age-specific death rates over many decades and then to use this information to project death rates over coming decades⁵ (Lee and Carter 1992; Alho 1998; Tuljapurkar et al. 2000; Wilmoth 1998; Olshansky et al. 2001; Lee 2001). In the early 1940s, when he was a student at Princeton University, the eminent demographer Ansley Coale developed and applied a version of this method (Notestein et al. 1944). Today vastly superior data resources are available⁶ and powerful, practicable methods have been developed to do more than Coale attempted (see e.g. Lee and Carter 1992; Alho 1998; Tuljapurkar et al. 2000). These methods use information about fluctuations in the speed of change in the past to estimate confidence bounds for the uncertainty enveloping life expectancy in the future. The official Japanese forecast, issued in 1997, for life expectancy (for males and females combined) in the year 2050 is 82.95 years (NIPSSR 1997). Projections based on the pattern of reductions in death rates in Japan since 1950 result in a life expectancy some 8 years longer, 90.91 years, with a 90% confidence range from 87.64 to 94.18 years (Tuljapurkar et al. 2000).

Progress in reducing mortality might be systematically slower than in the past. Officials could produce low life-expectancy scenarios to capture this eventuality. Then, however, they should also publish high scenarios that recognize that biomedical research may yield unprecedented increases in survival. Given the extraordinary linearity of the increase in best-practice life expectancy and given the ludicrous record of specious life-expectancy limits, the central forecast should be based on the long-term trend of sustained progress in reducing mortality.

13.2 Continuing Belief in Looming Limits

Faith in proximate longevity limits endures, sustained by *ex cathedra* pronouncements and mutual citations. In their quest to impose a cap on average longevity, students of mortality ignored essential research questions. Major changes in life expectancy hinge on improvements in survival at advanced ages, but comprehensive

⁵See footnote 4 above.

⁶See footnote 1 above.

analysis of the remarkable reductions since the mid-twentieth century in death rates after age 80 first flourished in the 1990s (Kannisto et al. 1994; Kannisto 1996; Vaupel 1997; Wilmoth et al. 2000; Vaupel et al. 1998). Hypothesized biological barriers to longer lifespans also first received systematic attention (and refutation) a decade ago (Vaupel et al. 1998; Carey et al. 1992; Curtsinger et al. 1992; Wachter and Finch 1997; Carey and Judge 2001). The impact of continuing mortality improvements on life expectancy attracted empirical and theoretical attention in the late 1980s, with refined methods developed over the past decade (Lee and Carter 1992; Alho 1998; Tuljapurkar et al. 2000; Vaupel 1986; Vaupel and Canudas Romo 2000). It now appears plausible that life expectancy in several post-industrial countries may approach or exceed 90 by the middle of this century (Tuljapurkar et al. 2000; Wilmoth 1998) and that half the girls born today in countries such as France and Japan may become centenarians (Vaupel 1998, 1997).

If the expectation of life in developed countries were approaching an imminent maximum, then the pace of improvement in mortality in the countries with the highest life expectancies would be slower than the pace in countries with shorter life expectancies. There is, however, no correlation between the level of life expectancy and the pace of improvement (Kannisto et al. 1994; Wilmoth 1997). Indeed, in the current life-expectancy leader, Japan, death rates are falling exceptionally rapidly. Furthermore, as life expectancy rose over the course of the twentieth century, the pace of mortality improvement at older ages *accelerated* (Kannisto et al. 1994; Kannisto 1996; Vaupel 1997, Wilmoth et al. 2000; Wilmoth 1997). Even after age 100, death rates are falling (Kannisto 1996; Vaupel 1997; Wilmoth et al. 2000). Female life expectancy is higher than the male level in long-lived countries, but female life expectancy is increasingly somewhat more rapidly (Kannisto et al. 1994; Wilmoth 1997).

Olshansky et al. (2001) emphasize a theoretical barrier: “entropy in the life table means that small but equal incremental gains in life expectancy require progressively larger reductions in mortality. . . . Projections based on biodemographic principles that recognize the underlying biology within the life table would lead to more realistic forecasts of life expectancy that reflect the demographic reality of entropy in the life table.” Entropy in the life table is merely the statistic

$$\int s(a, t) \ln s(a, t) da / \int s(a, t) da,$$

where $s(a, t)$ is the probability of surviving from birth to age a at age-specific death rates prevailing at time t . Contrary to Olshansky et al.’s claim, in countries with long life expectancies a continuing rate of decline in age-specific death rates of N percent per year will increase life expectancy at birth by about N years per decade (Vaupel 1986; Vaupel and Canudas Romo 2000). Note that steady *rates* of change in mortality levels produce steady *absolute* increases in life expectancy: this relationship may underlie the linear trend of record life expectancy. In any case, valid biodemographic principles impose no insurmountable barriers to longer lives

(Vaupel et al. 1998; Carey et al. 1992; Curtsinger et al. 1992; Wachter and Finch 1997; Carey and Judge 2001).

In sum, the past decade of mortality research has refuted the empirical misconceptions and purported theories that underlie the belief that the expectation of life cannot rise much further. In this article we have added a further line of cogent evidence. If life expectancy were close to its maximum, then the increase in the record expectation of life should be slowing. It is not. For 160 years, best-performance life expectancy has steadily increased by a quarter of a year per year, an extraordinary constancy of human achievement.

References

- Alho, J. (1998). A stochastic forecast of the population of Finland. *Statistics Finland Reviews*, 4(1). Helsinki: Tilastokeskus.
- Bourgeois-Pichat, J. (1952). Essai sur la mortalité biologique de l'homme. *Population*, 7(3), 381–394.
- Bourgeois-Pichat, J. (1978). Future outlook for mortality decline in the world. *Population Bulletin of the United Nations*, 11, 12–41.
- Carey, J. R., & Judge, D. S. (2001). Principles of biodemography with special reference to human longevity. *Population: An English Selection*, 13, 9–40.
- Carey, J. R., Liedo, P., Orozco, D., & Vaupel, J. W. (1992). Slowing of mortality rates at older ages in large medfly cohorts. *Science*, 258, 457–461.
- Cruijsen, H., & Eding, H. (2001). Latest national mortality forecasts in the European Union. In E. Taber, A. van den Berg Jeths, & C. Heathcote (Eds.), *Forecasting mortality in developed countries* (pp. 227–245). Dordrecht: Kluwer Academic.
- Curtsinger, J. W., Fukui, H. H., Townsend, D., & Vaupel, J. W. (1992). Demography of genotypes: Failure of the limited life-span paradigm in *Drosophila melanogaster*. *Science*, 258, 462–463.
- Demeny, P. (1984). A perspective on long-term population growth. *Population and Development Review*, 10(1), 103–126.
- Dublin, L. I. (1928). *Health and wealth*. New York: Harper.
- Dublin, L. I., & Lotka, A. J. (1936). *Length of life: A study of the life table*. New York: Ronald Press.
- Fogel, R. W., & Costa, D. L. (1997). A theory of technophysio evolution, with some implications for forecasting population, health care costs, and pension costs. *Demography*, 34(1), 49–66.
- Frejka, T. (1973). *The future of population growth*. New York: Wiley.
- Frejka, T. (1981). Long-term prospects for world population growth. *Population and Development Review*, 7(3), 489–511.
- Fries, J. F. (1980). Aging, natural death, and the compression of morbidity. *New England Journal of Medicine*, 303, 130–135.
- Fries, J. F. (1990). Compression of morbidity: Near or Far? *Milbank Quarterly*, 67, 208–232.
- Human Mortality Database*. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at: <http://www.demog.berkeley.edu/wilmoth/mortality>
- Kannisto, V. (1996). *The advancing frontier of survival*. Odense: Odense University Press.
- Kannisto, V., Lauritsen, J., Thatcher, A. R., & Vaupel, J. W. (1994). Reductions in mortality at advanced ages: Several decades of evidence from 27 countries. *Population and Development Review*, 20(4), 793–810.

- Keilman, N. (1997). Ex-post errors in official population forecasts in industrialized countries. *Journal of Official Statistics*, 13(3), 245–277.
- Keyfitz, N., & Flieger, W. (1968). *World population: An analysis of vital data*. Chicago: University of Chicago Press.
- Lee, R. D. (2000). Long-term population Projections and the US social security system. *Population and Development Review*, 26(1), 137–143.
- Lee, R. D. (2001). Predicting human longevity. *Science*, 292(5522), 1654–1655.
- Lee, R. D., & Carter, L. (1992). Modeling and forecasting the times series of U.S. mortality. *Journal of the American Statistical Association*, 87(419), 659–671.
- Martin, L. G., & Preston, S. H. (1994). *Demography of aging*. Washington, DC: National Academy Press.
- NIPSSR (National Institute of Population and Social Security Research). (1997). *Population projection of Japan, 1996–2050*. Tokyo: NIPSSR.
- Notestein, F., Taeuber, I. B., Kirk, D., Coale, A. J., & Kiser, L. K. (1944). *The future population of Europe and the Soviet Union* (Vol. 5, pp. 183–189). Geneva: League of Nations.
- Oeppen, J., & Vaupel, J. W. (2002). Broken limits to life expectancy. *Science*, 296(5570), 1029–1031.
- Olshansky, S. J., Carnes, B. A., & Cassel, C. (1990). In search of methuselah: Estimating the upper limits to human longevity. *Science*, 250(4981), 634–640.
- Olshansky, S. J., Carnes, B. A., & Désesquelles, A. (2001). Prospects for human longevity. *Science*, 291(5508), 1491–1492.
- Preston, S. H. (1974). Evaluation of postwar mortality projections in the United States, Canada, Australia, New Zealand. *World Health Statistics Report*, 27(1), 719–745.
- Riley, J. (2001). *Rising life expectancy: A global history*. Cambridge: Cambridge University Press.
- Siegel, J. S. (1980). Recent and prospective demographic trends for the elderly population and some implications for health care. In S. G. Haynes & M. Feinlied (Eds.), *Second conference on the epidemiology for aging* (NIH Publ. No. 80–969) (pp. 17–82). Bethesda: US Department of Health and Human Services.
- Tuljapurkar, S., Li, N., & Boe, C. (2000). A universal pattern of mortality decline in the G7 countries. *Nature*, 405(June), 789–792.
- United Nations. (1973). *World population prospects as assessed in 1968*. New York: UN.
- United Nations. (1985). *World population prospects as assessed in 1984*. New York: UN.
- United Nations. (1989). *World population prospects 1988*. New York: UN.
- United Nations. (1999). *World population prospects 1998*. New York: UN.
- United Nations. (2001). *World population prospects 2000*. New York: UN.
- Vaupel, J. W. (1986). How change in age-specific mortality affects life expectancy. *Population Studies*, 40(1), 147–157.
- Vaupel, J. W. (1997). The average French baby may live 95 or 100 years. In J. M. Robine et al. (Eds.), *Longevity: To the limits and beyond* (pp. 11–28). Berlin: Springer.
- Vaupel, J. W. (1998). Demographic thinking. *Science*, 280(5366), 986.
- Vaupel, J. W. (2000). Setting the stage: A generation of centenarians? *The Washington Quarterly*, 23(3), 197–200.
- Vaupel, J. W., & Canudas Romo, V. (2000). How mortality improvement increases population growth. In E. J. Docke et al. (Eds.), *Optimization, dynamics, and economic analysis: Essays in honor of Gustav Feichtinger* (pp. 345–352). Physica Verlag: Heidelberg/New York.
- Vaupel, J. W., Carey, J. R., Christensen, K., Johnson, T. E., et al. (1998). Biodemographic trajectories of longevity. *Science*, 280, 855–860.
- Wachter, K. W., & Finch, C. E. (Eds.). (1997). *Between zeus and the salmon: The biodemography of longevity*. Washington, DC: National Academy Press.
- Whelpton, P. K., Eldridge, H. T., & Siegel, J. S. (1947). *Forecasts of the population of the United States 1945–1975*. Washington, DC: US. Bureau of the Census.

- Wilmoth, J. W. (1997). In search of limits. In K. W. Wachter & C. E. Finch (Eds.), *Between zeus and the salmon: The biodemography of longevity* (pp. 38–64). Washington, DC: National Academy Press.
- Wilmoth, J. R. (1998). The future of human longevity: A demographer's perspective. *Science*, 280 (5362), 395–397.
- Wilmoth, J. R., Deegan, L. J., Lundstrom, J., & Horiuchi, S. (2000). Increase of maximum life-span in Sweden, 1961–1999. *Science*, 289, 2366–2368.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 14

Mortality Forecasts and Linear Life Expectancy Trends



Ronald Lee

14.1 Introduction

Two important articles on aggregate mortality trends were published in the spring of 2002, with important implications for our perspective on modeling, forecasting, and interpreting mortality trends. One such article was Oeppen and Vaupel (2002, henceforth OV), which shows a remarkable linear trend in the female life expectancy (at birth, period basis) of the national population with the highest value for this variable from 1840 to 2000. Of course the set of nations reporting credible life expectancy values has greatly expanded over this period, but that is unlikely to have mattered much for the results. Over this entire 160-year period, the record life expectancy consistently increased by 0.24 years of life per calendar year of time, or at the rate of 24 years per century. Extrapolation would lead us to expect a female life expectancy of around 108 years at the end of the twenty-first century.

A closely related article by White (2002) finds a linear trend in sexes-combined life expectancy for 21 industrial nations from 1955 to 1995, with an increase of 0.21 years of life per calendar year. White also finds that a linear trend in life expectancy gives a better fit to the experience of almost all the individual countries than does a linear trend in the age-standardized death rate, or the log of the age standardized death rate. He also found that when a quadratic time trend was fitted to the standardized rates, the coefficient on the squared term was significantly positive, indicating that the rate of improvement has been accelerating.

Research for this paper was funded by a grant from NIA, R37-AG11761. Tim Miller carried out all computations for this paper. Monique Verrier provided editorial assistance. I am grateful to workshop participants for helpful comments on an earlier draft.

R. Lee (✉)

Demography and Economics, University of California, Berkeley, CA, USA
e-mail: rlee@demog.berkeley.edu

Both OV and White discuss the processes of catch-up and convergence. OV notes that some countries converge toward the leader (e.g. Japan), some have moved away from it (e.g. the US in recent decades), and some move more or less parallel to it. White finds that nations experience more rapid e_0 gains when they are farther below the international average, and conversely, and therefore tend to converge toward the average. The variance across countries has diminished markedly over the forty years. However, there has been no tendency for the rate of increase of average e_0 to slow down. Based on the current position of the US, which is somewhat below the average (just as OV shows that the US is below the record line), White predicts that e_0 will grow a bit more rapidly than the average rate of 0.21 years per year, perhaps at 0.22 years per year. At this rate, the US would reach $e_0 = 83.3$ in 2030—about 1.5 years above the Lee-Carter (1992) forecast, and about 3.8 years above the Social Security Administration (2002) projection for that year. Extrapolation of the linear trend in either OV or White generates more rapid gains in future longevity than are foreseen by Lee-Carter (1992, henceforth LC), which projects increases of 0.144 years per year between now and 2030. This is only two-thirds as fast as 0.22 years per year in White, and 0.23 years per year in OV (averaging the female and male rates for OV).

Two major points are made in both articles. First, life expectancy (record or average) appears to have changed linearly over long periods of time. Second, national mortality trends should be viewed in a larger international context rather than being analyzed and projected individually. In this paper I will discuss both these points, and conclude with suggestions for incorporating them in forecasting methods. I will draw on the Human Mortality Database or HMD (at <http://www.mortality.org/>), to fit various models.

Figure 14.1 plots the OV maximum life expectancy together with that of the HMD and we see that they sometimes coincide, and sometimes the OV record exceeds the HMD, which includes fewer countries.

14.2 Linear Change in Life Expectancy over Long Historical Periods

Before reading the OV article, I had expected that the trajectory of record life expectancy over the past two centuries would have a tilted S shape, in which life expectancy began at first to increase slowly, then accelerated, and then decelerated in the second half of the twentieth century. If we go back far enough in time, we know that life expectancy had no systematic trend at all, although there might have been long fluctuations. We also can be pretty sure that initial gains in life expectancy, once the trend began, were slow. Based on the OV results, it appears that these portions of the history occurred out of our sight, before the start date of 1840. Indeed, Fig. 5 in the OV Supplementary Materials on the Science Web site plots English life expectancy over a longer period, and its trajectory conforms to this description.

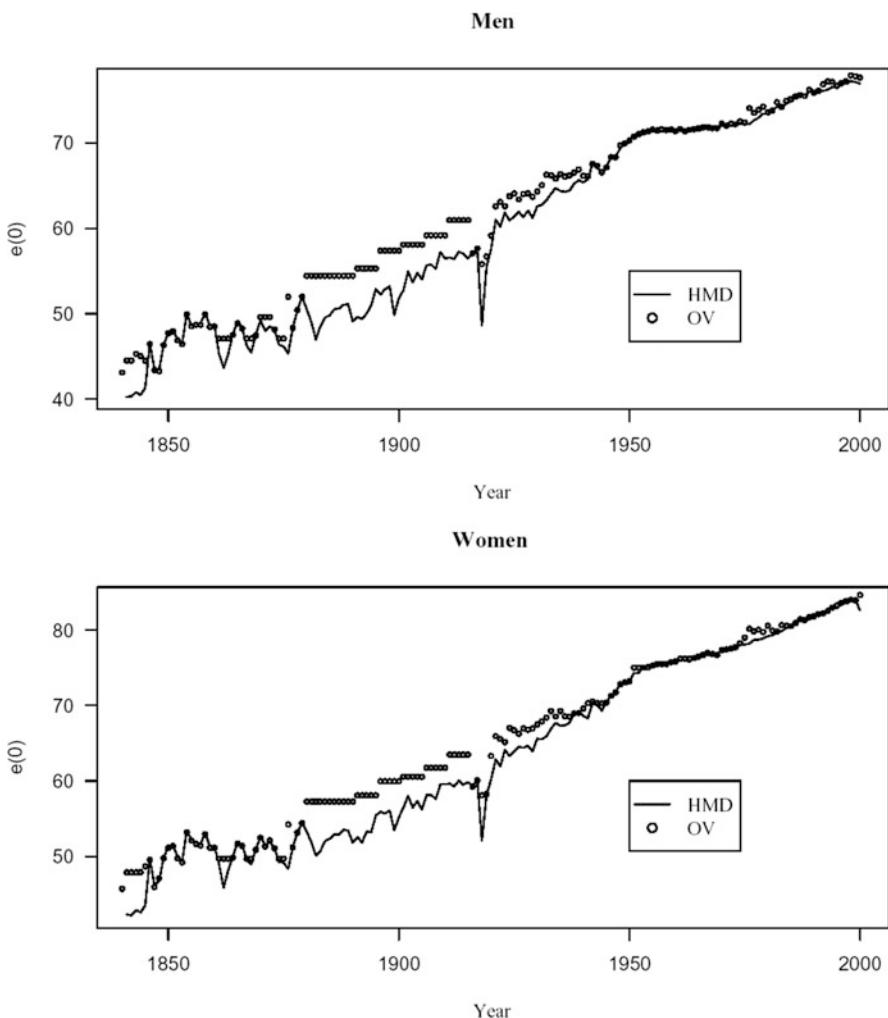


Fig. 14.1 Record life expectancy, by sex, from Oeppen-Vaupel and the Human Mortality Data-base, 1840 to 2000

OV do not actually test or explore the constancy of the slope for record life expectancy, so it is worth examining this point more carefully here. As a start, we can compute the average rates of life expectancy increase for the OV data by sex and sub-period, as follows:

Average Annual Rates of Decline of Record e_0 By Subperiod

	Females	Males	Average
1840–1900	0.24	0.24	0.24
1900–1950	0.27	0.26	0.27
1950–2000	0.23	0.15	0.19

From this we see that the regularity of the linear decline is not quite as strong as it appears from the striking figure in OV. For males in particular, there has been a noticeable deceleration over the past 50 years. For both sexes, there is a hint of the S shaped path that I had expected to see.

I have taken two more simple steps. First, I fitted a cubic polynomial to the data, and found that all three terms were significantly different than zero. The fitted curve, as shown in Fig. 14.2 for females, does have a slight S-shape. To see more clearly the implied rate of change, I plot the first derivative of the polynomial for females in Fig. 14.3. This suggests that the rate of change in fact increased substantially, more than doubling from 1840 to 1925 or so, and then substantially declining again thereafter, challenging the linear interpretation of the OV plot. Second, I calculated a 25-year moving average of the annual pace of increase for females, and this also is plotted in Fig. 14.3.

This less severe smoothing of the rate of change cautions us against drawing any firm conclusions from the data about linearity or nonlinearity. A case could be made for either.

If we accept that the OV trajectory is strikingly close to linear, then we are led to ponder why the record life expectancy might have risen in this way. After considerable thought, I find I have little useful to contribute on this important question. I find I am equally unable to explain the relative constancy of age-specific proportional rates of mortality decline, as summarized by the trend in the Lee-Carter (1992) k for the US since 1900, and the G7 countries since 1950 (Tulapurkar et al. 2000).

Of the two striking regularities, linear life expectancy trends and constant rate of decline of age-specific mortality, it is the linearity of life expectancy increase which I find most puzzling. In my mind, the risks of death (that is, the force of mortality or death rates, by age) are the fundamental aspect of mortality which we should model and interpret. One view, perhaps an incorrect view, is that period life expectancy is just a very particular and highly nonlinear summary measure, with little or no causal significance. If age-specific death rates (ASDRs) decline at constant exponential rates, then life expectancy will rise at a declining rate, at least for a long time.

This point is worth elaborating because OV, in the Supplementary Materials on the Science Web site, say: “Note that steady rates of change in mortality levels produce steady, absolute increases in life expectancy: This relationship may underlie the linear trend of record life expectancy.” I agree that ultimately, it is likely that life expectancy would rise linearly, once death rates below the ages which obey Gompertz’s Law have fallen to near zero, as Vaupel (1986) has pointed out. If θ is the Gompertz parameter (rate of increase of mortality with age in a period life table or cohort life table) and ρ is the annual rate of decline over time in mortality at all ages above, say, 50, then the rate of increase of e_{50} will be ρ/θ years per year (Vaupel 1986). However, there is substantial mortality at younger ages before Gompertz’s Law applies, particularly in the nineteenth century. There we would expect a “steady” rate of decline in death rates to lead to a declining rate of increase in life expectancy.

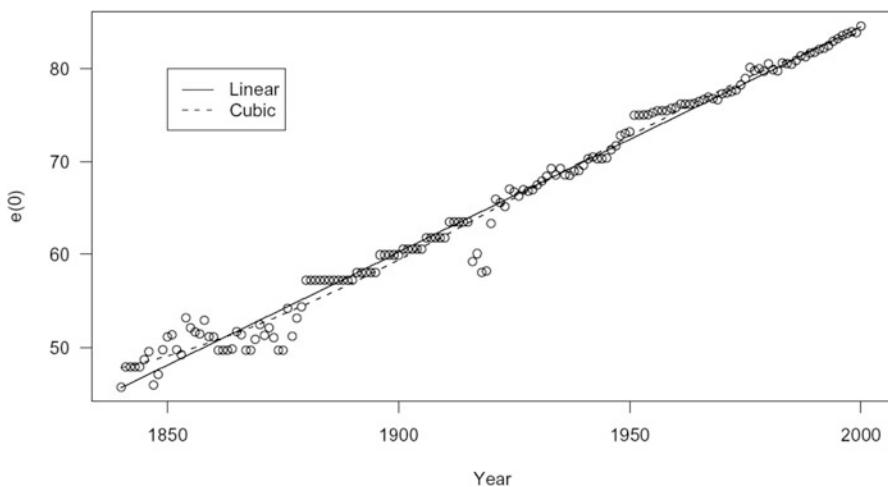


Fig. 14.2 Linear and cubic trends fitted to the Oeppen-Vaupel record female life expectancy, 1840–2000

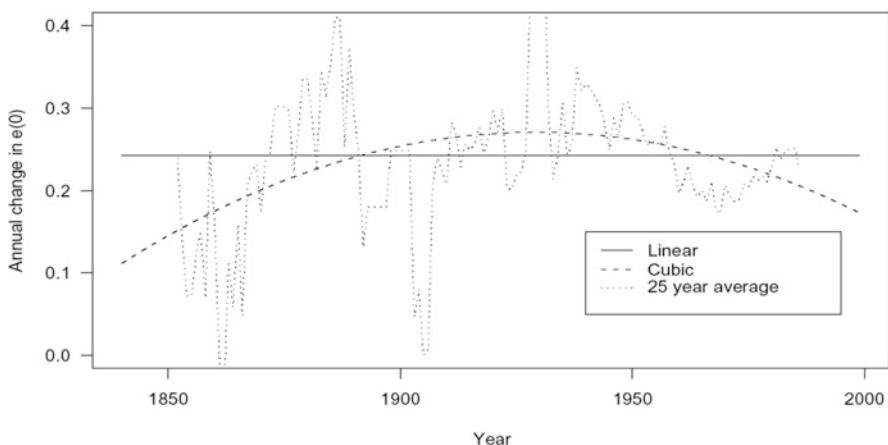


Fig. 14.3 Rate of change in female life expectancy calculated from linear and cubic fits to Oeppen-Vaupel record and 25-year moving average of change in record

These points are illustrated in Fig. 14.4, based on Swedish mortality experience. The average exponential rate of decline is calculated for each age-specific death rate for the period 1861 to 1961. This rate of decline is then applied to the initial age-specific death rates, and used to simulate them forward for 200 years. The resulting life expectancy is plotted in Fig. 14.3 along with the actual life expectancy. It can be seen that the simulated life expectancy trajectory is highly nonlinear, and its pace of improvement decelerates.

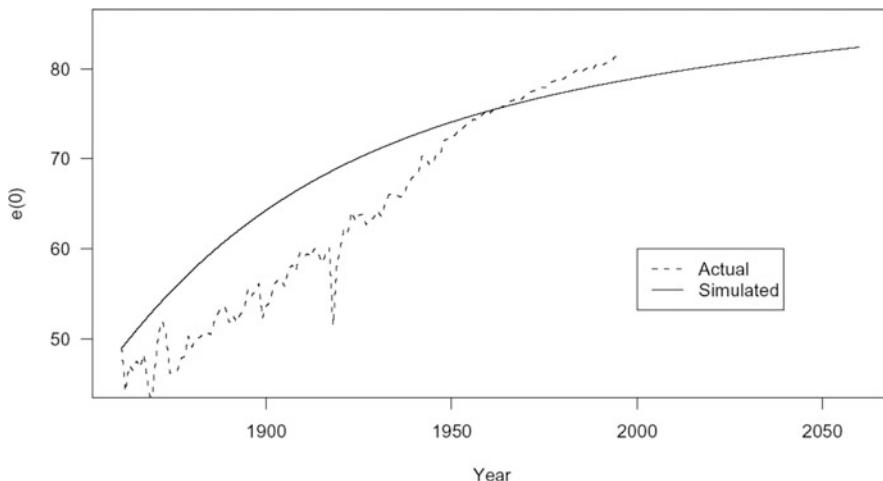


Fig. 14.4 Actual and simulated Swedish female life expectancy assuming constant proportional rates of decline for age-specific death rates, at average rates for 1861–1961

As time passes, the gains in life expectancy become more nearly linear, and for the last 50 years, are quite close to linear. By construction, the lines cross in 1961. Figure 14.4 shows that the constant exponential rates of decline in age-specific death rates could not account for the linearity of the increase in record e_0 since 1840.

When we look at the trajectories of the logs of the Swedish ASDRs from 1861 to 2000, they appear very far from linear, even if we restrict attention to the last 50 years, see Fig. 14.5 for selected rates. Most rates decline rapidly in some periods, and slowly in others, with patterns varying across the age span. One would not think to characterize these patterns as showing a constant rate of decline at each age. Yet this is a period over which the Lee-Carter model does a good job of fitting life expectancy, and projecting it within sample (Lee and Miller 2001). Evidently, the Lee-Carter method succeeds by picking out average tendencies from among a welter of variation, not by describing strong real-world regularities.

14.3 What Is Fundamental, Age at Death or Risk of Death?

The OV and White findings challenge the view that risks of death are fundamental, and age at death is derivative. If life expectancy (e_0) changes linearly, then rate of decline of death rates must be nonlinear, and in particular must be accelerating for at least some ages, as found by White for many of the 21 countries he analyzed. How can we reconcile the linearity of the change in e_0 with the fact that when LC models are fit, they have almost always revealed linear changes in k over rather long periods, such as a century in the US? To focus on the US case, there are two explanations. First, as the second figure in OV makes clear, over the twentieth century the US first

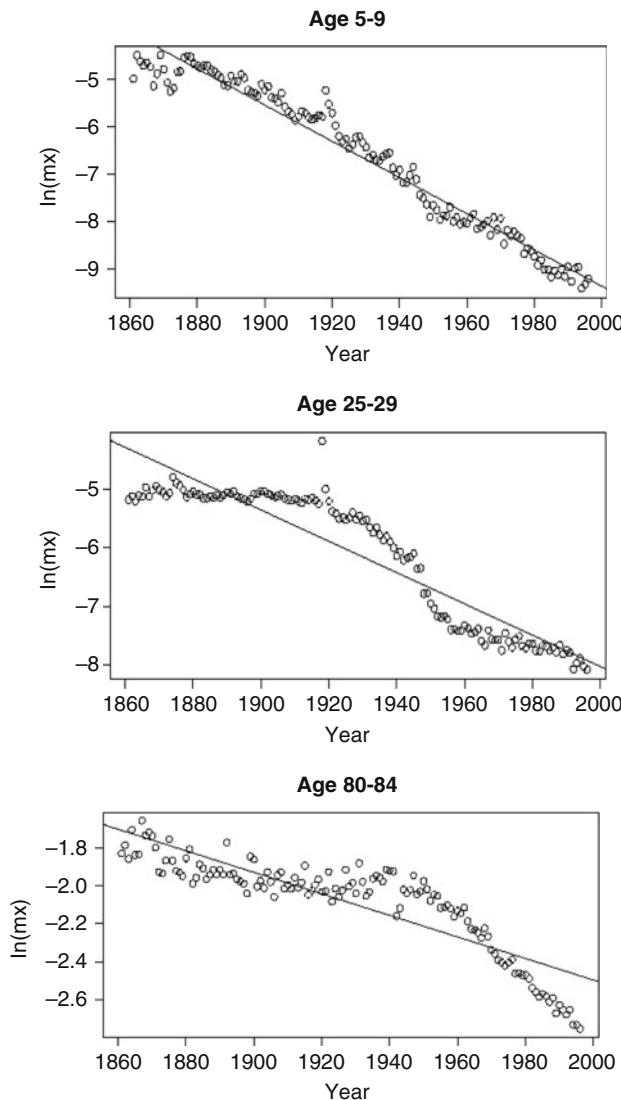


Fig. 14.5 Log of selected age-specific death rates for Swedish females, 1861–2000, showing irregular rates of decline

approached the record line, then briefly was close to being the leader, and finally fell away from the line starting in the 1960s. (This falling away very likely reflects the relatively early uptake of smoking in the US.) Since the trajectory of US e_0 in fact had the shape we would expect with a constant rate of decline in ASDRs, perhaps there is no puzzle to explain for the US case. But can the same story hold for all the G7 countries analyzed and projected by Tuljapurkar et al. (2000)? This brings us to

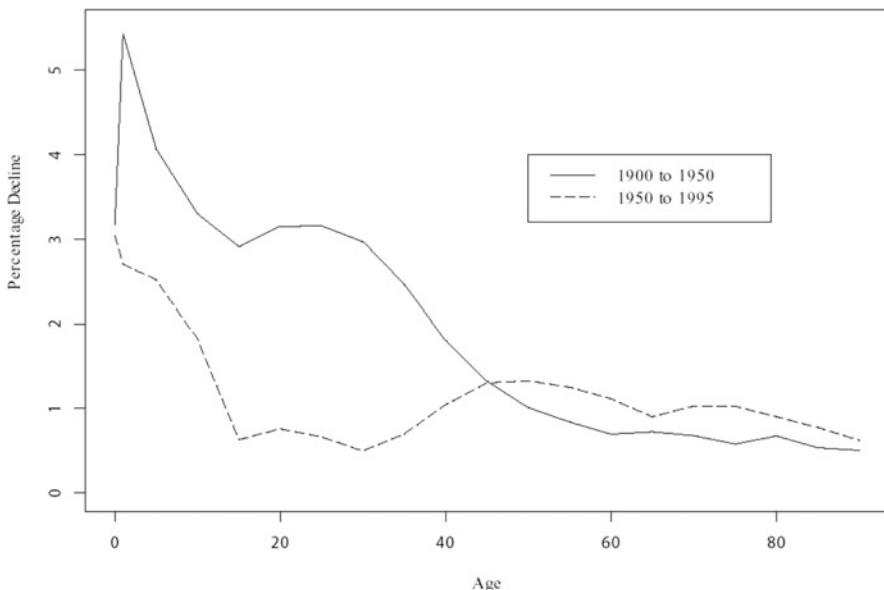


Fig. 14.6 Average annual reductions in age-specific death rates, US (sexes combined), showing the changing age pattern of decline

the second explanation, which is that contrary to the LC assumptions, the rates of decline have not been constant for each age, which is to say that the LC b_x coefficients have not been constant over the sample period. Instead, they have changed shape between the first half of the century, when the mortality decline was much more rapid for the young than for the old, and the second half, when there is little difference among the rates of decline above age 20 or so. Just when the ASDRs of the young became so low that their further decline could contribute little to increasing e_0 , the rates of decline at the older ages began to accelerate, as noted by Horiuchi and Wilmoth (1998). This tilting of the b_x schedule has meant that a given rate of decline of k can produce more rapid rates of increase in e_0 than would have been the case with the old b_x schedule. The tilting of the b_x schedule is shown for the US in Fig. 14.6, and for Sweden, France, Canada, and Japan in Fig. 14.7. In each case the annual rate of decline for mortality is plotted by age for the first and second halves of the twentieth century, except for Japan, for which the break point is 1975.

14.4 Using These Findings to Improve Mortality Forecasts

The first question is whether or not we should expect record e_0 to continue to increase at this rate in the future, and if so for how long? Since I do not understand why this linearity has occurred in the past, I have no reason to think it should, or

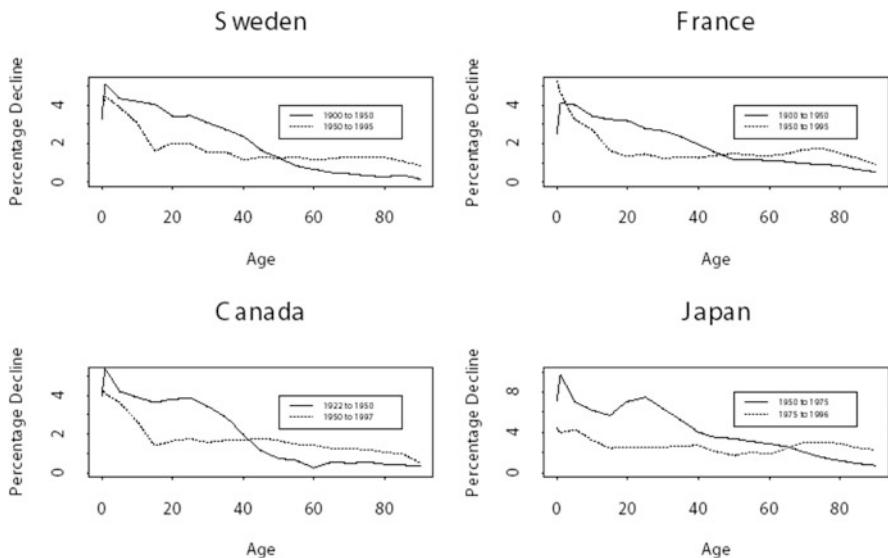


Fig. 14.7 Average annual reductions in age-specific death rates, selected low mortality countries (sexes combined), showing the changing age pattern of decline

should not, continue in the future. The regularity in the past invites the forecaster to assume it will continue in the future, at least for a while. Suppose then that we do assume it will continue. How can we use that assumption to mold our forecasts? This line of thinking leads us unavoidably to consider national mortality change in an international context, to which we now turn.

14.5 Considering National Mortality Change in an International Context

Let $E(t)$ be the best-practice life expectancy at time t . It is imperfectly estimated by the OV record series. The White average e_0 measure reflects a different concept. Let $e_i(t)$ be actual life expectancy at birth for country i in year t . I will consider a number of possible kinds of models describing the relation between changes in $e_i(t)$ and $E(t)$. I will write the equations in continuous time, but they are readily rewritten for discrete annual changes for purposes of estimation.

First Category of Models: All Countries Are Structurally Similar, But Start at Different Levels

$$de_i(t)/dt = \phi + \alpha(E(t) - e_i(t)) + \varepsilon_i(t) \quad (14.1)$$

Table 14.1a Estimated rate of convergence of national life expectancy to Oeppen-Vaupel record level in 18 countries of the Human Database (Eq. 14.1)

	Females, 1841–1999	Females, 1900–1999	Males, 1841–1999	Males, 1900–1999
O-V record gap (α)	0.0615** [0.0137]	0.0683** [0.0162]	0.0777** [0.0171]	0.0802** [0.0195]
Constant	0.0160 [0.0723]	0.0308 [0.0783]	-0.1368 [0.0918]	-0.1134 [0.0997]
Observations	1332	1155	1332	1155
Number of countries	18	18	18	18
Rho	-0.109	-0.126	-0.064	-0.076
R-squared	0.037	0.045	0.043	0.045

Estimates are based on panel corrected SE using Prais-Winsten regression (assuming first-order autocorrelation. SE of the coefficients are in brackets

**Significant at 1%

Here, life expectancy tends to increase at some constant rate ϕ , and in addition it tends to move a proportion α toward the best practice level (record level) $E(t)$ each year. It is also subject to a disturbance ε which could move it toward or away from this trajectory. This specification is consistent with the equation estimated by White. In estimation, I allow the $e_i(t)$ for each country to be autocorrelated ($e_i(t) = \rho e_i(t-1) + \eta_i(t)$) with all countries sharing the same autocorrelation coefficient ρ .

I fit this and later models to life expectancy series for 18 countries with relatively low mortality, with data drawn from the Human Mortality Database (HMD) at <http://www.mortality.org/>. The data series are of varying historical depth, with the shortest covering 29 years and the longest 159 years. Models are fit using an unbalanced design, so that the full range of data could be exploited. However, the estimation range is sometimes restricted to the period since 1900.

Table 14.1a reports estimates of α for females and for males, based on model specifications with and without autocorrelated errors, and using the OV record.

In all cases α is highly significantly different than 0, with values lying between 0.06 and 0.08, indicating a tendency for the life expectancy of the countries to converge towards the leader country. The half-life of a deviation from the record level is around 10 years ($e^{-10*0.07} = 0.5$). Here and throughout, results are very similar if the equation is estimated with no constant, so that the only source of life expectancy increase is catching up with the leader, or if there is no allowance for autocorrelated errors. Note that the R^2 is low at around 0.04, and that the estimated autocorrelation is negative, which is somewhat surprising. Table 14.1b is the same, except that it uses the HMD record life expectancy in place of OV. The results are also very similar, but with a slightly slower rate of convergence and lower R^2 .

Rather than taking the actual record e_0 from OV or HMD as an estimate of the target trajectory toward which life expectancy in all countries is tending, we can instead estimate the implicit unobserved target as part of fitting the model, as in the following equation:

Table 14.1b Estimated rate of convergence of national life expectancy to the highest level in the 18 countries of the Human Database (Eq. 14.1)

	Females, 1841–1999	Females, 1900–1999	Males, 1841–1999	Males, 1900–1999
HMD record gap (α)	0.0506** [0.0117]	0.0454** [0.0130]	0.0681** [0.0163]	0.0666** [0.0180]
Constant	0.1095 [0.0639]	0.1484* [0.0696]	-0.0229 [0.0793]	-0.0024 [0.0884]
Observations	1332	1155	1332	1155
Number of countries	18	18	18	18
Rho	-0.136	-0.153	-0.092	-0.101
R-squared	0.020	0.017	0.030	0.029

Estimates are based on panel corrected SE using Prais-Winsten regression (assuming first-order autocorrelation). SE of the coefficients are in brackets

*Significant at 5%; **Significant at 1%

Table 14.2 Estimated rate of convergence of national life expectancy to an annual implicit target in the 18 countries of the Human Database (Eq. 14.2)

	Females, 1841–1999	Females, 1900–1999	Males, 1841–1999	Males, 1900–1999
Current $e(0)$ (α)	0.0428** [0.0042]	0.0354** [0.0050]	0.0633** [0.0059]	0.0641** [0.0067]
Constant	2.5344** [0.2417]	3.1652** [0.2704]	3.0326** [0.3185]	4.2880** [0.3656]
Observations	1332	1155	1332	1155
Number of countries	18	18	18	18
Rho	-0.131	-0.171	-0.053	-0.045
R-squared	0.607	0.678	0.500	0.511

Estimates are based on panel corrected SE using Prais-Winsten regression (assuming first-order autocorrelation). SE of the coefficients are in brackets

**Significant at 10%

$$de_i(t)/dt = \phi + \gamma_t D_t - \alpha e_i(t) + \epsilon_i(t) \quad (14.2)$$

Here D_t is a period dummy for year t (else 0) and γ_t is its coefficient. γ_t/α gives the target trajectory, playing a role much like the OV record level. Results are reported in Table 14.2 (with estimates of γ not shown, to save space). Because the target is chosen to maximize its explanatory power, the R^2 is now much greater, while rates of convergence, α , are somewhat slower.

Figure 14.8 plots the estimated values of γ_t/α , corresponding to the implicit target trajectory. For comparison the record life expectancy for the HMD is also plotted. We see that the target trajectory lies above the maximum about half the time and also that the target trajectory is highly erratic, possibly with negative autocorrelation.

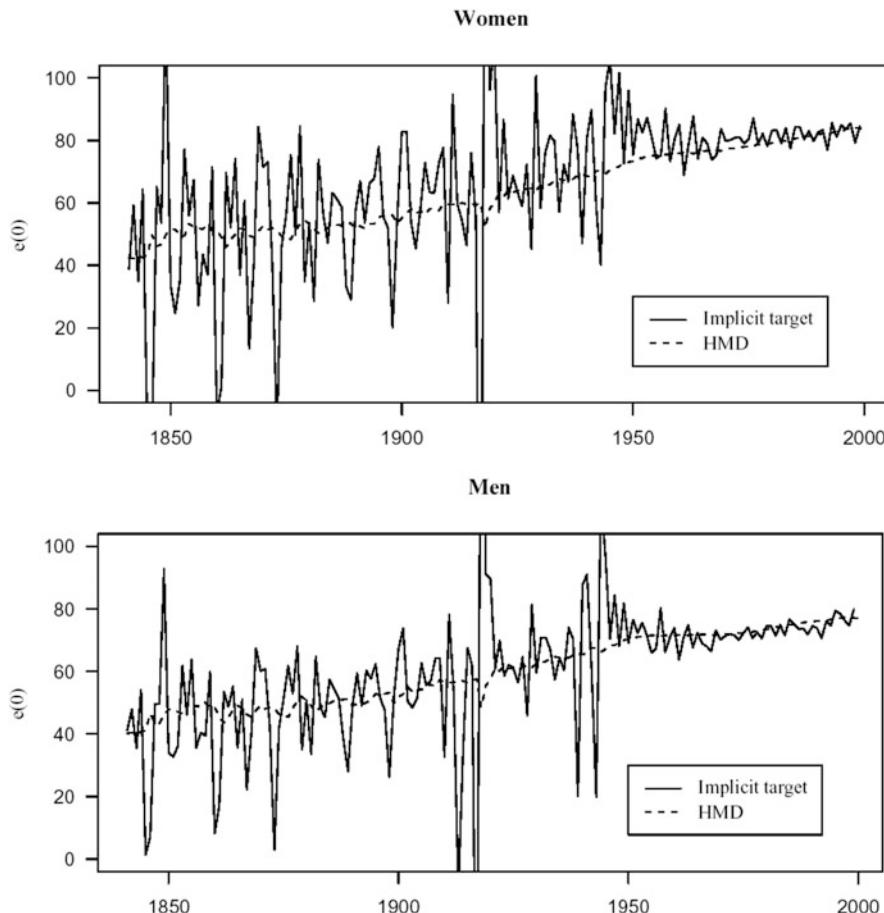


Fig. 14.8 Estimated implicit target of convergence (Eq. 14.2) in the 18 countries of the Human Mortality Database (erratic line), compared to the HMD record life expectancy (smooth line)

When life expectancy is generally above trend, as might happen in a year with a mild winter affecting many countries, for example, the regression will try to fit this by estimating a very high target value, and conversely. This will lead to an underestimate of the size of the convergence coefficient, α . To avoid these problems, it is desirable to impose a smoothness constraint of some kind on the target trajectory. Here I will take the simplest route, assuming that the target trajectory is a linear function of time, leading to the following equation:

$$de_i(t)/dt = \phi + \alpha(a + bt - e_i(t) + \epsilon_i(t)) \quad (14.3)$$

Table 14.3 Estimated rate of convergence of national life expectancy to a linear implicit target in the 18 countries of the Human Database (Eq. 14.3)

	Females, 1841–1999	Females, 1900–1999	Males, 1841–1999	Males, 1900–1999
Current $e(0)$ (α)	0.0704** [0.0138]	0.0586** [0.0153]	0.0865** [0.0166]	0.0845** [0.0184]
Year	0.0191** [0.0037]	0.0129** [0.0042]	0.0204** [0.0039]	0.0168** [0.0044]
Constant	−32.1870** [6.3273]	−20.8345** [7.4209]	−34.1528** [6.6860]	−27.0685** [7.6882]
Calculated: Year/ α	0.271307	0.220137	0.235838	0.198817
Observations	1332	1155	1332	1155
Number of countries	18	18	18	18
Rho	−0.071	−0.118	−0.037	−0.055
R-squared	0.042	0.045	0.052	0.055

Estimates are based on panel corrected SE using Prais-Winsten regression (assuming first-order autocorrelation). SE of the coefficients are in brackets

**Significant at 10%

The results are shown in Table 14.3. The estimated rate of convergence, α , is now slightly higher than in the first set of estimates. The rate of increase of the linear target trajectory is found by dividing the coefficient on “year” by the estimate of α , that is the coefficient on $-e_{i,0}$, which is also given in the table. The rate of increase in the target calculated in this way is slightly higher than for the record for OV or the HMD. For example, the gain per year in target e_0 estimated here for the whole period is 0.271 years per year, while in OV it is 0.243 years per year. Other comparisons are similar.

It is possible that countries that are twice as far from $E(t)$ may not converge twice as quickly. To allow for this, we can add a term that is quadratic in the size of the gap (the quantity in parentheses in Eq. 14.2). A negative coefficient on the quadratic term would indicate that the pace of increase in e_0 is less than proportionate to the size of the gap, and a positive coefficient that it is more than proportionate.

$$de_i(t)/dt = \phi + \alpha(E(t) - e_i(t)) + \beta(E(t) - e_i(t))^2 + \varepsilon_i(t) \quad (14.4)$$

The results of estimating this specification are given in Table 14.4a and 14.4b, and are unambiguous: In every case, the coefficient on the quadratic, β , is highly significantly greater than zero, and the coefficient on the linear term is negative. In order to interpret these coefficients, I show in Fig. 14.9 the derivative of the change in $e_i(t)$ with respect to the size of the gap, $E(t) - e_i(t)$.

Under the linear specification used earlier, this plot would be a straight line with height α . Here, however, we see that all the lines slope decisively upward to the right, indicating that the rate of convergence increases more than proportionately with the size of the gap. The initial negative values most likely reflect the limitations of the quadratic specification, rather than a true tendency of the rate

Table 14.4a Estimated quadratic rate of convergence of national life expectancy to Oeppen-Vaupel record level in the 18 countries of the Human Database (Eq. 14.3)

	Females, 1841–1999	Females, 1900–1999	Males, 1841–1999	Males, 1900–1999
O-V record gap (α)	−0.0372 [0.0338]	−0.0659 [0.0366]	−0.0518 [0.0308]	−0.0586 [0.0334]
(O-V record gap) ² (β)	0.0088** [0.0028]	0.0124** [0.0031]	0.0072** [0.0020]	0.0073** [0.0020]
Constant	0.1871 [0.0971]	0.2635* [0.1036]	0.2132* [0.1013]	0.2648* [0.1103]
Observations	1332	1155	1332	1155
Number of countries	18	18	18	18
Rho	−0.118	−0.153	−0.099	−0.114
R-squared	0.047	0.068	0.074	0.080

Estimates are based on panel corrected SE using Prais-Winsten regression (assuming first-order autocorrelation). SE of the coefficients are in brackets

*Significant at 5%; **Significant at 10%

Table 14.4b Estimated quadratic rate of convergence of national life expectancy to HMD record level in the 18 countries of the Human Database (Eq. 14.3)

	Females, 1841–1999	Females, 1900–1999	Males, 1841–1999	Males, 1900–1999
HMD record gap (α)	−0.0583* [0.0284]	−0.0687* [0.0292]	−0.0611* [0.0269]	−0.0809** [0.0296]
(HMD record gap) ² (β)	0.0119** [0.0029]	0.0125** [0.0029]	0.082** [0.0021]	0.0089** [0.0022]
Constant	0.2466** [0.0756]	−20.8345** [0.0819]	−34.1528** [0.0803]	−27.0685** [0.0911]
Observations	1332	1155	1332	1155
Number of countries	18	18	18	18
Rho	−0.148	−0.174	−0.110	−0.110
R-squared	0.036	0.038	0.065	0.072

Estimates are based on panel corrected SE using Prais-Winsten regression (assuming first-order autocorrelation). SE of the coefficients are in brackets

*Significant at 5%; **Significant at 10%

of change to decline as the gap increases in this low range. Most of the gaps, 90 to 95% of them, are less than 8 years. Only a few fall outside that range, and are subject to the higher sensitivities to the right on the plot. In future work it should be possible to examine the nonlinearity of the response better, drawing on data for Third World countries with higher mortality, but these have not yet been added to the HMD.

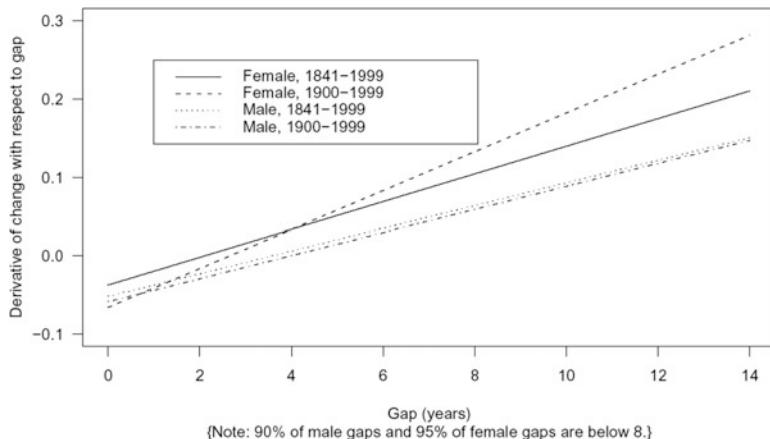


Fig. 14.9 Derivative of quadratic convergence to the Oeppen-Vaupel target: how the proportional effect of a gap increases with the size of the gap

14.6 Extensions

14.6.1 Heterogeneous Targets

If the foregoing models were the whole story, we would expect the life expectancies of countries to be distributed randomly around $E(t)$, since their mortality would have had decades or centuries in which to converge to $E(t)$ under the influences described in the equations. But of course, this is not the case. A more realistic model would take into account the heterogeneity of international experience, by incorporating additional factors that influence the level toward which each country's e_0 converges, which may not be the best practice level. I will call this modified target the idiosyncratic target. We can take it to equal $E(t) + \pi X(t)$, where X is a vector of relevant factors and π is a vector of coefficients. X includes relevant variables such as per capita income, educational attainment, nutritional measures, dietary measures, smoking behavior, and geographic/climatic conditions. πX expresses a deviation from the best practice level. Over time, $E(t)$ rises. If X remained constant, the target level would nonetheless increase with $E(t)$. More likely, πX also increases, indicating an additional source of increase in the target level of e_0 . πX could capture influences like those included in Preston's (1980) analysis, in which he fit socioeconomic models to international cross-sections of life expectancy, and then decomposed gains in life expectancy into movements along the πX curve with economic development, and upward shifts in the whole equation, which would here be reflected in the combination of convergence and a common growth rate, ϕ . The ϵ shocks could reflect political, military, weather, or epidemiological factors of a transitory nature. This model would be:

$$de_i(t)/dt = \phi + \alpha(E(t) + \pi X_{i,t} - e_i(t)) + \varepsilon_i(t) \quad (14.5)$$

Once again, it would be possible to estimate $E(t)$ as part of fitting the model, either unconstrained or constrained to have a linear trajectory. If estimated in this way it will reflect changes in the target net of socioeconomic progress, a concept closer to Preston's residual improvement of life expectancy. Country i will have a target or equilibrium life expectancy in year t of $E(t) + \pi X_{i,t}$, so heterogeneity in equilibria is now incorporated. Countries that are poor, smoke, eat a high cholesterol diet, have low education, or perhaps have a tropical climate, will tend towards lower levels of life expectancy.

14.6.2 Heterogeneous Rates of Convergence

It is also possible that different countries will have different rates of convergence, α . For isolated countries, or perhaps for very poor ones, or ones with very little transportation or communication infrastructure, α may be smaller.

We can take this into account by making α a function of a set of variables Z .

$$de_i(t)/dt = \phi + (\alpha + \delta Z_{i,t})(E(t) + \pi X_{i,t} - e_i(t)) + \varepsilon_i(t) \quad (14.6)$$

Z would include factors indicating the degree of integration of country i in the global community, and perhaps other factors bearing on the strength of government and the communications and transportation infrastructure in the country. It might be difficult to identify factors that belonged in Z rather than in X .

14.7 Forecasting Mortality

Let us assume that the linear trend in record or average life expectancy will continue. Then the next steps are straightforward. We use the linear trend to project the record life expectancy (or the target trend that was estimated as part of the convergence model). We will know the current life expectancy for a particular country of interest. We can use the appropriate or preferred equation for de_i/dt to estimate e_0 one year later, and then continue recursively. The projected e_0 will gradually approach the projected linear trend.

This procedure could be improved by using a model version which allowed for some heterogeneity, as in Eqs. (14.5) and (14.6). Not all countries will approach the same trend line, but each should approach a trajectory that is parallel to it. In these specifications, we would also have to consider the advisability of projecting changes in the X and Z variables, and methods for doing so.

The assumption of a pure linear trend could also be questioned, dropping the initial assumption. The central tendency (record, average, or other) could be modeled

as a stochastic time series, and forecasted in that way. That could certainly be done for the γ series, for example.

In general, the approach of forecasting mortality for individual countries in reference to the international context is very appealing, and I believe it is the natural way to go in future work. Whether this approach is applied to life expectancy itself, or to a Lee-Carter type k, or in some other way, will have to be settled by further research. In the meantime, these recent papers, and particularly OV, challenge our current perceptions of mortality change and expectations about future trends.

References

- Horiuchi, S., & Wilmoth, J. R. (1998). Deceleration in the age pattern of mortality at older ages. *Demography*, 35(4), 391–412.
- Human Mortality Database, University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at: <http://www.mortality.org> or <http://www.humanmortality.de>. Data downloaded on 08/27/02.
- Lee, R. D., & Carter, L. (1992). Modeling and forecasting the time series of U.S. mortality. *Journal of the American Statistical Association*, 87(419), 659–671.
- Lee, R. D., & Miller, T. (2001). Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography*, 38, 537–549.
- Oeppen, J., & Vaupel, J. W. (2002). Broken limits to life expectancy. *Science*, 296(5570), 1029–1031.
- Preston, S. (1980). Causes and consequences of mortality declines in less developed countries during the twentieth century. In R. Easterlin (Ed.), *Population and economic change in developing countries*. Chicago: University of Chicago Press.
- Social Security Administration, Board of Trustees. (2002). *Annual report of the board of trustees of the federal Old-Age and Survivors Insurance and Disability Insurance (OASDI) trust funds*. Washington, DC: U.S. Government Printing Office.
- Tuljapurkar, S., Li, N., & Boe, C. (2000). A universal pattern of mortality decline in the G7 countries. *Nature*, 405(June), 789–792.
- Vaupel, J. W. (1986). How change in age-specific mortality affects life expectancy. *Population Studies*, 40(1), 147–157.
- White, K. M. (2002). Longevity advances in high-income countries, 1955–96. *Population and Development Review*, 28(1), 59–76.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 15

Forecasting Life Expectancy: A Statistical Look at Model Choice and Use of Auxiliary Series



Juha M. Alho

15.1 Why Forecast Life Expectancy?

Let $\mu(x, t)$ be the hazard (or force) of mortality in age x at time t . Define $p(x, t)$ as the probability of surviving to age x , under the hazards of time t , or

$$p(x, t) = \exp\left(-\int_0^x \mu(y, t) dy\right).$$

Then, the expectation of the remaining life time in age $x \geq 0$, equals

$$e_x(t) = \int_0^\infty p(x+y, t) dy / p(x, t).$$

These are synthetic period measures, i.e., they are intended to summarize the chances of survival at time t . Life expectancy at birth, $e_0(t)$, is the most frequently used summary measure. Despite their popularity life expectancies are not directly used in cohort-component population forecasting. Instead, proportions of type

$$p(x+1)/p(x) = \exp(-\Lambda_x(t)),$$

where $\Lambda_x(t)$ is the increment of the cumulative hazard in age $[x, x+1]$, are used for proportions of survivors from exact age x to exact age $x+1$. Similarly, in the computation of present values of annuities, for example, a cohort perspective is

J. M. Alho (✉)

University of Joensuu, Joensuu, Finland

e-mail: juha.alho@helsinki.fi

necessary. In that case, the more relevant concept is the remaining life time of a person alive at exact age $x \geq 0$, at time t , which equals

$$c_x(t) = \int_0^\infty \exp\left(-\int_0^y \mu(x+u, t+u) du\right) dy.$$

Since mortality has typically declined, we expect that $e_x(t) \leq c_x(t)$. We note that even if life expectancies $e_x(t)$ have considerable descriptive value, they are of limited direct usefulness in population forecasting.

Taken together the values of $e_x(t)$ do determine the hazards $\mu(x,t)$ for a given t , but if only $e_0(t)$ is known, then infinitely many patterns $\mu(x,t)$'s would produce the same value $e_0(t)$. In special cases, such as a proportional hazards model ($\mu(x,t) = \mu(x)g(t)$ with $\mu(x)$ known) or a log-bilinear model of the Lee-Carter type ($\mu(x,t) = a(x) + b(x)g(t)$ with $a(x)$ and $b(x)$ known), a one-to-one correspondence exists (e.g., Alho 1989). In these cases forecasting $e_0(t)$ leads directly to estimates of age-specific mortality, but the assumption of known multipliers is strong. Given that the multipliers may change over time, it is not clear that this would, in practice, lead to a more accurate forecast of mortality hazards than forecasting the latter directly.

On the other hand, $e_0(t)$ might perform as an “auxiliary measure” if it behaves in a more time-invariant manner (e.g., Törnqvist 1949) than the age-specific series themselves. The recent finding of Oeppen and Vaupel (2002), in which the so-called best-practice life expectancy, i.e., the life expectancy of the country that is the highest at any given time, was shown to have evolved almost linearly for 160 years, points to this possibility. The first purpose of this paper is to establish the empirical relationship of the best-practice life expectancy to country-specific life expectancies in selected industrialized countries, during the latter part of the 1900's. Simple regression techniques will be used. The second purpose is to examine the statistical underpinnings of using best practice life expectancy as an auxiliary series for the prediction of the country-specific life expectancies.

15.2 Changes in Life Expectancy in 19 Industrialized Countries in 1950–2000

Oeppen and Vaupel (2002) show that the best practice life expectancy for females has followed remarkably well ($R^2 = 0.99$) the model:

$$\tilde{e}_0(t) = 45 + (t - 1840)/4,$$

for $t \geq 1840$. Could this “invariant” be used as an auxiliary series to improve accuracy?

To examine this question empirically we have collected data on female life expectancies for 14 European countries, Australia, Canada, Japan, New Zealand,

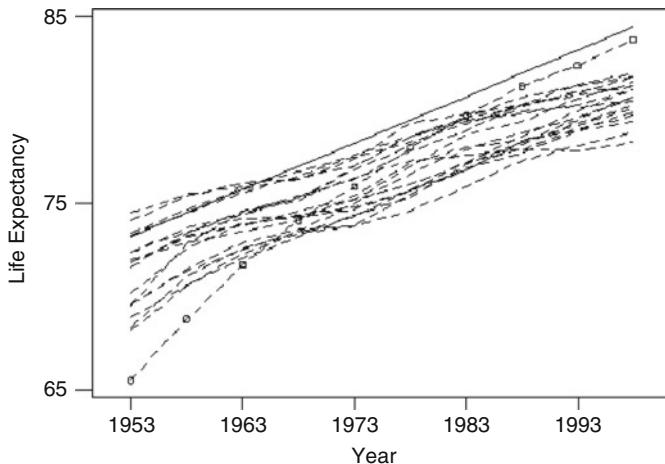


Fig. 15.1 Life expectancies in 19 countries (Japan with a circle), and the best practice life expectancy (solid)

and the United States, for the periods 1950–55, 1955–1960, . . . , 1995–2000 (United Nations 2000). For ease of exposition, we denote the 5 year periods as $t = 1953, 1958, \dots, 1998$. Denoting life expectancy at birth in country $i = 1, 2, \dots, 19$ by $e_{0,i}(t)$ we define the variables of interest as:

$$\text{early life expectancy } LE53(i) = e_{0,i}(1953);$$

$$\text{later life expectancy } LE78(i) = e_{0,i}(1978);$$

$$\text{deviance } Dev(i) = \tilde{e}_0(t) - e_{0,i}(t),$$

$$\text{early annual improvement } Early(i) = (e_{0,i}(1978) - e_{0,i}(1953))/25,$$

$$\text{later annual improvement } Later(i) = (e_{0,i}(1998) - e_{0,i}(1978))/20.$$

Figure 15.1 shows the life expectancies of the 19 countries together with the best practice line. Two facts stand out. First, Japan has behaved in a radically different manner from the rest of the countries. A formal test using Mahalanobis' distance (e.g., Afifi and Azen 1979, 282) also suggests that Japan is an outlier with a P-value <0.001. Second, all other countries appear to gradually veer off below the line. It is this set of 18 countries that we will be primarily concerned with in this paper.

To quantify the latter effect the following descriptive statistics were calculated for the 18 countries (Japan omitted):

Variable	N	Mean	Median	StDev
Dev53	18	-1.956	-1.550	1.942
Dev98	18	-3.928	-3.800	1.050

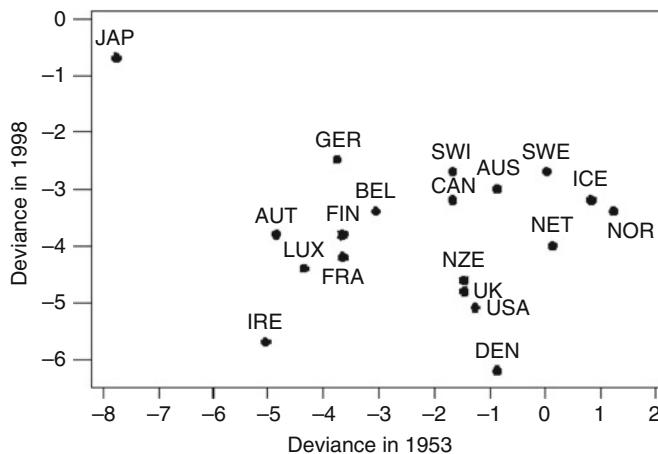


Fig. 15.2 Deviances in 1953 and 1998

Thus, the 18 countries that were an average of 2 years behind the best country in the early 1950's (the best country being a member of the set of 18!), have fallen 2 years further behind in approximately 45 years. We also see that the spread among the 18 countries has decreased by a half.

For reference later, we note that *had* one forecasted life expectancy 45 years ahead in the first part of the 1950's, by assuming that life expectancy will increase at the same rate as best practice life expectancy, then the average error in the 18 countries would have been 2 years.

Figure 15.2, which includes Japan, illustrates how different Japan is. However, it also reveals other interesting changes. For example, Denmark that was just under the best-practice line in the early 1950's has fallen a full 6 years behind. The neighboring countries of Iceland, Norway and Sweden also fell behind, but by "three years only". Thus, Denmark has, during a half a century, gradually distanced itself from the neighbors.

To examine country-specific changes more closely, we regressed the early improvement (Early) on life expectancy in the early 1950's (LE53), among the 18 countries. The estimated coefficients are:

Predictor	Coef	SE Coef	T	P
Constant	1.4708	0.3257	4.52	0.000
LE53	-0.017432	0.004537	-3.82	0.002

with $R^2 = 47.7\%$. Regressing later improvement (Later) on life expectancy in the late 1970's (LE78) yielded:

Predictor	Coef	SE Coef	T	P
Constant	2.5127	0.6091	4.13	0.001
LE78	-0.030312	0.007909	-3.83	0.001

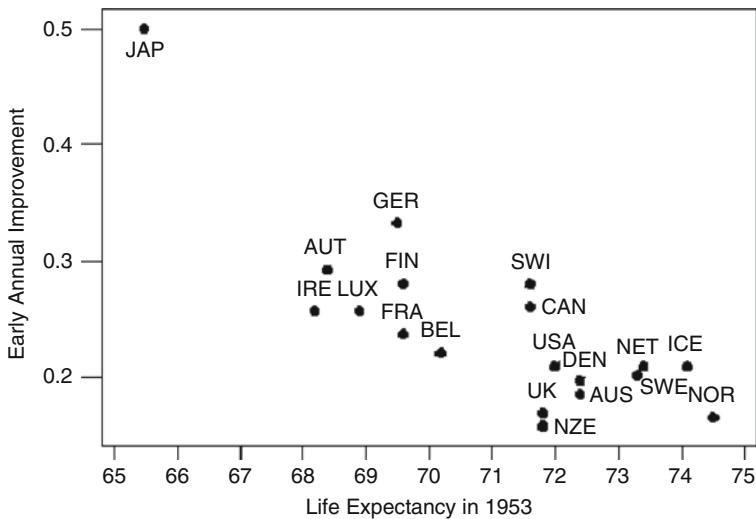


Fig. 15.3 Early annual improvements as a function of life expectancy in 1953

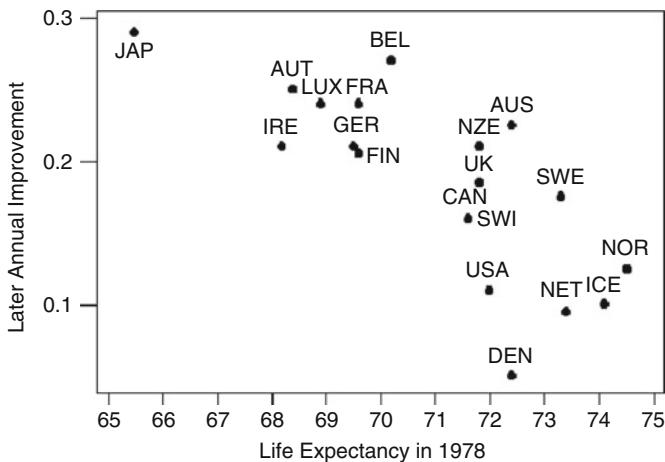


Fig. 15.4 Later annual improvements as a function of life expectancy in 1978

with $R^2 = 47.9\%$. Figures 15.3 and 15.4 illustrate the same phenomenon. We find that in both cases the countries that had high life expectancy grew, on average, slower than those with low life expectancy. The well-known phenomenon of “regression to the mean” explains part of the changes, but we cannot ignore the possibility that there would be a tendency of having a lower rate of improvement when starting from a high value.

We then examined the persistence of improvement among the 18 countries. Correlations (with P-values for the hypothesis of zero correlation in parenthesis) between Later, LE78, and Early were (Japan omitted):

	Later	LE78
LE78	-0.692 (0.001)	
Early	0.342 (0.165)	-0.081 (0.748)

This suggests that there may be some persistence. However, when Later is regressed on LE78 and Early, the coefficients are

Predictor	Coef	SE Coef	T	P
Constant	2.3514	0.5855	4.02	0.001
LE78	-0.029288	0.007525	-3.89	0.001
Early	0.3617	0.2163	1.67	0.115

with $R^2 = 50.2\%$ (adjusted for the number of explanatory variables). While the regression is marginally better than the one not including Early (with $R^2 = 47.9\%$), the effect of Early is small and not significant. The regression is compatible with the notion that current level rather than past improvement has had a systematic association with the later development.

Descriptive statistics on early and later improvement among the 18 countries are as follows (Japan omitted):

Variable	N	Mean	Median	StDev
Early	18	0.2280	0.2140	0.0490
Later	18	0.1789	0.1950	0.0618

Had these statistics been used to forecast life expectancy in the late 1970's for the late 1990's, the average error would have been $20(0.2280 - 0.1789) = 0.982$, as opposed to the average error of $20 \times (0.25 - 0.1789) = 1.422$ years that would have resulted from the use of the best practice line. I.e., the error of the latter forecast would have been about 50% higher.

We conclude that during 1950–2000, as life expectancy has increased, its annual improvement has gradually decreased. Based on Figs. 15.3 and 15.4 this holds for Japan, as well. The 18 countries have also come closer together, and they have fallen further behind Japan.

15.3 Conditions on the Usefulness of an Auxiliary Series

The model for the best-practice life expectancy says that (female) life expectancy at birth increases by 0.25 years every calendar year, but the 18 countries have fallen from 1.5 years behind in the 1950's to nearly 4 years behind in the late 1990's, on average. The deviance for the average of the 18 countries is a roughly linear function of time ($R^2 = 86.1\%$), and we estimate that the deviance has increased by about 0.05 years each calendar year. In 50 years time the best-practice line would imply an increase of 12.5 years, but if the average of the 18 countries continues to fall behind,

the increase would be less, or $12.5 - 0.05 \times 50 = 10.0$ years. In general, we might wish to establish an empirical relationship between the best practice line and the measure of interest, which we take here to be the average of the 18 countries.

Suppose there are some functions $f_j(t)$, $j = 0, 1, 2, \dots$, such that an invariant $g(t)$ is of the form

$$g(t) = \sum_{j=0}^m \alpha_j f_j(t).$$

Suppose the series of interest, say $e(t)$, is related to the invariant via

$$e(t) - g(t) = \sum_{j=0}^n \beta_j f_j(t) + \epsilon(t)$$

Where $\epsilon(t)$ is random with expectation $E[\epsilon(t)] = 0$. If $n \geq m$, then the *same* (e.g., generalized least squares) forecast for $e(t)$ is obtained by (a) modeling the difference $e(t) - g(t)$ and adding the result to $g(t)$ that is assumed to be known, or (b) by modeling $e(t)$ directly with the same explanatory variables $f_j(t)$, $j = 1, \dots, n$, but with modified coefficients $\gamma_j = \beta_j + \alpha_j$ (take $\alpha_j = 0$ for $j > m$). This follows from the fact that if the result of (a) is known, then the result of (b) can be deduced, and *vice versa*. Thus, in this case the knowledge of the invariant provides no help.

On the other hand, suppose $m > n$, or the invariant $g(t)$ behaves in a more complex manner than the deviance $e(t) - g(t)$. In this case, if the future values of the invariant can be assumed to be known for all t , we can reduce the dimensionality of the problem to m explanatory variables by modeling the deviance from the invariant. This can be of important practical use, especially if the future values of some of the functions $f_j(t)$, $j = n + 1, \dots, m$, are unknown. From this perspective having a linear invariant (with $m = 2$ only) is, paradoxically, the least helpful!

An alternative point of view is that if there is information about the difference $e(t) - g(t)$ that has not been reflected in the past values of the series $e(t)$, then such information can be introduced *via judgment* into forecasting. In the example at hand, suppose one believes that there is a feedback mechanism in operation such that if the life expectancy of a country falls sufficiently far behind the best-practice life expectancy, then corrective action will be taken by the society to reduce the deviance, in the future. This is a reasonable hypothesis, and presumably such an effect could manifest itself in the future. For example, even though Denmark has distanced itself from its neighbors for a half a century, perhaps later it will recoup some of the loss. More generally, if the 18 countries that have fallen behind Japan transform their life style in such a way that it resembles more that of Japan in terms of nutrition, job-security, attitude to leisure etc., then maybe they will begin to catch up. However, as this is a strong judgmental assumption that has to be defended by means other than statistical analysis, we will next pursue a number of alternatives that a statistical analyst might consider.

15.4 Model Choice

Figure 15.5 shows, in accordance with the earlier analyses, that the average improvement was higher in the early part of the observation period than in the later part. If the intention is to forecast until, say, 2050, the observation period is rather short, and alternative ways of viewing the trend are plausible. (a) Disregarding the first appearance, if we assume that the series is actually stationary, then the mean (*) is approximately the best predictor after a few years. (b) If we think that the series is a random walk, then the last observation (·) is the best predictor. (c) If we think that there is an exponentially linearly declining trend in the series, then the best prediction also declines exponentially (×). (d) If we think there is a linear trend, then the best predictor is the estimated linear line (+).

Forecasting as far as 2050, a choice between (a) – (d) can make a tremendous difference (this was pointed out in a more general context by Whelpton et al. 1947, already):

- using the historical average we expect to gain $50 \times 0.2062 = 10.3$ years;
- using the latest value we expect to gain $50 \times 0.15 = 7.5$ years;
- using the exponential trend we expect to gain 5.9 years;
- using linear trend we expect to gain 4.0 years.

All values are below the expected gain of 12.5 years derived from the linear model for the best practice life expectancy.

To distinguish between the models we can first examine the estimated variance of the residuals under models (a) – (d) and the best practice line model that assumes a

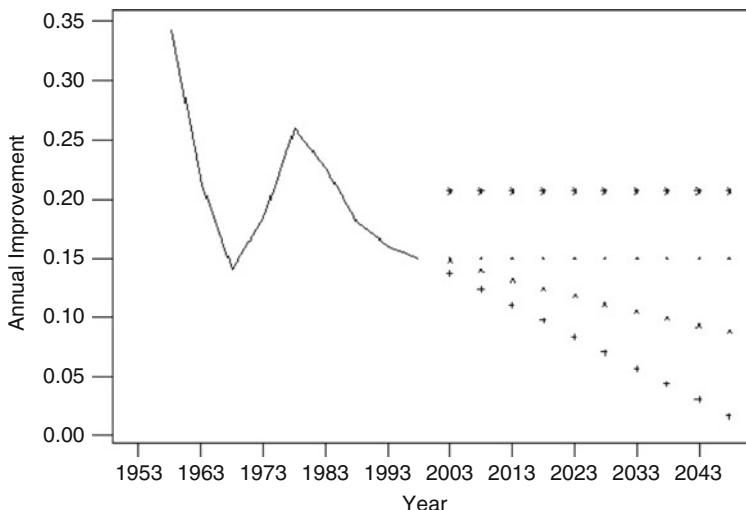


Fig. 15.5 Average annual improvement in average life expectancy during five-year periods, of the 18 countries (Japan excluded), in 1950–2000, and four forecasts based on historical average (*), last observed value (·), exponential trend (×) and linear trend (+)

constant rate of increase of 0.25 years per calendar year. The number of data points is $n = 10$ (from ten 5-year periods), and the number of estimates of annual increase is $n - 1 = 9$. The residual degrees of freedom in models (a) – (d) are 8, 8, 7, and 7, respectively. The best practice line model has 9 degrees of freedom, because it has no estimated parameters. Compared in this manner we find that the estimated variances of the residuals in the five models are 0.0041, 0.0042, 0.0031, 0.0031 and 0.0056. In view of Fig. 15.5, it is not surprising that the two regression models lead to the best fit. Similarly, it is not surprising that the last model with a rate coming from the outside of the data set fits the worst. The fact that the random walk model is not among the best is informative. Although the regression models fit the best, we recognize that the data period is short and one cannot take results of this type as decisive.

Another possibility is to try to find supporting evidence based on alternative approaches to the same problem. Here the “rates-to-life expectancy” comparison is available. The life expectancy of the Finnish women in 2000 was 81.0 years, or essentially the same as the average of 80.6 for the late 1990’s, of the 18 countries. A stochastic forecast (Alho 2002) that assumed the decline in age-specific mortality to continue in each age at the rate of the most recent 15 years lead to a median of female life expectancy in 2050 of 86.7, indicating a gain of 6 years. This agrees with the assumption of an exponential decline model (c). We will examine this model further.

Consider a function $e(t)$ such that $e(0) = A$ and $e'(t) = e^{\alpha - \beta t}$, $\beta > 0$, for $t \geq 0$. It follows that

$$e(t) = A + B(1 - e^{-\alpha t}),$$

where $B = e^\alpha / \beta$. Taking $t = 0$ to correspond to the late 1990’s, we have $A = 80.6$ and our empirical estimates can be translated to values $\alpha = -1.8472$ and $\beta = 0.01151$, which imply that $B = 13.7$. Under this model the average life expectancy of the 18 countries would never exceed $80.6 + 13.7 = 94.3$ years. For the year 2050 we would get the value $80.6 + 6.2 = 86.8$, for example. (The increase here is slightly larger than the 5.9 years given above, because the starting period is earlier.)

To complement the above point estimates we note that by using the so-called delta-method (e.g., Rao 1973, 385–6) we can compute a standard error for the estimate of B , as 9.4 years. Thus a 95% confidence interval for the additional improvement is quite wide, approximately 13.7 ± 18.4 years. From this, the 95% upper limit for the average life expectancy of the 18 countries would be about $94.3 + 18.4 = 112.7$ years. Of course, even under this model, individual people can live much longer.

Figure 15.6 has a graph of the past data together with a point forecast until the late 2040’s. Visually, the slight concavity smoothly continues from the past data to the point forecast.

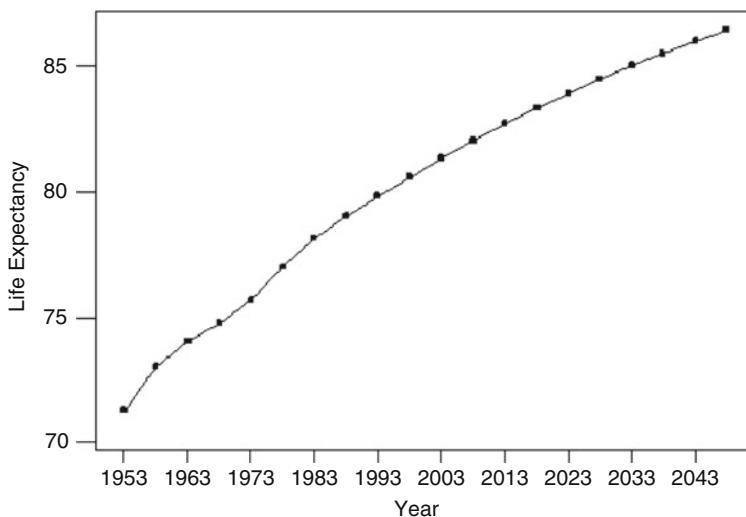


Fig. 15.6 Average life expectancy of the 18 countries in 1950–2000 continued with a forecast based on an exponential trend in annual improvements for 2001–2050

15.5 Concluding Remarks

We have investigated statistically the possible use of the best-practice life expectancy as an aid in forecasting the life expectancy of industrialized countries. The evidence shows that during the past 50 years this would have been overly optimistic. The results do not preclude the possibility that in the longer term a comparison to the best practice line might prove to be useful, but beliefs concerning this cannot be based on statistical analyses of the type we have conducted. Instead, arguments concerning processes, whose effects have not manifested themselves yet, are required.

Better fits would have been provided by models that incorporate the slowing down of improvement in life expectancy, among the countries studied. A model that assumed a geometric slowing down leads to an absolute upper bound for life expectancy, but estimates about this upper bound are statistically quite uncertain. The validity of such a model cannot be ascertained based on the short data period we consider.

Independently of whether life expectancy turns out to be approximately linear or concave (or convex!) in the long run, there may well be other periods besides the latter part of the twentieth century, in which groups of countries veer off the trend for decades. From the perspective of individual countries this possibility would have to be allowed in the construction of prediction intervals.

In case one is not willing to choose an appropriate model at all, one can try to assign probabilities to each model, and do model averaging (Draper 1995). This approach has the advantage of leading to more honest prediction intervals, as it does

not condition on a particular choice, but the disadvantage is that it requires the assignment of probabilities. It may be difficult to achieve a consensus on the latter.

References

- Afifi, A. A., & Azen, S. P. (1979). *Statistical analysis* (2nd ed.). New York: Academic.
- Alho, J. M. (1989). Relating changes in life expectancy to changes in mortality. *Demography*, 26, 705–709.
- Alho, J. M. (2002). *The population of Finland in 2050 and beyond* (Discussion Paper 826). Helsinki: The Research Institute of Finnish Economy.
- Draper, D. (1995). Assessment and propagation of model uncertainty (with discussion). *Journal of the Royal Statistical Society, B*, 45–97.
- Oeppen, J., & Vaupel, J. W. (2002). Broken limits to life expectancy. *Science*, 296(5570), 1029–1031.
- Rao, C. R. (1973). *Linear statistical inference and its applications* (2nd ed.). New York: Wiley.
- Törnqvist, L. (1949). Näkökohdat, jotka ovat määräanneet primääristen prognoosiolettamusten valinnan. In J. Hyppölä, A. Tunkelo & L. Törnqvist (Eds.), *Suomen väestöä, sen uusiutumista ja tulevaa kehitystä koskevia laskelmia* (Tilastollisia tiedontantoja 38). Helsinki: Tilastokeskus.
- United Nations. (2000). *World population prospects. The 2000 revision: Volume I: Comprehensive tables*. New York: United Nations.
- Whelpton, P. K., Eldridge, H. T., & Siegel, J. S. (1947). *Forecasts of the population of the United States 1945–1975*. Washington, DC: U.S. Bureau of the Census.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 16

Life Expectancy Convergence Among Nations Since 1820: Separating the Effects of Technology and Income



Jim Oeppen

16.1 Limits and Convergence in Life Expectancy

Figure 16.1 shows some details of the probable trajectories of limits and convergence for average life expectancy over the past four centuries. The curved line is an attempt to define the upper bound, or “best practice”, average life expectancy that could be achieved at any one time.¹ We can think of this as an evolving upper bound to the “technophysio” evolution of the human population, in the sense proposed by Fogel and Costa (1997). The bottom limit of the graph is drawn at an average life expectancy of 22.5 years, to approximate the lowest level that a population could experience and still be viable in the long term.² Today, even a country like Sierra Leone, with one of the lowest life expectancies recorded by the U.N., is close to the upper limit for a pre-1800 population.

If these two limits are plausible, then the history of average life expectancy over the last four centuries must lie between them. It is immediately apparent that the scope for absolute divergence after 1850 is much greater than before. Since the

I should like to gratefully acknowledge the enormous help I have received, from a large number of kind individuals and institutions, in building the demographic data-series for this paper. A full list of their names and the sources used will appear in the longer version of this paper. It is also clear that this topic could not have been addressed at all without the economic data published by Angus Maddison.

¹For details of its definition, and particularly its linearity after 1840, see Oeppen and Vaupel (2002) and associated Web material.

²The lower limit to viability is somewhat uncertain as it depends on assumptions about age-specific mortality patterns and about the dependence between mortality and reproductive health.

J. Oeppen (✉)

Max Planck Institute for Demographic Research, Rostock, Germany
e-mail: joeppen@health.sdu.dk

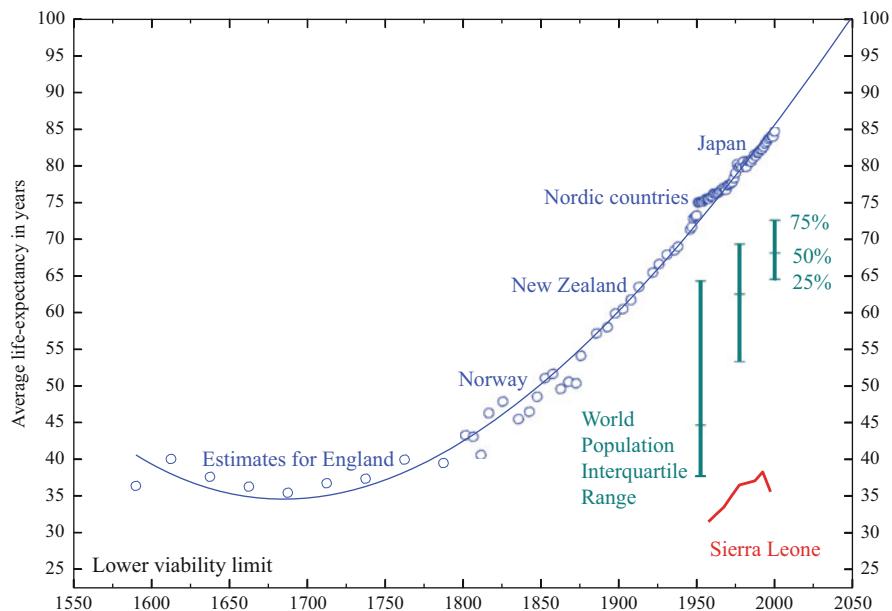


Fig. 16.1 Limits and convergence for national average female life expectancy at birth

middle of the twentieth century, data on life expectancy or U.N. estimates are available for most countries and convergence has been generally apparent, although there has been recent concern about sub-Saharan Africa and the former communist bloc. A particularly effective way of depicting this convergence was published by Wilson (2001).³ He weighted national averages by populations to estimate the concentration of life expectancy for the World population at three dates. The vertical bars show the inter-quartile range of life expectancy for countries containing half the World's population.

The bars emphasise three massive changes: rapid improvement in life expectancy, greater symmetry in the distribution, and the globalisation, or compression, of mortality experience. It also becomes clear that the years from 1850 to 1950 were probably the period of maximum diversity in the history of human mortality. After 1950 cross-sectional, or sigma, convergence is apparent, with compression from below, but the gap between the 75th percentile and the national “best-practice” limit seems to be growing.⁴ This might be interpreted as evidence of a new period of divergence. Today’s highest levels of life expectancy can only be achieved by reducing the mortality of the elderly and it may be that some countries will find this harder to achieve than the gains they made at younger ages.

So how do we explain these broad patterns of rising and converging life expectancy, bounded by one fixed and one evolving limit? Many economists have

³This article contains similar analyses of fertility.

⁴Of course, it is possible that this process began before 1950.

assumed that survival improvements follow automatically from economic development. This is supported by the observation that the rank ordering of countries across both variables is broadly similar and persistent over time, although the ratio of highest to lowest in income is massively bigger than that for mortality, suggesting a non-linear relationship.⁵ Thus rising world life expectancy could be a simple function of rising real incomes, but what role should be assigned to technology transfers?

A second problem is that while the mortality range is compressing across nations over time, the consensus seems to be that the same is not true for income, although this is a subject of debate among economists. One might seek to explain it as the approach to a fixed biological upper bound to average life expectancy, resulting in diminishing returns to wealth, but this common idea of a fixed and imminent limit receives no support in two recent investigations (Oeppen and Vaupel 2002; White 2002). We have also seen that the World population does not seem to be converging on the evolving upper bound.

For insights into the roles of wealth and technology in determining the levels and convergence of life expectancy we turn to the classic article in the field.

16.2 The Classic Article: Preston (1975)⁶

Speculation and assumption regarding the link between income and health persist. The first concrete, macro-level, study seems to have been undertaken by Preston (1975). In this article, Preston partitioned an historic fall in mortality into two factors: modern economic growth, and improvements in health technology. The first step was to find a function linking national income per head and life expectancy. Figure 16.2 is an updated version of Preston's graph – the original included cross-sections for the 1900s, 1930s, and 1960s, with logistic functions fitted to the latter two.⁷ The idea is that there is a level of technology, or universal production function,

⁵The stability over time of the rankings for survival is a curious phenomenon because there has been a massive “sectoral” redistribution since 1820 in the pattern of deaths by age. Why should a country like Norway maintain its position close to the top of the list, regardless of whether mortality is concentrated among infants, adults or the elderly – perhaps the equivalent of changing from an organic, to a mineral, to an information economy?

⁶Slightly longer versions detailing cause-of-death effects are contained in Preston (1976) and (1985).

⁷The data shown are for female life expectancy and GDP per capita, expressed in 1995 international Geary-Khamis dollars. Preston used life expectancy for the sexes combined and national income per head in 1963 U.S. dollars, truncating the horizontal axis below the income level reached by the four richest countries in 1960. His list of countries grows over time, and by 1960 overlaps heavily with the one used here, but is not an exact match, especially among the poorer nations. Preston fitted logistic curves to the un-truncated data with an *a priori* asymptote of 80 years and scaled each cross-section's income from 0 to 100. The lines shown here are polynomials in the log of GDP per capita. The 1960s outlier with poor survival but high income is Venezuela.

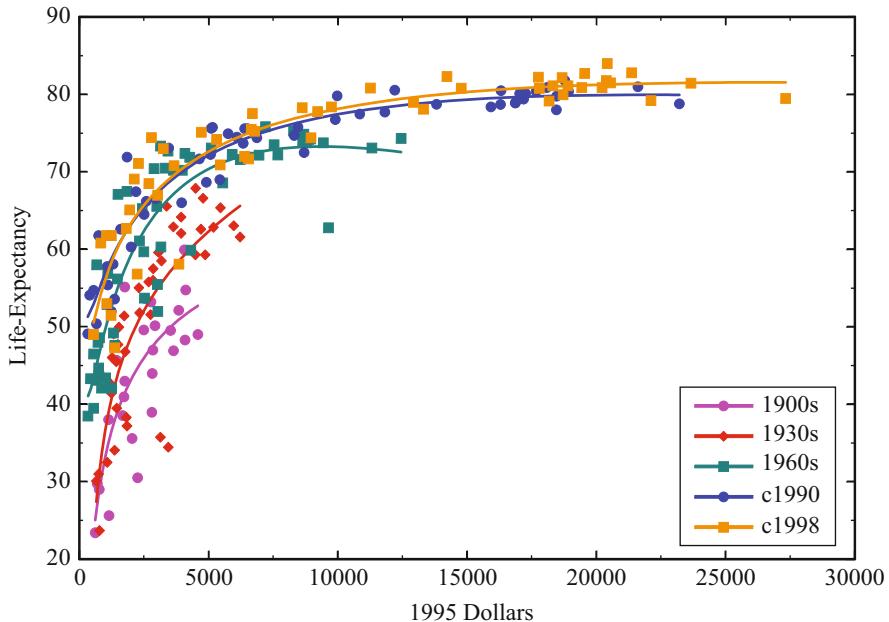


Fig. 16.2 Female life-expectancy at birth and GDP per capita

that links input (National Income *per capita*) to output (average life expectancy) at any one moment in time, but that these functions are subject to temporal shifts. For example, an input of 5000 dollars per head “produced” an average lifetime of 50 years with the technology prevailing in the 1900s, but the same amount of money realised about 70 years in the 1960s. The addition of data for the 1990s does not alter the basic finding and shows that the curve is still shifting upwards in the wealthier countries.

Easterlin (1996) has pointed out the similarities between Preston’s approach and a pioneering study of modern economic growth by Solow, who divided growth in output into two components: (1) input growth with fixed technology, and (2) shifts in the production function due to technological change. In between technological shifts, countries could still make gains by increasing inputs. Easterlin describes Preston’s study as “done independently of Solow’s work and no less deserving of classic status” (1996, p. 75).

Preston calculated the rise in life expectancy that would have occurred if health technology had been fixed at the 1930s curve, but real incomes per head had risen as observed. Subtracting the hypothetical gain due to income change from the total gain led to the conclusion that about 80% of the rise in life expectancy from 1930 to 1960 could not be attributed to increases in income. This model of mortality treated changes in health technology as an unidentified, exogenous component expressed as a function of time. In subsequent papers, Preston (1980) included variables measuring literacy and nutrition, showing that 70% of the mortality decline in

LDCs (excluding China) between 1940 and 1970 could not be accounted for by changes in income. For the period 1965 to 1979, again for LDCs, he has estimated that the technology share has fallen to 30% (Preston 1985).

16.3 Extending the Analysis

Preston's model of the linkage between health and national income offered a new way of looking at an old problem, but it still leaves some questions unanswered. The two cross-sections enclose a period of rapid technological change in health, including the introduction of antibiotics, so we have to question whether the dominance of technological change over income change is a long-term phenomenon. The "best-practice" line in Fig. 16.1 suggests that these cross-sections enclosed the only marked "step" in survival over the last century. The rest of the period suggests stable change at the top level. For countries lower down the list, some authors have suggested that the period of inter-war retrenchment in international trade may have deferred demographic innovations, accentuating "catch-up" in the immediate post-war period (see e.g. Bloom and Williamson 1998).

Another important question is whether the dynamics of the model are well specified. Is health technology simply scalable with income? The "sectoral" changes in the age-distribution of deaths required to move from a life expectancy of 50 to one of 80 have been associated with major shifts of emphasis; changes in the importance of infectious versus chronic diseases, public health measures versus medicine, and the changing share of responsibility between the individual and the state. It seems unlikely that every country will experience these changes in the same way. We should expect persistent, "national" effects within the overall relationship.

For this paper on convergence, Preston's major result is that if we assume a fixed and universal health technology function then the curves show that rising real incomes will lead to convergence in life expectancy, without the requirement that the real income distribution should itself converge. But since Preston argues that technological change dominated between the 1930s and the 1970s, a mechanism based only on income offers little insight into the full mechanisms of long-run convergence. Today there is a considerable literature on economic convergence, but almost no formal analyses of demographic convergence (Wilson 2001). Demographers have concentrated on transition models, with time-scale compression for the late-entrants. The discussion favours technology transfer rather than economic growth, and emphasises the countries that have achieved high life expectancy with relatively low incomes. As with economic growth, late-entrants are recognised to have experienced rates of life expectancy increase never seen in the pioneering countries, but this is not framed in an analysis of the lower costs of imitation compared with innovation.⁸

⁸For an economic approach, see van Elkan (1996).

This paper tries to expand the wealth of results that Preston's article generated into three new areas. Firstly, the temporal range of the data can be extended. Secondly, new statistical methods may allow us to learn more about the precise roles of time/technology and income. Finally, these same methods may provide more insights into country-level patterns, treated as residuals in the Preston model.

16.4 New Data

The desire to look at change in life expectancy over the long course of the health transition is severely restricted by the available data, and thus the resulting model will be unsatisfactory in many ways.⁹ Time and GDP *per capita* are poor proxies for the things we would really like to know.

This paper takes advantage of the collection of national time-series for 56 countries published by Maddison (1995).¹⁰ GDP per capita is used here, expressed in 1990 Geary-Khamis dollars to remove the effects of inflation and currency. The series were made comparable using a Purchasing Power Parity (PPP) approach rather than relying on exchange rates, which are often distorting.¹¹ For most advanced capitalist countries, there are estimates for 1820, 1850, and then annually from 1870 to 1994. For other countries, the starting dates and continuity are variable.

Despite the recent increase in the number of countries with annual life-table series, much of the life expectancy data is sporadic and covers varying periods.¹² The GDP data has been averaged to match the time spans of the life expectancy estimates, which are shown for females in Fig. 16.3. The nineteenth century has few low-survival countries and is largely confined to countries in Europe, or of European origin. Some Asian and Latin American estimates start in the first half of the twentieth century, but the U.N. estimates for African countries only begin in 1950. This results in a very unbalanced design with discontinuous data, and places a limitation on the kinds of models that can be used. Initial experiments in modelling showed that the 1918 flu epidemic created problems with the residuals, as did wars.

⁹For the years after 1950, a much richer selection of socio-economic variables is available. See Jonathan Temple's website at <http://www.bris.ac.uk/Depts/Economics/Growth/> for a guide to data and literature.

¹⁰The data extends over 56 countries and from 1820 to 1994. A more extensive set of countries from 1950 to 1998 is contained in Maddison (2001). The additional countries will allow this analysis to extend to the poorer nations of Africa, Asia, and Central and South America and it is being revised to incorporate these new data.

¹¹PPP attempts to compare currencies by their power to buy similar products.

Preston switched to PPP after his first article. Non-traded items are typically cheaper in poor countries and thus PPP adjusted wealth estimates are usually higher than those based on exchange rates for poor countries. This may explain why Preston found that a log model did not fit the low-income range.

¹²Collections of mortality data can be found by following the links at <http://www.demogr.mpg.de/>

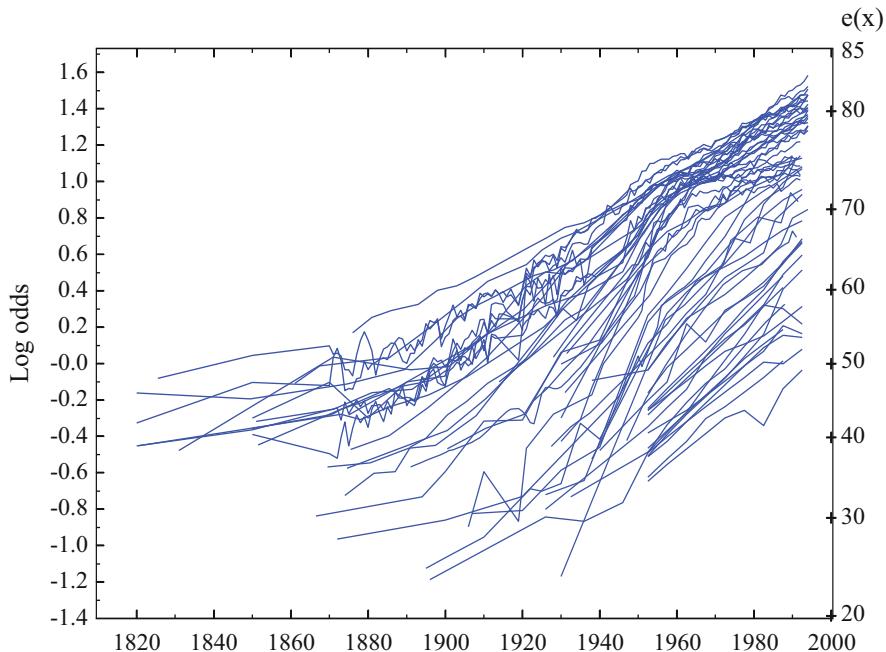


Fig. 16.3 Female life-expectancy at birth

An attempt has been made to identify the years when countries were subject to the direct and indirect effects of war and these, together with data for 1918, have been omitted.

The income data used in the model is log GDP *per capita* in real dollars, and the life expectancy data is expressed as the log-odds of survival assuming an upper limit on life expectancy of 100 years.¹³

Equation 16.1: Log Odds of Life Expectancy at Birth

$$\ln \text{odds}(e_0) = \ln(e_0/(100 - e_0))$$

The average number of years lived is divided by the average number of years “lost” assuming an upper limit on average life expectancy of 100 years. Taking the natural log means that the measure is unbounded on the positive and negative sides. This has the advantage of linearising the life expectancy data and removing any ceiling effect – something that Preston did with a logistic transform.

¹³This transformation is probably unnecessary and may be dropped in the final analyses.

16.5 National Effects: A Shopping Analogy

Although economists would be quick to point out that there is no real market, Preston introduced us to the idea of an international price/technology/quantity relationship for life expectancy. We can hold any one constant and think about the other pair. Five thousand dollars in 1960 “bought” 70 years of life expectancy, but the same money in 1900 could only buy 50 years. Most people are familiar with this from buying personal computers and other electrical goods. A thousand dollars today buys a greater quantity of computing power than it did 5 years ago because technology has shifted the supply curve upwards.

Is a model based on an international relationship sufficient to reproduce the data? Preston considered that there were also national relationships hidden within the international model, but he didn’t integrate them and treated them as residuals. He observed, for example, that Japan seemed to have very high life expectancy in the 1900s, relative to its income per head, and cites Taeuber’s explanation of “personal cleanliness and the assumption of health responsibility by government organizations as important factors in counteracting the adverse effects of poverty”(Preston 1975, n. 22, p. 236).

To extend the shopping analogy, suppose that technology is always a year ahead in Japan compared to the U.S. – as a result you get a better PC at any given price in Japan, because the technology/price relationship is higher.

Similarly, if delivery costs are high in a country, then the quantity/price ratio is pulled down for any given technology. Thus there may be persistent technological leads and lags, and pricing discounts and surcharges at the national level.

This viewpoint prompts new questions. For example, Norway seems to have always been close to the top of the life expectancy rankings, yet it was relatively poor by European standards at the start of the nineteenth century. Has it always been able to “buy” its life expectancy at a discount, perhaps because an egalitarian society gives a real meaning to “*per capita*” income, and more easily translates this into health for all the population? Or have they always had a technological lead, perhaps because literacy was so high, not gender specific, and created a tradition of investment in human capital? (Graff 1987, p. 375; Houston 1988, p. 135). Maybe both factors were at work and they were lucky? When they were poor, they knew how to control infant mortality at little cost. By the time mortality reduction required expensive items, like advanced health care for all and State support for the elderly, they had become rich.

Preston’s analysis was for the sexes combined, but we can also consider who is doing the “shopping” – a man or a woman? Why, in the modern era, do women seem to be able to buy more? Each discipline that addresses this question – from evolutionary biology to sociology – has its own explanation.

16.6 Multilevel Models¹⁴

In Preston's original paper, the model was fitted to cross-sectional data. This raises a number of familiar estimation problems. We can illustrate this in an intuitive way by examining the points in Fig. 16.1. It is likely that had lines been plotted for countries rather than cross-sections, a very different impression of the data would have been gained. Each country's data can be thought of as a number of repeated measures on a single unit or group. Preston's plot shows us the between-groups picture at three points in time, but we also need to understand the within-group change.

This paper uses a multilevel model to go beyond Preston's "series of cross-sections" approach. Multilevel models are designed to respect the hierarchical nature of both data and explanations. The textbook examples often use school data. Pupils may be tested several times; they are grouped within classes and by teachers, which are grouped within schools, districts, and so on. Historical demographers are used to seeing nested data on regions, villages, families, parents, siblings, and individuals. Treating the hierarchy explicitly solves a number of problems. Factors may be significant at one level and not at another, or they may work differently at two levels. For example, the income of a family may have one effect on infant mortality and the average income of the village may have another. One may reflect familial access to resources, and the other may affect community provision of health infrastructure. Many studies ignore the hierarchy and force all the variables to one level. Thus we might see person-level regressions of infant mortality, with population density, a higher level measure, as an explanatory variable.

Equation 16.2 shows a naïve regression model that we can use as a means of introducing the concepts of multilevel models.

Equation 16.2: Regression Model

$$\ln \text{odds}f_{j,t} = \beta_0 + \beta_1 \times t + \beta_2 \times \ln \text{GDP}_{j,t} + \varepsilon_{j,t}$$

This equation models the log odds of female life expectancy in all countries, indexed by j , and for all time periods t , as the sum of a constant, a function of time, a function of log GDP *per capita* in the country at time t , and an error term. This is a "one size fits all" model, and it probably wouldn't work very well, judging by the plots of the data. A standard method of extending such a model is to fit a separate intercept term, or constant, for each country. In bivariate regression, this results in a series of parallel lines, one for each country. In this regression, it leads to a family of parallel planes. The model is usually referred to as Analysis of Covariance, or ANCOVA. The intercepts are really weighted means of the extent to which a particular country differs from the international model. A similar model is RANCOVA, or Random ANCOVA, shown in Eq. 16.3. In this model, we have a

¹⁴For information on multi-level models, in increasing order of complexity, see Kreft and de Leeuw (1998), Snijders and Bosker (1999), and Goldstein (1995).

global intercept $\beta_{0,0}$ and an additive term, $\beta_{0,j}$, for each country j . The additive terms are estimated as a sample from a normal distribution with mean zero. They sum to zero, so they can also be regarded as a “national” residual that doesn’t vary with time.

Equation 16.3: RANCOVA Model

$$\ln \text{odds}f_{j,t} = (\beta_{0,0} + \beta_{0,j}) + \beta_1 \times t + \beta_2 \times \ln \text{GDP}_{j,t} + \varepsilon_{j,t}$$

This idea of national offsets, or residuals, about an international model can be extended to the other parameters. Equation 16.4a shows a multilevel model with country-specific offsets for all parameters – $\beta_{1,j}$ for time, and $\beta_{2,j}$ for income.

Equation 16.4a: Multilevel Model

$$\begin{aligned} \ln \text{odds}f_{j,t} &= (\beta_{0,0} + \beta_{0,j}) \\ &+ (\beta_{1,0} + \beta_{1,j}) \times t + (\beta_{2,0} + \beta_{2,j}) \times \ln \text{GDP}_{j,t} + \varepsilon_{j,t} \end{aligned}$$

Rearranging the terms to Eq. 16.4b reveals another view of the model and exposes its levels. The first line is an international model (or *fixed part* in multilevel model terminology); line two is a country-specific level expressed as offsets or residuals from the higher level (level 2); and finally there is a lowest level residual term (level 1). We now have a model that is interpretable at the international and national levels. It can also be forecast, in both the international and national components.¹⁵

Equation 16.4b: Multilevel Model Separated into a Fixed Part and Two Residual Levels

$$\begin{aligned} \ln \text{odds}f_{j,t} &= \beta_{0,0} + \beta_{1,0} \times t + \beta_{2,0} \times \ln \text{GDP}_{j,t} \\ &+ \beta_{0,j} + \beta_{1,j} \times t + \beta_{2,j} \times \ln \text{GDP}_{j,t} + \varepsilon_{j,t} \end{aligned}$$

To make the interpretation easier, Eq. 16.5 illustrates such a model, which we will pretend is for Finland, country number 6 in this dataset. The international intercept for the log odds is .45, but *ceteris paribus*, Finland always seems to lag behind by a factor of -0.01 . Overall, the measure of survival rises 0.15 for each additional unit of time, but Finland lags a little behind by -0.07 , with a combined effect giving a rise of $+0.08$ per year. On the other hand, Finland seems to be “buying” its survival at a discount of 0.26. Instead of the international gain of 0.19 per additional unit of income, it gets 0.45 for each extra income unit.

¹⁵This assumes that GDP forecasts are available. We might also use the model to back-project, or interpolate.

Equation 16.5: Illustrative Multilevel Equation for Country Number 6

$$\begin{aligned}\ln \text{odds}_6,t &= 0.45 + 0.15 \times t + 0.19 \times \ln \text{GDP}_{6,t} \\ &\quad - 0.01 - 0.07 \times t + 0.26 \times \ln \text{GDP}_{6,t} + \varepsilon_{6,t}\end{aligned}$$

Turning back to Eq. 16.4a, we can interpret the combined effects on the explanatory variables. β_1,j results in a growing lead when it is positive, and a growing lag when it is negative. Similarly, β_2,j is a discount when it is positive, and a surcharge when it is negative. As with the intercept in the regression model, we don't really know how to interpret a particular $\beta_{0,j}$ parameter. It could represent a fixed lead (lag), independent of time. Or if it were associated with income, then it is a stable rebate (cost). In practice, these two fixed possibilities cannot be logically differentiated.

This is a very simplified introduction to the use of the model in this context. It would take too long to recount the full properties of multi-level models here, but a number are particularly relevant. Firstly, the model recognises that the structure of the data is hierarchical. Each country's data is a set of repeated measures over time which share certain implicit factors that cannot be ignored. Secondly, the model does not require that the time points are evenly spaced, or that the data are known for all countries. This allows us to use unbalanced data.

16.7 Model Results

As expected, the model represented in Eq. 16.4 does not work very well. Examination of the residuals shows that they have significant patterns. The strategy adopted for this paper, which concentrates on the national effects, is to make the fixed or international part of the model as flexible as possible. This is an attempt to avoid having the failures of the fixed part interpreted as national effects. Conversely, the national component of the model has been kept as simple as possible, and the form used in that part of Eq. 16.4 is retained.

The final form of the international model uses a polynomial in time because the log transforms do not fully linearise the data. It is also clear that there are epochs in the data. For this reason, the data has been partitioned into pre-World War I, Inter-War, and post World-War II epochs for both time and income. Even then, the post-World War II reconstruction decade presents significant problems. This is an extremely important period in the diffusion of antibiotics and other health technologies, and a time of rapid economic change, so rather than delete the data points, affected countries have been identified and a special component fitted in the model. Finally, even with this expanded model, it was clear that the variance in the level 1 residuals seemed to be a negative function of income, so the model was expanded to allow complex level 1 variation to remove this heteroscedasticity (Goldstein 1995).

Table 16.1 Multilevel model for log odds of life expectancy

Variable	Female	Male
Intercept	-2.3420	-1.954
Time	-0.1073	-0.0961
Time squared	0.0179	0.0185
Time cubed	-0.0004	-0.0005
Inter-war dummy	-0.7258	-0.9735
Post-war dummy	0.9255	0.7637
Ln GDP pre-war	0.1999	0.1173
Ln GDP inter-war	0.2892	0.2403
Ln GDP post-war	0.1122	[0.0486] ^a
Post-war reconstruction dummy	-2.0817	-1.6653
Time * post-war reconstruction dummy	0.1349	0.1082
-2 * log likelihood	-5635	-4938

^aBracketed parameters are not significant

The parameter estimates for the fixed, or international, part of the model are shown in Table 16.1.¹⁶

This model suggests that the long-term upward trend in survival was depressed in the inter-war period, but jumped up after 1945. In both periods men seem to be in a worse position. Income terms for females are positive, particularly so in the inter-war period, but there is evidence for a post-war diminution in the effect of income, and this contrasts with Preston's conclusions for combined male and female life expectancy in LDCs. The inter-war period does seem to show a period of retrenchment – temporal gains slowed and wealth became more important. For men, income seems to be less important than for women and is not statistically significant after 1945. Broadly speaking, the parameters for females and males show similar patterns, although the model is less successful in explaining the male data.

16.8 National Patterns

The nation-specific component of the model for country j is

The Nation-Specific Component of the Multi-level Model

$$\beta_{0,j} + \beta_{1,j} \times t + \beta_{2,j} \times \ln GDP_{j,t}$$

¹⁶1810 has been subtracted from the Time variable in these models, to limit the scale of the polynomial terms.

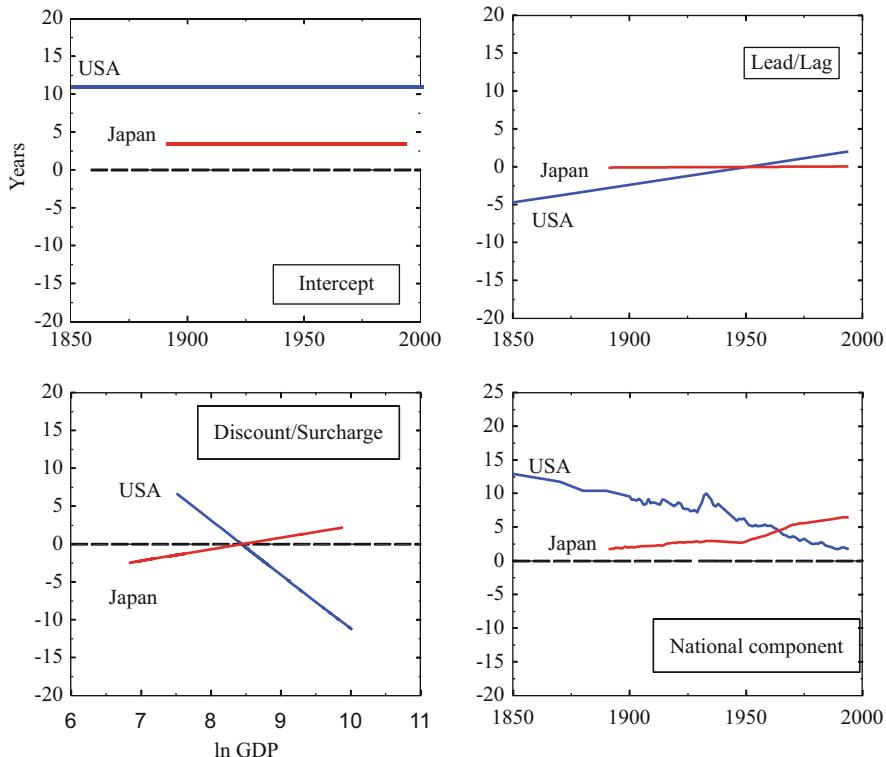


Fig. 16.4 Components of national level life-expectancy estimates

and this can be interpreted as the national “offset” that can be added to the international prediction to get the full prediction for any country.¹⁷ It tells us how one country is performing after controlling for the functions of income and time common to all countries. Figure 16.4 plots the three elements that sum to the national component for the USA and Japan. Because it is controlled out, the international level can be thought of as the horizontal line at zero years. The first panel shows the country-specific constants. Multilevel models are usually fitted to grand-mean centred data so that the variance of the intercept has meaning. The second panel shows the temporal components. Japan’s offset relative to overall temporal change seems to be fixed and approximately zero. The USA has a broadly similar “health technology”, if we can interpret it this way, which is not surprising. The real

¹⁷Normally, intercepts are estimated setting all the explanatory variables to zero. For most social science data, this leads to intercept values outside the plausible range of the data. For these data, it is not so extreme for the variable t , as 1810 was subtracted from the year, but the general position holds for GDP. In multilevel models, it is customary to centre the variables that are used at this level and this has been done here. Kreft and de Leeuw (1998) give an excellent description of the role of centring data in conventional regression, and multilevel models.

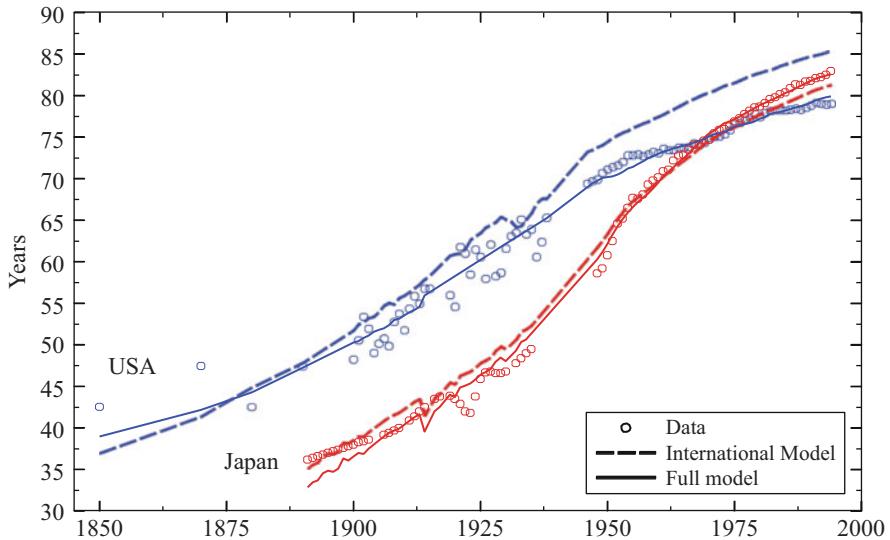


Fig. 16.5 Female life expectancy – model and data against time

difference in their contemporary position is shown in the third panel. The US has become progressively unable to translate log dollars into health. It is now about 10 years behind the position we would expect based on income alone, a figure slightly offset by a small advantage of about 2 years in technology. The American pattern is typical of the countries of Northwest European origin, but only the southern European countries come close to Japan's position of combining a high income with a good conversion rate. The sums of the three components that make up the national offset from the international model are shown in the fourth panel. The USA has regressed to the mean, but Japan shows some strengthening of a long-term advantage.

Figure 16.5 shows the raw data for Japan and the USA, together with dashed lines showing the life expectancy we would expect if they had zero offsets from the international model. As we might expect from the previous graph, this model fits quite well for Japan but overestimates US life expectancy for women today by about 5 years. The full model fits the US data quite well by setting diminishing returns to log GDP *per capita*. Figure 16.6 shows the same data plotted against income. The slope of the US response to log GDP seems to have changed around 1950, but it could be argued that both Japan and the US may now be back on the long-term path.

Figure 16.7 shows these offsets for all the countries, plotted against time. It is immediately apparent that the model is estimating convergence. The fixed part of the model predicts an overall life expectancy in the low twenties for 1820. This is both viable and plausible, and suggests that countries like Norway had a female life expectancy about 25 years above the international prediction. Over time they have progressively lost this advantage, until today it is less than 5 years. It also seems clear

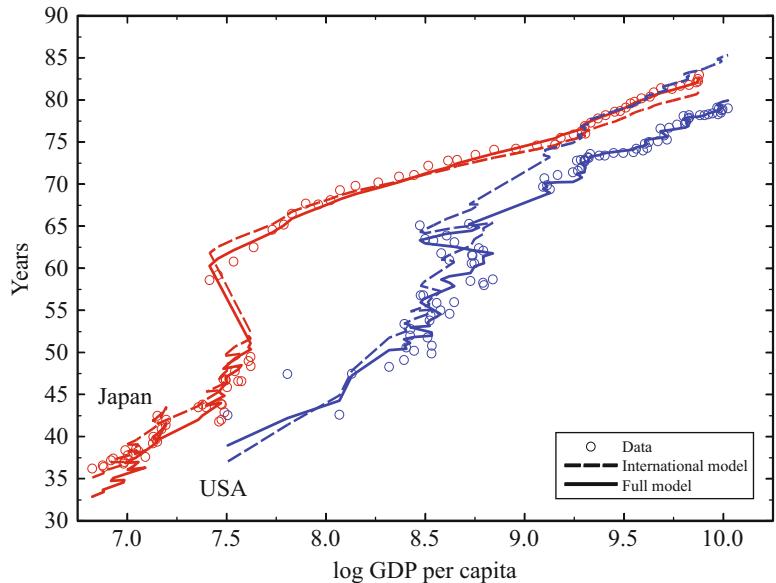


Fig. 16.6 Female life-expectancy: model and data against per capita income

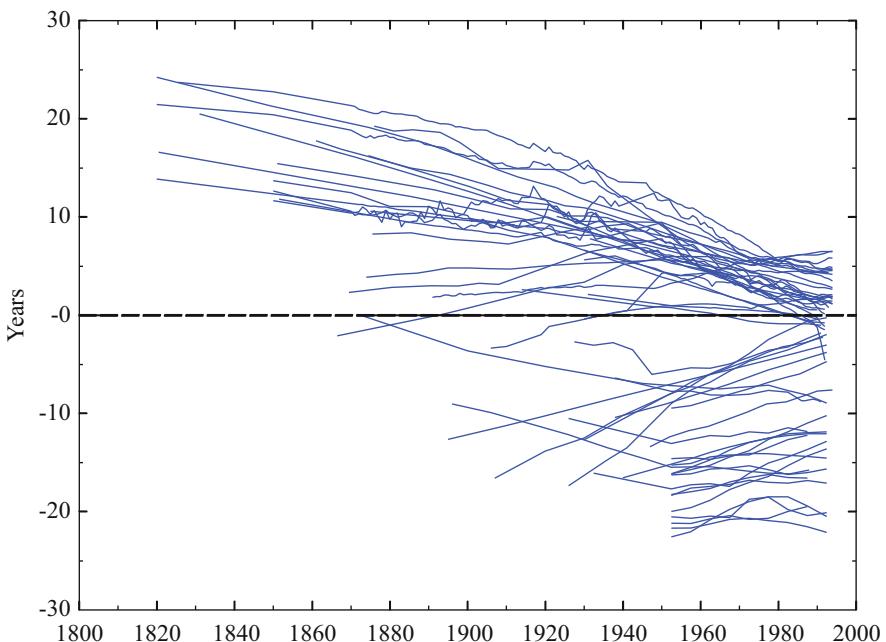


Fig. 16.7 National level female life-expectancy estimates against time

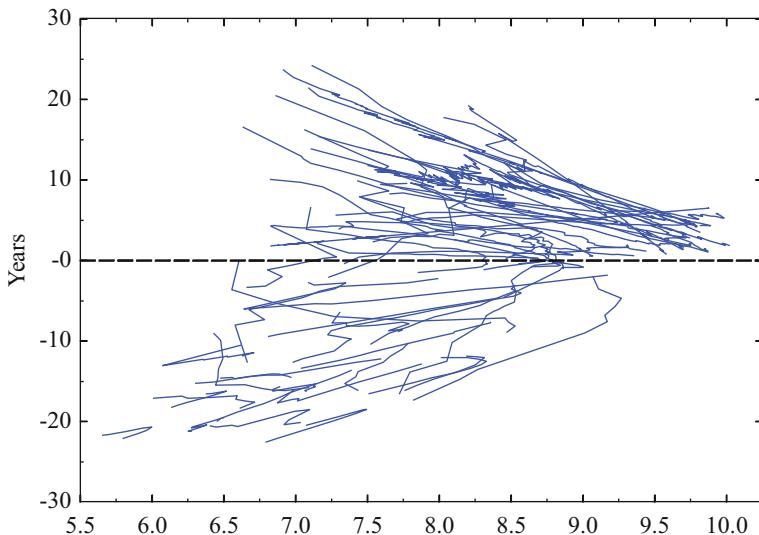


Fig. 16.8 National female life-expectancy estimates against GDP per capita

that some of the low life expectancy countries are also converging, although perhaps not at the same temporal pace when the disadvantage is greater than 10 years. It should also be noted that these data stop in 1994 and do not show the recent effects of the AIDS epidemic.

Plotting the same data against GDP in Fig. 16.8 is even more striking. Now it seems much clearer that there is overall convergence in life expectancy as income increases. There doesn't seem to be much evidence that there are different points of convergence for some countries, although the poorest ones are not well represented in Maddison's data.

In general, examination of the level 1 residuals shows that the fits for each country are very good, but there is temporal autocorrelation that has not been removed.¹⁸ An exception to the encouraging results concerns the former communist countries of central and Eastern Europe. Paradoxically, the model gives better fits in the 1990s during and after the transition problems than it does in the 1960s. There is a consistent pattern of under-estimation in that period. One plausible possibility is that the economies were difficult to quantify and the GDP estimates are too low. The alternative explanation is that command economies really could deliver good health at lower cost when the challenge was to keep infants, children, and adults alive. Perhaps it was the shift of focus to the elderly population that created difficulties.

¹⁸ Autocorrelation is ignored in this presentation, although the MLwiN software used is capable of dealing with autocorrelation in irregularly recorded time-series. See Goldstein et al. (1998).

16.9 Convergence

Many researchers have commented on the apparent convergence in life expectancy over time, and also identified a significant number of countries where the progress in health easily outruns their economic performance (Caldwell 1986). Despite this, there seems to have been no attempt to address convergence in a formal way, although there is an extensive literature in Economics on estimating convergence across countries (Jones 1998; Barro and Sala-I-Martin 1995). Among the many insights this provides for demographic studies is the distinction made between “sigma” and “beta” convergence. The former, as the name suggests, is assessed by calculating cross-sectional standard deviations (sigmas). Their evolution over time is then examined. If the trend is towards lower dispersion, it may be interpreted as convergence.¹⁹ Sigma convergence is a measure of what actually happened, but calculating it is difficult with the demographic data before 1950 because they are unbalanced and sporadic.

Beta convergence exists if the slope of a regression line over time is negatively related to the intercept. For example, if a country has a high starting position, or intercept, and a negative slope over time relative to the average, then it is an indication that this country may eventually converge towards the mean. Beta convergence is really a measure of the propensity to converge, holding other effects constant. While it may lead to sigma convergence, it may also be overtaken by changes in other variables or random shocks.

The multilevel model facilitates the estimation of beta convergence, because of the explicit national-level intercepts and slope parameters, but two departures from the conventional intercept-slope regression approach should be borne in mind when interpreting the results. Firstly, all the “national” parameters are estimated from zero-mean normal distributions, so they can be thought of as offsets from the fixed or international model. Secondly, the intercepts in this paper are estimated at the grand means of the data. This is not a requirement of the method, although conventional in multi-level modelling, but it means that the variance of the intercept distribution has meaning within the scales of the data. This allows one to compare the variances across parameters and convert the parameter estimates into z-scores for comparative purposes.

Returning to a consideration of Eq. 16.4b, the national parameter estimates $\beta_{0,j}$, $\beta_{1,j}$, and $\beta_{2,j}$ may reveal correlation. Suppose that the intercept terms, the $\beta_{0,j}$ are negatively correlated with the lead/lag parameters, the $\beta_{1,j}$. In this case a high (low) intercept will be associated with a falling (rising) national offset over time, and the country will tend to converge on the international model. The same applies to the $\beta_{0,j}$ and $\beta_{2,j}$ parameters. Negative (positive) correlation means that a country with a good starting discount will have a falling (rising) response with respect to income. Finally,

¹⁹There is a debate in the Economics literature on the difference between convergence and regression to the mean, see Friedman (1992) and Quah (1993).

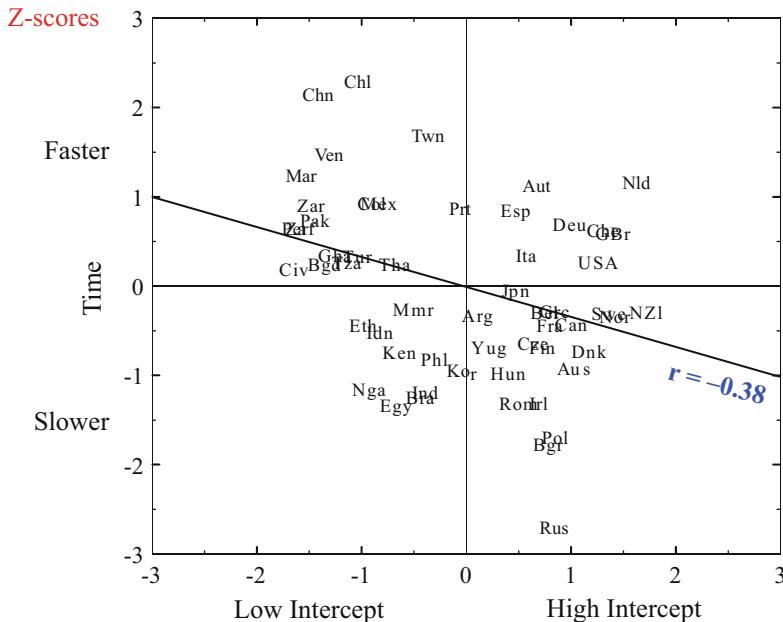


Fig. 16.9 National level parameters for female life expectancy

if $\beta_{1,j}$ and $\beta_{2,j}$ are negatively (positively) correlated, then relative gains with respect to one variable may be offset (reinforced) by the other.

Before examining convergence in the life expectancy data, we need to consider whether the income data are converging, although the Preston model reveals that this is not a necessary condition for convergence. If there were a simple linkage between income and survival, then economic convergence might be driving mortality convergence directly. Economic convergence is a subject of great debate and the general opinion seems to be that there is no unconditional convergence (Jones 1998). Some groups of countries seem to be converging within their own “club”, but it has been argued that this is conditional on the structures of their economies.²⁰ My own very simple attempts to look at convergence in the full Maddison GDP dataset, using a multilevel model without covariates, also indicate that there is no evidence of global convergence. In fact, divergence is suggested.

Figure 16.9 plots the national intercepts, $\beta_{0,j}$ against the parameters on Time, $\beta_{1,j}$, after they have been converted to z-scores. There is a negative correlation of -0.38 , which indicates a slight but statistically significant tendency for countries with high intercepts to decline against time, and *vice versa*, leading to convergence.²¹ There is some evidence of clustering by geographic areas and economic types. On this time

²⁰As an epidemiological example, we might speculate on whether the malarial and non-malarial nations could converge on different trajectories.

²¹The country abbreviations are shown in Table 16.2 at the end of the chapter.

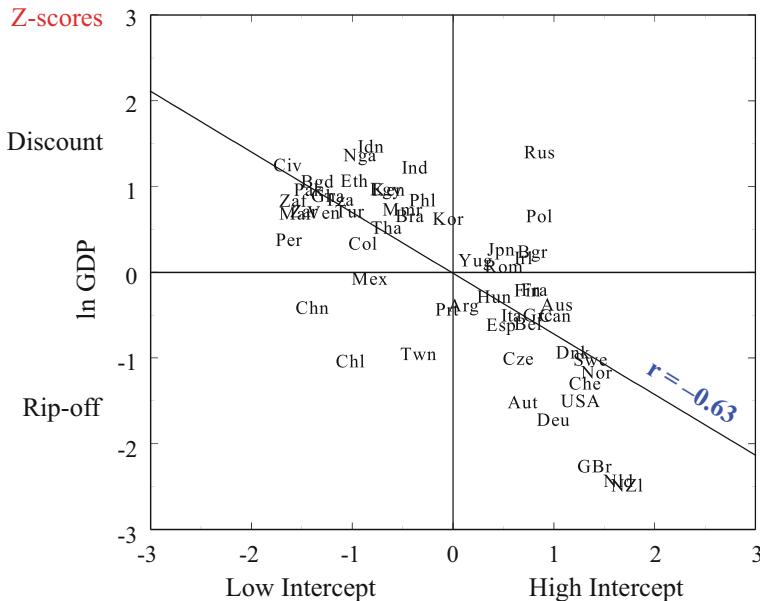


Fig. 16.10 National level parameters for female life-expectancy

scale, advanced economies with high life expectancies seem to be progressing faster. These correspond to the economies that Sachs and Warner identified as “open” to international trade in 1960 (Sachs and Warner 1995). By 1992, the few remaining “closed” economies are confined to a peripheral arc running from China to Egypt. The former communist countries are making slower progress.

Figure 16.10, with a correlation of -0.63 shows that income is the major factor driving beta convergence. As incomes rise, laggard countries are adding years faster than the leaders. Countries of north-west Europe, and their wealthier former colonies, seem to be doing badly, with Britain, the Netherlands, and New Zealand in the worst position. Figure 16.11 shows that the correlation between the time and income parameters, $\beta_{1,j}$ and $\beta_{2,j}$ is also negative and significant at -0.39 , indicating that the effects are offsetting to some degree. Two countries are worth highlighting across all three graphs. Japan is the only wealthy country that is close to the origin on all three scales. Ireland seems to be associated with the countries of the former communist bloc!

The full possibilities of this model are not covered here, but some work has been done. For exploratory purposes one can treat the parameter estimates as data. The intercept for females is strongly related to educational attainment as measured in 1985, so the broad ranking may be a long-run feature. There is some evidence that countries with mid-range average education of between 5 and 9 years have the positive time-parameters that indicate “catch-up”. The GDP parameter is strongly and negatively associated with education. It seems that the educated countries are in a phase when additional years of life expectancy are “expensive”.

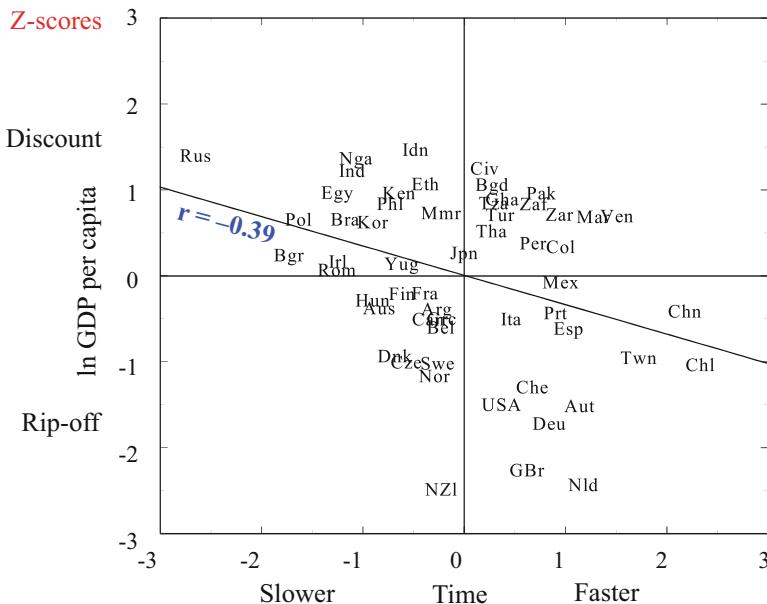


Fig. 16.11 National level parameters for female life-expectancy

The effect of income inequality on life expectancy has been a subject of debate (Wilkinson 1998). It could be that, with the international parameter controlling for the broad effect of GDP, the national parameter might be associated with the income distribution of the country. Countries where there is high inequality may have below par conversion of income into health. Using contemporary data from the World Bank on the share of income held by the lowest 20%, there seems to be no relationship with the GDP parameters. However, the more egalitarian countries do seem to have higher intercepts, but they also have lower rates of change over time. No relationships are significant when the data are restricted to the 17 countries described by Maddison as “advanced capitalist”, although the hypothesis is only expected to apply to richer countries.

Because the model is parameterised at the “national” level it would be possible to use country-specific GDP forecasts to forecast life expectancy. Another experiment is to enter the US GDP time-series into each country’s equation. I expected that this would lead to impossible values, but they looked plausible. For example, the model suggests that in the post-Independence era, Indian women would have had US life expectancy if they had had US incomes. The only forecast that exceeded the “best practice” line in Fig. 16.1 was for Russian women. In general, there seemed to be little evidence that there were “structural” limits in the fitted equations that would prevent life expectancy approaching the best levels if incomes grow. One of the insights from this exercise was that some countries

Table 16.2 56 countries

Country	Abbreviation	Country	Abbreviation
Argentina	Arg	Korea	Kor
Australia	Aus	Mexico	Mex
Austria	Aut	Morocco	Mar
Bangladesh	Bgd	Myanmar	Mmr
Belgium	Bel	Netherlands	Nld
Brazil	Bra	New Zealand	Nzl
Bulgaria	Bgr	Nigeria	Nga
Canada	Can	Norway	Nor
Chile	Chl	Pakistan	Pak
China	Chn	Peru	Per
Colombia	Col	Philippines	Phl
Cote d'Ivoire	Civ	Poland	Pol
Czechoslovakia	Cze	Portugal	Prt
Denmark	Dnk	Romania	Rom
Egypt	Egy	Russia	Rus
Ethiopia	Eth	South Africa	Zaf
Finland	Fin	Spain	Esp
France	Fra	Sweden	Swe
Germany	Deu	Switzerland	Che
Ghana	Gha	Taiwan	Twn
Greece	Grc	Tanzania	Tza
Hungary	Hun	Thailand	Tha
India	Ind	Turkey	Tur
Indonesia	Idn	UK	GBr
Ireland	Irl	USA	USA
Italy	Ita	Venezuela	Ven
Japan	Jpn	Yugoslavia	Yug
Kenya	Ken	Zaire	Zar

have trajectories that are independent of GDP. Chile and China, for example, have national parameters that are approximately equal to the international parameter on GDP, but with a negative sign, so that changing GDP has no effect at all.

16.10 Conclusion

Multilevel models offer considerable scope for disentangling effects in collections of unbalanced, repeated-measures data. These results are preliminary and designed to explore what can be done, rather than suggesting final interpretations. Preston's finding that time seems to be becoming less important in LDCs is contradicted in the international component of this model, where it is the income effects that seem to

have diminished in the post-war era, particularly for men. On the other hand, changes in income seem to be more important than health technology in explaining survival convergence. Breaking down the national level effects into their constituent components suggests that countries of Northwest European origin have translated a diminishing proportion of their gains in income into gains in health. Combined with the “catch-up” opportunities for the laggard countries, this has led to rapid convergence. Japan and the Southern European countries seem to be the exceptions to this diminishing return to log income.²² They seem to have emerged from the “pack” by maintaining a small long run advantage over the international position. The story of why women’s patterns are different from men’s will have to wait for another day.

References

- Barro, R. J., & Sala-i-Martin, X. (1995). *Economic growth*. London: McGraw-Hill.
- Bloom, D. E., & Williamson, J. G. (1998). Demographic transitions and economic miracles in emerging Asia. *World Bank Review*, 12(3), 419–455.
- Caldwell, J. C. (1986). Routes to low mortality in poor countries. *Population and Development Review*, 12(2), 171–220.
- Easterlin, R. A. (1996). *Growth triumphant: The twenty-first century in historical perspective*. Chicago: University of Michigan Press.
- Fogel, R. W., & Costa, D. L. (1997). A theory of technophysio evolution, with some implications for forecasting population, health care costs, and pension costs. *Demography*, 34(1), 49–66.
- Friedman, M. (1992). Do old fallacies never die? *Journal of Economic Literature*, 30, 2129–2132.
- Goldstein, H. (1995). *Multilevel statistical models* (2nd ed.). London: Arnold.
- Goldstein, H., et al. (1998). *A user's guide to MLwiN*. London: Institute of Education.
- Graff, H. J. (1987). *The legacies of literacy: Continuities and contradictions in western cultures and society*. Bloomington: IUP.
- Houston, R. A. (1988). *Literacy in early modern Europe: Culture and education 1500–1800*. London: Longmans.
- Jones, C. I. (1998). *Introduction to economic growth*. London: Norton.
- Kreft, I., & de Leeuw, J. (1998). *Introducing multilevel modeling*. London: Sage.
- Maddison, A. (1995). *Monitoring the world economy: 1820–1992*. Paris: OECD.
- Maddison, A. (2001). *The world economy: A millennial perspective*. Paris: OECD.
- Oeppen, J., & Vaupel, J. W. (2002). Broken limits to life expectancy. *Science*, 296(5570), 1029–1030.
- Preston, S. H. (1975). The changing relation between mortality and level of economic development. *Population Studies*, 29(2), 231–248.
- Preston, S. H. (1976). *Mortality patterns in national populations: With special reference to recorded causes of death*. London: Academic.
- Preston, S. H. (1980). Causes and consequences of mortality in less developed countries during the twentieth century. In R. Easterlin (Ed.), *Population and economic change in developing countries*. New York: NBER.
- Preston, S. H. (1985). Mortality and development revisited. *Population Bulletin U.N.*, 18, 34–40.
- Quah, D. (1993). Galton's fallacy and tests of the convergence hypothesis. *Scandinavian Journal of Economics*, 95(4), 427–443.

²²Measurement problems of the “black economy” in southern Europe should also be considered.

- Sachs, J. D., & Warner, A. M. (1995). *Economic reform and the process of global integration* (Brookings Papers on Economic Activity, pp. 1–118). Washington, DC: Brookings Institution.
- Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.
- Temple, J. See website at: <http://www.bris.ac.uk/Depts/Economics/Growth/> for a guide to data and literature.
- van Elkan, R. (1996). Catching up and slowing down: Learning and growth patterns in an open economy. *Journal of International Economics*, 41, 95–111.
- White, K. M. (2002). Longevity advances in high-income countries, 1955–96. *Population Development Review*, 28, 59–76.
- Wilkinson, R. (1998). *Unhealthy societies: The afflictions of inequality*. London: Routledge.
- Wilson, C. (2001). On the scale of demographic convergence, 1950–2000. *Population Development Review*, 27(1), 155–171.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 17

Linear Increase in Life Expectancy: Past and Present



Tommy Bengtsson

Improvements in human stature, real income and life expectancy have taken place at an unprecedented speed during the last 200 years. In the case of life expectancy at birth, the record has been broken at an amazingly constant pace since 1840. Females have continuously gained 2.92 months per year, males slightly less (Oeppen and Vaupel 2002). While the increase is considerable, with improvements in life expectancy of some 8 years from parent to child, it is the regularity of the advancement that is remarkable, not the speed as such. The reason is that the countries entering the mortality transition at a later stage in history tend to exhibit an even faster improvement. Japan, for example, experienced improvements of 6 months per year in life expectancy during its catch-up in the twentieth century. In China the corresponding figure was well above 1 year in the 1960s and 70s. Thus, while the increase of best-practice countries is not astonishing in itself, the linearity of the improvement certainly is. This raises obvious questions: What are the causes of this linear increase and how long can it be sustained? Other questions concern whether the observed linear increase in life expectancy can be used in forecasting life expectancy both for countries lagging behind and countries in the lead.

To Oeppen and Vaupel, the linear development of life expectancy suggests that the process of mortality reductions “should not be seen as a disconnected sequence of unrepeatable revolutions but rather as a regular stream of continuing progress” (Oeppen and Vaupel 2002:1029), referring to Lee and Carter (1992), and Tuljapurkar et al. (2000). Still, in the next sentence Oeppen and Vaupel state, with reference to Riley (2001), that mortality improvements are the result of a complex process of “advances in income, salubrity, nutrition, education, sanitation, and medicine, with the mix varying over age, period, cohort, place, and diversity” (Oeppen and Vaupel 2002:1029). Thus, on the one hand the advance in life

T. Bengtsson (✉)
Centre for Economic Demography, Lund University, Lund, Sweden
e-mail: tommy.bengtsson@ekh.lu.se

expectancy is a regular stream of continuous progress but on the other it is also an intricate interplay of a mixture of social, economic and medical factors, which sounds almost like a paradox that calls for further clarification. As starting point, I will take a closer look at the countries that are the leaders in life expectancy: a very small number indeed, consisting of nine countries altogether. What mixture of factors varying over *age, period, cohort, diversity of diseases*, and *place* has made them the global leaders of life length? I will then turn to the issue of causality.

17.1 Descriptive Overview

Starting with *age and period*, to quote Oeppen and Vaupel ([2002](#):1029), “most of the gain in life expectancy was due to large reductions in death rates at younger ages. In the second half of the 20th century, improvements in survival after age 65 propelled the rise in the length of people’s lives”. This is indeed one of the most well-known and universal facts regarding the historical mortality transition. It holds true both for countries that have experienced the mortality decline recently and for those where it started hundreds of years ago. In Sweden, an example of the latter, the level of infant mortality dropped almost without interruptions from the mid-eighteenth century onwards. The other Nordic countries showed a similar development (Bengtsson and Lundh [1999](#)). England and other parts of Western Europe, as well as North America, initially followed the same pattern of infant mortality decline in the eighteenth and the beginning of the nineteenth century but after that point in time, the decline levelled out. Alfred Perrenoud ([1984](#)) consequently differentiates between a Nordic model, with continued decline, and a West-European model, with an interrupted decline. From around the end of the nineteenth century to present day, however, all countries in the industrialized world have exhibited the same development of rapid decline in infant and child mortality, which was the main reason for the rise in life expectancy. Though death rates for the elderly started to drop already in the latter part of the nineteenth century, it was not until the mid-twentieth century that life expectancy was largely *propelled* by falling mortality at age 65 and above. Since the change from infant and child mortality to mortality among the elderly as the major explanation of the observed improvements in life expectancy has not yet occurred in many countries, most of the increase in life expectancy during the twentieth century in these countries has still been due to the decline in infant and child mortality.

Another striking and commonly-shared pattern is the change over time in the *diversity of diseases*. Here Oeppen and Vaupel refer to Bongaarts and Bulatao’s volume *Beyond Six Billion* ([2000](#)) and Riley’s book *Rising Life Expectancy: A Global History* ([2001](#)). In turn, Riley’s way of reasoning is much in line with the Epidemiological Transition Theory, according to which the pestilences and famines are followed by receding pandemics, and later by degenerative and man-made diseases (Omran [1971](#)). This line of theory argues that the changes are mainly the result of man’s control over his environment. How mortality patterns change over

time is not the object of much controversy; more so the causes behind these changes. General agreement prevails that decline in infant and child mortality starts off in the form of reduced mortality in highly virulent infectious diseases but is upheld by a decline in less virulent infectious diseases. The role of famines in this process is more debated but it is unlikely that it has influenced trends in infant and child mortality to any significant degree (Wrigley and Schofield 1981). Thus, it was a reduction in the mortality among children in highly virulent diseases, primarily a drop in smallpox mortality, together with further reductions in mortality in low virulent infectious diseases, which brought the best-practice countries to the lead in 1840.

The second crucial change came after the mid-twentieth century and consisted of the fall in mortality in ages 65 years and above. This decline was largely caused by reductions in chronic diseases (cardiovascular, cerebrovascular and some cancer diseases), possibly in combination with a general health amelioration due to improvements of conditions in early life in the beginning of the twentieth century. In fact, when the great mortality decline of Western Europe was first analysed by Derrick (1927) and Kermack et al. (1934), they clearly advocated the role of *cohort factors*. They believed that the decline in adult mortality in England, Wales, and Scotland, as well as in Sweden, to a large extent was the result of improved conditions in childhood. This argument later lost ground as the Demographic Transition Theory evolved and the process of modernization and other *period factors* came into focus (UN 1953). It was not until the late seventies that Preston and van de Walle (1978), Fridlizius (1989) and later Barker (1994) and Fogel (1994), brought these issues into play again. Today, there is an extensive and lively debate on the importance of early-life factors on mortality in later life vis-à-vis period factors.

Regarding *place*, it is striking that so few countries have been record-holders in female life expectancy over the past 160 years, as shown in Table 17.1. The linear line is based on the experience of only nine countries, starting off with Sweden and Norway in the first two decades after 1840, and followed by Australia in the 1860s. New Zealand came into the lead in 1876 and almost monopolised this position until

Table 17.1 The nine world record holding countries

Country	Years in lead	Starting year	Last year
Sweden	10	1840	1973
Norway	48	1841	1969
Australia	8	1861	1920
New Zealand	57	1876	1947
Denmark	2	1918	1919
Iceland	19	1941	1984
The Netherlands	1	1964	1964
Switzerland	1	1985	1985
Japan	15	1986	2000

Source: Oeppen and Vaupel (2002, Supplementary Material, Table 2)

1941, with the exception of a brief period between 1916 and 1919 when first Sweden and then Denmark surpassed all others. During the Second World War, Norway and Sweden, now in company with Iceland, showed the best-practice levels in life expectancy and remained leaders until 1986 when Japan took over the lead position. Except for 2 years, during which the Netherlands and Switzerland respectively out-performed all other countries, Japan has continued to hold the world record until the present day. Thus, three of the nine countries demonstrate outstanding life expectancy figures for very limited periods, i.e. a year or two, and for one of these, Australia, annual data is in fact lacking. Only two single values are used for Australia, of which one is used to cover a 7-year period in the 1860s and 1870s (Oeppen and Vaupel 2002, Supplementary Material, Table 2).¹ That leaves five countries which among them share the world record for 148 out of the 160 years that the analysis covers: Sweden, Norway, New Zealand, Iceland, and Japan. The focus will hence be placed on these five leading nations.

Sweden and Norway have the highest levels of life expectancy from 1840 until data become available for New Zealand in 1876. For a decade, the New Zealand level exceeds the two Nordic countries with almost 5 years. However, the improvement is somewhat faster in the two Nordic countries than in New Zealand, which means that they converge and finally catch-up in the beginning of the 1940s, as shown in Fig. 17.1. This is the period in which the decline in infant and child mortality drives up life expectancy. It is also the period when mortality in low virulent infectious diseases declines. Thus, the question is what makes these three countries stay ahead of all others as regards mortality in infectious diseases among infants and children during this 100-year period. Further, if comparing the three, why is New Zealand unable to improve as fast as Sweden and Norway and keep its superiority?

In the period after the Second World War, in the countries that started the mortality decline in the nineteenth century, life expectancy was mainly driven by reductions in mortality among the elderly due to a reduction in chronic diseases. Up until 1985, when Japan took over the lead, Iceland, and for some years Norway and Sweden, exhibit the highest life expectancy rates. It is noticeable that while New Zealand, Norway, Sweden, and also Iceland, level off as they come into the phase where life expectancy no longer mainly is driven by reductions in infant and child mortality, Japan does not. Once again, the question is what factors make these five countries more successful than the rest of the world.

¹The data that Oeppen and Vaupel use stems from Rowland and cover periods between 9 and 16 years in length. The value for 1861–63, 1867–68, and 1874–75, when Australia is in lead, is based on the average of the period 1860–1875, in which the value is lower than for Norway. Since we don't know if it is in the lead in any of these particular years or not, we leave Australia out of the following discussion.

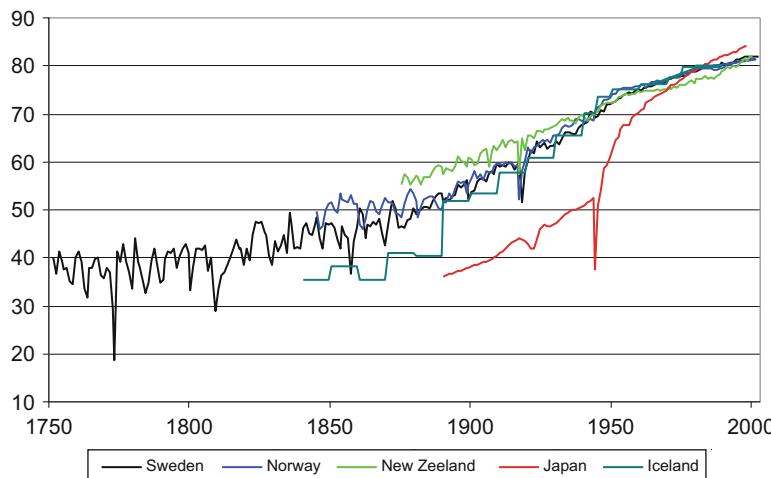


Fig. 17.1 Female life expectancy 1751–2001. (Sources: *The Human Mortality Database*, University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany), available at www.mortality.org or www.humanmortality.de (data downloaded on 7 October 2005); *Hagskinna. Icelandic Historical Statistics* (1997). Jónsson G. and M.S. Magnússon (eds.); and Nanjo and Kobayashi (1985))

17.2 Causes

Beyond Six Billion by Bongaarts and Bulatao (2000:117–123), which Oeppen and Vaupel refer to as regards changes in lethal diseases, not only gives an overview of the diversity of diseases during various phases but also supplies an explanation much in line with Riley (2001) and Omran (1971) and which can be summarised as follows:

1st stage 1700–1800.

Reduction in volatility/epidemics due to

1. more stable consumption (better storage and transportation) giving higher resistance
2. decline in virulence of pathogens

2nd stage 1800–1900.

Reduction in infectious diseases (influenza, pneumonia, bronchitis, TB, and smallpox) due to

1. increasing standard of living through improved health behaviour
2. public health measures including smallpox vaccination

3rd stage 1900–1960.

Reduction in infectious diseases due to

1. reduced exposure
2. reduced transmission

4th stage 1960–1996.

Reduction in chronic diseases (cardiovascular, cerebrovascular and some cancer diseases) due to

1. early detection and prevention
2. improvement in surgical procedures
3. refinement of medical therapies

The question is whether these factors explain why the five countries are in the lead and why the development of female life expectancy is linear. How well did Norway, Sweden and New Zealand do in terms of standard of living, improved health behaviour, and public health measures in comparison to other countries up until the Second World War?

Regarding Norway and Sweden, they did not perform well in terms of living standards in the latter 1800s. GDP per capita in 1870 (in 1990 international dollars) was low compared with the US, the Netherlands and England, as shown in Table 17.2.

While real wages for workers in Sweden increased after 1870 (Bengtsson and Dribe 2005; Jörberg 1971), we are less certain of the development prior to that. In fact, during the first part of the nineteenth century, the majority of people might even have been worse off than during the eighteenth century (Bengtsson and Dribe 2005). GDP per capita for Norway was even lower than for Sweden. In New Zealand, GDP per capita was well above the Nordic countries, to the extent that it was almost on a par with the most advanced European countries. While the ranking of New Zealand versus the two Nordic countries correctly reflects their position regarding life expectancy, the rest is contrary to what we could expect given the countries' economic performance.

The Nordic countries are today known for their flat income distribution. Historically, though, this has not been the case. The income distribution in the beginning of the nineteenth century – a period of commercialization of agriculture and rapid economic transformation – was most likely spread. Several indicators show that while the income for landowners increased, it decreased for non-landed groups making the latter more vulnerable to variations in food prices (Bengtsson 2004; Bengtsson and Dribe 2005).

Turning to *health behaviour*, we have no evidence that Norwegians and Swedes were better-off than other Europeans; rather the opposite. Malthus, for example,

Table 17.2 GDP per capita in 1870 (in 1990 international dollars)

Country	GDP/capita
United Kingdom	3191
Netherlands	2753
United States	2445
New Zealand/Australia/Canada	2339
Sweden	1664
Norway	1432

Source: Maddison (2001, Table B-21)

reports from travelling in Scandinavia that Swedes were dirty and poorly fed (James 1966:67). They did not even have proper inns for travellers. Malthus statement refers to the beginning of the nineteenth century but similar depictions can be found well into the 1930s, when the depreciatory concept *Lortsverige* (Dirty-Sweden) was coined (Nordström 1938).

Riley claims that Sweden was very successful in terms of public health measures (Riley 2001). He states that the vaccination campaigns eradicated smallpox. While it is true that Sweden experienced rather few deaths in smallpox mortality after vaccination started in 1801, most of the decline in smallpox mortality, nevertheless, took place before this time (Bengtsson 1998, 2001; Sköld 1996). Smallpox vaccination was not the only public health improvement undertaken. Other measures include breast-feeding campaigns, education of midwives, disease control, investments in water supply and sewage systems, and promotion of improved personal hygiene.

Without doubt, investments in water supply and sewage control did have effects on mortality in urban areas after c. 1880, even more so in more urbanized countries than Norway and Sweden, both of which both were predominantly rural countries at this time. In fact, most cities in both Norway and Sweden in the mid-nineteenth century, when the two were in the lead, were small; no more than large villages by contemporary West European standards. By the same standards, only the two capitals could rightly be defined as cities, which gave Norway and Sweden a comparative advantage at a time when water supply and sewage control still were insufficient in most urban regions. The urban toll was thus less heavy in less urbanized countries, like Sweden and Norway, than in the more developed parts of Western Europe.

As for other public health measures, such as the breast-feeding campaigns that took place in Sweden from the 1830s and onwards, a certain local impact in areas where breast-feeding was uncommon was reported (Brändström 1984), but the influence on national levels of infant mortality was slight. Incidentally, childhood mortality in fact went up in Sweden during the period when these campaigns started (Fridlizius 1984). Other measures, like the training of midwives, commenced about the same time as the breast-feeding campaigns. Taken together, it is likely that they had some impact on the infant mortality decline from around the 1830s onwards.

The question of improved storage of food and other undertakings aimed at stabilizing consumption which, according to Bongaarts and Bulatao (2000) were important for the mortality reduction in the eighteenth century, constitutes a controversial issue. Firstly, we have no evidence of such measures being taken in the Nordic countries and secondly, it is unlikely that such initiatives would have influenced trends in infant and child mortality to any significant degree (Wrigley and Schofield 1981).

To summarize, Sweden and Norway had no advantage in terms of living standards or equal income distribution in the nineteenth century when they exhibited the highest level of recorded life expectancy in the world. Neither were they particularly well-organized with regard to, for example, the poor relief system. However, several public measures, like the above mentioned education of midwives and breast-

feeding campaigns, were carried out in order to improve public health. They were also favoured by a low degree of urbanization. But perhaps most important of all, they were lucky to escape from highly virulent diseases, i.e. smallpox, even before vaccination programs started.

For New Zealand, which held the position of world leader in life expectancy for 57 years of the 160-year period, other factors were instrumental. In 1876, the first year data is available for New Zealand, there are indications showing that the country's white population was the most long-lived on Earth and that the level of living standard was well above the Nordic countries, as well as most other countries of the world. The situation also differed in that a large share of its rapidly increasing population was made-up by immigrants. Furthermore, not only is it reasonable to assume these immigrants to be of reasonably good health, in particular considering that they had travelled long distance, most of them were selected for their qualities and also underwent at least one health test, for phthisis (Pool and Cheung 2005). New Zealand also became a well-organized society with an expanding public sector. It can easily be compared to Britain in terms of health organisation and institutions but without the burden of large cities and with a selected population, thus almost as a natural experiment with the UK as a point of departure.² Still, however, improvements in life expectancy were slower than in Norway and Sweden. Thus, the lead for New Zealand is much a result of its initial superiority: a selected, well-fed, well-organized population in combination with a low disease load.

In comparing New Zealand with Norway and Sweden, and the three of them with the rest of the world, it is difficult to distinguish any "regular stream of continuing progress" that could explain why these three specific countries on a global scale perform so outstandingly in life expectancy. It is seemingly a basket of factors, partly different in New Zealand from the two Nordic countries, which make them stay at the top. They share the low degree of urbanization and a low disease load and differ in terms of economic resources. They are well-organized in that public schooling was introduced early and that they took public measures to improve health, but so did many other countries without advancing to the lead position and without contributing to the linear development of life expectancy.

Moving on to the latter part of the twentieth century, this period is characterised by *propelled* life expectancy due to declining death rates among the elderly, mainly as an effect of mortality reduction in chronic diseases (cardiovascular, cerebrovascular and some cancer diseases). According to Bongaarts and Bulatao (2000) this is brought about by early detection and prevention of diseases, improvements in surgical procedures, and refinements of medical therapies. The question is why Iceland and, post-1986, Japan perform better than all other countries within these areas.

Starting with Iceland – the world leader for 19 years between 1941 and 1984 – it has an historical development that sets it apart from the other Nordic countries.

²Jim Oeppen made this point during the workshop that this volume is based on. He also noticed that life expectancy on the English country side was on a par with New Zealand during the latter part of the nineteenth century.

While Iceland had a stagnant population throughout the nineteenth century, the other countries experienced steady population growth. In addition, mortality was higher and life expectancy was lower than in any of the other Nordic countries. The Icelandic life expectancy at birth during the 1880s is about 10 years less than Norway and Sweden. In the last decades of the nineteenth century and in the beginning of the twentieth century, however, mortality drops dramatically and in the 1940s, Iceland is on a par with Norway and Sweden, as shown in Fig. 17.1. Causing the catch-up is a fall in infant and child mortality. Due to the high infant and child mortality in the latter part of the nineteenth century, the proportion surviving to ages above 65 years in the 1940s was very low compared with the countries in the lead. Thus, the population at risk in Iceland is constituted in a different way than New Zealand, Norway and Sweden since the females in older ages represented a much smaller proportion of their birth cohorts than in the other three countries. As the difference with the other countries diminishes, Iceland follows the same pattern of life expectancy as Norway and Sweden. In some of the years in the post-World War II period their life expectancy is slightly higher, in some years it is lower.

Japan differs somewhat from the other countries and had it not been for this country, the linear development would have been interrupted by the 1980s since the life expectancy in Iceland, Norway, Sweden, and New Zealand is levelling off. Japan is a country with a very high income level, attained only over the last decades. Previous to its rapid economic development from the 1960s onwards, the situation was less favourable and life expectancy was rather low. In fact, Japan is an example of one of the newly industrialized countries that have shown a tremendous development during the course of the twentieth century. Life expectancy for women rose from 60 years in 1950 to 80 years in 1984. This is not a development of life expectancy by 2.92 but by 6.00 months per year! Is this due to early detection and prevention of diseases, improvements in surgical procedures, and refinements of medical therapies causing mortality among the elderly to decline? It is surely not; rather it is an exemplary case of the catch-up of a rapidly developing country. The increase in life expectancy is not mainly driven by mortality improvements solely among the elderly but within a wider age span.

The rapid change in Japan also has some characteristics in common with Iceland that are important in making comparisons with the old leaders. The rapid transition from high to low mortality within a short period of time means fewer elderly relatively speaking. For example, while 90% of the women born in Sweden in 1935 have reached the age of 65 years, the corresponding figure for Japan is only 60%³. Thus the population age structure in 2000 is entirely different in Japan comparative to Sweden. This has two implications. First, the Japanese women at higher ages contributing to the period life-expectancy in year 2000 are highly selected in comparison to the Swedish older women. If the “scarring” effect is

³Based on data from *Population Statistics of Japan* 2003, complied by National Institute of Population and Social Security Research, Japan, and Befolkningsutvecklingen under 250 år (*Population Development in Sweden in a 250-year perspective*), Demografiska rapporter 1999:2, Statistics Sweden.

smaller than positive selection effects then the Japanese elderly would be expected to be less fragile than their Swedish counterparts⁴. Second, due to its late fertility transition, the elderly are relatively speaking fewer in Japan compared with Sweden. While the proportion 65 years and over in Sweden reached 8% in 1900, the corresponding figure was reached in Japan more than 70 years later. Because the Japanese elderly are fewer, relatively speaking, the costs for pensions and care are smaller for the Japanese vis-à-vis the Swedish working generation.

17.3 Summary and Discussion

We started off by pointing out that very few countries, only nine, have held the leading position in life expectancy at birth over the last 160 years, and only five of them did so for more than a few years. Typically, new countries caught up with and replaced former best-practice countries for several decades. However, we can note that this has not only been the case for life expectancy but has also occurred in other areas of human activity, such as economic performance. The total number of leaders in GDP per capita corresponds in size to the number of leaders in life expectancy, albeit they are not the same countries. The list of GDP leaders, ranked from high to low according to years at the lead and for roughly the same period, is as follows: the USA, Australia, Switzerland, the UK, and New Zealand (Maddison 2001). The fact that this list is different from that for life expectancy, emphasizes the relevance of Easterlin's conclusions that the mortality decline since the mid-nineteenth century was not mainly driven by economic development (Easterlin 1999). Instead, there are a variety of factors that lie behind the improvements in life expectancy, economic factors constituting only one of these.

High life expectancy in Sweden and Norway in the mid-nineteenth century was due to low infant and child mortality, in particular in infectious diseases. It was not the result of private or societal investments, strong nation-states, high living standards or an equal income distribution. What favoured the Nordic countries was instead a low degree of industrial development and urbanization combined with a low disease load. When New Zealand took over the lead position in 1876 (the first year for which there is data), it shared these favourable characteristics of the two Nordic countries, in addition to other favourable conditions such as high income levels and a positively selected population.

Mortality in all age-groups dropped quickly in the early twentieth century but the improvements in life expectancy were still mainly due to the decline in infant and child mortality. During this period it was largely due to societal investments of reducing disease exposure; directly through eradication programs, indirectly through water purification systems and better transportation. This was the case for the most developed countries. For example, the previously large differences between hospital wards in many US cities disappear within a few decades (Fogel 2004). The same

⁴See Preston et al. 1998, p. 1232.

process was at hand in countries like Singapore, Sri Lanka, Argentina, Costa Rica, Chile, Cuba, and Uruguay, which later affected the mortality transition (Bongaarts and Bulatao 2000:124). Still, none of these countries reached the highest level of life expectancy. New Zealand remained in the lead and it was instead the Nordic countries, making the same sorts of societal investments, which caught up in the 1940s. Thus, societal investments in infrastructure and high per capita income alone were not that important in determining life expectancy at birth. If this were the case, the US, the UK, Australia, and New Zealand would be at the top.

Iceland had been lagging far behind the leading nations when it started its catch-up which brought it to lead in the 1940s, and for Japan this was even more the case before it became the world leader in the 1980s. Both countries, especially Japan, had a very small elderly population when reaching the top position in terms of life expectancy. Today, life expectancy in the developed countries is not entirely the result of low infant and child mortality, as used to be the case, but instead old age mortality has become important. Medical care and private investments in health are also of greater significance nowadays. Thus Japan had the advantage of rapid improvements in living standards and societal investments in infrastructure, facilitated by access to new types of medical care, all of which they share with the most advanced countries of the world. In addition, they have a smaller group of elderly in the population. This means that the elderly are a selected group. Since the proportion of elderly is lower in Japan due to its late fertility decline, it also means, *ceteris paribus*, that Japanese senior citizens will have access to more care resources than their equals in countries that have experienced an earlier fertility transition.

Why then has life expectancy in the flow of best-practice countries followed a linear trend for a 160 year period? It is hard to find any specific continuous process that explains such a development. It resembles more the outcome of a diversity of processes, of which some but not all are directly related to human activity. If one compares with economic performance, it is reasonable to ask why life expectancy does not follow the exponential trends of economic output, the signum of economic growth ever since Malthus wrote his first essay, but instead *only* a linear trend? As regards economic development, more resources create more resources for good investments, thus generating an exponential trend. In the Malthusian world, population expanded at an exponential rate and production at a linear rate at best. Here we are confronted with the opposite situation: The economy expands at an exponential rate while the population stagnates and life expectancy develops at a linear rate. The question is then ‘why have we been less successful in investments in health than in economic growth’?

It is reasonable that the best-practice path of linearity could be used for forecasting life expectancy for countries still in the catch-up phase, assuming that they have the economic means and incentives to invest in best-practice technology. Countries with a rapid catch-up, like Iceland and Japan, also has the advantage of having a relatively smaller proportion of elderly for several decades, constituting a sort of population momentum, which works in their favour and which helps them to stay on the path of linear advancement. The momentum, however, disappears after some time and then it is likely that they, like their predecessors at the lead, face a downward bend in life expectancy. Few countries finally became world leaders,

still they tend to be replaced and their life expectancy curve levels off. Thus, new countries are catching up and replacing the former leader and contributing to the linear development of life expectancy.

A parallel within economics can be drawn with the so called Cardwell's Law (Mokyr 1990) economics, which states that no country can maintain technological lead for very long. It implies turnover at the top, just as the case was with regard to life expectancy. Several other economists could be referred to, such as Alexander Gerschenkron who developed the concept of Economic Backwardness (Gerschenkron 1966). Perhaps even more to the point are the concepts of the institutional economist Thorstein Veblen, who coined the phrases "the penalty of taking the lead" and "the advantage of borrowing the technological arts". Both phrases refer to the disadvantage of old investments, or in the case of human capital, "aged" capital vis-à-vis new capital. These concepts therefore seem highly relevant to apply when evaluating the significance of low proportions of elderly (i.e. "aged" capital) in populations that has more recently undergone demographic transition, as for example Japan. Will then the future pace of life expectancy in Japan slack as a result of the penalty of taking the lead (population ageing) as has been the case for previous world leaders? And will new countries take the advantage of a backward position and reach the top ranks in terms of life expectancy? My answer is 'yes' on both of these questions.

References

- Barker, D. J. P. (1994). *Mothers, babies, and disease in later life*. London: BMJ Publishing Group.
- Befolkningsutvecklingen under 250 år* (Population Development in Sweden in a 250-year perspective), Demografiska rapporter 1999:2. Stockholm: Statistics Sweden.
- Bengtsson, T. (1998). The great mortality Decline: Its causes and consequences. In C.-E. Núñez Clara-Eugenio (Ed.), *Debates and controversies in economic history*. Madrid: Fundación Rámon Areces.
- Bengtsson, T. (2004). Mortality and social class in Four Scanian parishes. In T. Bengtsson, C. Campbell, J. Z. Lee, et al. (Eds.), *Life under pressure. Mortality and living standards in Europe and Asia* (pp. 1700–1900). Cambridge, MA: MIT Press.
- Bengtsson, T., & Dribe, M. (2005). New evidence on the standard of living in Sweden during the eighteenth and nineteenth centuries: Long-term development of the demographic response to short-term economic stress. In R. C. Allen, T. Bengtsson, & M. Dribe (Eds.), *Living standards in the past: New perspectives on well-being in Asia and Europe*. Oxford: Oxford University Press.
- Bengtsson, T., & Lundh, C. (1999). Child and infant mortality in the Nordic Countries Prior to 1900 (Lund Papers in Economic History, No. 66). Lund: Department of Economic History, Lund University.
- Bongaarts, J., & Bulatao, R. A. (Eds.). (2000). *Beyond six billion. Forecasting the world's population*. Washington, DC: National Academy Press.

- Brändström, A. (1984). *De kärlekslösa mödrarna. Spädbarnsdödligheten i Sverige under 1800 talet med särskild hänsyn till Nedertorneå*. Umeå: Umeå University.
- Derrick, V. P. A. (1927). Observations on (1) errors in age in the population statistics of England and Wales and (2) the changes in mortality indicated by the national records. *Journal of the Institute of Actuaries*, 58, 117–159.
- Easterlin, R. A. (1999). How benevolent is the market? A look at the modern history of mortality. *European Review of Economic History*, 3(3), 257–294.
- Fogel, R. W. (2004). *The escape from hunger and premature death* (pp. 1700–2100). Cambridge: Cambridge University Press.
- Fridlizius, G. (1984). The mortality decline in the first phase of the demographic transition: Swedish experiences. In T. Bengtsson, G. Fridlizius, & R. Ohlsson (Eds.), *Pre-industrial population change*. Stockholm: Almqvist & Wiksell International.
- Fridlizius, G. (1989). The deformation of cohorts: Nineteenth-century mortality in a generational perspective. *Scandinavian Economic History Review*, 37, 3–17.
- Gerschenkron, A. (1966). *Economic backwardness in historical perspective. A book of essays*. Cambridge, MA: The Belknap Press of Harvard University Press.
- Hagskinna. *Icelandic Historical Statistics*. (1997). Guðmundur Jónsson & Magnús S. Magnússon (Eds.). Reykjavík: Statistics Iceland.
- James, P. (Ed.). (1966). *The travel diaries of T.R. Malthus*. Cambridge: The Syndics of the Cambridge University Press.
- Jörberg, L. (1971). *A history of prices in Sweden 1732–1914* (Vol. II). Lund: Gleerups.
- Kermack, W. O., McKendrick, A. G., & McKinley, P. L. (1934). Death rates in Great Britain and Sweden: Some general regularities and their significance. *Lancet*, 226, 698–703.
- Lee, R. D., & Carter, L. (1992). Modeling and forecasting the time series of U.S. mortality. *Journal of the American Statistical Association*, 419(87), 659–671.
- Maddison, A. (2001). *The world economy. A millennial perspective*. Paris: OECD.
- Mokyr, J. (1990). *Lever of riches*. Oxford: Oxford University Press.
- Nanjo, Z., & Kobayashi, K. (1985). *Cohort life tables based on annual life tables for Japanese nationals covering the years 1891–1982*. Tokyo: Nihon University Population Research Institute.
- Nordström, L. (1938). *Lort-Sverige*. Stockholm: Kooperativa förbundet.
- Oeppen, J., & Vaupel, J. W. (2002). Broken limits to life expectancy. *Science*, 296(5570), 1029–1031.
- Omran, A. R. (1971). The epidemiological transition. A theory of the epidemiology of population change. *Millbank Memorial Fund Quarterly*, 49(4), 509–538.
- Perrenoud, A. (1984). The mortality decline in a long term perspective. In T. Bengtsson, G. Fridlizius, & R. Ohlsson (Eds.), *Pre-industrial population change*. Stockholm: Almqvist & Wiksell International.
- Pool, I., & Cheung, J. (2005). Why were New Zealand levels of life expectation so high at the dawn of the twentieth century? *Genus*, 61(2), 9–33.
- Population Statistics of Japan. (2003). *Complied by the national institute of population and social security research*. Tokyo: Japan.
- Preston, S. H., & van de Walle, E. (1978). Urban french mortality in the 19th century. *Population Studies*, 32, 275–297.
- Preston, S. H., Hill, M. E., & Drevenstedt, G. L. (1998). Childhood conditions that predict survival to advanced ages among African-Americans. *Social Science & Medicine*, 47(9), 1231–1246.
- Riley, J. (2001). *Rising life expectancy: A global history*. Cambridge: Cambridge University Press.
- Sköld, P. (1996). *The two faces of smallpox: A disease and its prevention in eighteenth and nineteenth-century Sweden*. Umeå: The Demographic Database, Umeå University.

- Tuljapurkar, S., Li, N., & Boe, C. (2000). A universal pattern of mortality decline in the G7 countries. *Nature*, 405(June), 789–792.
- United Nations. (1953). *The determinants and consequences of population trends* (Population Studies, No. 17). New York: UN.
- Wrigley, E. A., & Schofield, R. S. (1981). *The population history of England 1541–1871. A reconstruction*. London: Edward Arnold.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Part IV

Causes of Death

Tommy Bengtsson and Kaare Christensen

This part addresses the question of how information on changes in patterns in the cause of death can be used to improve mortality forecasting. While the increase in life expectancy was largely propelled by the decline in infant and child mortality up until the middle of the twentieth century, it has since then been sustained by the decline in old age mortality. The improvement in life expectancy among the elderly is mainly due to progress in combating chronic diseases. Mortality in cardiovascular and cerebrovascular diseases and some forms of cancer has declined. While improved medical care – earlier detection, improved surgery, and better therapies – is a major factor behind this, it is not the only one. Changes in working conditions, life styles and improvements early in life have contributed to the reduction of mortality in chronic diseases, as have other factors. In this volume, the focus is on the changes in the patterns of cause of death.

The chapter by Graziella Caselli, Jacques Vallin, and Marco Marsili discusses the usefulness of making extrapolations of past trends in major diseases. They discuss the problems related this method. In spite of clear drawbacks in using this information for extrapolation, they do not categorically reject it since it can provide a fairly realistic overview of what is behind trends and in doing so alert policy makers of possible effects if these trends continue. Other means of making use of causes of death information for forecasting are discussed as well, including making use of information from other countries.

Måns Rosén starts from an epidemiological perspective, discussing the relationships between incidence, prevalence and mortality. In addition to examining possibilities of extrapolating past trends in cause specific mortality, Rosén brings up the central discussion of whether prolongation of life leads to compression or

T. Bengtsson

Centre for Economic Demography, Lund University, Lund, Sweden

e-mail: tommy.bengtsson@ekh.lu.se

K. Christensen

Institute of Public Health, University of Southern Denmark, Odense, Denmark

prolongation of morbidity. While many studies in the past have concluded that compression is dominating, some recent Swedish studies indicate that this may not be the case. Thus it may well be that improvements in medical care and therapy will lead to an increase in the demand for health expenses but not necessarily health care in this group.

The third and final chapter, by Richard Willets, discusses how analyses of mortality by cause of death will influence forecasts in the UK. His conclusion is that, despite well-documented difficulties in making cause of death projections in the past, there still appears to be a good case for continuing to do so. This is particularly the case when predicting mortality among the not very old elderly, say those below 80 years of age. It can also be used to test "extreme" scenarios. Thus while using information on causes of death in making mortality forecasts has proven to be difficult, there still is a substantial potential to be gained.

Chapter 18

How Useful Are the Causes of Death When Extrapolating Mortality Trends. An Update



Graziella Caselli, Jacques Vallin, and Marco Marsili

Old age and adult mortality have over the last decades enjoyed a remarkable decline throughout the western world, posing the researcher with new challenges and opening up fresh horizons in life expectancy trends. The recent drop in mortality may be largely traced to the unexpected decline in cardiovascular diseases and certain cancers. Thus it could be hoped that in the future these trends would continue and extend to include other causes where, for the moment, little change has occurred. Such a hypothesis is all the more realistic in view of the fact that recent changes are linked, not just to advances in more efficacious medical treatment, but also to a growing awareness on the part of the general public regarding questions of health and the crucial role played by life style and behaviour. These include improved dietary habits, for example, a better attitude to risk factors, particularly to smoking, alcohol abuse, dangerous driving, etc. This awareness, which prevails among more recent, well-informed and better educated cohorts, not only produces immediate results, but maybe even more so in the future, should this spare coming generations the accumulation of risks which were and continue to be the burden particularly of older cohorts.

These considerations have increasingly encouraged researchers to refute the timid claims regarding future mortality generally made by Institutes of Statistics when

This paper is an update of Caselli and Vallin (1999a) (in French) and Caselli and Vallin (1999b) (in English).

G. Caselli (✉)

Department of Demography, University of Rome “La Sapienza”, Rome, Italy

e-mail: graziella.caselli@uniroma1.it

J. Vallin

INED, Paris, France

M. Marsili

ISTAT, Rome, Italy

producing population estimates (Vallin 1989, 1992; Vallin and Meslé 1989; Meslé 1993; Caselli 1993; van Poppel and de Beer 1996) and to seek to take better account of more recent progress when estimating future mortality trends. This has led to including causes of death as a component of mortality (Benjamin and Overton 1985; Caselli and Egidi 1992; Wilmoth 1996) and to seek methods to account for the cohort effect, and indeed to combine the two at times (Caselli 1996).

More complex data or more sophisticated methods are not themselves a guarantee for better results. Numerous experiences of this nature have ended up more as a disappointment than anything else. Our goal here is to focus on the advantages and disadvantages of taking causes of death into consideration when making mortality estimates and to explore the results of the different possible methods. It is beyond the scope of this paper to take a stand regarding the present debate on life expectancy outcomes or even to contribute to this. Rather, our task is to establish whether, by refining the methods, the results of a simple extrapolation of past trends could be improved, without making future hypotheses and irrespective of those directly stemming from an analysis of past trends.

The first obstacle one meets when projecting mortality trends cause by cause depends on the fact that even if there is one cause for which mortality increases, this will inevitably, sooner or later, depending on the relative importance of this cause, lead to a general increase in mortality for all causes, the overall perspective thus being more pessimistic than that yielded by extrapolating total mortality, as we will show below. In other words, it is almost not worthwhile considering mortality outlooks by cause if we are unable to “predict” the inflection points or the changes in the direction of the evolution curve. Therefore the question which must be posed is if by some means, when using the model of past trends, we can predict such changes in the trends.

To do so we will focus on the England & Wales male population and on mortality risks between 60 and 85 years. Opting for this population will help focus on mortality trends among the elderly, these being more sensitive to changes described above, and elude the thorny question of life expectancy thresholds, which to our mind calls for an entirely different approach.

When dealing with causes of death, for the sake of clarity, obviously only a limited number of groups of specific causes may be referred to, albeit with adequately diverse recent trends to be able to highlight the difficulties involved and evaluate the possible solutions. Five sufficiently descriptive causes were selected:

- cardiovascular diseases,
- bronchial and lung cancers,
- digestive cancers,
- other tumours,
- other diseases and violent deaths.

This classification is particularly suited to England & Wales as it includes a cause, bronchial and lung cancer, for which male mortality underwent a sharp rise followed by a decline, from the 1970s.

A reference period also had to be selected to elaborate a model of past trends. It was decided to focus alternatively on a long series, 1950–2000, which includes the period where mortality from bronchial and lung cancers was steep, as well as a shorter series (1981–2000), showing more recent trends.

The estimations made were obtained by extrapolating the logarithms of age specific mortality rates, which vary according to the number and types of variables considered to adapt the data sources.

Having, first of all, highlighted the absurdity of extrapolations based on a simple linear adjustment of a chronological series of age specific mortality rates (referred to here as the “*linear*” model), we will then try to obtain better results by gradually refining the modelling of the data series. Thus three increasingly complex models will be explored. First, while keeping to the approach where an independent adjustment is made for each chronological series of rates by age, an effort will be made to improve the outcome by selecting the best curve possible to adjust the data series (referred to here as the “*least squares*” method). Then, a model elaborated by Ronald Lee and Lawrence Carter, referred to here as “*Lee-Carter*”, will be used, where the logarithm of age specific mortality rates is a function of age as well as of period. Finally, thanks to a solution described elsewhere (Caselli 1993; Burgio and Frova 1995), a third component, that is the cohort effect, will be considered, using the “*APC*” model (age, period, and cohort).

However, to judge the comparable validity of these different approaches, extrapolations using older series must be compared with reality as it occurred. We will do this by using data from 1950–1980 to make projections for 1981–2000, which can then be compared with real mortality trends.

18.1 Extrapolation of Mortality by Cause Risks Absurdity

Figure 18.1 describes the results of a simple logarithmic extrapolation for mortality rates for all causes (the “*linear*” model), for each of the five age groups considered here (from 60–64 to 80–84 years), until the year 2050, based on data for 1950–2000, and shows a mortality projection which ignores individual trends for each cause of death. Average life expectancy between 60 and 85 years for an English male passes from 18.1 years in 2000 to 20.0 years in 2050, in other words a two-year gain.

Figure 18.2, on the other hand, illustrates the results of summing similar type extrapolations performed separately for each group of causes. A systematic increase in total mortality immediately occurs at older ages between 75 and 85 years, while the trend of reduction for ages between 60 and 75 years is less important than that obtained from the extrapolation for all causes, to such an extent that the average number of years one could expect to live between 60 and 85 years remains quite stable (around 18.0) over all the projection period (Table 18.5). Not only is this absent increase in survival at older ages hard to believe, it also appears somewhat absurd as sooner or later it yields mortality rates twice as high as the present for the highest ages. The problem, as we know, stems from the fact that causes of death are

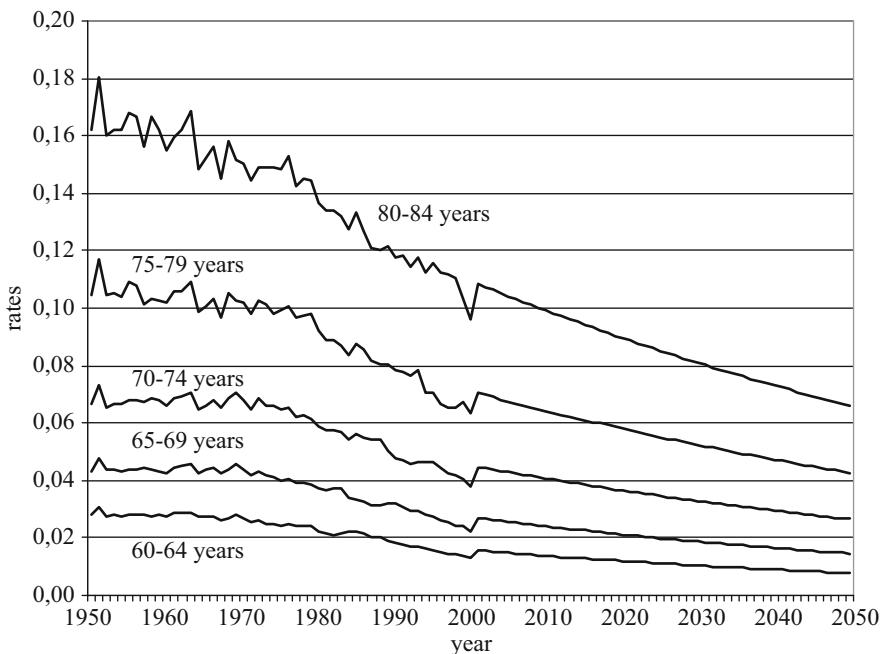


Fig. 18.1 Extrapolation of mortality rates for all causes by age group 2001–2050, based on a “linear” adjustment of data for 1950–2000 (England and Wales, males)

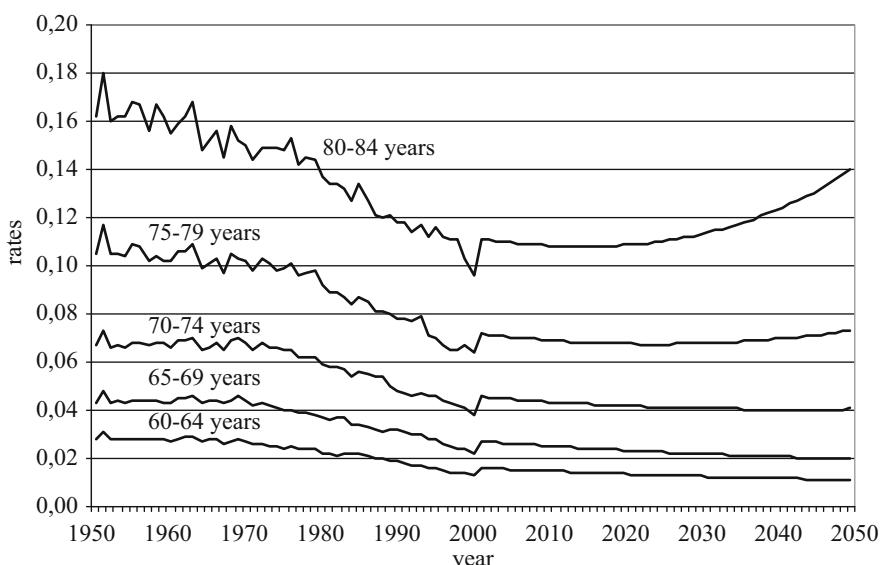


Fig. 18.2 Mortality trends by age group 2001–2050 obtained by summing specific rates by cause extrapolated using a “linear” adjustment of 1950–2000 data (England and Wales, males)

included where mortality trends were rising during a large part of the period of reference. This is the case with bronchial and lung cancers, as well as “other tumours”, where unfavourable trends are contrasted with favourable trends in cardiovascular diseases and digestive cancers (Fig. 18.3).

According to this outline, the impact of bronchial and lung cancers on total mortality comparative rates at 60–84 years would rise from 9.9% in 2000 to 30.9% in 2050, while that of cardiovascular diseases would fall from 45.7 to 23.4% (Table 18.1)!

No doubt this example is too extreme. Obviously, for England & Wales no one would dream of extrapolating bronchial and lung cancer mortality trends for

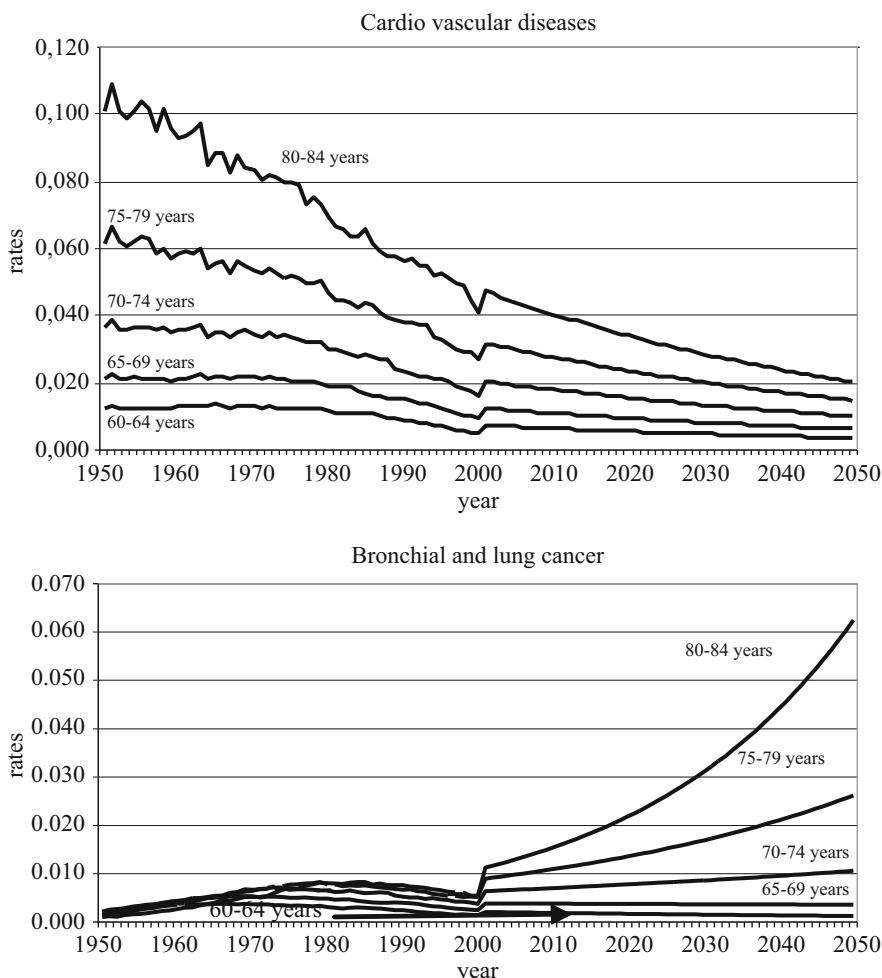
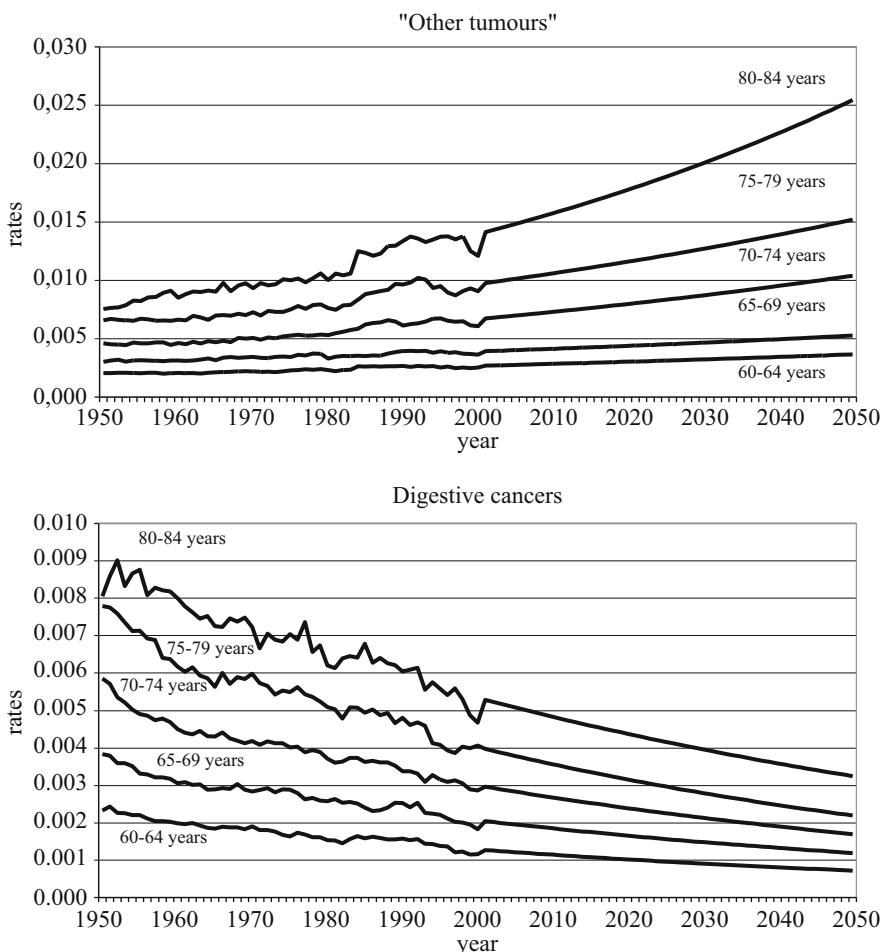


Fig. 18.3 Extrapolation of mortality rates by age group 2001–2050, for 4 groups of causes, the trends of which are in contrast, using a “linear” adjustment of 1950–2000 data (England and Wales, males)

**Fig. 18.3** (continued)**Table 18.1** Percent of each group of causes as part of the standardized mortality rates for all causes at 60–84 years, in 2000 and 2050, following an extrapolation using a “linear” adjustment of 1950–2000 rates, and then 1981–2000 rates (England & Wales, males)

Group of causes of death	2000	2050	
		Base 1950–2000	Base 1981–2000
Cardiovascular diseases	45.7	23.4	16.2
Bronchial and lung cancers	9.9	30.9	3.6
Digestive cancers	7.7	4.0	5.6
Other tumours	16.9	24.0	37.6
Other diseases and violent deaths	19.8	17.7	37.1
TOTAL	100.0	100.0	100.0

2001–2050 by a linear adjustment of the entire period 1950–2000, when in fact a reversal trend occurred in the early 1970's.

Thus fresh calculations were made, restricting the adjustment of past trends to the period 1981–2000. The results are visibly improved for bronchial and lung cancer, as this time mortality for this cause decidedly follows a downward trend for all age groups (Fig. 18.4). However, the problem is still not solved as mortality from “other tumours” increases for all ages. Therefore, in the final calculation the sum of the extrapolations by cause generate a reduction in overall mortality (Fig. 18.5). There is no doubt, given this scenario, that the total number of years lived between 60 and 84 years increases, rising from 18.1 years in 2000 to 20.2 years in 2050. However, this rise is less rapid than when total mortality is extrapolated (reaching 22.0 years), but continues to fall if the extrapolation continues beyond 2050 and does not top for the oldest old high levels, as in the previous instance.

18.2 Would More Sophisticated Methods Be Any Better?

Could we do any better with more sophisticated methods? The first attempt to be made, while keeping to the approach which adjusts only one dimension of mortality (chronology), is to choose, should it exist, an adjustment curve which is more appropriate than a simple straight line.

18.2.1 A Better Adjustment of Chronological Series of Rates by Age

Here a choice was made between four classic functions (straight line, parabole, hyperbole, logistic) which offered the least sum of the square distances to the observed values being selected. Thus, for bronchial and lung cancers, for example, the parabolic method was opted for, as this would effectively appear to prolong more satisfactorily observed trends in mortality by age (Fig. 18.6).

Again, the fact that mortality tends towards zero is obviously disputable. Unfortunately, for “other tumours”, the “least-squares” are obtained by the straight line method and we come up again against the same problem which arose previously, albeit not quite as quickly, where a major cause such as bronchial and lung cancers has been totally eliminated. Thus, we have merely delayed the march of time towards the unlimited increase in mortality for older ages (as in Table 18.2), but in 2050, the mortality profile by cause is much more deformed than in the previous instance, with cardiovascular diseases are no longer at the top of the list, falling from 46% in 2000 to 5.5% in 2050, while the impact of “other tumours” is remarkably increased, from 17% to 55%, keeping the lead, to such an extent that the role of bronchial and lung cancer is eliminated.

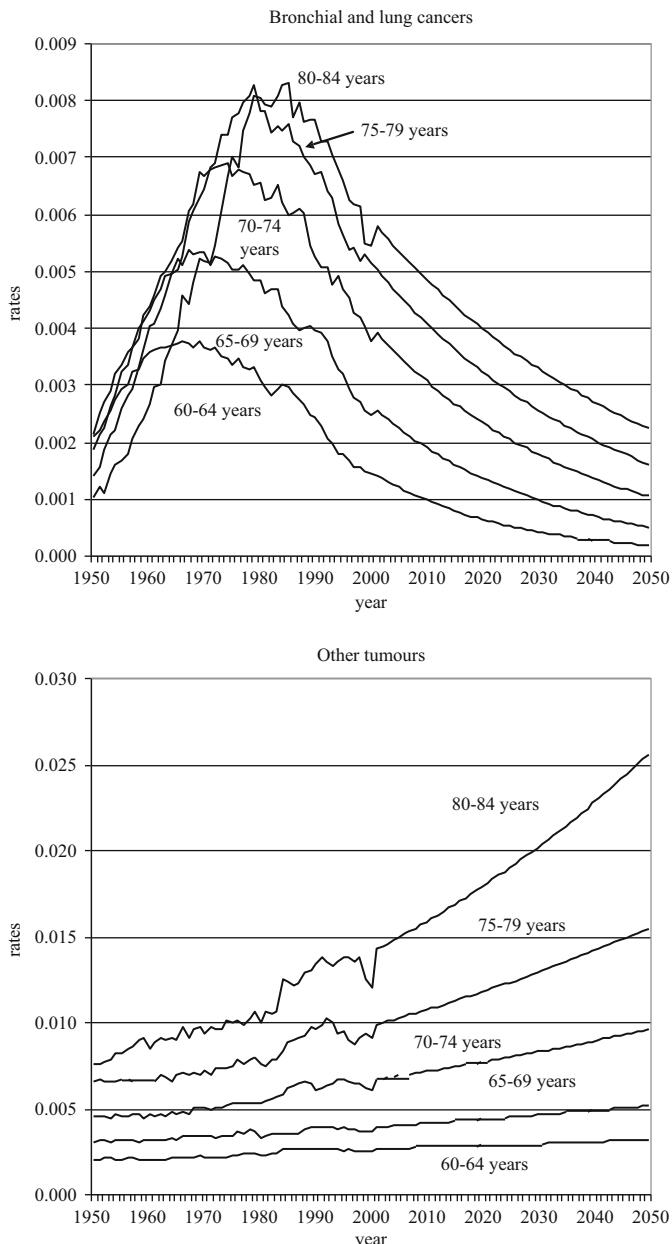


Fig. 18.4 Extrapolation of mortality rates by age group 2001–2050, for bronchial and lung cancers and for “other tumours”, using a “linear” adjustment of 1981–2000 data (England & Wales, males)

If the data observed are only adjusted for the most recent period (1981–2000), bronchial and lung cancers remain largely unchanged, but this tends to modify the changes foreseen for “other tumours” and thus delay the moment in which these raise

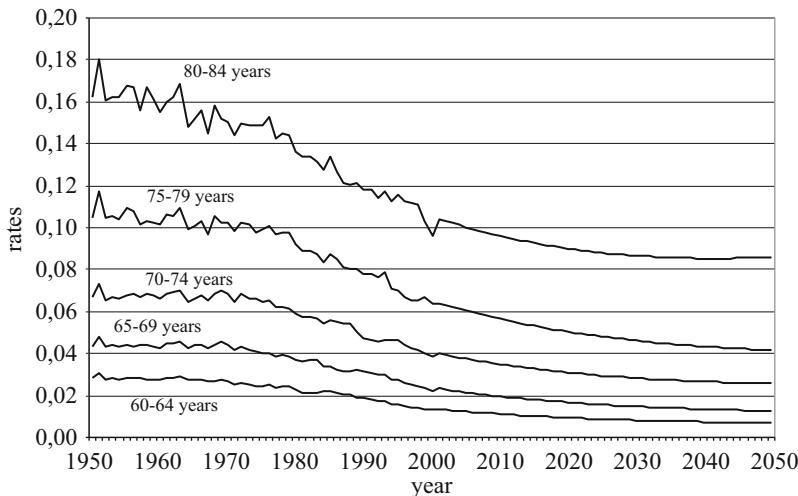


Fig. 18.5 Mortality trends by age group 2001–2050, obtained by summing rates by cause extrapolated using a “linear” adjustment of 1981–2000 data (England & Wales, males)

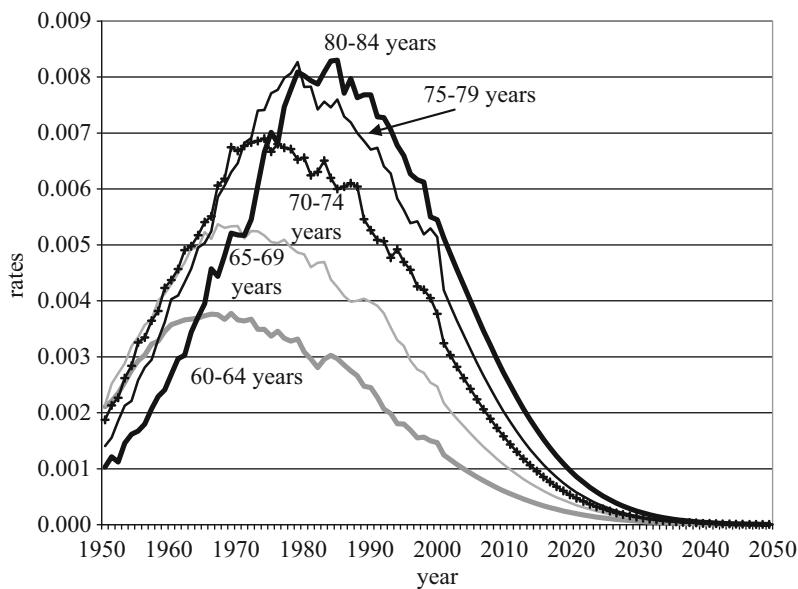


Fig. 18.6 Extrapolation of age group mortality rates for bronchial and lung cancers, using a “least-squares” adjustment of 1950–2000 data (England & Wales, males)

the sum of the total rates by cause. The mortality profile by cause for 2050 is thus considerably modified, with an increase to 16% for cardiovascular diseases and a decrease to 38% for “other tumours”.

Table 18.2 Percent of each group of cause as part of the standardized mortality rates for all causes at 60–84 years, in 2000 and in 2050, after a “least-squares” extrapolation of 1950–2000 rates and then 1981–2000 rates (England & Wales, males)

Group of causes of death	2000	2050	
		Base 1950–2000	Base 1981–2000
Cardiovascular diseases	45.7	5.5	15.9
Bronchial and lung cancers	9.9	0.0	3.5
Digestive cancers	7.7	6.9	5.6
Other tumours	16.9	55.3	37.6
Other diseases and violent deaths	19.8	32.3	37.4
TOTAL	100.0	100.0	100.0

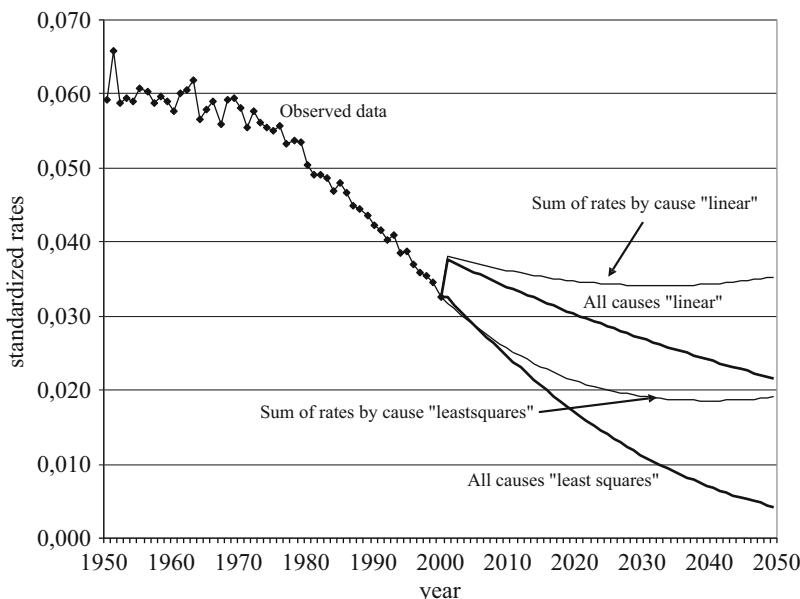


Fig. 18.7 Results compared, in terms of standardized mortality rates at 60–84 years, “linear” and “least-squares” models, for mortality for “all causes” and the “sum of rates by cause” based on observed data for 1950–2000 (England & Wales, males)

Figures 18.7, 18.8, and 18.9 report the different outcomes obtained to date regarding standardized mortality rates at 60–84 years, referring alternatively to the periods 1950–2000 and 1981–2000.

Based on data observed for 1950–2000, the improvement gained by using the “least-squares” method to adjust the curve, generates overall within the limits of the extrapolation period explored here, a trend in the sum of mortality rates by cause which is clearly less preposterous than that obtained with a strictly “linear” model even though still far removed from that yielded by the direct extrapolation of mortality for all causes. According to the sum of the extrapolations by cause, the

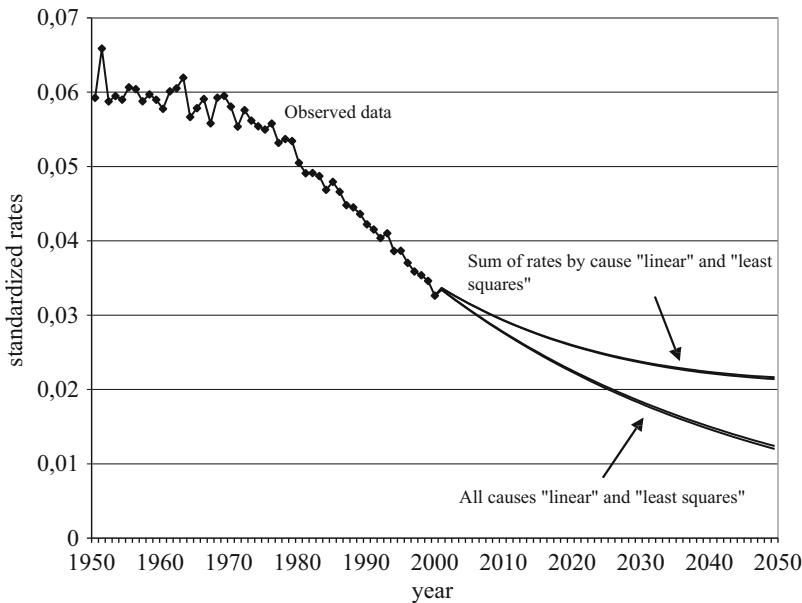


Fig. 18.8 Compared results, in terms of standardized mortality rates at 60–84 years, “linear” and “least-squares” models, for mortality for “all causes” and the “sum of rates by cause” based on observed data for 1981–2000 (England & Wales, males)

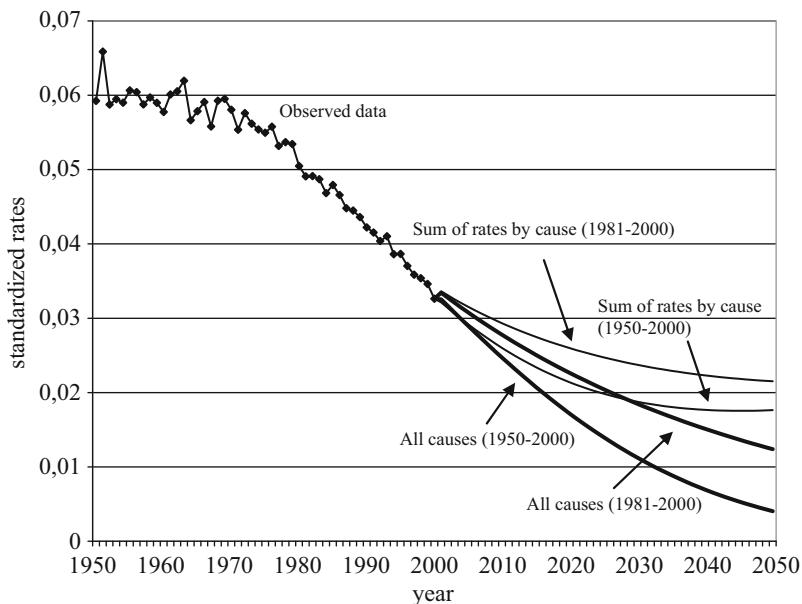


Fig. 18.9 Results, in terms of standardized mortality rates at 60–84 years, of the Lee-Carter model for mortality for all causes and the sum of the rates by cause, according to the reference period used (1950–2000 and 1981–2000) (England & Wales, males)

mean number of years lived between 60 and 85 years rises from 18.1 in 2000 to 20.6 for 2050, compared with 24.0 obtained with the direct extrapolation of mortality for all causes.

Nonetheless, it should be pointed out that the reference period used for the adjustment can notably change the end result. If this is limited to the most recent period, the role (favourable) played by trends in cardiovascular diseases is more quickly obliterated than that (unfavourable) played by “other tumours” (Fig. 18.8). Surprisingly, in 2050, by summing the extrapolations by cause the average number of years lived would be exactly the same as in the previous instance (20.2 years), and this time, too, it is lower than that obtained by a direct extrapolation of mortality for all causes (22.1).

One could, while maintaining the same approach, whereby a model is elaborated using a period component of age rates, attempt a further refinement, by choosing for each cause of death not only the best adjustment function but also the reference period which would best reflect recent trends. The limits of such an approach emerge fairly quickly, which risks being over-subjective and in any case fails to solve the problem of the impossibility of foreseeing an eventual reversal of the upward trends in “other tumours”.

18.2.2 “Age-Period” Adjustment (Lee-Carter Model)

In order to continue, more complex models are needed, which take into account other aspects of mortality, possibly able to anticipate trends already germinating in certain available data sources. First of all, using the model proposed by Ronald Lee and Lawrence Carter (1992), we will perform our extrapolations using a combination of past information on age and period. This stochastic model may be denoted by:

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

where, of course, $m_{x,t}$ is the mortality rate at age x at times t , a_x , b_x , and k_t , the model’s parameters, and $e_{x,t}$ the stochastic error, so that the average $E(e_{x,t})$ is equal to zero and the variance $V(e_{x,t})$ is constant. When the model is adjusted by the least-squares method, the interpretation of the parameters is very simple: the adjusted value of a_x is strictly equal to the average of $\ln(m_{x,t})$ for the period, so that b_x represents change in mortality age structure and k_t period trends. Regardless of whether the extrapolation is based on overall data observed between 1950 and 2000 or only on those for the most recent period (1981–2000), the outcomes obtained for each group of causes is little different from those obtained using the classic adjustment of the least squares: cause profiles in 2050 in Table 18.3 are more or less the same as those in Table 18.2.

As before, when looking back, two facts are quite remarkable. On the one hand the result obtained by directly extrapolating mortality for all 18.1 years lived between 60 and 85 years in 2000 rises to 22.0 years in 2050, when referring to the period 1981–2000, instead of only 23.9 when referring to the period 1950–2000.

Table 18.3 Percent of each group of causes as part of the standardized mortality rates for all causes at 60–84 years, in 2000 and in 2050, after extrapolating with the Lee-Carter method 1950–2000 rates and 1981–2000 rates (England and Wales, males)

Group of causes of death	2000	2050	
		Base 1950–2000	Base 1981–2000
Cardiovascular diseases	45.7	4.0	16.2
Bronchial and lung cancers	9.9	0.8	3.6
Digestive cancers	7.7	7.9	5.6
Other tumours	16.9	54.4	38.1
Other diseases and violent deaths	19.8	32.9	36.5
TOTAL	100.0	100.0	100.0

However, on the other hand, a cause by cause extrapolation considerably reduces over time these differences, until by summing the rates by cause extrapolated, in 2050 we obtain, respectively, 20.1 and 20.8 years lived, depending on the reference period considered. This occurs, as previously was the case, so that with this model as with the standard adjustment of the least squares, a marked increase may be foreseen in mortality for other tumours. Finally, this model, despite the fact that it is much more sophisticated, contributes little more than that offered by the standard adjustment of the least squares.

18.2.3 “Age-Period-Cohort” Adjustment (APC Model)

Are further refinements necessary when using an “APC” model based on the combined effects of age, period and cohort? APC models have been used chiefly to interpret past mortality trends (Osmond and Gardner 1982; Hobcraft et al. 1982; Osmond 1985; Caselli and Capocaccia 1989; Wilmoth et al. 1990). Their application in mortality forecasts is more recent (Caselli 1996) or limited to certain specific causes. Burgio and Frova (1995), based on the fact that, generally speaking, the mortality risk, m , may be expressed as a function $m = f(Z\Theta)$ of factors $Z = (z_1, \dots, z_n)$ and the parameters $\Theta = (\theta_1, \dots, \theta_k)$, hypothesised that the logarithms of the mortality rates could be adjusted using a polynomial function of age, period and cohort:

$$\ln(y_{t,x}^*) = a + a(x) + p(t) + c(t - x)$$

with:

$$\ln(y_{t,x}^*) = a + \sum_i b_i x^i + \sum_j c_j t^j + \sum_k d_k (t - x)^k,$$

for

Table 18.4 Percent of each group of causes as part of the standardized mortality rates for all causes at 60–84 years, in 2000 and 2050, after extrapolating with the “APC” model rates for 1950–2000 and for 1981–2000 (England & Wales, males)

Group of causes of death	2000	2050	
		Base 1950–2000	Base 1981–2000
Cardiovascular diseases	45.7	3.4	9.9
Bronchial and lung cancers	9.9	0.0	1.5
Digestive cancers	7.7	7.3	16.6
Other tumours	16.9	55.6	46.6
Other diseases and violent deaths	19.8	33.6	25.4
TOTAL	100.0	100.0	100.0

$$i = 1, \dots, h_1, j = 1, \dots, h_2 \text{ and } k = 1, \dots, h_3$$

In this function, $y_{t,x}^*$ denotes the theoretical value of mortality rates at age x during the year t (total or by cause) and $a, b_1, \dots, b_{h_1}, c_1, \dots, c_{h_2}, d_1, \dots, d_{h_3}$ are the parameters estimated by the least-squares method.

While this adequately describes past trends, it is not directly applicable to forecasts, to the extent that it does not pretend to prognosticate short-term fluctuations, translated by variations of the “period” parameter. For this reason the authors subdivided this parameter into two additive components, a basic movement, described by the straight line uniting the points relative to the first and last observations, and deviations in this trend. To perform the extrapolation they simply prolonged the basic movement, presuming deviations equal to zero in the basic trend.

The cause profile for 2050, for the reference period 1950–2000 (Table 18.4), is very similar to that obtained for the previous two attempts (Tables 18.2 and 18.3). What can be noted is a slightly larger impact of “other tumours” (55.6%) compared with a lesser impact of “cardiovascular diseases” (3.4%). On the other hand, results differ when, in the projection by cause, the more recent reference period 1981–2000 is taken. An important role is played by “other tumours” (47% as opposed to 38%), compared with a lesser impact of “other diseases” (25% as opposed 37%), while that of tumours of the digestive trace increases (17% compared with 6%). Nonetheless, regarding the number of years lived between 60 and 85 years (Table 18.5), the outcome of the APC approach for the years 1981–2000 is particularly interesting. Only with the APC model is the number of years lived according to the sum of the extrapolations by causes (23.2 years) close to that obtained with the direct extrapolation of mortality for all causes (24.0 years). It can be clearly seen that the APC model, which takes into account cohort effects, is better able to embrace the complexities of more recent trends.

Figure 18.10 compares trend estimates of the sum of standardized mortality rates by cause for each of the four models used here, applied alternatively to the two periods 1950–2000 and 1981–2000.

Table 18.5 Trends from today to 2050 in the number of years lived between 60 and 84 years, according to the model and the reference period used (England & Wales, males)

Reference period and model used	Observed values		Extrapolated values		
	2000	2015	2030	2050	
Reference period: 1950–2000					
<i>Linear model:</i>					
Total all causes	18.1	18.2	19.0	20.0	
Sum of specific rates by cause	18.1	17.7	18.0	18.1	
<i>Least-squares model</i>					
Total all causes	18.1	20.3	22.3	24.0	
Sum of specific rates by cause	18.1	19.8	20.5	20.6	
<i>Lee-Carter model</i>					
Total all causes	18.1	20.2	22.2	23.9	
Sum of specific rates by cause	18.1	19.6	20.5	20.8	
<i>APC model</i>					
Total all causes	18.1	20.3	22.3	24.1	
Sum of specific rates by cause	18.1	19.8	20.7	20.8	
Reference period: 1981–2000					
<i>Linear model:</i>					
Total all causes	18.1	19.5	20.7	22.0	
Sum of specific rates by cause	18.1	19.0	19.7	20.2	
<i>Least-squares model</i>					
Total all causes	18.1	19.5	20.8	22.1	
Sum of specific rates by cause	18.1	19.1	19.8	20.2	
<i>Lee-Carter model</i>					
Total all causes	18.1	19.5	20.7	22.0	
Sum of specific rates by cause	18.1	19.0	19.7	20.1	
<i>APC model</i>					
Total all causes	18.1	20.3	22.3	24.0	
Sum of specific rates by cause	18.1	20.1	21.8	23.2	

Compared with the results of the “linear” model applied to the entire period 1950–2000, this is largely unaware of a further acceleration in the 1980’s mortality decline among the elderly, particularly regarding cardiovascular diseases. This predicts almost constant mortality levels, if not a slight increase toward 2040, while all the other cases on the figure (comprising the “linear” model applied to the period 1981–2000) all appear to have grasped the drop in mortality for this cause although the intensity tends to vary. In other words, at this level of appreciation, choosing the right reference period is very important.

Nonetheless, if further refinement is sought, two aspects may be noted. Even when applied to the entire period 1950–2000, the results of the “APC”, “Lee-Carter” and “least-squares” models are not different from each other and the same as for the “linear” model when restricted to the most recent period and, thus, offer greater resistance should a poor choice be made regarding the reference period.

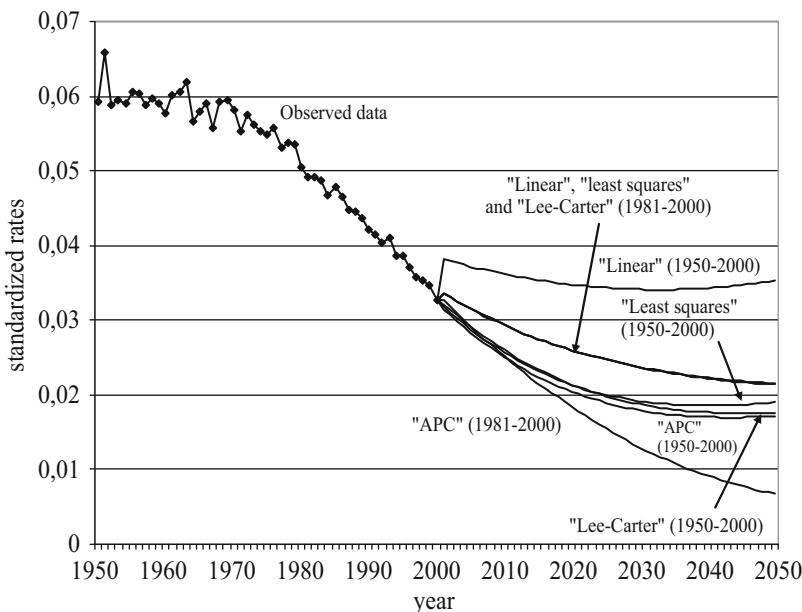


Fig. 18.10 A comparison, in terms of standardized mortality rates at 60–84 years, of the four approaches used (“linear”, “least-squares”, “Lee-Carter”, and “APC” models), of the sum of the rates by cause, according to the reference period used for the extrapolation (1950–2000 and 1981–2000) (England & Wales, males)

Finally, in each instance, whether for one reason or another, when attempting an extrapolation over the long term, undoubtedly it is advisable to use the most sophisticated model, the APC model, the only one to take into account the cohort effect and thus has the advantage of being able to detect the variety of changes which occur during the entire period. The divergence between the results obtained arises when accounting for recent or current reversal of certain tendencies. The actual performance of the different projections may be appreciated even more clearly if focus is given to how a specific cause has developed for which a fresh reversal has been recorded. This can be seen in Fig. 18.11, illustrating patterns for bronchial and lung cancers. Leaving aside the obvious absurdity of the application of the “linear” model to the entire period 1950–2000, it can be seen at which point this model is distinguished from the other three. When the reversed trend has been evident for 10 years or more, the results of all the projections are fairly similar. Of course what can be seen are the same nuances noted above for the sum of the rates by cause, but these are more attenuated. The trend, less pronounced in causes such as “other tumours”, is more decisive at this level.

However, coming back to our question: is it worth considering the cause of death? This exercise, which is purely a forecast, does not suffice to provide an answer. Nonetheless, two comments are worth making. If a long reference period is opted for (1950–2000), one blatant result is that, by taking into account the causes of death, the

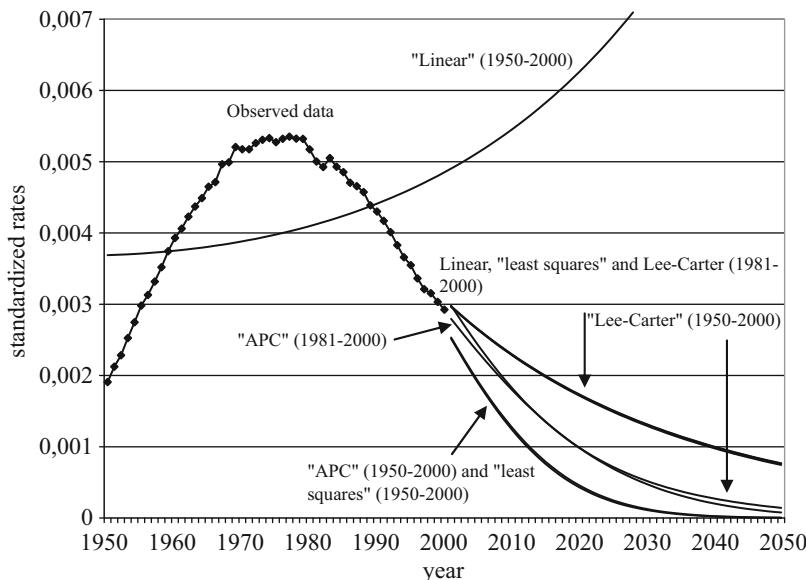


Fig. 18.11 A comparison of comparative mortality rates at 60–84 years, of the four approaches used (“linear”, “least-squares”, “Lee-Carter” and “APC” models), for mortality from bronchial and lung cancers, according to the reference period used for the extrapolation (1950–2000 and 1981–2000) (England & Wales, males)

results of the “linear” model are more pessimistic than others, with a “stagnation” in the number of years lived between 60 and 85 years around 18.1 (in 2050), compared with 20.0 years obtained by directly extrapolating rates for all causes (Table 18.5). With the other three approaches used only slight differences arise when the cause of death is considered, with the number of years lived between 60 and 85 years just topping 20.6–20.8 in 2050. It should be noted that for each of the three models, the sum of the extrapolated rates by cause is even less favourable than that obtained by directly extrapolating mortality for all causes (24 years instead of 20.6 and 20.8).

If the reference period is confined to the end of the observation period (1981–2000) the situation is reversed for the “linear” model which generates a sizeable increase in the number of years lived between 60 and 85 years (20.2 in 2050), but, again, this result is visibly lower than the result obtained by extrapolating mortality for all causes (22.0 in 2050). On the other hand, with the “least-squares” and “Lee-Carter” models the outcome of the projection by cause is not very different from that which is got using the longer period of reference and, for these models, too, the number of years lived is lower than that for all causes. Results for the more recent reference period regarding the application of the APC model are decidedly more interesting. As will be recalled, values for years lived in 2050 differ little among each other according to whether we consider the sums of rates extrapolated by cause or the extrapolation of mortality for all causes (23.2 compared with 24.0).

These results may be easily explained. In the first instance (the long reference period), major importance is given to the role played by reversed mortality from bronchial and lung cancers. This is quite well accounted for relatively speaking by the more sophisticated extrapolation by cause models, but not by the “linear” model, which by spreading the effects of the changing situation over the entire period, ignores the substantial decline in mortality for this cause. More importantly, it completely overlooks this decline among the oldest old that has only occurred quite recently (see Fig. 18.3). In the second instance (more recent, shorter period), where reversed mortality from bronchial and lung cancers is “recognized” by all of the models, differences mainly arise with regard to how they perceive the role played by “other tumours”, which neither the “least-squares” nor the “Lee-Carter” models were able to apprehend fully, while only the APC model managed to grasp these changes.

18.3 The Models Put to the Proof

While providing food for thought, a comparison of the different projections does not help us objectively in assessing how meaningful it is to take into account the causes of death nor the validity of the models used to do so. What it does show us are the differences among the results obtained and to suppose that this or that result is more or less plausible. To determine whether a quality leap has occurred one can estimate the model on an earlier period and compare the model’s projections with how reality has unfolded thereafter. This is our approach.

It turns out that for any extrapolation the period opted for is of paramount importance. We saw that if the period selected is too long, or too short, the risk is that the different trends underway will not be detected. Thus it was decided to apply the models to the period 1950–1980 and compare extrapolations for the period 1981–2000 to reality.

In this case it is clear that, regardless of the model used, apart from the APC model the extrapolation of mortality for all causes largely underestimated the drop in mortality (Fig. 18.12a). It is equally astonishing to see to what extent the results of the first three models are confounded: Absolutely nothing in from the trends in mortality rates by age for all causes in the 1960’s and 1980’s was captured by the refinements in these models. All yield a little less than 16 years to live between the ages of 60 and 85 years in 2000, instead of the 18.1 years observed (Table 18.6).

Considering the first three models, the picture is not better when causes of death are considered (Fig. 18.12b): despite differences in outcome among the models, none of them corresponded at all to the reality. Each of them underestimated the fall in mortality. This underestimation, as expected, is totally exaggerated in the “linear” model (just about 15 years to live between 60 and 85 years). Regarding “least squares” and “Lee-Carter” models, it is better to avoid working on a cause-by-cause approach and, thus, the projection was notably improved, although none of them succeeded in arriving at a realistic result (Table 18.6). Moreover, the APC

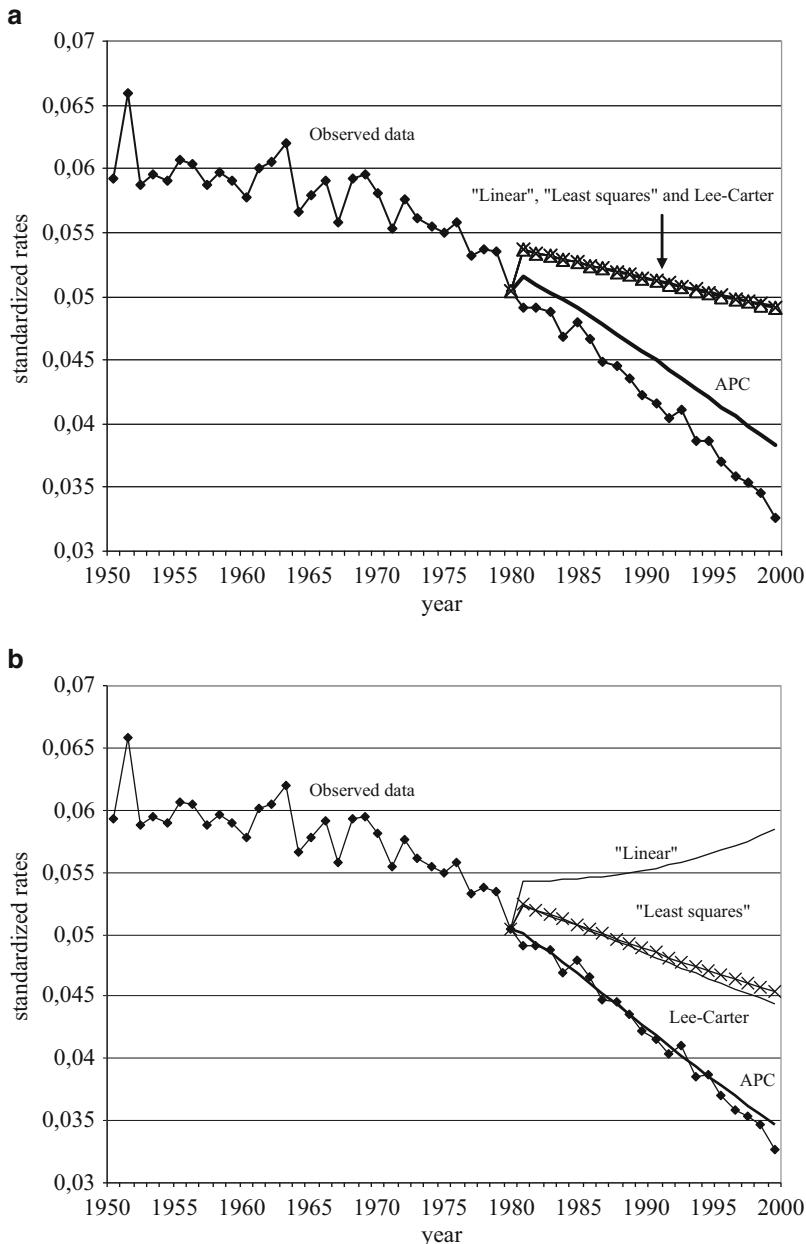


Fig. 18.12 Extrapolations for 1981–2000 of trends for 1950–1980 according to the four models, compared with real trends (England & Wales, males) **(a)** Direct extrapolation of mortality for all causes, **(b)** Sum of the extrapolations for mortality for all causes

Table 18.6 Number of years lived between 60 and 85 years in 2000: comparison between observed values and those obtained by extrapolating the data for 1950–1980, according to the four models (England & Wales, males)

Model	Observed values	Extrapolated values	
	2000	Total all causes	Sum of specific rates by cause
“Linear” model	18.1	15.7	14.9
“Least-squares” model	18.1	15.7	16.3
“Lee-Carter” model	18.1	15.8	16.2
“APC” model	18.1	17.2	17.6

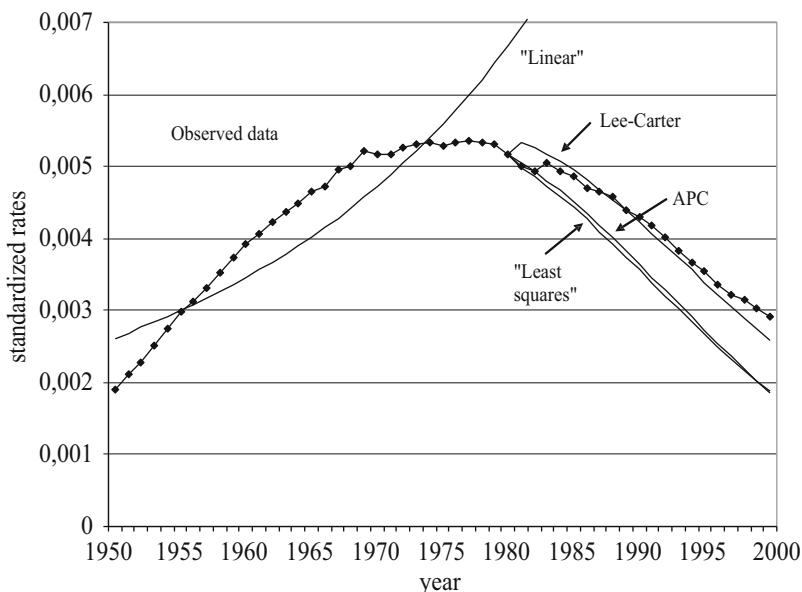


Fig. 18.13 Extrapolations for 1981–2000 of trends for 1950–1980 in bronchial and lung cancer mortality, according to the four models, compared with real trends (England & Wales, males)

model is the only one that approached reproducing reality. In particular when considering cause by death in the years 1981–2000 values often coincided with those observed (Fig. 18.12b), while for the year 2000 survival between 60 and 85 years differed by half a year.

However, when considering the results obtained by cause, it is clear that the APC model is not always better in capturing the renewed decline in mortality from bronchial and lung cancers (Fig. 18.13). The linear model naturally gave the most far-fetched results, extrapolating a preposterously high mortality rate, while on the other hand, the “Lee-Carter” projections best reflected the changing trends.

Even given this success the “Lee-Carter” model may not be conferred universal acclaim as of yet. Indeed, although the decline in bronchial and lung cancer mortality was the main reason for the rapid improvement in mortality trends in the 1970’s and

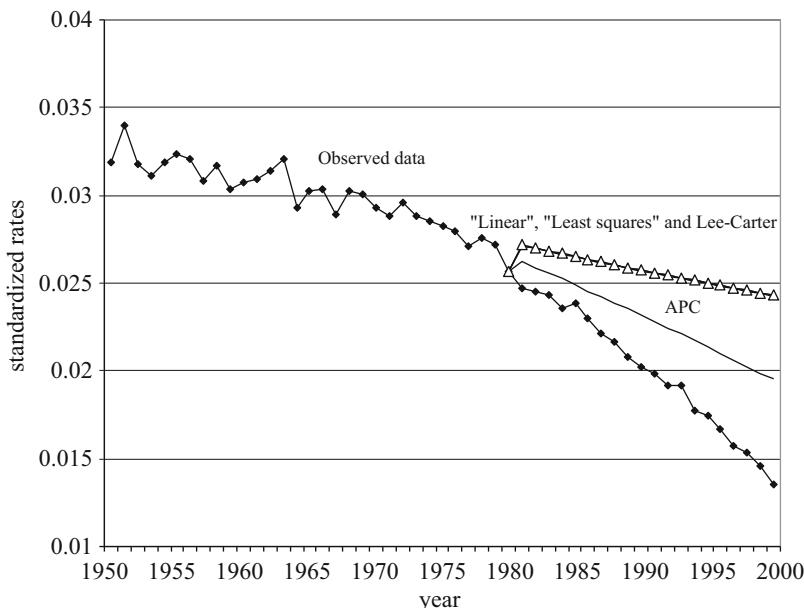


Fig. 18.14 Extrapolations for 1981–2000 of trends for 1950–1980 in mortality from cardiovascular diseases according to the four models, compared with real trends (England & Wales, males)

1980's, it was not the only reason. In fact, no single model, not even the APC model, is capable of fully apprehending this accelerated decline, because the "buds" of this even were not contained in any of the parameters of the models (Fig. 18.14). Otherwise what one finds for diseases of the cardio-vascular system is a perfect overlapping of the results of the first three models for the extrapolation of mortality for all causes (Fig. 18.12a).

In other words, there is no advantage in taking into account the causes of death to extrapolate mortality except in the case where future trends go strictly hand in hand with cohort phenomena, for example in the case of behaviour patterns with regard to smoking. In this case, the APC model performs best. No extrapolation model can foresee trends, the premises of which are not detectable in a reading of past trends.

18.4 Conclusion

Finally, if the aim is to foresee as realistically as possible mortality for all causes, by extrapolating past tendencies, we must make do with only extrapolating mortality rates for all causes. This is not to say that the idea of extrapolating mortality by cause is to be completely rejected. This can be useful from two points of view: to provide a fairly realistic overview of the consequences of cohort effects (in which case the APC model is out in front), as well as to alert policy makers on the effects to be

expected should past trends be prolonged over time (in which case the “linear” model suffices).

The extrapolation of past trends is not the only means of making forecasts. The future may also be fairly realistically based on observed data or that foreseen for elsewhere. Experiences of other countries may be used, where trends have already occurred similar to those one imagines will come to pass in the countries under focus. England was a precursor with regard to smoking habits and their experience may be used to anticipate reverse trends in bronchial and lung cancers, even if only based on current tobacco consumption. Moreover, the effects of recent policies may also be considered. A vaccination programme in a developing country may not be overlooked when estimating future mortality trends. One can, moreover take into account epidemiological facts which are already well-known, but whose effects on mortality are not yet evident. Perhaps even trends in the AIDS epidemic will help us estimate fairly precisely expected mortality over the next few years using only tendencies among the seropositive population. In each of these instances, working with a cause-by-cause model is to be favoured.

To make models, extrapolate trends, is all very well. However, the most complex method is not necessarily the best. The truth may be summed up by two sayings: The only good tools are those which are fashioned to suit the purpose and it is better to dream with your eyes open than make models with your eyes closed.

Acknowledgement The authors thank Dr. Maura Simone to her active contribution to the data processing required by this study.

References

- Benjamin, B., & Overton, E. (1985). Prospects for mortality decline in England and Wales. *Population Trends*, 42, 22–29.
- Burgio, A., & Frova, L. (1995). Projections de mortalité par cause de décès: extrapolation tendancielle ou modèle âge-période-cohorte. *Population*, 50(4/5), 1031–1052.
- Caselli, G. (1993). L'évolution à long terme de la mortalité en Europe. In A. Blum & J.-L. Rallu (Eds.), *European Population, II. Demographic Dynamics/Démographie européenne, II. Dynamiques démographiques* (pp. 111–164). Paris: INED/John Libbey. (Also published in English: OPCS: Studies on medical and population subjects, n° 56, London, 1994).
- Caselli, G. (1996). Future longevity among the elderly. In G. Caselli & A. Lopez (Eds.), *Health and mortality among elderly populations* (pp. 235–265). Oxford: Clarendon Press.
- Caselli, G., & Capocaccia, R. (1989). Age, period, cohort, and early mortality: An analysis of adult mortality in Italy. *Population Studies*, 43, 133–153.
- Caselli, G., & Egidi, V. (1992). New frontiers in survival: The length and quality of life. In *Human resources in Europe at the dawn of the 21st century/Le capital humain européen à l'aube du 21e siècle* (pp. 95–120). Luxembourg: Eurostat. (Proceedings of an International Conference, Luxembourg, 27–29 novembre 1991).
- Caselli, G., & Vallin, J. (1999a). Est-il utile de prendre en compte les causes de décès pour extrapolier les tendances de la mortalité? In C. Wattelar & J. Duchêne (Eds.), *Le défi de l'incertitude. Nouvelles approches en perspectives et prospectives démographiques*

- (pp. 85–110). Louvain-la-Neuve: Institut de démographie de l'UCL, Academia-Bruylant et L'Harmattan, Chaire Quételet.
- Caselli, G., & Vallin, J. (1999b). How useful are the causes of death when extrapolating mortality trends? In V. Egidi (Ed.), *Towards an integrated system of indicators to assess the health status of the population* (pp. 121–150). Rome: ISTAT.
- Hobcraft, J., Menken, J., & Preston, S. (1982). Age, period, and cohort effects in demography: A review. *Population Index*, 48(1), 4–43.
- Lee, R. D., & Carter, L. (1992). Modelling and forecasting US mortality. *Journal of the American Statistical Association*, 87(419), 659–675.
- Meslé, F. (1993). Analyse comparée des projections nationales en Europe. In J. Vallin (Ed.), *L'avenir de l'espérance de vie* (pp. 5–22). Paris: INED. (Congrès et Colloques n° 12).
- Osmond, C. (1985). Using age, period, and cohort models to estimate future mortality rates. *International Journal of Epidemiology*, 14(1), 124–129.
- Osmond, C., & Gardner, M. (1982). Age, period, and cohort models applied to cancer mortality rates. *Statistics in Medicine*, 1, 245–259.
- Vallin, J. (1989). L'avenir de l'espérance de vie vu à travers les projections de l'INSEE. *Population*, 44(4/5), 930–936.
- Vallin, J. (1992). Quelles hypothèses peut-on faire sur l'évolution future de la mortalité ? In G. Tapinos (Ed.), *La France dans deux générations: population et société dans le premiers tiers du XXIe siècle* (pp. 98–114). Paris: Arthème Fayard.
- Vallin, J., & Meslé, F. (1989). À long terme, l'écart d'espérance de vie entre hommes et femmes devrait diminuer. *Population*, 44(6), 1244–1251.
- van Poppel, F., & de Beer, J. (1996). Evolution of standard mortality projections for the elderly. In G. Caselli & A. Lopez (Eds.), *Health and mortality among elderly populations* (pp. 288–312). Oxford: Clarendon Press.
- Wilmoth, J. (1996). Mortality projection for Japan: A comparison of four methods. In G. Caselli & A. Lopez (Eds.), *Health and mortality among elderly populations* (pp. 266–287). Oxford: Clarendon Press.
- Wilmoth, J., Vallin, J., & Caselli, G. (1990). When does a cohort's mortality differ from what we might expect ? *Population, English selection*, 2, 93–126. (Also published in French: "Quand certaines générations ont une mortalité différente de celle que l'on pourrait attendre." *Population*, 44(2), 335–376).

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 19

Forecasting Life Expectancy and Mortality in Sweden – Some Comments on Methodological Problems and Potential Approaches



Måns Rosén

19.1 Introduction

Since mortality is affected by innumerable factors in society, an inter-disciplinary approach seems most appropriate. My contribution and point of departure starts from an epidemiological perspective and from the overall objective of the Swedish Centre for Epidemiology, i.e. to monitor public health in Sweden.¹ An advantage of epidemiology is the close link to public health and medicine as well as its focus on analyses of risk factors and the search for causal chains between risk factors, diseases and mortality. Mortality forecasting is a well-established discipline in demography, but maybe less developed within epidemiology. Still, there have been attempts to forecast mortality within the field of epidemiology (see e.g. Wilhelmsen et al. 2004; Gunning-Schepers 1989; Gunning-Schepers et al. 1989; Kruijshaar et al. 2002; Conroy et al. 2003). Usually, epidemiologists have focused on estimating mortality for specific causes of death (Wilhelmsen et al. 2004; Conroy et al. 2003) but there are also attempts to predict total mortality (Gunning-Schepers 1989; Kruijshaar et al. 2002). A common application has been to predict coronary heart mortality based on data on risk factors, e.g. smoking, level of cholesterol and blood pressure in the population (Wilhelmsen et al. 2004; Conroy et al. 2003). Knowledge of risk factor patterns is therefore an essential element in epidemiology. The risk factor approach will be discussed later. First, some comments on the outline of this paper.

¹National Board of Health and Welfare (2003).

M. Rosén (✉)

Centre for Epidemiology, National Board of Health and Welfare, Stockholm, Sweden
e-mail: mans.rosen@comhem.se

I will start by introducing the two most widely used measures in epidemiology, i.e. incidence and prevalence, and the relationships between those two measures and mortality. Second, I will give some general comments on pros and cons with different options for mortality forecasting. The different options discussed are extrapolating mortality trends, predicting disease-specific causes of death, predicting mortality trends based on potential elimination of causes of death or predicting mortality based on risk factors or other developments in the community. Third, some methodological problems will be discussed. Finally, I will advocate a risk factor based approach and speculate about future mortality and longevity based on our attempts to monitor public health in Sweden.

19.2 The Relationships Between Incidence, Prevalence and Mortality

Incidence is defined as the number of new cases of a disease during a specified time period while prevalence is the total number of people with a disease at a specific point in time. It is well illustrated by a bath tube (Fig. 19.1) where the water coming through the tap is the incidence and the water in the bath tube is the prevalence. The prevalence is affected by the incidence, but also by the number of people cured or deceased.

Those who die will no longer belong to the population at risk while those who are cured still belong to the population at risk. The cured survivors will have a probability of contracting a new disease. Primary prevention may influence incidence while prevalence is more of a measure of the total disease burden for society.

Mortality is affected by the incidence, which could be divided into different components from demographic characteristics (number of people, population growth and changing age distribution) to the risk of having/attracting a disease (based on risk factor patterns). Mortality is also influenced by the chance of surviving a disease. All these components are important to consider in mortality forecasting.

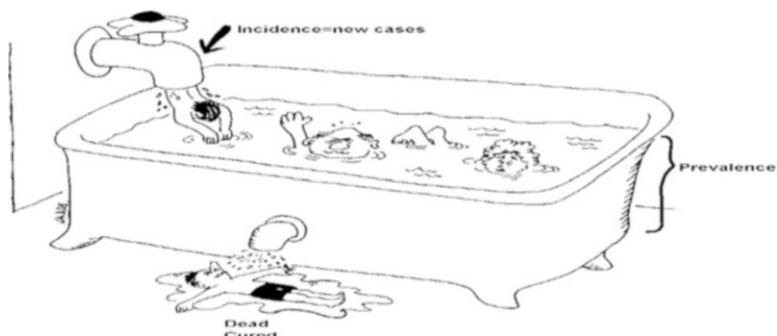


Fig. 19.1 The relationships between incidence, prevalence, mortality and cured

19.3 Extrapolating Mortality Trends or Predicting Disease-Specific Causes of Death

Life expectancy has increased impressively during the past 150 years. In Sweden, life expectancy for a man has improved from 35 years in the beginning of the 1800s to 61 years in the 1920s and up to 77.91 years in 2003. For women, life expectancy is 82.43 years in 2003. This success story seems to be never ending. In the 1980s, many believed there was little potential for improvement, but they were wrong. Still, it seems unlikely and atheoretical to believe this can persist forever.

History can also show us the danger of only extrapolating existing trends. A recent and dramatic story is the development of mortality and life expectancy in Russia.² Between 1970 and 1985 life expectancy in Russia was quite stable around 68 years (WHO). Between 1985 and 1987 it rose to 70 years followed by a substantial drop to about 64 years. Several studies have analysed the reasons to this dramatic and rapid change. The main explanations suggested are economic and social instability as well as changes in alcohol consumption (Shkolnikov et al. 2001; Notzon et al. 1998; Nemtsov 2002). The anti-alcohol campaign, launched in 1985, and the market reforms launched in 1992 were associated with large and rapid changes of alcohol consumption in Russia (Nemtsov 2002).

Trends in life expectancy among women in Denmark and the Netherlands can serve as other examples of the danger of only extrapolating trends. Since 1970, there is a steady increase in life expectancy for men both in Denmark and the Netherlands (WHO). However, extrapolating trends from the early 1970s would highly overestimate the longevity of women in Denmark and the Netherlands. Danish women increased their life expectancy substantially from about 76 years in 1970 to 78 years in 1977 followed by no increase at all up to 1995. After 1995 life expectancy among women in Denmark has started to increase again. For women in the Netherlands, life expectancy increased substantially up to 1990, but has thereafter not followed the increasing trends of many other western European countries. These changes in trends indicate clearly that the risk factor patterns of women in these two countries have been different than in other European countries.

The danger of extrapolating mortality trends is also evident when studying some disease-specific causes of death in Sweden. Lung cancer mortality among men increased substantially from the 1950s up to around the end of the 1970s followed by a decrease in both incidence and mortality (Fig. 19.2).

This trend break could easily have been anticipated if declining smoking rates had been considered. Smoking rates among men started to decline in the early 1960s accompanied with a trend break for lung cancer about 20 years later. Smoking rates among women have increased up to the late 1970s followed by a small decrease in smoking rates. So far, no shift in lung cancer rates among women can be seen. However, lung cancer rates among women seem to be levelling off.

²WHO Europe Health for All database, see <http://www.who.dk>; Shkolnikov et al. (2001); Notzon et al. (1998); Nemtsov (2002).

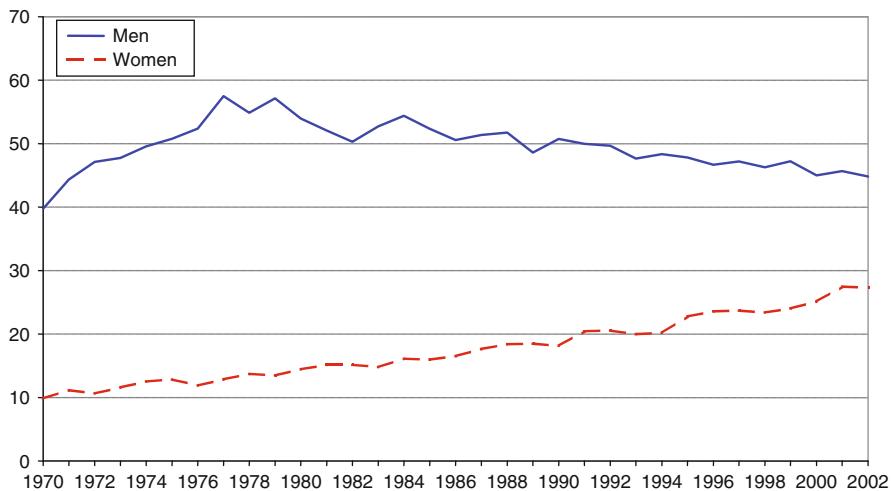


Fig. 19.2 Trends in lung cancer mortality in Sweden, 1970–2002

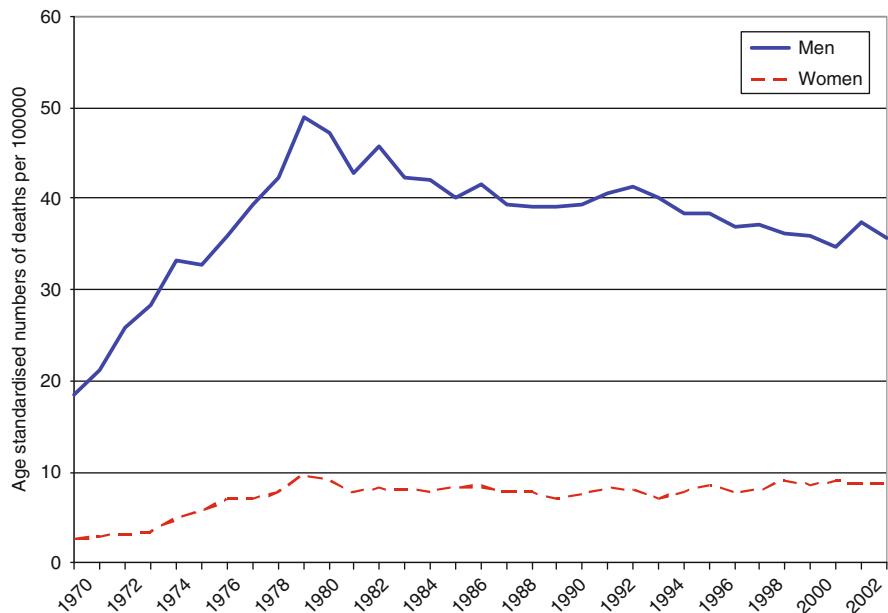


Fig. 19.3 Alcohol-related mortality in Sweden 1970–2002

Alcohol-related mortality rose dramatically after the abolition of the Swedish rationing system in 1955 and it was first around 1980 a decreasing alcohol mortality trend was noticed (Fig. 19.3). This trend break was probably due to intensified efforts in society as a whole. Cohorts born in the 1960s and 1970s also seem to be very healthy cohorts with low smoking rates and moderate alcohol consumption.

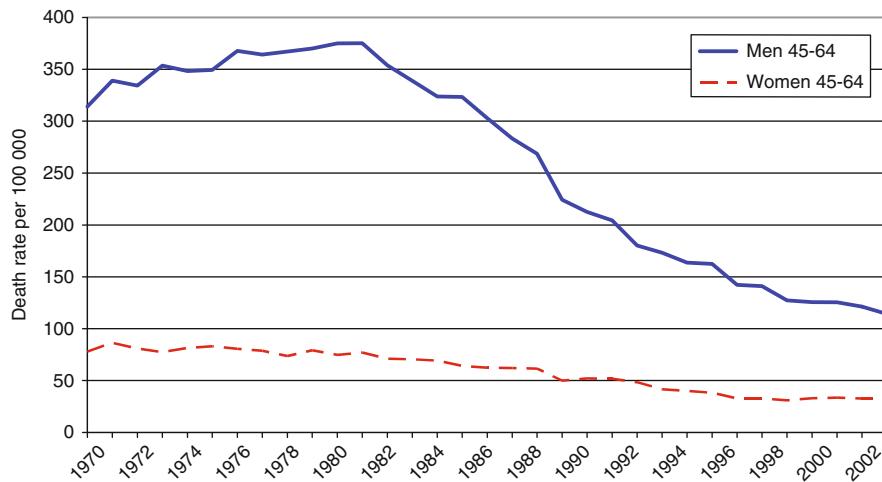


Fig. 19.4 Ischaemic heart disease mortality in Sweden 1970–2002, 45–64 years

The development of acute myocardial infarction and other coronary heart diseases among middle-aged men is another example of a trend break (Fig. 19.4). This trend break took place in the beginning of the 1980s and was due to several changes in risk factors, especially the decline in smoking rates among men. The level of serum cholesterol has also decreased in the Swedish population contributing to a decreasing trend in coronary heart mortality. All these examples clearly indicate caution in respect to merely extrapolating mortality trends.

19.4 Predicting Mortality Based on Potential Elimination of Causes of Death

To gain an idea of how great the potential is for increasing life expectancy one can do hypothetical calculations of how much it would increase if a disease no longer led to death (Curtin and Armstrong 1988; Haglund and Rosén 2001). In the Swedish Public Health Report of 2001 such calculations have been made (Haglund and Rosén 2001). The results are summarised in Table 19.1, which shows that the elimination of cardiovascular disease as a cause of death is the single most important step to prolong life expectancy followed by cancer. For cardiovascular disease more than 5 years could be gained for men by eliminating this disease group. Many may be surprised by the small gains obtained by eliminating traffic accidents (3 months for men and 1 month for women) or infectious diseases (1 month).

Social factors play an important role in the etiology of diseases and for mortality predictions. Upper white-collar workers have the lowest mortality. If the death risk for the whole population between 25 and 74 were reduced to the same level as for upper white-collar workers, men's life expectancy would have been 2 years and

Table 19.1 Changes in life expectancy in months through elimination of certain diseases or risk factors in Sweden 1993–1997

Disease/risk factor	Men	Women
Infections	1	1
Cancer	34	37
Cardiovascular diseases	64	48
Injuries and accidents	12	6
Traffic accidents	3	1
Tobacco	16	8
Social inequalities	29	17

Source: Haglund and Rosén (2001)

5 months longer and women's life expectancy 1 year and 5 months longer (Haglund and Rosén 2001).

19.5 Predicting Mortality Based on Development of Risk Factors

Predicting mortality based on social developments and predictions on risk factor changes seems most appropriate since these are the driving forces for mortality. The major problems are the lack of knowledge we have concerning all risk factors affecting all diseases. The three most important risk factors for coronary heart disease (CHD) are smoking, hypertension and high blood cholesterol levels. However, 247 risk factors for CHD have been suggested in the scientific literature (Hopkins and Williams 1981). It is impossible to make predictions for all these and many of them are not very well evidence based. Still, the three major risk factors explain quite a large proportion of CHD deaths and it is therefore much easier to predict the future CHD trends than to predict mortality for other causes of death, e.g. cancer where the knowledge base is more limited. Since about half of all deaths are caused by cardiovascular disease, it seems meaningful to make mortality predictions based on the risk factor development of this disease group.

19.6 Methodological Problems in Predicting Mortality Based on Risk Factor Predictions

In this paper I advocate a risk factor prediction approach to mortality forecasting. I hope the earlier presentation convincingly has shown the advantages of this approach in comparison with extrapolating mortality trends.

However, several methodological problems still exist. Four problems could be highlighted. Relative risks vary over time and by regions, latency times differ, co-morbidity and competing causes of death complicate the predictions and the lack of appropriate risk factor data limit the possibilities.

In the case of coronary heart disease, longitudinal studies from different parts of the world have displayed the same major and independent risk factors, but with varying relative risks (Wilhelmsen et al. 2004; Conroy et al. 2003; Empana et al. 2003). The Framingham risk functions based on U.S. populations overestimate the absolute coronary heart disease risk of middle-aged men when they are applied to different European populations (Empana et al. 2003). A problem in estimating mortality trends is the long latency times between exposure to risk factors and when the individuals are struck by the disease. For smoking and lung cancer latency time is usually more than 20 years of smoking. These kinds of considerations must be taken into account when making mortality predictions. However, the greatest problem in mortality modelling is usually lack of reliable risk factor data. Our own experiences of testing the Dutch mortality model (Gunning-Schepers 1989) on Swedish data showed the lack of risk factor data even in a data affluent society like Sweden.

19.7 Future Mortality and Longevity

As a simple exercise, Rosén and Haglund (2002) estimate future life expectancy in Sweden, not based on sophisticated dynamic population models, but merely on assumptions about risk factor developments and general knowledge about public health, recent successes in health care and the potential of eliminating certain causes of death (Table 19.2).

Social differences in mortality are large even in economically well-developed countries like Sweden. The reasons for these differences are multi-factorial and are most likely due to an accumulation of health risks during the whole life-cycle. Lower socio-economic groups have usually lower birth weights, have been brought up in more disadvantaged areas, have less education, smoke more, eat more unhealthy products, have more often monotonous work or are more often unemployed. However, history has shown that lower socio-economic groups will eventually reach the life expectancy of higher socio-economic groups, but that they are always 10–20 years behind. Eliminating the present social differences in health seems therefore a realistic scenario.

Table 19.2 Future estimated life expectancy (years) in Sweden

Intervention	Men	Women
Life expectancy, year 2000	77.4	82.0
Eliminating social differences	+2.5	+1.5
Improved lifestyle	+1.5	+1.5
Improved medical care	+1.5	+0.5
Optimist supplement	+1.0	+1.0
Estimated life expectancy	84.0	87.0

Source: Rosén and Haglund (2002)

It is also obvious that eliminating cardiovascular disease has the greatest impact on longevity. This is an area where we have evidence based knowledge of risk factors and great potential for primary prevention. Since the 1990s medical technologies have had a success story in developing life-saving interventions in the field of coronary heart disease. Altogether, this implies a high potential for improving longevity by reducing mortality for cardiovascular disease. We estimated a gain of 1.5 years due to improved lifestyle, mainly reduced smoking rates, and further gains due to improved medical technologies of 1.5 years for men and 0.5 years for women. The larger estimated gain for men is due to the fact that medical interventions will influence cardiovascular disease most, which is a larger burden for men than women. Finally, we added an optimist supplement of 1 year for improvements not foreseen by our estimates.

19.8 Implications for the Future

Mortality forecasting plays an important role for development and maintenance of national and private insurance schemes. However, there are also other social and economic consequences of changing mortality trends. A lively discussion has been whether prolonging lives may lead to compression or expansion of morbidity (Thorslund et al. 2004). Many studies in the past have indicated decreasing morbidity and improved functional status among the elderly, i.e. supporting the hypothesis of compression of morbidity. Recent studies in Sweden show, however, deteriorating health in some aspects among the elderly (Thorslund et al. 2004; Rosén and Haglund 2005). This development supports the hypothesis that we now are going from healthy survivors to sick survivors due to improvement in health care (Rosén and Haglund 2005). Since the late 1980s, new and very effective life-saving drugs and treatments have been developed, especially in the field of cardiovascular disease. This has had tremendous effect on survival among patients with acute myocardial infarction, heart failure and diabetes. Those surviving will, however, live with their chronic diseases and demand more care than earlier “healthy” survivors.

References

- Conroy, R. M., Pyörälä, K., Fitzgerald, A. P., Sans, S., Menotti, G., De Bacquer, D., et al. (2003). Estimation of ten-year risk of fatal cardiovascular disease in Europe: The SCORE project. *European Heart Journal*, 24, 987–1003.
- Curtin L. R., & Armstrong, R. J. (1988). “United States life tables eliminating certain cases of death. U.S. decennial life tables for 1979–81.” Vol 1, No 2. *DHHS Pub. No (PHS) 88-1150-2*, Public Health Service. Washington, DC: U.S. Government Printing Office.
- Empana, J. P., Ducimetiere, P., Arveiler, D., et al. (2003). Are the Framingham and PROCAM coronary heart disease risk functions applicable to different European populations? The PRIME study. *European Heart Journal*, 24, 1903–1911.

- Gunning-Schepers, L. (1989). The health benefits of prevention: Simulation approach. *Health Policy*, 12, 1–255.
- Gunning-Schepers, L., Barendregt, J. J., & Van Der Maas, P. J. (1989). Population interventions reassessed. *Lancet*, 1, 479–481.
- Haglund, B., & Rosén, M. (2001). Public health in the future. *Scandinavian Journal of Public Health*, 58, 231–239.
- Hopkins, P. N., & Williams, R. R. (1981). A survey of 246 suggested risk factors. *Atherosclerosis*, 40, 1–52.
- Kruijshaar, M. E., Barendregt, J. J., & Hoeymans, N. (2002). The use of models in the estimation of disease epidemiology. *Bulletin World Health Organisation*, 80, 622–628.
- National Board of Health and Welfare. (2003). *A finger on the pulse. Monitoring public health and social conditions in Sweden 1992–2002*. Stockholm: Centre for Epidemiology, National Board of Health and Welfare.
- Nemtsov, A. V. (2002). Alcohol-related human losses in Russia in the 1980s and 1990s. *Addiction*, 97, 1413–1425.
- Notzon, F. C., Komarov, Y. M., Ermakov, S. P., Sempos, C. T., Marks, J. S., & Sempos, E. V. (1998). Causes of declining life expectancy in Russia. *JAMA*, 279, 793–800.
- Rosén, M., & Haglund, B. (2002). Hur länge kommer vi att leva? *Framtider*, 1, 16–19.
- Rosén, M., & Haglund, B. (2005). From healthy survivors to sick survivors – Implications for the twenty-first century. *Scandinavian Journal of Public Health*, 33(2), 151–155.
- Shkolnikov, V., McKee, M., & Leon, D. A. (2001). Changes in life expectancy in Russia in the mid-1990s. *Lancet*, 357, 917–921.
- Thorslund, M., Lennartsson, C., Parker, M. G., & Lundberg, O. (2004). De allra äldstas hälsa har blivit sämre. Könnskillnaderna är stora – kvinnorna mår sämre visar nya data. *Läkartidningen*, 101, 1494–1499.
- WHO Europe Health-for-All-Database., see <http://www.who.dk>
- Wilhelmsen, L., Lappas, G., & Rosengren, A. (2004). Risk of coronary events by baseline factors during 28 years follow-up and three periods in a random population sample of men. *Journal of Internal Medicine*, 256, 298–307.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 20

How Analysis of Mortality by Cause of Death Is Currently Influencing UK Forecasts



Richard Willets

The purpose of this paper is to examine the potential benefits of cause of death analysis in the context of projecting future mortality rates in the UK.

In the first section of the paper the main features of recent mortality trends in the UK are briefly described. In the second methods currently used to project mortality in the UK are outlined. Current issues and topics for research are also discussed. In section three potential causes of the “UK cohort effect” are listed and the role of cigarette smoking, in particular, is discussed. A model of mortality which includes a year of birth component is discussed in section four. It is argued that models such as this can be used to analyse mortality from different causes and this analysis can have important benefits. Conclusions and implications are given in section five.

Throughout the paper most emphasis is placed on understanding and modelling mortality trends for older adults. This part of the age range is currently the focus of most research in the UK and has the greatest financial significance in terms of its impact on pension schemes and public finances.

20.1 Mortality Improvement in the UK

In common with many developed countries round the world, the UK has recently experienced substantial reductions in mortality rates. The pace of improvement, especially at older ages, has accelerated strongly as the figures in Table 20.1 demonstrate.

Table 20.1 shows that broadly the same fall (circa 20%) in the rate of mortality for males aged 65–74 has occurred in successive periods of 68, 17, 10 and then 6 years.

R. Willets (✉)
Willets Consulting Limited, Minneapolis, MN, USA

Table 20.1 Reduction in the mortality rate for males aged 65–74 in the England & Wales population since 1901

Time period	Reduction in the mortality rate for males aged 65–74
1901–1969 (68 years)	19.4%
1969–1986 (17 years)	21.6%
1986–1996 (10 years)	19.4%
1996–2002 (6 years)	21.1%

Source: The Office for National Statistics

The pace of change at the beginning of the twenty-first century has therefore been more than ten times as rapid as that seen in the first seven decades of the twentieth century.

This simple example illustrates the extent to which the pace of change in mortality rates at older ages has improved over time. More generally, we have seen a trend towards faster improvements at older ages, but less rapid change at younger ages. This feature of mortality change has applied to many developed countries and is sometimes referred to as the “aging of mortality improvement” (Wilmoth 1997).

Figure 20.1a, b illustrate this trend of more rapid improvement at increasingly advanced ages by comparing average annual rates of mortality improvement¹ over the past four decades with rates for the previous 50 years.

In terms of individual causes of death, the single most important driver of these accelerated improvements has been the substantial reduction in heart disease mortality seen in recent decades. This is illustrated by Fig. 20.2a, which shows the crude rate of heart disease mortality for men aged 65–74.

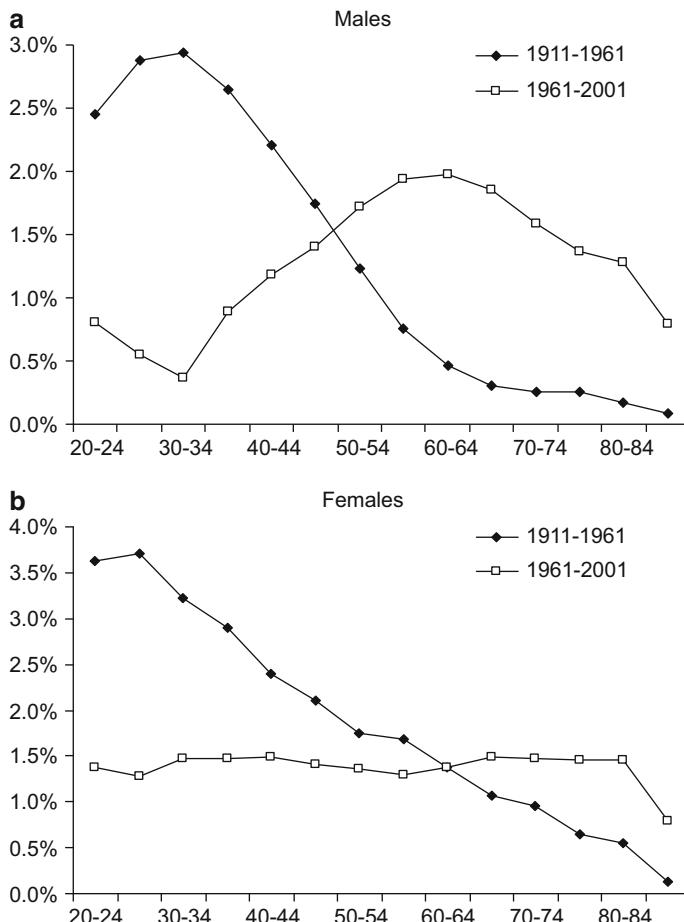
Figure 20.2a shows that the death rate from heart disease for men aged 65–74 has fallen by almost 60% since 1985. The reduction is equivalent to an annual rate of improvement of 4.8% p.a.

The major contributor to the decline in heart disease mortality is believed to be reduced cigarette smoking prevalence (Kelly and Capewell 2004). However, reduction in population blood pressure and cholesterol levels and improvements in treatment have also played a significant role. These positive trends have comfortably outweighed the impact of adverse trends in obesity, diabetes and lack of physical activity, which together contributed approximately 8000 extra deaths in England & Wales between 1981 and 2000 (Unal et al. 2004).

There have also been substantial reductions in other leading causes of death, such as stroke and cancer, at these ages, as illustrated by Figs. 20.2b and 20.2c.

However, the major contributor to the recent rapid improvement in mortality at older ages has been heart disease. Willets et al. (2004) showed that over half of the recent mortality improvement for men in their 60s in England & Wales was due to heart disease alone. The figures in Table 20.2 also show that most of the remainder of the improvement was due to reductions in stroke and cancer mortality.

¹Throughout this paper the term “mortality improvement rate” is taken to mean the rate of change in the mortality rate at a given age from year to the next, i.e. $1 - m(x,t)/m(x,t-1)$, where $m(x,t)$ is the central mortality rate for age x and time t .



Source: The Office for National Statistics

Fig. 20.1 Average annual mortality improvement rates, England & Wales population, 1911–2001.
(a) males, (b) females. (Source: The Office for National Statistics)

It is also worth noting that at younger ages, such as the 30–39 age group, improvements in mortality due to heart disease, stroke and cancer have been more than offset by adverse trends in other causes, notably those linked to drug and alcohol abuse.

In addition to this trend of accelerating improvements at older ages, UK patterns of mortality change have also been influenced by the feature sometimes referred to as the “UK cohort effect”.

Figure 20.3a, b show how the pace of improvement (the average annual reduction in mortality rates) has varied by year of birth in successive 10-year periods. Three features are evident in both figures:

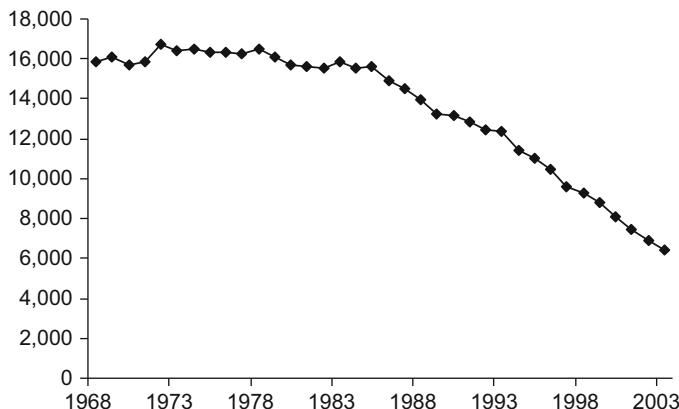


Fig. 20.2a Heart disease deaths per 1,000,000, ages 65–74, England & Wales population, 1968–2003, males. (Source: The Office for National Statistics)

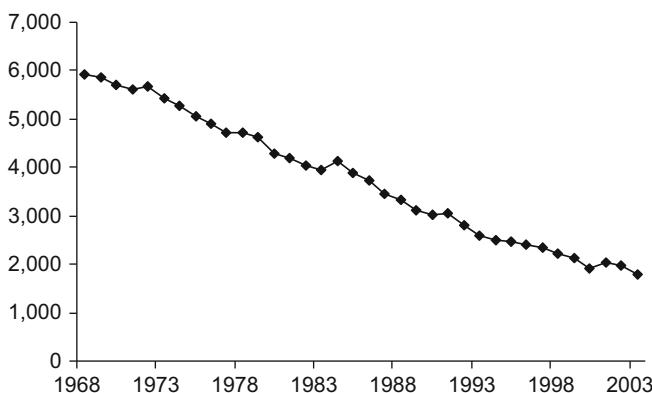


Fig. 20.2b Stroke deaths per 1,000,000, ages 65–74, England & Wales population, 1968–2003, males

- The pace of improvement has been consistently higher for people born in the period 1925–1945 than for people born either side of this generation (this being the so-called “UK cohort effect”). In particular more rapid improvement has been a feature of mortality change for those born in the period 1925–1935, with a secondary peak around 1945.
- Many of the peaks and troughs in the rates of improvement are replicated over time, showing the enduring significance of year of birth in determining the pace of change at different ages.
- The pace of improvement within birth cohorts has increased over time. In other words, the pace of improvement has been accelerating in recent decades, after allowing for the impact of the cohort effect.

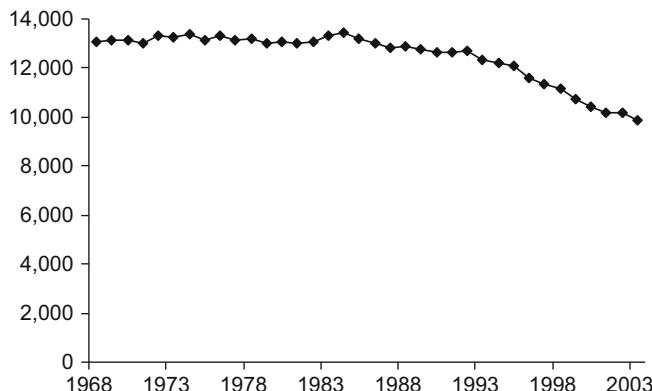


Fig. 20.2c Cancer deaths per 1,000,000, ages 65–74, England & Wales population, 1968–2003, males. (Source: The Office for National Statistics)

Table 20.2 Breakdown of contributions to the overall rate of mortality improvement over the period 1989–2001, by cause, England & Wales population, males

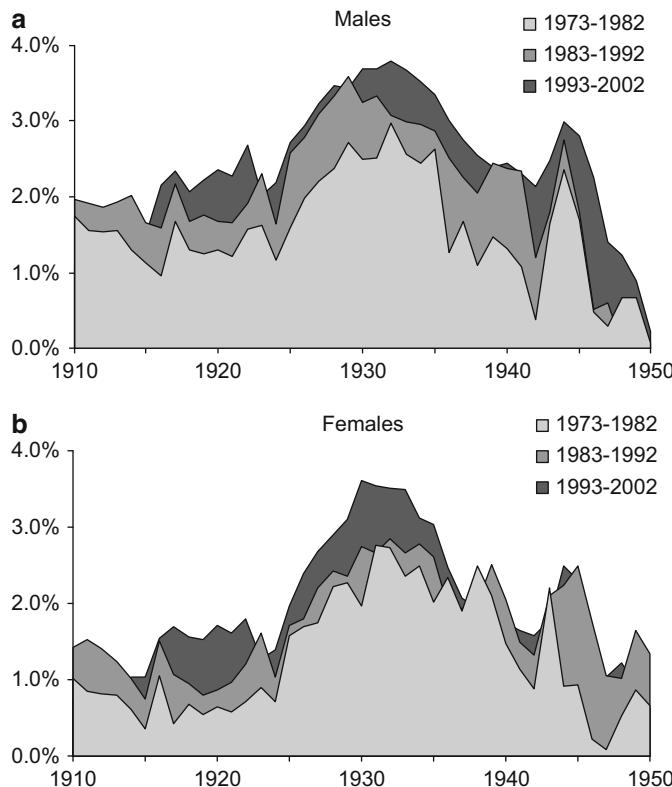
Cause	Ages 30–39	Ages 60–69
Heart disease	+0.4%	+1.8%
Stroke	+0.3%	+0.3%
Cancer	+0.1%	+0.9%
AIDS	+0.3%	0.0%
Drugs/alcohol/liver disease	-0.8%	0.0%
Violent/accidental	-0.3%	+0.3%
All others	0.0%	+0.4%
Aggregate	-0.1%	+3.4%

It is also worth noting that a similar effect can be seen in other developed countries. Data from the Human Mortality Database maintained by the University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany) was analysed for 17² developed countries (all those included on the database excluding those in Eastern Europe). For each of the 17 countries data from 1950 to 2003 for ages 40 to 89 (subject to the years available in each case) were used to calculate average rates of mortality improvement by year of birth.

The results are illustrated by Fig. 20.4. It is notable that the pace of improvement has been significantly more rapid for males born in or around 1935–1940 than generations born before or after this period.³

²Austria, Belgium, Canada, Denmark, England & Wales, Finland, France, Italy, Japan, Netherlands, New Zealand, Norway, Spain, Sweden, Switzerland, USA and West Germany.

³It is also worth noting that the trough in the pace of improvement for the “1920-born” generation is actually far lower than Figure 20.4 suggests. Without the impact of smoothing, the figure for this year is minus 7.8% p.a. This feature is likely to have been caused by the impact of the 1919 influenza pandemic.



Source: The Office for National Statistics

Fig. 20.3 Rate of mortality improvement by year of birth and 10-year period, England & Wales, smoothed using 7-year rolling averages. (a) males, (b) females. (Source: The Office for National Statistics)

20.2 Current Methodologies and Research in the UK

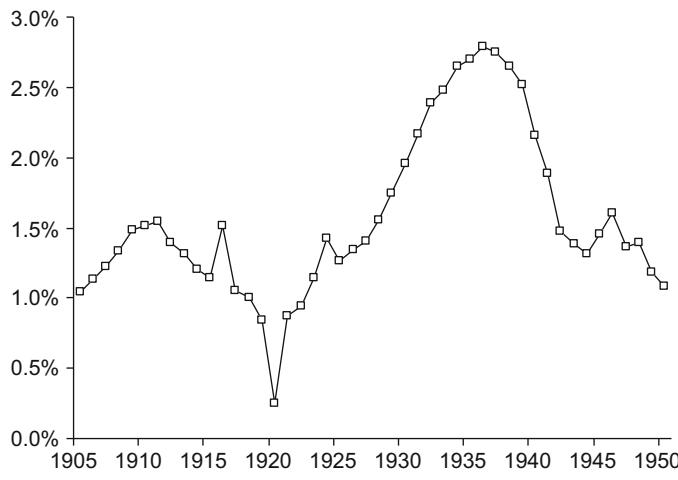
Mortality projections for the UK population are currently produced by the Government Actuary's Department (GAD). The projections assume that "current" rates of mortality improvement – based on the most-recent trends in aggregate mortality – will converge with target rates over a 25-year time-frame.

The latest projection, the so-called "2002-based" projection (GAD 2003), assumed target rates of improvement of 1.0% p.a. for both males and females.

The rates of improvement are projected on a cohort basis for generations born prior to 1947.

A number of variant projections are also made using alternative improvement scenarios.

Projections of future mortality for pensioners and annuitants in the UK are produced by the Continuous Mortality Investigation (CMI), a research group of



Source: Human Mortality Database

Fig. 20.4 Average annual rate of mortality improvement by year of birth for 17 developed countries, males, ages 40–89, data from 1950 to 2003, figures smoothed using rolling averages. (Source: Human Mortality Database)

the UK Actuarial Profession. The last official projection, published in conjunction with the “92 series” of mortality tables for pensioners and annuitants, is based on assumed rates of future improvement in historic trends by age group (CMI 1999). Rates of improvement in the so-called “CMIR17” basis were assumed to diminish over time, consistent with the idea of ultimate (or minimum) rates of mortality.

In 2002 interim “cohort” projections were published by the CMI (CMI 2002) which combined the rates of change in the CMIR17 basis with blocks of rapid improvement consistent with the projection of the UK cohort effect into advanced ages. Three variant projections were produced which differed in the extent to which the cohort effect was assumed to be projected forwards into the future.

Both the GAD and CMI projection methodologies project aggregate rates of mortality, rather than using a cause-of-death methodology.

Cause-of-death modelling has not, and is not, generally favoured as an approach for projecting future mortality rates in the UK. The 1976-based GAD projection of UK population mortality did model future improvements for 10 distinct groups of causes of death. However, this methodology was not adopted for future projections and a major review of the projection methodology for the UK population (GAD 2001) concluded that: “*projections of mortality should not be carried out by cause of death*”.

A similar review paper published by the CMI (CMI 2004) sought feedback from the UK Actuarial Profession on the methodology to be adopted for future mortality projections. One question it asked was whether projections should be carried out on an aggregate or a cause-of-death basis. The response was overwhelmingly in favour of an aggregate methodology.

Arguments against cause-of-death projections included:

- interactions between different causes are difficult to model, especially at older ages;
- medical/research effort will shift as the relative importance of different causes changes over time;
- there are problems with classifying the true cause of death in the very old; and
- changing methods of cause classification over time can distort trends.

Much of the research on future mortality improvement currently being carried out in the UK is being driven by regulatory change in the UK insurance industry which favours a stochastic approach to modelling risk.

Major topics of research include all forms of stochastic mortality modelling, such as Lee-Carter and variants, and advanced methods of smoothing mortality surfaces such as p-splines (CMI 2005).

However, two big questions at the time of this writing – at least for the UK insurance and pensions industries – are:

- whether the UK cohort effect will be projected forwards into the future as the 1925–45 generation ages; and,
- whether the pace of improvement in mortality rates for older ages will continue to accelerate.

For instance, a Guidance Note recently published by the UK Actuarial Profession (2004) states that in determining the capital requirements of an insurance company:-

the ICA [Individual Capital Assessment] should consider firstly, with justification, how any historically observed trends (including cohort effects) might continue, or might continue to accelerate or decelerate.

It is difficult to see how such a justification could be obtained without a consideration of the underlying causes of mortality trends, such as the cohort effect.

20.3 Understanding the “UK Cohort Effect”

A number of possible causes for the UK cohort effect have been discussed. These include:

- patterns of cigarette consumption;
- the adverse impact of World War II;
- patterns in birth rates;
- the development of the UK Welfare State after World War II;
- the impact of diet in early life/maternal malnutrition; and
- early life exposure to infectious disease/lifetime levels of inflammation.

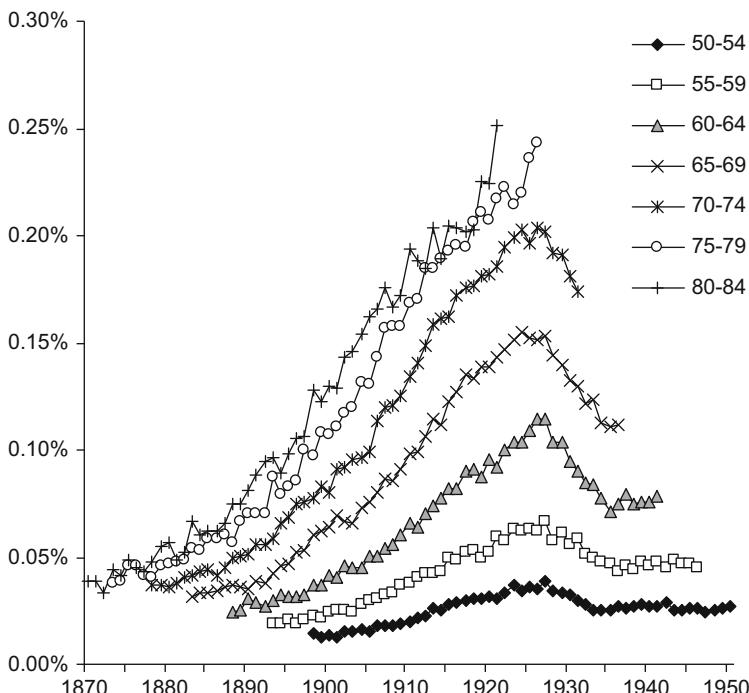
These possible causes are discussed in more detail in Willets (2004). However, to illustrate the value of considering mortality trends for different causes of death, the impact of cigarette smoking will be analysed in more detail here.

Cohort effects in lung cancer mortality rates have been well-documented in recent decades (see, for example, Caselli 1996). Indeed, in women especially, the trends in lung cancer mortality in the UK have been described as providing “*an almost perfect example of a cohort effect*” (Office for National Statistics 1997).

Figure 20.5 shows rates of lung cancer mortality for females in England & Wales by year of birth. It can clearly be seen that the rate of lung cancer deaths at each age group has peaked for those women born in or around 1925.

A similar pattern can be seen for males in England & Wales, with the peak rates of lung cancer mortality occurring for men born in or around 1905.

This data closely matches the pattern shown in figures for lifetime consumption of cigarette tar by year of birth (Lee et al. 1990). There is, therefore, strong evidence that trends in lung cancer mortality by year of birth are correlated with trends in cigarette consumption by year of birth. As a result it is a relatively straightforward task to project future rates of lung cancer mortality for mature generations of UK lives.



Source: The Office for National Statistics

Fig. 20.5 Rate of lung cancer mortality for females in England & Wales by year of birth, using data from 1950 to 2003. (Source: The Office for National Statistics)

As a result of this clear link between lifetime smoking behaviour and mortality from one of the major smoking-related causes of death, it is sometimes argued that the UK cohort effect is unlikely to be projected forwards far into the future.

This argument is based on the suggestion that the UK cohort effect has been largely caused by past patterns in smoking, but that cigarette smoking prevalence has now stabilised in the UK (Office for National Statistics 2004). Furthermore it is argued that smoking-related causes of death (such as lung cancer) are less significant in relative terms at older ages.

In order to explore whether this theory is supported by experience, it is useful to consider historic trends in different causes of death.

20.4 Modelling Mortality by Cause of Death

Tables 20.3a, 20.3b, and 20.3c illustrate the pattern of mortality improvement in three major causes of death for females in England & Wales, namely lung cancer, heart disease and breast cancer. In each case the average annual rate of improvement has been derived for successive periods of 10 years using log linear regression on cause-specific mortality rates.

Table 20.3a clearly shows that the rate of improvement in lung cancer mortality has been particularly rapid for a group of females born in the same period, who are ten years older in each successive 10-year period. The shaded figures relate chiefly to women born in or around 1930. It is notable that the pace of improvement for this cohort has remained relatively constant over time and also that the pace of improvement for adjacent age groups has been far lower.

The equivalent figures for heart disease mortality are given in Table 20.3b.

The pattern of improvement is somewhat different for heart disease. It is evident that the pace of improvement has accelerated over time for all birth cohorts, but that the most rapid pace of change has applied, consistently, to a much wider range of birth years than was the case for lung cancer.

In the case of breast cancer the pace of improvement has also accelerated over time, but was generally greatest for those aged under 50 in 1973–1983, those aged under 60 in 1983–1993 and those aged under 70 in 1993–2003.

Tables 20.3a, 20.3b, and 20.3c only give an approximate indication of cohort effects, by considering age-related improvements in successive periods of time. A more formal analysis can be achieved by modelling mortality rates using an approach in which year of birth parameters are included.

Such a model was constructed using a database of England & Wales population experience for the period 1968–2003. This period covers the years when deaths were classified using ICD8, ICD9 and ICD10; three versions of the International Classification of Diseases.

Raw deaths data for years from 1968 to 2000 were taken from the *20th Century Mortality* database produced by the Office of National Statistics (ONS). More recent data relating to the period 2001–2003 were taken from the *twenty-first Century Mortality* database (ONS).

Table 20.3a Average annual rate of lung cancer mortality improvement in successive 10-year periods, England & Wales population, females

Age group	1973-1983	1983-1993	1993-2003
40-44	+1.5%	-1.9%	+4.2%
45-49	+3.7%	+0.3%	+0.4%
50-54	+0.4%	+0.2%	+0.5%
55-59	-2.7%	+3.8%	-0.5%
60-64	-3.7%	+0.3%	+1.3%
65-69	-4.7%	-1.9%	+3.5%
70-74	-4.8%	-2.6%	+0.7%
75-79	-4.5%	-3.1%	-2.0%
80-84	-4.3%	-3.5%	-2.4%

Source: The Office for National Statistics

Table 20.3b Average annual rate of heart disease mortality improvement in successive 10-year periods, England & Wales population, females

Age group	1973-1983	1983-1993	1993-2003
45-49	+3.6%	+5.4%	+1.4%
50-54	+0.6%	+5.3%	+5.0%
55-59	-0.1%	+4.8%	+6.6%
60-64	+0.4%	+3.2%	+7.1%
65-69	+0.7%	+2.2%	+7.2%
70-74	+1.1%	+2.1%	+6.3%
75-79	+1.2%	+1.6%	+5.2%
80-84	+1.5%	+1.0%	+4.7%

Source: The Office for National Statistics

Mid-year population estimates were also taken from the most up-to-date ONS publications which incorporate the most recent revisions resulting from the 2001 Census results (October 2004).

For each calendar year (1968–2003) death numbers, split by 5-year age groups (up to 80–84) by gender and cause of death, were divided by the equivalent mid-year population estimates. Hence, central mortality rates for 5-year age bands were derived.

Table 20.3c Average annual rate of breast cancer mortality improvement in successive 10-year periods, England & Wales population, females

Age group	1973-1983	1983-1993	1993-2003
45-49	+0.5%	+1.6%	+3.9%
50-54	-0.6%	+1.6%	+3.5%
55-59	-1.1%	+1.7%	+2.8%
60-64	-0.7%	+0.6%	+2.9%
65-69	-0.5%	+0.2%	+3.2%
70-74	-1.1%	+0.3%	+2.6%
75-79	-0.6%	-0.5%	+1.3%
80-84	-1.0%	-1.1%	+1.9%

Source: The Office for National Statistics

Using these central mortality rates for each age group (x) and calendar year (t), mortality improvement rates were calculated for age groups between 40–44 and 80–84 inclusive, i.e.

$$\text{improvement rate, } \delta(x, t) = 1 - m(x, t)/m(x, t-1)$$

Each improvement rate was then assigned to one central year of birth. For example, the improvement rate for the 60–64 age group, for calendar year 2003, was assigned to year of birth 1941.

Various models can be constructed to decompose the *rates of improvement* for different causes using a combination of age, period and cohort factors. Age-period-cohort models have been widely used by epidemiologists and demographers to model mortality rates (see, for example, Tabeau 2001). These models are commonly fitted to log mortality rates. However, for this purpose, the rates of improvement themselves (i.e. the $\delta(x, t)$ terms) have been modelled. This approach is felt to produce results which are relatively easy to interpret and adapt for the projection of future mortality rates.

One feature of age-period-cohort models is that they do not provide a unique solution because of the interdependence of the three terms. There are various strategies to overcome this “identification problem”. However, for this particular paper, it was decided to consider the results of a simplified version of the model only, i.e. one with just period and cohort terms:-

$$\delta(x, t) = \beta(t) + \gamma(t-x)$$

where t = calendar year, $t-x$ = year of birth, $\sum w\gamma(t-x) = 0$, w = a weighting factor for each cohort taken as the number of deaths observed for that cohort.

This approach can be justified because the model fits rates of improvement rather than log mortality rates.

The missing age term is of far less significance than would be the case with a traditional age-period-cohort model. In fact, this age term can be seen as equivalent, in very broad terms, to the $b(x)$ term in the Lee-Carter model (Lee and Carter 1992), where:

$$\log m(x, t) = a(x) + b(x)k(t)$$

Most significantly there is not a clear pattern to the model residuals by age and time, which would indicate a poor fit.

The model can be fitted using a weighted least squares approach applied directly to actual and expected improvement rates or by using a maximum likelihood or minimum chi-squared function derived for the underlying mortality rates. All three approaches give similar results. In this instance results derived by applying the maximum likelihood approach have been used.

The two functions (i.e. period and cohort) derived from fitting the simplified model to cause-specific mortality data for females are given in Figs. 20.6a, 20.6b, 20.6c, 20.6d, 20.6e, and 20.6f. In each case rolling averages were used to identify underlying patterns in the data. In the case of the period function, the improvement rates for 1984 and 1993 were removed from the analysis as they were distorted by changes in the methodology used to assign a main cause to a death certificate. Likewise the rates for 2001 were also removed because ICD10 was first applied as a method of cause classification in this year.

It is interesting to note that the pattern of the cohort function is very different for lung cancer and heart disease. There is clear evidence that the cohort effect applies to later-born generations in the case of heart disease. This does not correlate well with trends in lung cancer or cigarette consumption by generation.

Another way of exploring how year of birth factors have influenced trends in different causes of death is to analyse how well a basic Lee-Carter model fits mortality rates for different birth years. This approach is illustrated by Fig. 20.7 in which fitted and actual rates are compared. For the three causes of death analysed the Lee-Carter model systematically over-estimated mortality rates for those born in 1935–1945, consistent with the impact of the cohort effect. However, it was again evident that this over-estimation applied to a significantly earlier generation in the case of lung cancer than for the other causes.

It is worth considering some of the characteristics of the causes of death in relation to cigarette smoking.

A review paper by Lee (2000) concluded that the relative lung cancer risk among current smokers was 10–20 times that of those who have never smoked. Furthermore, it was found that it generally took ex-smokers 20–25 years after giving-up to reduce the additional risk by 75%.

On the other hand, a similar review paper on heart disease risk (Lee 2001) concluded that the average relative risk of current smokers to those who have never smoked was 212%. Furthermore, it took ex-smokers 5–9 years after quitting to reduce this additional risk by 75%.

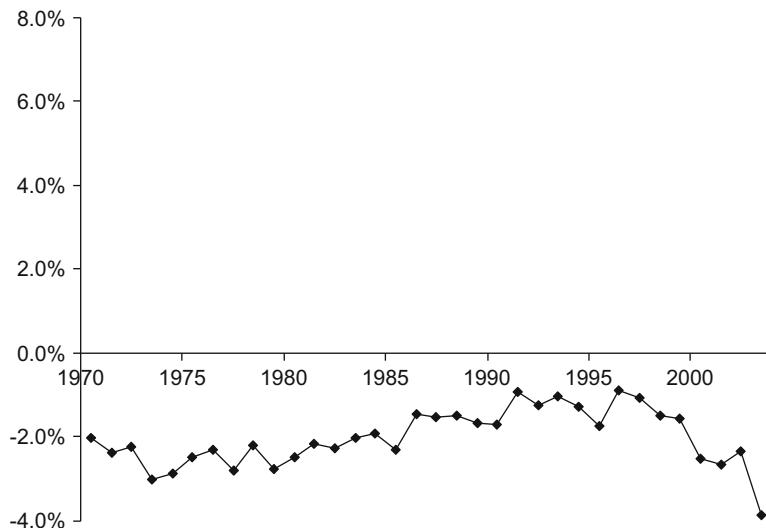


Fig. 20.6a Period function $\beta(t)$ derived by fitting a model of mortality improvement to lung cancer rates for females, England & Wales, 1968–2003

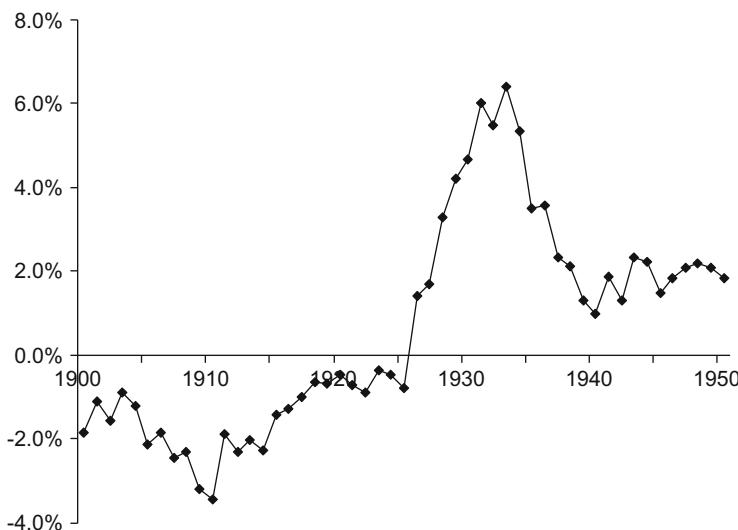


Fig. 20.6b Cohort function $\gamma(t-x)$ derived by fitting a model of mortality improvement to lung cancer rates for females, England & Wales, 1968–2003

Thus, it can be argued, historic patterns of smoking are much more likely to cause cohort effects in lung cancer than heart disease mortality.

There is also evidence that breast cancer mortality improvements have been faster for those born after 1925 than for those born before date. This pattern is unlikely to

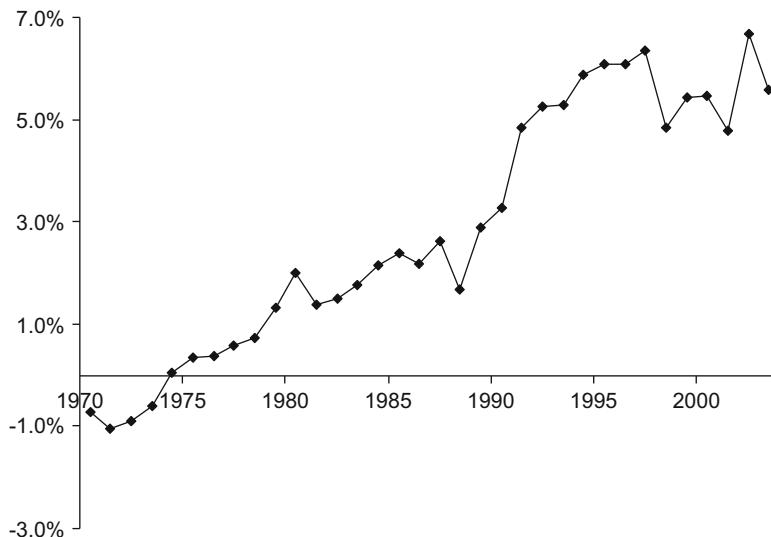


Fig. 20.6c Period function $\beta(t)$ derived by fitting a model of mortality improvement to heart disease rates for females, England & Wales, 1968–2003

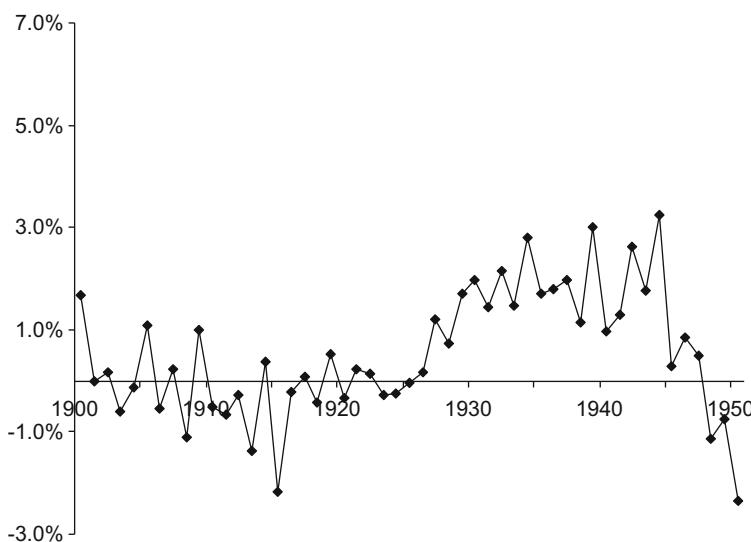


Fig. 20.6d Cohort function $\gamma(t-x)$ derived by fitting a model of mortality improvement to heart disease rates for females, England & Wales, 1968–2003

be due to changing patterns of smoking behaviour as smoking is not considered to be a major risk factor in breast cancer. A review paper by McPherson et al. (2000) made the statement that “*smoking is of no importance in the aetiology of breast cancer*”.

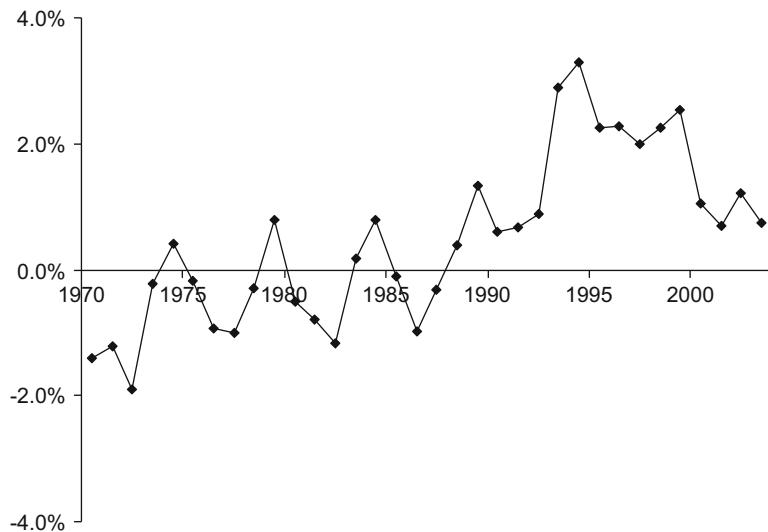


Fig. 20.6e Period function $\beta(t)$ derived by fitting a model of mortality improvement to breast cancer rates for females, England & Wales, 1968–2003

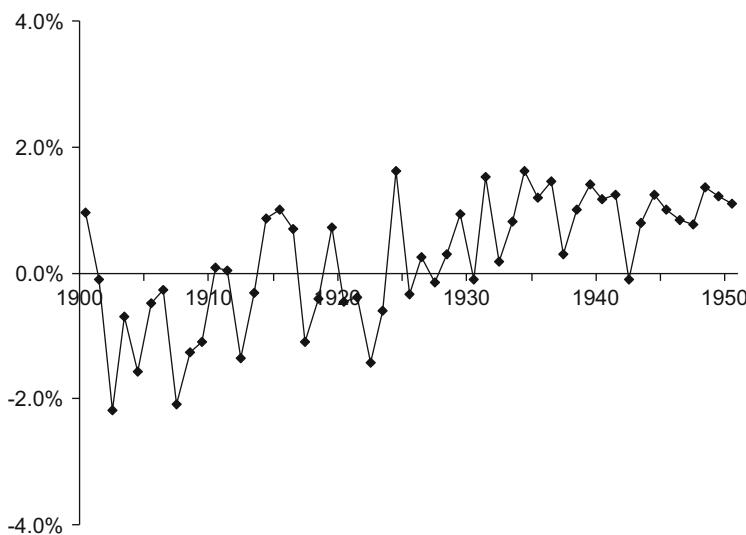


Fig. 20.6f Cohort function $\gamma(t-x)$ derived by fitting a model of mortality improvement to breast cancer rates for females, England & Wales, 1968–2003

The observed cohort effect may be partly due to the fact that the NHS Screening Programme for breast cancer was initiated in 1988. This was aimed – initially – at women aged 50–65, so would have most benefited those born in the 1930s and 1940s. However, it is notable that improvements in breast cancer mortality were also

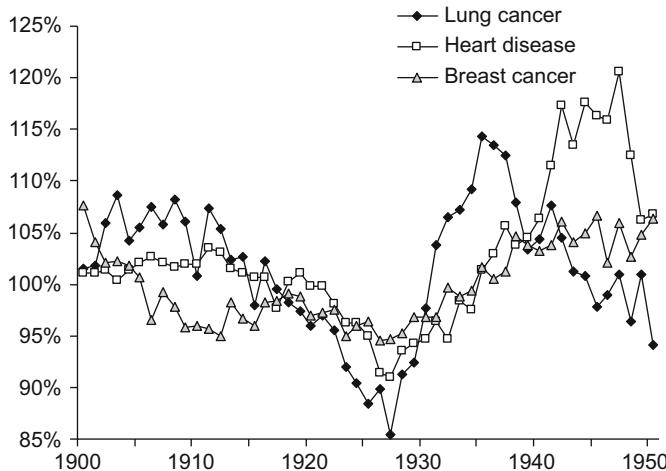


Fig. 20.7 Ratio of “expected” to actual mortality rates derived using a Lee-Carter model fitted to mortality rates for lung cancer, heart disease and breast cancer for females, England & Wales, 1968–2003, averaged by year of birth

relatively high (compared with other age groups) for women aged in their 30s in the 1970s and in their 40s in the 1980s.

It can therefore be argued that prevalence of cigarette smoking from one generation to the next has certainly been one factor which has driven the UK cohort effect and that, as a result, there is a degree of inevitability in some element of likely future improvement, especially for mortality at older ages from conditions strongly linked to smoking.

However, trends in heart disease and breast cancer mortality suggest that smoking may not be the only factor. In Willets (2004) it is argued that there appear to be two ‘sub-cohorts’ of the 1925–45 cohort: an earlier group where the improvements are largely due to smoking and a later one where other factors, such as diet in early life or exposure to infectious diseases, may have played a greater role.

The key point for this paper is that analysis of mortality trends by cause of death can play a vital part in determining the factors driving mortality trends, such as the cohort effect. Furthermore it is argued that such an understanding allows trends to be appropriately allowed for in the projection of future mortality rates.

20.5 Implications and Conclusions

In section four it was argued that in order to understand trends and observed features in aggregate mortality, trends in individual causes of death need to be analysed.

This understanding is necessary because subjective judgments are always made when projecting future rates of mortality, no matter what method is selected. Even the most mechanical method applied to aggregate mortality rates requires decisions

to be taken. Specifically, the precise structure of the model needs to be decided and the period of past data on which to base the future projection needs to be chosen.

In projecting future mortality rates for UK pensioners, it is necessary to form a view on (at least) the following points in deriving a suitable methodology:

- Should the model be parameterised using year of birth or attained age components (or both)?
- If year of birth parameters are used, should their effect reduce with time, increasing age or neither?
- Should the general pace of improvement continue at its current pace or be assumed to accelerate or decelerate over time? Should it revert to a long-term average rate and, if so, what time period should be used to calculate that average?
- Should the pace of improvement for males and females converge over time? If so, how long should the period of convergence last?
- Should the pace of improvement be assumed to be faster for (for example, higher socio-economic class) pensioners than for the average population? If so, how much faster and at which ages?

An understanding of the forces driving historic trends is an essential element of making decisions of this nature.

In fact, despite the historical experience of using cause-of-death projections and the well-documented difficulties, there nevertheless appears to be a good argument for utilising cause-of-death projections in making forecasts.

In forecasting mortality rates for those under the age of (say) 80 it can be instructive to divide deaths into a small number of cause-groupings, perhaps those with very strong historic trends, and compare the results with equivalent aggregate projections.

Cause-of-death modelling can also be a good methodology to test “extreme scenarios”, which are becoming of increasing interest to insurance regulators and capital markets. Such an approach tends to be welcomed by users of such projections, who can see the methodology “grounded in reality.” It can also provide a suitable mechanism for allowing for expert medical opinion in different diseases.

References

- Caselli, G. (1996). Future longevity among the elderly. In G. Caselli & A. D. Lopez (Eds.), *Health and mortality among elderly populations*. Oxford: Clarendon Press.
- Continuous Mortality Investigation. (1999). *CMI Report 17*. London: CMI.
- Continuous Mortality Investigation. (2002). *An interim basis for adjusting the '92 series' mortality projections for cohort effects* (Working Paper No. 1). London: CMI.
- Continuous Mortality Investigation. (2004). *Projecting future mortality: A discussion paper* (Working Paper No. 3). London: CMI.
- Continuous Mortality Investigation. (2005). *Projecting future mortality: Towards a proposal for a stochastic methodology* (Working Paper No. 15). London: CMI.
- Government Actuary's Department. (2001). *National population projections: Review of the methodology for projecting mortality*. London: GAD.

- Government Actuary's Department. (2003). *'2002-based' UK population projections*. London: GAD.
- Human Mortality Database., University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at <http://www.mortality.org> or <http://www.humanmortality.de>. (data downloaded on 26 May 2005).
- Kelly, M. P., & Capewell, S. (2004). *Relative contributions of changes in risk factors and treatment to the reduction in coronary heart disease mortality* (Health Development Agency Briefing Paper). London.
- Lee P. N. (2000). *A review of the epidemiology of lung cancer related to active smoking*. See <http://www.pnlee.co.uk>
- Lee P. N. (2001). *A review of the epidemiology of heart disease related to active smoking*. See <http://www.pnlee.co.uk>
- Lee, R. D., & Carter, L. (1992). Modelling and forecasting US mortality. *Journal of the American Statistical Association*, 87(419), 659–671.
- Lee, P. N., Fry, J. S., & Forey, B. A. (1990). Trends in lung cancer, chronic obstructive lung disease and emphysema death rates for England & Wales, 1941–85, and their relation to trends in cigarette smoking. *Thorax*, 45, 657–665.
- McPherson, K., Steel, C. M., & Dixon, J. M. (2000). Breast cancer – Epidemiology, risk factors and genetics. *British Medical Journal*, 321, 624–628.
- Office for National Statistics. (1997). *The health of adult Britain 1841–1994*. London: HMSO.
- Office for National Statistics. (2004). *Living in Britain; Results from the General Household Survey*. London: HMSO.
- 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* (European studies of population) (Vol. 9). Dordrecht: Kluwer Academic Publishers.
- UK Actuarial Profession. (2004). *Individual capital assessment*. Version 1.1. Guidance Note 46.
- Unal, B., Critchley, J. A., & Capewell, S. (2004). Explaining the decline in coronary heart disease mortality in England & Wales between 1981 and 2000. *Circulation*, 109, 1101–1107.
- Willets, R. C. (2004). The cohort effect: Insights and explanations. *British Actuarial Journal*, 10(4), 833–877.
- Willets, R. C., Gallop, A. P., Leandro, P. A., Lu, J. L. C., Macdonald, A. S., Miller, K. A., Richards, S. J., Robjohns, N., Ryan, J. P., & Waters, H. R. (2004). Longevity in the 21st century. *British Actuarial Journal*, 10(4), 685–832.
- Wilmoth, J. R. (1997). In search of limits. In K. Wachter & C. Finch (Eds.), *Between Zeus and the Salmon: The biodemography of longevity*. Washington, DC: National Academy Press.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Part V

Cohort Factors: How Conditions in Early Life Influence Mortality Later in Life

Tommy Bengtsson

The chapters of the final part address the question of how cohort factors such as nutrition and diseases in early life affect health and mortality in later life and how information on cohort effects best can be used to improve mortality forecasting. The idea that conditions both in infancy and early childhood and during the fetal stage have an impact on adult health dates far back in time, at least to the seventeenth century. It was advocated again in the 1930s, when demographers and epidemiologists became aware of the great mortality decline that started in the West around 1800. While cohort factors were of substantial importance up until the 1920s, it seems that period factors, such as improved water supply, sanitation, and antibiotics – in other words factors that influence health status of all age groups at the same time – gained importance thereafter. Cohort factors were revitalised again in the 1990s by the work of David Barker and his group and from that time on, there has been an upsurge in the number of studies on this topic. While many give support to the idea that early life conditions present in the fetal stage and the first years of life are important for health later in life, others question this conclusion. The question posed here is, if children very early in life are “programmed” for a certain health status, due to their mother’s health and prevailing living conditions, can this information be used to improve mortality forecasting? To approach this issue, this part gives an overview of findings and criticism of how conditions in early life influence adult and old age mortality, as well as some examples of recent studies.

The first chapter, by Martin Lindström and George Davey Smith, provides an overview of recent research in this area, with special emphasis on Sweden. In doing so they not only give a summary of empirical findings, historical and contemporary, but also of the various mechanisms that have been attributed to the link between health in early life and later in life. The mortality decline in three specific diseases, respiratory tuberculosis, haemorrhagic stroke, and bronchitis, which have accounted

for two-thirds of the total decline in mortality in ages 15–64 years from the mid-nineteenth century to the first decade of the twentieth century in Britain, have all demonstrable influences from infancy and childhood. The timing of improvements of conditions during infancy and childhood and health improvements later in life is precisely as expected for this period. Such a high degree of specificity and timing, however, is often lacking in analyses of contemporary data. Thus, while cohort factors were partly important for the decline from high to low mortality, their importance today needs to be verified further.

Kaare Christensen, in the second chapter, evaluates Barker's "fetal origins hypothesis" using data on twins and famines. Considerable evidence exists of the significant association between fetal growth and later life health outcomes, such as blood pressure and cardiovascular mortality. Firstly, the question is how large is this effect? Secondly, is the effect causal, or do other factors, such as genes or socioeconomic conditions, lie behind the association? Christensen begins by pointing out that the association between retardation of fetal growth and health in later life is not corroborated without exception, referring to several studies of famines. He then turns to twin studies, which are particularly interesting in this respect since twins usually experience retarded fetal growth and on average are 900 g lighter at birth than single children. Using the Danish twin register, which includes a large population with a long follow-up period, Christensen and his colleagues find no differences between twins and singles with regard to all-cause or cardiovascular mortality, and a similar study for Sweden confirms this result. Twins often differ in birth weight, which means that it is possible to control for genetic factors. Christensen and colleagues also use data for the US and find a very modest impact of birth weight on blood pressure, confirming previous analyses using twin data. Taken together, the effect of birth weight on blood pressure, after controlling for genetic factors, is too weak to be used to predict future mortality.

In the third chapter, Gabriele Doblhammer focuses on how factors related to season of birth influence mortality in later life comparing the United States, Austria, Denmark, and Australia. In all four populations, significant differences in lifespan exist by month of birth. Those born in the spring generally face lower life expectancy, likely due to nutritional factors, possibly also due to a lower disease load. The difference diminishes over time, which is consistent with improved diets but also a decline in morbidity in infectious diseases. Taken together, improvements in diet and reduction of infectious diseases in the beginning of life gave improved health later in life and contributed to the overall increase in life expectancy.

The fourth and final chapter, by Tommy Bengtsson and George Alter, analyses how conditions of nutrition and diseases in the first year of life affect old age mortality in historical populations in Belgium and Sweden. They make use of longitudinal individual level demographic data with information on socioeconomic factors at household level combined with data on food prices. They find that children born in years of excess infant mortality, generally due to outbreaks of epidemic diseases, face higher mortality in older ages. For Belgium, high food prices have a similar effect.

To summarize, there is a considerable amount of evidence as regards the ways health incidents in early life influence health later in life and mortality in certain causes of death. There is also evidence as regards its timing for England during the mortality decline from the mid-nineteenth century up to the 1920s and possibly also for Sweden and some other western countries. For the period afterwards, the results are less conclusive. While many studies show significant associations between conditions in early life and health later in life, others do not. Whether this association is causal has also been questioned, as has its overall impact on the mortality decline.

Chapter 21

A Life Course Perspective to the Modern Secular Mortality Decline and Socioeconomic Differences in Morbidity and Mortality in Sweden



Martin Lindström and George Davey Smith

During the past 200 years, most countries in the world have experienced a great increase in life expectancy. The timing of the onset of this decrease in mortality and corresponding increase in life expectancy has vastly differed between different countries, and this is true for the pace of the development as well. Some countries have still not achieved the life expectancy experienced by the most developed countries already 100 years ago or even earlier. Some countries have even experienced a backlash in the form of declining life expectancy in the 1990s due to, for instance, unemployment and alcohol consumption in some eastern European countries, and the HIV/AIDS epidemic in some African countries. Nevertheless, the general picture of improvement remains massively impressive, and in Sweden life expectancy has increased continuously for more than 200 years.

The secular mortality decline may be explained by a multitude of causes rather than one single cause. These causes may be categorised in different ways and from varying perspectives. Some causal effects (exposure leading to disease) are direct or short-term effects, others are long-term effects. The causes do not only include direct, *period effects* on mortality and survival such as the immediate effects of outbreaks of infectious diseases, the presence of endemic infections, as well as current hygiene, income, nutrition, housing and health care conditions. The causes also include long-term, sometimes very long-term, effects. The latter group of long-term causal mechanisms by which risk factors and protective factors affect health and disease many years later are often referred to as *cohort effects*, because different

M. Lindström (✉)

Department of Clinical Sciences, Malmö University Hospital, Lund University, Lund, Sweden

Centre for Economic Demography and Department of Economic History, Lund University,
Lund, Sweden

e-mail: martin.lindstrom@med.lu.se

G. D. Smith

Department of Social Medicine, University of Bristol, Bristol, UK

birth cohorts are exposed to different sets of risk factors as well as protective factors during their childhood and adolescence that affect their health in later life.

This contribution will deal with long-term cohort or early life effects on disease in later life, their biological mechanisms in general (not only in Sweden) and their implications for socioeconomic differences in mortality, with particular reference to Sweden. It will also shortly deal with the possibility of making predictions concerning future mortality based on cohort mortality and its indicators in Sweden and in other countries. In the following sections, we will discuss the early life effects and then their biological mechanisms without particular reference to Sweden.

21.1 The Secular Mortality Decline: Early Life and Cohort Explanations and Their Indicators

By the term *period effect*, we mean effects on health and survival caused by health determinants (see above) with a short time period between exposure and health/disease outcome (exposure factors which affect the risk of disease may promote either salutogenesis or pathogenesis). For instance, most infectious diseases give rise to symptoms in the very short term (hours-days-weeks) after the initial exposure to infection. However, for some infectious diseases such as tuberculosis (caused by mycobacterium tuberculosis) and leprosy (caused by mycobacterium leprae), the time interval from exposure to disease/symptoms may be much longer (months-years) due to the slow pace of multiplication of the pathogen in the infected human host. Other diseases, especially non-infectious chronic diseases such as many cancers and forms of cardiovascular disease, may have much longer latent periods, i.e. time intervals from exposure of determinants to disease amounting to several decades.

For some diseases, the time lag between exposure and disease may even range from early life (intra-uterine or first year(s) of life) exposure resulting in old-age morbidity and mortality. This set of factors is causally related to the mortality decline and concerns the effects of cohort or early life events on mortality in later life. The general idea behind the notion of *cohort effects* is that varying forms of stress or heavy disease load on the different organs or organ systems in the human body experienced in early life, most importantly during pregnancy and the first year(s) of life after birth, may “program” the organs to increased susceptibility to various diseases much later in life. However, the notion of cohort or early life causes of disease in later life is not restricted to purely biological mechanisms. Early life events of psychological significance experienced in early life may also give rise to psychological problems or certain persistent personality traits in adulthood (Suomi 1997).

The cohort or early life explanation was proposed by Kermack et al. 1934 (see also Davey Smith and Kuh 2001). They studied age-specific mortality in England, Wales, Scotland and Sweden. Their conclusion was that reductions attained at any particular time in the death rates of the various age groups depended primarily on the

date of birth of the individuals, and only secondarily on the actual year of death. The essential beneficial effects on health and survival among adults and older persons were mainly caused by a decrease in disease load achieved in these birth cohorts during early childhood several decades earlier, according to Kermack et al. (1934).

The past decades have witnessed a renewed interest in the cohort or early life approach to disease in Sweden (Bengtsson and Lindström 2000, 2003) as well as in other countries (Preston et al. 1998), particularly in chronic disease epidemiology (Kuh and Ben-Shlomo 1997, 2004; Galobardes et al. 2004). This approach has been heavily supported particularly by the work of Barker and colleagues. They have both hypothesised and investigated the early life preconditions for later life development of cardiovascular diseases and the metabolic syndrome, i.e. coronary/ischaemic heart disease, hypertension, adverse levels of blood cholesterol and lipids, stroke, type II diabetes mellitus, and overweight/obesity (i.e. the components of what some call the “metabolic syndrome”). The causal mechanism behind these diseases induced in early life is suggested to be inadequate cellular development in utero due to lack of sufficient nutrition (Barker 1994, 1995, 1997, 1998, 2001). The concept of down-regulation of fetal growth has been developed further into the nutritional programming (or fetal origins) hypothesis. According to this hypothesis the development of cardiovascular and other diseases in later life depends on whether fetal growth retardation due to insufficient nutrition is “proportionate” or “disproportionate”. The “disproportionate” growth retardation induced by insufficient nutrition during the mid and late trimesters of pregnancy seems to be responsible for cardiovascular diseases later in life, while the “proportionate” growth retardation of the first trimester is not (Barker 1995), although this distinction has later been tuned down by Barker (1998).

Not all evidence suggests an exclusive or even important role of malnutrition in the fetal origins hypothesis. For instance, maternal tuberculosis also impairs fetal growth (Riley 2001). The famine in rural Finland from 1866 to 1868 tripled death rates but did not alter the survivors’ lifespans (Kannisto et al. 1997).

The original disease load mechanism proposed by Kermack et al. has been developed and further investigated. Later Fridlizius (1989) suggested that the development of diseases in later life might be due to exposure to certain infectious diseases. For example exposure to smallpox in the late eighteenth century and exposure to scarlet fever in the mid nineteenth century, in the first 5 years after birth, resulted in reduced immunity against other diseases throughout life and thus a higher susceptibility to getting other infectious diseases in later life. In neither case did susceptibility to disease in adulthood seem to have been connected with nutrition in early life, because the risks of being infected with for instance smallpox and scarlet fever are to a high extent independent of nutrition (Rotberg and Rabb 1985). However, some findings of recent empirical investigations have suggested an association between nutrition and morbidity and mortality of scarlet fever epidemics in the Sundsvall region in northern Sweden (Curtis 2004). In contrast to nutrition, Fridlizius suggested deranged immunological balance between some specific infectious agents and the human host, which has implications for later life experiences of disease (Fridlizius 1989).

In recent years, the rather unspecified mechanisms suggested by Fridlizius have received some support from the bio-medical literature. Chronic inflammatory mechanisms may drive much of the influence of early life infections on later morbidity and mortality. Populations living in high mortality contexts are highly exposed to a wide variety of infectious diseases. Such populations also have high risks of acquiring chronic infectious diseases such as tuberculosis (Lawn et al. 2000) and infections caused by *Escherichia coli* and *Helicobacter pylori* (Cadwgan et al. 2000). These diseases lead to chronically elevated levels of inflammatory markers such as C-reactive protein, interleukin-6, tumour necrosis factor- α and fibrinogen that may mediate between early life infection and later life chronic disease morbidity and mortality (Finch and Crimmins 2004). Thus, reduced morbidity and mortality from infectious diseases in populations experiencing the great mortality decline could produce decreases in exposure to these markers of inflammation. Whether these inflammatory mediators actually have causal influence on chronic disease risk is not established (Timpson et al. 2005).

Helicobacter pylori is an established cause of peptic ulcers, and is associated (although maybe not causally) with coronary heart disease (Harvey et al. 2002). Infections caused by *Helicobacter pylori* are most commonly contracted in infancy and childhood and they persist throughout life. *Helicobacter pylori* infections are now declining in most low-mortality countries due to improvements in public health and hygiene (Li et al. 2000).

Exposure to infections during the fetal, perinatal and postnatal stages may affect both anatomical/organ development and development of the immune system. The effects of infections during the fetal stage depend on a number of fetal and maternal factors such as nutrition, genetic factors, fetal development stage and anatomical factors. Other examples of such infections are influenza and rubella. One example of a causal association between postnatal infection and adult disease is the association between Hepatitis B and primary liver cancer (Hall and Peckham 1997). A contemporary study from the USA on Americans aged 55–65 years shows that infectious disease during childhood multiplied the incidence of lung conditions, such as emphysema and bronchitis, by four in the 55–65-year age group. Non-infectious diseases showed much weaker associations with adult disease (Blackwell et al. 2001). It thus seems plausible that the prenatal and postnatal development of the lungs and the immune system are sensitive to critical events which may influence susceptibility to infections, allergic reactions or toxic exposures, but the exact and specific timing and critical periods for such early life influence on health later in life remain to be disentangled (Strachan 1997). A study of children born in 1921–1935 in Scotland also shows reduced lung capacity (in 1986) for those who experienced pneumonia before the age of 2 years (Shaheen 1997). Factors in utero and during the first years of life may affect the development of asthma later in childhood and adulthood (von Mutius 2001). A review of the effects on human lifespans of the inflammation/infection exposure in early life has proposed a “cohort morbidity phenotype” which represents inflammatory processes that persist from early age into adult life (Finch and Crimmins 2004). Early life experience of diarrhoea with subsequent dehydration may plausibly lead to higher blood pressure, a risk factor for

cardiovascular diseases in general and haemorrhagic stroke in particular in later childhood and adulthood, a hypothesis which has been found to be supported by some empirical findings (Davey Smith et al. 2006; Lawlor et al. 2006).

There is also some support in the literature of an effect of both nutrition and disease load (particularly infectious diseases) in early life. Unfavourable early life conditions generally seem to cause permanent biological damage, resulting in higher mortality in later life (Doblhammer and Vaupel 2001). The results of a large sample study of 15 million US deaths between 1989 and 1997 has also suggested effects of season of birth on mortality risk in later life. Being born during a season of hardships is associated with higher mortality in later life (Doblhammer 1999; Doblhammer 2008). Seasonal differences in exposure to infectious disease in early life are associated with mortality in adult life. Seasonal differences in the nutrition of the mother during pregnancy also seem to affect mortality in later life (Doblhammer 2002). A study from contemporary rural Gambia has shown that higher mortality levels are explained by permanently damaging effects during early life of disease exposure as well as malnutrition during the yearly dry-season. Both the damaging effects of disease load and malnutrition during the fetal stage of development are by some authors (Moore et al. 1997) attributed to effects on the immune system, a conclusion that may be supported by historical data (Bengtsson and Lindström 2000, 2003). Several recent studies, however, have cast doubt on this conclusion (Simondon et al. 2004; Moore et al. 2004).

Finch and Crimmins (2004) have recently argued that the inflammatory infection and nutrition hypotheses are not competing or contradictory but complementary in linking two mechanisms of morbidity in early and later life. For example, childhood diarrheas impair cardiac muscle synthesis (Hunter et al. 2001), which could explain associations of infant diarrhea with later cardiovascular disease (Blackwell et al. 2001). Slowed infant growth in the Barker hypothesis might consequently hypothetically be explained by inflammatory reactions in combination with impaired nutrient absorption. There is growing evidence from historical data (1766–1894) in Sweden in support of the disease load (particularly infectious diseases) mechanism suggested in two articles by Bengtsson and Lindström (2000, 2003).

There is also a rapidly accumulating amount of evidence in support of the early life conditions or life course approach in general from modern data (Kuh and Ben-Shlomo 2004; Kuh and Hardy 2002; Davey Smith 2003). The relative abundance (compared to historical data) and diversity of variables in modern data make it possible to attempt to understand the interactions between different determinants and successive exposures during the life course. It should thus be noted that modern data support not only the critical period model, which may be exemplified by the already referred to fetal-origins hypothesis. Modern data also support models following Omran's assumptions concerning multicausality and interaction of different causal factors in demography and epidemiology (Omran 1971). In contrast to the simpler mono-causal critical period model and fetal-origins hypothesis, the accumulation of risk model assumes that effects accumulate over the life course, although some particular developmental periods may entail greater susceptibility (Ben-Shlomo and Kuh 2002). Harmful effects on health may increase with the duration and/or number

of harmful exposures. Exposure to poor socioeconomic conditions may for instance lead to additive effects of experiencing low socioeconomic position during different parts of the life course, which may influence the risk of several diseases (Heslop et al. 2001). The accumulation of risk may also be due to the clustering of exposures (Ben-Shlomo and Kuh 2002).

In modern times, chronic diseases dominate the disease patterns both when it comes to morbidity and mortality. Such diseases include for instance cardiovascular diseases, cancers, rheumatoid arthritis, thyroiditis, and musculoskeletal disorders. Coronary heart disease is a good example (Davey Smith and Lynch 2005). It manifests itself during adulthood and old age, but the disease process starts many years earlier with the gradual development of atherosclerosis. This development begins with fatty streaks in the artery walls of children (Berenson et al. 1987). Arterial lesions are also evident in young men suffering from violent death (Strong et al. 1999). Risk factors for coronary heart disease include blood cholesterol levels, smoking, obesity, diabetes mellitus, hypertension, oral contraceptive use among women, psychosocial factors, mental illness, chronic infection/inflammation, coagulation factors, and air pollution (Marmot and Elliot 2005). Several studies have demonstrated that unfavourable pre-adult measures of cholesterol, blood pressure and adiposity are associated with increased intimal-medial thickness, which is a presymptomatic measure of coronary heart disease (Li et al. 2003; Raitakari et al. 2003; Davey Smith and Lynch 2005). These risk factors do not only affect coronary heart diseases in a mono-causal way, but they may also affect coronary heart disease by interacting with each other in order to increase or attenuate each other's effects on the disease aetiology leading to coronary heart disease.

21.2 Historical Trends and Socioeconomic Mortality Differences in a Life Course and Cohort Perspective

The research area that concerns the mortality decline entails a number of important issues that can each contribute to the understanding of the modern mortality decline and its complexity. The eradication of smallpox mortality (Sköld 1996a, b) and the variations in sex differences in mortality (Willner 1999) have been thoroughly investigated and discussed. Another issue concerns socioeconomic mortality differences and socioeconomic differences in the short term as well as secular mortality decline. This socioeconomic gradient to this day remains apparent, despite the development of the modern welfare state and active policies to redistribute income in many countries, e.g. Sweden. In fact, during the past two decades, Sweden has witnessed a continuous decline in age specific mortality rates in most age intervals and a corresponding increase in life expectancy. This mortality decrease is observed in all socioeconomic groups in Swedish society. However, the decrease has been more pronounced in higher socioeconomic strata (high education, high income, non-manual employees in higher positions according to occupational status) than

in lower socioeconomic strata, which has resulted in increasing socioeconomic differences in life expectancy in Sweden during the late 1980s, 1990s and early 2000s (National Public Health Report 2001, 2005).

It is often stated that socioeconomic mortality gradients, with the poor having worse health and increased risk of death compared to the rich, are ubiquitous phenomena, having always existed everywhere. This is an erroneous assumption, however (Davey Smith 2003). Reviewers (e.g., Macintyre 1998) often start with well-known historical examples, such as when Chadwick assembled data from different areas of Great Britain, and generalise to all situations. Chadwick's data, however, did suggest large socioeconomic differences in mortality in the first decades of the nineteenth century in Britain. The socioeconomic differences existed within many UK locales, although the high socioeconomic position gentry and professional population only lived on average 35 years in Liverpool compared to 55 years in Bath. The corresponding average for the labourer and artisan class was 15 and 25 years, respectively (Chadwick 1842; Wohl 1983). Although data from Geneva indicate presence of socioeconomic mortality differences in pre-modern society (sixteenth century) (Perrenoud 1975) and data from an English township 1650–1830 also suggest permanent presence of socioeconomic mortality differences (King 1997), the generalisation by MacIntyre concerning the presence throughout history of socioeconomic differentials in mortality contrasts to important extent with the observation by Livi-Bacci (1991). According to Livi-Bacci, rudimentary older data from England suggest the absence of socioeconomic differentials in mortality in England from approximately 1550 to ca. 1750 (Livi-Bacci 1991). The data that Livi-Bacci refers to are calculations of life expectancy from demographic data on English peers (Hollingsworth 1977) compared with life expectancy of the total English population calculated from the Wrigley and Schofield reconstitution data (Wrigley and Schofield 1981). In fact, the ducal families in England seem to have had a somewhat lower life expectancy than peers in general as well as the general population during the period prior to 1750. This pattern remains even after the increased risk of violent causes of death (including the "Agincourt" factor, i.e. the death-in-combat factor) are taken into account (Hollingsworth 1957). Furthermore, the reigning families of Europe seem to have had a life expectancy of 34 years in the sixteenth century, 30.9 years in the seventeenth and 37.1 years in the eighteenth century, i.e. life expectancies which fairly well correspond with the life expectancies of the general population in the corresponding countries during the same period. In the city of Rouen, fluctuations in grain prices during the ancien régime had a similar effect in various social classes (Galloway 1987).

A similar pattern has been observed in the parishes in the Scania Demographic Database in southern Sweden, where fluctuations in grain prices also had strong and similar effects in all social classes before the agrarian revolution in the early nineteenth century. In contrast, the onset and progress of the agrarian revolution resulted in both weaker associations between short-term fluctuations in grain prices and mortality. It also resulted in increasing socioeconomic differentials in the mortality response to fluctuations in grain prices, as the more prosperous segment of the population seems to have become much less exposed to the effects of the

fluctuations (Bengtsson 2000, 2004). These observations seem to constitute further proof in support of the notion that social differences in mortality were small or absent. Furthermore, the observations support the notion that socioeconomic differences in mortality increased during the eighteenth century because of the agrarian revolution.

Sweden started to gather and record demographic and socioeconomic data (including mortality), different measures of socioeconomic position and, in many parishes, causes of death for the whole country already in 1749. Hence, it is possible to go further back in time in Sweden than in probably any other country in the investigation of reliable demographic and socioeconomic data in order to better understand the dynamics of socioeconomic differences in longevity.

One explanation for the lack of socioeconomic differences in mortality in the rudimentary data presented by, for example, Livi-Bacci for England, may be that epidemic and endemic infectious diseases dominated the disease and mortality panorama in the general population, which is certainly not the case today. In many pre-modern societies, population density seems to have been positively associated with mortality due to increased risk of disease (i.e. infectious disease) exposure in densely populated areas. For instance, the remarkable healthiness of many frontier settlements in colonial North America in spite of their comparatively primitive material living conditions must have been partly due to the infrequent contact with others (Wrigley et al. 1997). The virulence of many such epidemic and endemic infectious diseases, e.g. smallpox, malaria, plague, typhoid, tetanus, yellow fever, encephalitis and poliomyelitis, are not at all influenced (or only minimally affected) by nutritional factors such as total energy intake, nutritional contents of the food and physical habitus. Other infectious diseases such as typhus, diphtheria, staphylococcus infections, streptococcus infections, influenza, syphilis and systemic worm infections are only affected by such nutritional factors to a limited or variable extent (Rotberg and Rabb 1985). This means that the upper socioeconomic strata (i.e. the nobility) must have been exposed to risks of disease and death from common infections prevailing at that time to the same extent as members of the lower social strata. In fact, as social contacts and networks of the upper strata most likely were more extensive than among the lower classes, the exposure in those groups may even have been higher than in the lower strata. As many of the infectious diseases mentioned above decreased in importance during the time period studied, all age-specific mortality rates declined and life expectancy increased. Consequently, other diseases and diagnoses more related to nutritional status and the protecting effects of higher socioeconomic position increased in relative importance as causes of morbidity and mortality, which would have served to increase socioeconomic differences in morbidity and mortality during the period under study. The result would be an increase in socioeconomic mortality differences and thus increased socioeconomic differentials in life expectancy.

In modern times, chronic diseases with long latent, asymptomatic phases between the induction/onset of the disease and the first symptoms dominate the patterns of morbidity and mortality in developed countries. Socioeconomic differences according to social characteristics such as occupational status, education and income

are well-known and have been described extensively both in Sweden (National Public Health Report 2005) and other countries (Marmot 2004; Davey Smith et al. 1990; Kaplan and Keil 1993) regarding morbidity and mortality in a wide variety of diseases. A recent review of the literature on the association between socioeconomic circumstances during childhood and cause-specific mortality during adulthood shows similar results. Adverse socioeconomic conditions during childhood were positively associated with increased all-cause mortality (in 18 of 22 studies), overall cardiovascular mortality (in 5 of 9), coronary heart disease mortality (in 7 of 10), stroke (in 4 of 6), and accidents and violence (in 3 of 5 studies). No such associations were found for rheumatic heart disease mortality (only 1 study) and overall cancer mortality. For lung- and smoking-related cancer mortality, respiratory disease mortality, suicides, alcohol- and illegal drug-related mortality only few studies showing no associations or studies showing diverse results concerning the association between childhood socioeconomic circumstances and cause-specific mortality were demonstrated (Galobardes et al. 2004). It thus seems that the association between childhood socioeconomic circumstances and risk of cardiovascular diseases in adulthood is particularly important in explaining life course effects on adult mortality (Galobardes et al. 2006a, b).

In Sweden only a few studies concerning socioeconomic conditions in childhood and health in adulthood have been conducted, but new data sets have been developed (Stenberg et al. 2007). Birth order position within the same family had statistically significant consequences for the health and survival (overall mortality) over the life course (Modin 2002). Socioeconomic inequities in overweight seem to reflect the cumulative influence of multiple adverse circumstances experienced from adolescence to young adulthood (Novak et al. 2006). Several Swedish studies demonstrate statistically significant associations between disadvantaged socioeconomic conditions during childhood as well as adverse socioeconomic mobility, and aspects of cardiovascular diseases such as all-cause and overall cardiovascular mortality (Rosvall et al. 2006), coronary heart disease (Wamala et al. 2001), myocardial infarction (Hallqvist et al. 2004), and carotid atherosclerosis (Rosvall et al. 2002). In one Swedish study, IQ in early childhood was found to be unrelated to adult cancer mortality (Batty et al. 2007). Childhood conditions such as family disruption and child abuse were found to be unrelated to adult sense of coherence (Krantz and Östergren 2004). The markedly few results from Sweden thus still seem to be consistent with other findings from the international literature.

21.3 Cohort Effects on Mortality and Mortality Predictions: Indicators and Models

A number of models exist to forecast future mortality in populations (Bengtsson and Keilman 2003). There are several reasons why these models should include a historical and long-term perspective on mortality and the development of

age-specific mortality. First, living conditions, i.e. living standards and diet, public health institutions and medicine and other areas relevant for the physical well-being of the population, improve from one period to the next. Such changes in living conditions are termed *period* effects. Second, the health and remaining lifespan of people living today are determined not only by contemporary period factors but also by living conditions earlier in life. Living conditions during childhood may affect health in later life through *cohort* effects on mortality. Third, the prediction of future mortality calls for a multivariate approach, including not one but a multitude of factors to predict mortality. These factors include long-term early life factors (Bengtsson 2003).

It thus seems obvious that early life and cohort factors should be included in the models when making predictions concerning future mortality. The crucial question is what indicators to use in order to assess how early life and cohort factors influence future mortality. The original work by Kermack and colleagues (1934) analysed the relationship between early life mortality, including infant (0–1 year) mortality, and its association with the age-specific mortality of different birth cohorts later during their life courses. Age-specific mortality is now commonly used as an indicator of mortality trends (United Nations 1999). Given the plausibility and scientific evidence for early life effects on cohort mortality presented earlier in this paper, age-specific mortality seems to be an obvious choice of indicator for making predictions concerning future mortality in a population when considering early life cohort effects. Infant mortality seems to be the most crucial measure of all age-specific mortality intervals in this respect (Bengtsson et al. 1998).

Fogel (1994) has used height as an indicator of early life effects on life expectancy and health in later life. In fact, recently both age-specific early life mortality (including infant mortality) and height have been demonstrated to be associated with mortality in later life using historical data from birth cohorts born before the twentieth century in four North European countries (Crimmins and Finch 2006).

Timing and specificity are key factors in life course epidemiology. Davey Smith and Lynch (2004) have pointed out that the mortality decrease in the three specific diseases respiratory tuberculosis, haemorrhagic stroke and bronchitis may have accounted for approximately two-thirds of the total decline in mortality for men and women aged 15–64 from the middle of the nineteenth century to the first decade of the twentieth century in Britain. Some other specific diseases including stomach cancer and rheumatic heart disease may account for some of the residual decline. These diseases have demonstrable influences from infancy and childhood, which have already been discussed. The timing for this time period when it comes to early life/cohort effects is also very good. Underlying factors such as decrease of child labour, increase in real wages, improved nutrition and increased height, a decrease in the proportion of working mothers, a decrease in family size, and improved housing conditions are also present for this period (Davey Smith and Lynch 2004). There is often a lack of such a high degree of specificity and timing in modern data. Specific exposures and outcomes should always be identified as well as the exact timing. The high availability of data in Sweden will plausibly make this task possible to accomplish in the years to come.

References

- Barker, D. J. P. (1994). *Mothers, babies, and disease in later life*. London: British Medical Journal Publishing Group.
- Barker, D. J. P. (1995). Fetal origins of coronary heart disease. *British Medical Journal*, 311, 171–174.
- Barker, D. J. P. (1997). Fetal nutrition and cardiovascular disease in later life. In M. G. Marmot & M. E. J. Wadsworth (Eds.), *Fetal and early childhood environment: Long-term health implications*. *British Medical Bulletin* 53(1), 96–108.
- Barker, D. J. P. (1998). *Mothers, babies, and health in later life*. London: Churchill Livingstone.
- Barker, D. J. P. (2001). The malnourished baby and infant. *British Medical Bulletin*, 60, 69–80.
- Batty, G. D., Modig Wennerstad, K., Davey Smith, G., Gunnell, D., Deary, I. J., Tynelius, P., & Rasmussen, F. (2007). IQ in early adulthood and later cancer risk: Cohort study of one million Swedish men. *Annals of Oncology*, 18(1), 21–28.
- Bengtsson, T. (2000). Inequality in death: Effects of the agricultural revolution in southern Sweden, 1765–1865. In T. Bengtsson & O. Saito (Eds.), *Population and economy. From hunger to modern economic growth* (pp. 301–334). Oxford: Oxford University Press.
- Bengtsson, T. (2003). The need for looking back in time when predicting future mortality trends. In T. Bengtsson & N. Keilman (Eds.), *Perspectives on mortality forecasting* (Social Insurance Studies, No. 1: Current Practice) (pp. 87–97). Stockholm: Swedish National Insurance Board.
- Bengtsson, T. (2004). Mortality and social class in four Scanian Parishes, 1766–1865. In T. Bengtsson, C. Campbell, J. Z. Lee, et al. (Eds.), *Life under pressure. Mortality and living standards in Europe and Asia, 1700–1900* (pp. 135–172). Cambridge, MA: MIT Press.
- Bengtsson, T., & Keilman, N. (Eds.). (2003). *Perspectives on mortality forecasting* (Social Insurance Studies, No. 1: Current Practice). Stockholm: Swedish National Insurance Board.
- Bengtsson, T., & Lindström, M. (2000). Childhood misery and disease in later life: The effects on mortality in old age of hazards experienced in early life, southern Sweden, 1760–1894. *Population Studies*, 54, 263–277.
- Bengtsson, T., & Lindström, M. (2003). Airborne infectious diseases during infancy and mortality in later life in southern Sweden, 1766–1894. *International Journal of Epidemiology*, 32, 286–294.
- Bengtsson, T., Saito, O., Reher, D., & Campbell, C. (1998). Population and the economy: From hunger to modern economic growth. In C.-E. Núñez (Ed.), *Debates and controversies in economic history* (pp. 69–143). Madrid: Editorial de Estudios Ramon Areces, SA.
- Ben-Shlomo, Y., & Kuh, D. (2002). A life course approach to chronic disease epidemiology: Conceptual models, empirical challenges and interdisciplinary perspectives. *International Journal of Epidemiology*, 31, 285–293.
- Berenson, G. S., Srinivasan, S. R., Freedman, D. S., Radhakrishnamurthy, B., & Jr Dalferes, E. R. (1987). Atherosclerosis and its evolution in children. *American Journal of Medical Science*, 294, 429–440.
- Blackwell, D. L., Hayward, M. D., & Crimmins, E. M. (2001). Does childhood health affect chronic morbidity in later life? *Social Science and Medicine*, 52, 1269–1284.
- Cadwgan, A. M., Watson, W. A., Laing, R. B., MacKenzie, A. R., Smith, C. C., & Douglas, J. G. (2000). Presenting clinical features and C-reactive protein in the prediction of a positive stool culture in patients with diarrhoea. *Journal of Infection*, 41(2), 159–161.
- Chadwick, E. (1842). *Report of an enquiry into the sanitary conditions of the Labouring population of Great Britain*. London: Poor Law Commission.
- Crimmins, E. M., & Finch, C. E. (2006). Infection, inflammation, height and longevity. *Proceedings of the National Academy of Sciences of the United States of America*, 103(2), 498–503.
- Curtis, S. M. (2004). Nutrition and scarlet fever mortality during the epidemics of 1860–1890: The Sundsvall region. *Social History of Medicine*, 17(2), 199–221.
- Davey Smith, G. (2003). *Health inequalities: Lifecourse approaches*. Bristol: Policy Press.

- Davey Smith, G., & Kuh, D. (2001). William Ogilvy Kermack and the childhood origins of adult health and disease. *International Journal of Epidemiology*, 30, 696–703.
- Davey Smith, G., & Lynch, J. (2004). Commentary: Social capital, social epidemiology and disease aetiology. *International Journal of Epidemiology*, 33, 691–700.
- Davey Smith, G., & Lynch, J. (2005). Lifecourse influences on coronary heart disease. In M. Marmot & P. Elliott (Eds.), *Coronary heart disease epidemiology: From aetiology to public health*. Oxford: Oxford University Press.
- Davey Smith, G., Shipley, M. J., & Rose, G. (1990). Magnitude and causes of socioeconomic differentials in mortality: Further evidence from the Whitehall Study. *Journal of Epidemiology and Community Health*, 44, 265–270.
- Davey Smith, G., Leary, S., & Ness, A. (2006). Could dehydration in infancy lead to high blood pressure? *Journal of Epidemiology and Community Health*, 60, 142–143.
- Doblhammer, G. (1999). *Longevity and month of birth: Evidence from Austria and Denmark* (Demographic Research, Vol. 1(3)). Rostock: Max Planck Institute for Demographic Research.
- Doblhammer, G. (2002). *Differences in life-span by month of birth for the United States: The impact of early life events and conditions on late life mortality* (MPIDR Working Paper WP 2002–019). Rostock: Max-Planck Institute for Demographic Research.
- Doblhammer, G. (2008). The month of birth: Evidence for declining but persistent cohort effects in lifespan. In T. Bengtsson (Ed.), *Perspectives on mortality forecasting* (Social Insurance Studies, No. 5: Cohort factors: How conditions in early life influence mortality later in life). Stockholm: Swedish Social Insurance Agency.
- Doblhammer, G., & Vaupel, J. W. (2001). Lifespan depends on month of birth. *Proceedings of the National Academy of Sciences of the United States of America*, 98(5), 2934–2939.
- Finch, C. E., & Crimmins, E. M. (2004). Inflammatory exposure and historical changes in human life-spans. *Science*, 305, 1736–1739.
- Fogel, R. W. (1994). The relevance of Malthus for the study of mortality today: Long-run influences on health, mortality, labour-force participation and population growth. In K. Lindvall & H. Landberg (Eds.), *Population, economic development and the environment. The making of our common future* (pp. 231–284). Oxford/New York: Oxford University Press.
- Fridlizius, G. (1989). The deformation of cohorts. Nineteenth-century mortality in a generational perspective. *Scandinavian Economic History Review*, 37(3), 3–17.
- Galloway, P. R. (1987). *Population, prices and weather in preindustrial Europe*. Berkeley: University of California.
- Galobardes, B., Lynch, J. W., & Davey Smith, G. (2004). Childhood socioeconomic circumstances and cause-specific mortality in adulthood: Systematic review and interpretation. *Epidemiologic Reviews*, 26, 7–21.
- Galobardes, B., Davey Smith, G., Jeffreys, M., & McCarron, P. (2006a). Childhood socioeconomic circumstances predict specific causes of death in adulthood: The Glasgow student cohort study. *Journal of Epidemiology and Community Health*, 60, 527–529.
- Galobardes, B., Smith, G. D., & Lynch, J. W. (2006b). Systematic review of the influence of childhood socioeconomic circumstances on risk for cardiovascular disease in adulthood. *Annals of Epidemiology*, 16, 91–104.
- Hall, A. J., & Peckham, C. S. (1997). Infections in childhood and pregnancy as a cause of adult disease – Methods and examples. In M. G. Marmot & M. E. J Wadsworth (Eds.), *Fetal and early childhood environment: Long-term health implications*. *British Medical Bulletin*, 53(1), 10–23.
- Hallqvist, J., Lynch, J. W., Bartley, M., Lang, T., & Blane, D. (2004). Can we disentangle life course processes of accumulation, critical period and social mobility? *Social Science and Medicine*, 58(8), 1555–1562.
- Harvey, R. F., Spence, R. W., Lane, J. A., Nair, P., Murray, L. J., Harvey, I. M., & Donovan, J. (2002). Relationship between the birth cohort pattern of *Helicobacter pylori* infection and the epidemiology of duodenal ulcer. *Quarterly Journal of Medicine*, 95(8), 519–525.
- Heslop, P., Smith, G. D., Macleod, J., & Hart, C. (2001). The socioeconomic position of women, risk factors and mortality. *Social Science and Medicine*, 53, 477–485.

- Hollingsworth, T. H. (1957). A demographic study of the British ducal families. *Population Studies*, 11(1), 4–26.
- Hollingsworth, T. H. (1977). A note on the medieval longevity of the secular peerage. *Population Studies*, 29(1), 157.
- Hunter, R. J., Patel, V. B., Miell, J. P., Wong, H. J., Marway, J. S., Richardson, P. J., & Preedy, V. R. (2001). Diarrhea reduces the rates of cardiac protein synthesis in myofibrillar protein fractions in rats in vivo. *Journal of Nutrition*, 131, 1513–1519.
- Kannisto, V., Christensen, K., & Vaupel, J. W. (1997). No increased mortality in later life for cohorts born during famine. *American Journal of Epidemiology*, 145(11), 987–994.
- Kaplan, G., & Keil, J. E. (1993). Socioeconomic factors and cardiovascular disease: A systematic review of the literature. *Circulation*, 88, 1973–1998.
- Kermack, W. O., McKendrick, A. G., & McKinlay, P. L. (1934, March). Death rates in Great Britain and Sweden: Some regularities and their significance. *Lancet*, 223, 698–703.
- King, S. (1997). Dying with style: Infant death and its context in a rural industrial township 1650–1830. *Social History of Medicine*, 10(1), 3–24.
- Krantz, G., & Östergren, P. O. (2004). Does it make sense in a coherent way? Determinants of sense of coherence in Swedish women 40 to 50 years of age. *International Journal of Behavioral Medicine*, 11(1), 18–26.
- Kuh, D., & Ben-Shlomo, Y. (1997). *A life course approach to chronic disease epidemiology. Tracing the origins of ill-health from early to adult life*. Oxford: Oxford University Press.
- Kuh, D., & Ben-Shlomo, Y. (2004). *A life course approach to chronic disease epidemiology. Tracing the origins of ill-health from early to adult life* (2nd ed.). Oxford: Oxford University Press.
- Kuh, D., & Hardy, R. (2002). *A life course approach to Women's health*. Oxford: Oxford University Press.
- Lawlor, D. A., Davey Smith, G., Mitchell, R., & Ebrahim, S. (2006). Adult blood pressure and climate conditions in infancy: A test of the hypothesis that dehydration in infancy is associated with higher adult blood pressure. *American Journal of Epidemiology*, 163, 608–614.
- Lawn, S. D., Obeng, J., Acheampong, J. W., & Griffin, G. E. (2000). Resolution of the acute phase response in West African patients receiving treatment for pulmonary tuberculosis. *International Journal of Tuberculosis and Lung Disorders*, 4(4), 340–344.
- Li, X., Kolltveit, K. M., Tronstad, L., & Olsen, I. (2000). Systemic diseases caused by oral infection. *Clinical Microbiology Review*, 13(4), 547–558.
- Li, S., Chen, W., Srinivasan, R. S. R., Bond, M. G., Tang, R., Urbina, E. M., & Berenson, G. S. (2003). Childhood cardiovascular risk factors and carotid vascular changes in adulthood: The Bogalusa Heart Study. *Journal of the American Medical Association*, 290, 2271–2276.
- Livi-Bacci, M. (1991). *Population and nutrition. An essay on European demographic history*. Cambridge: Cambridge University Press.
- Macintyre, S. (1998). Social inequalities and health in the contemporary world: Comparative overview. In S. Strickland & P. S. Shetty (Eds.), *Human biology and social inequality* (Symposia of the Society for the Study of Human Biology). Cambridge: Cambridge University Press.
- Marmot, M. (2004). *Status syndrome. How your social standing directly affects your health*. London: Bloomsbury.
- Marmot, M., & Elliot, P. (Eds.). (2005). *Coronary heart disease epidemiology. From Aetiology to public health* (2nd ed.). Oxford: Oxford University Press.
- Modin, B. (2002). Birth order and mortality: A life-long follow-up of 14,200 boys and girls born in early 20th century Sweden. *Social Science and Medicine*, 54(7), 1051–1064.
- Moore, S. E., Cole, T. J., Poskitt, E. M. E., Sonko, B. J., Whitehead, R. G., McGregor, I. A., & Prentice, A. M. (1997). Season of birth predicts mortality in rural Gambia. *Nature*, 388, 434.
- Moore, S. E., Fulford, A. J. C., Streatfield, P. K., Persson, L. Å., & Prentice, A. M. (2004). Comparative analysis of patterns of survival by season of birth in rural Bangladeshi and Gambian populations. *International Journal of Epidemiology*, 33, 137–143.

- National Public Health Report.* (2001). Stockholm: National Board on Health and Welfare.
- National Public Health Report.* (2005). Stockholm: National Board on Health and Welfare.
- Novak, M., Ahlgren, C., & Hammarström, A. (2006). A life-course approach in explaining social inequity in obesity among young adult men and women. *International Journal of Obesity*, 30 (1), 191–200.
- Omran, A. R. (1971). The epidemiologic transition. A theory of the epidemiology of population change. *The Milbank Memorial Fund Quarterly*, 49, 509–538.
- Perrenoud, A. (1975). L'inégalité sociale devant la mort à Genève au XVIIIème siècle. *Population*, 30(Special number), 239.
- Preston, S. H., Hill, M. E., & Drevenstedt, G. L. (1998). Childhood conditions that predict survival to advanced ages among African-Americans. *Social Science and Medicine*, 47, 1231–1246.
- Raitakari, O. T., Juonala, M., Kähönen, M., Taittonen, L., Laitinen, T., Mäki-Torkko, N., Järvisalo, M. J., Uhari, M., Jokinen, E., Rönnemaa, T., Åkerblom, H. K., & Viikari, J. S. A. (2003). Cardiovascular risk factors in childhood and carotid artery intima-media thickness in adulthood: The cardiovascular risk in young Finns Study. *Journal of the American Medical Association*, 290, 2277–2283.
- Riley, J. C. (2001). *Rising life expectancy. A global history.* Cambridge: Cambridge University Press.
- Rosvall, M., Östergren, P. O., Hedblad, B., Isacsson, S. O., Janzon, L., & Berglund, G. (2002). Life-course perspective on socioeconomic differences in carotid atherosclerosis. *Arteriosclerosis, Thrombosis and Vascular Biology*, 22(10), 1704–1711.
- Rosvall, M., Chaix, B., Lynch, J. W., Lindström, M., & Merlo, J. (2006). Similar support for three different life course socioeconomic models on predicting premature cardiovascular mortality and all-cause mortality. *BioMed Central Public Health*, 6, 203.
- Rotberg, R. I., & Rabb, T. K. (1985). The relationship of nutrition, disease, and social conditions: A graphical presentation. In R. I. Rotberg & T. K. Rabb (Eds.), *Hunger and history. The impact of changing food patterns and society.* Cambridge/London: Cambridge University Press.
- Shaheen, S. (1997). The beginnings of chronic airflow obstruction. In: M. G. Marmot & M. E. J. Wadsworth (Eds.), Fetal and early childhood environment: Long-term health implications. *British Medical Bulletin*, 53(1), 58–70.
- Simondon, K. B., Elguero, E., Marra, A., Diallo, A., Aaby, P., & Simondon, F. (2004). Season of birth is not associated with risk of early adult death in rural Senegal. *International Journal of Epidemiology*, 33, 130–136.
- Sköld, P. (1996a). *The two faces of smallpox: A disease and its prevention in eighteenth and nineteenth century Sweden.* Thesis, Umeå University, Umeå.
- Sköld, P. (1996b). From inoculation to vaccination: Smallpox in Sweden in the eighteenth and nineteenth centuries. *Population Studies*, 50, 247–262.
- Stenberg, S. A., Vägerö, D., Österman, R., Arvidsson, E., von Otter, C., & Janson, C. G. (2007). Stockholm Birth Cohort Study 1953–2003: A new tool for life-course studies. *Scandinavian Journal of Public Health*, 35(1), 104–110.
- Strachan, D. P. (1997). Respiratory and allergic diseases. In D. Kuh & Y. Ben-Shlomo (Eds.), *A life course approach to chronic disease Epidemiology. Tracing the origins of ill-health from early to adult life* (pp. 101–120). Oxford: Oxford Medical Publications.
- Strong, J. P., Malcom, G. T., McMahan, C. A., Tracy, R. E., Newman, W. P., III, Herdrick, E. E., & Cornhill, J. F. (1999). Prevalence and extent of atherosclerosis in adolescents and young adults: Implications for prevention from the pathobiological determinants of atherosclerosis in Youth Study. *Journal of the American Medical Association*, 281, 727–735.
- Suomi, S. J. (1997). Early determinants of behaviour: Evidence from primate studies. In M. E. Marmot & M. E. J. Wadsworth (Eds.), Fetal and early childhood environment: Long-term health implications. *British Medical Bulletin*, 53(1), 170–184.
- Timpson, N. J., Lawlor, D. A., Harbord, R. M., Gaunt, T. R., Day, I. N. M., Palmer, L. J., Hattersley, A. T., Ebrahim, S., Lowe, G. D. O., Rumley, A., & Davey Smith, G. (2005).

- C-reactive protein and its role in metabolic syndrome: Mendelian randomisation study. *Lancet*, 366, 1954–1959.
- United Nations. (1999). *World population prospects*. The 1998 revision. Volume I: Comprehensive Tables, ST/SER.A/177. New York: United Nations.
- von Mutius, E. (2001). Paediatric origins of adult lung disease. *Thorax*, 56, 153–156.
- Wamala, S., Lynch, J. W., & Kaplan, G. A. (2001). Women's exposure to early and later life socioeconomic disadvantage and coronary heart disease risk: The Stockholm Female Coronary Risk Study. *International Journal of Epidemiology*, 30(2), 275–284.
- Willner, S. (1999). *Det svaga könet? Kön och vuxendödlighet i 1800-talets Sverige*. Linköping: Linköping University.
- Wohl, A. S. (1983). *Endangered lives: Public health in Victorian Britain*. London: JM Dent.
- Wrigley, E. A., & Schofield, R. (1981). *The population history of England 1541–1871: A reconstruction*. Cambridge: Cambridge University Press.
- Wrigley, E. A., Davies, R. S., Oeppen, J. E., & Schofield, R. S. (1997). *English population history from family reconstitution 1580–1837*. Cambridge: Cambridge University Press.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 22

Early Life Events and Later Life Health: Twin and Famine Studies



Kaare Christensen

22.1 Introduction

During more than a decade, the relation between early life conditions and late life health has been one of the major topics in the epidemiological literature. The interest for the connection between early life conditions and later life health is by no means new. As early as in the sixteenth century, Francis Bacon suggested that the nutrition in the womb and the first year of life is very important for later life health. In more recent times, especially the classical paper from 1934 by Kermack et al., and Forsdahl's work in the 1970s (Forsdahl 1977), have been important contributions. However, since the early 1990s, one of the main advocates for the importance of early life conditions for later life health has been David Barker and his group in Southampton. They revitalized the so-called "fetal origins hypothesis" and have produced an impressive series of papers on the topic. The studies were motivated by this intriguing hypothesis proposed by Barker and co-workers which asserts that a baby's nourishment before and during infancy programmes its susceptibility to cardiovascular diseases as well as several other diseases and adverse outcomes, ranging from diabetes mellitus to cancer and suicide. There is evidence that an association exists between fetal growth and later life health outcomes such as blood pressure and cardiovascular mortality. The key question is, however, whether it is fetal nourishment or other factors such as genes or socioeconomic conditions that cause the association. Some studies suggest that socioeconomic confounding cannot explain the association between fetal growth and cardiovascular mortality (Leon et al. 1998), but they are few, and even fewer studies have evaluated the influence of genetic confounding (Figs. 22.1 and 22.2).

A major concern about the work on the fetal origins hypothesis has been the wide range of exposure proxies and outcomes. Among the exposure proxies measured are

K. Christensen (✉)

Institute of Public Health, University of Southern Denmark, Odense, Denmark
e-mail: KChristensen@health.sdu.dk

Fig. 22.1 The fetal origins hypothesis states that the observed negative correlation between birth weight and cardiovascular disease risk is due to early life nutrition

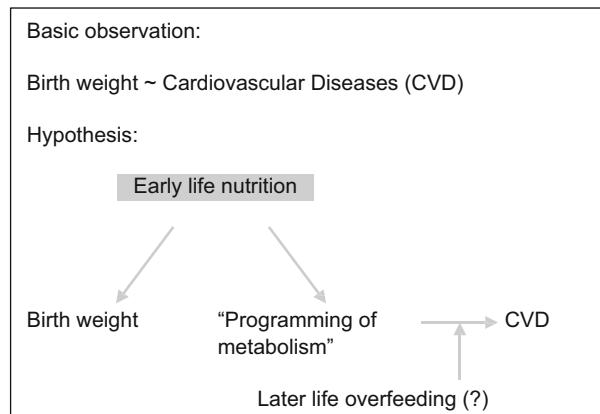
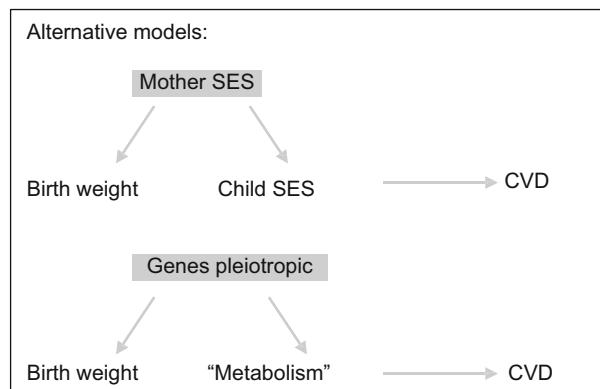


Fig. 22.2 Alternative explanations for the observed negative correlation between birth weight and cardiovascular disease risk



birth weight, birth length, ponderal index, abdominal circumference, and corresponding measurements have been performed at age 1 year. The measurements have been related to outcomes such as survival, cardiovascular diseases, hypertension, diabetes, suicide, and even to getting married (Barker 1994, 1998). Furthermore, these relations have been investigated in hundreds of data sets and very often in subtypes like obese mothers or children with rapid catch-up growth etc. This increases the risk of type 1-errors, i.e. statistically significant random findings, considerably, and has therefore created a certain amount of scepticism (Kramer and Joseph 1996). To avoid some of these problems, in our group at the University of Southern Denmark, we have focused on famine and twin studies which are dealing with extreme conditions in early life, and we have examined how these conditions affect later life health.

22.2 Famine Early in Life and Later Life Health

During the late 1990s, three reports have addressed the influence of prenatal exposure to famine on health in later life. We studied 161,744 individuals born during the 1866–1868 Finnish famine and found, on the basis of a comparison with more than 600,000 individuals born before and after the famine, that nutritional deprivation in utero has no effect on survival in adult life (Kannisto et al. 1997). Stanner et al. (1997) investigated a broad range of coronary heart disease and diabetes mellitus risk factors among 169 people exposed to malnutrition in utero during the siege of Leningrad in 1941–1942 and nearly 400 born before or outside the area of the siege. They found no association between intrauterine malnutrition and glucose intolerance, dyslipidaemia, hypertension, or cardiovascular disease in adulthood. Finally, Ravelli et al. (1998), studied 279 individuals who were exposed to malnutrition in utero during the Dutch hunger winter 1944–1945 and nearly 425 controls born before and after. They found an association between intrauterine exposure to famine and decreased glucose tolerance in adults aged around 50 years. However, it was disappointing that Ravelli and colleagues, among all the suggested outcomes related to fetal nourishment, initially only reported glucose tolerance. Later reports showed that, for example, blood pressure showed no association with blood pressure in the Dutch famine study (Roseboom et al. 1999).

22.3 Later Life Health for Twins

Twin studies are especially interesting because twins experience severe growth retardation in the uterus, especially in the third trimester and, furthermore, studies of twins can control for the influence of the mother's socioeconomic status and for the effect of genes. Twins experience considerable retardation in intrauterine growth – for example, they are on average more than 900 g lighter than single children at birth (MacGillivray et al. 1988; Kline et al. 1989). Phillips (1993) has argued that even small birth weight differences in twins could reflect important differences in intrauterine conditions important for programming of diseases later in life, because the mean birth weight among twins is already considerably lower than in singletons. Therefore, birth weight differences in twins offer a unique opportunity to test the “fetal origins hypothesis”.

This also raises a question important to twin researchers: Does the reduced growth pattern in the last trimester make twins more vulnerable in adult life with an increased risk of cardiovascular diseases, diabetes mellitus and other assumingly “programmed” diseases? If so, then twin studies may be a poor model for studying these diseases because the causal field of the diseases could be different from that of singletons.

If twins are “programmed” due to the considerable growth retardation during the third trimester, one could expect an increased mortality and especially an increased cardiovascular mortality in adulthood for twins compared to the general population. Both a Swedish and our earlier Danish twin study found similar mortality patterns among twins and singletons in adulthood (Vågerö and Leon 1994; Christensen et al. 1995). However, the Swedish twin study (Vågerö and Leon 1994) had a limited follow-up period with no distinction between monozygotic (MZ) and dizygotic (DZ) twins (the latter having slightly higher birth weights on average), and our Danish twin study (Christensen et al. 1995) did not include causes of death.

Therefore we studied cause-specific mortality of 19,986 Danish twin individuals from the birth cohorts 1870–1930 followed from 1952 through 1993 (Christensen et al. 2001). Despite the large sample size and follow-up period, we were not able to detect any difference between twins and the general population with regard to all-cause mortality or cardiovascular mortality. Hence, the intrauterine growth retardation experienced by twins does not result in any “fetal programming” of cardiovascular diseases.

22.4 Twins and Genetic Confounding

The potential for genetic confounding in relation to the fetal origins hypothesis has been illustrated by Dunger et al. (1998), who showed that variation in the insulin gene (INS VNTR) is associated with fetal growth. Based on studies of fetal insulin secretion and monogenic diseases, Hattersley and Tooke (1999) proposed that genetically determined insulin resistance contributes substantially to the association of low birth weight with diabetes, hypertension and vascular diseases and named this hypothesis “the fetal insulin hypothesis”.

We used the Minnesota Twin Family study (Iacono et al. 1999) to test the potential influence of genetic confounding on the association between birth weight and systolic blood pressure, which is the best documented association between fetal growth and later life health outcome (Barker 1998; Kuh and Ben-Shlomo 1997; Law and Shiell 1996; Taylor et al. 1997). The effect of genetic confounding was evaluated by analysing individual twin data as well as intrapair differences in birth weight and systolic blood pressure. This approach enables controlling for the effect of all genetic factors in monozygotic pairs and on average half of the genetic factors in dizygotic pairs as well as environmental maternal effects. Two recent twin studies (Poulter et al. 1999; Dwyer et al. 1999) using a similar design did not find evidence for a genetic component to the association, but as pointed out in the accompanying editorial (Leon 1999), the number of monozygotic twin pairs in these studies was sparse.

We used the following statistical analysis:

As in Hopper and Seeman (1994), for each twin, i , of a pair ($i = 1,2$) let Y_i be systolic blood pressure and X_{1i} = birth weight and X_{2i} = current weight.

Let

$$Y_i = a_0 + a_1 X_{1i} + a_2 X_{2i} + E_i \quad (22.1)$$

where E_i represents measurement error and effects specific to twin i . Each of the coefficients a_1 and a_2 represents the strength of a linear association between the blood pressure and a corresponding variable. The intrapair difference is

$$D = Y_1 - Y_2 = a_1 D_1 + a_2 D_2 + E \quad (22.2)$$

where $D_j = X_{j1} - X_{j2}$ ($j = 1, 2$) and $E = E_1 - E_2$. From (22.2) it can be seen that the same coefficients a_1 , a_2 can be estimated by regressing D against D_1 , D_2 , and constraining the fitted line to pass through the origin (because (22.2) does not have an intercept term). This second regression approach controls for age, sex and genetic factors (all in monozygotic twins and on average half in dizygotic twins).

From the Minnesota Twin Family Study (Iacono et al. 1999) we included 1311 pairs of adolescent twins, and we found a negative association between birth weight and systolic blood pressure in the overall sample. The regression coefficient after controlling for current weight was -1.88 mm Hg/kg (SE 0.61), which corresponds to results from previous studies of singleton adolescents. The regression coefficient fell to -0.64 mm Hg/kg (SE 0.86) when the intrapair analyses were used. The largest reduction was observed among monozygotic twins: from -2.44 mm Hg/kg (SE 0.75) in the overall monozygotic twin sample to -1.06 mm Hg/kg (SE 1.14) in the analyses of the within monozygotic pair differences.

22.5 Overview

In 2002, Huxley and co-workers published an overview of the available data on the relation between birth weight and later life blood pressure. This relation has been put forward as one of the most consistent and strong evidence of the association between birth weight and later life health. This overview found a clear indication of publication bias as smaller studies reported a large effect while larger studies reported a small effect. Furthermore, it was somewhat disturbing to see that the hypothesis-generating group consistently reported a larger effect than other research groups.

Huxley et al. (2002) also summarized the available interpair comparison studies in twins and found, based on the above-mentioned study of the Minnesota twins and on other studies, that the confidence interval for the effect estimated by intrapair differences in twins was not different from zero. Even if the largest estimate of the relation between birth weight and blood pressure is accepted, that is to say a 2 mm Hg increase in blood pressure for every increase in birth weight by 1 kilo, this is clearly not an important factor on the individual level. Theoretically, it might be an important finding for understanding the etiology of blood pressure, but from a public

health perspective, the change of 2 mm Hg per 1 kilo difference in birth weight would not call for intervention. This view is underlined by the observation by Huxley and colleagues who draw attention to the fact that increase in birth weight also correlates with an increase in weight which might have an even stronger effect on the adult blood pressure than the inverse relation between birth weight and blood pressure, when current weight is controlled for. At present, therefore, too little is known about the effect, and the evidence of the correlation between early life growth and later life health and survival is not strong enough to have any practical importance for forecasting models.

References

- Barker, D. J. P. (1994). *Mothers, babies, and disease in later life*. London: British Medical Journal Publishing Group.
- Barker, D. J. P. (1998). *Mothers, babies and health in later life*. Edinburgh: Churchill Livingstone.
- Christensen, K., Vaupel, J. W., Holm, N. V., & Yashin, A. I. (1995). Twin mortality after age 6: Fetal origins hypothesis versus twin method. *British Medical Journal*, 310, 432–436.
- Christensen, K., Wienke, A., Skytthe, A., Holm, N. V., Vaupel, J. W., & Yashin, A. I. (2001). Cardiovascular mortality in twins and the fetal origins hypothesis. *Twin Research*, 4, 344–349.
- Dunger, D. B., Ong, K. K., Huxtable, S. J., Sherriff, A., Woods, K. A., Ahmed, M. L., Golding, J., Pembrey, M. E., Ring, S., Bennet, S. T., & Todd, J. A. (1998). Association of the INS VNTR with size at birth. ALSPAC Study Team. Avon longitudinal study of pregnancy and childhood. *Nature Genetics*, 19(1), 98–100.
- Dwyer, T., Blizzard, L., Morley, R., & Ponsonby, A. L. (1999). Within pair association between birth weight and blood pressure at age 8 in twins from a cohort study. *British Medical Journal*, 319(7221), 1325–1329.
- Forsdahl, A. (1977). Are poor living conditions in childhood and adolescence an important risk factor for arteriosclerotic heart disease? *British Journal of Preventive Social Medicine*, 31, 91–95.
- Hattersley, A. T., & Tooke, J. E. (1999). The fetal insulin hypothesis: An alternative explanation of the association of low birthweight with diabetes and vascular disease. *Lancet*, 353, 1789–1792.
- Hopper, J. L., & Seeman, E. (1994). The bone density of female twins discordant for tobacco use. *New England Journal of Medicine*, 330, 387–392.
- Huxley, R., Neil, A., & Collins, R. (2002). Unravelling the fetal origins hypothesis: Is there really an inverse association between birthweight and subsequent blood pressure? *Lancet*, 360(9334), 659–665.
- Iacono, W. G., Carlson, S. R., Taylor, J., Elkins, I. J., & Macque, M. (1999). Behavioral disinhibition and the development of substance use disorders: Findings from the Minnesota Twin Family Study. *Development and Psychopathology*, 11, 869–900.
- Kannisto, V., Christensen, K., & Vaupel, J. W. (1997). No increased mortality in later life for cohorts born during famine. *American Journal of Epidemiology*, 145, 987–994.
- Kermack, W. O., McKendrick, A. G., & McKinlay, P. L. (1934). Death rates in Great Britain and Sweden: Some general regularities and their significance. *Lancet*, 226, 698–703.
- Kline, J., Stein, Z., & Susser, M. (1989). *Conception to birth*. New York: Oxford University Press.
- Kramer, M. S., & Joseph, K. S. (1996). Enigma of fetal/infant origins hypothesis. *Lancet*, 348, 1254–1255.
- Kuh, D., & Ben-Shlomo, Y. (1997). *Life course approach to chronic disease epidemiology. Tracing the origins of ill-health from early to adult life*. Oxford: Oxford University Press.

- Law, C. M., & Shiell, A. W. (1996). Is blood pressure inversely related to birth weight? The strength of evidence from a systematic review of the literature. *Journal of Hypertension*, 14, 935–941.
- Leon, D. A. (1999). Twins and fetal programming of blood pressure. *British Medical Journal*, 31, 1313–1314.
- Leon, D. A., Lithell, H. O., Vägerö, D., Koupilova, I., Mohsen, R., Berglund, L., Lithell, U.-B., & McKeigue, P. M. (1998). Reduced fetal growth rate and increased risk of death from ischaemic heart disease: Cohort study of 15,000 Swedish men and women born 1915–29. *British Medical Journal*, 317, 241–245.
- MacGillivray, I., Campbell, D. M., & Thompson, B. (Eds.). (1988). *Twinning and twins*. London: Wiley.
- Phillips, D. I. (1993). Twin studies in medical research: Can they tell us whether diseases are genetically determined? *Lancet*, 341, 1008–1009.
- Poulter, N. R., Chang, C. L., MacGregor, A. J., Schneider, H., & Spector, T. D. (1999). Association between birth weight and adult blood pressure in twins: Historical cohort study. *British Medical Journal*, 319, 1330–1333.
- Ravelli, A. C. J., van der Meulen, J. H. P., Michels, R. P. J., Osmond, C., Barker, D. J. P., Hales, C. N., & Bleker, O. P. (1998). Glucose tolerance in adults after prenatal exposure to famine. *Lancet*, 351, 173–177.
- Roseboom, T. J., van der Meulen, J. H. P., Ravelli, A. J., van Montfrans, G. A., Osmond, C., Barker, D. J. P., & Bleker, O. (1999). Blood pressure in adults after prenatal exposure to famine. *Journal of Hypertension*, 17(3), 325–330.
- Stanner, S. A., Bulmer, K., Andrès, C., Lantseva, O. E., Borodina, V., Poteen, V. V., & Yudkin, J. S. (1997). Does malnutrition in utero determine diabetes and coronary heart disease in adulthood? Results from the Leningrad siege study, a cross-sectional study. *British Medical Journal*, 315, 1342–1349.
- Taylor, S. J., Whincup, P. H., Cook, D. G., Papacosta, O., & Walker, M. (1997). Size at birth and blood pressure: Cross sectional study in 8–11 year old children. *British Medical Journal*, 314, 475–480.
- Vägerö, D., & Leon, D. (1994). Ischaemic heart disease and low birth weight: A test of the fetal origins hypothesis from the Swedish Twin Registry. *Lancet*, 343, 260–263.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 23

The Month of Birth: Evidence for Declining but Persistent Cohort Effects in Lifespan



Gabriele Doblhammer

23.1 Introduction

In the second half of the nineteenth century, mortality started to decline and this decline seemed to follow a cohort pattern. A cohort pattern is direct evidence for the effect of early-life circumstances on adult mortality: improvements in the living environment early in life lead to reduced mortality during the whole life course. Early studies for England, Wales and Scotland (Kermack et al. 1934, 2001) suggest that, as far as mortality up to the year 1925 is concerned, the year of birth had more predictive power than the year of death. Around 1925, the responsible factors for the mortality decline changed and the year of birth lost its predictive power. Period factors became more and more important. In a recent study, Davey Smith and Kuh (2001) updated the table of relative death rates for England and Wales from the Kermack paper. They showed that from 1925 onwards at younger ages, death rates fell faster than predicted based on birth cohort regularities, whereas at older ages, mortality declined at a much slower rate than predicted.

Elo and Preston (1992) point out that the early studies were probably so successful in demonstrating cohort effects because they were based on mortality data from a time period before the process of mortality decline on a period level had begun. Once both cohorts and periods began to show mortality improvements, it became more difficult to separate the two.

The cohort decline at the beginning of the epidemiological transition is not without contest either. In their study of the Swedish mortality decline, Kermack et al. (1934, 2001) admitted that the cohort effect is not as clear as it was for England and Wales. The authors argue that a rectangular block dating from 1855 onwards and

G. Doblhammer (✉)

University of Rostock, Rostock, Germany

Rostock Centre for the Study of Demographic Change, Rostock, Germany
e-mail: doblhammer@rostockerzentru.de

affecting the age groups 10 to 30 years was the primary cause of disturbances in the cohort decline. If the block is omitted, a cohort decline is observed. This result could not be replicated, however, in a study of Swedish mortality between 1778 and 1993 by Vaupel et al. (1997). The conclusion according to Vaupel et al. is that “the pattern is clearly more complex than a pure cohort-effect model would suggest” (Vaupel et al. 1997: 63).

As a result of the loss of the predictive power of cohort factors around 1925, the emphasis turned towards period factors such as advances in medical technology, life-style, smoking, physical activity and diet. In the 1970s, however, Forsdahl (1973, 1977, 1978) observed that regional differences in adult lung cancer and heart disease were not related to contemporary differences in lifestyle, smoking behaviour, or socioeconomic status but rather to differences in regional infant mortality during childhood and youth of the cohorts under study. His study is now considered the starting point of a large and productive area of research that tries to link early-life conditions to the manifestation of chronic disease later in life. The discussion about early-life effects on health at adult ages gained momentum with studies conducted by the Southampton group of Barker and colleagues (Barker 1994; Barker and Osmond 1986a, b, 1987). The group developed the fetal-origins hypothesis of adult disease (also known as the ‘Barker hypothesis’), which suggests that coronary heart disease at adult ages results from poor conditions in utero caused by inadequate nutrition on the part of the mother and infectious diseases she suffered during pregnancy. Since inadequate nutrition of the fetus is reflected in low birth weight, the Barker hypothesis claims that growth retardation in utero leads to low birth weight and to an increased risk of chronic disease later in life. It seems that the main connection between birth weight and heart disease later in life is systolic blood pressure – infants with a low birth weight experience increased systolic blood pressure at adult ages.

The fetal-origins hypothesis has led to a large amount of research that generally concludes that low birth weight is associated with an increased risk of heart disease at adult ages and that low-birth weight infants suffer from increased systolic blood pressure later in life. The interpretation of these outcomes has been repeatedly challenged, however. The main idea underlying the fetal-origins hypothesis is that a critical period exists early in life and that negative effects during this period cannot be reversed later in life. Critics of the hypothesis frequently bring forward the argument that birth weight is confounded with socioeconomic status. Negative social factors in the early-life environment may set people onto life trajectories that negatively affect their health over the whole life course. Therefore, the almost universally observed relationship between birth weight and the risk of chronic disease later in life may be an outcome of the whole life course rather than the result of a critical period early in life (Joseph and Kramer 1996; Kramer 2000).

This criticism leads to the question whether one can find an indicator for the prenatal and early postnatal environment that is not related to the life-course. Birth weight certainly does not fulfil this criterion and, in addition, it is not widely available. Studies that use birth weight or other direct indicators of early-life circumstances are usually based on hospital data, which are invariably subject to selection bias. Moreover, their sample sizes tend to be modest.

The fetal-origins hypothesis suggests that nutrition and infectious diseases during the pregnancy of the mother are responsible for growth retardation in the infant, which leads to an increased risk of heart disease at adult ages. Both nutrition and infectious diseases are highly seasonal: respiratory infections peak in the autumn and winter, and gastrointestinal infections peak during warm periods of the summer months. The availability of fresh fruits and vegetables – and thus of micronutrients – tends to change according to the seasons of the year. An indicator that reflects the seasonally changing environment during the prenatal and early postnatal period is month of birth.

Epidemiological research on the underlying factors of schizophrenia has long used month of birth as an indicator for early-life circumstances that affect the risk of schizophrenia later in life. This line of research dates back to Ellsworth Huntington, who in 1938 published his book about seasonality (Huntington 1938), in which he describes the relationship between the seasons of the year and social, psychological, and demographic phenomena. By 1997, more than 250 studies about the month-of-birth effect in schizophrenia had appeared and many more are still being published (Torrey et al. 1997). Most of the research on the relationship between month of birth and the incidence of diseases has been conducted for mental disorders, in particular schizophrenia and bipolar disorders. The season-of-birth effect has also been studied for autistic disorder, Alzheimer patients, anorexia nervosa patients, and for diseases of the nervous system such as Parkinson's disease, multiple sclerosis and epilepsy. Recently much attention was for example given to the month-of-birth effect in insulin dependent childhood diabetes. For a review of studies about the relationship between month of birth and certain diseases, see Doblhammer (2004). Many of these studies suggest that virus infections in utero or in the first few months of life are responsible for the increased risk of developing a certain disease. None of these studies, however, provides concrete evidence for a specific causal mechanism.

Although widespread evidence exists concerning the month-of-birth effect for certain diseases, little attention has been given to the question whether there is a correlation between the month of birth and lifespan and whether this relationship has changed over cohorts. If the month-of-birth effect decreases in younger cohorts then this suggests that the influence of the very first period of life on adult lifespan has reduced over time. Two recently published studies point in this direction. A study of young adults in rural Gambia found a significant difference in survival to age 45 by month of birth (Moore et al. 1997). When the authors repeated their study on a rural Bangladeshi population, they could not detect an effect of season of birth on survival (Moore et al. 2004). Furthermore, a study of early adult death in rural Senegal also failed to find any influence between month of birth and survival at young adult ages (Simondon et al. 2004). Both Moore and colleagues and Simondon and colleagues suggest that the negative findings in the Bangladeshi and Senegal population are due to cohort effects. While the data for Gambia are based on births during the 1950s and 1960s, the Bangladeshi and Senegal data include individuals born between 1974 and 2000 (Bangladesh), and 1962 and 2000 (Senegal).

This article takes up the question whether the season-of-birth effect in life-span has changed over time and whether it is less important for more recent than for older

cohorts. To answer this question the article first describes the month-of-birth pattern in lifespan for selected countries of the Northern (Austria, Denmark, and United States) and Southern Hemisphere (Australia). Second, a cohort analysis is performed based on the Danish register data and of consecutive US census rounds that include information about the season of birth. Third, the underlying causal mechanisms behind the month-of-birth effect are reviewed and finally, outcomes are discussed in the light of mortality forecasting.

23.2 Data

The optimal data to test for differences in lifespan by season of birth are longitudinal data. Birth cohorts born in a specific season are followed from birth to death and life expectancy can be calculated using simple life-table methods. Such data rarely exist however. The data that are closest to this requirement are register data from the Scandinavian countries. The Danish data used in this study consist of a mortality follow-up of all Danes who were at least 50 years old on 1 April 1968. This totals 1,371,003 people, who were followed up to week 32 of 1998. The study excludes 1994 people who were lost to the registry during the observation period. Among those who are included in the study, 86% (1,176,383 individuals) died before week 32 of 1998; 14% (192,626 individuals) were still alive at the end of the follow-up.

Population registers do not exist for Austria, Australia and the US, where only individual death records are available. Exact dates of birth and death are known for in all 681,677 Austrians who died between 1988 and 1996 and for 219,820 native-born Australians who died between 1993 and 1997 at ages 50+. Two data sources are used for the United States. First, US death records for the years 1989 to 1997, which include place of birth, are the basis for the analysis of the month-of-birth pattern in lifespan. Second, the three US census rounds 1960, 1970, and 1980 are used to study cohort patterns. These three census rounds are the only rounds that include information about the quarter of birth. Data are extracted from the “Public Use Microdata Samples”, which are accessible under <http://usa.ipums.org/usa/>. The extract is restricted to the native-born white US population aged 0 to 100. For whites, this gives a sample size of 1,490,444 in 1960, 1,672,107 in 1970, and 1,812,839 in 1980.

23.3 Methods

For Denmark, both the risk population and the number of deaths are known which means that it is possible to estimate remaining life expectancy at age 50 based on life tables that were corrected for left truncation. This was achieved by calculating occurrence and exposure matrices that take into account an individual’s age on 1 April 1968. For example, a person who was 70 at the beginning of the study and who died at age 80 enters the exposures for ages 70 to 80 but is not included in the

exposures for ages 50 to 69. The central age-specific death rate is based on the occurrence-exposure matrix. The corresponding life-table death rate is derived by means of the Greville Method (Greville 1943).

For Austria, Australia and the United States, the population at risk is unknown, which means that lifespan by month of birth cannot be estimated based on simple life-table techniques. For these three countries, remaining lifespan at age 50 was therefore estimated by calculating the average of the exact ages at death. It has been pointed out that using mean age at death as an approximation for life expectancy may lead to serious bias in the observed month-of-birth pattern (Gavrilov and Gavrilova 2003). It is well known that mean age at death does not correctly estimate life expectancy in non-stationary and non-extinct populations. The emphasis of this study, however, is not life expectancy per se but the month-of-birth pattern in life expectancy, which generally should not be affected. There is one exception, however. If the seasonal distribution of births has changed over time then not only life expectancy but also the month-of-birth pattern is biased when estimated based on mean ages at death. If more people proportionally are born in spring in younger cohorts than in older cohorts, then for a given time period the mean age at death will be biased downward for those born in spring. In the case of Austria and Denmark, the changes in the seasonal distribution of births over time are minor, however, and the effect on the month-of-birth pattern is negligible (Doblhammer 2004).

The Danish register data consist of sufficiently large numbers of exposures and deaths to distinguish between age and cohort effects. Two ten-year birth cohorts are followed over an age-span of 20 years. The birth cohort April 1908 to March 1918 enters the study period between the ages of 50 and 59 and 11 months. They are followed from age 60 to age 79 in order, theoretically, to allow each member of the cohort to reach each age. The second cohort is aged 60 to 69 and 11 months at the 1968-baseline and is followed from age 70 to age 89. This specification allows the study of age-specific death rates at ages 70 to 79 for both cohorts. Conditional on surviving to age 70, all individuals are followed from age 70 to age 79. Those who survived age 79 are treated as censored. Mortality of the two cohorts between ages 70 and 79 is modelled by a proportional hazard model with the baseline hazard following a Gompertz function,

$$\mu(x|y_i) = ae^{b_0x}e^{b_1y_1+b_2y_2+b_3y_3} \quad (23.1)$$

where $\mu(x|y_i)$ is the force of mortality at age x conditional on covariates y_i , a is the age-independent level of mortality and b_0 the increase in mortality over age. The three indicator-variables, y_1 , y_2 and y_3 , denote the quarter of birth and take the value one if a person is born in a specific quarter, and zero otherwise; the first quarter is defined as reference group. The parameters a , b_0 , b_1 , b_2 , and b_3 are estimated by maximising the likelihood function. For each of the two cohorts, a separate model is estimated.

To estimate differences in survival according to the quarter of birth based on the US censuses for the year 1960, 1970, and 1980, a method called Survival-Attributes

Assay (Christensen et al. 2001) is applied. This method uses cross-sectional data on “fixed-attributes” to estimate the effect of a fixed trait on survival.

Let N_x be the number of people at age x . Let p_x be the proportion of x -year-olds who have some fixed attribute such as the season of birth. Let p_{x+n} be the proportion at age $x+n$. Let s be the conditional survival probability from age x to age $x+n$ for the individuals who have the fixed attribute. Let S be the conditional survival probability from age x to $x+n$ for the entire cohort.

Then, because

$$p_x N_x s = N_x S p_{x+n} \quad (23.2)$$

it follows that

$$s = S \frac{p_{x+n}}{p_x} \quad (23.3)$$

Thus, the relative risk of surviving from age x to age $x+n$ for people born in a specific quarter is the ratio of their observed proportions in the two cross-sections. The proportion of the population within 10-year age groups that is born in a certain quarter of the year is followed over the three census rounds.

This method relies on the assumption that the $x+20$ -year-olds in the third cross-section were similar to the x -year-olds in the first cross-section 20 years earlier; e.g. the 70-year-olds in 1980 are the survivors of the 60-year-olds in 1970 and the 50-year-olds in 1960. In other words, the change in the proportion of the fixed attribute over an age range of 20 years is solely due to age effects and not affected by cohort effects. Differences in the survival probability of ages that are further apart than 20 years can be both due to cohort and age effects. The method of Survival-Attributes Assays therefore, does not permit a clear distinction between age and cohort effects. The main advantages of the method are, however, that it does not require the calculation of death rates in order to verify the month-of-birth effect and that it can be used to study differences in survival at ages where death rates are low and therefore subject to random fluctuations. Thus, it can be used to study whether differences in survival by season of birth not only exists among today’s elderly but also among more recent cohorts.

23.4 Results

23.4.1 Differences in Lifespan in the United States, Austria, Denmark and Australia

A similar relationship between month of birth and lifespan exists in all of the Northern Hemisphere countries (Fig. 23.1). Adults born in the autumn (October–

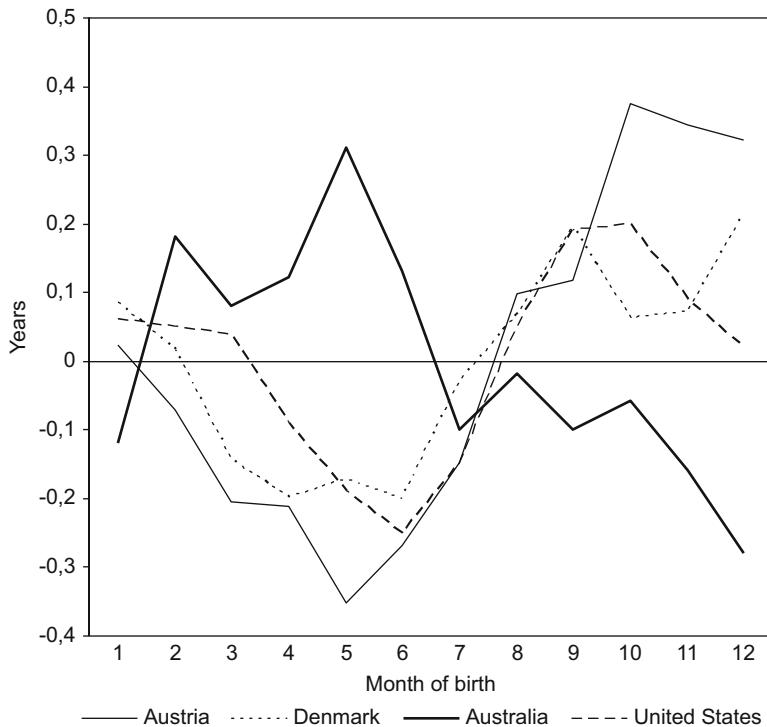


Fig. 23.1 Deviation from average remaining lifespan at age 50 for people born in a specific month in Austria, Australia, Denmark and the United States

December) live longer than do adults born in the spring (April–June). The difference in lifespan between the spring- and autumn-born is twice as large in Austria (0.6 years) as in Denmark (0.3 years).

In Denmark for those born in the second quarter, lifespans are 0.19 ± 0.05 years shorter than average; for those born in the fourth quarter they are 0.12 ± 0.04 years longer than average. This difference is statistically significant (Cox-Mantel statistic: $p < 0.001$). Also in Austria the deviation in mean age at death is highly significant (Bonferroni test: $p < 0.001$) for those born in the second and the fourth quarters. The lifespans of people born between weeks 14 and 26 are 0.28 ± 0.03 years below average; lifespans of those born between weeks 40 and 52 are 0.32 ± 0.03 years above average. A highly significant difference in mean age at death by month of birth exists for US decedents who died between 1989 and 1997. Those born in June and July die about 0.44 years earlier than the October-born. The pattern in the Northern Hemisphere is mirrored in the Southern Hemisphere. The mean age at death of people born in Australia in the second quarter of the year is 78.0; those born in the fourth quarter die at a mean age of 77.65. The difference of 0.35 years is statistically significant (Bonferroni test: $p < 0.001$).

Table 23.1 Parameter estimates and 95% confidence intervals of the Gompertz models estimated for Danes born in a specific quarter for birth cohorts 1898–1909 and 1909–1918 at ages 70–79

	Parameter estimate	95% LCI	95% UCI
Cohort 1898–1909			
1st quarter	1.00		
2nd quarter	1.03	1.02	1.04
3rd quarter	1.00	0.99	1.02
4th quarter	0.98	0.96	0.99
Ln(a)	-3.44	-3.45	-3.43
$b_0 \times 102$	8.59	8.42	8.76
Cohort 1909–1918			
1st quarter	1.00		
2nd quarter	1.08	0.96	1.21
3rd quarter	1.05	0.94	1.18
4th quarter	1.06	0.95	1.19
Ln(a)	-3.51	-3.62	-3.39
$b_0 \times 102$	8.08	7.49	8.68

Source: The Office for National Statistics

23.4.2 Changes in the Month-of-Birth Pattern over Cohorts in Denmark

Table 23.1 contains the parameter estimates $\ln(a)$ and b_0 and the odds ratios of the Gompertz models for the two cohorts 1889 to 1908 and 1909 to 1918. In the older cohort (1889 to 1909), there exists a significant excess mortality of those born in the second quarter (+3% as compared to those born in the first quarter). People born in the fourth quarter experience the lowest mortality risk (-2% compared to those born in the first quarter). In the younger cohort (1909 to 1918), the differences in mortality by quarter of birth become statistically not significant.

23.4.3 Changes in the 20-Year Survival Probability by Quarter of Birth in the United States

Figure 23.2 shows the relative risks of the 20-year survival probabilities conditional on age for people born in a specific quarter compared to the average population.

At younger ages, the mortality advantage of the autumn-born and the disadvantage of the spring-born is minor. For males, it is a maximum of 1% over an age range of 20 years. In other words, up to the age of 40 the conditional survival probability of surviving the next 20 years is about 1% higher for the autumn-born than for the average population; it is 1% lower for the spring-born. Differences in the 20-year survival start to accelerate from the age group 40–49 onwards, when those born in the fourth quarter have a higher chance of 2.4% to survive the next 20 years; this

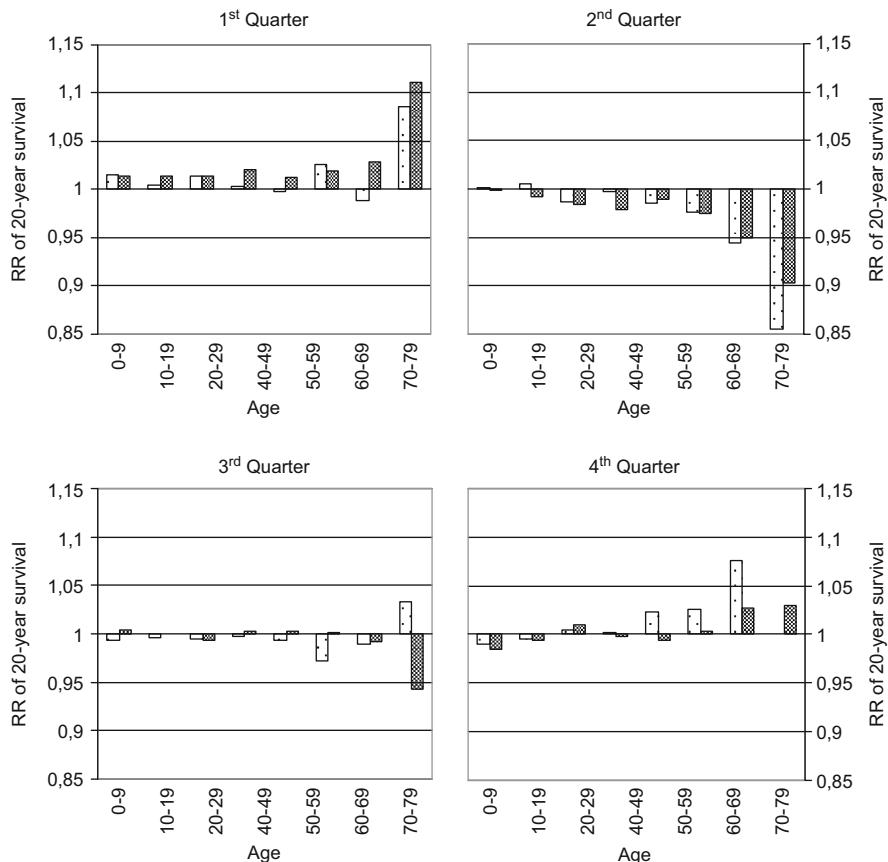


Fig. 23.2 Relative risk of 20-year survival by age in the 1960 US census by sex (males = white-dotted bars, females = grey bars) and quarter of birth

advantage increases to 7.8% for ages 60–69. The disadvantage in the 20-year survival of people born in the second quarter starts at ages 50–59 and is about minus 2.8%. It increases to minus 6.6 per cent for the age group 60–69. Similar trends emerge for women.

23.5 Discussion

In all four populations, Austria, Australia, Denmark and United States, significant differences in lifespan exist by month of birth. Those born in spring generally face a lower life expectancy than those born in autumn. This finding is independent of the Hemisphere as is shown by the Australian result.

The US death data contain detailed information about the state of birth, education, marital status, and race of the deceased (Doblhammer 2004). Among the white US population, the age-standardised peak-to-trough difference in the month-of-birth pattern increases from the North to the South while the basic pattern remains unchanged. The difference is smallest in New England, with 0.31 years and largest in the East South Central Region, with 0.86 years. The differences in the West are intermediate. The age-standardised differences in lifespan by month of birth vary significantly according to education levels. The difference between the spring trough and the winter peak is 0.62 years for those with a low education and 0.38 years for the highly educated. There exists a highly significant difference in the month-of-birth pattern by marital status. The difference between the peak and the trough is largest for the never-married (0.62 of a year), and smallest for the married (0.40 of a year); the widowed (0.45 of a year) and the divorced (0.44 of a year) are intermediate. The month-of-birth pattern of the 1.7 million US African Americans differs significantly from that of the white population. It differs not only with respect to the number of years between the trough and the peak (0.57 years) but also with respect to the shape of the curve. The mean age at death is highest for those born between January and March. As in the case of the white population, the mean age at death for African Americans is lowest for decedents born in July.

In a multivariate analysis of ages at death in the United States (Doblhammer 2004), the main effects of sex, month of birth, education, race, region of birth, marital status and all the two-way interactions of the variables are highly significant. This implies that the regional differences in the month-of-birth pattern are neither due to differences in education nor to differences in race but they exist independently of them. Overall, above factors explain 25% of the variation in ages at death. About 86% of the model explanation is due to the effect of marital status and only 0.4% to the effect of month of birth. The large majority (70%) of this 0.4% result from the interactions of month of birth with region of birth and race.

What are the causal mechanisms behind the month-of-birth effect on life-span? One frequently raised concern is that the month-of-birth effect reflects the seasonal distribution of deaths rather than the seasonal changes in the early-life environment. More specifically, the concern is that the interaction between the seasonal distribution of deaths and the monthly increase in adult mortality causes a month-of-birth pattern. This hypothesis has been already widely discussed in the research about the month-of-birth effect in schizophrenia, whose incidence is seasonal and whose risk increases with age. Two studies (Doblhammer and Vaupel 2001; Doblhammer 2004) have shown that, although month of birth, age, and month of death influence mortality simultaneously, they are independent of each other.

A second, frequently raised concern is that the month-of-birth effect is caused by socioeconomic differences in the seasonal distribution of births. The number of births is distributed seasonally over the year with the exception of only a few populations. If the seasonality in births is partly driven by the preference of couples for giving birth in certain seasons of the year, then this preference may differ between social groups. In schizophrenia research, this explanation is generally

known as the procreational habits theory. Individuals with schizophrenia may have a procreational pattern that differs from those of the non-schizophrenic population (Torrey et al. 1997). On basis of the 1981 census for Austria, it was possible to refute this hypothesis (Doblhammer 2004). This is also true for the deadline hypothesis. Starting school is usually tied to reaching a certain age before a certain deadline. Children who are born shortly after the deadline have to wait an additional year before they can start school and will therefore be among the oldest of their classmates. This may pose a special advantage compared to those who are born shortly before the deadline, who will thus always be among the youngest. However, since the mean age at death of the autumn-born is higher than that of the spring-born, the deadline hypothesis cannot explain the month-of-birth effect on the lifespan.

Public health experts at the beginning of the twentieth century felt that the health status of mothers and whether mothers breastfed their babies were the two most important factors determining the survival of an infant, followed by housing, sanitation and general poverty (Preston and Haines 1991). The health status of pregnant women depended largely on their diet and on the general disease load. Breastfeeding the infant is related primarily to a lower incidence of infectious diseases of the gastrointestinal tract, which historically is the major cause of infant mortality. Danish data on historical infant mortality between the years 1911 and 1915 show that it is the spring-born who experience higher mortality in their first year of life (Doblhammer 2004). The standardised death rate of the June-born infants is 30 per cent higher than the death rate of the December-born. This finding implies that those factors that contributed to the high infant mortality of the past are also the factors that cause the differences in lifespan by month of birth.

Nutrition is highly seasonal. Diet at the beginning of the twentieth century did not much resemble contemporary dietary patterns. People ate less meat, fruits, and vegetables and more starchy staple food. The first vitamins were not discovered until 1911, and in the early 1900s, nutritionists were even opposed to greens, which were considered to require more bodily energy for digestion than they provided. Although severe malnourishment was not widespread, people had inadequate nutrition – particularly during the winter and early spring. Peak growth of the fetus in utero occurs during the third trimester. For infants born in spring, the third trimester coincides with a period of largely inadequate nutrition; for those born in the autumn it coincides with a period of plenty.

The effect of nutrition early in life on adult health is highly contested. Studies that looked at the old-age mortality of cohorts born shortly after periods of famine, which were thus presumably marked by severe malnutrition of the mother during the gestational period of their unborn, did not find any differences (Kannisto et al. 1997). Two studies about the long-term effects of severe starvation during the siege of Leningrad come to contrary results (Stanner et al. 1997; Sparén et al. 2003). The effect of the Dutch famine in 1944–1945 on later life disease and mortality is explored in a series of studies (Rosenboom et al. 2001, 2000a, b). The authors find that mortality up to age 18 was higher for those born before the famine and those exposed to the famine in the third trimester. Between the ages of

18 and 50, however, no effect of prenatal exposure to the famine could be demonstrated. Thus, the evidence is weak concerning the effect of nutrition during gestational age on mortality later in life.

The incidence of infectious diseases depends on the climate and on the seasons of the year. The incidence of waterborne infectious diseases, which affect mainly the gastrointestinal tract, is correlated with warmer temperatures and flooding. Peak climatological temperatures coincide with the incidence of foodborne diseases. Many childhood diseases are highly seasonal; airborne diseases affecting the respiratory tract usually peak in autumn and winter. Historically, people born in years with extremely high infant mortality caused primarily by whooping cough and smallpox tend to have higher mortality later in life (Bengtsson and Lindström 2003).

Infant mortality at the turn of the twentieth century was mainly caused by exogenous factors, in particular infectious disease. Infants born in spring had an increased risk to die from infectious disease during their first year of life. The month-of-birth pattern in adult lifespan suggests that those who survived were debilitated and suffered from higher mortality during adult ages.

Explanations other than nutrition and infectious disease have also been brought forward to explain the month-of-birth effect. One of the first to study the influence of the month of birth on the lifespan was Elsworth Huntington, who formulated the hypothesis in 1936 that high temperatures at the time of conception weaken the “germ plasma” of the parents, with negative effects on the development of the foetus. Recent research has shown that the sperm quality of men who work outdoors does indeed decrease during periods of high temperatures (Centola and Eberly 1999). A related hypothesis is that hot summers are the cause of protein deficiencies at the time of conception (Pasamanick 1986). This hypothesis is clearly ruled out on basis of the US death data. The United States consists of six major climatic zones with very different climatic conditions. Since the US death data contain the state of birth, it is possible to correlate the peak-to-trough difference in lifespan by month of birth for people born in a specific state with maximum and minimum temperature and with the maximum difference in temperature. It appears that no correlation exists between the peak-to-trough difference and the temperature variables, neither for total mortality nor for major causes of death (Doblhammer 2004).

Another explanation is that seasonal changes in the hours of daylight influence the human endocrine functions and that the month-of-birth effect might be caused by variations in the internal chemistry or neural development brought about by the seasonal variations in light (Wehr 1998; Turnquist 1993; Quested 1991; Morgan 1978; Jongbloet 1975; Pallast et al. 1994).

23.6 Conclusion

The analysis of the Danish register data and the consecutive US census rounds shows that the differences in lifespan by month of birth have become smaller over time. This is consistent with the explanation of nutrition and infectious disease since both

have considerably improved over time. Although diet still differs between spring and fall, the difference in the nutritional value is much smaller than at the beginning of the twentieth century. In addition, the epidemiological transition has reduced infectious disease to a minor cause of death.

Ample evidence exists that the health of today's elderly is scarred by negative events that they experienced during their pre-natal or early post-natal life. This study shows that already among the elderly, those born in more recent years are less affected by seasonal early life factors and that the month-of-birth effect has become smaller. This finding suggest that in more recent cohorts period factor as opposed to cohort factors have gained increased importance and that period factors may therefore predominantly determine future gains in life expectancy.

On the other hand, the US census rounds indicate that differences in the 20-year survival still exist at young ages between 1960 and 1980. In addition, a study of twins born in the 1970s in Minnesota (Doblhammer 2004) shows that the seasonal pattern in birth weight – a widely used indicator for growth retardation in utero – is positively correlated with the month-of-birth pattern in the mean age at death of decedents aged 50+ who were born in Minnesota. Thus, there exists evidence that the seasonal fluctuations in the early life environment of recent cohorts have still an effect on life expectancy.

Huge gains have been made in the health environment during the very first period of life during the last century. Infant mortality has fallen drastically during the last century. Since infant mortality in the early twentieth century was primarily due to exogenous factors such as infectious disease, the decline in infant mortality points to a largely improved health environment of infants and children. However, in the 1980s and 1990s researchers have repeatedly pointed out that decreasing poverty among the elderly has led to increasing poverty in childhood (Preston 1984), particularly in the United States. A recent article by Komlos and Baur (2004) finds that in the most recent decades the height of Americans has been lagging behind that of Europeans while at the beginning of the twentieth century they were the tallest of the world. The authors even present some evidence that heights have been stagnating among US men and might actually be decreasing among females born in the 1960s. Height primarily reflects the socioeconomic and epidemiological environment during childhood and adolescence. It is significantly correlated with health and longevity at adult ages. Height and life expectancy rises together.

In an ageing society with too few children and an ever increasing proportion of the old and very old the danger exists that resources in general and social transfers in particular are channelled towards the elderly which may lead to increasing poverty among children. In his 1984 article, Preston wrote, “that the transfers from the working age population to the elderly are also transfers away from children [...].” Thus, childhood poverty may become an ever more widespread phenomenon leading to a deterioration of the social and epidemiological environment early in life. In this case, cohort factors may gain importance again in mortality forecasting.

References

- Barker, D. J. P. (1994). *Mothers, babies and diseases in later life*. London: British Medical Journal Publishing Group.
- Barker, D. J. P., & Osmond, C. (1986a). Infant mortality, childhood nutrition, and ischemic heart disease in England and Wales. *Lancet*, 10(1), 77–81.
- Barker, D. J. P., & Osmond, C. (1986b). Childhood respiratory infection and adult chronic bronchitis in England and Wales. *British Medical Journal*, 293(6557), 1271–1275.
- Barker, D. J. P., & Osmond, C. (1987). Death rates from stroke in England and Wales predicted from past maternal mortality. *British Medical Journal*, 295(6590), 83–86.
- Bengtsson, T., & Lindström, M. (2003). Airborne infectious diseases during infancy and mortality in later life in southern Sweden, 1766–1894. *International Journal of Epidemiology*, 32, 286–294.
- Centola, G., & Eberley, S. (1999). Seasonal variations and age-related changes in human sperm count, motility, motion parameters, morphology, and white blood cell concentration. *Fertility and Sterility*, 72, 803–808.
- Christensen, K., Orstavik, K. H., & Vaupel, J. W. (2001). The X-chromosome and the female survival advantage. An example of the intersection between genetics, epidemiology and demography. *Annals of the New York Academy of Sciences*, 954, 175–183.
- Davey Smith, G., & Kuh, D. (2001). Commentary. William Ogilvy Kermack and the childhood origins of adult health and disease. *International Journal of Epidemiology*, 30, 696–703.
- Doblhammer, G., & Vaupel, J. W. (2001). Lifespan depends on month of birth. *Proceedings of the National Academy of Sciences of the United States of America*, 98(5), 2934–2939.
- Doblhammer, G. (2004). *The late life legacy of very early life*. Heidelberg: Springer.
- Elo, I. T., & Preston, S. H. (1992). Effects of early-life conditions on adult mortality: A review. *Population Index*, 58(2), 186–212.
- Forsdahl, A. (1973). Momenter til belysning av den høye dødelighet, Finnmark Fylke. *Tidsskrift for Norske Lægeforeninga*, 93, 661–667.
- Forsdahl, A. (1977). Living conditions in childhood and subsequent development of risk factors for arteriosclerotic heart disease. *Journal of Epidemiology and Community Health*, 32, 34–37.
- Forsdahl, A. (1978). Are poor living conditions in childhood and adolescence an important risk factor for arteriosclerotic heart disease? *British Journal of Preventive and Social Medicine*, 31, 91–95.
- Gavrilov, L. A., & Gavrilova, N. S. (2003). Early-life factors modulating lifespan. In S. A. S. Rattan (Ed.), *Modulating aging and longevity* (pp. 27–50). Dordrecht: Kluwer Academic Publishers.
- Greville, T. N. E. (1943). Short methods of constructing abridged life tables. *Record of the American Institute of Actuaries*, 32(1), 29–42.
- Huntington, E. (1938). *Season of birth*. New York: Wiley.
- Jongbloet, P. H. (1975). The effects of pre-ovulatory overripeness of human eggs on development. In R. J. Blandau (Ed.), *Aging Gametes* (pp. 300–329). Basel: S. Karger.
- Joseph, K. S., & Kramer, M. S. (1996). Review of the evidence of fetal and early childhood antecedents of adult chronic disease. *Epidemiologic Reviews*, 18(2), 158–174.
- Kannisto, V., Christensen, K., & Vaupel, J. W. (1997). No increased mortality in later life for cohorts born during famine. *American Journal of Epidemiology*, 145(11), 987–994.
- Kermack, W. O., McKendrick, A. G., & McKinlay, P. L. (1934). Death-rates in Great Britain and Sweden. Some general regularities and their significance. *Lancet*, 226, 698–703.
- Kermack, W. O., McKendrick, A. G., & McKinlay, P. L. (2001). Death-rates in Great Britain and Sweden. Some general regularities and their significance. *International Journal of Epidemiology*, 30, 678–683.

- Komlos, J., & Baur, M. A. C. (2004). From the tallest to (one of) the fattest: The enigmatic fate of the American population in the 20th century. *Journal of Economics and Human Biology*, 2, 57–74.
- Kramer, M. S. (2000). Invited commentary: Association between restricted fetal growth and adult chronic disease. Is it causal? Is it important? *American Journal of Epidemiology*, 152(7), 605–608.
- Moore, S. E., Cole, T. J., Poskitt, E. M. E., Sonko, B. J., Whitehead, R. G., McGregor, I. A., & Prentice, A. M. (1997). Season of birth predicts mortality in rural Gambia. *Nature*, 338, 434.
- Moore, S. E., Fulford, A. J. C., Streatfield, P. K., Persson, L. A., & Prentice, A. M. (2004). Comparative analysis of patterns of survival by season of birth in rural Bangladeshi and Gambian populations. *International Journal of Epidemiology*, 33, 137–143.
- Morgan, R. (1978). Seasonality in schizophrenia (letter). *Lancet*, April, 830.
- Pallast, E. G. M., Jongbloet, P. H., Straatman, H. M., & Zielhuis, G. A. (1994). Excess seasonality of births among patients with schizophrenia and seasonal ovopathy. *Schizophrenia Bulletin*, 129, 355–361.
- Pasamanick, B. (1986). Seasonality of schizophrenia births (letter). *American Journal of Orthopsychiatry*, 56, 168–169.
- Preston, S. H. (1984). Children and the elderly. Divergent path's for America's dependents. *Demography*, 21(4), 435–457.
- Preston, S. H., & Haines, M. R. (1991). *Fatal years. Child mortality in late nineteenth-century in America*. Princeton: Princeton University Press.
- Quested, D. J. (1991). Incidence rates of schizophrenia (letter). *British Journal of Psychiatry*, 159, 587–588.
- Roseboom, T. J., van der Meulen, J. H. P., Osmond, C., Barker, D. J., Ravelli, A. C., & Bleker, O. P. (2000a). Plasma lipid profiles in adults after prenatal exposure to the Dutch Famine. *American Journal of Clinical Nutrition*, 72, 1101–1106.
- Roseboom, T. J., van der Meulen, J. H. P., Osmond, C., Barker, D. J., Ravelli, A. C., Schroeder-Tanka, J. M., van Montfrans, G. A., Michels, R., & Bleker, O. (2000b). Coronary heart disease after prenatal exposure to the Dutch Famine, 1944–45. *Heart*, 84, 595–598.
- Roseboom, T. J., van der Meulen, J. H. P., Barker, C. O., D. J. P., Ravelli, A. C. J., & Bleker, O. P. (2001). Adult survival after prenatal exposure to the Dutch famine 1944–45. *Paediatric and Perinatal Epidemiology*, 15(3), 220–225.
- Simondon, K. B., Elguero, E., Marra, A., Diallo, A., Aaby, P., & Simondon, F. (2004). Season of birth is not associated with risk of early adult death in rural Senegal. *International Journal of Epidemiology*, 33, 130–136.
- Sparén P., Vägerö, D., Shestov, D. B., Plavinskaja, S., Parfenova, N., Hoptiar, V., Paturot, D., & Galanti, M. R. (2003). Long term mortality after severe starvation during the siege of Leningrad: Prospective cohort study. *British Medical Journal*, <https://doi.org/10.1136/bmj.37942.603970.9a> (published 5 December 2003).
- Stanner, S. A., Bulmer, K., Andres, C., Lantseva, O. E., Borodina, V., Poteen, V. V., et al. (1997). Does malnutrition in utero determine diabetes and coronary heart disease in adulthood? Results from the Leningrad siege study, a cross sectional study. *British Medical Journal*, 315, 1342–1348.
- Torrey, E. F., Miller, J., Rawlings, R., & Yolken, R. H. (1997). Seasonality of births in schizophrenia and bipolar disorders: A review of the literature. *Schizophrenia Research*, 28, 1–38.
- Turnquist, K. (1993). Second-trimester markers of fetal size in schizophrenian (letter). *American Journal of Psychiatry*, 150, 1571–1572.

- Vaupel J. W., Wang, Z., Andreev, K. F., & Yashin, A. I. (1997). *Population data at a glance. Shaded contour maps of demographic surfaces over age and time* (Odense Monographs on Population Aging, 4). Odense: Odense University Press.
- Wehr, T. (1998). Effect of seasonal changes in daylength on human neuroendocrine function. *Hormone Research*, 49, 118–124.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 24

Early-Life Conditions and Old-Age Mortality in a Comparative Perspective: Nineteenth Century Sweden and Belgium



Tommy Bengtsson and George Alter

24.1 Introduction

Kermack et al. (1934) proposed the cohort explanation in their analysis of the aggregated mortality decline in England, Wales, Scotland, and Sweden. Their conclusion was that reductions in the death rates of the various age groups attained at any particular time depended primarily on the individuals' date of birth, and only indirectly on the particular year under consideration. The essential effects on health and survival of adults and older persons were mainly caused by improvements and beneficial effects on their respective birth cohorts during childhood several decades earlier. Jones (1956) brought up the cohort approach anew and recently it has gained focus again through works in medicine but also in historical demography (Barker 1994; Elo and Preston 1992; Fogel 1994; Fridlizius 1989; Kuh and Ben-Shlomo 1997; Preston et al. 1998; Finch and Crimmins 2004). The plausible causal relationships between early-life experiences and old-age mortality have been discussed, with special attention to intrauterine cellular development and cellular development during early childhood. Robert Fogel (1994) has proposed several plausible causal mechanisms that connect malnutrition in utero and during early life to chronic diseases in later life. These propositions are also supported by the work of Barker (1994, 1995) who suggested that the preconditions for coronary heart disease, hypertension, stroke, diabetes, and chronic thyroiditis are initiated in utero without

T. Bengtsson (✉)

Centre for Economic Demography, Lund University, Lund, Sweden

e-mail: tommy.bengtsson@ekh.lu.se

G. Alter

University of Michigan, Ann Arbor, MI, USA

Inter-University Consortium for Political and Social Research (ICPRS), University of Michigan,
Ann Arbor, MI, USA

becoming clinically manifest until much later in life. In contrast, Jones (1956), and later Fridlizius (1989), in his analysis of the aggregated mortality decline in Sweden, proposed that the genesis of disease in later life could be due to exposure to certain infectious diseases in the first years of life. Fridlizius argued that this was caused by life-long reduced immunity, which consequently gave higher general risk of other infectious diseases in later life. The knowledge of the medical mechanisms today seems to be more in favour of permanent retardation of organs due to infections rather than immunological mechanisms (Finch and Crimmins 2004). Either way, it implies that factors other than nutrition are important early-life determinants of mortality in later life, because the outcome of some other important infectious diseases, like smallpox, is almost completely unrelated to the nutritional status of the infected individual (Rotberg and Rabb 1985).

In two recent essays, Bengtsson and Lindström (2000, 2003) investigated the different cohort hypotheses using longitudinal data for individuals instead of aggregated data like Kermack et al., Fridlizius, and many others. Bengtsson and Lindström analysed demographic and economic data for individuals and households and combined with community data on food prices and disease load. The analyses included variables measuring conditions during the fetal stage and the first year of life, as well as the disease load and access to nutrition (food prices). Strong support was found for the hypothesis that the disease load experienced during the birth year had a consistent impact on mortality in later life (Bengtsson and Lindström 2003). This was particularly the case with the outcome of airborne infectious diseases during old age (Bengtsson and Lindström 2000). The hypotheses that access to nutrition and the disease load on mothers during the fetal stage had impacts on mortality in later life were not supported and neither was the hypothesis about the effects of access to nutrition during the first year of life (Table 24.1)

The influence of short-term variations in infant mortality on old-age mortality was due to both cycles in infant mortality and trend (Bengtsson and Lindström 2003). A test of potential nonlinearity of the cycle effect reveals that only years with particularly high disease load had an impact on mortality in later life (Bengtsson and Lindström 2003). Highly virulent infectious diseases, especially smallpox and whooping cough, dominated infant mortality in those years (Bengtsson and Lindström 2003). These diseases are so virulent that the outcome shows no social

Table 24.1 Estimation of effects of infant mortality rate at birth on mortality in ages 55–80 years in various diseases

Death cause	Deaths	Rel. risk	Wald P-value
Airborne infectious disease	343	4.65	0.00
Non-infectious diseases	208	3.41	0.05
Old-age mortality	339	1.46	0.45
All causes of death	1400	1.80	0.03

Source: Bengtsson and Lindström (2000: table 5)

Note: The model controls for sex, birth year, birthplace, socioeconomic status, time period, and logarithm of current rye prices in four Scanian parishes, 1766–1894. Number of deaths = 1400

gradient, and they probably penetrated the entire population. Smallpox, which was mainly a childhood disease during the eighteenth century, was not only highly virulent but also equally mortal with an overall fatality rate of about 15% (Sköld 1996: 70–75; Riley 2001). The eighteenth-century pattern of total childhood dominance in smallpox mortality indicates that during this period the majority of the adult population had already been exposed to smallpox as children, and had survived. The fact that airborne infectious diseases in the first year of life particularly affected old-age mortality may imply that exposure to airborne infectious diseases during the first year of life may make individuals more vulnerable throughout life. Cohorts exposed during infancy to such infectious diseases may thus be much more susceptible to high morbidity and mortality rates in old age than cohorts exposed later in childhood to epidemics of smallpox and other infectious diseases. The causal biological mechanisms in early life that might explain these significant associations are only partly known (for further discussion, see Bengtsson and Lindström 2003; Lindström and Davey Smith 2008).

In another recent essay on the effect of different forms of community variables, Bengtsson et al. (2002), analyse early-life links in further detail. The essay supports the earlier conclusion that the functional form of the effect of exposure is nonlinear and includes valuable information about the causal link.

Furthermore, it shows that while both observed and unobserved characteristics at family level – whether the result of genetics or of shared experience – had a significant impact on mortality in later life, their inclusion in the models only marginally alters previous estimates.

The main question in this paper is whether these historical findings for rural Sweden are a local, or possibly a Nordic, phenomena, or if we find the same pattern outside Sweden. We have therefore replicated the study for southern Sweden with one for the rural parish of Sart in eastern Belgium. The population we analyze in Sart are born in the period 1799 to 1846 and followed until 1899. For southern Sweden, we analyze the population born between 1750 and 1840, followed to 1895. The longitudinal demographic data on individuals and household socioeconomic data from parish registers were combined with community data on food costs and disease load, using a Cox regression analysis for the mortality among ever-married persons in ages 55–80 years. In addition to trying to answer the main question, we will briefly discuss whether information about conditions early in life can be used to improve mortality forecasting.

24.2 Models

We use a *proportional hazards model* (Cox 1972). This means that we assume that a relative effect on mortality of any covariate is constant over age. The model allows time-varying covariates. It is very important to check the assumptions behind this model, especially the proportionality assumption. We have therefore routinely tested

all models for deviations from the proportionality assumption.¹ The test we have used is based on the correlation between $\log(t)$ and the Schoenfeld residuals for each covariate. A large correlation indicates that the corresponding coefficient varies with time, which means that the hazards are not proportional. We found no signs of nonproportionality for any of the covariates or globally.²

24.3 Data for Scania

The Swedish data come from the Scanian Demographic Database, which consists of records of births, marriages, deaths, and migrations for nine rural

parishes and one town situated in the southernmost part of Sweden. The material for two of the parishes dates back to 1646 and for the others to the 1680s. The publicly available records end in 1895. Four of the rural parishes – Hög, Kävlinge, Halmstad, and Sireköpinge – are included in this study. The parish register material is of high quality and shows few gaps for births, deaths and marriages. Migration records are less plentiful, but continuous series exist from the latter part of the eighteenth century. Information concerning farm size, property rights, and various other items from the poll tax records and land registers, are linked to the family reconstitutions based on the parish records of marriages, births, and deaths.

Our interest in life-course effects on later-life mortality further limits our dataset. We need information about socioeconomic conditions at birth not only for those born in the parish but also for in-migrants. Data needed to create information for the socioeconomic condition at birth of an in-migrant can be obtained from the birth parish, but is generally available only after 1829. We have therefore had to limit the period of our analyses to 1829–1894.

The sampled parishes are compact in their geographical location, showing the variations that could occur in peasant society with regard to size, topography, and socioeconomic conditions, and they offer good, early source material. The entire area was open farmland, except northern Halmstad, which was more wooded. Halmstad and Sireköpinge were noble parishes, while freehold and crown land dominated in Kävlinge and Hög. The parishes each had 200–500 inhabitants in the latter half of the eighteenth century. The agricultural sector in Sweden, and Scania, became increasingly commercialised in the beginning of the nineteenth century. New crops and techniques were introduced. Enclosure reforms and other reforms in the agricultural sector influenced the population growth, in particular in Sireköpinge, which experienced rapid population growth. In Kävlinge, the establishment of several factories and railroad transportation led to rapid expansion from the 1870s onwards (see Bengtsson 2001 for more details).

¹We used the function ‘cox.zph’ in the ‘survival’ package in R.

²For a more detailed description of the test, see Therneau and Grambsch (2000: chapter 6, 127–152).

The social structure of the agricultural sector is often difficult to analyse since differences of wealth between the various categories of farmers and occupations are unclear and subject to change with the passage of time. Land was the most important source of wealth in the societies we analyse. Data from land registers on types of tenure is limited and therefore must be combined with information from poll tax records concerning farm size in order to arrive at a better understanding of each household's access to land. The category *peasant* includes freeholders, tenants on crown land, and tenants on noble land as well as a few tenants on church land. We only include peasants with farms larger than 1/16 *mantal* in this category since it has been argued that peasants with smaller farms were not self-supporting. *Mantal* was not a measure of the actual size of the farm but a tax-assessment unit based on potential productivity. The few persons belonging to the nobility are also included in this group. The second group includes farmers with land smaller than 1/16 *mantal*, crofters, and the landless workers, the latter being in majority (see Bengtsson and Lindström 2000 for more details). Thus, we are only differentiating between two social groups: those with land enough to feed a family and those who need to work for someone else to be able to support a family.

The nineteenth century was a period of considerable social change in the countryside. It has been described as a period of proletarianization and pauperization (see Lundh 1983 for an overview). The numbers of landless increased (Carlsson 1968). The downward mobility was significant since many children of farmers were unable to obtain a farm themselves. This was true both for Sweden in general and for the area we study (Lundh 1998). Downward mobility was also common among the elderly, since many either sold their farms or gave them to their children. However, they could still be rather well off since the new owner of their farm often had to look after them in accordance with special contracts (*undantagskontrakt*). Not only did social stratification increase at the beginning of the nineteenth century, the economic condition of the landless worsened. They were, for example, more vulnerable to short-term economic stress than they were both before and after this period (Bengtsson and Drike 2005).

The nineteenth century was also a period of rapidly expanding population in Scania as well as in Sweden in general (see Bengtsson 2001). Fertility rates were rather stable and mortality fell, first among infants and children, later among adults and the elderly. During the period we study, the crude death rate for ages 55–80 years was declining in the four parishes, as in the rest of Sweden. Life expectancy of Swedish women was the highest in the world (about 45 years around 1830) and remaining life expectancy at age 55 was about 16 years. The figures for men were several years lower. The corresponding figures for our four parishes are slightly higher than for Sweden.

Mortality in ages 55–80 years varied markedly from 1 year to the next and showed a downward trend, as was the case nationally in Sweden (Jones 1956; Fridlizius 1989). The models that we apply include a number of variables: sex, whether a person has in-migrated to the parish or not, the parish of residence, birth year, current food prices, season of birth and four other variables as indicators of conditions in early life (see Bengtsson and Lindström 2000, 2003 for details). The

infant mortality rate in the year of birth, a time-varying community variable, is used as a fixed early-life covariate. It measures the disease load during first year of life. Here the variation from 1 year to the next is also large but diminished somewhat over time. The trend is upward for most of the eighteenth century and downward from the 1780s onwards. Thus, old-age mortality is preceded by a decline in infant mortality by about 70 years. In order to separate effects of the trend from the effects of occurring cycles in infant mortality, we have constructed two variables: a trend variable, constructed using a Hodrick-Prescott filter and a variable designed to pick-up cycles, measured as deviations from the trend. We have then categorized the variable “infant mortality rate cycle” into five groups, the upper one at 0.12.³

The aggregated indicator of the food prices is included in the regressions as a time-varying communal covariate (Bengtsson 1989, 1993). We use the deviation from the log trend in rye prices as an indicator. This means that the aggregated economic information is used as a time-varying covariate common to all individuals in the risk set at each point in calendar time.⁴ Both price at birth and price at the time of conception are included in the models. Thus, we estimate the effects of food availability both during pregnancy and during first year of life. We use local prices of rye, the most common crop, referring to the conditions in the fall, and we estimate the effects of food prices during the subsequent year (see Bengtsson 2001 for more details). Finally, we include mortality in ages 20–50 years as an indicator of the disease environment during pregnancy.

24.4 Data for Sart

Our Belgian data come from the municipality of Sart, located in the Ardennes region, close to the German border. Although the commune is geographically quite large compared to other Belgian communes, it has always been sparsely populated (42 inhabitants per square kilometer in 1846). The territory of Sart includes part of the area known as the “Hautes Fagnes,” a high plateau of peat bogs and forest where agriculture has always been marginal. The population of Sart resided in a half dozen hamlets on the northern slopes of the Fagnes. Each one remains an island surrounded by forests. The area was very poor in the nineteenth century (Alter et al. 2004b).

The rural population of the Ardennes was mostly composed of smallholders on middle-sized farms. At the beginning of the nineteenth century, much of the land in Sart was held in common, and recognized members of the community had a variety of rights on common lands and forests (Vliebergh and Ulens 1912). An 1847 law encouraged the sale of common land and the formation of larger holdings, but the few large estates that resulted were abandoned shortly thereafter (Vliebergh and

³For further discussion of the functional form, see Bengtsson et al. (2002).

⁴We use the free software R (Hornik 2002), and the R package eha, which can handle time- varying community covariates in a simple way (Broström 2002).

Ulens 1912: 62). Agricultural techniques were primitive at the beginning of the century, and farmers depended heavily on forests for both wood and feeding livestock. The population of Sart grew from about 1800 persons at the beginning of the nineteenth century to about 2500 in 1850, and there are signs of increasing population pressure (Alter et al. 2004a). After 1850, the area was strongly affected by the rapid industrialization of the region. Sart is less than 20 km from Verviers and 40 km from Liège, two important centers of the Industrial Revolution. Out-migration to industrial centers increased rapidly after 1850 (Oris and Alter 2001). The combination of slow population decline and the introduction of new agricultural techniques raised incomes, which are reflected in rapidly rising property values.

Sart was chosen as a research site because we have exceptionally complete demographic records covering most of the nineteenth century. In 1811, the population of Sart was recorded in a register showing the names of all persons arranged in households. This register was updated to reflect changes due to births, marriages, deaths, and migration, and a new register was opened in 1843. After the national census of 1846, population registers were implemented in all of Belgium (Alter 1988). Since the population register records dates of in- and out-migration, we are able to reconstruct the population at risk at every point in time.

Previous studies have already revealed the patterns of mortality in Sart during the nineteenth century (Alter et al. 2004a; Alter and Oris 2000). The crude death rate remained approximately constant until 1870, when a sustained decline began. The influence of several major epidemics is readily apparent, including a typhus epidemic in 1816–17 following the Napoleonic Wars. Life expectancy remained stable around 42 years before 1850 and increased slowly in the second half of the century. Improvements in mortality began among children aged 1 to 10 in the period 1867–1880, and spread to adolescent and adult ages after 1880 (Alter et al. 2004a). Infant mortality, however, was largely unaffected until the last decade of the century. The infant mortality rates show no trend during the period examined. Mortality after age 55 did improve after 1850 as in Scania.

We use annual prices of oats as an indicator of economic stress in Sart⁵. Oats were by no means the preferred grain, because it could not be made into bread.⁶ As the least expensive grain, however, oats were the last resort of the poor. Previous studies have shown that an increase in the price of rye was often associated with lower mortality in Sart (Oris et al. 2005). Prices were determined on international markets, and farmers benefited from high prices as long as they had substitutes for high-priced

⁵We have done analyses with both oats and rye, and the response was more sensitive to oats in this case. Both grains were grown in Sart, but prices were determined on international markets. In some cases we find that the grains have opposite signs when they are both included in the regression equation. The reason, we believe, is that farmers in Sart had the choice of selling their rye in urban markets when the price rose. Even if this meant that they ate more oats (the cheaper grain), it would increase their welfare. When the price of oats went up, the poor in Sart were severely affected, because (as we say in the paper) oats was their last resort.

⁶Prices for 1798 to 1830 were derived from Deprez (1948). Prices after 1831 come from Gadisseur (1990).

grains. An increase in the price of oats had a large impact on the poor, however, because they had nowhere else to turn. Prices of oats tended to be more stable than other grains, but there were large fluctuations in prices between 1800 and 1820. There were constant demands for men and supplies when Eastern Belgium was part of the Napoleonic Empire. A severe economic crisis accompanied by a typhus epidemic followed Napoleon's defeat.

24.5 Results

Table 24.2 shows the abridged results for a basic model with four early-life factors included as well as a number of control variables. The full results for Sart are shown in Appendix Table 24.5, and for Scania in Bengtsson and Lindström (2000). Our focus is whether the effect of the disease load in year at birth has a strong impact on mortality in later life both in Scania and in Sart. This, however, is not the case. Instead, we find that the impact of food prices in year of birth is important in Sart; a 10% decline in food prices reduces mortality in older ages by 14%. This is the average effect for the entire population and since we can safely assume that lower strata were most negatively affected by high food prices the effect must have been even larger among the poor (Bengtsson 2004; Alter et al. 2004a).

Sources: For Scania: Bengtsson and Lindström (2003: table 2). For Sart: Appendix Table 24.6.

Table 24.3 shows the estimation results of a similar model in which the disease load in the first year of life is included as a single early-life variable but where it has been divided into a trend and a cycle component. We still find no support for the hypothesis that the disease load in first year of life has an impact on old age mortality in Sart, as it was in Scania⁷. This is, in fact, hardly surprising since the infant mortality rate shows almost no trend in Sart.

Turning to Table 24.4, showing possible nonlinear effects of the disease load in birth year on old age mortality, we find that children born in years of very high infant mortality in Sart also face higher mortality later in life. The result is precisely the same as for Scania, despite the fact that the study for Scania covers a much longer period, one in which infants often were exposed to a highly virulent disease – smallpox. Thus, normal variation in the disease environment at birth from 1 year to the next has no impact on old-age mortality in Sart, only the extreme years have an effect in later life.

⁷There were no smallpox epidemics in Sart because there was a strong vaccination campaign in the early 19th century. There were isolated cases of smallpox later in the century, including a small epidemic in 1871, because physicians did not realise that the immunity conveyed by vaccination is not permanent.

Table 24.2 Estimation of effects of infant mortality rate at birth, crude death rate at ages 20–50 years 9 months prior to birth, and cycles in the logarithm of rye prices both 9 months prior to birth and at birth

Covariate	Scania		Sart	
	Rel.risk	p-value	Rel.risk	p-value
Infant mortality rate at birth	2.93	0.00	1.00	0.33
Crude death rate in ages 20–50 at conception	0.67	0.91	0.99	0.62
Food-prices up 10% at conception	0.65	0.60	0.86	0.43
At birth	0.19	0.86	1.39	0.09
No. of deaths	1398		573	

Sources: For Scania: Bengtsson and Lindström (2003: table 1). For Sart: Appendix Table 24.5

Note: The model controls for socioeconomic status, sex, birth year, birth season, birthplace, parish of residence, and logarithm of current rye prices in four Scanian parishes, 1766–1894, and in Sart, 1854–1899

Table 24.3 Estimation of effects of cycles and trend in infant mortality rate (IMR) at birth

Covariate	Scania		Sart	
	Rel.risk	p-value	Rel.risk	p-value
IMR at birth trend	4.85	0.01	1.01	0.36
Cycle	2.35	0.01	1.00	0.97
No. of deaths	1398		573	

Note: The model controls for sex, birth year, birth season, birthplace, parish of residence, socioeconomic status, crude death rate at ages 20–50 years 9 months prior to birth, and cycles in the logarithm of rye prices 9 months prior to birth, at birth, and currently in four Scanian parishes, 1766–1894, and in Sart, 1854–1899

Table 24.4 Estimation of nonlinear effects of cycles in infant mortality rate at birth

Covariate	Scania		Sart	
	Rel.risk	p-value	Rel.risk	p-value
IMR at birth				
Very low	1.03	0.90	–	–
Low	1.01	0.93	1.14	0.26
Normal	1.00	–	1.00	–
High	0.91	0.28	0.89	0.28
Very high	1.51	0.02	1.58	0.02
No. of deaths	1398		573	

Note: The model controls for sex, birth year, birth season, birthplace, parish of residence, socioeconomic status, crude death rate at ages 20–50 years 9 months prior to birth, and cycles in the logarithm of rye prices 9 months prior to birth, at birth and currently in four Scanian parishes, 1766–1894, and in Sart, 1854–1899

Sources: For Scania: Bengtsson and Lindström (2003: table 3). For Sart: Appendix Table 24.7

24.6 Discussion

In a large body of studies, the conclusion is that conditions very early in life, in the fetal stage and in the first year of life, have an impact on health and mortality in later life. Robert Fogel (1994) has proposed several plausible causal mechanisms that connect malnutrition in utero and during early life to chronic diseases in later life. Fogel's propositions have been supported by the work of Barker (1994, 1995) who has suggested that the preconditions for coronary heart disease, hypertension, stroke, diabetes, and chronic thyroiditis are initiated in utero but do not become clinically manifest until much later in life. To the extent that the damage caused by malnutrition in early life shows up later in life, we label it "permanent" damage. Recent reviews of the evidence are given by Finch and Crimmins (2004) and Lindström and Davey Smith in this volume.

For Scania, we have previously shown that the disease load in the birth year affected mortality in ages 55–80 years, while we have found no support for either the in utero or the diseases environment during pregnancy hypotheses (Bengtsson and Lindström 2000, 2003; Bengtsson et al. 2002). The question we raise in this paper is whether this is a "Swedish" phenomenon or if it can be found elsewhere. On the one hand, we find support for the hypothesis of disease load in year at birth also for nineteenth century Sart, a rural parish in eastern Belgium. On the other hand, not only the disease load but also the food availability in year at birth affected old age mortality in Sart. Neither in Sart nor in Scania have we found any in utero effects.

As regards the impact of the disease load during infancy, in Sart we find a threshold effect but no basic effect (linear) and no trend effect. When we analyze the period 1829–1894 in Scania, we find threshold effects almost identical to those in Sart (Bengtsson and Broström 2003). Over the longer period from 1766 to 1894 the results for Scania are driven both by the trend in infant mortality, which also appears in analyses of aggregated data (Jones 1956; Fridlizius 1989; Finch and Crimmins 2004), and by extreme values. In the eighteenth century, most years with very high infant mortality were years with high smallpox mortality (Bengtsson and Lindström 2003: Table 24.4) but smallpox was much less important in nineteenth-century Sart – and the sub-period 1829–1894 for Scania. Thus, other epidemic diseases like whooping cough, for Scania, and typhus and cholera, for Sart, had effects in old age mortality similar to those of smallpox.

Another interesting finding is that food prices in the year of birth also affected later life mortality in Sart. Could it be that when highly virulent diseases (smallpox) disappear, socioeconomic conditions gain in importance, at least relatively speaking but perhaps also in absolute terms? If this is the case, then it fits well with the observation that socioeconomic factors were gradually becoming more and more important for mortality levels during the course of the nineteenth century.

Finally, we turn to a brief discussion of the use of early-life indicators to improve mortality forecasting. Several studies have shown that the disease load early in life can cause permanent damage that shows up in increased mortality risks later in life. This was the case for Sweden, using aggregate level data, as well as for a rural

population in the very south of Sweden, using micro level data. This was also the case for Sart in eastern Belgium, also using individual level data, as shown in this study. Several other recent studies show similar results for other parts of Europe, as well as for Canada (for an overview, see Beise et al. 2006; Schuster and Sunder 2005). Taken together, these findings suggest that the disease load early in life can be used to predict mortality later in life. Finch and Crimmins (2004), however, find that whereas mortality in the first year of life is of most importance for mortality risks at older ages in some countries, mortality in childhood is more important in other countries. This constitutes a problem since it implies different mechanisms. If the mechanism in play had been the same and remained stable across populations, its value as a predictor of future mortality would certainly be much higher.

Acknowledgement This work has been done within three projects: Early-Life Conditions, Social Mobility, and Longevity: Social Differences and Trends in Adult Mortality in Sweden, 1650–1900, financed by the Swedish Council for Working Life and Social Research; Early-Life Conditions, Social Mobility, and Health in Later Life, financed by the Bank of Sweden Tercentenary Foundation, and Early-Life Conditions, Social Mobility, and Longevity, P01 AG18314, financed by the National Institute of Aging, USA. The Swedish data come from the Scania Demographic Database, which is a collaborative project between the Regional Archives in Lund and the Department of Economic History at Lund University.

Appendix

Table 24.5 Sart, 1855–1899. Estimation of effects of infant mortality rate at birth, crude death rate at ages 20–50 years 9 months prior to birth, and cycles in the logarithm of out prices both 9 months prior to birth and at birth, controlling for socioeconomic status, sex, birth year, birth season, birth place. The number of deaths is 573

Subjects	658	Rel.risk	p-value	Mean
Deaths	573			
Time at risk	8770.889			
Log likelihood	−3070.84			
p	0			
Female	0.91	0.28	0.47	
Year of birth	1.04	0.00	1821.42	
Birth season:				
Jan–Mar	1.00			27.21
Apr–Jun	1.09	0.49		25.11
Jul–Sep	0.94	0.61		21.64
Oct–Dec	1.20	0.10		26.03
Wealth after age 55				
None	1.00			36.63
Low	1.19	0.07		43.47
High	0.70	0.00		19.91

(continued)

Table 24.5 (continued)

Born in Sart	1.13	0.24	0.70
IMR year of birth	1.00	0.33	169.86
Adult Death Rate year before birth (difference from trend)	0.99	0.62	0.12
Oats year of birth (difference from trend in logs)	1.39	0.09	0.00
Oats year before birth (difference from trend in logs)	0.86	0.43	0.00

Table 24.6 Sart, 1855–1899. Estimation of effects of cycles and trend in infant mortality rate at birth controlling for socioeconomic status, sex, birth year, birth season, birth place. The number of deaths is 573

		Rel.risk	p-value	Mean
Subjects	658			
Deaths	573			
Time at risk	8770.889			
Log likelihood	−3071.97			
P	0			
Female	0.90	0.24	0.47	
Year of birth Birth season:				
Jan–Mar	1.04	0.00	1821.42	
	1.00		27.21	
Apr–Jun	1.07	0.58	25.11	
Jul–Sep	0.93	0.58	21.64	
Wealth after age 55	Oct–Dec	1.21	0.09	26.03
	None	1.00		36.63
	Low	1.22	0.04	43.47
	High	0.70	0.00	19.91
Born in Sart		1.14	0.23	0.70
IMR Trend year of birth		1.01	0.36	161.80
IMR Cycle year of birth		1.00	0.97	8.07

Table 24.7 Sart, 1855–1899. Estimation of threshold effects of infant mortality rate at birth controlling for socioeconomic status, sex, birth year, birth season, birth place. The number of deaths is 573

		Rel.risk	p-value	Mean
Subjects	658			
Deaths	573			
Time at risk Log likelihood P	8770.889			
	−3068.41			
	0			
Female	0.90	0.21	0.47	
Year of birth Birth season:				
Jan–Mar	1.04	0.00	1821.42	
	1.00		27.21	

(continued)

Table 24.7 (continued)

	Apr–Jun	1.07	0.56	25.11
	Jul–Sep	0.94	0.60	21.64
Wealth after age 55	Oct–Dec	1.21	0.09	26.03
	None	1.00		36.63
	Low	1.22	0.04	43.47
	High	0.72	0.01	19.91
Born in Sart		1.14	0.23	0.70
IMR in year of birth	Low	1.14	0.26	21.92
	Normal	1.00		47.49
	High	0.89	0.28	21.19
	Very high	1.58	0.02	9.41

References

- Alter, G. (1988). *Family and the female life course: The women of Verviers, Belgium, 1849–1880*. Madison: University of Wisconsin Press.
- Alter, G., Neven, M., & Oris, M. (2004a). Mortality and modernization in Sart and surroundings, 1812–1900. In T. Bengtsson, C. Campbell, J. Z. Lee, et al. (Eds.), *Life under pressure: Mortality and living standards in Europe and Asia, 1700–1900* (pp. 173–208). Cambridge, MA: MIT Press.
- Alter, G., Neven, M., & Oris, M. (2004b). Stature in transition: A micro-level study from nineteenth-century Belgium. *Social Science History*, 28(2), 231–247.
- Alter, G., & Oris, M. (2000). Mortality and economic stress: Individual and household responses in a nineteenth-century Belgian village. In T. Bengtsson & O. Saito (Eds.), *Population and economy. From hunger to modern economic growth* (pp. 335–370). Oxford: Oxford University Press.
- Barker, D. J. P. (1994). *Mothers, babies, and disease in later life*. London: British Medical Journal Publishing Group.
- Barker, D. J. P. (1995). Fetal origins of coronary heart disease. *British Medical Journal*, 311, 171–175.
- Beise, J., Desjardin, B. & Johansson, K. (2006). *Early life conditions and adult mortality in the St. Lawrence Valley, Canada, 1700–1850*. Paper presented at the Annual Meeting of the Population Association of America, Los Angeles, March 30–April 1, 2006.
- Bengtsson, T. (1989). *Real wage variations and adult mortality: Life events in Västanfors, 1750–1859*. Paper presented at the IUSSP General Conference, New Delhi.
- Bengtsson, T. (1993). Combined time-series and life event analysis: The impact of economic fluctuations and air temperature on adult mortality by sex and occupation in a Swedish mining parish. In D. Reher & R. Schofield (Eds.), *Old and new methods in historical demography* (pp. 239–258). Oxford/New York: Clarendon Press/Oxford University Press.
- Bengtsson, T. (2001). Mortality: The great historical decline. International Encyclopedia of the Social and Behavioral Sciences, Elsevier Science Ltd, pp. 10079–10084. Available online at: <http://www.sciencedirect.com/science/referenceworks/0080430767>
- Bengtsson, T. (2004). Mortality and social class in four Scanian parishes, 1766–1865. In T. Bengtsson, C. Campbell, J. Z. Lee, et al. (Eds.), *Life under pressure. Mortality and living standards in Europe and Asia, 1700–1900* (pp. 135–171). Cambridge, MA: MIT Press.

- Bengtsson, T., & Broström, G. (2003). *How conditions in early-life and social mobility influence mortality in later life: Southern Sweden, 1829–1894*. Paper presented at the SSHA conference, November 13–16, 2003, Baltimore, USA.
- Bengtsson, T., & Broström, G., & Lindström, M. (2002). *Effects of conditions in early-life on old age mortality in southern Sweden, 1766–1894: Functional form and frailty*. Paper presented at the ESSHC, February 27–March 2, 2002, Den Haag, The Netherlands.
- Bengtsson, T., & Dribe, M. (2005). New evidence on the standard of living in Sweden during the eighteenth and nineteenth centuries: Long-term development of the demographic response to short-term economic stress. In R. C. Allen, T. Bengtsson, & M. Dribe (Eds.), *Living standards in the past. New perspectives on well-being in Asia and Europe* (pp. 341–372). Oxford: Oxford University Press.
- Bengtsson, T., & Lindström, M. (2000). Childhood misery and disease in later life: The effects on mortality in old age of hazards experienced in early life, southern Sweden, 1760–1894. *Population Studies*, 54, 263–277.
- Bengtsson, T., & Lindström, M. (2003). Airborne infectious diseases during infancy and mortality in later life, southern Sweden, 1766–1894. *International Journal of Epidemiology*, 32, 286–294.
- Broström, G. (2002). *MLife*. <http://tal.stat.umu.se/~gb/MLife/>. Umeå: Umeå University.
- Carlsson, S. (1968). *Yrken och samhällsgrupper: den sociala omgrupperingen i Sverige efter 1860*. Stockholm: Almqvist & Wiksell.
- Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B*, 34, 187–220.
- Deprez M. (1948). Essai sur le mouvement des prix et des revenus en Belgique au début du 19e siècle (1798–1830). Unpublished thesis, University of Liège.
- Elo, I. T., & Preston, S. H. (1992). Effects of early-life conditions on adult mortality: A review. *Population Index*, 58(2), 86–212.
- Finch, C. E., & Crimmins, E. M. (2004). Inflammatory exposure and historical changes in human life-spans. *Science*, 305, 1736–1739.
- Fogel, R. W. (1994). The relevance of Malthus for the study of mortality today: Long-run influences on health, mortality, labour-force participation and population growth. In K. Lindvall & H. Landberg (Eds.), *Population, economic development, and the environment: The making of our common future* (pp. 231–284). Oxford/New York: Oxford University Press.
- Fridlizius, G. (1989). The deformation of cohorts: Nineteenth-century mortality in a generational perspective. *Scandinavian Economic History Review*, 37, 3–17.
- Gadisseur, J. (1990). *Le produit physique de la Belgique, 1830–1913*. Brussels: Palais des Académies.
- Hornik, K. (2002). The R FAQ. Available online at <http://www.ci.tuwien.ac.at/~hornik/R/>. ISBN 3-901167-51-X.
- Jones, H. B. (1956). A special consideration of the aging process, disease, and life expectancy. In J. H. Lawrence & C. A. Tobias (Eds.), *Advances in biological and medical physics* (Vol. 4, pp. 281–337). New York: Academic.
- Kermack, W. O., McKendrick, A. G., & McKinlay, P. L. (1934). Death rates in Great Britain and Sweden: Some regularities and their significance. *Lancet*, 1, 698–703.
- Kuh, D., & Ben-Shlomo, Y. (1997). *A life course approach to chronic disease epidemiology*. Oxford: Oxford University Press.
- Lindström, M., & Davey Smith, G. (2008). A Life Course Perspective to the Modern Secular Mortality Deline and Socioeconomic Differences in Morbidity and Mortality in Sweden. In T. Bengtsson (Ed.), *Perspectives on Mortality Forecasting, Social Insurance Studies, No. 5: Cohort factors: How conditions in early life influence mortality later in life*. Stockholm: Swedish Social Insurance Agency.
- Lundh, C. (1983). “Levnadsstandarden – indikatorer och mått. Engelsk och svensk debatt om lönearbetarnas villkor 1750–1850.” Meddelande från Ekonomisk-historiska institutionen, no. 23. Lund: Dept. of Economic History, Lund University.

- Lundh, C. (1998). Servant migration in the early nineteenth century. *Journal of Family History*, 24, 53–73.
- Oris, M., & Alter, G. (2001). Paths to the city and roads to death: Mortality and migration in East Belgium during the industrial revolution. In I. Devos & M. Neven (Eds.), Recent Work in Belgian Historical Demography 19th and Early 20th centuries. Special issue. *Revue Belge d'Histoire Contemporaine*, 31(3–4), 453–495.
- Oris, M., Neven, M., & Alter, G. (2005). Individuals and communities facing economic stress: A comparison of two rural areas in 19th century Belgium. In R. C. Allen, T. Bengtsson, & M. Dribe (Eds.), *Living standards in the past. New perspectives on well-being in Asia and Europe* (pp. 373–402). Oxford: Oxford University Press.
- Preston, S. H., Hill, M. E., & Drevenstedt, G. L. (1998). Childhood conditions that predict survival to advanced ages among African-Americans. *Social Science and Medicine*, 47, 1231–1246.
- Riley, J. C. (2001). *Rising life expectancy: A global history*. Cambridge: Cambridge University Press.
- Rotberg, R. I., & Rabb, T. K. (1985). *Hunger and history: The impact of changing food patterns and society*. Cambridge/London: Cambridge University Press.
- Schuster K., & Sunder, M. (2005). Stature and twist of fate: Evidence on the biological standard of living from the lives and deaths of 19th-century Bavarian men. Manuscript, Dec. 16, 2006.
- Sköld, P. (1996). *The two faces of smallpox: A disease and its prevention in eighteenth- and nineteenth-century Sweden*. Umeå: The Demographic Data Base, Umeå University.
- Therneau, T. M., & Grambsch, P. M. (2000). *Modeling survival data: Extending the Cox model*. New York: Springer.
- Vliebergh, E., & Ulens, R. (1912). *L'Ardenne. La Population Agricole Au XIXe Siècle*. Brussels: Librairie Albert Dewit.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

