Ph.D. thesis

Demographic Surfaces: Estimation, Assessment and Presentation, with Application to Danish Mortality, 1835-1995

Kirill Andreev

Center for Health and Social Policy



Faculty of Health Sciences
University of Southern Denmark

1999

CONTENTS

PREFACE 1

CHAPTER 1 Estimating Survivors at High Ages from Data on Deaths

1.1 Introduction 3
1.2 Terminology and notation 3
1.3 Review of available methods 4
1.3.1 Extinct cohort method 4
1.3.2 Survival ratios method (SR) 5
1.3.3 Das Gupta's method (DG) 8
1.3.4 Method of Coale and Caselli (CC) 14
1.3.5 Relation of the Das Gupta to the Coale-Caselli method 17
1.4 Mortality projection methods 18
1.4.1 Age-specific decline of mortality (MD) 18
1.4.2 Projecting population and mortality trends constrained for observed death counts (DC) 20
1.5 Comparisons 21
1.5.1 Estimating absolute survivor counts 21
1.5.2 Estimating survivor population distribution 25
1.6 Conclusions 28

CHAPTER 2 The Quality of Oldest-Old Mortality Data

2.1 Introduction 31
2.2 Age heaping 32
2.2.1 Ratio of q_{80} / q_{81} 33
2.2.2 Age heaping at age 100 33
2.2.3 Lexis maps of the local test for mortality deviations 36
2.2.4 Benchmark mortality procedures 39
2.2.5 Results of the age-heaping test 40
2.3 Age misreporting 45
2.3.1 Introduction 45
2.3.2 Lexis maps of death distributions 50

2.3.3 Lexis maps of death distribution ratios 54
2.3.4 Logistic procedure 56
2.3.5 Application of logistic procedure 59
2.4 Discussion 65

CHAPTER 3 The Danish Mortality Database

3.1 Introduction 68
3.2 Database structure 69
3.3 Original data 71
3.3.1 Population 71
3.3.2 Deaths 71
3.4 Construction of the database 72
3.4.1 Deaths 72
3.4.2 Population 77
3.5 Danish demographic statistics 82
3.6 Major indicators of Danish population changes 85
3.7 Conclusion 89

CHAPTER 4 A Descriptive Analysis of the Danish Population

4.1 Introduction 91
4.2 A descriptive analysis of the Danish Population 91
4.2.1 Mortality 91
4.2.2 Mortality Progress 99
4.2.3 Compression of mortality 102
4.2.4 Sex ratio of mortality 105
4.2.5 The oldest-old population 108
4.3 Mortality differences between Denmark, Sweden, the Netherlands and Japan 110
4.3.1 Excess Danish Mortality 110
4.3.2 Analysis of cause-specific mortality 114
4.3.3 Time trends in cause-specific mortality 121
4.4 Discussion 131

CHAPTER 5 Overview of the program Lexis 1.1

5.1 Introduction 136
5.2 Program design 137
5.2.1 Contour map construction 137
5.2.2 Graphic design 139
5.2.3 The Lexis map document 141
5.2.4 Map Editor 142
5.2.5 Text editor 148
5.3 Graphical user interface (GUI) 149
5.3.1 Mouse interface 149
5.3.2 Tabbed dialog boxes 149
5.3.3 Drag and drop support 150
5.4 Making a new map 150
5.5 Technical data 151
5.6 Distribution and copyright 151

Summary 153
Danish Summary 156
References 158

APPENDIX

Appendix Table 2.1 The mortality databases used in data quality checks 164
Appendix Table 3.1 Raw population data 165
Appendix Table 3.2 Raw death counts data 166
Appendix Table 3.3 Earlier publications of Danish population statistics 166
Appendix Table 3.4 The average deviation between the genuine and interpolated death distributions for the years 1916, 1921-1940 167
Appendix 4.1 Estimating mortality progress surfaces 168
Appendix 4.2 Kernel smoothing of Lexis maps 169
Appendix 4.3 Estimating mortality ratio surfaces 169
Appendix Table 4.1 List of causes of deaths selected for the analysis of mortality differences 171

LIST OF TABLES

Table 1.1 Survivor estimates by age groups 23
Table 1.2 Rank distributions of survivor estimate methods by country 25
Table 1.3 Survivor estimates adjusted to census totals and by age group 28
Table 1.4 Rank distributions of survivor estimate methods adjusted to census totals by country 28

Table 2.1 Age-heaping defects revealed by the benchmark mortality procedure 46
Table 2.2 Fit of the logistic procedure to the proportion of deaths at age 100+ out of deaths at age 80+. Female populations, year 1980 60
Table 2.3.1 Age exaggeration in mortality databases. Ages 80+ 63
Table 2.3.2 Age exaggeration in mortality databases. Ages 80–99 64

Table 3.1 The death distribution within the age group 95–99 and in the year 1916 and 1921–1940 74
Table 3.2 The number of deaths above age 100 76
Table 3.3 Annual rates of increase in Danish life expectancy in the selected periods 89

Table 4.1 Life expectancy in the beginning of 20th century 99
Table 4.2 Proportion of the life table deaths in Denmark 105
Table 4.3 Improvements in life expectancy in the period from 1970 to 1995, 24 countries 111
Table 4.4 Decomposition of excess Danish mortality by causes of deaths for the period 1985–1993 117
Table 4.5 Decomposition of excess Danish mortality by aggregated causes of death for the period 1985–1993 120

Table 5.1 Lexis map types 139

LIST OF FIGURES

Fig. 1.1 Lexis diagram 4
Fig. 1.2 Survival ratios method 6
Fig. 1.3 Das Gupta method 9
Fig. 1.4 Mortality implied by Das Gupta method vs. mortality observed in the K-T database 10
Fig. 1.5 The ratio of mortality observed in the K-T database to the mortality implied by the Das Gupta method 11
Fig. 1.6 The relative error of survivor estimates, England and Wales, 1952, females, Das Gupta method 13
Fig. 1.7 Time rates for period 1986-1995, an aggregate of 13 countries, females 15
Fig. 1.8 Age specific mortality progress by decades, an aggregate of 13 countries, females 16
Fig. 1.9 Mortality projection for cohort crossing year 1970 at age 99. Illustration to MD method 19
Fig. 1.10 Relative errors of survivor estimates 22
Fig. 1.11 Relative errors of survivor estimates adjusted to census totals 26

Fig. 2.1(a) Ratio of q80 to q81, males 34
Fig. 2.1(b) Ratio of q80 to q81, females 35
Fig. 2.2 Local test of mortality deviations compared with adjacent ages 38
Fig. 2.3 Deviation from the general mortality pattern at a given year and age 41
Fig. 2.4 Death distribution changes in Sweden, females 52
Fig. 2.5 Swedish female death distributions from year 1900 to 1995 52
Fig. 2.6 Cumulative death distribution changes over time 53
Fig. 2.7 Ratio of death distributions 55

Fig. 3.1 Illustration of the Danish database structure 70
Fig. 3.2 Deviation between the original and the reconstructed populations 81
Fig. 3.3 Changes in the Danish population from 1835 till 1996 86
Fig. 3.4 Changes in the age structure of the Danish population 86
Fig. 3.5 Danish life expectancy 88

Fig. 4.1 Danish mortality rates....92
Fig. 4.2 Mortality progress, %....101
Fig. 4.3 Death distribution, %....103
Fig. 4.4 Sex ratio of Danish mortality....106
Fig. 4.5 Ratio of the population distribution to the average levels in 1835–1920....109
Fig. 4.6 Mortality ratio, Denmark to Sweden, Denmark to the Netherlands and Denmark to Japan....113
Fig. 4.7(a) Disadvantageous trends in Danish cause-specific mortality, males....123
Fig. 4.7(b) Disadvantageous trends in Danish cause-specific mortality, females....127
Fig. 4.8 Trends in alcohol and tobacco consumption in Denmark, Sweden, the Netherlands and Japan....133

Fig. 5.1 Translation of data matrix to Lexis map element....138
Fig. 5.2 The example of a Lexis map object....140
Fig. 5.3 The example of Plot frame object....140
Fig. 5.4 The example of Scale object....140
Fig. 5.5 The Lexis Map Appearance dialog....143
Fig. 5.6 The Plot Frame dialog....143
Fig. 5.7 The Scale dialog....144
Fig. 5.8 Illustration of the menu command Edit|Smart scale....145

PREFACE

The work presented here was carried out from 1996 to 1999. It was begun at the Center for Health and Social Policy, Odense University Medical School, Denmark and completed at the Max Planck Institute for Demographic Research, Rostock, Germany.

The dissertation includes five chapters and an accompanying CD-ROM. In the first chapter I focus on the estimation of survivors of non-extinct cohorts at advanced ages. In the second chapter I perform a quality assessment of the oldest-old databases. The databases are included in the Odense Archive of Population Data on Aging established in 1992 at the Odense University Medical School. The third chapter is devoted to the estimation of Danish mortality surfaces for all ages in the period 1835–1996. In the forth chapter I discuss the evolution of Danish mortality and compare it with that of Sweden, the Netherlands, and Japan. I also discuss the cause-specific mortality differences between these countries. In the last chapter I provide a description of the program Lexis, which I developed for producing demographic contour maps. The accompanying CD-ROM contains Lexis maps of quality evaluations, graphs of cause-specific mortality trends, and the program Lexis itself.

I am grateful to James W. Vaupel, Anatoli I. Yashin and Otto Andersen for promoting my work on this project and for being excellent supervisors. I wish to thank Roger Thatcher, Väinö Kannisto, Shiro Horiuchi and Hans Chr. Johansen for numerous discussions of demographic problems; John Wilmoth, Hans Lundström, Ewa Tabeau, Frans Willikens and Michael Væth for providing mortality data; Bernard Jeune and Axel Skytthe for their general support and encouragement. I am also grateful to Ivan Iachine for many fruitful discussions about the Lexis software. Two earlier versions of Lexis were developed by Bradley A. Gambill and Wang Zhenglian, and some of the concepts in the current version build on this earlier work.

I wish to thank Karl Brehmer for helping to edit the text of this thesis and Silvia Leek for help with the graphics. I am grateful to Kirsten Gauthier for help with the Danish translation of the summary. I also extend my thanks to the entire staff of the CHS at Odense University and of the Max Planck Institute for Demographic Research for their overall support of this project.

My research was supported in part by grants from the U.S. National Institute on Aging (P01-08761) and the Danish Research Councils. Other support was provided by the Max Planck Institute for Demographic Research.

Finally, I wish to thank my wife, Mila Andreeva, who helped me considerably with the preparation of the CD-ROM, and our two sons Maksim and Fedja for letting me use our home computer.

Rostock, Germany, 1999 Kirill Andreev

CHAPTER 1

Estimating Survivors at High Ages from Data on Deaths

1.1 Introduction

In many countries it is difficult to obtain reliable data on population counts at advanced ages from official statistics (Thatcher, 1992; Coale and Kisker, 1990; Elo and Preston, 1994). The common defect is the exaggeration of age, both the age recorded in censuses and the registered age at death, which leads to untrustworthy low mortality estimates at advanced ages. As death registration is considered much more reliable than population enumeration in censuses (Kannisto, 1994; Condran et al., 1991), my focus in this chapter is the estimation of population counts from death counts alone. This chapter reviews four published methods used to achieve this goal and suggests two new methods which can be superior in certain circumstances.

We have undertaken this study analyzing the data from the Kannisto-Thatcher (K-T) database on population and death counts at older ages (Kannisto ,1994). The primary data used for this purpose are death counts classified by single age, year and cohort for males and females and for different countries. Most of the data sets start in the year 1950 and continue up to the present time; the common age to start mortality time series is 80 but for some countries, for example Sweden, Denmark, England and Wales, Finland data are available for longer periods and cover more ages. The methods described below are developed for closed populations where the only factor responsible for the population attrition is mortality and no migration flows are present in the data[1].

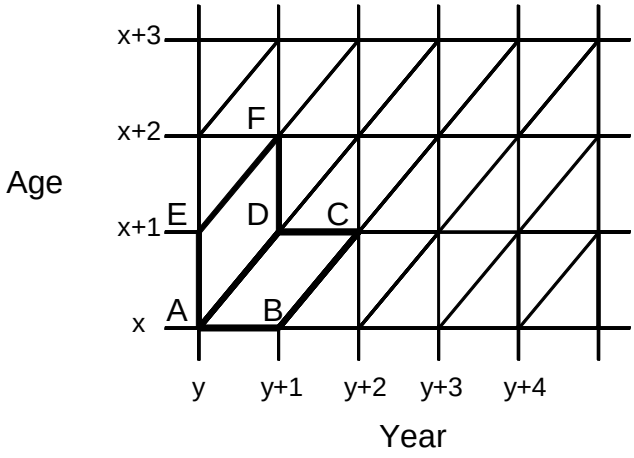
1.2 Terminology and notation

Consider a typical Lexis diagram (Fig. 1.1). The individuals $N_{x,y}$ who reached the exact age x during the calendar year y correspond to the line AB on the Lexis diagram. Thus $N_{x,y}$ is the population at risk at age x and year y. The individuals who die before reaching the age $x+1$ out of the population at risk $N_{x,y}$ correspond to the parallelogram ABCD and we denote it[2] as $D_{x,y}$.

[1] This condition is usually satisfied for ages 80+ (Kannisto, 1994)

[2] Sometimes this quantity is called "age last birthday"

Figure 1.1 Lexis Diagram



The individuals aged $[x, x+1]$ at January 1st year y correspond to the line AE on the Lexis diagram. We denote this quantity as $\tilde{N}_x$ and the corresponding death counts[3] (parallelogram AEFD) as $\tilde{D}_x$.

There are two mortality measures associated with these numbers. The first one is the age-specific probability of dying computed as a ratio of the number of deaths to the associated population at risk $q_{x,y} = \frac{D_{x,y}}{N_{x,y}}$ ($\tilde{q}_{x,y} = \frac{\tilde{D}_{x,y}}{\tilde{N}_{x,y}}$). The second measure is the force of mortality $\mu_{x,y} \approx -\ln(1-q_{x,y})$ ($\tilde{\mu}_{x,y} \approx -\ln(1-\tilde{q}_{x,y})$).

1.3 Review of available methods

1.3.1 Extinct cohort method

Let N_x be the population at risk at age x in some cohort and D_x be the number of deaths between ages x and $x+1$. At advanced ages migration is negligible and we can use the following relation $D_x = N_x - N_{x+1}$. To obtain N_x from the data on deaths we can take the sum of all deaths starting

[3] The deaths $\tilde{D}_x$ belong to the same year of birth and occur in the same year. This quantity "current year" minus "year of birth" is described by V. Kannisto as "cohort age" and by Das Gupta as "calendar age".

with age x to the highest age ω with the observed death counts $N_x = \sum_{i=x}^{\omega} D_i$. This method is known as the method of extinct generations and it was pioneered by Vincent in 1951. The application of this method is limited to the extinct cohorts and the reconstruction of the population at risk for the whole array of death counts is not possible because for the younger cohorts the number of survivors at age $\omega+1$ is not zero. If we have, for example, the data until the year 1990, the only population counts for cohorts crossing this year at age, say, 105 and above can be computed by this method. For younger cohorts crossing the year at ages below 105 it is not possible to obtain population estimates because the cohorts are not extinct. These cohorts form the lower triangle with incomplete demographic data and several methods to produce survivor estimates for non-extinct cohorts have been proposed. These methods can be considered as complementary methods for the method of extinct generations to produce estimates for the whole array of mortality data. Another advantage of these methods is that they provide alternative population estimates to the official numbers which are often being of doubtful quality (Das Gupta ,1990).

1.3.2 Survival ratios method (SR)

The survival ratios method was used extensively by Kannisto (1994) in his work on the compilation of the Kannisto-Thatcher database. Every non-extinct cohort crossing year y at age x has a 'survival ratio' that is the ratio of current survivors to the death counts in the last k years $R = \frac{\tilde{N}_{x,y}}{\sum_{i=1}^{k} D_{x-i,y-i}}$. The number of deaths is known, so the idea is that if we can estimate R from past experience then we can use it to estimate the number of survivors in the current cohort (Thatcher, personal communication). This method is based on assumption that the 'survivor ratios' or, equally, k-ages survival ${}_k\tilde{s}_{x-k,y-k} = \frac{\tilde{N}_{x,y}}{\tilde{N}_{x-k,y-k}}$ in two or more subsequent cohorts is the same. Suppose that we have to estimate survivors $\tilde{N}_{x,y}$ at age x and in the year y. Using the equation for k-ages survival yields

$$\tilde{N}_{x,y} = \frac{{}_k\tilde{s}_{x-k,y-k}}{1-{}_k\tilde{s}_{x-k,y-k}} \sum_{i=1}^{k} D_{x-i,y-i} \qquad (1.1)$$

In this method the unobserved survival ${}_k\tilde{s}_{x-k,y-k}$ is replaced by the average survival from age $x-k$ to x observed in m preceding cohorts

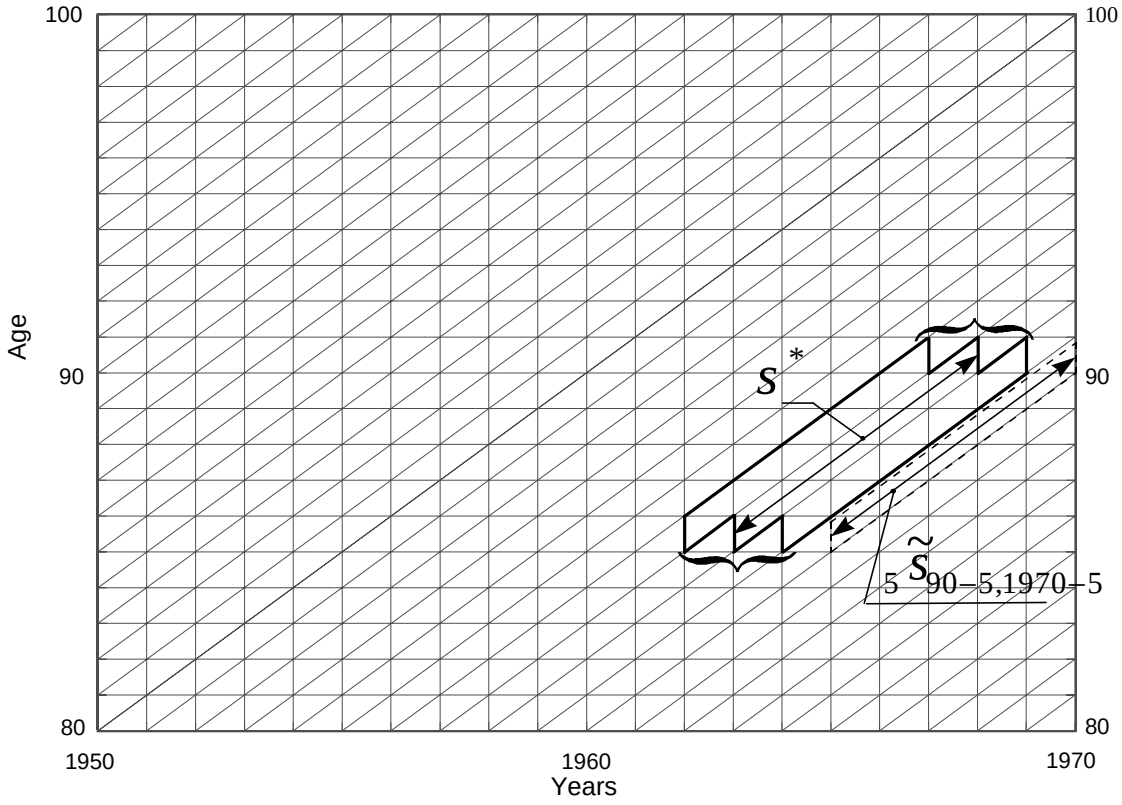
$$s^* = \frac{\sum_{i=1}^{m} \tilde{N}_{x,y-i}}{\sum_{i=1}^{m} \tilde{N}_{x-k,y-k-i}} \tag{1.2}$$

and the number of survivors is computed as

$$\tilde{N}_{x,y} = \frac{s^*}{1-s^*} \sum_{i=1}^{k} D_{x-i,y-i} \tag{1.3}$$

In order to start our estimates we need to select the highest age ω with non-zero survivor counts. To do that we compute the average number of deaths above the highest age at death in the last five years. If this number is higher than 0.5, we select this age as the first age having a non-zero survivor count and set the survivor counts for ages above it to zero. If this number is less than 0.5, we step down to the lower age and repeat the procedure. Finally, we apply the extinct cohort method for all cohorts with known survivor counts (Kannisto, personal communication).

Figure 1.2 Survival ratios method.



At this stage we are able to apply the SR method to obtain survivor estimates for ages below ω. The number of cohorts m in equation (1.2) could be one or more, and the number of years k can be taken as five or more to get the number of deaths between the age $x-k$ and the age x more than one hundred. These precautions allow us to reduce the variation in the survival and obtain stable series of survivor estimates. Once the survivor estimate for the age $\omega-1$ is computed, we repeat this procedure for age $\omega-2$. Fig. 1.2 illustrates this.

The population estimates produced by this procedure become increasingly lower than the actual numbers of survivors as we proceed to the lower ages. This phenomenon, which Kannisto calls the 'drag effect', is attributed to the mortality decline at older ages (Kannisto, 1993). He also suggested several ways to cope with this problem.

The first is that the quality of the official population estimates is believed to be acceptable at lower ages while at higher ages the population counts are considerably overestimated. In this case the SR method will produce lower survivor counts at higher ages compared with official figures. The two series of estimates intersect at about age 95 and Kannisto suggest this age as a good point to switch from the SR estimates to the official figures.

The second possibility is that an additional parameter can be introduced in the method to account for mortality decline. Kannisto suggested including correction coefficient c in equation (1.3):

$$\overline{N}_{x,y} = c\tilde{N}_{x,y} \tag{1.4}$$

The constant c is interpreted as the ratio of the odds in the current cohort to the odds of survival in the preceding m cohorts. If mortality is declining, this constant would be higher than one, and by selecting an appropriate value of c we can make a correction for the mortality decline which is not captured by equation (1.3). If accurate census totals are available for the high ages we can constrain our survivor estimates to agree with the observed numbers. In this context the constant c is a parameter of this method which can be estimated to meet the census constraints.

The third possible way to improve the method would be to estimate survival trends in the preceding cohorts and to use the prediction of the survival to produce survivor estimates. If the observed trend is not significant we can use the mean value of the survival as in the original algorithm. I should note that there is a trade-off between the significance and the reliability of the survival projection. To obtain a reliable projection we need to take as few cohorts as possible, while, on the other hand, to reach statistical significance we need to take as many cohorts as

possible. In my exercise with the female data in England and Wales significant trends in the survival were observed at ages below 95 using ten cohorts to make projection. In contrast I did not find any significant trends applying the method to the Danish population. In small countries like Denmark the survival trends are concealed by the high variation of the observed mortality rates and we need to increase the length of the time series to get significant estimates of the trends.

1.3.3 Das Gupta's method (DG)

Das Gupta developed a variant of the method of extinct generation in order to revise age distribution in the United States at age 85 and over in 1980, by race and sex. The reason for doing this was the strong evidence of age overstatement in the 1980 census population. If we use the census population and the observed death counts in the year 1980 to compute death rates at advanced ages, all race-sex groups depict an erratic bell-shaped pattern of mortality while the evidence coming from more accurate data suggests that mortality at advanced ages should increase rather smoothly with age. At the time he was working on this problem, the death counts for years 1980–88 were available and he needed to estimate the population at 1989, January 1st to apply the extinct cohort method to reconstruct the census population in the year 1980.

Before constructing the new estimates he also made some adjustments to the data. Firstly, he computed proportions of deaths by single year of age using Medicare data and distributed the total number of deaths above age 70 and for the years 1980–1988 according to these proportions. The total number of deaths was obtained from the NCHS[4]. This procedure implies a) completeness of the coverage of death registration of the elderly population provided by the NCHS and b) no misreporting of age at death into or out of ages 70 and over. The Medicare data are assumed to represent the true pattern of death distribution at ages 70+ because of the legal requirement that the enrollees must be 65 years old or older when they enroll. He also converted Medicare data from "calendar age" to "age last birthday" by averaging two successive ages, distributed deaths with unknown age and sex-race attributes and, finally, applied a 3-year moving average smoothing to correct for possible age-heaping.

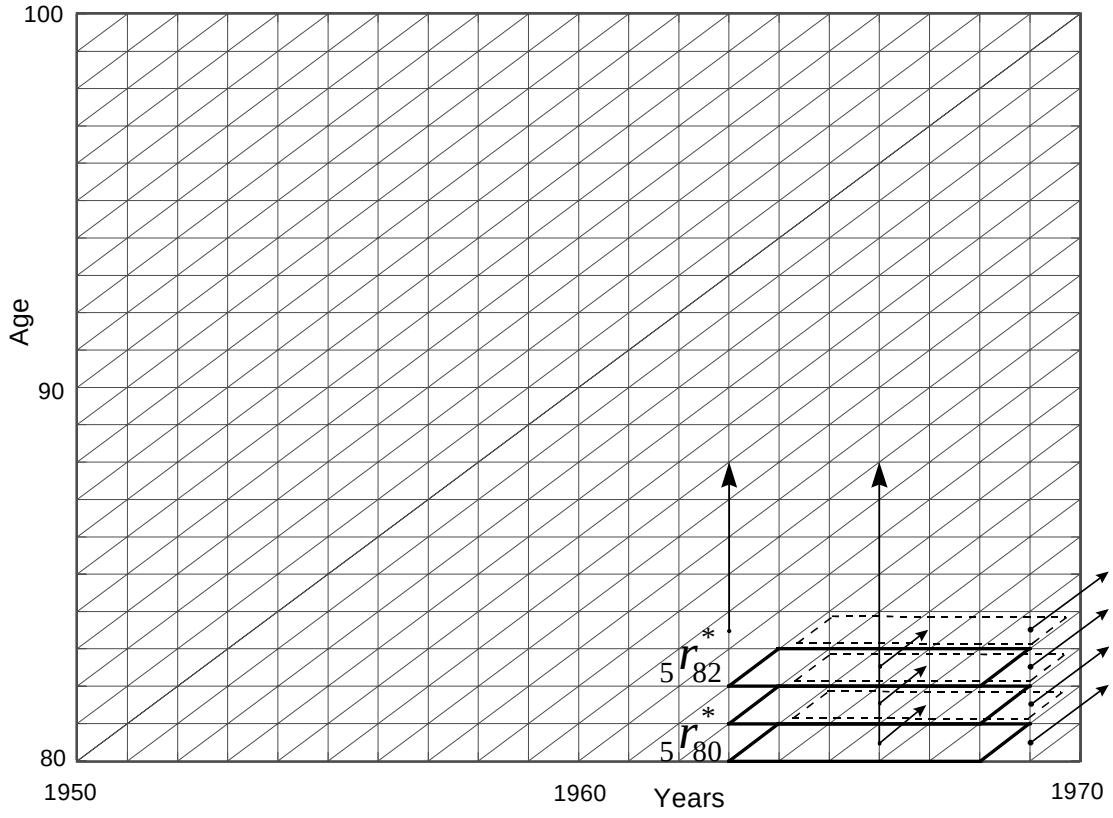
In order to apply the method of extinct generations for estimating population in 1980 given deaths up to 1988, Das Gupta computes the number of deaths which are still to come in the cohorts reaching age 85 and over in 1988. If we know, for example, the number of deaths $D_{x,y}$ (ABCD parallelogram, Fig. 1) at age x and in the year y, the deaths at the next age and in the same cohort

[4] National Center for Health Statistics

can be computed by applying the cohort death ratio $D_{x+1,y+1} = r_{x,y} D_{x,y}$. The quantity $r_{x,y}$ is not observed because no deaths are observed beyond the year 1988. Das Gupta substitutes the $r_{x,y}$ with the cohort death ratios observed in the last four years ${}_k r_x^* = \sum_{i=1985}^{1988} D_{x+1,i} \Big/ \sum_{i=1984}^{1987} D_{x,i}$. Index k is equal to four in this case and I omit it to simplify notation.

As he computes one-for-all cohort death ratios r_x^*, he applies these ratios to project deaths in the cohorts reaching age 85 and over in the year 1988. Subsequently, he uses the extinct cohort method to estimate the population in 1980. Finally, Das Gupta adjusts his estimates of population by multiplying them by a constant factor for the totals at ages 85+ to agree with the corresponding totals in the U.S. census for each race-sex group. Fig. 1.3 illustrates the Das Gupta method.

Figure 1.3 Das Gupta's method.



As Das Gupta did not apply his method to reliable population data to provide any evidence as to how the method performs and what quality of estimates should be expected, I explore his method more deeply and apply it to the data from the K-T database.

At first glance the procedure employed by Das Gupta seems to use the projected deaths to estimate the population at risk. Despite this impression, as noted by Thatcher (1993), this method

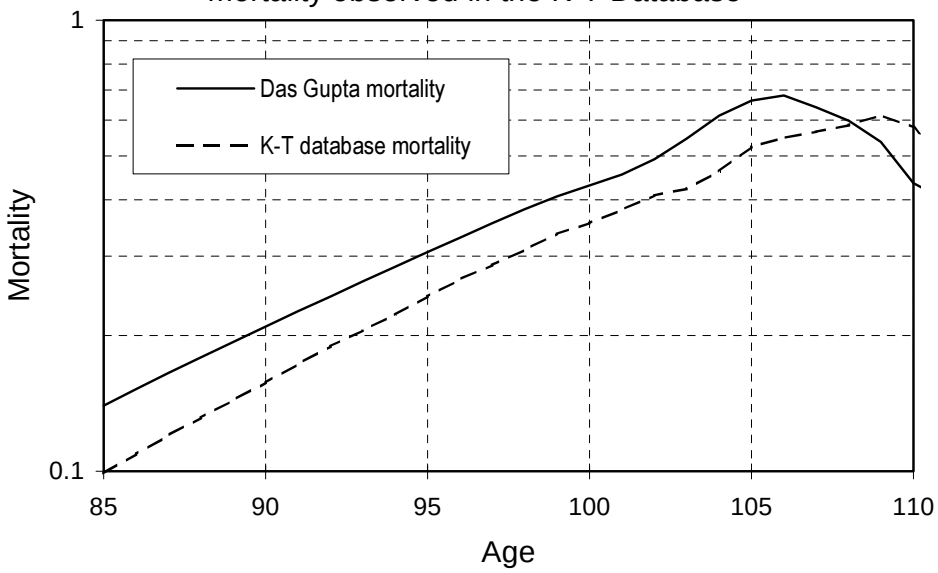
does not rely on any mortality predictions but uses the currently observed deaths to estimate the mortality in the current year.

Let $D_{x,y}$ be the death counts at age x and in the last year y with available death counts. Given the sequence of r_x^* we can compute the expected number of deaths at age $x+n$ as $D^*_{x+n,y+n} = D_{x,y}\prod_{i=x}^{x+n-1} r_i^*$. Applying the extinct cohort method we can compute the population at risk $N_{x,y}$ and, consequently, the age specific probability of dying $q^*_{x,y} = \dfrac{D_{x,y}}{D_{x,y} + \sum_{i=1}^{\omega} D^*_{x+i,y+i}}$ implied by this procedure. It can be shown that the expression for mortality implied by the Das Gupta method q_x^* reduces to

$$q_x^* = \frac{1}{1 + r_x^* + r_x^* r_{x+1}^* + \ldots + r_x^* r_{x+1}^* \cdots r_{\omega-1}^*} \tag{1.5}$$

From equation (1.5) one can see that the q_x^* depends entirely on the r_x^* sequence and the future is not involved at all. Thus, given the r_x^* sequence one is able to compute q_x^* and vice versa. The crucial condition of the successful application of the Das Gupta method would be how well the implied mortality q_x^* approximates the unobserved current death rate.

Figure 1.4 Mortality implied by Das Gupta method vs mortality observed in the K-T Database



Later in the text I show that the DG method produces higher mortality estimates of the current mortality rate if mortality in the population is declining over time. In this respect the bias in survivor estimates produced by the DG method is similar to the bias of the SR method in which the

survivors at lower ages are underestimated. In the section devoted to the Coale-Caselli method I discuss the theoretical basis for this bias.

Figure 1.5(a) The ratio of mortality observed in the K-T database to the mortality implied by the Das Gupta method. An aggregate of 12 countries. Males.

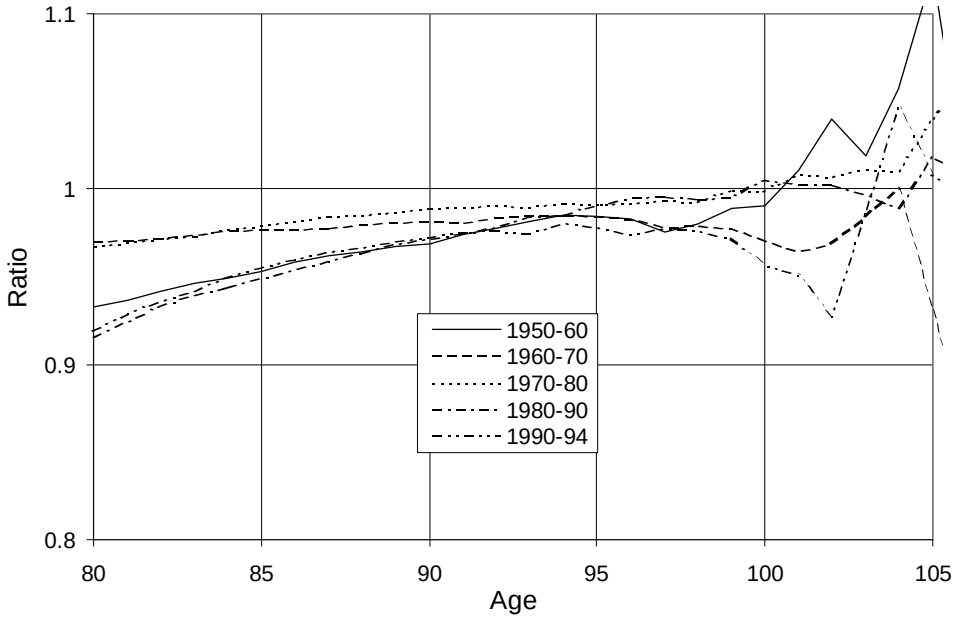
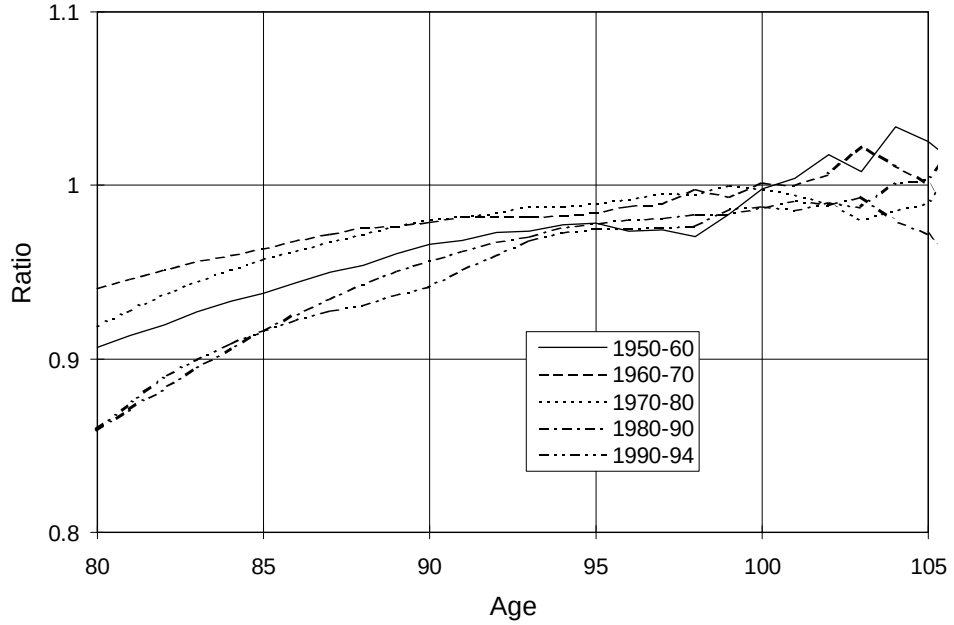


Figure 1.5(b) The ratio of mortality observed in the K-T database to the mortality implied by the Das Gupta method. An aggregate of 12 countries. Females.



The first sign of the overestimation of the current mortality rate came from the Das Gupta article itself. I took the ratios which Das Gupta published for the white males in the USA and computed the mortality implied by these ratios using equation (1.5). For the same period of time I

also computed the mortality estimates in an aggregate of 13 countries[5] from the K-T database. Fig. 1.4 shows the result.

The US mortality appears to be higher than the mortality observed in the K-T database despite the evidence that the US oldest-old mortality might be the lowest in the world (Manton and Vaupel, 1995). The bell-shaped pattern of US mortality at ages above 100 is also suspicious and can probably be attributed to age exaggeration in the death registration at advanced ages.

In order to assess the performance of this method more rigorously, I computed the decennial life tables for an aggregate of 12 countries with reliable data from the K-T database and simultaneously estimated mortality using observed cohort death ratios for the same period. Fig. 1.5 shows the ratio of the observed mortality to the mortality implied by the Das Gupta method.

As this figure shows, the mortality implied by the Das Gupta method is always higher than the observed mortality for virtually at all ages below 100. The difference is more pronounced for the younger ages and for the recent periods. The female mortality is underestimated to a higher degree than male mortality.

Using this evidence I conclude that Das Gupta method does not capture the current mortality rate very well but overestimates the current mortality rate by 10–20%, particularly at lower ages. The survivor estimates produced by this method, similar to those of the SR method, would be lower than the actually observed population counts.

What is the reason for such results? The answer lies in the very nature of the Das Gupta method itself. He applies the current death ratios to the current death counts and the implication of this procedure is that the death ratios do not change over time. Such a situation can arise only in two cases. The first is that the mortality stays constant over time. The second possibility is that the population rate of increase is equal to the mortality rate of decline, so the rate of the death counts changes over time is zero.

In order to explore whether these assumptions are satisfied, I have analyzed the death ratio trends for most of the countries included in the K-T database. The analysis shows that the ratios were steadily increasing for the period 1950–1994 rather than staying constant. Thus the death ratios in the cohort crossing age x in the current year would be higher than those observed above this age. By substituting the lower death ratios we overestimate the current mortality because the denominator of (1.5) is underestimated.

[5] Austria, Denmark, England & Wales, Finland, West Germany, France, Iceland, Italy, Japan, Netherlands, Norway, Sweden, Switzerland

To improve this method one needs to build a projection of the observed death ratio trends to the years beyond the last year with observed data as, for example, in the method of Labat and Dekneudt. Another approach would be to forecast the age-specific mortality progress function in the year for which we would like to produce survivor estimates.

Finally, we should note another important factor, discussed by Thatcher (1993), which affects the estimates of mortality in the Das Gupta method, namely, the annual mortality fluctuations. The method uses death counts only in the last year but the deaths in this year could be abnormally high because of influenza epidemics or harsh winters, like the year 1951 in England and Wales. In this case the survivor counts could be considerably overestimated despite the general downwards bias of this method. To illustrate this point, I computed the survivor estimates for the year 1952 and for the female data in England & Wales using death counts observed in 1951 and the cohort death ratios pooled over the five last years. The deviation of estimated survivor counts from observed counts is computed by means of relative error:

$$\delta_x = \frac{\hat{N}_x - \tilde{N}_x}{\tilde{N}_x} \cdot 100\% \qquad (1.6)$$

where $\tilde{N}_x$, $\hat{N}_x$ are observed and estimated survivor counts, respectively. The results are shown in Fig. 1.6.

Figure 1.6 The relative error of survivor estimates, England & Wales, 1952, Females.



The estimated survivor counts are about 15–20% higher than the actually observed numbers for ages 80–95 because of the abnormally high mortality in the year 1951. We conclude that before applying

this method one should check the mortality conditions in the last year, for example, by analyzing the total number of deaths in the last year and the adjacent years.

1.3.4 Method of Coale and Caselli (CC)

The method proposed by Coale and Caselli stems from the general relationship of the dynamics of closed populations derived by Bennett and Horiuchi (1981)

$$\mathcal{N}(x,y) = \int_x^{\infty} \mathcal{D}(t,y) e^{\int_x^t \xi(u,y)du} dt \tag{1.7}$$

where $\mathcal{N}(x,y)$ and $\mathcal{D}(x,y)$ are the population and death density surfaces and $\xi(x,y) = \frac{1}{\mathcal{N}(x,y)} \frac{\partial \mathcal{N}(x,y)}{\partial y}$ is the time rate of population increase. The equation (1.7) tells us that the population at age x in any year y can be computed from the death counts and the age-specific rates of population increase over time observed in this year. This equation cannot be applied directly because $\xi(x,y)$ are not observed in the population. Let $\nu(x,y) = \frac{1}{\mathcal{D}(x,y)} \frac{\partial \mathcal{D}(x,y)}{\partial y}$ be the time rate of death changes and $\eta(x,y) = \frac{1}{\mu(x,y)} \frac{\partial \mu(x,y)}{\partial y}$ be the time rate of mortality changes. Based on identity $\mathcal{D}(x,y) = \mu(x,y)\mathcal{N}(x,y)$ the following relation holds $\xi = \nu - \eta$. This decomposition was used by Coale and Caselli to compute the survivor estimates. They estimated $\nu(x)$ from the observed death counts in the current year y and applied a linear model for $\eta(x)$ because the mortality progress function $\eta(x)$ operating in the same year y is not observed as well as $\xi(x)$. In their model $\eta(x)$ linearly declines from some level at age 80 to zero at the highest age attained. The initial level of mortality progress at age 80 is unknown and they choose it in a such way that the calculated survivor counts are consistent with the census totals observed in the current year.

In the K-T database the death counts are available by single year of age and the following discrete approximation can be used to estimate the population at risk at age x in the current year:

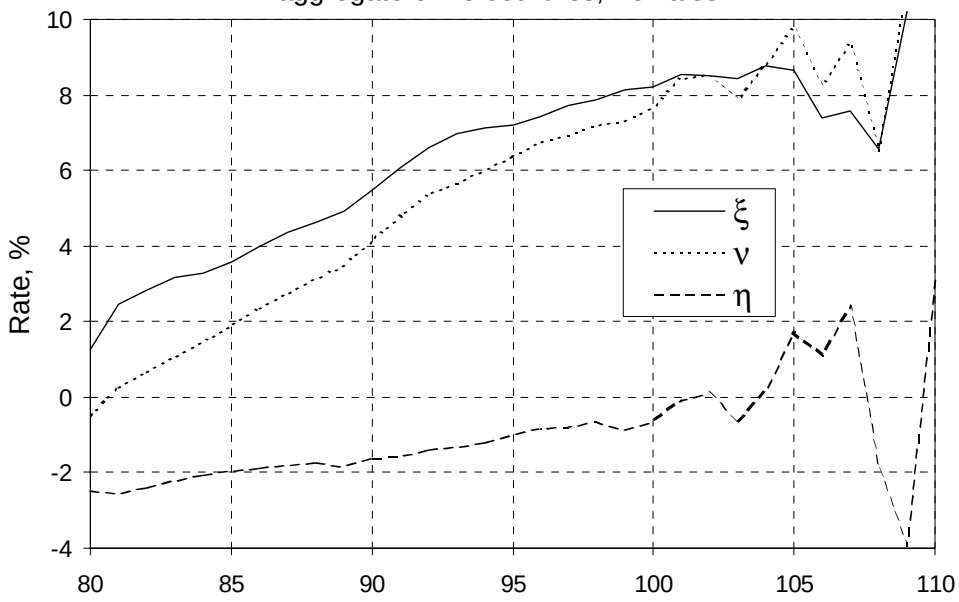
$$N_x = N_{x+1} e^{1\xi_x} + D_x e^{1\xi_x/2} \tag{1.8}$$

The method is designed to be applied only if the correct census totals are available and there are no errors in death registration. The method was developed for situations where the population structure is considered less accurate because of errors in the individual records but no gross transfers in or out of age group 80+ occurred.

Another crucial assumption of this method is the model used to describe age-specific mortality improvement. The model specification is very important in modern populations with declining mortality and increasing population of the oldest-old. Such circumstances lead to lower trend rates in death counts at lower ages compared with those for population counts and mortality. In expression $\nu = \xi + \eta$ the first term is positive but the second term is negative so they are working in opposite directions. Thus the population rates of increase at lower ages are determined mostly by the model rather than by the trends in observed death counts.

To illustrate this point I computed the rates for an aggregate of 13 countries from the K-T Database and for the period 1986–1995. Fig. 1.7 shows the result

Figure 1.7. Time rates for period 1986-1995
An aggregate of 13 countries, Females

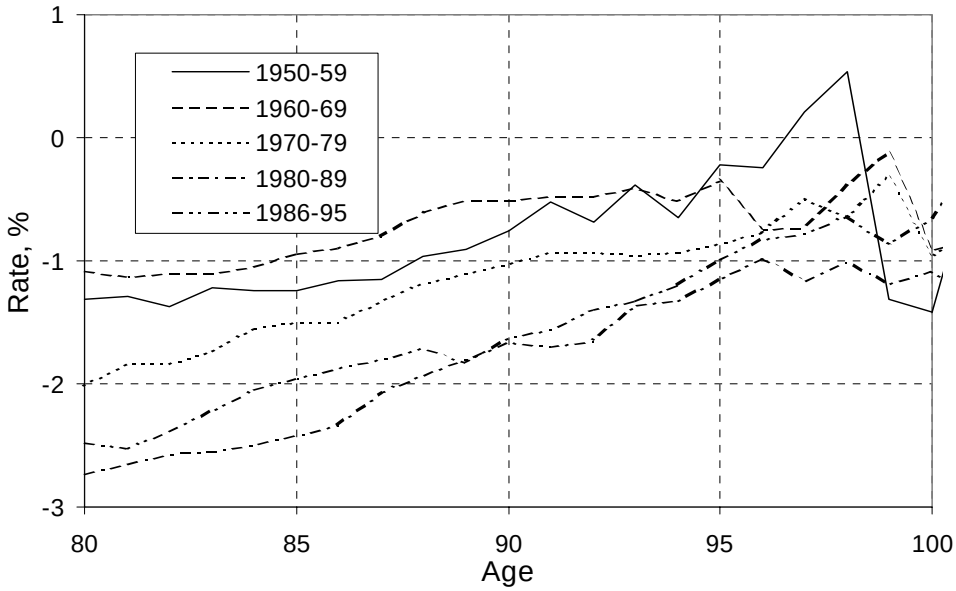


As it is seen from this figure, the ν is quite small around the ages close to 80 and the ξ is completely determined by η. If we fail to capture the age-specific pattern of η prevailing in the current year, the population estimates for the lower ages will be less reliable than the estimates for the higher ages.

The age-specific pattern of η itself is close to the pattern proposed by Coale and Caselli. The rate of mortality improvement is higher at lower ages and lower at higher ages. The validation of linear relationship is a more subtle matter and it is not discussed in their article. To shed light on the age-specific patterns of mortality improvement I applied Poisson regression to an aggregate of

13 female populations[6] from the K-T Database. The model was fitted by single age and by 10 year time periods.

Figure 1.8 Age specific mortality progress by decades
An aggregate of 13 countries, Females



The results (Fig. 1.8) indicate significant deviations from the linear relationship for earlier periods. The age-specific pattern is closer to the logistic curve than to a straight line. In the 1960s, for example, mortality improvement was about 1% at ages 80–84 and 0.5% at ages 88+, with the linear change from 1% to 0.5% at ages from 84 to 88. The most recent decades are closer to the linear pattern but still some leveling off is observed at the higher ages. In the 1980s, for example, mortality improvements at ages over 96 were almost the same.

The other important observation following from Fig. 1.8 is that mortality progress at old ages was not uniform during the period from 1950 to 1995. The rates of improvement in the 1950s were higher than the rates of improvement in the 1960s and the rates of improvement in the 1980s were higher than the rates of improvement in the most recent years.

The analysis of age-specific mortality improvement leads to the conclusion that the linear model could be a reasonable approximation for periods starting with the year 1970 while for earlier periods its suitability is more doubtful.

Finally, I should note that this method, like the DG method, is also vulnerable to the annual mortality fluctuations because it uses the death counts only from the last year. In the case of the CC method however, it is of lesser importance because the estimates obtained by this method are always constrained to the census totals.

[6] Austria, Denmark, England & Wales, Finland, West Germany, France, Iceland, Italy, Japan, Netherlands, Norway, Sweden and Switzerland.

1.3.5 Relation of the Das Gupta to the Coale-Caselli method

Having reviewed the Coale-Caselli model I turn to its relation to the Das Gupta method. Let $\lambda(x,y)=\frac{\partial \ln \mathcal{D}(x+u,y+u)}{\partial u}\Big|_{u=0}$ and $\theta(x,y)=\frac{\partial \ln \mathcal{D}(x,y)}{\partial x}$ be the rates of change of death density surface in cohort and age directions, respectively. Using the following relations between rates $\lambda=\theta+\nu$ and $\nu=\xi+\eta$ we can replace ξ with $\nu-\eta$ and ν with $\lambda-\theta$ in (1.7):

$\mathcal{N}(x)=\int_x^\infty \mathcal{D}(t)e^{\int_x^t \lambda(u)-\theta(u)-\eta(u)du}dt$. All functions are taken at the same point of time. By definition $\mathcal{D}(t)=\mathcal{D}(x)e^{\int_x^t \theta(u)du}$ and finally

$$\mathcal{N}(x)=\mathcal{D}(x)\int_x^\infty e^{\int_x^t \lambda(u)-\eta(u)du}dt \qquad (1.9)$$

If, for example, the mortality progress in the current year is zero, $\eta\equiv 0$, the equation (1.9) can be approximated by $N_x=\sum_{i=x}^{\infty} D_x\prod_{j=x}^{i}(1+{}_1\lambda_j)$. The quantity $(1+{}_1\lambda_j)$ corresponds to the Das Gupta cohort death ratios.

It also follows from this example that the Das Gupta method does not take into account the current mortality progress η, implying that it is zero. This implication constitutes the main bias of the Das Gupta method because mortality at older ages is known to have been declining over the last half century (Kannisto ,1994).

In order to improve the DG method one needs to employ some model of mortality progress prevailing in the current year, as, for example, in the CC method. If the mortality estimation is straightforward from the available demographic data, the estimation of mortality progress surface $\eta(x,y)$ is more complicated and no reliable demographic methods addressing this problem have been developed so far.

Finally, I should point out another source of errors in the CC and DG methods. The ratios computed from the observed death counts are not centered on the current year because we do not observe the deaths after this point in time. So they are substituted with the ratios computed from the last few years preceding the current year. This makes them imprecise and introduces an additional error in the estimates.

1.4 Mortality projection methods

The problem of survivor estimation is equivalent to the problem of mortality estimates in the incomplete triangle of demographic data. In this section an attempt to build mortality projection models to compute survivor estimates is undertaken. Both methods presented here use the past information to predict mortality in the cohorts with unknown survivors counts.

1.4.1 Age-specific decline of mortality (MD)

Let Y be the year for which we would like to produce the survivor estimates and ω be the highest age with non-zero survivor counts $\tilde{N}_{\omega,Y}$. The procedure to select ω is described in the section devoted to the SR method. Using the extinct cohort method we can easily compute population and consequently mortality for all cohorts crossing year Y at age ω and above. The MD method makes a mortality projection for the cohort crossing year Y at age $\omega-1$ and uses projected mortality and death counts observed in this cohort to produce the survivor estimates at age $\omega-1$. As the survivor estimates are obtained I compute the population at risk by the extinct cohort method and repeat the procedure for the age $\omega-2$.

Suppose that we need to make a mortality projection for the cohort z crossing year Y at age $X=Y-z-1$. In order to do this I fit a loglinear model for every age x from x_0 (usually 80) to $X-1$ and for n cohorts preceding z:

$$\ln \tilde{\mu}_{x,y^*-y} = \beta_{0x} + \beta_{1x} y \tag{1.10}$$

where $y^* = z + x + 1$ is the year for which I would like to make a mortality projection and y changes from 1 to n. To obtain parameter estimates I maximize the following loglikelihood function:

$$L = \sum_{y=1}^{n} \tilde{D}_{x,y^*-y} \ln \tilde{q}_{x,y^*-y} + (\tilde{N}_{x,y^*-y} - \tilde{D}_{x,y^*-y}) \ln(1-\tilde{q}_{x,y^*-y}) \tag{1.11}$$

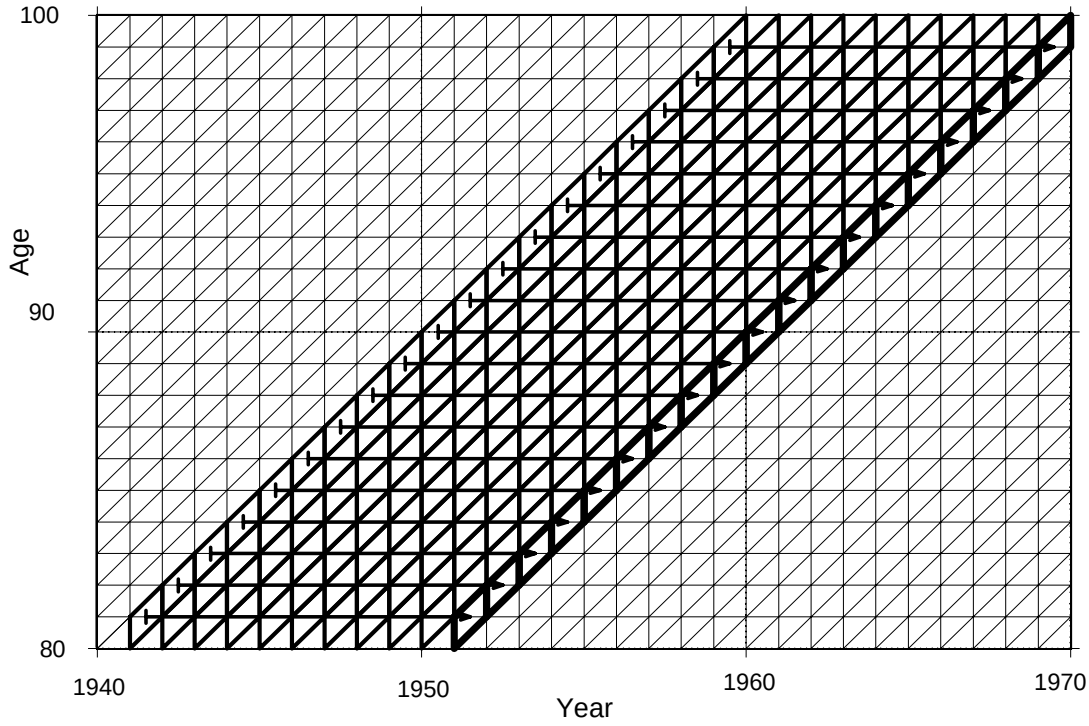
where $\tilde{q} = 1 - e^{-\tilde{\mu}}$. Once the parameter estimates $\hat{\beta}_{0x}$ and $\hat{\beta}_{1x}$ are obtained, I can compute predicted cohort age-specific probabilities of dying $\tilde{q}_x^*$ using equation (1.10). Finally, I calculate the survivor counts in cohort z:

$$\tilde{N}_{X,Y} = \frac{{}_{X-x_0}s_{x_0}}{1-{}_{X-x_0}s_{x_0}} \sum_{i=x_0}^{X-1} \tilde{D}_{i,z+i+1} \tag{1.12}$$

where ${}_{X-x_0}s_{x_0} = \prod_{i=x_0}^{X-1}(1-\tilde{q}_i^*)$ is the estimated survival ratio.

This method is illustrated in Fig. 1.9. The figure shows how the survivor estimates for the year $Y = 1970$ and the age $X = 99$ are obtained. The number of cohorts n used to make the mortality projection is 10.

Figure 1.9 Mortality projection for cohort crossing year 1970 at age 99. Illustration to MD method.



I note that because the number of cohorts n is constant, the estimates for lower ages X are less reliable because the proportion of estimated mortality rates used to make the projection increases. We can use the whole array of mortality rates available at each step but in this case the linear model might be inappropriate and we need to use a more involved procedure to predict mortality in the current cohort. I have done a pilot investigation into how well a cubic spline performs in the modeling of age-specific mortality trends. Though the mortality trend was fitted very closely by the cubic spline, the mortality projections turned out to be much worse than those produced by the model discussed above. Some additional constraints should be imposed on the spline functions to obtain the more reliable mortality projections.

Another interesting extension of this method would be the modeling of the observed mortality surface instead of age-specific mortality trends. It would allow us to obtain more precise

parameter estimates by reducing the number of parameters used to fit the past mortality trends and produce the more smooth mortality projections.

1.4.2 Projecting population and mortality trends constrained for observed death counts (DC)

Both this method and the MD method aim at projecting mortality in incomplete cohorts using the observed past information. The main differences from the MD method are a) the population levels of the preceding cohorts are modeled simultaneously with mortality trends and b) the number of cohorts n used to compute prediction is not constant while it is adapted to the observed variation in the population at risk.

Suppose as before that z is the cohort for which we would like to produce survivor estimates and the mortality in the preceding cohorts is known. Let x be the age for which we fit the DC model. The variables z and x uniquely define the year y^* which the cohort z crosses at age x. I illustrate the model by obtaining mortality projection for age x. The age index is omitted later to simplify notation. The observed past information is the series of the population at risk $\tilde{N}_y$ and the number of deaths $\tilde{D}_y$ in n preceding cohorts. I use the loglinear models for population at risk

$$\ln n_y = \alpha_0 + \alpha_1 y \tag{1.13}$$

and mortality rates

$$\ln \mu_y = \beta_0 + \beta_1 y \tag{1.14}$$

In order to estimate parameters of this model we need to maximize the following loglikelihood function

$$L = \sum_{y=y^*-n}^{y^*-1} \tilde{N}_y \ln n_y - n_y + \tilde{D}_y \ln q_y + (\tilde{N}_y - \tilde{D}_y) \ln(1-q_y) \tag{1.15}$$

subject to constraint $\tilde{q}_{y^*} n_{y^*} = \tilde{D}_{y^*}$. This constraint tells us that the projected mortality and population at risk are consistent with the observed death counts.

The number of cohorts n used to fit this model depends on the variation in the observed population at risk. I start fitting the model with some small number of n like 4 and use the likelihood ratio test to test the null hypothesis $\alpha_1 = 0$. If the null hypothesis is accepted I increase the n by one and refit the model. The procedure is repeated until the significance is reached. The final parameter estimates are used to build mortality projection $\tilde{q}_x^*$ for the cohort z at age x using equations (1.13) and (1.14). As mortality projections are obtained the rest of procedure coincides with the MD method.

1.5 Comparisons

In order to compare the different methods I computed the survivor estimates at 1970, January 1st for the female data in Sweden, Denmark, England and Wales. The death counts for these countries are available up to the year 1995 and the population at 1970, January 1st can be computed entirely by the extinct cohort method. This precaution provides us with a reliable benchmark population for the comparison of the methods.

In my analysis I distinguish two different problems. The first one is when the census totals above age 80 are not available. This is the most common case in the oldest-old mortality data. The second case is when the accurate census totals above age 80 are available from the vital statistics and I can constrain the estimates to be in agreement with these numbers. As noted by Kannisto the accurate census totals are produced only by the countries with operating population registers like Denmark or Sweden but in this case the survivor counts are known and we do not need to carry out any estimation. All methods except the CC procedure can be applied in both cases so the estimates for the CC method are presented only in the section devoted to the estimation of population distribution, not the absolute numbers of survivors.

1.5.1 Estimating absolute survivor counts

My focus in this section is the estimation of the absolute number of survivors above age 80. I applied the SR, DG, MD and DC procedures to obtain survivor estimates for the female data in Denmark, Sweden, England and Wales. The estimated series start at age 85 for the SR method and at age 82 for all other procedures. Fig. 10 shows the relative error of the estimates δ computed by equation (1.6). The estimated counts by five year age groups and the corresponding relative errors are given in the Table 1.1.

Figure 1.10 (a) Relative errors of survivor estimates
Denmark, Females, 1970, January 1st



Figure 1.10 (b) Relative errors of survivor estimates
Sweden, Females, 1970, January 1st

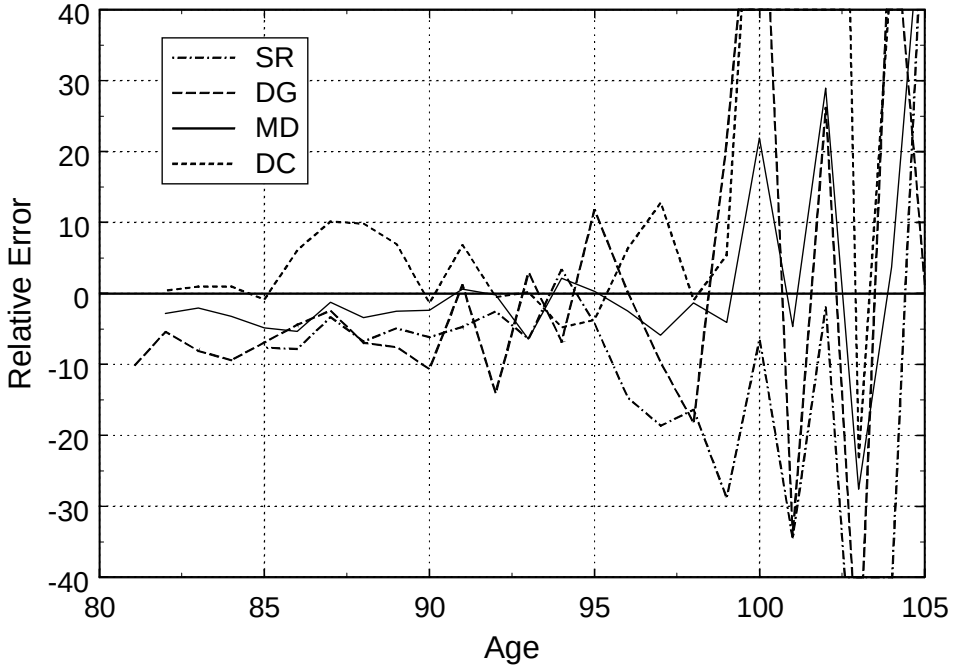


Figure 1.10 (c) Relative errors of survivor estimates
England and Wales, Females, 1970, January 1st

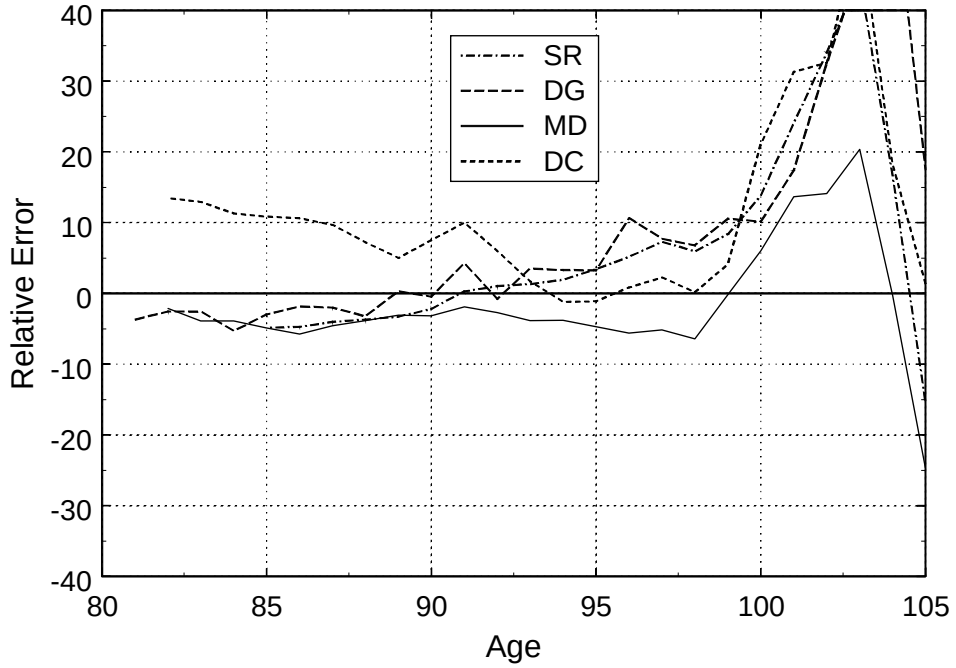


Table 1.1 Survivor estimates by age groups.

The bold items show the lowest absolute relative error in the age group. The methods were applied to female populations.

	Age	85–89		90–94		95–99		100+	
	Method	Population	Rel. Error	Population	Rel. Error	Population	Rel. Error	Population	Rel. Error
Denmark	Observed	15,891		4,099		586		27	
	SR	12,506	-21.3	3,585	-12.5	494	-15.6	32	18.6
	DG	12,571	-20.9	3,452	-15.8	451	-23.0	32	20.1
	MD	13,157	-17.2	3,625	-11.6	472	-19.5	23	**-15.9**
	DC	15,135	**-4.8**	4,039	**-1.5**	624	**6.5**	21	-22.5
Sweden	Observed	29,192		8,061		1,202		78	
	SR	27,317	-6.4	7,710	-4.3	1,055	-12.2	61	-22.2
	DG	27,568	-5.6	7,559	-6.2	1,232	2.5	94	20.6
	MD	28,089	**-3.8**	7,951	-1.4	1,179	**-1.9**	85	**8.8**
	DC	30,773	5.4	8,119	**0.7**	1,234	2.7	180	130.6
England &	Observed	227,376		68,329		11,347		935	
Wales	SR	217,680	-4.3	68,269	**-0.1**	11,928	5.1	1,112	18.9
	DG	222,509	**-2.1**	69,398	1.6	12,118	6.8	1,112	18.9
	MD	216,909	-4.6	66,342	-2.9	10,792	-4.9	1,005	**7.5**
	DC	248,334	9.2	72,558	6.2	11,395	**0.4**	1,168	25.0

The SR, DG and MD methods applied to the Danish population (Fig. 1.10(a)) produced very low survivor estimates, especially for ages below 90. The underestimation error reached up to 25–30% at ages below 85. Only DC method estimates, with relative error within a 10% band, are close to the observed counts. Such poor performance of the other methods can be explained by a rapid decline in Danish mortality in the 1960s. The average rate of mortality improvement at ages above 80 was about 2.1% compared with 1.75% in Sweden and 1.4% in England & Wales. The Danish rate of mortality improvement was especially high in the period from 1965–1970 reaching a peak of about 4% per year. It led to the sharp fall in the death counts series which were increasing until that time. This fall was caught by the DC model while all other models failed to capture this mortality decline and as a result produced the significantly lower survivor estimates.

The application of the methods to the Swedish population were more successful, with the relative errors being about 5% for the lower age groups (see Table 1.1). In this case the MD method shows the best performance compared with all other methods. The DC method produced a highly overestimated population after the age of 100 but the survivor counts below age 85 are reproduced notably well. The DG and SR methods generally produced lower survivor counts than those of the observed data and the MD method estimates.

The results of estimation of the English and Welsh population show more systematic patterns of deviations. The DG and SR procedures produce higher survivor estimates for the ages above 92 and lower estimates for the years below that age. The population below the age of 90 is best approximated by the DG procedure with an underestimation error of 2.1%. The SR and MD methods show similar patterns of deviation for this age interval but the relative errors are twice as high. The population in the age group 95–99 is the best approximated by the DC method but for all other age groups the method produces significantly higher population counts when compared with the other methods. The population in the age group 85–99, for example, is overestimated by 9.2 while it is underestimated by 4% by all the other methods.

Following my intention to rank the models in order of overall performance I computed the relative rank statistics. For each age, with the survivor estimates available for all models, I assigned a rank depending on the absolute relative error. The method producing the smallest error for a particular age receives the rank zero. Then I pool the ranks over methods and divide the results by the total rank. Thus each method receives the value between 0 and 1 indicating its relative performance. The method which consistently produces the smallest errors will receive the lowest rank and vice versa. The sum of all relative ranks is equal to one. Table 1.2 shows the results.

Table 1.2 Rank distributions of survivor estimate methods by country.

The bold items show the method with the lowest rank. The methods were applied to female populations.

	Method			
Country	SR	DG	MD	DC
Total	0.2903	0.2769	**0.1962**	0.2366
Denmark	0.3241	0.3426	0.2500	**0.0833**
Sweden	0.3254	0.2937	**0.0952**	0.2857
England & Wales	0.2319	**0.2101**	0.2464	0.3116

In the case of the Danish population the DC method received the lowest rank. That is consistent with the results shown in Fig. 1.10(a) and in Table 1.1. In the case of England and Wales the DG method shows the best performance closely followed by the SR and MD procedures. In the other two cases the lowest ranks were received by the MD procedure. This suggests that in comparison with the other methods, the MD procedure was superior.

1.5.2 Estimating survivor population distribution

In this section we assume that the correct census totals are available for ages above 85 so the survivor estimates can be adjusted by taking advantage of this additional information. This additional information allows us to produce closer approximations to the unobserved survivor counts because now we are concerned only with the estimation of the population distribution not the absolute counts. I applied the methods, including the CC procedure, to the same data as above and constrained the estimates to be in agreement with the population aged 85 and above. The survivor estimates produced by the DG, MD and DC methods were prorated to meet this total; in the SR method I used the correction coefficient to fulfill this requirement and the CC method was applied without any modifications. The parameter of this method was chosen according to the recommendations of Coale and Caselli.

The results are shown in Fig. 1.11. The first observation is that in case of Sweden and Denmark all methods reveal a comparable performance. The relative errors are centered around zero and no systematic deviations from this pattern are observed. The exception to this observation is the survivor estimates produced by the DC method for ages above 100 in the case of Swedish data. The numbers are appreciably higher when compared with the observed survivors.

Figure 1.11(a) Relative errors of survivor estimates adjusted to census totals
Denmark, Females, 1970, January 1st

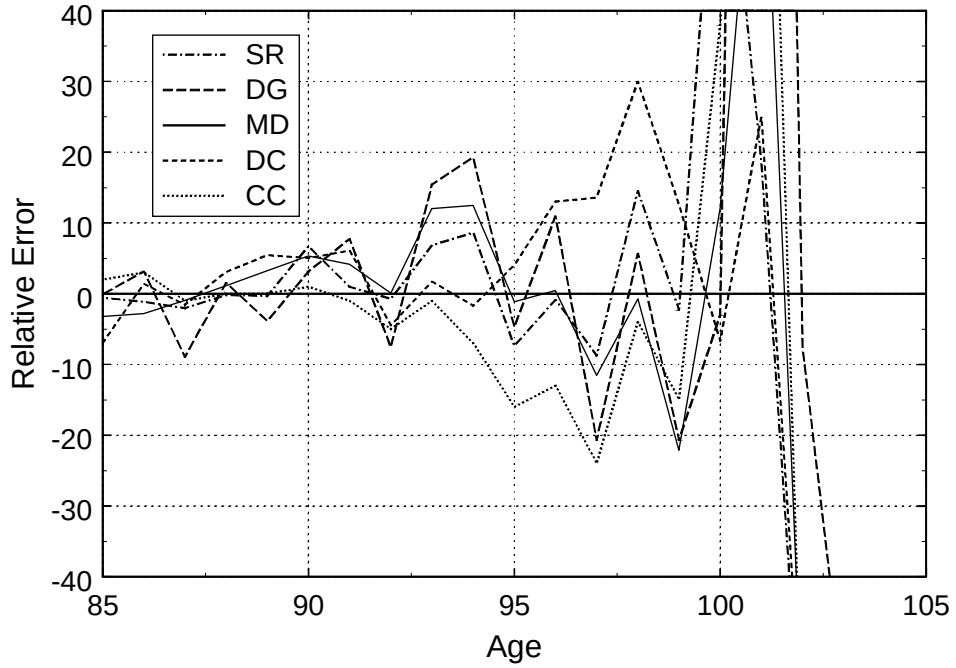


Figure 1.11(b) Relative errors of survivor estimates adjusted to census totals
Sweden, Females, 1970, January 1st

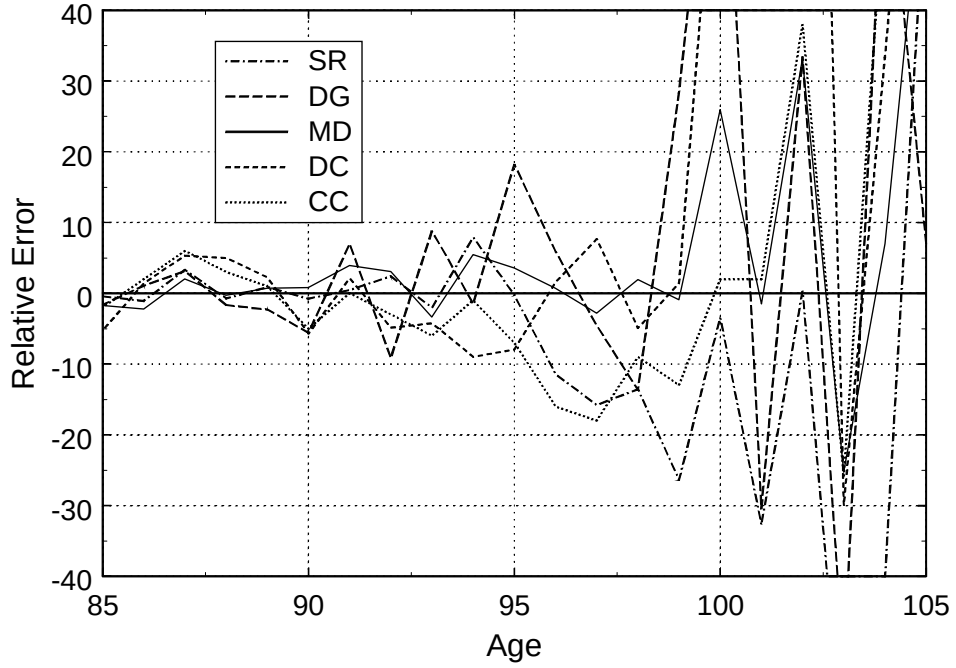
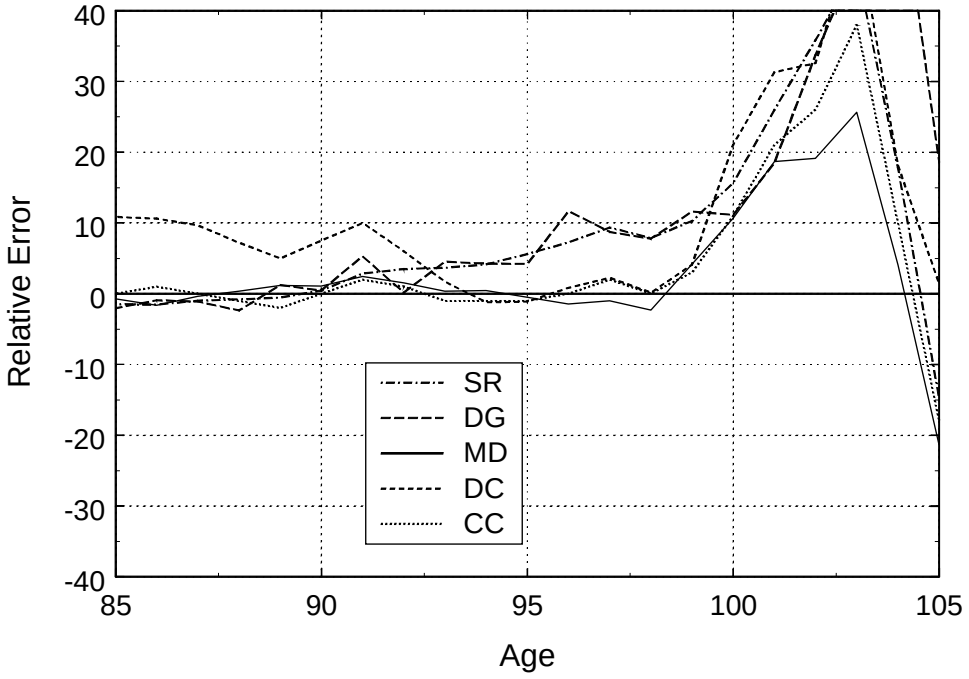


Figure 1.11(c) Relative errors of survivor estimates adjusted to census totals
England & Wales, Females, 1970, January 1st



In contrast to Swedish and Danish data systematic deviations from the actual survivor counts are observed in the case of English and Welsh data set. The SR and DG methods yield increasingly higher survivor estimates starting with age 90. The MD, CC and DC methods show the same pattern starting with age 98. In addition, the DC method differs from all the other procedures in producing higher numbers for ages below 92. In conclusion I note that the population between ages 85–98 is well estimated only by the CC and MD procedures.

Table 1.3 shows the observed and estimated population by age groups. Applied to the English and Welsh data, the CC method approximated the observed population very well compared with other methods. In the case of Denmark and Sweden there is no such outstanding model and all procedures show roughly the same performance. The average error of estimates lies approximately within 1–2% for 85–89 age group, 5% for 90–94 age group, 8–10% for 95–99 age group and 10–30% for 100+ age group. These numbers can serve as a general guideline for the magnitude of error of the estimated survivor counts.

Finally, I computed the relative rank statistics as described above. The results are summarized in Table 1.4. The rank of the MD model pooled over all data sets is the lowest among all methods, which suggests that this model is the most appropriate one for producing the survivor estimates in this case.

Table 1.3 Survivor estimates adjusted to census totals and by age groups.

The bold items show the lowest absolute relative error in the age group. The methods were applied to female populations.

	Age	85–89		90–94		95–99		100+	
	Method	Population	Rel.Error	Population	Rel.Error	Population	Rel.Error	Population	Rel.Error
Denmark	Observed	15,891		4,099		586		31	
	SR	15,747	**-0.91**	4,262	3.98	564	**-3.79**	34	9.18
	DG	15,693	-1.24	4,310	5.14	563	-3.85	40	30.61
	MD	15,693	-1.25	4,324	5.49	563	-3.94	27	-12.68
	DC	15,736	-0.97	4,200	2.46	649	10.77	22	-29.79
	CC	16,046	0.97	4,034	**-1.60**	495	-15.53	33	**6.14**
Sweden	Observed	29,192		8,061		1,202		78	
	SR	29,257	0.22	8,118	**0.71**	1,095	-8.86	62	-20.01
	DG	29,142	**-0.17**	7,990	-0.88	1,302	8.32	99	27.51
	MD	29,015	-0.61	8,213	1.88	1,218	**1.29**	88	12.37
	DC	29,419	0.78	7,762	-3.71	1,180	-1.86	172	120.44
	CC	29,595	1.38	7,792	-3.34	1,059	-11.93	87	**11.79**
England &	Observed	227,376		68,329		11,347		935	
Wales	SR	224,714	-1.17	69,980	2.42	12,165	7.21	1,128	20.68
	DG	224,588	-1.23	70,046	2.51	12,231	7.79	1,122	20.02
	MD	226,421	-0.42	69,252	1.35	11,266	-0.72	1,049	**12.17**
	DC	248,334	9.22	72,558	6.19	11,395	0.43	1,168	24.97
	CC	227,107	**-0.12**	68,449	**0.18**	11,355	**0.07**	1,077	15.17

Table 1.4 Rank distributions of survivor estimate methods adjusted to census totals by country.

The bold items show the method with the lowest rank. The methods were applied to female populations.

	Method				
Country	SR	DG	MD	DC	CC
Total	0.1935	0.2516	**0.1242**	0.2500	0.1806
Denmark	**0.1556**	0.2611	0.1667	0.2111	0.2056
Sweden	0.1476	0.2571	**0.0952**	0.2667	0.2333
England & Wales	0.2652	0.2391	0.1174	0.2652	**0.1130**

1.6 Conclusions

The problem of estimating the survivors of non-extinct cohorts from the data on deaths is equivalent to the problem of estimating mortality in the incomplete triangle of the demographic data. The initial data are the observed death counts and the mortality experience of the earlier

cohorts. Therefore, the success with which we can estimate population counts depends entirely on how well we can make a mortality projections for this incomplete triangle using the observed data. We should also draw a clear distinction between two different situations. The first situation is when the accurate census totals for the high ages are available and can be used in computations. The first assumption underlying this situation is that the census gives a correct total for high ages while the population structure is distorted by misreporting at individual ages. The second assumption is that there are no gross transfers between the high and low age groups. A different situation arises when no accurate population counts are available and we need to estimate the absolute numbers of survivors. As noted by Kannisto, the second situation is the most common case in the oldest-old mortality data, while the first one can be considered as a very special case.

Numeric comparisons I have made suggest the superiority of the MD model for both problems. Although in some circumstances the other models can show a better performance, like the CC model as applied to English and Welsh data, the MD model produces generally good results and can be applied in both situations. The SR and DG methods reveal a somewhat average performance compared with the MD method and they can be recommended for comparison with the MD procedure. These two methods have a general downwards bias as applied to populations with declining mortality and the survivors at lower ages (<95) are expected to be underestimated by about 10%.

As demonstrated above (estimating the survivors in Denmark, Females in the year 1970) all methods can fail if the mortality was declining very rapidly in the period where no direct mortality estimates are available. The underestimation errors can reach 30% at lower ages and the current mortality rate would be consequently overestimated by the same amount. In the case of Denmark the sharp mortality decline in the years adjacent to the year 1965 is manifested by a sharp decline in death count trends at ages below 90. Because there is no reason to believe that the decline is attributed to the lower cohort sizes we can assume that it caused by the drop in mortality rates and apply the DC procedure which takes advantage of the observed death counts using them as a constraint in the model. The DC model performs best in this case while the results produced by other methods are imperfect. One should be careful about applying this method in cases where there is a sharp drop in the population at risk as, for example, in the cohorts born during WWI. The performance of this method in this case is subject to further evaluation.

The method of Coale and Caselli (CC) demonstrated superior results when applied to the data for England & Wales. In other cases the performance of this method is comparable with that of other procedures. The application of this method is rather limited because it cannot be applied if

there are no census totals available. Employing this method one should pay attention to the two following problems. The first one is that the method uses the deaths counts only in the latest year and this year can be a year of exceptionally high or low mortality because of the annual mortality fluctuations. The second one is that the size of the cohorts crossing the current year can vary significantly across the cohorts because of the birth counts variation and the possible variation in the geographical coverage of the vital statistics. Though the estimates in this method are constrained to the census totals the results can be seriously distorted in both cases. The DG method is subject to the same drawbacks because it also uses the death counts from the latest year. I conclude that before applying the CC and DG methods one should check if the conditions necessary for successful application of these methods are met.

The MD method demonstrated both accurate and stable results in situations where census totals were available and not available. In both situations the method received the lowest rank aggregated over all data sets which suggests that its application is worthwhile. This method uses prior information to build mortality projections for the incomplete triangle of demographic data. Afterwards the estimates are computed on the cohort basis making it free of the drawbacks related to the annual mortality fluctuations inherent in the CC and DG methods. This method certainly should be considered if one has to choose among the different methods of estimating the number of survivors. If more data are available one can use more complicated models to depict the observed mortality trends and build a mortality projection using these models. Though, as my experience with the cubic spline shows, the models that fit the observed data better do not necessarily lead to better mortality projections.

For countries with small populations and consequently with high variation in the observed mortality and death counts, we can develop a model which fits a smooth mortality surface to the observed mortality rates and build our projection using the estimated mortality surface. I think this approach is promising for further developments and it could be extremely useful in the field of mortality projections.

Finally, I note that in order to obtain more empirical evidence, the methods have to be applied to all data sets for which reliable population and death statistics exist, particularly in countries with population registers where we can use the more recent years to test the procedures.

CHAPTER 2

The Quality of Oldest-Old Mortality Data

2.1 Introduction

It is well known that mortality estimates at old ages are often hampered by various problems (Kannisto, 1993, 1994; Thatcher, 1992, 1993; Coale and Kisker, 1986, 1990; Elo and Preston, 1994; Condran et al., 1991). Age misreporting is usually present both in censuses and in death registration statistics. The most common manifestations of the data quality problems are implausible age-specific mortality fluctuations and abnormally low mortality estimates at higher ages. The first problem is usually attributed to age heaping, the tendency to round age at death to numbers ending in five or zero. A number of tests have been developed, such as Whipple's index, to assess the plausibility of the age distribution in censuses. As I show below, the same problem occurs in death registration statistics but the heaping might occur at different numbers and prevail only in certain periods of time. The second problem is usually related to the general prevalence of age exaggeration among the oldest-old, the propensity of old people for overstating their age. This leads to the underestimation of death rates at older ages and to a tendency for them to level off or even decline with age. It is commonly recognized that this problem becomes more severe as age increases and, further, that older data are more prone to contain errors than more recent statistical data.

Although age exaggeration is the most common form of age misreporting, there are also other patterns of age misstatement that can lead to abnormally low mortality estimates at advanced ages (Preston et al., 1997). Even if the proportion of death counts misreported from a particular age to the lower age group is higher than the proportion misreported to the upper age group, the resulting mortality estimates at higher ages would be lower than the actual values. The reason for this is that the distributions of age at death taper off very rapidly at older ages, so the absolute number of deaths allocated into the upper age groups would be higher than those allocated to the lower age groups, thus producing heavier-tailed death distribution and, consequently, lower mortality estimates at older ages.

In this chapter I assess the quality of data collected in the Kannisto-Thatcher (K-T) database on population and death counts at older ages (Kannisto, 1994). The population at risk in this database is estimated by the extinct cohort method (Vincent, 1951), so the mortality estimates are

free of the errors introduced by population statistics, which are commonly recognized as being of lower quality than death registration statistics (cf. e.g. Kannisto, 1994; Condran et al., 1991). My main concern was to assess the plausibility of the death distributions and the resulting mortality estimates since errors in mortality estimates can only reflect errors in the death distributions - not errors in the population at risk. The K-T database contains data from about thirty countries classified by sex, cohort, age and year at death. Most of the data sets start in the year 1950 and at age 80 but for some countries, such as Sweden and Denmark the data are available for all ages and start well back in the 19th century. The huge volume of data requires a compact presentation, which can best be accomplished by advanced visualization tools such as those discussed by Vaupel at el. (1998). I decided to present the results of my analysis with the help of Lexis maps as they allow us to reduce considerably the volume of presented material while keeping the details untouched. All maps produced during the work on this project were created with the help of the program Lexis, which was developed by K. Andreev.

2.2 Age heaping

As noted above, age heaping is a well-known problem in demography and a number of tests, such as Whipple's index, were developed to assess the plausibility of the age distribution. The direct application of these methods for oldest-old mortality is not possible, however, because of a rapid change of age distribution and a high degree of stochastic variation at older ages. Thus, new tests for age heaping are needed.

Consider a cohort of individuals for which the number of deaths D_x and population at risk N_x are recorded by the single age x. Suppose, that proportion α of deaths with true ages $x-1$ and $x+1$ is reported to be of age x. Thus, the death counts are misreported from two adjacent ages with probability α. Suppose also, that the population at risk is computed by the extinct cohort method. In this case the number of deaths reported at age x will be $D_x^* = D_x + \alpha(D_{x-1} + D_{x+1})$ and the population at risk will be $N_x^* = N_{x+2} + D_x + D_{x+1} + \alpha D_{x-1}$. These equations show that both the death counts and the population at risk are distorted by misreporting. Let $q_x = \frac{D_x}{N_x}$ be the actual age-specific probability of dying and $q_x^* = \frac{D_x^*}{N_x^*}$ be the probability observed in the population with

inaccurate data. Taking the derivative $\frac{\partial q^*}{\partial \alpha}$ we see that mortality rates at ages with heaping will be higher ($q_x^* \geq q_x$) than those observed in an error-free population while mortality at adjacent ages will be lower ($q_{x\pm1}^* \leq q_{x\pm1}$). Using this observation a number of statistical tests and graphical methods for age-heaping tests can be constructed.

2.2.1 Ratio of q_{80} / q_{81}

The first method suggested by Kannisto (1993) is based on the ratio of the age-specific probabilities of dying at ages 80 and 81: q_{80} / q_{81}. Because most of the countries in the K-T database begin with age 80 the comparison cannot be based on the ages surrounding 80. The mortality increase observed at age 80 in the K-T database suggests that the ratio should be close to 0.915 for males and 0.9 for females[7]; any high upward deviations from this ratio should lead us to suspect age heaping. Fig. 2.1 shows the ratios calculated from the decennial life tables.

Abnormally high ratios are observed in New Zealand (Maori) and Portugal up to the year 1965. Less striking, but still evident age heaping is observed in Ireland, Spain (until the year 1970), New Zealand (non-Maori) (until the 1970s), England and Wales (in the 1910s), Latvia (until the 1970s) and in the Netherlands (NSO)[8] in the middle of the 19th century. A relatively small amount of age heaping is observed in Canada (1950s), Australia (1960s) and Estonia(1950s).

2.2.2 Age heaping at age 100

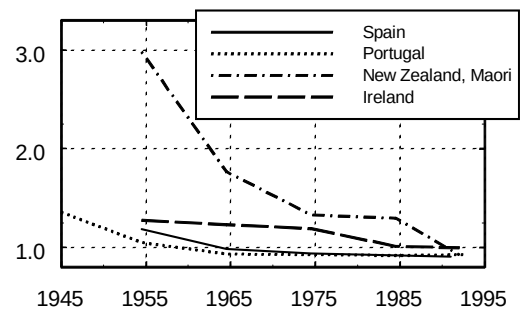
The second method suggested by Kannisto (1993) deals with age heaping at age 100. He argues that the ratio $\frac{4q_{100}}{q_{98} + q_{99} + q_{101} + q_{102}}$ is slightly below 1.0 if mortality at these ages increases according to the Heligman-Pollard model (Heligman and Pollard, 1980). He applies this procedure together with a graphic display to the years 1970–1990 and finds some evidence of age heaping for France, West Germany and New Zealand and to a lesser extent for Switzerland and Australia. He also observes that age heaping is more pronounced for males than for females.

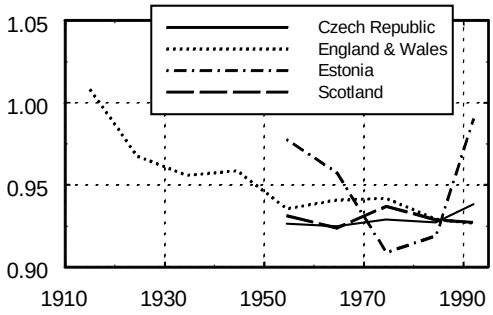
[7] These numbers are from a period life table computed for Nordic countries for the years 1950-1990.

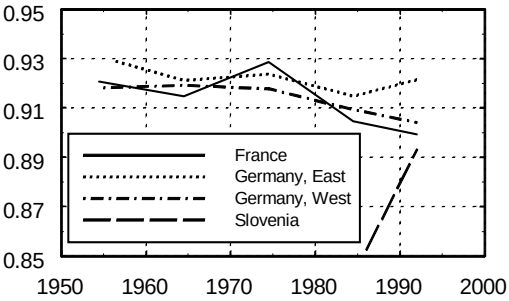
[8] The data for the Netherlands are originally from the Central Statistical Bureau of the Netherlands. The construction of the mortality database was carried out by Tabeau at al. (1994). The data for the Netherlands collected by Kannisto are stored in a different database.

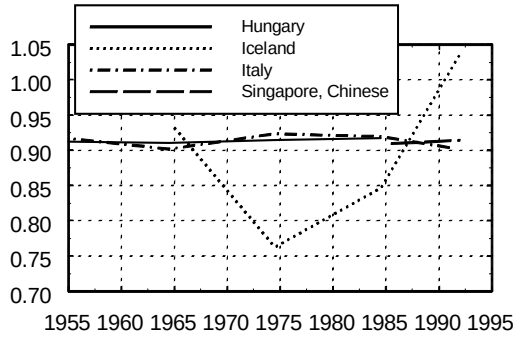
Figure 2.1(a) Ratio of q_{80} to q_{81}

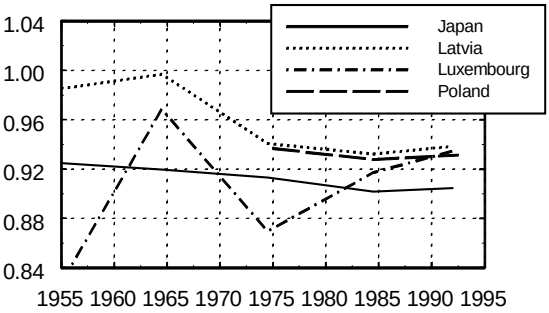
Kannisto-Thatcher database, Males

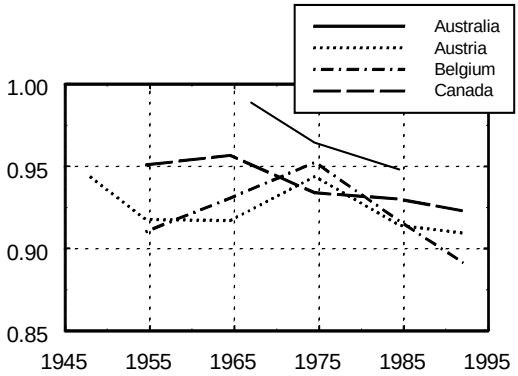


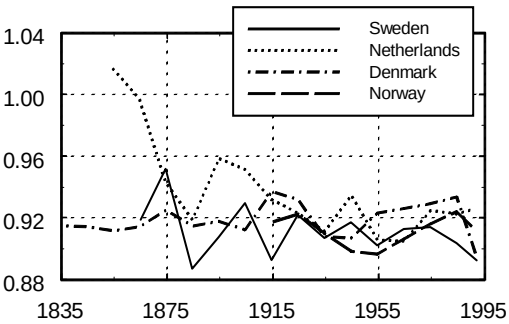












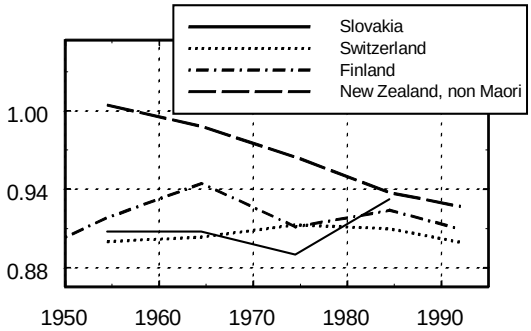
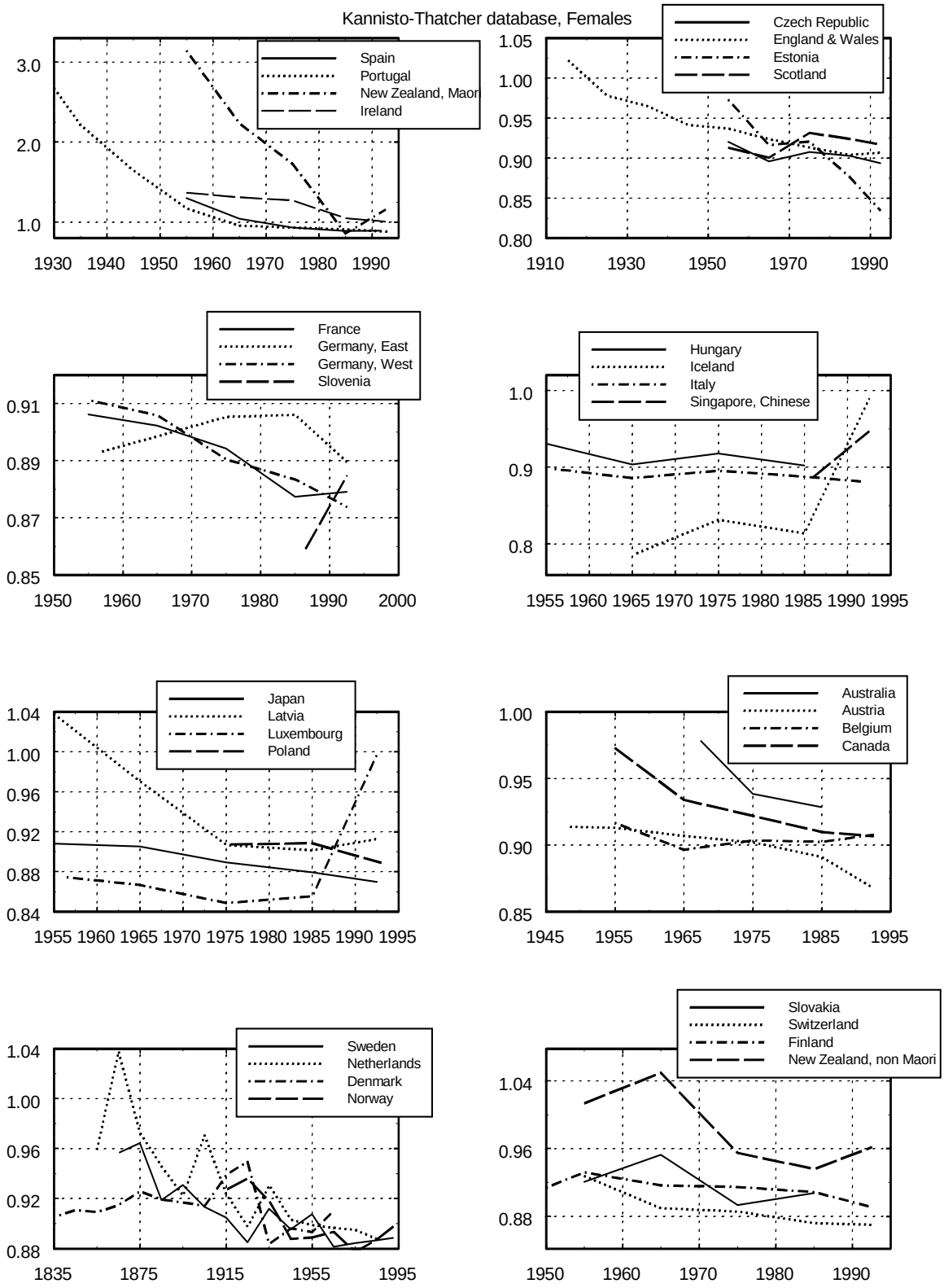


Figure 2.1(b) Ratio of q_{80} to q_{81}

Kannisto-Thatcher database, Females



2.2.3 Lexis maps of the local test for mortality deviations

As briefly discussed by Vaupel et al. (1998), Lexis maps may be useful in data quality checks. Using the Lexis map display device we can check every value in the database for possible errors and easily see at a glance where the problems occur. The basic assumption of this method is that mortality surfaces change smoothly over age and time, so the mortality at given age and year is approximately the same as the mortality at adjacent ages or years.

Let $D_{x,y}$ be the number of deaths at age x and year y (this quantity is depicted by a rectangle on the Lexis diagram) and $T_{x,y}$ be the corresponding total time lived by all individuals of age x and in the year y. A good approximation of $T_{x,y}$ would be the number of persons aged $[x, x+1]$ in the middle of the year y (Chiang, 1984). We can use the following statistics to test whether the mortality in the year y and age x deviates significantly from the mortality observed at surrounding ages and years:

$$X = \frac{m_{x,y} - \frac{1}{n}\sum_i m_i}{\sqrt{Var(m_{x,y}) + \frac{1}{n^2}\sum_i Var(m_i)}} \tag{2.1}$$

where $m_{x,y} = \frac{D_{x,y}}{T_{x,y}}$ is the mortality rate observed in the year y and age x, and $Var(m_{x,y}) = \frac{D_{x,y}}{T_{x,y}^2}$ is the large sample variation of the mortality rate estimate (cf. e.g. Keiding, 1990). I did not specify exactly which ages and years should be taken to compute m_i because this depends on what particular test one has in mind.

The statistic has an asymptotic standard normal distribution and the large deviations of the tested mortality rate from the average mortality rate observed at adjacent years and ages correspond to the large deviations of X. However, the population at risk may be so large that even very small deviations can be viewed as statistically significant, especially if the population of the country is large, such as Japan, for example. Therefore, in order to limit our attention to large deviations in the mortality rates we can additionally compute the relative deviation from the observed mortality rate as

$$R = \frac{m_{x,y}}{\sum_i m_i} - 1 \tag{2.2}$$

Mortality $m_{x,y}$ is considered suspicious if both (2.1) and (2.2) are significant. Equation (2.1) will help to remove the stochastic variation at higher ages while equation (2.2) helps to eliminate the

insignificant mortality fluctuations at lower ages where $T_{x,y}$ is large. The test can be applied to every $m_{x,y}$ covered by the data set (except those on the boundaries) and a Lexis map of the suspicious mortality rates can be produced.

I applied this procedure to all databases listed in Appendix Table 2.1 and produced Lexis maps for each database - for males and females separately. These maps are included on the accompanying CD-ROM[9] and can be viewed with the program Lexis, which is also provided on the CD-ROM.

I divided all outcomes of X and R statistics into five groups:

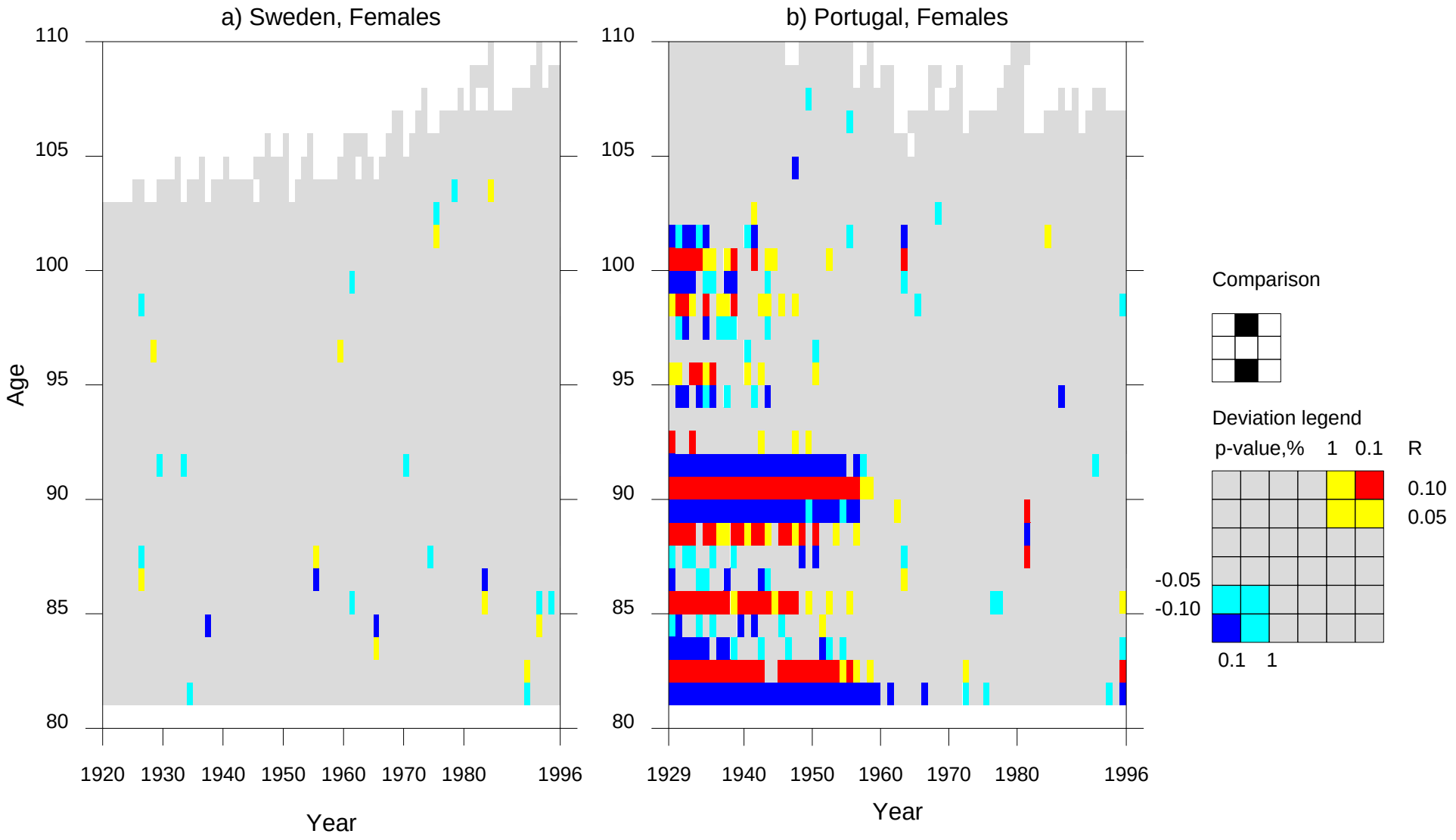
1. large negative deviations: $R < -0.1$ and X is significant at the 1% level (two-tailed test)
2. small negative deviations: $R < -0.05$ and X is significant at the 1% level (excluding group 1 outcomes)
3. large positive deviations: $R > 0.1$ and X is significant at the 1% level
4. small positive deviations: $R > 0.05$ and X is significant at the 1% level (excluding group 3 outcomes)
5. non-significant deviations

Fig. 2.2 shows an example of quality check maps for the female data in Sweden and Portugal. As indicated in the box labeled “Deviation legend”, the color blue is used to display large negative deviations, cyan for small negative deviations, yellow for small positive deviations, and red for large positive deviations. Areas with the non-significant deviations are depicted in gray, and white indicates the years and ages where it was impossible to carry out this test. The small box labeled “Comparison” shows the ages and years used to compute X statistics. My intention was to test the data for age heaping, and I performed the comparison using two adjacent ages as indicated by two black rectangles; the tested mortality rate is located in the middle of this rectangle.

The interpretation of the maps is straightforward. The colors red and yellow highlight the ages with age heaping; the mortality observed at these ages is much higher than that observed at adjacent ages. The colors blue and cyan accentuate the ages with the lower mortality, i. e. the ages contributing to heaping. Looking at Fig. 2.2 I conclude that the most significant age heaping is observed in the female population of Portugal at ages 82, 85, 90, 95, 100 and in the years before 1960, while the Swedish data pass this test.

[9] The files for this test are in folder \quality\ltmd. The Lexis maps for males start with ‘m’ and for females start with ‘f’; the rest of the file name coincides with the abbreviation listed in Appendix Table 2.1.

Figure 2.2 Local test of mortality deviations compared with adjacent ages.



The method can be also applied to test for possible year heaping. In this case the comparison can be performed with two adjacent years but unlike the previous case, the results are ambiguous. It is well known that the period effects of mortality can be highly exceptional such as those produced by the Spanish Influenza epidemic in 1918. Annual fluctuations of the oldest-old mortality caused by severe winters or influenza epidemics are also quite common. Such abnormal mortality conditions would appear on the quality maps as grave data errors. One should always check for historical explanations and if these explanations fail, then the reliability of the data should be brought into question.

Sometimes the comparison with adjacent ages will give rise to some intriguing cohort patterns, such as for male data in Japan (BMD). These cohort effects could be the manifestations of exceptional demographic conditions for the cohorts in question rather than errors in the data. The cohort born in 1945 in Japan appears to have lower mortality rates after age five than the two adjacent cohorts. As noted by Shiro Horiuchi (personal communication) the year 1945 was one of the worst years in the history of 20th-century Japan, with the lowest fertility rates and the highest rates of infant mortality: this cohort suffered from exceptional demographic conditions earlier in life. In order to get an answer to the question why the mortality of the cohort was lower later in life, a more refined analysis is needed.

2.2.4 Benchmark mortality procedures

The method described above has the advantage of being easily computed, so the instant quality check can be performed quickly and then conveniently presented in the form of a Lexis map. However, the test has two serious limitations. The first one is that the increase in mortality with age is not taken into account. This is specially important for older ages, where mortality curves are particular steep. The second limitation is that the method does not allow for any quality evaluations at the edges of the observed mortality surfaces, which is of special concern at age 80, the starting age for most data sets included in the K-T database. In order to overcome these difficulties, the following model has been developed.

The basic idea is to compare mortality at a given age with the general mortality pattern prevailing in the current year. We can fit the general mortality pattern by using a model and estimate the deviation of mortality at a given age from the mortality predicted by the model.

A number of mortality models used to describe the age pattern of oldest-old mortality have been analyzed by Thatcher et al. (1998). They came to the conclusion that the gamma-Makeham

model[10] is the most appropriate model for depicting the pattern of oldest-old mortality. Using their results the benchmark mortality model can be specified as follows.

Suppose that mortality in the current year follows the gamma-Makeham model

$$\bar{\mu}(x) = \frac{ae^{bx}}{1+\sigma^2 \frac{a}{b}\left(e^{bx}-1\right)} + c \tag{2.3}$$

and x^* is the age for which we would like to estimate the deviation from the mortality predicted by the model. We can use the following model of mortality to fit the observed death rates:

$$\mu(x) = \bar{\mu}(x), \quad x \notin [x^*, x^*+1] \tag{2.4}$$

$$\mu(x) = e^{\beta}\bar{\mu}(x), \quad x \in [x^*, x^*+1]$$

In this case parameter β is a measure of mortality deviation at a given age from those implied by the model and the quantity e^{β} can be interpreted as the relative risk at age x^*. We can also use the likelihood ratio procedure to test if the estimate of β significantly deviates from zero. The model can be fitted to every year and age covered by a database and the estimates of the relative risks can be presented by a Lexis map.

2.2.5 Results of the age-heaping test

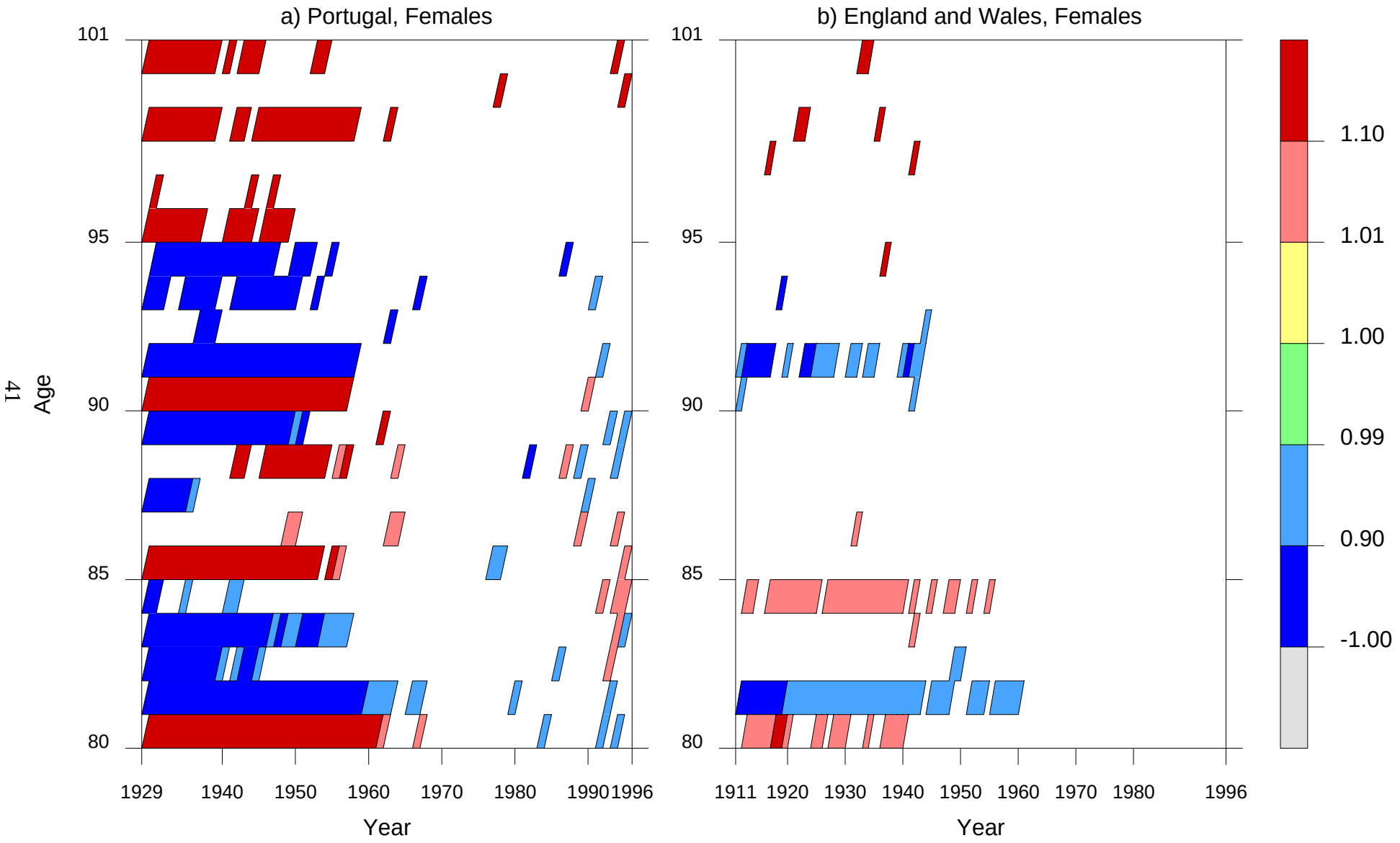
I applied the benchmark mortality test to all the data sets listed in Appendix Table 2.1 (except the USA) and produced Lexis maps of the relative risk estimates. Overall, 74 databases were processed and for each database a Lexis map was constructed. The whole set of Lexis maps comprises about 85,000 estimates of the relative risks and it can be found on the accompanying CD-ROM[11].

The model was fitted to every age from 80 to 100 for all years covered by the database. Fig. 2.3 shows the Lexis maps for the female data in Portugal and England & Wales. One horizontal parallelogram on the Lexis map corresponds to the estimate of the relative risk ($e^{\hat{\beta}}$) at the given year and age. If the estimate of β is not significant at the 1% level, the parallelogram was painted white; otherwise shades of red are used for upwards mortality deviations and blue for downwards mortality deviations. Sometimes it is impossible to carry out the estimation because the data are not available for a particular year and age. Such cases appear as gray rectangles on the Lexis maps.

[10] There refer to it as a "logistic model"

[11] The maps are located in \quality\bmmt folder.

Figure 2.3 Deviation from the general mortality pattern at a given year and age.



The benchmark mortality procedure was applied to test for age-heaping defects in the K-T database.
The gamma-Makeham model was used to fit the general mortality pattern. Significance level 1%.

Now I turn to the interpretation of the Lexis maps. For illustration I discuss the results of the age-heaping test carried out for the female population of Portugal. The results are shown in Fig. 2.3(a), which comprises altogether about 1,400 parameter estimates. Mortality at ages where the number of the reported deaths is abnormally high is considerably elevated and this elevation is manifested by the red horizontal lines stretching over these years. Estimates of the model show that mortality until the 1960s and at ages 80, 85, 90, 95, 98 and 100 is significantly higher than the general mortality pattern fitted by the model. Starting with the 1970s the horizontal lines disappear as a result of improvements in the official statistics.

The long blue horizontal lines disclose the ages with lower mortality. The significant fraction of death counts at these ages has been misclassified due to age heaping, so the number of deaths at these ages is much lower as it would be expected in the error-free population. This leads to significantly lower mortality estimates, which are depicted by shades of blue in Fig. 2.3(a).

Fig. 2.3(a) provides us with an instant data quality assessment for the whole female Portuguese database. I conclude that until the 1960s, the Portuguese data seem to be of a little use because of the high degree of age heaping. In the middle of the 1960s the quality of the data improved significantly and reliable mortality estimates can be computed starting with the 1970s.

Fig. 2.3(b) illustrates another pattern of age heaping which is frequently observed on the Lexis maps. Sometimes mortality at a given age appears to be lower than the expected mortality over a period of several years, but no mortality elevations are observed at other ages. In the female data for England and Wales, mortality at age 91 is consistently lower for the years from 1911 to 1950. A possible explanation for this pattern is that an appreciable proportion of deaths at age 91 was reported as having occurred at age 90. Since the death distribution decreases very rapidly at advanced ages, the absolute number of misreported death counts might be insufficient for an appreciable elevation of mortality at age 90 while mortality at age 91 is significantly reduced. One could call this low-degree age heaping, which can be observed at various other ages as well. The female populations of England & Wales, Australia and Canada are the most striking examples of populations having such a defect in the data.

This procedure resulted in noticeable diagonal mortality deviations also in the case of female data for Japan, Italy and male data for Belgium. In order to judge whether it is genuine cohort effects that are being observed rather than the heaping of date of birth data, a more elaborated analysis is required. We need to look at the cohort mortality at lower ages which are not covered by the databases, explore the vital registration system operating in the period with the observed cohort

effects, check for earlier life mortality conditions of the cohorts, etc. Such an analysis would beyond the scope of this chapter.

Finally, attention must also be given to the data problems revealed by this test which cannot be directly classified as age-heaping problems. I have certainly found some defects in the data but the mechanisms behind the errors seem to be different those that lead to age heaping. For example, female mortality in Italy at ages 97–99 and in the years 1952–70 appears to be significantly higher than the mortality predicted by the model. As there is no immediate explanation for this result, this case requires further clarification.

Another example is the elevation of mortality at ages 91–96 and in the years 1864–1875 in the Norwegian data, in both male and female populations. If we look at the age-specific mortality profile in the year 1865, say, we will see a sharp rise in mortality at ages 90–95 and a comparable mortality drop at higher ages. This defect appears as the red blemishes on both Norwegian maps. Additionally, significant age heaping is found at age 80 in the period 1846–1930, especially in the female population. I conclude that for earlier years the Norwegian data are most certainly flawed but this cannot be completely attributed to age heaping.

Table 2.1 summarizes the results of this age-heaping test. The most severe data problems were found in Portugal, Spain and Ireland. Age-heaping defects in Portugal exist until the 1970s, when the quality of statistics was improved considerably. The data for recent years are much more accurate. The similar improvement in Spanish statistics took place somewhat later, close to the 1980s, but the data for age 100+ are not available after 1981. For earlier periods the data for both countries are of a little use because of the notable age-heaping defects. The Irish data have problems with age heaping throughout the database coverage - from 1950 to 1992 - especially at ages 80, 84 and 86.

Another group of countries for which I found inaccuracies in the data includes Australia, Canada, Chile, France, England and Wales, West Germany, Italy, Poland, Norway (NSO), New Zealand (Maori) and New Zealand (non-Maori). Detailed information concerning which years and ages are affected by this defect can be obtained by exploring the country-specific Lexis maps. Here, I comment briefly on some countries listed in Table 2.1.

In the Australian, Canadian and English & Welsh populations we observe long horizontal blue lines which are produced by the lower mortality. In the female data for England and Wales the mortality drop at age 81 ranges from 10% in the years 1911–1920 to 5% in the 1950s. As the quality of vital statistics improves over time the defect diminishes and completely disappears after the year 1960. As discussed above it seems possible that this pattern can be attributed to mild age heaping at

age 80. In those years, the age at death was registered in complete years, whereas later the exact year of birth of the deceased was registered. Recording the year of birth offers less temptation to use round number. It seems probable that age heaping exists in the data for the earlier years (R. Thatcher, personal communication).

In the Australian data the defect is most evident in the female population and in the years 1966–1980. The drop in the death rate at age 81 is approximately 10%; a value comparable with those of England and Wales. Canadian female data exhibit the similar defect both at age 81 and 91 for the period prior to 1970. The male population in Canada is affected to a much lesser degree and the defect is virtually invisible on the map. It also worth noting that an analogous mortality drop at age 81 is observed in New Zealand (non-Maori) but the estimates of mortality deviations are not significant at the 1% level due to the small size of the population. If one creates a map that includes all estimated relative risks, the defect becomes apparent. As before, it is observable only for the female population.

My quality check of Polish data reveals significant downward mortality deviations at ages 98 and 99. This can be attributed to heaping at age 100 but it is impossible to check this hypothesis due to the lack of data above age 100. I conclude that the death rates at ages 98 and 99 are distorted but closer inspection would require additional work.

The quality checks of French and German (West) data disclose significant age heaping at age 100. The number of deaths reported at age 100 appear too high in all populations, and it is well depicted by the red horizontal lines at age 100 apparent on all maps. And in the German data, the age 99 is affected as well. The ages contributing the deaths reported for the age 100 are expected to have a lower mortality, and this drop in mortality is also evident on the maps.

In the Italian data a perceptible rise of mortality was detected at age 99 for males and at ages 97–99 for females in the period from 1952 to 1970 as depicted by the red blemishes on the Lexis maps. The mortality is up to 50% higher than the predictions of the test and the female population is affected more than the male population. Another interesting feature of the Italian maps is the elevated mortality of the 1880 cohort. Mortality in the age range 80 to 90 is 5% higher than the general mortality level in this period. The phenomenon is clearly manifested in both Lexis maps by the red diagonal lines in the 1960s, and it vanishes after the age 90. The values of the estimated parameters for the male and female populations are comparable.

The data for New Zealand were collected separately for the Maori and non-Maori populations. For the non-Maori population the drop in mortality at age 81 is reminiscent of the quality checks for England and Wales. Similar drop for the Maori population appeared to be non-

significant at the 1% level and it is not visible on the maps. In addition, the very flat and sometimes declining age-specific mortality schedules observed during the fit of the model to the Maori population are indicative of a considerable overstatement of age at death in these data. This leads to the conclusion that the Maori data in New Zealand are of doubtful quality even no significant mortality deviations were detected.

The goal of the analysis presented here was to provide a quality assessment of the K-T database as concerns unusual age specific mortality fluctuations typical of age heaping. As it turns out, many of the data sets are affected by this defect but the degree of age heaping varies substantially from country to country and from year to year. At the present time a researcher should use the erroneous data with caution: he or she should conduct a prior evaluation to determine how sensitive final conclusions are to the errors that might be present in the data. Finally, I should note that I made no attempt to correct the faulty data here, as this would be beyond the scope of this work.

2.3 Age misreporting

2.3.1 Introduction

As noted by Kannisto (1993) it is a characteristic of old people and their family members to be proud of their age and they are thus apt to overstate it. It also widely recognized that this tendency becomes stronger with increasing age and that men usually exaggerate their age to higher degree than women (the latter observation is less well accepted; Dechter et al. (1991), for example, found a different pattern in their analysis). This natural tendency to overstate one's age has been alleged to account for the age exaggeration in censuses. Because age at death is not self-reported, one should expect that the extent of age exaggeration in death registration statistics is lower than in population statistics. Here deliberate misreporting is likely to be overweighed by errors due to a lack of knowledge of the decedent's age. Under such circumstances gross transfers both in upper and lower age groups are possible. As discussed by Preston et al. (1997), even if the proportion of deaths misreported in the lower age group is higher than the proportion misreported into the higher age group, the mortality estimates based on such data might be still lower than the actual numbers because the death distribution decreases very rapidly with age. Rosenwaike and Logue (1983) also support the idea that the age at death on the death certificates could be misreported in both

Table 2.1 Age-heaping defects revealed by the benchmark mortality procedure.

The Lexis maps related to this table are stored in the folder '\quality\bmmt'. The Lexis map for the data set 'Australia, Males' is stored in the file 'maustl.lex' where 'm' stays for males ('f' for females) and 'austl' is the abbreviation of this data set as shown in Appendix Table 2.1.

Country	**Data Problems**				**Comments**
	Males		**Females**		
	Year	**Age**	**Year**	**Age**	
Australia	1965–80	81	1965–80	81	Age heaping is more evident in the female data set than in the male data set
Austria					No defects
Belgium					No defects
Canada			1950–65	81, 91	
Chile	1983–88	99			
Czech Republic					No defects
Denmark					The data were interpolated from 5-year age groups before the year 1910
England & Wales	1911–50 1911–70	80, 84 81, 91	1913–37 1911–80	80 81, 84, 91	Age heaping is more evident in the female data set than in the male data set
Estonia					No defects
Finland					No defects
France	1950–85	100	1950–80	100	
Germany, East					
Germany, West	1960–82 1968–95	99 100	1956–82 1969–88	99 100	
Hungary					No defects
Iceland					No defects
Ireland	1950–80 1950–87	80 81, 84, 91	1950–86 1950–80	80, 81, 84, 86, 90, 91	Noticeable age heaping
Italy	1952–70	99	1953–67	97, 98, 99	Cohort effects

Table 2.1 (cont.).

Country	Data Problems				Comments
	Males		Females		
	Year	Age	Year	Age	
Japan					
Japan (BMD[12])					Cohort effects are found in the female population
Latvia					No defects
Luxembourg					No defects
The Netherlands (NSO[13])					No defects
Netherlands					No defects
New Zealand (Maori)					Age heaping is observed at age 80 but it is not significant at the 1% level. The fitted mortality curves are very flat. In some years mortality declines with age: year 1950 for males; years 1959, 1968, 1975 for females
New Zealand (non-Maori)					Age heaping at age 81 is found in the female population but not significant at the 1% level.
Norway					No defects
Norway (NSO[13])	1846–60	80	1846–80 1846–1930	80 80	High mortality elevation is observed in the years 1860–80 and at ages 88–96; both for males and females Age heaping
Poland	1989–95 1976–84	98 99	1971–96	99	Age heaping

[12] Berkley Mortality Database

[13] National Statistical Office

Table 2.1 (cont.).

Country	Data Problems				Comments
	Males		Females		
	Year	Age	Year	Age	
Portugal	1940–63 1940–52 1941–54 1929–54	80,81 85, 88, 90 98 100	1929–67 1929–56 1929–56	80, 81 83, 85, 90 95, 98, 100	Noticeable age heaping
Scotland					No defects
Singapore, Chinese					No defects
Slovakia					No defects
Slovenia					No defects
Spain	1950–73 1950–67 1950–61 1950–74	80, 81, 91 83, 84 88, 90 96, 98, 99	1950–74 1950–69 1950–69	80, 81, 84 88, 90, 91 93, 98, 99	Noticeable age heaping
Sweden (BMD[12])					No defects
Sweden					No defects
Switzerland			1973–88	100	Insignificant age heaping in the female population

directions. In their study they found a greater tendency for the reported age to be older than the actual age of decedent.

The data collected in the K-T database were acquired from death registration statistics, and the population at risk was computed by the extinct cohort method. Thus, the errors in the mortality estimates can only reflect errors in the reported death counts and, whatever the exact mechanism of the misreporting, we expect erroneous data to manifest themselves in lower observed mortality rates, a suspiciously low slope of age-specific mortality profiles, bell-shaped patterns of age-specific death rates, higher proportions of death counts at higher ages compared with the data from reliable sources, implausible time trends in the death rates and the percentiles of the death distributions, etc.

A number of quality tests aimed at detecting age exaggeration problems were suggested by Kannisto (1993). The first procedure he calls the 'pyramid' test, which is a visual test of death distributions. He argues that mortality at ages close to 100 is roughly 50% and that the ratio of number of deaths at age 100 to those at age 101 should be close to 50% as well. For younger ages, where mortality is lower, this ratio is less than 50% and approximately equal to 65%. Thus, one can plot the death count pyramids based on decennial life tables and compare them with the expected pattern. Long-tailed distributions will indicate data which are likely to have been affected by age exaggeration.

The second test suggested by Kannisto is the ratio of deaths at age 100+ to those at age 85+ and the ratio of deaths at age 105+ to those at age 100+. High values of these ratios indicate that the data is of dubious quality, and ratios close to the average levels mean that the data is of acceptable quality. He also provides two benchmarks for comparison: a) the ratios which are observed in a stationary population with the age-specific mortality schedule computed from an aggregated life table for all countries and b) the ratios which are observed in a stable population growing at 3.5% per a year with the same mortality schedule. The ratios observed in growing populations with the same mortality will be lower than in the stationary populations because the population distribution is steeper. Furthermore, the ratios can be adjusted to the life expectancy at age 80: the higher the life expectancy is, the larger the observed ratio should be.

The third test discussed by Kannisto is the analysis of age-specific mortality schedules. In his work he analyzed the slope of the logit of mortality rates. As noted above, age exaggeration results in an underestimation of mortality levels, especially at the highest ages. Thus, the slope of mortality computed from the faulty data would be lower than that of estimates from accurate data - this can be ascertained by fitting this model to the raw data. The model turned out to fit oldest-old mortality very well for recent periods (Thatcher, 1998), and it has the advantage that the logits of

mortality rates can be easily computed from life tables. There is also some evidence that mortality progress has been higher in recent decades at lower ages than at higher ages (Kannisto et al., 1994). If this pattern of mortality progress has in fact been prevailing during this period, the mortality curves observed in the most recent years will be steeper than the mortality curves observed in the 1950s, for example. Consequently, we should observe a negative correlation between the level of mortality at age 80 and the slope of the mortality curve. This correlation can provide an additional correction to the analysis of mortality slopes. In his work, though, Kannisto did not find any significant correlation between the slope of mortality and the mortality level at age 82.

The last method proposed by Kannisto is the analysis of the sex ratio of deaths at advanced ages. He argues that age overstatement is more common for men than for women and that it leads to abnormal sex ratios at advanced ages. In his analysis Kannisto computes the sex ratio of deaths at ages 80–99 and 100+ for the period 1970–90 and concludes that the data for New Zealand (Maori) are most certainly affected by age exaggeration.

The quality check methods proposed below can be viewed as an extension of Kannisto's work aimed at providing a more detailed and comprehensive analysis of mortality databases. I will focus on the analysis of the observed death distributions because death counts are the basic data used for computing the death rates at older ages. I will also demonstrate the usefulness of Lexis maps in this type of demographic analysis.

2.3.2 Lexis maps of death distributions

The death distribution observed in a given year depends on the mortality rates and population structure prevailing in this period. In the 20th century mortality at advanced ages has declined steadily while the population of elderly has grown remarkably. Fig. 2.4 illustrates the changes in the distribution of deaths in the 80+ female population of Sweden arising from these demographic conditions.

At the beginning of the century mortality was rather high and only 15% of Swedish female deaths occurred after age 80. As indicated by the short-dashed line in the panel A, the proportion of deaths at age 80 is the highest among all series and the distribution of deaths in this period decreases remarkably rapidly with age. A snapshot of the death distribution in the period 1985–95 reflects a pattern quite different from that of, say, beginning of the 20th century or in the 1950s. The mortality decline which took place in the twentieth century shifted the mode of the death distribution beyond age 80, at the same time compressing the distribution itself around the mode. At the present time about 60% of all female deaths in Sweden occur after age 80, which is a four-fold increase since the

beginning of the century. The distribution observed in the 1950s takes a somewhat intermediate place between those observed at the beginning of the century and the present period. Interestingly, the proportion of deaths at age 87 have not changed at all since 1900.

The Lexis map shown in Fig. 2.5 provides more reach and a more detailed picture of the evolution of Swedish female death distribution over the 20th century. The death distribution was computed by single year and for all ages starting in 1900. The first important observation that we need to be aware of for further analysis is that the modal age at death has been increasing since 1900; the most major changes took place in the 1960s (gains prior to this period were close to zero). This observation helps to explain the remarkably different patterns of death distributions shown in Fig. 2.4(a). Another important observation is that the contour lines at ages after 80 have been uniformly increasing since 1900, providing us with a simple pattern of changes in percentiles of death distribution over time. In other words, the cumulative death distribution for ages after 80 computed for recent years lies over those computed for the preceding years: the pattern is illustrated in Fig. 2.4(b).

Based on this regular trend, the self-consistency check of oldest-old data can be constructed as follows. For each year we simply compute the observed distribution of deaths and plot the results as a Lexis map. The irregular trends in contour levels will disclose the imperfect data. The cumulative death distributions were computed for the female data in Portugal and the resulting contour map is presented in Fig. 2.6(a). For comparison, a similar map for the female population of Sweden is shown in Fig. 2.6(b).

On comparing the Portuguese map with the Swedish map, we can see at once that the age at death is highly exaggerated in Portugal in the period prior to 1970. This is clearly demonstrated by the decreasing contour lines at ages after 95 in the period from 1929 to 1970, which is opposite to the pattern observed in the Swedish map - and it is also contrary to our expectations.

The method discussed here has the advantage that it can be instantly computed and an overall assessment of data quality can be easily obtained. The drawback to this procedure is that the only severe problems with age misreporting can be safely detected. For this reason I developed more refined methods of analysis, which are discussed below.

Figure 2.4 Death distribution changes in Sweden, Females.

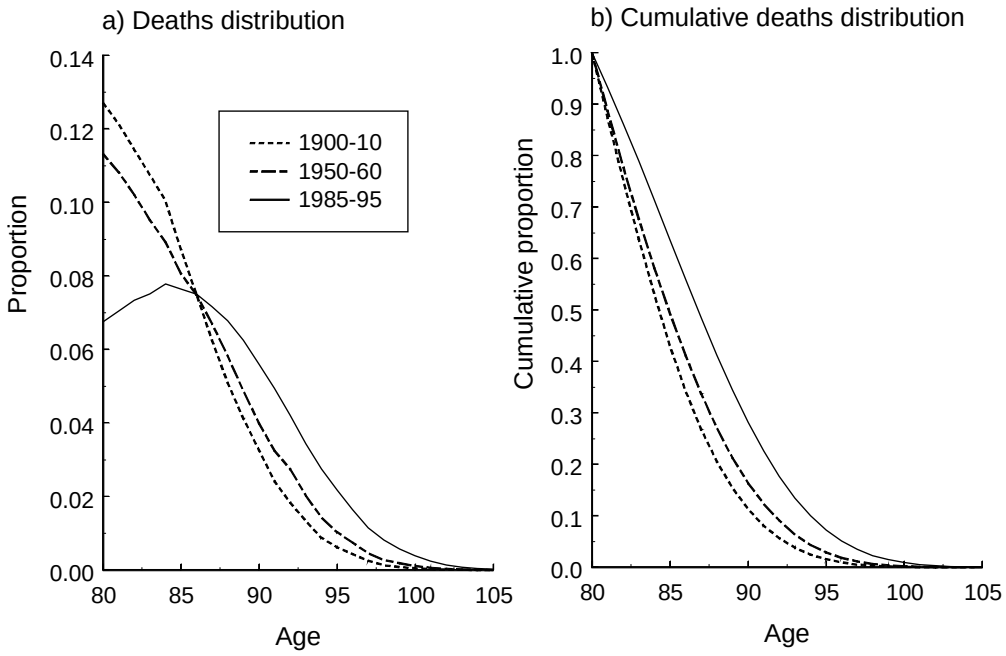


Figure 2.5 Swedish female death distributions from year 1900 to 1995.

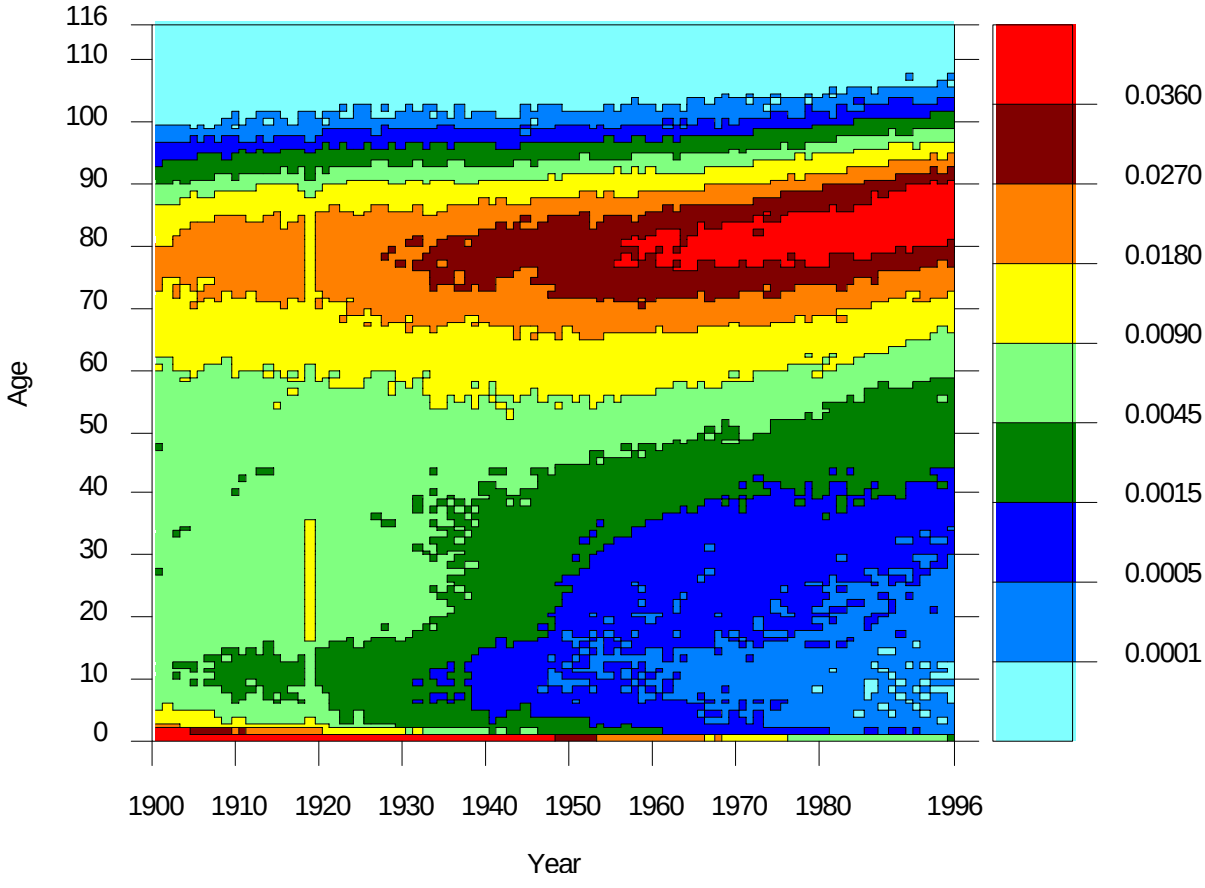
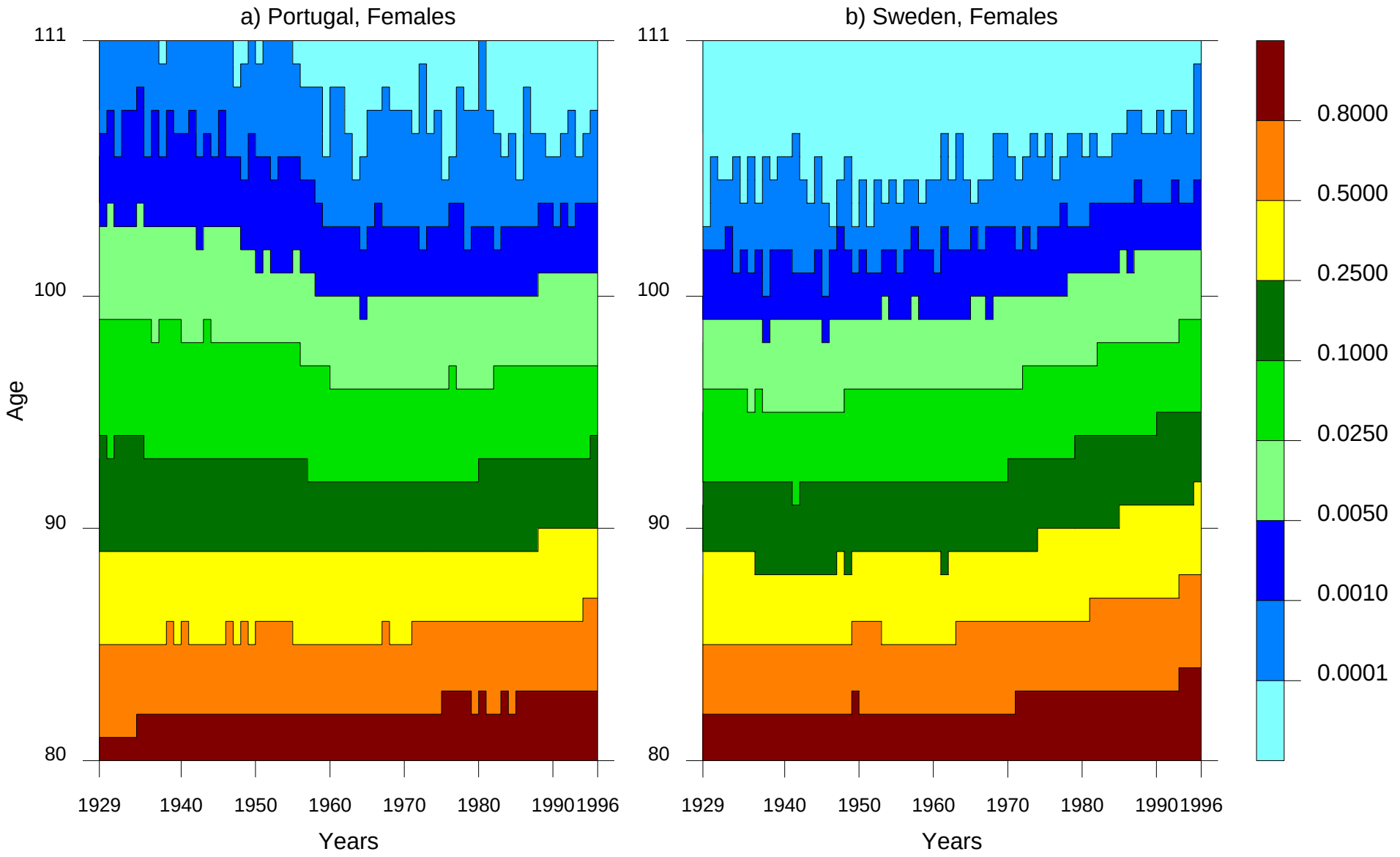


Figure 2.6 Cumulative death distribution changes over time.



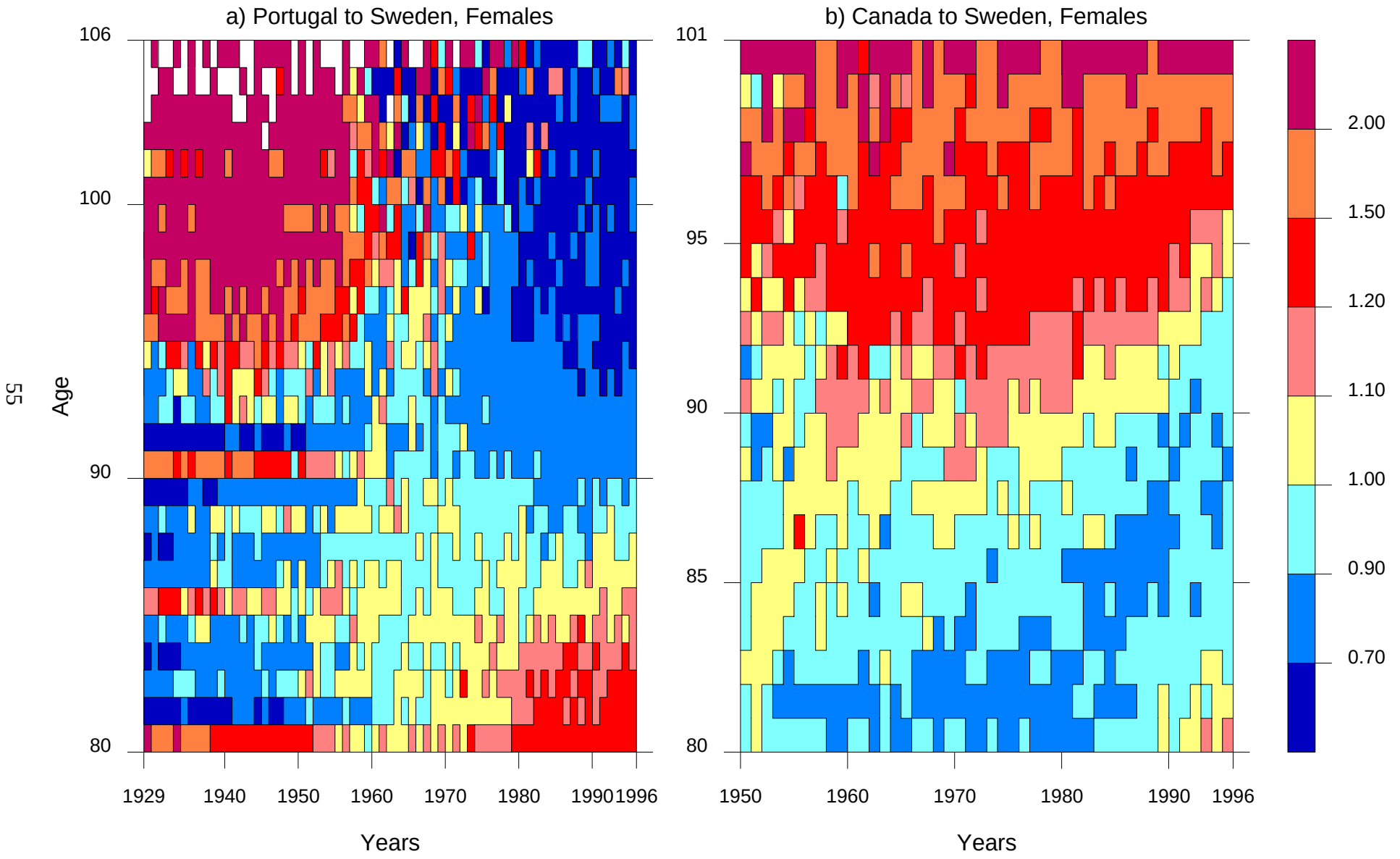
2.3.3 Lexis maps of death distribution ratios

In order to assess the plausibility of the observed death distribution we can compare it with some known accurate distribution of deaths. The main challenge when using this approach is the choice of the benchmark distribution, because the current death distribution depends on the mortality rates operating at this particular moment in time and the past experience of the population summarized in the population structure. If we choose another country for comparison there will always be differences between countries because of different past and present demographic conditions. In this respect, only striking differences should be considered suspect.

Fig. 2.7 shows the ratio of death distributions in the female data in Portugal and Canada to those in Sweden. In this example, the Swedish data are used as a standard because of their widely recognized reliability. Both maps demonstrate how striking the differences could be. The red blemish in the upper left corner of the Fig. 2.7(a) is the area where proportions of deaths in Portugal are more than double those in Sweden. In contrast, the Portuguese proportions become lower than Swedish ones starting with the 1970s, first at ages 89–95 and then at ages 89–105. This complete turn-about can be explained by improvements in Portuguese statistics, and the lower proportions at higher ages can be explained by higher mortality in Portugal compared with Sweden. This analysis suggest a pattern of age exaggeration in the period from 1929 to 1970 in Portugal, a conclusion which we already come to in a different context.

Fig. 2.7(b) shows a different pattern of age exaggeration. The proportions of deaths in Canada are consistently higher than those in Sweden but the difference is not so large as in case of Portugal. Most ratios above age 90 lie in the range of 20% to 60%, and only the proportions of deaths at age 100+ differ by more than double. In contrast to the Portuguese pattern, there are no trends in the contour lines to be observed on this map. It seems that the Canadian data are highly distorted by age misreporting, unless, of course, the mortality at the highest ages is in fact exceptionally low.

Figure 2.7 Ratio of death distributions.



2.3.4 Logistic procedure

The logistic procedure I propose here is an extension of the death distribution ratio method. As noted above, age misreporting commonly produces heavier tails of the death distribution, so if the proportion of deaths observed in one population is more significant than those observed in other populations, this is a symptom of age misreporting problems. My goal will be to classify the populations by proportions of deaths observed at a particular age.

In order to do this I can employ logistic regression (Hosmer and Lemeshow, 1989). The exact procedure I use is a special case of logistic regression, which is the reason why I call it 'logistic procedure' instead of 'logistic regression'. For every year, age and population I will estimate a single number showing how the proportion of deaths observed in a certain population and at a particular age and year deviates from the proportions observed in other populations. The results can be arranged by country and conveniently presented by a Lexis map showing parameter estimates specific to that country.

Suppose that in some year y we observed deaths by single age from age a_1 to age a_2 and a^* is the tested age. The logistic regression is used for analysis of binary data. If we denote the outcome variable by Y, then the logit $g(x)$ of the probability $\pi(x) = \Pr(Y = 1|x)$ is assumed to be a linear function of covariates:

$$g(x) = \ln \frac{\pi(x)}{1-\pi(x)} = \beta_0 + \overline{\beta}\overline{x} \tag{2.5}$$

where β_0 is the general proportion at age a^* and $\overline{x}$ and $\overline{\beta}$ are the vectors of covariates and regression coefficients, respectively. In our case, $Y = 1$ if the death is recorded at age a^* and $Y = 0$ otherwise. In this analysis I will use only dummy covariates to take into account the population included in the regression.

I now turn to the problem of fitting this model. Applying the procedure to each year and age could be computationally quite complex because we need to refit the model many times in order to select the significant variables. Fortunately, in our case the maximum likelihood equations can be solved easily. Let D_i^* be the number of deaths recorded at age a^* and $\overline{D_i}$ be the total number of deaths in the range from a_1 to a_2 and in the i th country. To find the parameter estimates we need to maximize the following loglikelihood function

$$L = \sum_i D_i^*(\beta_0 + \overline{\beta}\overline{x}) - \overline{D_i} \ln(1 + e^{\beta_0 + \overline{\beta}\overline{x}}) \tag{2.6}$$

Firstly, consider the simplest case when no covariates are included at all. The regression equation includes only the general proportion β_0 or grand mean in the usual regression notation. Equation (2.6) reduces to

$$L = \sum_i D_i^* \beta_0 - \overline{D}_i \ln(1 + e^{\beta_0}) \tag{2.7}$$

and the estimate of β_0 is $\hat{\beta}_0 = \ln \frac{\pi_0}{1-\pi_0}$, where $\pi_0 = \frac{\sum_i D_i^*}{\sum_i \overline{D}_i}$. The denominator of the latter equation is the total number of deaths observed in the current year and the numerator is the total number of deaths observed at age a^*; both death counts are pooled over all included populations. To summarize, the estimate of β_0 is simply the logit of proportion of deaths at age a^*.

In the second example I consider the regression equation with one dummy variable. The dummy variable is equal to 1 if the death counts belong to the jth population, otherwise it is 0. To obtain the parameter estimates we need to solve the following system of equations

$$\frac{\partial L}{\partial \beta_0} = \sum_i D_i^* - \overline{D}_i \frac{e^{\beta_0 + \beta_j x_j}}{1 + e^{\beta_0 + \beta_j x_j}} = 0 \tag{2.8}$$

$$\frac{\partial L}{\partial \beta_j} = \sum_i D_i^* x_j - \overline{D}_i \frac{x_j e^{\beta_0 + \beta_j}}{1 + e^{\beta_0 + \beta_j}} = 0 \tag{2.9}$$

where $x_j = 1$ if $j = i$ and 0 otherwise. Taking this into account the equation (2.9) can be simplified to:

$$\frac{\partial L}{\partial \beta_j} = D_j^* - \overline{D}_j \frac{e^{\beta_0 + \beta_j}}{1 + e^{\beta_0 + \beta_j}} = 0 \tag{2.10}$$

and the equation (2.8) can be decomposed into two terms:

$$\frac{\partial L}{\partial \beta_0} = \left[D_j^* - \overline{D}_j \frac{e^{\beta_0 + \beta_j x_j}}{1 + e^{\beta_0 + \beta_j x_j}} \right] + \left[\sum_{i \neq j} D_i^* - \overline{D}_i \frac{e^{\beta_0 + \beta_j x_j}}{1 + e^{\beta_0 + \beta_j x_j}} \right] = 0 \tag{2.11}$$

It follows from (2.10) that the first term is 0 and, finally, the estimate of β_0 is

$$\hat{\beta}_0 = \ln \frac{\pi_0}{1-\pi_0} \tag{2.12}$$

where $\pi_0 = \sum_{i \neq j} \frac{D_i^*}{\overline{D}_i}$ is the proportion of deaths at age a^* pooled over all populations except the jth.

Substituting (2.12) into (2.10) yields the estimate for β_j:

$$\hat{\beta}_j = \ln \frac{\pi_j}{1-\pi_j} - \ln \frac{\pi_0}{1-\pi_0} \tag{2.13}$$

where $\pi_j = \frac{D_j^*}{D_j}$ is the proportion of deaths at age a^* observed in the jth population. Thus β_j is the difference between logits of proportions in the jth population and the proportion observed in other populations.

The same interpretation can be extended to the multivariate case if a few covariates are included in the regression. In this case β_0 denotes the logit of proportion of deaths observed at age a^* in the populations which are *not* included as covariates in equation (2.6). My assumption is that this group of populations provides a reliable estimate of proportion of deaths in a given year and at a given age and can thus serves as a standard for comparison between the populations.

The estimated coefficients β_j are interpreted as the difference between the logits of proportion observed in the standard group and the population in question. Positive values of $\hat{\beta}_j$ tell us that the proportion of deaths observed at age a^* in the jth population is higher than the proportion observed in the main group of populations and negative values tell us that is lower. Our main concern involves positive values, since they can be manifestations of age-misstatement errors.

The logistic procedure discussed here also permits us to select the significant covariates automatically. In order to arrive at the final regression equation without having to refit the regression manually, the following procedure is proposed. I start with the estimation of β_0 only, which is the logit of proportion of deaths at age a^* pooled over all populations. The next step is to include one dummy variable for each population and estimate β_j for each country. The significance of the estimate is tested with the likelihood ratio test at the 1% significance level. Among all significant estimates I select the population whose estimate of β_j has the maximum positive value and include it in the regression. The next significant covariate is selected in a similar fashion except that the sign of the estimate must be the opposite of that included in the previous step. This precaution allows us to avoid the dominance of large countries like the USA, which has very high proportions of deaths at advanced ages. If we do not change the sign of the included variable we might end up with a regression where the β_0 is essentially the proportion observed in the USA while all other coefficients are significant and negative. The reason for this is the large population size of this country.

Table 2.2 shows the fit of the logistic procedure to the death distribution of female populations in the year 1980. The deaths at age 100+ were aggregated into a single age class labeled '100+', and we are comparing the proportion of deaths at age 100+ D^* out of deaths recorded at age 80+ $\overline{D}$. In the USA, for example, the number of deaths $\overline{D}_j$ recorded at age 80+ in the year 1980 is 299,205 and the number of deaths D_j^* at age 100+ is 4,903. The estimates for countries in the range from Denmark to Switzerland (Table 2.2) are not significant, and in this example the countries in this group form the standard population whose proportion at age 100+ is compared with all others. The main difference between the logistic procedure and the method of death distributions is that the standard population is given *a priori* rather than it emerging from the computations itself.

The proportions in the countries at the top and bottom of Table 2.2 deviate significantly from the benchmark proportion and the estimated coefficients show the extent of deviation. The highest positive deviation is observed in the USA, where the proportion of deaths above age 100 is three times higher than the benchmark proportion. The proportions in England & Wales, Australia, France and Spain are also significantly higher than the benchmark proportion but the differences are not so striking as in the USA.

Table 2.2 also includes two additional columns: a) the exponent of the estimated coefficient and b) the ratio of proportion observed in a particular country to the benchmark proportion. It is evident from Table 2, that the values in these two columns are very close to each other even if the relation between them is nonlinear:

$$\frac{\pi_j}{\pi_0} = e^{\hat{\beta}_j} \frac{1+e^{\hat{\beta}_0}}{1+e^{\hat{\beta}_0+\hat{\beta}_j}} \tag{2.14}$$

For the range of values of β_j in Table 2.2, the approximation is rather good, so the exponent of the estimated coefficient β_j can be interpreted as the ratio of proportions of the j th population to the standard proportion.

2.3.5 Application of logistic procedure

I have applied the logistic procedure to the countries listed in Appendix Table 2.1. In some countries the deaths above age 100 are not available so I produced two sets of estimates.

In the first set I checked the proportions by single age in the distribution of deaths at ages above 80. This implies that all deaths above age 80 are available in the country included in the estimation. The deaths above age 100 were aggregated into a single age group 100+. The starting

Table 2.2 Fit of the logistic procedure to the proportion of deaths at age 100+ out of deaths at age 80+. Female populations, year 1980, $a^* = 100+$, $\hat{\beta}_0 = -5.99$ and $\pi_0 = 5.3 \cdot 10^{-3}$. All estimates are significant at the 1% level.

Country	Covariate	$\overline{D}$	D^*	$\hat{\beta}_j$	$e^{\hat{\beta}_j}$	π_j / π_0
USA	FUSACB	299,205	4,903	1.1465	3.15	3.11
England & Wales	FENWAL	125,183	1,012	0.4315	1.54	1.54
Australia	FAUSTL	19,443	150	0.3844	1.47	1.47
France	FFRANC	130,839	890	0.2575	1.29	1.29
Spain	FSPAIN	56,928	380	0.2385	1.27	1.27
Denmark	FDENMA	11,000	70			1.21
Iceland	FICELA	296	3			1.92
Ireland	FIRELA	5,866	31			1.00
The Netherlands	FNLEWA	22,222	130			1.11
New Zealand (non-Maori)	FNZNON	4,544	33			1.38
Norway	FNORJK	8,787	53			1.15
Portugal	FPORTU	16,683	62			0.71
Sweden	FSWWIL	19,769	105			1.01
Switzerland	FSWITZ	12,999	51			0.75
Italy	FITALY	110,887	388	-0.4105	0.66	0.66
Japan	FJAPAN	122,731	397	-0.4894	0.61	0.61
Finland	FFINLA	7,698	24	-0.5263	0.59	0.59
Austria	FAUSTR	20,701	52	-0.7430	0.48	0.48
Belgium	FBELGI	23,376	57	-0.7728	0.46	0.46
Germany, West	FGERMW	149,797	365	-0.7735	0.46	0.46

year is 1861, where we have data for four countries (Denmark, Sweden, the Netherlands and Norway). In the more recent years the estimates based on data from 25 countries.

The second set of estimates is based on the death distribution at ages in the range from 80 to 99. The starting year is 1950 and the data from 34 countries were used in the calculations.

In both cases the logistic procedure was applied by single year and the exponents of the estimated coefficients for each country were plotted as Lexis maps. The estimates were computed separately for males and females and a total of 50 Lexis maps were produced for the first set and 68 maps for the second set of estimates. The Lexis maps can be found on the CD-ROM in the folders '\quality\lp01' and 'quality\lp02' for the first and second set, respectively. Every folder on the CD-

ROM contains a file README.TXT with the names of the countries and the corresponding Lexis maps. The most important results are summarized in Table 2.3.1 and Table 2.3.2.

Exceptionally high proportions of deaths at older ages are found in the USA, Canada, Portugal, Spain, New Zealand (Maori) and Norway. The Maori population of New Zealand is rather small and most of the estimates are not significant. Nevertheless, if one plots the cumulative death distributions versus the distribution observed in the group of reliable countries, it is apparent that the proportions of deaths at older ages in New Zealand (Maori) are considerably higher and I conclude that the data are seriously distorted by age-misstatement errors.

The data for the USA appear to have extremely high proportions at older ages and this evidence supports the widespread belief that severe age-misstatement errors were commonplace in the United States in the period in question. The upward trend in the contour lines observable on the Lexis maps might be an indication of improvements in data quality, but the difference is still very large even at the end of the period we have data for.

On the other hand, the differences in death proportions can reflect the fact that the death rates at older ages are considerably lower in the USA than in other countries (Manton and Vaupel, 1995). Additionally, I note that the differences in death distributions can be attributed to differences in the population structures as well. Moreover, the death distribution of the USA includes deaths from both the white and black population and it can be distorted only by errors in the black population, since these data are considered less reliable than for the white population (Preston et al., 1996).

Canadian data reveal a pattern quite similar to that of the United States except that proportions of deaths at higher ages are lower than in the USA. As in the USA there is perhaps a slight trend in the contour lines. Either there are similarities between the errors in death registration systems of the two countries or there is in fact a North American phenomenon of extremely low oldest-old mortality. To determine which of these two explanations is more likely deeper analysis is required. For now it seems more plausible that the data for both countries suffers from serious age misreporting.

The problems with age misreporting observed in Spain and Portugal occurred in the same period of time when severe age heaping was found in the data. Starting with the 1970s the quality of data improved considerably, approaching that of other countries. It leads us to conclude that before 1970 the data for both countries are of limited use and researchers should be encouraged to utilize the data starting with 1970 only.

According to Kannisto (personal communication), despite improvements in vital statistics in recent years, the Spanish data are still very unreliable and death rates have not reached yet a credible level compared to those of Portugal. After the monarchy was overthrown in 1910 in Portugal, a permanent and very strict population register was established in 1911. At that time, the ages of many middle-aged and old people were not recorded accurately. Consequently, these errors exist in the data in the following decades. When these cohorts die off around 1970–1980, the data become reliable. The case of Portugal is unique and different from Spain. In Spain the improvements in data quality were gradual and less dramatic than in Portugal in the 1970s. The death rate (x1000) for ages 80–99 in the period 1990–1993 was, for example, 137 in the male population of Spain while it was 171 in Portugal and 150 in Norway, Sweden and Denmark combined. In the female populations the numbers are 104, 132 and 104, respectively. Clearly, that mortality rates in Spain are much lower than in Portugal and their levels are comparable with those of Nordic countries. The difference, however, is likely to be due to age-misreporting errors for the Spanish population.

The situation of the Norwegian data is somewhat surprising. Normally, the data from Nordic countries are considered reliable because the vital registration systems were introduced so early. Nevertheless, unexpectedly high proportions of deaths at ages 90+ are consistently observed for the Norwegian population from 1861 and to 1960. It seems that the age-at-death data were grossly exaggerated during this period and one should be cautious about using the data prior 1960. Remembering the drawbacks of Norwegian data detected in the age-heaping analysis, I must state that current estimates of the Norwegian mortality surface (especially for earlier years) are not completely reliable and more work should be done to correct the shortcomings in the data.

Certain irregularities in the age-specific patterns of the death distribution have also been found in the Italian data in the period from 1955 to 1969. The shape of the age distribution is notably different from that observed in the adjacent years. By coincidence, the total number of deaths at ages above 80 also differs from the figures from the WHO mortality database[14] and in exactly the same period of time. It is probable that some operational mistakes occurred during the compilation of the Italian data. This is a problem which has to be investigated more closely in collaboration with the Italian National Institute of Statistics, which provided the data for the K-T database.

[14] http://www.who.int/whosis/mort/

Table 2.3.1 Age exaggeration in mortality databases. Ages 80+.

a) The logistic procedure was applied to the death distribution at age above 80

b) A blank field in the "Age Exaggeration" column means that no data defects were detected

c) The Lexis maps can be found in '\quality\lp01'

Country	Death series	Age exaggeration
Australia	1968–1985	very light, females, years 1980–86
Austria	1947–1995	
Belgium	1974–1995	
Canada	1985–1995	heavy, both sexes, years 1985–95 and ages 95+ (see comments in the text)
Denmark	1861–1995	
England & Wales	1911–1995	
Estonia	1990–1995	
Finland	1878–1995	
France	1950–1995	
Germany, East	1990–1995	
Germany, West	1970–1995	
Iceland	1961–1995	females, years 1980–1995 and ages 90+ (not likely to be due to age misreporting)
Ireland	1950–1992	
Italy	1952–1993	both sexes, a strange pattern is to be observed for years 1955–1969
Japan	1950–1995	
The Netherlands (NSO)	1850–1993	Male proportions are a little high at ages 95+ and in the years 1985–93
New Zealand (Maori)	1950–1995	heavy, both sexes, 1950–80[15]
New Zealand (non-Maori)	1950–1995	light, females, 95+
Norway (NSO)	1861–1993	both sexes, years 1861–1960 and ages 90+
Portugal	1929–1995	heavy, both sexes, years 1929–70[16] and ages 95+
Slovenia	1983–1995	
Spain	1950–1980	heavy, both sexes, 1950–1970 and ages 95+
Sweden (BMD)	1861–1995	
Switzerland	1950–1995	
USA	1962–1990	heavy, both sexes, all years (see comments in the text)

High proportions of deaths are also to be seen in Iceland, especially in the female population for the period 1980–95 and for ages above 90. At first glance, this suggests that there are severe age-misreporting problems at older ages. However, there are some arguments which support the hypothesis that the proportions of deaths can indeed be high due to the evolution of mortality in this country. First of all, the registration system in Iceland is known for its reliability and the quality of

[15] Absolute counts for this population are very small in order to provide a reliable statistical inference.

[16] Male data for Portugal are available from 1940 onwards.

Table 2.3.2 Age exaggeration in mortality databases. Ages 80–99.

a) The logistic procedure was applied to the death distribution at ages 80–99

b) A blank field in the "Age Exaggeration" column means that no data defects were detected

c) The Lexis maps can be found in '\quality\lp02'

Country	Death series	Age exaggeration
Australia	1967–1990	very light, females, years 1980–89 and ages 95+
Austria	1947–1995	
Belgium	1950–1995	
Canada	1950–1995	heavy, both sexes, years 1950–95 and ages 95+ (see comments in the text)
Chile	1983–1987	males, heaping at age 99, years 1983–87
Czech Republic	1950–1995	
Denmark	1950–1995	
England & Wales	1950–1995	
Estonia	1950–1995	
Finland	1950–1995	
France	1950–1995	
Germany, East	1954–1995	
Germany, West	1956–1995	
Hungary	1950–1990	
Iceland	1961–1995	females, years 1980–1995 and ages 90+ (not likely to be due to age misreporting)
Ireland	1950–1992	
Italy	1952–1993	males, a strange pattern is to be observed for years 1955–1969
Japan	1950–1995	
Latvia	1950–1994	females, years 1950–70 and ages 95+
Luxembourg	1956–1995	
The Netherlands	1950–1995	Male proportions are a little high at ages 95+ and in the years 1980–90
New Zealand (Maori)	1950–1995	heavy, both sexes, 1950–80[15]
New Zealand (non-Maori)	1950–1995	both sexes, years 1950–95 and ages 96+
Norway	1950–1995	heavy, both sexes, years 1950–60 and ages 90+
Poland	1972–1995	
Portugal	1950–1995	heavy, both sexes, years 1950–1960 and age 95+
Scotland	1950–1995	
Singapore (Chinese)	1982–1995	
Slovakia	1950–1989	
Slovenia	1983–1995	
Spain	1950–1993	heavy, both sexes, years 1950–70 and ages 95+
Sweden	1950–1995	
Switzerland	1950–1995	
USA	1962–1990	heavy, both sexes, years 1962–90 and ages 90+ (see comments in the text)

vital statistics is comparable with that of other Nordic countries. Secondly, we note that the high proportions of deaths are to be observed only for recent periods, while for earlier periods the proportions are comparable with the commonly observed proportions. This pattern is completely different from that found in Portugal, for example. Thirdly, the oldest-old mortality in Iceland seems to have been the lowest in the world until the 1990s (Kannisto, 1994). This mortality regime can in turn lead to a flattening of the death distribution curve, so the proportions of deaths at high ages in recent years will be appreciably higher than those in other countries. All of these arguments make it seem likely that the high proportion of deaths observed in Iceland are due to a specific mortality regime rather than to age misreporting at older ages.

Less severe problems with age misreporting have also been found in other populations, and the reader can get more information on year-age patterns, time trends and the magnitude of the estimated age exaggeration by referring to the Lexis maps.

2.4 Discussion

Mortality data collected in the K-T database have the advantage that the cohort survival histories for ages above 80 can be reconstructed entirely from the recorded death counts by the extinct cohort method. This eliminates the necessity of using census population data, which are usually of poorer quality than death registration data, to produce mortality estimates. Nevertheless, there still exist some problems with mortality data collected in this way. Because the data collection is carried out over several decades of time and the quality of statistics tends to improve over time, I decided to use Lexis maps for my quality checks. This approach allows me to assess the entire mortality database and see at a glance were the problems occur. To do this I developed the program Lexis to help me with the visualization of large arrays of demographic information and developed new methods of data quality checks which permit me to test each value in the databases against possible errors.

In my investigation I focused on two main problems which occur frequently in oldest-old data. The first problem is age heaping, which is usually defined as the propensity to report age at death rounded off to a year ending in zero or five. I found severe age heaping in Portugal, Spain and Ireland. For Portugal and Spain the data are not affected in a uniform way. The data from the 1970s onwards are of much better quality and the age-heaping problem vanished in recent years. Improvements in death registration statistics are also visible on Lexis maps for Ireland, but they took place somewhat later, in the mid-1980s. Less severe but nonetheless significant age-heaping defects were found in Australia, Chile, England and Wales, Italy, West Germany, France and

Poland. Some patterns like those observed in Norway (NSO) can not be clearly classified as age heaping although the irregularities of the age-specific mortality pattern certainly point to the presence of some distortions in the data. For more detailed information, the interested reader can refer to the Lexis maps included with this chapter.

The second problem I addressed here is age misreporting. The age at death could be deliberately exaggerated or simply misreported. Generally age misreporting is characterized by an age-misreporting matrix (Preston et al., 1997) which is essentially a probability distribution of the genuine age at death reflecting the operational errors, incomplete statistical data and other uncertainties in the data collection. Both directions of misreporting, to lower and to higher ages, are possible but as discussed above, in most cases the distorted death distribution will have heavier tails than the actual distribution of deaths. Following this observation, I developed a procedure for comparing the proportions of deaths reported at a particular age in all countries included in the K-T database. The procedure allows us to assess whether or not the proportion reported in a particular country deviates significantly from the generally observed proportions. The results of this analysis can be reported in an intelligent way by a Lexis map, which brings out immediately the suspicious areas in the database.

Implausibly high proportions of deaths have been found in Canada, the USA, New Zealand (Maori), Portugal, Spain and Norway (NSO). In the last four countries the errors were found only for earlier years while more recent data are consistent with the commonly observed proportions. This pattern can be traced to improvements in the quality of vital statistics which occurred in the early 1970s[17]. On the other hand, the results for the USA and Canada are somewhat puzzling since only slight trends over time in proportions of deaths at advanced ages are to be observed. The two interpretations are possible in this case. The first is that the death registration statistics in both countries is imperfect and the figures produced by the statistical offices give unreliably low mortality estimates at advanced ages. The second possibility is that the data are accurate but this would mean that oldest-old mortality in North America is in reality exceptionally low compared with other developed countries (Manton and Vaupel, 1995). The second interpretation seems less plausible than the first but it is worthy of further investigation.

In addition, less severe age misreporting problems have been found in New Zealand (non-Maori), Ireland, Australia and Latvia. Errors in the last three countries are present mostly in the

[17] The data for New Zealand (Maori) are not of satisfactory quality until the 1980s.

female populations while the male data seem to be of a better quality[18]. Some irregularities are also apparent in Italian data, in both the male and female populations. However, they seem to be produced by errors in operational procedures rather than errors in death registration statistics. The interested reader can find a great deal of material by exploring the Lexis maps.

In principal, the data quality assessment performed here should not be considered conclusive because any unusual demographic conditions may produce country-specific patterns that will differ significantly from the commonly observed values. Nevertheless, in so far as no other evidence is known the results presented here should be interpreted as inaccuracies in the data, which must in its turn be treated accordingly.

The methods presented here can be divided into two groups. The first extends Kannisto's ideas of qualitative analysis by using the more powerful technique of the Lexis contour map to facilitate the analysis of the entire mortality database. The second introduces formal statistical procedures of hypothesis testing (e.g. benchmark mortality procedures, the logistic procedure) and provides quantitative measures of the inaccuracies in the data. Here, too, the Lexis contour maps can be employed to present the results of statistical analysis.

The next step should be to explore opportunities to improve the quality of the data. The correction of the drawbacks I have found in the data will present a significant challenge because the original data used to compute the aggregated numbers published in the official statistics are not available. Further work should focus on communication with the national statistical offices in order to detect misreporting patterns behind the faulty data and to develop methods for correcting the errors. Another approach would be to develop new statistical models which permit us to estimate amount of age heaping or age misreporting along with the correction of erroneous data. Some of the possible avenues to follow in future research are the Monte Carlo simulations of different patterns of age misreporting, smoothing methods for the Lexis diagram, and backward mortality projections.

[18] Another explanation is that the sample size of the male population is too small for reliable statistical inference.

CHAPTER 3

The Danish Mortality Data Base

3.1 Introduction

We have 17th-century parish registrars and their far-sighted collection of statistical data – which has been carried on by government statisticians to the present day – to thank for the fact that we are now in a position to study different aspects of the human life span. One of the major topics in demographic research is the continuing mortality decline in European countries. Although many researchers have been studying the mortality transition of the last two centuries in Europe, "our understanding of historical mortality patterns, and of their causes and implications, is still in its infancy" (Schofield and Reher 1991).

Research in this area is usually hampered by the lack of reliable long-term mortality series. Danish population statistics, which embody an enormous amount of relevant data extending well back into the seventieth century, are an important exception. This makes the compilation all heterogeneous sources of information to construct a consistent database on Danish mortality a worthwhile endeavor.

One of the outstanding features of the Danish statistics is that one can use them to reconstruct the mortality evolution by a single age, year and cohort. Deeper insights into the nature of mortality development can be gained by studying age-specific and cohort-specific mortality trends rather than by simply using the crude mortality indicators. The techniques for studying such data range from relatively simple graphical methods such as Lexis contour maps (Vaupel et al., 1998) to sophisticated statistical models.

The long series of cohort mortality can be highly important in studies on the influence of different genes on the human life span. In these studies the age dynamics of gene proportions is analyzed, which requires that mortality estimates exist for cohorts born a hundred years ago and earlier (Yashin et al., 1998).

The fact that we can compute period and cohort life tables for different years and ages means that we can establish a mortality benchmark for the wide range of epidemiological studies in which the mortality of the Danish population as a whole is compared with the mortality in selected groups of individuals (Christensen, K. et al., 1995). In addition, the Danish population counts estimated by

single year and age are very important since they provide the estimates of exposure time for calculations of the different epidemiological rates.

Furthermore, similar mortality data going back to the mid-nineteen century are available for Sweden, Norway and the Netherlands. Thus, it is possible to make a comparison between countries based on long-term age-specific mortality trends. A first step in this direction is the estimation of mortality ratio surfaces of different countries. It would help to look at mortality differences from another perspective. The estimation of ratio surface of Danish mortality to other countries is another incentive for carrying out this project. The age-specific differences in mortality are usually less well-known and this analysis might reveal some hidden details of such differences.

Besides, we should note that the estimates of Danish mortality and population structure can be quite useful in the area of mortality and population projections.

Finally, the primary goal of this project is to construct a consistent database of Danish mortality. Section 3.2 provides the necessary information about the database structure and its coverage. In section 3.3 we discuss the available raw data used for database compilation and section 3.4 brings together the methods that have been used for achieving the desired level of data completeness and aggregation. Then, I provide a brief historical review of Danish population statistics together with the crude indicators of Danish population changes.

3.2 Database structure

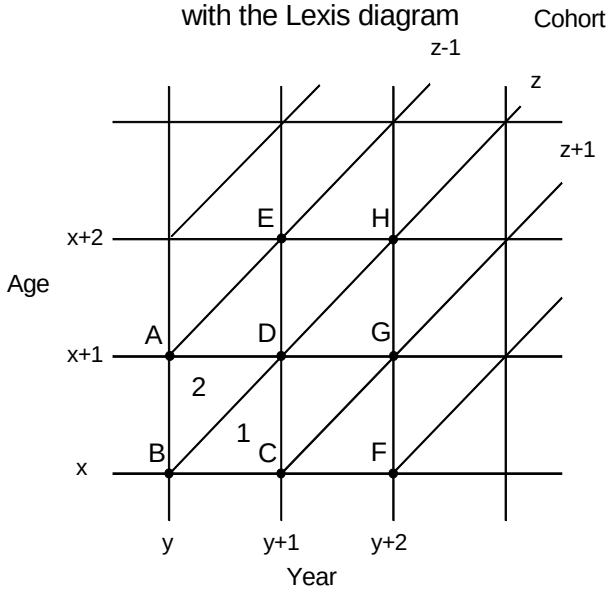
The right choice of database structure can substantially reduce both the cost of data retrieval and the basic computational operations performed on the data. Based on our experience the following database structure is suggested:

COHORT	AGE	POPULATION	DEATHS	TIMING	YEAR
…		Example of records		…	…
1940	30	32,592	13	1	1970
1939	30	31,256	16	2	1970
…	…	…	…	…	…

As the data are collected by years the database is kept sorted by YEAR, AGE and TIMING, and each year includes the same number of ages.

The Lexis diagram shown in Fig. 3.1 illustrates the rationale for the proposed structure.

Figure 3.1 Illustration of the Danish database structure with the Lexis diagram



The database includes six fields and each record is used to store the information about one Lexis triangle (TIMING). Timing “1” corresponds to the triangle BCD and the timing “2” to the triangle ABD. The column DEATHS contains the number of deaths in these Lexis triangles. For example, 13 deaths occurred in the cohort z=1940, year y=1970 and age x=30. The 16 deaths which occurred in the same year and age but in the previous cohort z=1939 belong to timing 2 (triangle ABD). Consequently, the sum of the deaths in timings 1 and 2 is the number of deaths recorded in the given year and age (rectangle ABCD).

The interpretation of numbers stored in the field POPULATION depends on the timing order. If the timing is 1, the population at risk (‘cross-age’ population) is recorded, otherwise it is the population on January 1st (‘cross-year’ population) that is listed. In our example the population numbers are depicted by lines BC and BA on the Lexis diagram and equal to 32,592 and 31,256, respectively.

The database structure presented here is not optimized for the size. The variables YEAR, COHORT, AGE and TIMING are linearly dependent and we can compute, for examples, the YEAR variable as YEAR = COHORT + AGE + TIMING - 1. In other words, one of these four variables can safely be omitted without any loss of information, but given the importance of all fields, they are kept in the database intentionally since this permits a significant reduction in the time of computations.

Data stored in such a format can be used for calculations of virtually any demographic indicators: central death rates, period and cohort life tables, mortality progress indicators and so on. In addition, knowing the exact absolute population and death counts permits us to test statistical

hypotheses and to construct confidence intervals for the computed values. A detailed discussion of the Lexis diagram can be found in Impagliazzo (1984) and Tabeau et al. (1994). In the latter report the different observational planes used by the national statistical offices are discussed as well.

3.3 Original data

3.3.1 Population

Before 1906 Danish population data are scanty, consisting of the censuses which were held every five or ten years. The censuses held in 1801 and 1834 tabulate population by ten-year age groups and those held between 1834 and 1870 by five year groups. As of 1870 the population is tabulated by single-age groups, which makes these censuses entirely suitable for our requirements. The major part of the work, therefore, is concentrated on reconstructing the single-age distribution and providing population estimates for the periods between censuses. All Danish censuses in the nineteenth century were dated February 1st, with the only exception in 1834 (February 18th). The database format requires that the population estimates are for January 1st, so an additional population adjustment has to be made.

Starting in 1906, Danmarks Statistik[19] provides population estimates by single year of age, which can be added directly to the database. For the years 1906–1940 the population at higher ages is given by open age class 85+, leaving the single-age distribution unknown. In this case the population can be estimated indirectly by the extinct cohort method (Vincent, 1951). Some data uncertainties exist also in the period from 1932 to 1940, as the population counts available for these years were rounded off to hundreds. Appendix Table 3.1 summarizes the available raw population statistics.

3.3.2 Deaths

The information about available data on deaths is included in the Appendix Table 3.2. The period from 1943 to the present time does not require any additional manipulations – the death counts can be added directly to the database. For the period from 1921 to 1942, the data are available in the same degree of detail, with the exception that deaths for ages above 100 were aggregated into the single-age group 100+. The deaths in this age group have to be separated by a single year of age. The death counts recorded in this group are very small, and the influence of the separation

procedure on mortality estimates at lower ages is negligible.

The data become less abundant as we move back to earlier years. In the period 1916–1920 the death counts are given by single year and age (see Fig. 3.1, rectangle ABCD). Here the death counts have to be split between cohorts in some reasonable way.

In the period 1835–1915, the death counts are given only in five-year age groups. These data have to be separated by a single year of age and afterwards by cohort to fit the database standard.

3.4 Construction of the database

3.4.1 Deaths

1921–1996

Data on deaths available for these years fit the database structure entirely and were added directly to the database with the exception of the open age class 100+ from 1921–1942. The separation of the 100+ group is discussed below.

1916–1920

Death counts for the years 1916–1920 are given by single year of age. Before adding these data to the database we needed to separate them between the two cohorts that constitute the Lexis triangle. I did this by splitting deaths evenly between cohorts at ages after one. This seems to be a reasonable albeit not a perfect solution at the moment. The separation of cohort deaths at age zero can be achieved more satisfactorily because more detailed statistics are available for this age. The procedure, which I applied to all years where it was possible, is described below.

1911–1915

Deaths for the years 1911–1915 were published by five-year age groups. Along with the death counts given by single year, the aggregated death counts for years 1911–1915 were published by single year of age. The available single-year-of-age death distribution was used to allocate deaths by single age from 1911 to 1915. Subsequently the death counts were split evenly between the cohorts.

[19] National statistical office of Denmark.Danmarks Statistik, Sejrøgade 11, 2100 København Ø. E-mail: dst@dst.dk. Internet: www.dst.dk.

1835–1910

The deaths for this period are aggregated by five-year age groups, which then must be separated by single year of age. As the main intention was to stick to the original data as closely as possible, I selected interpolation as the proper tool for carrying out this task. By using interpolation instead of statistical graduation techniques we can store the death counts in Lexis triangles bound to the five-year totals published in the official statistics. In order to obtain the original aggregated data one can simply sum up the death counts in the Lexis triangles constituting the age group. Naturally, the annual series of the total number of deaths computed from the database will coincide with the total number of deaths found in the official statistics.

Before proceeding to the interpolation, a suitable interpolation method must be selected and its suitability for our problem tested. The performance of the different interpolation methods depends heavily on the interpolated function and our goal is to select a method which can be reliably applied to the cumulative distributions of deaths observed in Denmark in the nineteenth century.

The methods of interpolation and separation employed by actuaries and demographers are discussed in Shryock et al. (1993). They describe the most frequently used methods of oscillatory interpolation, such as Karup-King's Third-Difference Formula, Sprague's Fifth-Difference Formula, and Beers's Six-Term Ordinary Formula, all of which have been used for years to deal with such problems. All these methods are rooted in the polynomial interpolation. They differ only in the number of knots on the interval, boundary constraints and the degree of the interpolating polynomial.

Another appealing method of polynomial interpolation stems from modern developments in numeric analysis which led to the emergence of spline interpolation techniques. Application of spline functions to demographic problems can be found in McNeil et al. (1972). Dierckx (1993) provides systematic introduction to spline theory and discusses the methods of efficient manipulations and numerically stable computations of spline functions. As discussed by Dierckx, any spline can be expressed as a linear combination of b-splines. Therefore the problem of finding an interpolating spline is equivalent to the problem of finding the b-spline coefficients. Once the coefficients have been computed, the interpolated values are easily evaluated by means of the linear combination of b-splines. It is also worth noting that the derivatives and integrals of spline functions can be also calculated in an efficient manner.

In order to test these methods, the death counts with known single-year-of-age distribution of deaths were aggregated into five-year age groups and then interpolated back into the groups by

single year of age. Before carrying out this test the Lexis map of the distribution deviations was computed for the years 1835–1995 to select the period with roughly the same distribution of the grouped death counts as in the years 1835–1915. The visual analysis shows that the deviations lie within 50% prior to 1940 except for the years surrounding the influenza epidemic of 1918. The death distribution in the period starting with the year 1940 is quite different from those observed in the nineteenth century because of rapid mortality changes in the immediately preceding decades. In the end, the years 1916, and 1921–1940 were selected for testing the interpolation methods.

The procedures were applied to the cumulative death distribution starting at age 5 and ending at age 100, with data points available every five years. The high-order derivatives for the spline function at the boundaries were set to zero, thus providing for a natural spline interpolation procedure. Once the interpolated data have been computed, we can assess the deviation of the interpolated distributions from genuine death distributions by one of five widely used methods. The results are shown in Table 3.4 in the Appendix. The bold-faced values in each row of the table show the minimal deviation among all interpolation schemes. It is evident from the table that the cubic spline interpolation is superior to other procedures.

I must note that all methods produced negative values for some ages in the last age group (95–100) because of a rapid function change in this age interval. To circumvent this problem, different boundary conditions were imposed to the spline functions at age 100. This averted negative interpolated values, but the death distribution within this group still exhibited an implausible pattern in comparison with the original distribution.

The next step was to analyze the time trends in this distribution using the linear regression model. It turned out that the trends were not significant at all ages, so the average death distribution shown in Table 3.1 was applied to separate the deaths in this age group.

Table 3.1 The death distribution within the age group 95–99 and in the year 1916 and 1921–1940[20]

Age	95	96	97	98	99
Males	0.401815	0.267665	0.170289	0.104261	0.055970
Females	0.382720	0.259440	0.114684	0.114684	0.071235

Age zero

Death statistics for the first year of life are more detailed than is required by this database.

[20] The years 1917-1919 were excluded because of abnormal mortality conditions.

Starting with the year 1855, for example, the deaths are recorded by the following periods: 0–1 month, 1–2, 2–3, 3–6, 6–9 and 9–12 months. Using these data the deaths in the Lexis triangles can be computed more accurately than for all other ages.

Let x_1 be the upper limit of the age interval and x_2 the lower limit. Assuming that the deaths are distributed evenly in the interval $[x_2, x_1]$, the proportion of deaths occurring in the older and younger cohort will be $\pi_1 = (x_1 + x_2)/2$ and $\pi_2 = 1 - \pi_1$ accordingly. Applying these equations for all age intervals and summing up the deaths, we obtain the death counts by cohort for the first year of life.

In the period from 1835 to 1854 such detailed statistics are not available. In this case the average distribution of deaths observed in the years 1855–1879 was used to split the death counts by cohorts.

The data for the first year of life were aggregated instead of separating the death counts by single age and cohort (as was done for all other ages). Thus some information has been lost, and one should be aware of the fact that the database is not planned for use in studies of infant mortality where the more detail data can be exploited. Still, the mortality at age zero is necessary for the computation of aggregated demographic characteristics summarizing the experience of the whole age range, e.g., life expectancy at birth.

Ages 100+

In the period 1835–1854 deaths at ages above 100 were published by the following age groups: 100–105, 105–110 and 110+. From 1855 to 1942 they are given as a single group 100+. To separate the age group 100+ one needs to make an assumption about mortality at such advanced ages because the direct computation of mortality estimates is not possible – not even using the data from other countries. It is evident from Table 3.2 that the absolute death count numbers are very small and the use of complicated separation procedures would hardly influence the mortality estimates at lower ages.

Bearing that in mind, the deaths were separated with the help of exponential distribution[21], which implies that mortality is constant at ages after 100, with the level of mortality described by a single parameter λ. The parameter λ was estimated by fitting this model to the period life table for 1950–1970. For the male population the estimate of λ was 0.7783 and for females it was 0.6653.

[21] Death counts for the years 1835-1854 were aggregated into the single 100+ age group before the separation.

Table 3.2 The number of deaths above age 100.

	Period						
	1835–39	1840–49	1850–59	1860–69	1870–79	1880–89	1890–99
Males	7	19	13	4	7	3	8
Females	16	26	27	15	25	22	25
	1900–09	1910–19	1920–29	1930–42			
Males	6	9	20	35			
Females	40	32	36	68			

Distribution of deaths by Lexis triangles

For the years prior to 1920 the death counts by single age must be separated between the cohorts contributing the deaths into the two Lexis rectangles. I used the simplest approach here: the deaths were divided evenly between the cohorts. This assumption is not normally justified, especially for older ages[22] where the mortality rates are particularly high. We must discuss it in more detail as it is directly related to the mortality estimates.

It is clear that the proportion of deaths in the triangle BCD to the deaths in the rectangle ABCD (Fig. 3.1) depends on the current population structure and the current mortality rate – and neither is available until the database is completed. The first approach would be to:

a) estimate the current mortality rate and the population structure using the uniform distribution of deaths in the Lexis triangles;
b) develop a statistical model which takes into account the dependence of the distribution on age and year;
c) estimate the model and use the predictions from this model to redistribute the deaths between cohorts;
d) re-estimate the current population and mortality rate using redistributed death counts, and then repeat steps c) and d) until convergence is reached.

It appears that this procedure would be promising, but we did not explore it in more detail because it would be of little practical importance for the construction of the Danish mortality database.

Another approach would be to develop a linear model for the proportion of deaths in one of two Lexis triangles and estimate it using the known data collected on the cohort basis. The

[22] The differences are highest at age zero but in this case more detailed statistics are available for the estimation of separation factors.

predictions resulting from this model will be used further to allocate data by cohort in the data with unknown proportions of deaths. This approach was used, for example, by Condran et al. (1991) and by Wilmoth in his work on Swedish data[23], where he presented seven linear models useful in the analysis of the proportion of deaths in the Lexis triangles. The situation is complicated by the fact that we actually need to build a backward projection, since there are no detailed data available for the nineteenth century. Wilmoth used Swedish deaths for the years 1901–1991 to estimate the model and then he derived the predicted proportions of lower triangles deaths for the years 1751–1900 based on the model predictions in the year 1910, with the corrections for birth counts.

As was stated above, the data on deaths in Denmark are available by cohort from 1921 onwards and by five-year age groups for 1835–1915. The use of more complicated models to separate death counts by cohort improves the overall quality of mortality estimates for the period prior 1921 hardly at all, since for most years the death counts are already interpolated from the original five-year age group data. For this reason I split the death counts evenly between the cohorts and made no attempts to estimate the separation factors.

3.4.2 Population

1976–1996

The population counts for these years have been published as estimates for January 1st and for all ages by single year of age. I have included these population counts in the database without any modifications, with the exception of the cohorts for which I computed the estimates by the extinct cohort method (see below).

Population estimates for this period stem from the Central Personal Register, which was established in 1968. Since that time every resident of Denmark has a CPR number and the information about him is stored in the databases of Danmarks Statistik. Based on this information Danmarks Statistik has been publishing annual estimates of the Danish population since 1976, as the success of CPR had become evident.

1906–1975

The population estimates for these years were obtained by Ulla Larsen directly from Danmarks Statistik. The population is that of January 1st, and it is given by single year of age. The estimates are based on the information available from the census questionnaires along with

[23] See the online documentation at http://demog.berkeley.edu/wilmoth/mortality/

additional non-published data. More specific information about the source of these data and the procedures used to produce the estimates is not available. At advanced ages the population counts were aggregated into a single age group: 85+ for the years 1906–1940, 100+ for the years 1941–1970, and 99+ for 1971–1975. These age groups do not pose any problems, since the population at such ages can be computed by extinct cohort method.

1870–1901

Population by single age is available for this period from censuses held in 1870, 1880, 1890 and 1901 (see also Appendix Table 3.1).

The first problem is that the censuses were carried out on February 1st rather than on January 1st as required by the database. Therefore, one needs to correct the census population by taking into account population trends over time. To make the adjustment a simple regression model $\ln \tilde{N}(y) = \beta_0 + \beta_1 y$ was fitted separately to each age x, with $\tilde{N}(y)$ being the 'cross-year' population in the year y (1870.085[24], 1890.085, ... 1906, 1907, ... 1917). All time series stop just before the influenza epidemic and include only cohorts with a loglinear increase in birth counts[25]. This restriction seems to be reasonable because the series of $\tilde{N}(y)$ are highly correlated with the birth count of the corresponding cohort and because the number of births fell markedly in 1910 both for males and females. This drop in the number of births can be clearly traced in the population structure and can worsen both the fit of regression and the adjustment we must now make.

Finally, the population estimates for January 1st were calculated by linear interpolation of the census population using the age-specific derivatives predicted by the regression for January 1st.

Another problem that needed to be addressed here was the estimation of population between censuses. I did this in a standard way by using the natural balance equation. The population $\tilde{N}_{x,y}$ aged x at the time of first census y will be aged $x+\Delta$ at the time of second census $y+\Delta$, where Δ is the time between censuses. We know the values $\tilde{N}_{x,y}$, $\tilde{N}_{x+\Delta,y+\Delta}$ from the censuses and we know the number of deaths $D_{z,\Delta}$ in the cohort $z = y - x - 1$ crossing the year y at age x during period Δ from death statistics. Given these numbers we are able to compute the inconsistency error between them:

---

24 The fractional part of these numbers reflects the fact that the censuses were taken on February 1st

25 The cohorts 1835-1909 (R^2=0.981) for males and 1835-1908 (R^2=0.980) for females.

$$\delta_{x,y} = \tilde{N}_{x,y} - D_{z,\Delta} - \tilde{N}_{x+\Delta,y+\Delta} \qquad (3.1)$$

In the ideal case, i.e. if the population is closed for migration, the error δ is zero. In real populations it can deviate significantly from zero because of migration or because of inaccuracies in the census population which can be produced, for example, by different coverage in two censuses; or by errors in the recorded age at death in the period between censuses.

In my procedure the population between censuses was estimated with the help of the natural balance equation, and the error δ was distributed evenly among the Lexis triangles of the cohort z in the period from y to $y+\Delta$. An alternative name for this procedure is 'intercensal cohort survival method' (cf. e.g. Wilmoth, BMD documentation[23]).

Sometimes this method produces negative population numbers in the period between censuses. Such unacceptable results are related mainly to the following three sources of errors:

a) errors in the census population counts and death registration;
b) errors introduced by the interpolation procedure;
c) invalid assumption of even distribution of the error among Lexis triangles (this is closely related to the age-specific patterns of migration).

The problem was not explored more deeply in our case as the negative numbers occurred at ages where the extinct cohort population can be computed.

We should also note that the error δ is particularly high in this period because of high emigration, mostly to America (Hvidt, 1971). The database permits calculations of the total migration numbers, which are consistent with those given in Matthiessen (1970).

1834–1869

Population statistics for this period are also available only from censuses – and the counts are given in five-year age groups (Appendix Table 3.1). Before applying the natural balance equation method we need to estimate the single-age population structure using available population and death count data. I rigorously tested two methods before applying the superior one to the real data.

The first method is the combination of the natural balance equation and the extinct cohort method. In this method some known single-age population is projected back to the time of the previous census using the death counts available for this period. Any migration that may have occurred in this period is not taken into account. The resulting population distribution at the time of the previous census is used then to prorate the official aggregated population.

Let $\tilde{N}_{x,y}$ be the population aged x at the beginning of year y and in the cohort $z = y - x - 1$. Then the population at the time of previous census is

$$\tilde{N}_{x-\Delta,y-\Delta} = \tilde{N}_{x,y} + D_{z,\Delta} \tag{3.2}$$

where Δ is the time between censuses and $D_{z,\Delta}$ is the number of deaths in the cohort z during period Δ. Using estimated single-age proportions $\pi_{x,y-\Delta} = \frac{\tilde{N}_{x,y-\Delta}}{\sum_x \tilde{N}_{x,y-\Delta}}$ in the year $y - \Delta$ it is easy to separate the available census data by single year of age.

The second method I tested is the interpolation of the cumulative population distribution by the natural cubic spline. This involves the computation of b-spline coefficients and the evaluation of interpolating spline by every year of age. If the software is available, this can be performed effortlessly.

In order to test the methods on the available data, I aggregated the single-age population for the period from 1925 to 1974 into the age groups of the 1834, 1840 and 1860 censuses and then reconstructed it again by single year of age. I applied the method of the natural balance equation with step Δ equal to ten years. That is, the population in the year 1925 was reconstructed using the population of 1935 as a pivot. Subsequently I computed the deviation $\delta = \sum_x \frac{(\tilde{N}_x - \hat{\tilde{N}}_x)^2}{N_x}$ between the original and the reconstructed populations and plotted it in Fig. 3.2.

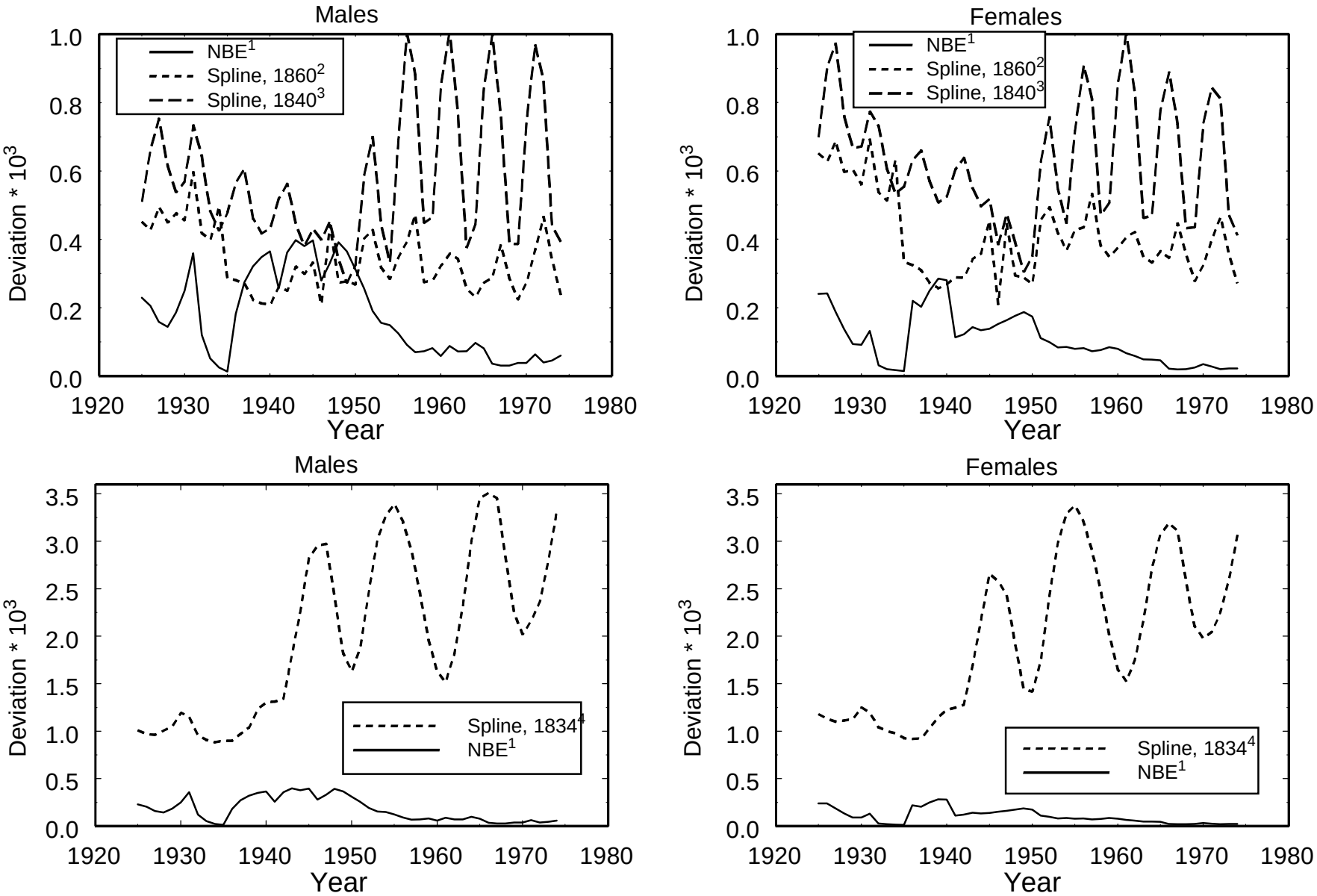
It is apparent from Fig. 3.2 that the method of the natural balance equation with the small number of exceptions reproduces the genuine population more accurately than the spline interpolation procedures, especially if the population is given in the broader age groups, as in the 1834 census.

I therefore estimated the single-age distribution of population for this period by means of the natural balance equation method. The gaps between censuses were filled in using the same procedure as in the period from 1870 to 1901.

Extinct cohort population

The extinct cohort method (Vincent, 1951) is widely recognized as producing reliable population estimates at older ages, where migration can be safely ignored. In this database the

Figure 3.2 Deviation between the orginal and the reconstructed populations.



[1]The method of the natural balance equation.

[2]The spline interpolation of the deaths distribution of 1860 census.

[3]The spline interpolation of the deaths distribution of 1840 census.

[4]The spline interpolation of the deaths distribution of 1834 census.

extinct cohort population estimates were computed for all ages above 80. The last cohort with extinct population was 1887 for males and 1883 for females.

The procedure for Denmark is more complicated than the standard method because the coverage of Danish statistics was changed in 1920. In this year South Julland (Sønderjulland) became a part of Denmark, increasing the population of the country by about 163,000. The population of South Julland was enumerated as a standalone geographical area in the 1921 census and death counts were included in the official statistics starting with the year 1921. For this reason, population estimates in the cohorts crossing year 1920 have to exclude the deaths that occurred in this part of country in the years prior to 1921. This fraction of deaths was taken to be equal to the population of South Julland on January 1st 1921, which was published by single age in the 1921 census.

'Cross-Age' Population

The calculation of the 'cross-age' population or population at risk N (Fig. 3.1, line BC) is based on the assumption of even migration distribution:

$$N_{x,y} = \frac{1}{2}\left((\tilde{N}_{x-1,y} - {}^{2}D_{x-1,y}) + (\tilde{N}_{x,y+1} + {}^{1}D_{x,y})\right) \tag{3.3}$$

where $\tilde{N}_{x,y}$, ${}^{1}D_{x,y}$, ${}^{2}D_{x,y}$ are the population estimates at January 1st, death counts in timing one and two, and in the year y and at age x, respectively.

These population estimates are particularly useful for computing both period and cohort life tables constructed by the cohort method or in mortality models where the error has binomial distribution.

3.5 Danish demographic statistics

The Danish population and death statistics are rooted in the seventeenth century, when parish registers became compulsory. In this section I list the demographic events relevant to the present work in chronological order. Information about early Danish parish registers can be found in Johansen (1998). The information presented here is based mostly on Matthiessen (1970) and Impagliazzo (1984).

- **1645–1646** - parish registers of births, deaths and marriages maintained by the clergy became compulsory by rescript. The territory of Denmark was covered only partially in the following few decades.

- **1735** - summary statistics of parish registers became available annually in the form of a statistical publication called “General Extract”.
- **1769, August 15th. First census.** Census information was presented in summary tables. The population was divided by sex, and age was reported by six groups for the ages under 48 and by an open age class 48+. Marital status was recorded as married and non-married. Occupational status was divided into nine groups. Although the enumerated population was "de jure population", some temporarily absent persons, e.g. sailors, may have been omitted. Some military personnel was also excluded from the enumeration for security reasons.
- **1775** - A prescribed schedule of vital statistics was introduced. Clergy used this schedule to fill in deaths by sex and 10-year age groups, and births by sex and legitimacy. Starting in 1783 the number of marriages was also included.
- **1787, July 1st. Second census.** This census was similar to that of 1769, with the exception that the names of individuals were recorded as well.
- **1796** - The first statistical office (Tabelkontoret) was founded. This office conducted the 1801 census. The office was abolished in 1819 in favor of the statistical commission (Tabelkommisionen).
- **1800** - Births reported by clergy were divided into the categories live-births and stillborns.
- **1801, February 1st. Third census.** The population was enumerated by 10-year age groups. Statistical reports of this census were published together with reports of the 1834 census.
- **1829** - Introduction of the death certificate.
- **1834, February 18th. Fourth census.** This is the first census conducted by the Tabelkommisionen. The population was enumerated by 10-year age groups. The results of this census were published in the first statistical publication (Tabelværket, 1st series, 1st volume).
- **1835** - The distribution of marriages by broad age groups was introduced. Deaths became recorded by the following age groups: below 1 year, 1–2 years, 3–4 years, 5–9 years, etc. Such detailed death statistics made possible the calculation of reliable mortality estimates.
- **1840, February 1st. Fifth census.** The population was recorded by five-year age groups and by single age for ages under five. This is the first census in which the population was tabulated by five-year age groups.
- **1845, February 1st. Sixth census.**
- **1850** - The national statistical office was founded (Statens Statistiske Bureau, later Det Statistiske Department, and presently Danmarks Statistik).

- **1850, February 1st. Seventh census.**
- **1855, February 1st. Eighth census.**
- **1860, February 1st. Ninth census.** The birth distribution by age of mother was introduced.
- **1864, Autumn.** Sønderjylland (hertugdømmet Slesvig) became part of Germany. About 55,000 people emigrated from this region in 1867–1900, the major part to America and a smaller part to Denmark.
- **1870, February 1st. Tenth census.** For the first time the population was reported by single age. The island of Ærø became part of Denmark with the peace treaty of 30 October 1864 and was included in the census statistics.
- **1877** - Birth certificates were required everywhere in Denmark.
- **1880, February 1st. Eleventh census.**
- **1890, February 1st. Twelfth census.**
- **1901, February 1st. Thirteenth census.**
- **1906, February 1st. Fourteenth census.**
- **1911, February 1st. Fifteenth census.**
- **1911** - Individual data on birth, marriage and death were sent from clergy to the national statistical office, thereby abolishing the former schedule of vital statistics.
- **1916, February 1st. Sixteenth census.**
- **1920, June 15th** - Sønderjylland (hertugdømmet Slesvig) became part of Denmark, thereby increasing the total population by about 163,000 people.
- **1921, February 1st. Seventeenth census.**
- **1925, November 5th. Eighteenth census.**
- **1930, November 5th. Nineteenth census.**
- **1935, November 5th. Twentieth census.** In this census questionnaires were distributed to all individuals.
- **1940, November 5th. Twenty-first census.**
- **1945, June 15th. Twenty-second census.**
- **1950, November 7th. Twenty-third census.**
- **1955, October 1st. Twenty-fourth census.**
- **1960, September 26th. Twenty-fifth census.**
- **1965, September 27th. Twenty-sixth census.**
- **1968** The Central Population Register (CPR) was established. The process of registering

statistical information became continuous. The establishment of the CPR led to the abolishment of the questionnaire-based census.

- **1970, November 9th. Twenty-seventh census.** This is the last census which used questionnaires.
- **1976, January 1st. First CPR based census.**
- **1981, January 1st. Second CPR based census.**

...

Information on vital statistics has been published in the Table Works (Statistisk Tabelværker) since the year 1801. The first publication covers the period from 1801 to 1833 and was published together with the 1801 and 1834 censuses. All other publications cover five-year periods. The population-statistical report (Befolkningens Bevægelser) has been published since 1931 on an annual basis. To complete the picture of Danish statistics, I have reproduced the table of former publications of Danish statistics from Befolkningens Bevægelser 1995 in the Appendix Table 3.3. Two other important sources of population statistics should be mentioned:

1) "Causes of Death in Denmark" (Dødsårsagerne i Danmark), which is published by Danish Ministry of Health (Sundhedsstyrelsen);
2) "Population by provinces" (Befolkningen i kommunerne pr. 1. Januar), issued by Danmarks Statistik.

3.6 Major indicators of Danish population changes

Data on the total Danish population by sex and year are shown in Fig. 3.3. In the period from 1835 to 1996, the male population increased from 605,300 to 2,592,200 and the female population from 619,000 to 2,658,800, which corresponds to an annual rate of increase of 0.9%, or approximately 25,000 people, per year. Females outnumbered males for the whole period of observation and especially in the last two decades. This can be explained by the highest gap ever between male and female mortality observed in this last period. Persistent growth of Danish population continued until 1981, when the total population started to decline. This decline lasted until 1986, when the population began to grow again. The population leap in 1920 resulting from the reunification of Denmark and South Julland is also clearly visible on the graph.

Declining mortality and fertility, two attributes of the demographic transition, had a profound influence on the age structure of the Danish population. Fig. 3.4 shows the striking

Figure 3.3 Changes in the Danish population from 1835 till 1996.

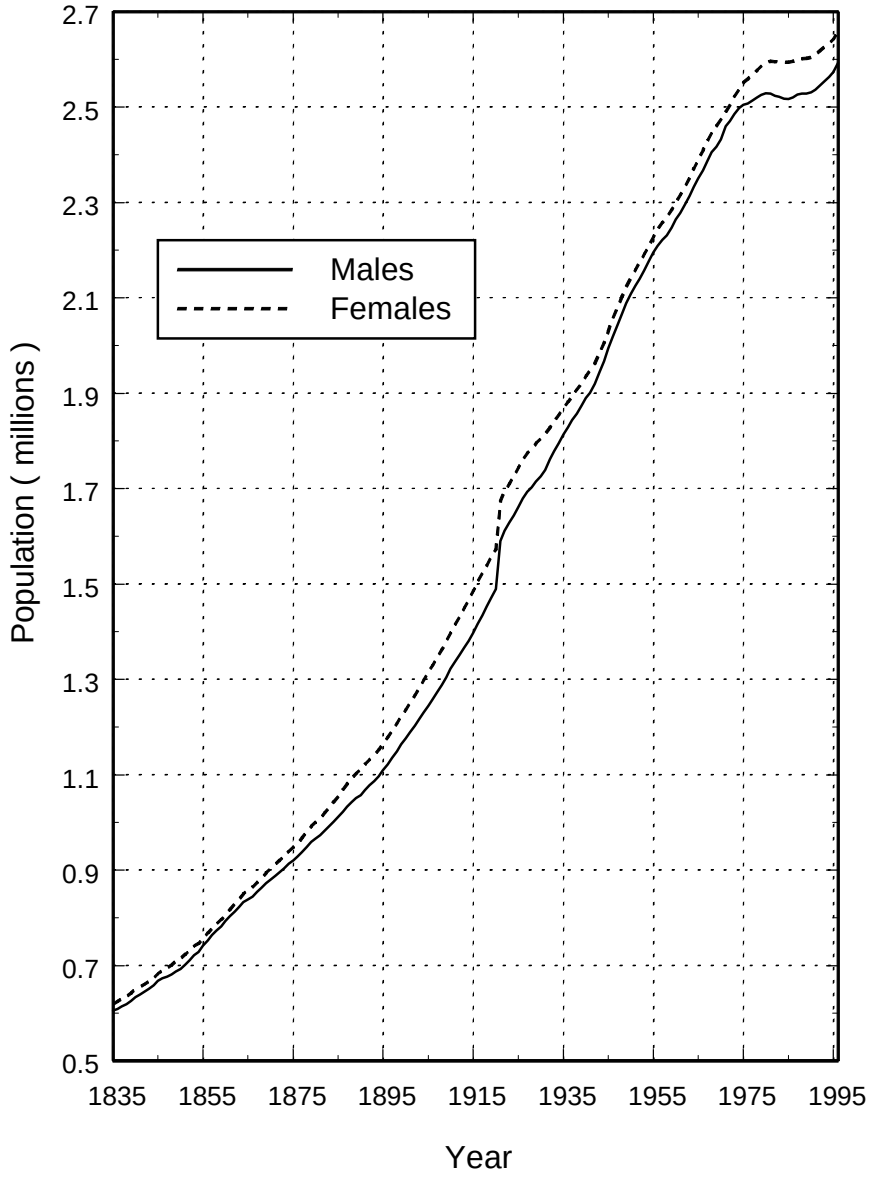
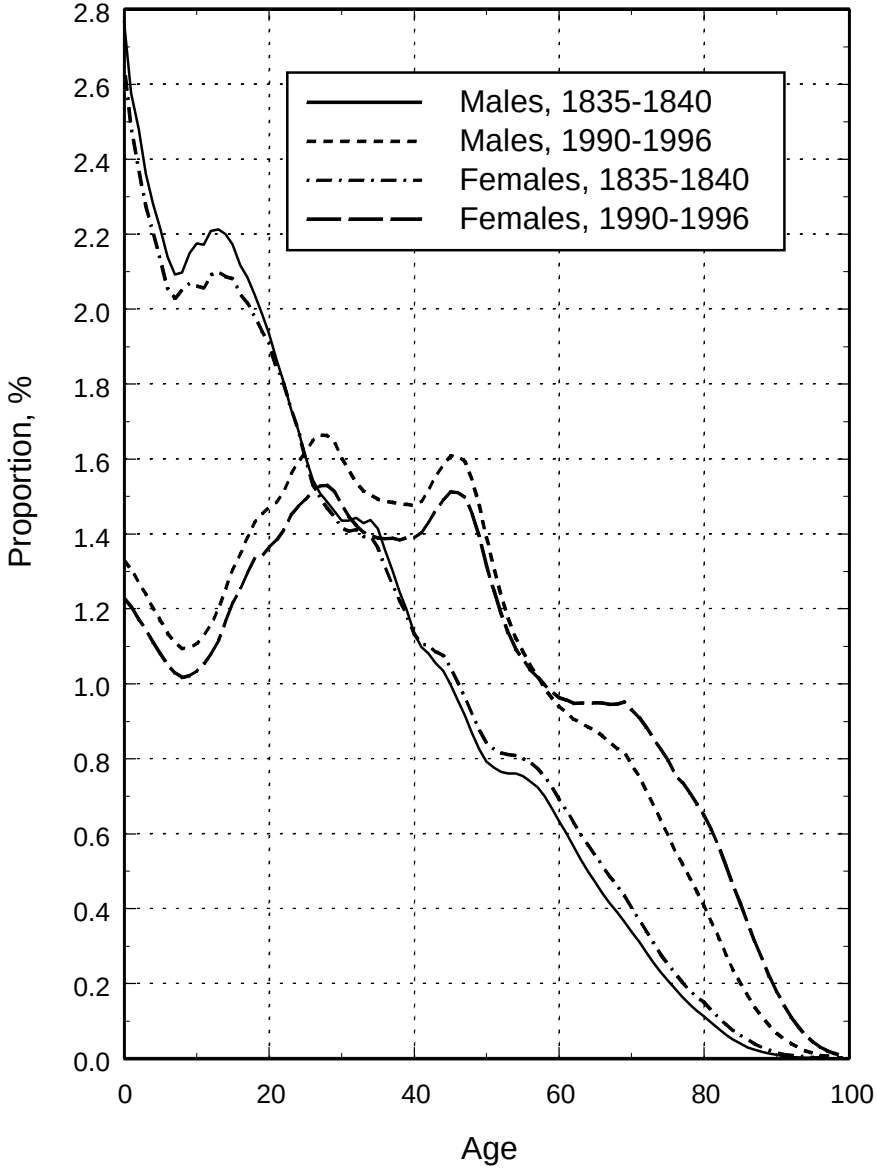


Figure 3.4 Changes in the age structure of the Danish population.



differences between 1835–1840 and 1990–1996 age structures. The first is characterized by a high proportion of children and young people while the proportion of oldest old (80+) is negligible. In contrast, the contemporary age structure of the population exhibits substantially reduced proportions of young people and dramatically increased proportions of the oldest-old. In the male population the proportion of children aged 0–10 dropped 50 percent while the proportion of males aged 80 quadrupled and that of 90-year-olds rose by a factor of seven. The changes in the female age structure are even more impressive, the proportion of 90-year-olds, for example, is 11 times higher than in 1835–1840.

Life expectancy conventionally summarizes the changes in the mortality regime as an overall measure of mortality. To follow the changes in life expectancy I computed the single-year period life tables and plotted the life expectancy at birth in Fig. 3.5.

As indicated by this figure, Danish life expectancy has undergone remarkable changes since the middle of the nineteenth century. In the year 1835 males lived an average of 40 years and females 42 years. By 1994 these figures had risen to 73 and 78 years of age, respectively.

Until the 1870s the rate of increase remained relatively moderate. Fitting the linear regression model

$$e_0(y) = \beta_0 + \beta_1 y \tag{3.4}$$

to the trends in the life expectancy gives 0.041 for males and 0.026 for females (Table 3.3). Curves are jagged due to the frequent epidemics of infectious diseases which plagued the country at the time. Consequently, the standard error of estimates is high and the estimates are clearly not significant.

Persistent increases of life expectancy started in the 1870s. Until the year 1950 this increase was fairly strong, with an annual rate of increase of 0.315 for males and 0.314 for females. Starting in the 1950s mortality improvement decelerated, especially for males. The annual rates of increase in life expectancy for the different periods are shown in Table 3.1, which includes the estimates of $\hat{\beta}_1$ of the model (3.4).

Figure 3.5 Danish life expectancy.

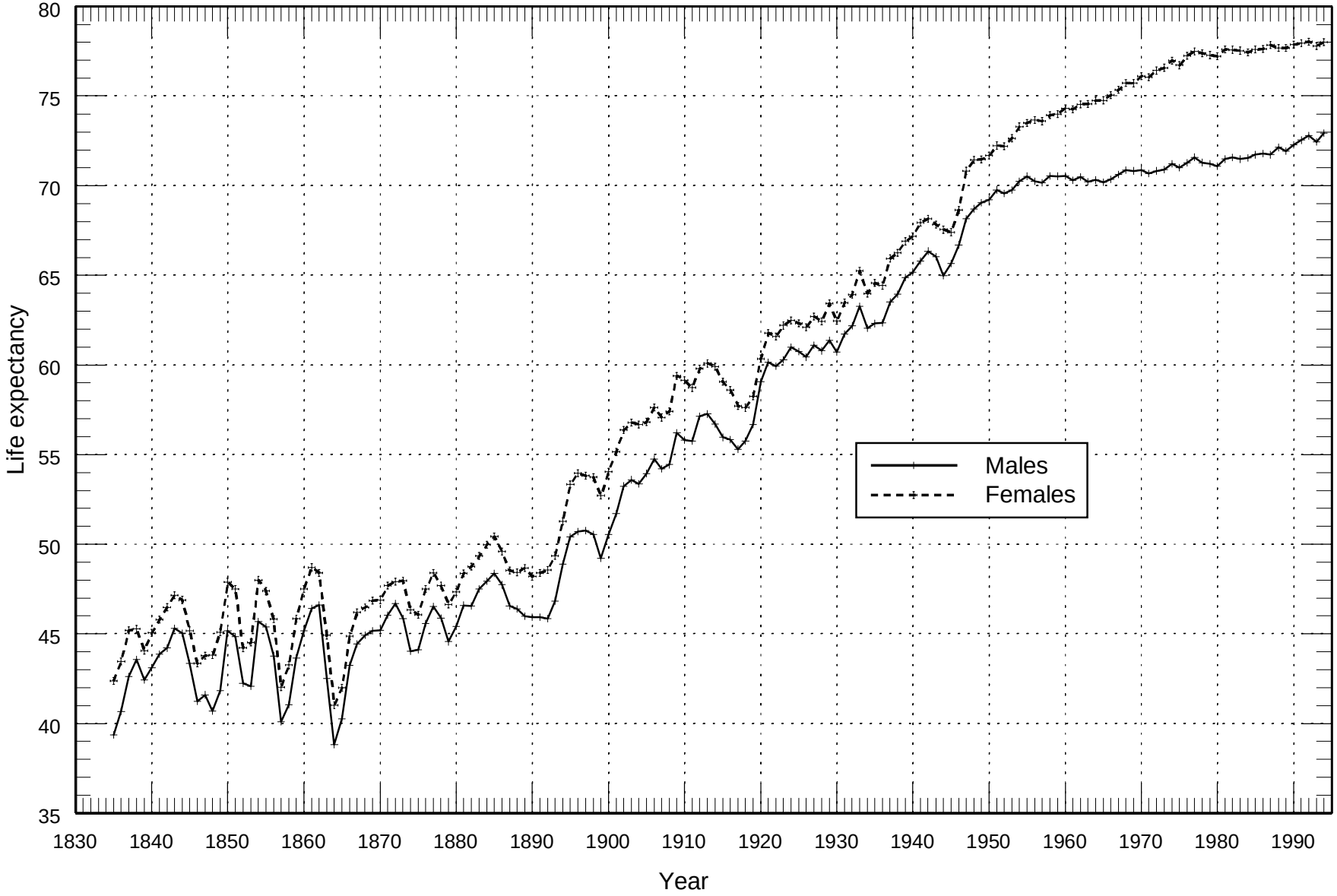


Table 3.3 Annual rates of increase in Danish life expectancy in the selected periods.

Period	Males	Females
1835–1869	0.041 (0.034)	0.026 (0.033)
1870–1949	0.315 (0.007)	0.314 (0.006)
1950–1979	0.055 (0.005)	0.189 (0.005)
1980–1994	0.113 (0.010)	0.044 (0.007)

As indicated by this table, the difference in the male rate of increase before and after the year 1950 is about 6-fold and the difference in the female rate of increase before and after the year 1980 is about 4-fold.

The difference between male and female life expectancy can also be clearly followed in Fig. 3.5. For the whole period of observation, female life expectancy was always higher than male life expectancy. In the period from 1835–1950 there are no systematic changes: the male–female difference is irregular, hovering between one and four years.

Starting with the 1950s a large gap between male and female life expectancy began to manifest itself. This gap reached its peak in the year 1980 and then started to decline. For Denmark this development is attributed to the stagnation in male life expectancy while female life expectancy continued to increase. In the 1980s the pattern was reversed: male life expectancy began a steady increase while female life expectancy stagnated at a constant level. This means that the gap between male and female mortality has begun to decrease in recent years, contrary to the tendency observed 10 years ago.

3.7 Conclusion

The Danish database described here shows the potentialities for research on cohort-, age- and time-specific mortality changes. The database includes observations on two demographic characteristics: population and death counts, which were collected for the period from 1835 to 1996 and from age zero to the highest age attained. The database is organized by single year of time, age and cohort, which permits us to focus on the subtle age-specific and time-specific details, thereby refining the results of demographic analysis. The methods utilized for the construction of this database can be employed for the construction of similar databases for other countries with rich demographic statistics, such as Sweden, Finland, Norway and the Netherlands, as well. This will make it possible to perform comparisons between countries with the emphasis on age- and time-specific mortality

differences, rather than having to use crude mortality indicators.

The organization of the database permits the calculation both of period and cohort life tables for any period covered by the database. Additionally, I performed a consistency check between the official Danish life tables and those computed from the database, and the two series of estimates are in good agreement with each other. The official life tables for the period from 1931 to 1995 were kindly given to me by Michael Væth and the life tables for the period from 1991 to 1995 were taken from publications of Danmarks Statistik.

My analysis of Danish population changes demonstrated that the total population has increased 5-fold since 1835, reaching a figure of over five million in the mid-1990s. The changes in age structure of the population are even more remarkable. The most striking observations are the increase in the population of elderly and the transition from a young society to an aging society.

The data on males and females are kept separately in the database, which makes it possible to study sex differences in survival with a focus on age-specific mortality differences. The age-specific analysis would shed more light on the phenomena discussed in the previous section: the gap between male and female mortality and the gradual narrowing of the gap between male and female life expectancy in the most recent years.

Life expectancy gains during the last few decades have been very moderate in Denmark compared to other developed countries (Middellevetidsudvalg ,1993). The more detailed data compiled in this database permit us to explore the less well-known age- and time-specific mortality differentials in Denmark and other countries. In conclusion, the data collected here would be extremely useful in an analysis of the excess of Danish mortality observed in recent decades.

CHAPTER 4

A Descriptive Analysis of Danish Population

4.1 Introduction

Danish demographic statistics permit a reliable estimation of the Danish population surface for the period from 1835 to 1996 and for all ages (Chapter 3). In this section I will give an overview of the evolution of the Danish population with the main focus on age-specific changes in the force of mortality. In addition, I will compare the Danish mortality development with that of Sweden, the Netherlands and Japan, and I will discuss the cause-specific mortality differences between these countries. Due to the large amount of data analyzed, most results will be presented in the form of Lexis maps following the approach of Caselli et al. (1985), Caselli et al. (1987), Vaupel et al. (1998).

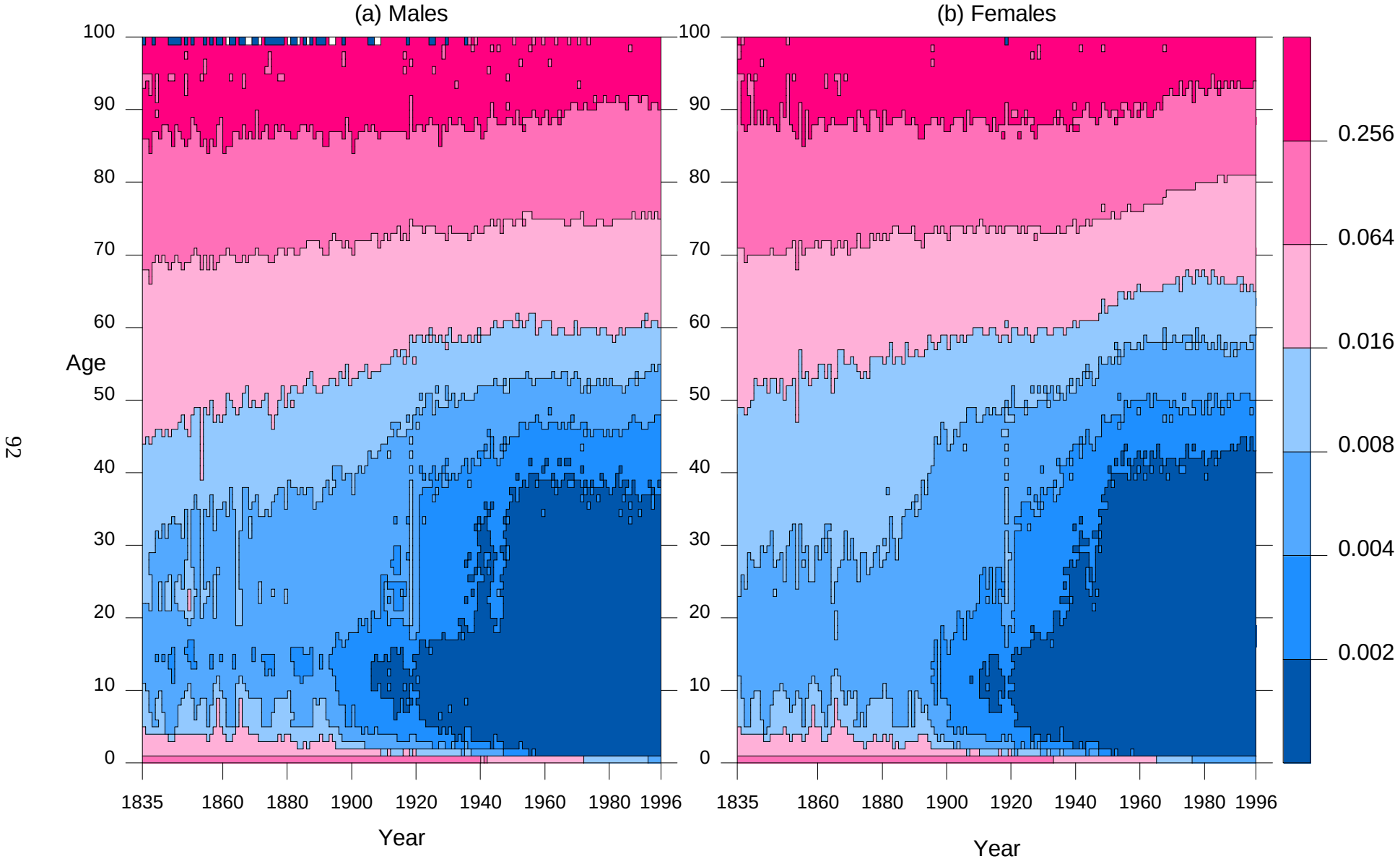
4.2 A descriptive analysis of the Danish Population

4.2.1 Mortality

In order to grasp the evolution of the entire mortality surface of Denmark, the central death rates have been computed by single year and age, and plotted in Fig. 4.1. The death rate is assumed to be constant over a Lexis rectangle, and it is painted in a single color on the Lexis maps. In other words, the Lexis map presented here is plotted without any use of interpolation techniques, so it reflects the original data we have.

The scale shown on the right divides the whole surface into seven areas. Each area on the map corresponds to a certain range of mortality levels. The colors of the scale have been selected according to recommendations by Cleveland (1994). The color encoding used for producing this map has two functions. First, by changing the shade of a fixed hue we can perceive an ordering of a quantitative variable, i.e., mortality rates. Second, by using two different hues (magenta and cyan) we can achieve another perceptual goal: the boundaries between map regions can be clearly perceived. In contrast to Cleveland, I decided to portray low mortality values with cyan (a cold color) and high mortality values with magenta (a warm color). This should help to improve our perception of ordering in mortality rates since it corresponds to the color encoding used in

Figure 4.1 Danish Mortality Rates.



geographical maps, where ocean depth is portrayed in tones of blue. From now on we will refer to a map region by the hue used to fill it in and by the scale level which separates this region from the adjacent region with the lower values. For example, in Fig. 4.1 all death rates higher than 0.256 are colored in magenta. We will refer to this color as magenta (0.256). For the map region including the lowest values we will refer by the scale level as (<0.002). On this map the area with the lowest death rates is colored in dark cyan and we will refer to it as dark cyan (<0.002). As one can see in Fig. 4.1, such low levels of mortality did not start to emerge until the beginning of the 20th century in the age group 10–15.

The trends in the contour lines allow us to follow the evolution of mortality over time. The contour line itself shows the location of a particular level of mortality over age and time. For example, on the female mortality map the contour line corresponding to the mortality level of 0.008 starts at age 23 in 1835 and ends at age 59 in 1995. This means that the risk of death for a 23-year-old female in the year 1835 was equal to the risk of death of a 59-year-old female in 1995. This shows an impressive age-specific shift in the mortality level.

Because human mortality increases uniformly starting at approximately 30 years of age, the shift of contour lines into the higher ages would portray the progress in mortality. The slope of the contour lines reflects the rate of this progress. Childhood mortality reductions can be seen in the shrinkage of the area of very high mortality in the first years of life. Especially striking is the reduction of infant mortality ${}_1q_0$, which fell from 148 per 1,000 in the 1855–1865 to 8 in 1985–1995 for males and from 124 to 6 for females.

We observed that mortality decreased generally at almost all ages but that this progress was not uniform over age or over time. This permits us to identify the timing of mortality changes and their age-specific features. In addition, we note that the overall progress in mortality did not follow the same pattern in the male and female populations, especially in the second half of this century. The onset of the rapid mortality decline occurs in the late 1890s. This is the time during which the areas of the low mortality began to form. These areas are portrayed in blue and dark blue in Fig. 4.1 and correspond to mortality levels below 0.4% and 0.2% respectively. The mortality reductions prior to this period are somewhat less regular except for ages around fifty. The latter generalization should viewed with caution, however, as the overall quality of mortality estimates in the 19th century is less reliable than in the 20th century. Mortality progress at older ages (70+) was very slow and no appreciable gains in mortality reductions can be observed until the 1950s.

Female mortality at young and middle ages (0–60) fell noticeably from the 1890s to the 1950s, while male mortality gains were more moderate. Starting with the 1950s male mortality

stagnated. This is evident in Fig. 4.1, where the contour lines (e.g. 0.008) rose until 1950 but then remained at a constant level or even declined. In the 1990s there have been some positive changes in the mortality dynamics as it is indicated by the upward bend in the contour lines. For example, the death rate at age 65 declined from a level of 54 per 1000 in 1835 to 25 in 1945 and then increased to 30 in 1970; in the period 1990–1995 it was again at a level of 25 per 1000. The stagnation in mortality can be seen on the female map as well, but it occurred later in time. This difference in mortality dynamics between the sexes had a major impact at the emergence of the gap between male and female mortality - a matter which we will discuss later on. It is important to note that infant mortality has continued to decline over time and has now reached the lowest level in the history of the Danish population.

It is somewhat surprising that starting in the 1950s, when reductions in middle-age mortality were low, considerable progress occurred at older ages (70+). Again, male mortality reductions lagged well behind those of females but the progress in both populations is apparent on the Lexis maps. The gains in the older age groups were not as exceptional as the mortality reductions in childhood and middle ages. Nevertheless, this observation is rather important because it shows that the elimination of premature deaths at young and middle ages is not the only factor contributing to an increase in the human life span.

Cohorts that reached 90 in the 1970s were born in the 1880s - a time when childhood and infant mortality fell dramatically. It might be the case that progress at advanced ages can be attributed to improved health conditions in childhood. Additional research is required in order to test this hypothesis.

Another important issue in relation to reductions in oldest-old mortality is the fact that the quality of statistics improves over time. It has been shown by Preston *et al.* (1997) that age misreporting (not necessarily age exaggeration) at advanced ages results in lower mortality estimates computed from erroneous data. This means that improvements in oldest-old mortality can be masked by age misreporting which might be present in the data for earlier periods.

Period effects are also clearly manifested in Fig. 4.1. They can be traced in the long vertical lines of exceptional mortality. The high ridges of mortality in 1853 and 1918, for example, reflect the aftermath of epidemics of cholera and Spanish influenza. The Second World War appears as an elevated mortality rate at ages 18–35 on both maps, but the excess of mortality is clearly higher for males than for females.

The other strength of the Lexis map is that it reveals age-specific mortality differences which might otherwise go unnoticed. Consider the influenza epidemic of 1918 - the most dramatic

occurrence in the civilized world in the 20th century (with the exception, of course, of the First and Second World Wars). The impact of this epidemic on overall mortality is clearly visible in Fig. 4.1, although it is almost imperceptible from the trends in crude death rates. The crude death rate in 1918 was 13.2 per 1,000 per annum for males and 12.9 for females, while the average mortality rates in the years 1916, 1917, 1919 and 1920 were 13.4 and 13.0, respectively. The difference between the rates is negligible, which gives one the impression that the mortality regimes in 1918 and in adjacent years were similar. On the other hand, if we compute the crude mortality rate for ages 20–40 only, the difference is marked: 10.3 versus 5.6 for males and 9.0 versus 5.4 for females.

Any presentation of the Danish mortality surface would not be complete without a discussion of the factors governing the two most important features of these maps: a) the decline in mortality at the end of 19th century and b) the stagnation in mortality in the last decades of the 20th century. I will discuss the first phenomenon only briefly here because a full examination of it falls well outside of the scope of this study. I will explore the second phenomenon in more detail later on, as there are more statistical data are available which enables us to make comparisons between countries and to investigate differences in cause-specific mortality and in social-economic variables.

There is a vast amount of literature devoted to mortality decline in the 19th century and at the beginning of the 20th century but there has hitherto been no systematic study of the mortality transition in Denmark. The most prominent work in this field is perhaps the monograph of Matthiessen (1970), which focuses on the construction of total mortality, fertility and migration schedules for Denmark by five-year age groups. However, the underlying factors that can shed light on the observed mortality trends received only little attention in his work.

Studies in historical demography indicate that the decline in mortality at the beginning of the 20th century is mainly due to a decline in mortality from infectious diseases; especially the decline in deaths from tuberculosis and diphtheria played an important role. Caselli, for example, argues that the decline in respiratory tuberculosis accounted for over half the gains in life expectancy between 1871 and 1911 in England and Italy (Schofield (Ed.) et al., 1991). Nevertheless, it is unrealistic to select these diseases as unique factors behind the mortality transition. The reductions in death rates from other infectious diseases such as the plaque, smallpox, cholera, typhus, typhoid fever, measles, whooping cough and malaria were also substantial and the decline in respiratory diseases such influenza, bronchitis and pneumonia played a significant role, as well. At the same time mortality from diseases of the circulatory system and cancer increased, which led to a change in the structure of cause-specific mortality from infectious to degenerative diseases.

Data on cause-specific mortality for European countries of acceptable quality are available going back to about the middle of the 19th century. These data have been extensively exploited in historical demographic studies because of their accessibility. It is obvious that the examination of long-term trends in cause-specific mortality is only the first step in a demographic analysis, since the trends by themselves do not reveal the causative mechanisms of the mortality decline. In view of this fact, many explanations have been put forward. All of them are based largely on known facts and I will discuss the most important ones.

McKeown (1976) argues that the mortality decline can be mainly attributed to improvements in nutrition during the 19th century. Nutrition seems to have a strong influence on the incidence, severity and lethality of such diseases as tuberculosis, bacterial diarrhoea, cholera, measles and, to some extent, diphtheria and influenza. At the present time, malnutrition, especially protein-energy malnutrition, is thought to be linked to impairments of the immune system, particularly to the thymus gland and lymphoid tissues (Lunn in Schofield (Ed.) et al., 1991).

Other studies demonstrate that the decline in mortality took place chiefly due to improvements in the sanitary environment and public hygiene, which are usually associated with drainage and sewage disposal, a sufficient supply of safe drinking water, with clean and paved streets. The example of sewage conditions can be found in a survey of six European countries conducted by Thomas Legge (1896) in the earlier 1890s[26]. In Copenhagen, for example, the sewage disposal system was far from meeting the standards of the time: sewage conducted straight into the harbor. In addition, in some sections of Christiania (district of Copenhagen) drainage and pavement had not been completed. Johansen and Boje (1986) provided another example of living conditions in Odense at the beginning of the 19th century, where sewage flowed down the street into a trench. They described conditions in Hans Jensens Stræde, where H. C. Anderson was born and where H. C. Anderson Hus (city museum) is now located. Today this street is the biggest tourist attraction in Odense.

Public health measures and effective governmental interventions also played a significant role. The classic example is the outbreak of cholera in Hamburg in 1892. The epidemic affected at least 16,926 of a total population of 625,000, and more than 8,605 people died of the disease (officially reported numbers). In contrast, only six cholera deaths were reported in Bremen. The number of cholera deaths in Hamburg exceed the number of deaths from this disease in all previous

[26] Woods, Robert. Public Health and Public Hygiene. In Schofield (Ed.) et al., 1991; 233-247.

epidemics together. The local government was completely responsible for the epidemic in that it ignored the first cases of the disease so as not to disrupt trade and business life in the city. The population was not informed about the protection measures recommended by Koch (in fact, no proper attention was paid to his instructions against cholera at all). In contrast, the medical authorities in Bremen were convinced about Koch's recent discoveries and of the effectiveness of protective measures such as quarantine, isolation, water and milk boiling, hand disinfecting, and the avoidance of crowding. Before the epidemic a hospital had been built in Bremerhaven and a disinfection plant acquired. When the disease struck, the population was immediately informed about protective measures and the proper instructions were distributed. The results are self-evident (Bourdelais, P.; Woods, R. in Schofield (Ed.) et al., 1991).

Another group of factors which is frequently discussed in connection with mortality decline is the rising standard of living and improvements in housing and working conditions. Dr. Edward Smith (1876) wrote: '… the peasant, gaining immunity from his open-air existence, may escape the noxious results of stagnant drains and even of impure water; but it is his sleeping accommodation which produces the most insidious (and often fatal) results upon his health. Overcrowding has probably killed more than all other evil conditions whatever.'[27] Improvements in housing conditions have usually been accompanied by legislative acts which set the standards for new buildings, e.g., the Housing Act of 1858 in Denmark or the Act of 1902 in France. On the other hand, mortality was consistently lower in rural than in urban areas despite the generally worse housing conditions. The main reason seems to be that peasants spent most of their time working outside in the fresh air so their exposure to environmental hazards was lower than for town workers. It has been suggested 'that the house itself was not the principle determining factor' and that there are other factors which are closely linked with poor housing conditions such as poor sanitation, malnutrition, etc. (John Burnett[27]).

Advances in medical science also played an important role in mortality decline. The introduction of a vaccine against smallpox in 1796 by Edward Jenner, the isolation of quinine in 1820 by Caventon and Pelletier (malaria treatment), Koch's discovery of the bacterial nature of cholera in 1884, the work of Behring on an immunization against diphtheria, the discoveries of Louis Pasteur, which had a profound influence on public health through the establishment of

[27] Burnett, John. Housing and the Decline of Mortality. In Schofield (Ed.) et al., 1991.

principles of pasteurization, antisepsis and asepsis - all these advances contributed indisputably to the observed mortality decline.

The role of medical intervention seems, however, to be less significant than the dissemination of medical knowledge and rules of public hygiene among people. For example, McKeown (1976) argued that advances in medical science cannot be credited as being the principle factor responsible for the decline in mortality since many diseases were already declining long before effective medical therapy had become available. The first antidiphtheritic serum was available in Denmark in the summer of 1895 but Thorvald Madsen (1956), who helped to prepare the serum, noted that the mortality rates had already fallen before it had become available. In view of this fact Madsen and Madsen (1956) attributed the decline in mortality from this disease in the years around 1895 to changes in the type of diphtheria bacillus rather than to the introduction of serum therapy (Lancaster, p110, 1990). Jean N. Biraben in his work *Pasteur, Pasteurization, and Medicine* (in Schofield (Ed.) et al., 1991) states 'In Western Europe mortality had begun to fall during the 1870s, but its decline reached unprecedented levels from 1885 onwards. As it is clear …, it was not vaccines or sera which were responsible for this fall that has continued into our own period, but the spread of cleanliness, disinfection, antisepsis and asepsis'.

Other factors which have been put forward to explain the decline in mortality are changes in disease virulence, changes in climate, rising levels of social income and even the influence of sun activity. Attempts to separate factor-specific influence and to assign some numeric measure to the contribution of each individual factor to the decline in mortality are hampered by the lack of reliable data and the gap in knowledge about causative mechanisms. All historical demographers seem to agree that this is an unrealistic and futile task.

There has been less published on historical Danish developments than on other countries despite the rich volume of statistical data. Death counts by cause, for example, have been publishing for urban areas since 1860 and for the whole country since 1921. More scanty and less reliable data on causes of death can be found in parish reports (Johansen, 1996). Andersen (1973) has put forward the agricultural reforms as the principal factor behind the decline in mortality from 1735 to 1839 (more modern periods have not been analyzed in his work). He also emphasizes the importance of economic growth, smallpox vaccination, and improvements in hygiene and housing conditions. He argues that hospitals did not contributed to the decline: on the contrary, admission to a hospital increased the risk of becoming infected. Lancaster (1990) has maintained that the experience of Denmark is similar to that of most European countries whereas the other

Scandinavian countries should be treated as isolated areas. Unfortunately, this statement is not supported by any statistical material.

The analysis of the Danish mortality surface suggests that the Danish population was among the mainstream of late 19th century European mortality transitions. Moreover, there is some evidence that Denmark was ahead of many countries and that Danish gains in life expectancy were significantly higher than elsewhere. For example, Vallin (in Schofield (Ed.) et al., 1991) discusses life expectancy in different European countries on the eve of the First World War. It follows from his analysis that life expectancy in Denmark was the highest in Europe. Part of Vallin's table is reproduced in Table 4.1.

Table 4.1 Life expectancy in the beginning of 20th century[28].

Country	Period	Life expectancy at birth
Denmark	1911–15	57.7
Norway	1911–21	57.2
Sweden	1911–20	57.0
Netherlands	1910–20	56.1
Ireland	1910–12	53.8
England and Wales	1910–11	53.5
Switzerland	1910–11	52.3
France	1908–13	50.4

4.2.2 Mortality Progress

Current progress in mortality is an important indicator for demographers since it shows the tendency of death rates to increase or decline. Information about mortality progress is frequently used in mortality and population projections. By using the data from the Danish mortality database it is possible to estimate the surface of mortality progress and to discover the age-year domains with different mortality trends. The Lexis map shown in Fig. 4.2 is quite new in demographic research in the way it permits us to look at mortality changes over time.

The procedure used for estimating the mortality progress surface is described in Appendix 4.1. Mortality progress rates have been estimated for every year and age using 5 preceding and 5 following years. Thus, a single estimate of mortality progress is based on 11 death rates centered at the year for which the estimate is produced. In Fig. 4.2 only estimates based on the complete 11-

[28] Reproduced from Vallin, J. in Schofield R. (Ed.) et al., 1991, p47.

year time series are shown, so the map covers the period from 1840 to 1990, which is smaller than the period covered by the mortality database.

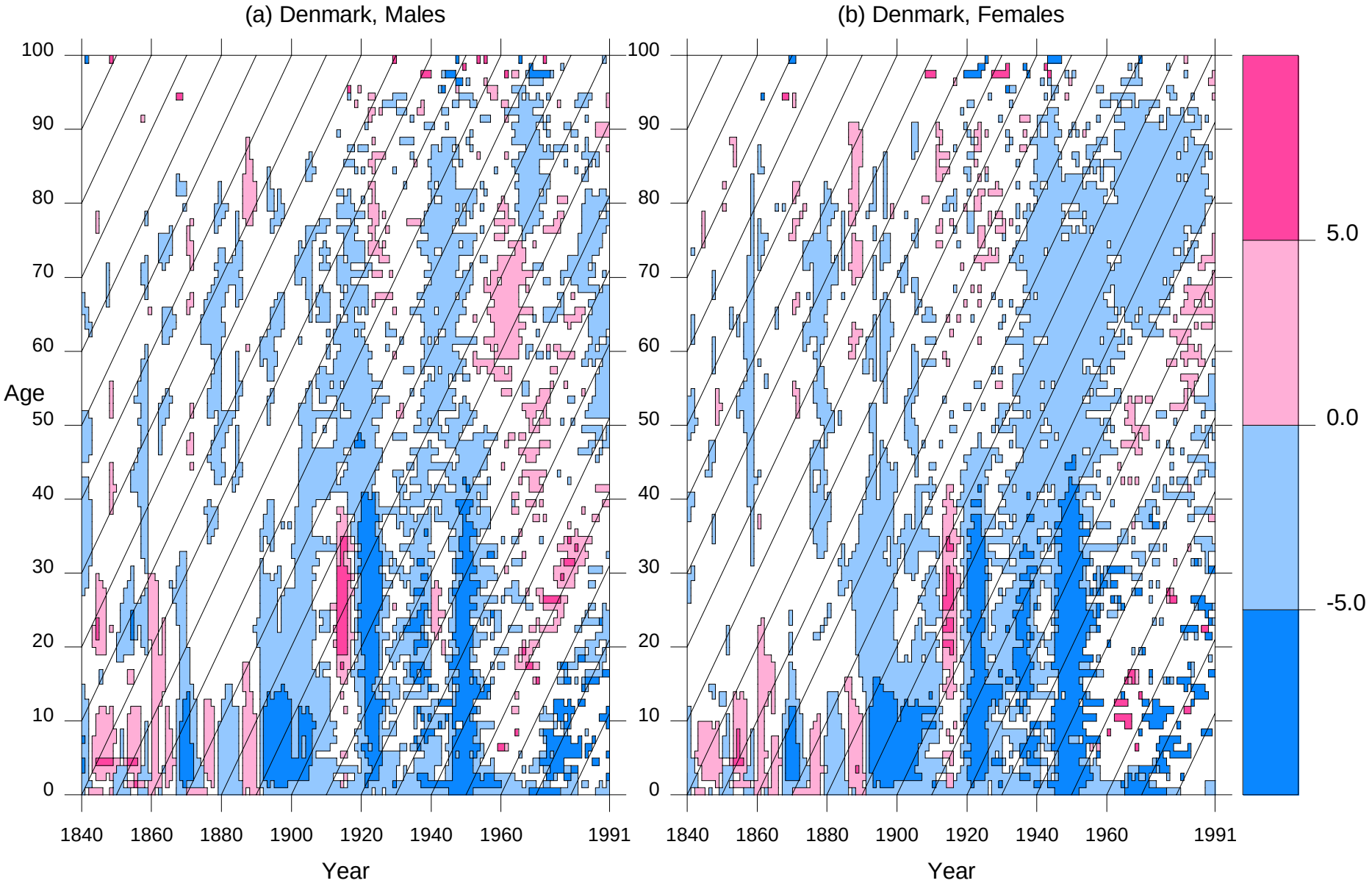
The scale shown on the right divides all mortality progress estimates into four areas. Light magenta (0.0) and dark magenta (5.0) depict the age-year domains in which mortality was increasing; light magenta (0.0) is used for areas with a rate of increase less than 5% and dark magenta (5.0) for areas with a rate of increase over 5%. Light cyan (-5.0) and dark cyan (<-5.0) show areas with declining mortality. The rate of decline is less than 5% for the light cyan areas and more than 5% for the dark cyan areas. The color white corresponds to the estimates which were not significant at the 10% level or where the procedure could not produce an estimate because of a lack of the data.

We turn now to the discussion of the main features of the Lexis maps. The dark cyan blur at ages 0–15 and in the years around 1900 marks the onset of the persistent mortality decline in the Danish population. We can see that high rates of mortality improvement became evident in the early 1890s, both in the male and female populations. Prior to that time mortality changes were of a sporadic nature, with distinctly expressed periods of increasing and declining mortality. The rates of improvement were highest (> 5%) at ages 1–15, with a peak 8–10% at about age 5. Progress of up to 2% per year is also evident at ages up to 40 in the female population and up to 30 in the male population. At higher ages the improvements were less significant.

Over time the area with high rates of mortality improvement spread out to higher ages, and rates of progress above 5% are to be observed up to the age of 40 and until the middle of the 1950s. This drastic mortality decline was interrupted only twice during these 60 years: first by the influenza epidemic of 1918 and second by the Second World War. The pattern of mortality decline at these ages was similar for males and females - but not above the age of 40. After 40 appreciable mortality progress (1–5%) can be observed in the periods 1900–1920 and 1940–1950 for males and in 1935–1960 for females.

Starting in the 1960s mortality decline decelerated significantly, and there was even some mortality increase, as indicated by the light magenta clearly visible on the maps. An especially strong rise in male mortality (1.5%) is to be seen in the period 1955–1970 and at ages 55–75. I conclude that the late 1950s mark the start of stagnation in the Danish mortality decline. However, stagnation

Figure 4.2 Mortality Progress, %.



was not uniform over age. During this period, striking mortality progress can be observed at older ages. Especially eye-catching is the mortality decline (about 2%) in the female population at ages 70–95 in the period 1940–1990. The rate of decline in the male population was less significant: a comparable level of decline is visible only in the years surrounding 1970. The highest rates of progress in oldest-old mortality are found in the period 1965–1969, where the rate of progress at age 80 was about 3% for males and 5% for females.

In the late 1980s another change in the dynamics of death rates took place. Male mortality at ages 50–80 started to decline while female mortality at ages 60–80 began to increase. In addition, the rate of mortality progress at older ages fell to zero, indicating the onset of stagnation in oldest-old mortality. In view of the importance of mortality progress at advanced ages for the projection of the oldest-old population, I computed standardized mortality rates for ages 80+ in order to survey mortality trends for the period 1990–1995. In the female population death rates remained at a constant level after the year 1990, and in the male population a slight increase can be observed.

To complete the depiction of mortality progress, cohort lines have been added to the Lexis maps. It can be seen in Fig. 4.2(b) that the female mortality increase runs along the cohort lines concentrating around 1920. In the male population a similar effect is noticeable for the cohorts born around 1950. This pattern calls for explanation, but so far there have been no demographic studies which link the rate of mortality progress to events that occurred earlier in life and to cohort-specific characteristics.

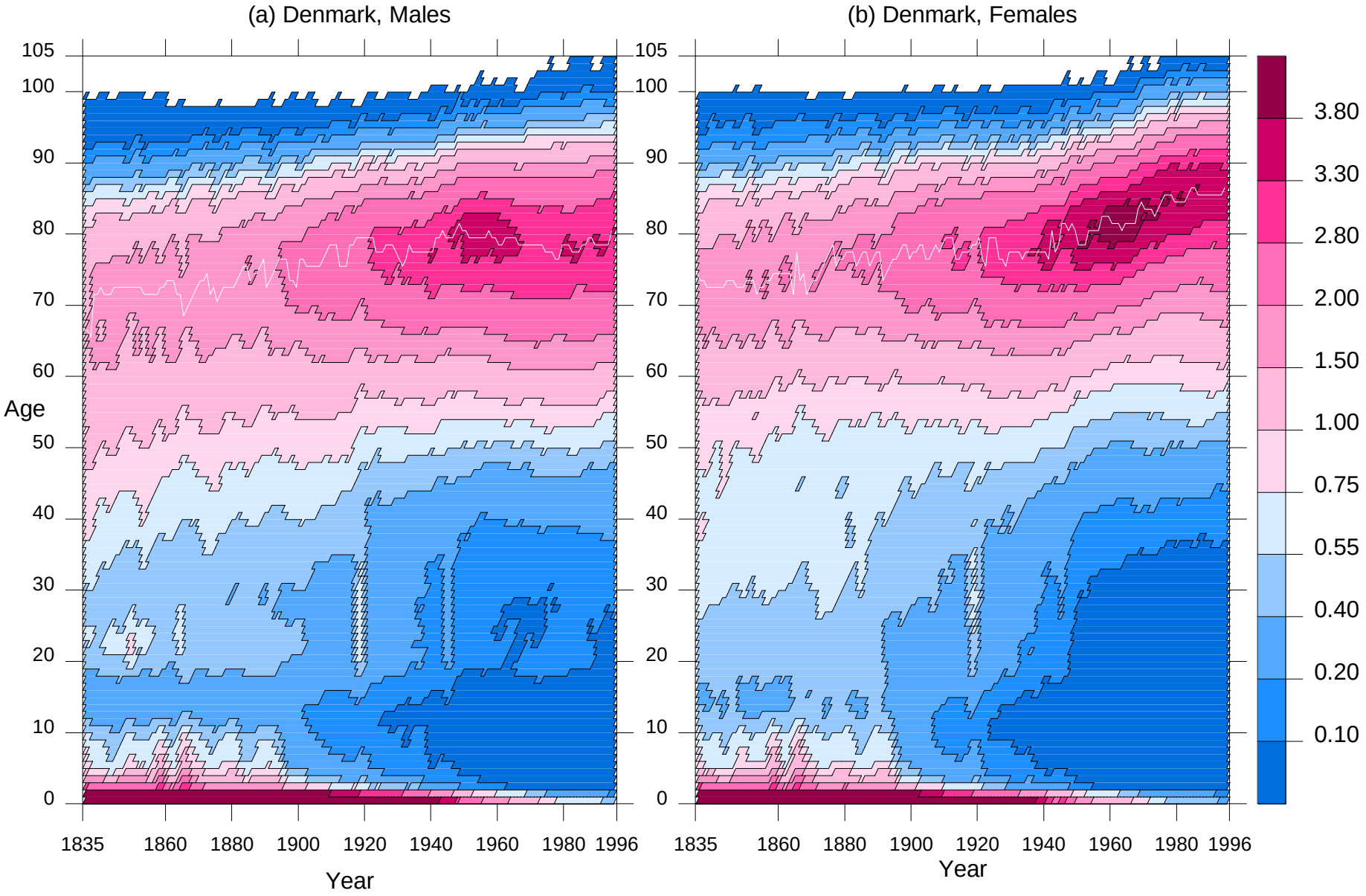
4.2.3 Compression of mortality

There is a lively debate in the demographic literature about whether or not the maximal life span of humans is fixed and whether or not there has been a rectangularization of the survival curve in recent decades (e.g. Fries, 1980; Aarssen and de Haan, 1994; Kannisto et. al., 1994; Curtsinger, et. al., 1992). The data from the Danish mortality database allow us to examine trends in the distribution of deaths, thus providing a historical overview of changes over age and time.

To produce the maps shown in Fig. 4.3, the period life tables were computed by single year and d_x columns were extracted and plotted as a Lexis map. The life tables were constructed by the cohort method. In addition, the death distribution surface has been smoothed on a 3x3 matrix with weights generated by bivariate Epanechnikov kernel (Appendix 4.2).

The scale legend shows the percentage levels of the death distribution, i.e., all values of d_x (%) falling in the same range are plotted with the same color as delineated by the scale. For example, the maximum female death distribution at older ages in 1835 is observed at age 74; the

Figure 4.3 Death distribution, %.



Smoothed with 3x3 bivariate Epanechnikov kernel

percentage of life table deaths at this age is about 1.66%. In 1994 the maximum is at age 86, where the proportion of deaths is 3.6%. In order to emphasize the evolution of the maximum of the death distribution, the white line connecting the ages with the maximal proportions of the life table deaths was added to the maps. Since we are concerned with the compression of mortality, the maximum of the death distribution at adult ages has been stressed rather than infant mortality.

As is evident from Fig. 4.3, the mode of life table death distribution at older ages has increased substantially since the middle of the 19th century. For males it has increased from 70 to 78 years of age and for females from 73 to 85 (1.5 times higher than for males). However, the pattern of increase was not parallel for both sexes. The main increase in the mode in males occurred before 1940. There have been no significant changes since that year. For the female population the increase has been more persistent over time and has even accelerated since 1940.

The pattern of mortality compression is also revealed by Fig. 4.3. The areas in cyan tones in the lower right-hand corner correspond to exceptionally low levels of death density. In the period 1970–1994 the number of deaths below age 50 was about 8% for males and 5% for females, whereas in the 19th century these numbers were 47.5% and 45.5%, respectively. If we exclude deaths at age 0, the difference is still marked: 7.5% (39.5%) for males and 4.5% (38.5%) for females. This huge difference resulted from the rapid progress in the mortality of children and young adults. Progress at high ages has been less dramatic. Undoubtedly, this pattern of mortality progress led to a compression of the death distribution at age about 80, which prevailed until the mid-1960s both for males and females. As more and more deaths have become concentrated at ages close to 80, new domains of high death density have emerged on the Lexis maps. These areas appear in magenta in Fig. 4.3.

However, along with the process of mortality compression, the proportion of deaths at very old ages has increased as well. On the Lexis maps this is indicated by the rising contour lines at age 80 and above. Until the 1960s mortality progress at the highest ages was negligible compared with that in the lower age groups. The death distribution was becoming more and more compressed at age 80 despite the increasing proportion of deaths at ages 80 and above. However, starting with this period, mortality reductions at age 70 and above became more appreciable and had a profound influence on the tail of the death distribution. As a result, the area with the highest death density disappeared on both maps and the age at maximal death density moved toward the higher ages in the female population.

These findings are summarized in Table 4.2. The observation period has been divided into four time periods and the proportion of life table deaths computed for ages below 75, 75–85 and for

age 85 and above. We can see the phenomenon described in the previous paragraph manifested in the trends in the proportion of deaths in the age group 75–85. It increased until the 1970s and then started to decline. The proportion of deaths in the lowest age group declined continuously (except for males in the last period) and the proportion in the highest age group increased consistently.

Table 4.2 Proportions of the life table deaths in Denmark.

	Males			Females		
Period	0–75	75–85	85+	0–75	75–85	85+
1835–1900	81.6	14.3	4.1	77.1	16.8	6.0
1900–1950	63.8	26.3	9.9	59.8	28.2	12.0
1950–1970	51.5	32.5	16.1	40.2	36.8	23.0
1970–1995	51.2	31.1	17.7	33.4	32.0	34.6

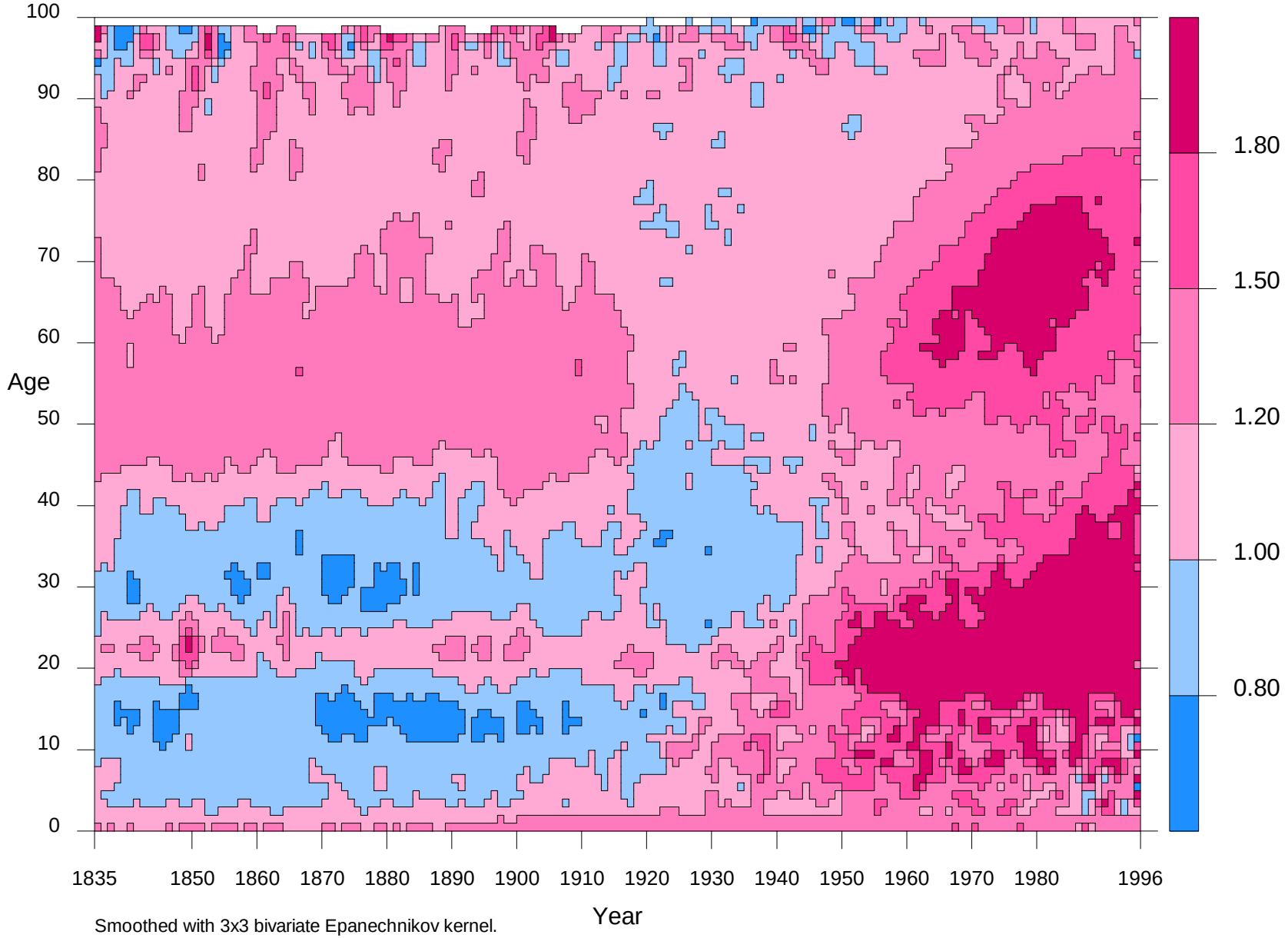
In sum, the findings from Danish mortality data suggest that there have been two processes operating simultaneously in recent decades: compression of mortality and reduction of oldest-old mortality. The first process was dominant in the earlier periods, and the death distribution became more compressed over time, reaching its peak in the 1960s. After that time the decline in oldest-old mortality took over and the proportions of deaths at the oldest ages rose substantially, which reduced the level of compression. Moreover, the age with the highest death density (at adult ages) has been gradually increasing over time, especially in the female population. For males the mode of death distribution stagnated in the 1940s and even declined in the later 1970s. In the last decade there has been an increase.

4.2.4 Sex ratio of mortality

As described above, the decline in mortality in recent decades was greater for the female than for the male population. In order to examine the sex differences in survival more closely, a surface of sex ratios of Danish mortality was estimated, cf. Fig. 4.4. The mortality ratio surface was computed with the kernel estimation procedure (Appendix 4.3), using a 3x3 smoothing matrix and Epanechnikov bivariate kernel weights. At boundaries where no complete data for smoothing are available - that is, at age zero and in the years 1835 and 1995 - the ratio of age-specific death rates was plotted instead of kernel estimates.

The scale divides the surface into 6 areas. The colors light and dark cyan are used to depict excess of female mortality, while magenta tones are used for the areas with excess male mortality. The level of equal mortality in two populations can be followed with the contour line at level 1, which demarcates the magenta and cyan domains. Besides a qualitative description, the scale also

Figure 4.4 Sex ratio of Danish mortality.



Smoothed with 3x3 bivariate Epanechnikov kernel.

provides information about the extent of excess mortality. The dark magenta (1.80) area, for example, comprises ratios where male mortality was 80% higher than female mortality.

As can be seen in Fig. 4.4, there are three distinct periods with clearly different patterns of male-female mortality differences. Until the 1920s females had a disadvantage at ages 5–20 and 25–40 with the excess mortality of about 20%. This can be largely attributed to the complications in connection with childbearing, but it is not the complete explanation. There is an excess of male mortality in infancy, in the earlier twenties and at ages over 40. The most significant differences were found at ages 45–65, where male death rates outnumbered female rates by about 20–50%.

This pattern of survival was remarkably stable over a period of 90 years. The first signs of changes in the mortality regime did not start to become evident until the early 1920s. The period from 1920 to 1950 is characterized by minimal sex differences in mortality. Even if female mortality was generally lower than male mortality, the sex ratios fall within the 20% range, which makes this period remarkable in that there was a high degree of similarity in the mortality regimes of both populations.

The end of the Second World War clearly marks the onset of new regime in the sex differences in mortality. Already in the 1950 male death rates outnumbered female rates at virtually all ages. Especially high differences are to be observed at ages 50–60 and at ages close to 20. These two age groups acted as starting points for two areas of excess male mortality that emerged later in time: one at adult ages and another at young adult ages. Both areas are colored dark magenta (1.8), which corresponds to an excess of male mortality of over 80%.

At the young ages the area with excessive male mortality has been spreading over time to cover more ages. At present time it encompasses the age group 15–40. For comparison, in 1950 excess male mortality of more than 80% occurred only at ages 18–23.

In the age group 50–60, the gap between male and female mortality rose over time, simultaneously moving to the higher ages, i.e. somewhat along the cohort lines. The sex differences reached a peak in 1980 and at age 70, and then decline. In this period male death rates were approximately double those of females. Toward the 1990s, the area with the highest excess of male mortality disappeared completely. This pattern is clearly demonstrated by the dark magenta (1.80) oval in the upper right-hand corner of Fig. 4.4.

The decline in sex differences at ages 60–80 was the main reason for a convergence in life expectancies for the male and female populations of Denmark in recent years. Death rates at young and young adult ages are very low nowadays and changes in these rates affect life expectancy at birth in a less notable way. The convergence in life expectancies can also be observed in other

countries as well. To reveal the age-specific mortality differences, I have produced similar mortality-sex-ratio maps[29] for Sweden[30] and the Netherlands[31]. It turns out that the global pattern of mortality ratios is strikingly similar between countries. There are still some differences between the maps but there are far more similarities to be observed. It might be the case that there are certain factors that affect mortality in some uniform way, thereby maintaining the fixed pattern of male-female differences over various countries.

4.2.5 The oldest-old population

The decline in mortality together with the decline in fertility had a profound impact on the population structure of Denmark. The contemporary population distribution is characterized by reduced proportions of young ages and increased proportions of the elderly. Fig. 4.5 shows the changes in the population distribution relative to the average levels of 1835–1920. To produce Fig. 4.5, the single age population distribution on 1 January was divided by the average distribution in the years 1835–1920 and subsequently smoothed on a 3x3 matrix with Epanechnikov weights (Appendix 4.2). The period 1835–1920 was selected after visual examination of changes in the population distribution. Until 1920, time trends in the distribution had been rather moderate and the distribution itself relatively stable.

Since the 1920s the population distribution has changed dramatically both for male and females. An especially marked increase is visible in the proportions of oldest-old; these emerging areas are colored dark magenta in Fig. 4.5. The magenta (5.00) and dark magenta (10.0) areas correspond to the ages where proportions are 5 and 10 times greater than in 1835–1920. The increase in the proportions of oldest-old has been accompanied by corresponding reductions in proportions of ages below 30 (by a factor about 1.5–2). This phenomenon is portrayed by the blue areas in Fig. 4.5 - the areas where the proportions are lower than in 1835–1920. The influence of the ‘baby-boom’ on the population structure is also clearly manifested by the strong diagonal patterns starting in the late 1940s.

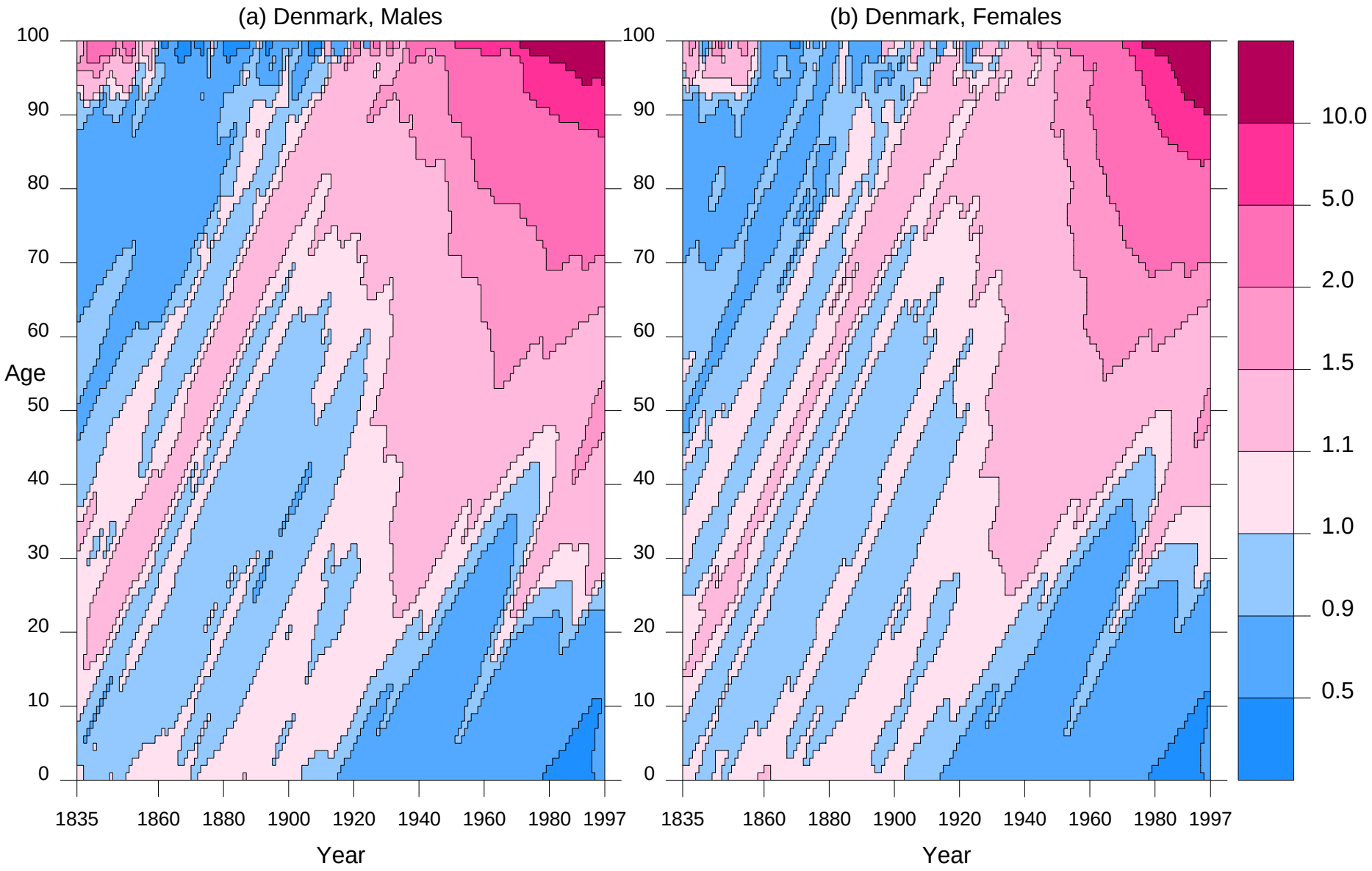
---

[29] The maps can be requested from the author at andreev@demogr.mpg.de.

[30] The data were made available by J. Wilmoth, Berkley Mortality Database, USA.

[31] The data were made available by E. Tabeau, NIDI, the Netherlands.

Figure 4.5 Ratio of the population distribution to the average levels in 1835-1920.



4.3 Mortality differences between Denmark, Sweden, the Netherlands and Japan

4.3.1 Excess Danish Mortality

As it was discussed in the previous sections, Danish life expectancy gains have been fairly moderate in the last few decades. This mortality development is atypical for the developed countries, and we shall explore it here in more detail. Table 4.3 shows the increase in life expectancy in OECD countries in the period from 1970 to 1995. Danish male life expectancy rose by 1.8 year and female life expectancy by 1.9 year. Among the male population only Poland and Hungary exhibit lower life expectancy gains; in the female population the Danish gains were the lowest of all countries. Denmark occupied the 4th (males) and the 6th (females) places in the table in 1970, and the 18th and 19th places, respectively, in 1995.

This development has not gone unnoticed, and in 1993 the Danish Ministry of Health undertook a large study to find out the reasons for this adverse trend in Danish life expectancy. As a result fourteen books were published in 1993 and 1994 (Sundhedsministeriets Middellevetidsudvalg, 1993) with the focus on cause-specific death rates for broad age groups and on social-economic and life style differences between Denmark and European countries. Despite the large volume of the material presented, the age-specific mortality differences did not received the proper amount of attention in this study. Nonetheless, this analysis can shed some light both on which age groups have experienced more excess mortality and on the time when the problems started to emerge.

Age specific differences in Danish survival can be revealed by estimating the mortality ratio surfaces (Appendix 4.3). In the present analysis the mortality ratio surfaces of Denmark to Sweden, the Netherlands and Japan[32] were estimated for the year 1950 and onwards and for age 30 and above, separately for males and females. Fig. 4.6 shows the ratios of the central death rates, which were significant at the 1% level. The values of the excess mortality can be followed with the scale legend.

[32] Data for Japan were made available by J. Wilmoth, Berkley Mortality Database.

Table 4.3 Improvements in life expectancy in the period from 1970 to 1995[33]

	Males	1970	1995	Diff.		Females	1970	1995	Diff.
1	Mexico	58.2	69.5	11.3		Mexico	62.5	76.0	13.5
2	Korea	59.8	70.0	10.2		Korea	66.7	76.0	9.3
3	Australia	67.4	75.0	7.6		Japan	74.7	82.8	8.1
4	Japan	69.3	76.4	7.1		Portugal	71.0	78.6	7.6
5	Austria	66.5	73.5	7.0		Australia	74.2	80.9	6.7
6	Portugal	65.3	71.5	6.2		Greece	73.6	80.3	6.7
7	United Kingdom	68.6	74.3	5.7		Austria	73.4	80.1	6.7
8	Germany	67.4	73.0	5.6		Spain	75.1	81.2	6.1
9	Belgium	67.8	73.3	5.5		France	75.9	81.9	6.0
10	France	68.4	73.9	5.5		Belgium	74.2	80.0	5.8
11	Luxembourg	67.0	72.5	5.5		Germany	73.8	79.5	5.7
12	New Zealand	68.3	73.8	5.5		Luxembourg	73.9	79.5	5.6
13	United States	67.1	72.5	5.4		Switzerland	76.2	81.7	5.5
14	Greece	70.1	75.1	5.0		Ireland	73.2	78.5	5.3
15	Switzerland	70.3	75.3	5.0		New Zealand	74.6	79.2	4.6
16	Ireland	68.5	72.9	4.4		United Kingdom	75.2	79.7	4.5
17	Sweden	72.2	76.2	4.0		United States	74.7	79.2	4.5
18	Czech Republic	66.1	70.0	3.9		Sweden	77.1	81.5	4.4
19	Norway	71.0	74.8	3.8		Czech Republic	73.0	76.9	3.9
20	Netherlands	70.9	74.6	3.7		Netherlands	76.6	80.4	3.8
21	Spain	69.6	73.2	3.6		Norway	77.3	80.8	3.5
22	Denmark	70.7	72.5	1.8		Poland	73.3	76.4	3.1
23	Poland	66.6	67.6	1.0		Hungary	72.1	74.5	2.4
24	Hungary	66.3	65.3	-1.0		Denmark	75.9	77.8	1.9

Light magenta (1.0) is used for the values of excess Danish mortality below 30%, median magenta (1.3) for excess in the range of 30–50% and dark magenta for mortality ratios with values over 50%. The cyan tones show areas where Danish mortality was in fact lower than in another country, e.g. Japan. Because interpretation of the contour maps is straightforward only a brief discussion of each of the 6 maps is given here:

a) **Denmark to Sweden, Males.** In the period 1950–1960 mortality in both countries was virtually the same and no significant deviations are to be observed. Starting in the 1960s the first signs of the excess mortality at ages close to 60 became evident. The pattern of excess Danish mortality was rather sporadic and the values were about 20%. Starting in the 1980s the situation worsened and the area of excess mortality spread to the higher and lower ages. At the same time the mortality difference at age 60 rose to 40%. Up to 1996 there is a tendency of increasing mortality differences and no reverse trends are perceptible.

[33] Source: OECD Health Data 1997. OECD Health Policy Unit 2, rue André Pascal F-75775 Paris Cedex 16. Web site: http://www.oecd.org/.

b) **Denmark to Sweden, Females**. The onset of systematic excess mortality lies somewhat later than for males. With the high degree of confidence we can point to the late 1960s - the time when the Danish female mortality started to outnumber Swedish mortality. The dynamics of the process is essentially the same as for males, i.e., the excess spread out to cover more ages and the differences in the middle of the excess mortality area were aggravating. However, the process was more rapid and led to higher mortality differences in the most recent years. In the year 1995, for example, excess female mortality at ages 50–70 was more than 50%, whereas no such level of excess is found in the male populations.

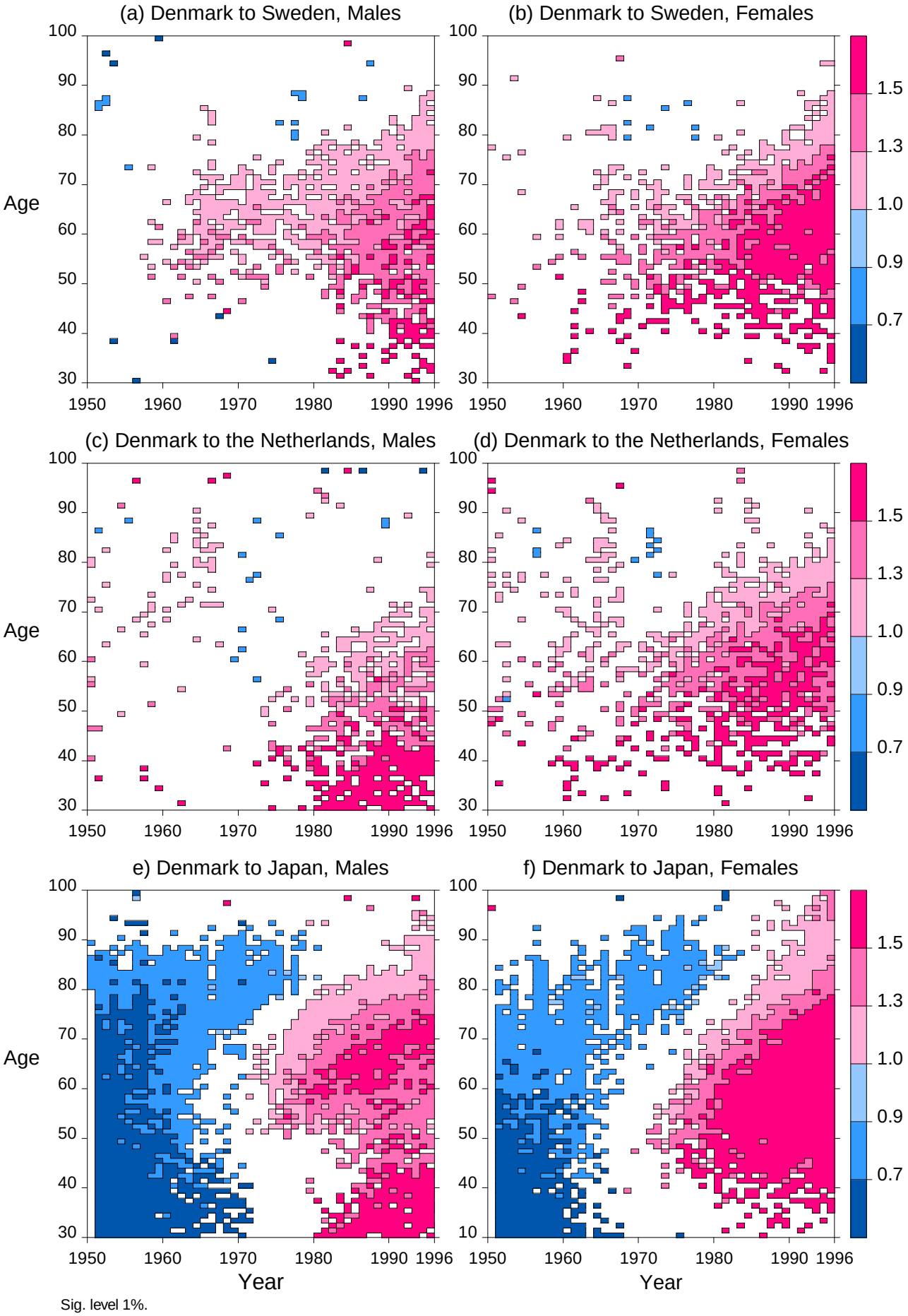
c) **Denmark to the Netherlands, Males**. Excess Danish mortality is also visible on this map but contrary to the Swedish comparison, the excess is observed at ages 30–50 and it starts in the 1980s. At older ages excess mortality is less marked. Here the difference in mortality is the lowest of all 6 comparisons discussed here.

d) **Denmark to the Netherlands, Females**. The general pattern of excess mortality is quite similar to the Swedish pattern. It is evident that the process starts in the earlier 1970s and spreads out in the course of time. The magnitude of the excess is less dramatic but it is concentrated in the same age groups as in the Danish-Swedish case.

e) **Denmark to Japan, Males**. The pattern of mortality ratio in the period 1950–1970 is quite different from that observed in comparisons with European countries. During this time mortality in Denmark was much lower and the excess of Japanese mortality is observed at virtually all ages except the oldest-old. Starting in the 1970s the pattern was completely reversed due to a rapid decline in Japanese mortality. Excess of Danish mortality began to occur at age 60 in the earlier 1970s and spread out over time. Concurrently, another area of abundant mortality began to occur at young ages. In the most recent years, the excess of Danish mortality is apparent at all ages below 90, but especially in the age group 65–80 and 30–45, where the mortality differences are over 50%.

f) **Denmark to Japan, Females**. The pattern of mortality differences observed on this contour map is very close to that found for Japanese males. Until 1970 Denmark had an advantage in mortality over Japan. Then an area of excess Danish mortality at ages 50–60 started to appear. The excess of Danish mortality is the most drastic and the most spread out of all comparisons presented in Fig. 4.6. In 1995 mortality in Denmark was more than 50% higher at all ages in the range from 40 to 80. In contrast to the findings concerning with the male populations there is no distinct mortality excess at young adult ages; mortality seems to be higher at ages 30–35 but the

Figure 4.6 Mortality Ratio.



Sig. level 1%.

difference is less marked.

To summarize my findings, I conclude that despite certain differences in the pattern, excess Danish mortality is similar in all country comparisons presented here. The first indications of excess mortality emerged in the later 1960s and in the narrow age group 50–60. Over time, mortality developments in the countries chosen for comparison led to an expansion of the area of abundant Danish mortality to the higher and lower ages. This tendency remained unchanged until the last year for which there is data available. In the female populations the expansion to the higher ages has been more dramatic than to the lower ages but the most striking differences are still observed at about the age of 60 - just as was the case 25 years ago. In the male populations the pattern differs from country to country.

4.3.2 Analysis of cause-specific mortality

The mortality database maintained by the World Health Organization (WHO) makes it possible for us to analyze cause-specific mortality differences between countries. These data have been used extensively in studies devoted to the stagnation of Danish life expectancy (Sundhedsministeriets Middellevetidsudvalg, 1993; Bjerregaard and Juel, 1993; Bjerregaard and Juel, 1994). The analysis presented in this section has two main objectives. The first is to decompose excess Danish mortality by causes of deaths. The second is to survey the trends in cause-specific mortality using the most recent data.

For this purpose I have downloaded the mortality data from the WHO web site (www.who.int/whosis/mort/) and extracted the relevant country-specific data from the raw mortality files. I then created an abridged list of 25 causes of deaths (Appendix Table 4.1) and aggregated the death counts using this custom classification. The causes of deaths included in the abridged list were selected after a careful examination of scientific publications related to the current analysis.

The method of excess mortality decomposition I used here is very simple, and the interpretation of the results is based on the assumption of independence of causes of death. The excess mortality for a given period and in a certain age group can be computed as

$$\rho = \left(\frac{m}{m'} - 1 \right) \cdot 100\% \qquad (4.1)$$

where m and m' are the central death rates (cf. e.g. Chiang, 1984) in Denmark and the country selected for comparison, respectively. The mortality ratio is then:

$$\frac{m}{m'} = \frac{\sum_i m_i}{\sum_i m_i'} = 1 + \frac{\sum_i \Delta_i}{\sum_i m_i'} \tag{4.2}$$

where m_i is the mortality rate from the i th cause of death and $\Delta_i = m_i - m_i'$ is the absolute difference in mortality from the i th cause of death in Denmark and another country. Finally,

$$\rho = \sum_i \rho_i \tag{4.3}$$

where $\rho_i = \frac{\Delta_i}{\sum_i m_i'} \cdot 100\%$ is the contribution of the i th cause of death in the total excess mortality.

As can be seen in Fig. 4.6, the most striking differences in mortality are observed at ages 50–70, and this pattern has remained more or less stable since 1970. Following this observation I applied this method to decompose mortality differences in Denmark and other countries for the period 1985–1993 for 4 age groups: 50–54, 55–59, 60–64 and 65–69. Table 4.4 shows the computed values of ρ_i by cause of death, age group, sex, and country. The last row of the table shows the total excess mortality (%) as computed by Equ. 4.3. Positive values of ρ_i, which indicate a significant contribution to excess mortality (>2%), have been highlighted to facilitate the reading of this table. I must note that two sets of estimates of excess mortality (based on WHO data and on the data published by the national statistical offices) are in a agreement here for all populations but the Netherlands. For the male and female populations of the Netherlands, excess Danish mortality is somewhat higher if it is estimated from the WHO data.

Table 4.4 contains a great deal of material which the interested reader will wish examine for himself. Here a few general comments:

a) **Denmark vs. Sweden, Males.** The main contribution to excess Danish mortality stems from higher mortality from lung cancer (group 2). Approximately 25% of the overall excess mortality can be attributed to this case of death. Higher mortality from the residual group of neoplasm (8) is also noticeable in all age groups, but its contribution is less substantial and it decreases with age. In the lower age groups mortality from ischaemic heart disease (9), cirrhosis of the liver (18) and suicide (21) is also significantly higher in Denmark except in the highest age group, where the difference is barely perceptible. Excess mortality from respiratory diseases (15) increases sharply with age, and it accounts for about 15% of the total excess mortality at ages 60–69.

b) **Denmark vs the Netherlands, Males.** The two most important causes of death with higher Danish mortality are the residual neoplasm group (8) and ischaemic heart disease (9). The latter

accounts for about 20% of total excess mortality at ages 50–54 and 45% at ages 64–69; the residual neoplasm group makes up approximately 15% of excess mortality. In contrast to comparisons with other populations, mortality from lung cancer is actually lower in Denmark than in the Netherlands. This finding is rather surprising since higher mortality from lung cancer in Denmark has been observed in all comparisons - for both the male and female populations. In the age group 50–54, where excess Danish mortality was the highest, significant differences are to be seen in death rates from cirrhosis of the liver (18) and suicide (21). The contribution of these causes of death to excess mortality declines with age and becomes negligible in higher age groups.

c) **Denmark vs Japan, Males.** Here I have found striking differences in mortality from ischaemic heart disease (9), which is responsible for about 70% of excess Danish mortality. The next most important cause of death is lung cancer (2), which is also substantially higher in Denmark than in Japan. The third important cause contributing to excess mortality is diseases of respiratory system (15).

d) **Denmark vs Sweden, Females.** The highest mortality differences exist in the area of cancer, especially lung cancer (2), residual cancers (8) and breast cancer (7). Altogether these 3 causes of death account for about 40% of observed excess mortality. Other important contributors to excess Danish mortality are respiratory diseases (15) and ischaemic heart disease (8). The contribution to excess mortality from the group ‘Bronchitis, emphysema and asthma’ is noticeably higher than from ischaemic heart disease, and it is as important as lung cancer at ages 65–69. At ages 50–59 we see marked differences in mortality from cirrhosis of the liver (18) and suicide (21). At higher ages mortality differences from these causes of death are less pronounced. An excess of mortality from cerebrovascular disease is also noticeable although its contribution seems to be less significant than from the causes mentioned above.

e) **Denmark vs the Netherlands, Females.** The general structure of cause-specific excess mortality is quite similar to that found when comparing Danish and Swedish data. However, mortality differences in ischaemic heart disease at high ages and the differences in suicide mortality in the lower age groups are slightly higher than those found in comparison to Sweden.

f) **Denmark vs Japan, Females.** Here, too, we have similar causes of death as in Sweden and the Netherlands, i.e., lung cancer (2), breast cancer (7), residual malignant neoplasm (8), ischaemic heart disease (9), and respiratory diseases (15). At ages 50–59 a substantial contribution is also provided by the differences in mortality from cirrhosis of the liver (18) and suicide (21).

The residual group of diseases (25) includes the remaining causes of death which have not been classified in any of the other 24 categories. As can be seen in Table 4.4 these residual causes of death provide an appreciable contribution to excess Danish mortality for all countries and for all age groups. The relative importance of the causes of death in (25) reflects above all the fact that different diagnostic and coding practices have been adopted in the different countries. In the case of Denmark, this is a relatively high proportion of deaths that are classified as unknown or ill-defined cases.

Table 4.4 Decomposition of excess Danish mortality by causes of deaths for the period 1985–1993. Males.

The table shows the contribution of a particular cause of death to the total excess mortality(%). A description of the causes of death together with the WHO category numbers is provided in Table 4.1 of the Appendix. Causes of death contributing more than two percent to excess mortality appear in boldface type.

Ages	50–54			55–59			60–64			65–69		
Cause	Sw	Nl	Jp	Sw	Nl	Jp	Sw	Nl	Jp	Sw	Nl	Jp
1	1.12	1.25	-0.08	0.47	0.69	-0.70	0.11	0.11	-1.15	-0.12	-0.05	-1.50
2	**5.86**	-1.46	**7.11**	**8.49**	-0.89	**9.96**	**9.48**	-1.10	**10.11**	**8.87**	-1.71	**9.27**
3	0.02	0.40	0.79	0.13	0.78	1.66	-0.23	0.71	**2.52**	-0.34	1.25	**3.94**
4	0.27	-0.23	0.06	0.78	0.45	0.65	1.04	0.57	1.14	1.11	0.31	1.40
5	0.50	-0.25	-6.63	0.15	-0.65	-7.50	0.01	-0.82	-7.98	-0.30	-0.79	-7.80
6	0.52	0.64	-0.27	0.87	1.18	0.20	0.82	1.12	0.38	1.11	1.31	1.12
7	0.09	0.09	0.09	0.08	0.09	0.09	0.07	0.07	0.08	0.05	0.05	0.06
8	**6.71**	**5.53**	**2.57**	**5.18**	**4.52**	-0.87	**5.03**	**4.22**	0.47	**3.82**	**3.05**	**2.32**
9	**4.70**	**8.18**	**26.12**	**4.17**	**10.56**	**32.27**	**2.16**	**10.34**	**36.63**	0.62	**10.76**	**38.95**
10	-0.70	-2.43	-6.10	-0.55	-2.32	-5.15	0.01	-2.13	-4.48	-0.28	-2.72	-4.98
11	1.31	1.95	-6.11	0.58	1.66	-4.91	0.95	1.80	-3.37	0.88	1.52	-2.67
12	0.58	0.76	1.35	0.17	0.13	1.49	0.77	0.33	**2.77**	0.42	-0.04	**3.35**
13	0.86	0.98	1.03	1.15	1.25	1.47	1.21	1.45	1.92	1.14	1.37	**2.04**
14	-1.36	0.22	-1.25	-1.05	0.38	-1.58	-0.74	0.38	-2.52	-0.70	0.35	-3.92
15	1.40	1.42	2.00	**3.02**	**2.41**	**3.99**	**4.47**	**2.51**	**6.04**	**5.30**	**2.18**	**7.49**
16	0.01	-0.02	0.02	0.01	0.01	0.03	0.06	0.03	0.08	0.01	0.00	0.06
17	0.39	0.37	-0.29	0.10	0.19	-0.64	0.18	0.29	-0.80	0.26	0.19	-1.12
18	**4.67**	**5.96**	-0.12	**2.89**	**4.25**	-0.59	1.63	**2.39**	-0.44	1.03	1.26	-0.26
19	-0.10	0.42	0.13	-0.03	0.39	0.15	0.19	0.44	0.38	0.18	0.45	0.47
20	-0.58	-0.01	-0.84	-0.05	0.12	-0.52	0.06	-0.11	-0.48	0.09	-0.19	-0.29
21	**3.16**	**7.10**	1.87	1.54	**3.50**	0.88	1.42	**2.26**	1.32	0.70	1.23	0.75
22	1.10	1.09	-0.16	0.35	0.49	-0.54	0.44	0.51	-0.19	0.28	0.25	-0.27
23	-0.17	0.55	-0.14	-0.25	0.31	-0.19	-0.02	0.40	0.12	0.11	0.32	0.26
24	-2.91	**2.68**	-0.35	-1.87	1.48	-0.64	-1.49	0.62	-0.97	-0.71	0.31	-0.96
25	**9.88**	**8.95**	**14.56**	**9.34**	**6.85**	**12.67**	**7.97**	**4.80**	**10.61**	**6.44**	**2.96**	**8.86**
Total	37.35	44.15	35.34	35.66	37.85	41.69	35.60	31.19	52.19	29.98	23.63	56.59

Table 4.4 (cont.) Females.

Ages	50–54			55–59			60–64			65–69		
Cause	Sw	Nl	Jp	Sw	Nl	Jp	Sw	Nl	Jp	Sw	Nl	Jp
1	0.26	0.30	-0.47	0.28	0.43	-0.49	0.04	0.22	-0.85	-0.10	0.13	-0.77
2	**8.58**	**9.21**	**14.04**	**12.58**	**12.63**	**19.31**	**9.65**	**10.24**	**15.27**	**7.15**	**9.15**	**10.80**
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	1.65	1.19	**2.32**	**2.19**	1.89	**3.51**	1.84	1.37	**3.23**	1.34	0.78	**2.83**
5	-0.20	0.12	-7.28	0.42	0.73	-5.82	-0.23	-0.04	-5.94	-0.54	-0.35	-5.49
6	0.78	1.20	0.34	0.92	1.46	0.93	1.11	1.74	1.48	0.84	1.52	1.28
7	**9.14**	**4.41**	**19.81**	**7.77**	**4.37**	**18.98**	**5.28**	**2.35**	**14.47**	**3.72**	1.72	**10.47**
8	**8.52**	**13.31**	**20.19**	**9.23**	**14.28**	**23.97**	**8.63**	**13.74**	**22.22**	**4.50**	**9.80**	**16.50**
9	**4.08**	**4.13**	**11.92**	**5.60**	**6.54**	**18.71**	**5.45**	**9.57**	**25.07**	**3.57**	**10.50**	**27.32**
10	-0.37	-1.57	-4.97	-0.59	-1.89	-5.00	0.17	-1.45	-4.72	-0.17	-2.48	-6.19
11	**2.84**	**3.00**	-3.47	**2.45**	**2.64**	-2.96	1.94	**3.17**	-2.15	1.69	**2.75**	-2.43
12	1.15	1.39	1.81	0.78	1.20	1.96	0.96	1.47	**2.50**	1.06	1.90	**3.16**
13	1.44	1.65	1.87	1.68	1.61	**2.25**	1.66	1.73	**2.44**	1.44	1.95	**2.56**
14	-0.08	0.58	-0.48	-0.19	0.60	-0.95	-0.41	0.55	-1.41	-0.39	0.69	-2.24
15	**4.70**	**5.06**	**7.05**	**7.04**	**7.63**	**11.04**	**7.78**	**8.60**	**12.74**	**6.85**	**7.62**	**11.00**
16	0.04	0.03	0.05	0.12	0.07	0.17	0.07	0.02	0.12	0.08	0.06	0.13
17	0.15	0.15	-0.38	0.30	0.45	-0.16	0.20	0.39	-0.34	0.05	0.26	-0.59
18	**3.61**	**4.34**	**4.02**	**2.23**	**2.97**	1.85	1.54	**2.12**	0.54	0.69	0.98	-0.70
19	0.61	0.95	1.00	0.66	0.86	1.05	0.73	1.01	1.22	0.67	1.05	1.17
20	0.12	0.27	0.40	0.47	0.44	0.57	0.44	0.27	0.68	0.39	0.13	0.40
21	**6.14**	**8.33**	**6.98**	**3.97**	**5.09**	**4.51**	**2.63**	**3.05**	**2.51**	1.55	1.98	1.01
22	0.56	0.88	0.25	0.12	0.40	-0.16	0.19	0.27	-0.18	0.31	0.22	-0.13
23	0.20	0.24	0.38	0.34	0.27	0.53	0.48	0.52	0.85	0.43	0.70	1.03
24	1.27	**4.27**	**3.74**	0.00	1.95	1.34	-0.31	1.07	0.32	-0.07	0.58	-0.28
25	**9.67**	**8.91**	**15.36**	**10.68**	**7.79**	**15.87**	**8.76**	**5.41**	**13.29**	**8.45**	**4.29**	**11.91**
Total	64.86	72.37	94.48	69.05	74.39	111.02	58.62	67.41	103.35	43.52	55.95	82.75

If precise diagnostics were possible, one might expect that the absolute contributions to excess Danish mortality from the specific causes of deaths would be even higher since more deaths would be allocated to the specific disease categories. However, the relative contribution of a particular cause of death to excess Danish mortality in this case might thereby change.

In sum, the results presented in Table 4.4 should be considered suggestive but not conclusive. The problem we are facing here is rooted in the quality of data on cause-specific mortality. In addition to the large group of residual diseases, the results related to the diseases of the circulatory and the respiratory systems (Appendix Table 4.1, chapters III and IV) should be viewed with greater caution than others (Juel and Sjol, 1995; Bjerregaard and Juel, 1993).

Yet another source of errors is the different classifications of diseases used by countries submitting data to the WHO. This sometimes makes it difficult to restore time trends of specific causes of deaths. Denmark used the 8th revision of the International Classification of Disease (ICD)

from 1969 to 1993 while in the Netherlands and Japan this classification was used only up until 1979 and in Sweden until 1987. After these dates death counts were reported in these countries using the 9th revision of the ICD. In contrast, Denmark never made of use the 9th revision of the ICD but has used the 10th revision since 1994. An example of problems associated with the transition from the 8th to the 9th revision: in Japan this resulted in an abrupt jump in death rates from the 10th cause of death (other forms of heart disease); a similar jump is also noticeable in the Netherlands but not in Sweden.

To minimize the effect of problems associated with misclassification and to improve the overall quality of results, I aggregated the causes of deaths by disease categories included in the chapters of Appendix Table 4.1. By aggregating the data it is possible to obtain more reliable results, but the structure of those causes of death that provide contributions to excess Danish mortality will be less detailed. I repeated the procedure described above using these broader categories of diseases. In addition, I computed the relative contribution of a particular cause of death and included it in Table 4.5. The highlighted items in Table 4.5 are causes of death that provide the highest contribution to the excess Danish mortality.

As is evident from Table 4.5, there are striking similarities among the results for female populations. In all countries and at all ages excess mortality from cancer (group II) has made the highest contribution to the observed mortality differences. The numbers range from 40 to 50% of total excess mortality. Cardiovascular diseases (III) and respiratory diseases (IV) take second and third place, respectively, in order of importance. However, at ages 50–54 the most important contribution (after cancer) comes from mortality from accidents, poisonings, and violence (VI), leaving the cardiovascular and respiratory diseases behind.

The results obtained for male populations are less homogeneous between the countries. However, in case of Sweden the pattern is similar to that observed in the female populations, i.e., the main contribution is attributed to cancer mortality followed by cardiovascular mortality and mortality from respiratory diseases. Generally, about 45% of excess mortality is related to cancer. We also note that the importance of respiratory diseases rises significantly with age and that this is the age group 65–69 where mortality differences are more pronounced. In addition, at ages 50–54 the group of digestive diseases, including cirrhosis of the liver, constitutes a significant part of excess mortality.

As follows from Table 4.5, the male population of Japan has a striking advantage of lower mortality from cardiovascular diseases. The differences in death rates from cancer also provide a positive contribution to excess Danish mortality but this is less marked. Approximately 60% of

Table 4.5 Decomposition of excess Danish mortality by aggregated causes of death for the period 1985–1993. Males.

The table shows both the absolute and relative contribution (%) of a particular group of diseases to the total excess mortality. A description of causes of death included in a particular group of diseases is provided in Appendix Table 4.1. The items in boldface are causes of death providing maximal contributions to excess mortality.

Ages	50–54			55–59			60–64			65–69		
Chapter	Sw	Nl	Jp	Sw	Nl	Jp	Sw	Nl	Jp	Sw	Nl	Jp
Absolute contribution to excess mortality(%)												
I	1.12	1.25	-0.08	0.47	0.69	-0.70	0.11	0.11	-1.15	-0.12	-0.05	-1.50
II	**13.97**	4.73	3.70	**15.68**	5.48	4.19	**16.22**	4.77	6.72	**14.31**	3.47	10.32
III	6.75	9.44	**16.28**	5.52	**11.29**	**25.17**	5.11	**11.79**	**33.47**	2.78	**10.88**	**36.69**
IV	0.44	2.00	0.48	2.08	2.99	1.80	3.97	3.21	2.79	4.88	2.72	2.51
V	4.00	6.37	-0.82	2.81	4.77	-0.96	1.88	2.72	-0.54	1.31	1.53	-0.07
VI	1.19	**11.42**	1.22	-0.24	5.78	-0.48	0.34	3.80	0.29	0.37	2.11	-0.22
VII	9.88	8.95	14.56	9.34	6.85	12.67	7.97	4.80	10.61	6.44	2.96	8.86
Total	37.35	44.15	35.34	35.66	37.85	41.69	35.60	31.19	52.19	29.98	23.63	56.59
Relative contribution to excess mortality (%)												
I	2.99	2.83	-0.23	1.33	1.83	-1.68	0.31	0.34	-2.21	-0.40	-0.20	-2.65
II	**37.40**	10.71	10.48	**43.98**	14.48	10.04	**45.55**	15.28	12.88	**47.75**	14.70	18.23
III	18.08	21.38	**46.07**	15.47	**29.82**	**60.37**	14.34	**37.79**	**64.14**	9.28	**46.06**	**64.84**
IV	1.17	4.53	1.35	5.83	7.91	4.33	11.14	10.30	5.35	16.27	11.53	4.44
V	10.71	14.43	-2.33	7.87	12.59	-2.30	5.29	8.71	-1.04	4.36	6.46	-0.13
VI	3.18	**25.86**	3.46	-0.67	15.27	-1.15	0.97	12.18	0.56	1.25	8.93	-0.38
VII	26.46	20.26	41.19	26.19	18.09	30.38	22.40	15.40	20.33	21.49	12.52	15.66
Total	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table 4.5 (cont.) Females.

Ages	50–54			55–59			60–64			65–69		
Chapter	Sw	Nl	Jp	Sw	Nl	Jp	Sw	Nl	Jp	Sw	Nl	Jp
Absolute contribution to excess mortality(%)												
I	0.26	0.30	-0.47	0.28	0.43	-0.49	0.04	0.22	-0.85	-0.10	0.13	-0.77
II	**28.47**	**29.44**	**49.42**	**33.11**	**35.35**	**60.88**	**26.28**	**29.41**	**50.73**	**17.00**	**22.63**	**36.38**
III	9.14	8.60	7.16	9.92	10.10	14.96	10.19	14.50	23.13	7.59	14.61	24.42
IV	4.80	5.83	6.24	7.26	8.74	10.10	7.64	9.57	11.11	6.59	8.64	8.30
V	4.35	5.57	5.41	3.36	4.27	3.48	2.71	3.40	2.45	1.76	2.17	0.87
VI	8.17	13.72	11.35	4.43	7.71	6.22	2.99	4.92	3.50	2.22	3.49	1.63
VII	9.67	8.91	15.36	10.68	7.79	15.87	8.76	5.41	13.29	8.45	4.29	11.91
Total	64.86	72.37	94.48	69.05	74.39	111.02	58.62	67.41	103.35	43.52	55.95	82.75
Relative contribution to excess mortality(%)												
I	0.40	0.41	-0.50	0.41	0.58	-0.44	0.07	0.32	-0.82	-0.23	0.24	-0.94
II	**43.90**	**40.68**	**52.31**	**47.95**	**47.53**	**54.84**	**44.84**	**43.62**	**49.08**	**39.06**	**40.44**	**43.97**
III	14.10	11.88	7.58	14.36	13.57	13.47	17.39	21.50	22.38	17.44	26.12	29.51
IV	7.40	8.06	6.61	10.52	11.75	9.10	13.03	14.19	10.75	15.15	15.44	10.03
V	6.70	7.70	5.73	4.87	5.74	3.13	4.63	5.05	2.37	4.05	3.87	1.05
VI	12.60	18.96	12.01	6.42	10.37	5.60	5.10	7.29	3.39	5.11	6.23	1.97
VII	14.90	12.31	16.26	15.47	10.47	14.30	14.94	8.02	12.86	19.42	7.66	14.40
Total	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

excess mortality (Denmark vs. Japan) can be attributed to cardiovascular diseases, 15% to malignant neoplasms and about 4% to respiratory diseases. The Japanese levels of mortality from other causes of death are comparable with the Danish levels.

Compared to the male population of the Netherlands, the most notable mortality differences are to be observed in cardiovascular mortality (III). Death rates from this cause of death account for 30% of the total mortality differences at ages 55–59 and about 45% at ages 64–69. However, at ages 50–54 the highest contribution comes not from the cardiovascular diseases but from accidents (VI), which accounted for 25% of total excess mortality. The rest of excess Danish mortality is equally divided among all other categories apart from infectious diseases - and cancer, which is somewhat more important, especially at older ages.[KA6]

4.3.3 Time trends in cause-specific mortality

The analysis presented in the previous section helped us to highlight the most important causes of death contributing to excess Danish mortality in middle age. Another question of principal interest involves trends in death rates from specific causes of death. I have computed the series of death rates for the period from 1970 to 1993 by 4 age groups for all countries included in the analysis. The year 1970 was chosen because it marks the emergence of the area of excess mortality, as can be seen in Fig. 4.6. In addition, all countries used the 8th revision of the ICD at that time, which permits us to avoid certain classification problems in the earlier years. The year 1993 is the latest year for which Danish data were available at the time this chapter was written. Altogether, 200 plots[34] have been analyzed and 44 that show the disadvantageous trends in Danish mortality are presented in Fig. 4.7. It must be emphasized that the causes of deaths discussed below have been selected in order to shed light on Danish excess mortality. In other words, only causes of death where Danish mortality is higher are discussed here. The reader interested in the trends of all causes of death can explore the graphs provided on the CD-ROM for himself. Another approach for the analysis of cause-specific mortality trends can be found in Andreev et al. (1997).

Male mortality from lung cancer (2) has been steadily increasing in Denmark, Sweden and Japan but not in the Netherlands, where a moderate decline in mortality can be observed (Fig.

[34] The plots with the trends in cause-specific mortality are provided on accompanying CD-ROM. The files are stored in HTML format and can be viewed with any Web browser. If your CD-ROM drive is assigned D: letter, open D:\CAUSES\CAUSE.HTM to start browsing.

4.7(a)). During the whole period Danish mortality has been double that of Sweden and Japan but appreciably lower than that of the Netherlands. The decline in Dutch mortality led to the convergence of mortality levels in the Danish and Dutch populations, so there is much less difference between the two countries at the beginning of 1990s than in the 1970s. Moreover, there is a certain drop in the death rates at ages 50–54 both in Denmark and the Netherlands; this can perhaps be attributed to certain cohort effects, but this hypothesis requires additional elaboration.

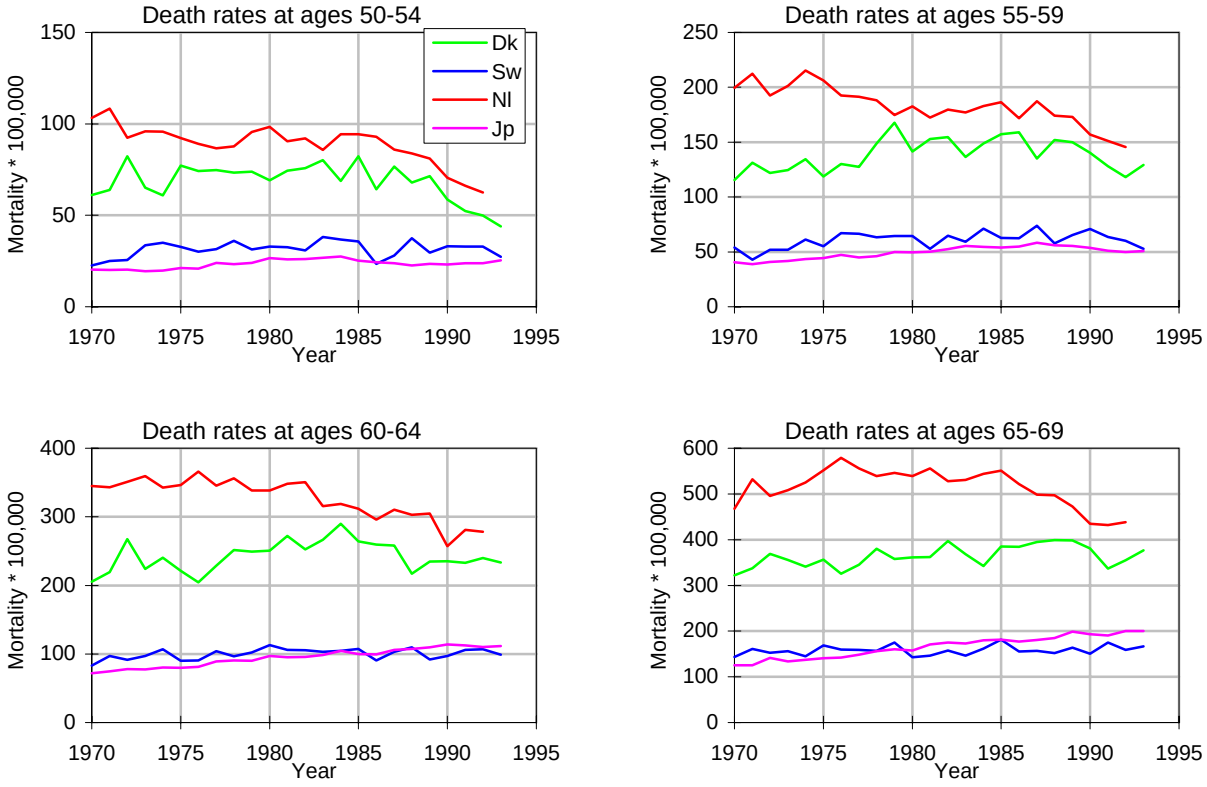
Trends in death rates from ischaemic heart disease (9) followed almost the same trajectory in all European countries. Until 1980 the death rates remained at an approximately constant level, but then a persistent decline can be observed in all populations. The rate of decline was appreciable, and by the year 1993 the level of mortality had dropped to nearly half that of 1980. The level of mortality now differs considerably between countries and age groups. Even though it followed the same pattern of decline, in the lower age groups (50–54) Danish mortality was generally higher than that of Sweden and the Netherlands. In contrast, at higher ages (60–69) Danish and Swedish mortality curves are very close, while Dutch mortality is significantly lower. The exceptionally low level of Japanese mortality makes the position of this country outstanding in comparison. Mortality in Japan has also declined but its level was significantly lower (4- to 6-fold) than in European countries.

Regarding respiratory diseases, we observe that there are no notable trends in Danish mortality from this cause of death. This is true of Sweden as well, although mortality in Denmark was on average 2.5 times higher than in Sweden. If we look at the Netherlands, in the early 1970s mortality in both countries was nearly the same but in the early 1990s Danish death rates were about 50% higher because of reductions in Dutch mortality. The death rates in Japan also show a downward trend, but their level was comparable with the Swedish level in the 1970s, which is appreciably lower than in the Netherlands. Because of this decline, the level of Japanese mortality in recent years has been the lowest of all countries included in the comparison.

In the case of cirrhosis of the liver (18) the trend in Danish mortality is the opposite of that are observed in the other countries. I found a substantial and uniform increase in all age groups in Denmark, while mortality in Sweden dropped sharply in the early 1980s and mortality in Japan remained either constant (50–59) or declining (60–69). Even though Japanese mortality

Figure 4.7(a) Disadvantageous trends in Danish cause-specific mortality. Males.

Malignant neoplasm of the trachea, bronchus and lungs(2), Males



Ischaemic heart disease(9), Males

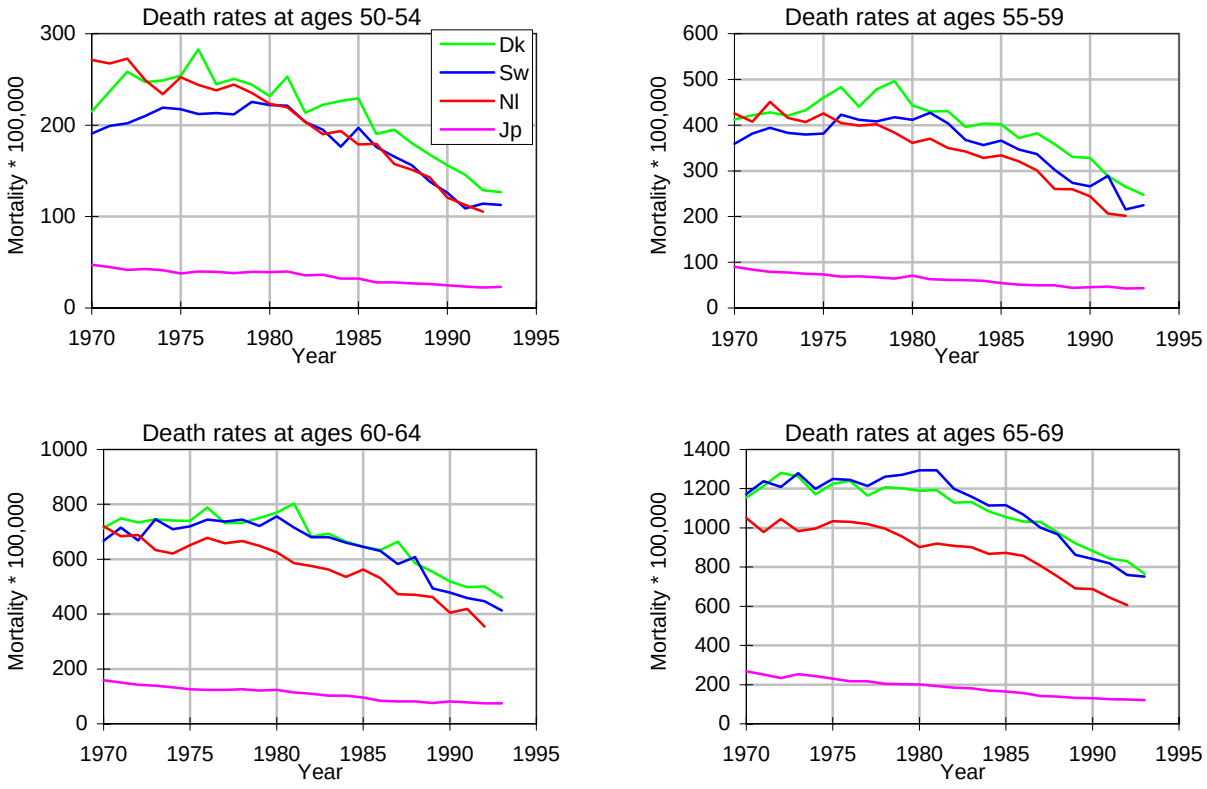
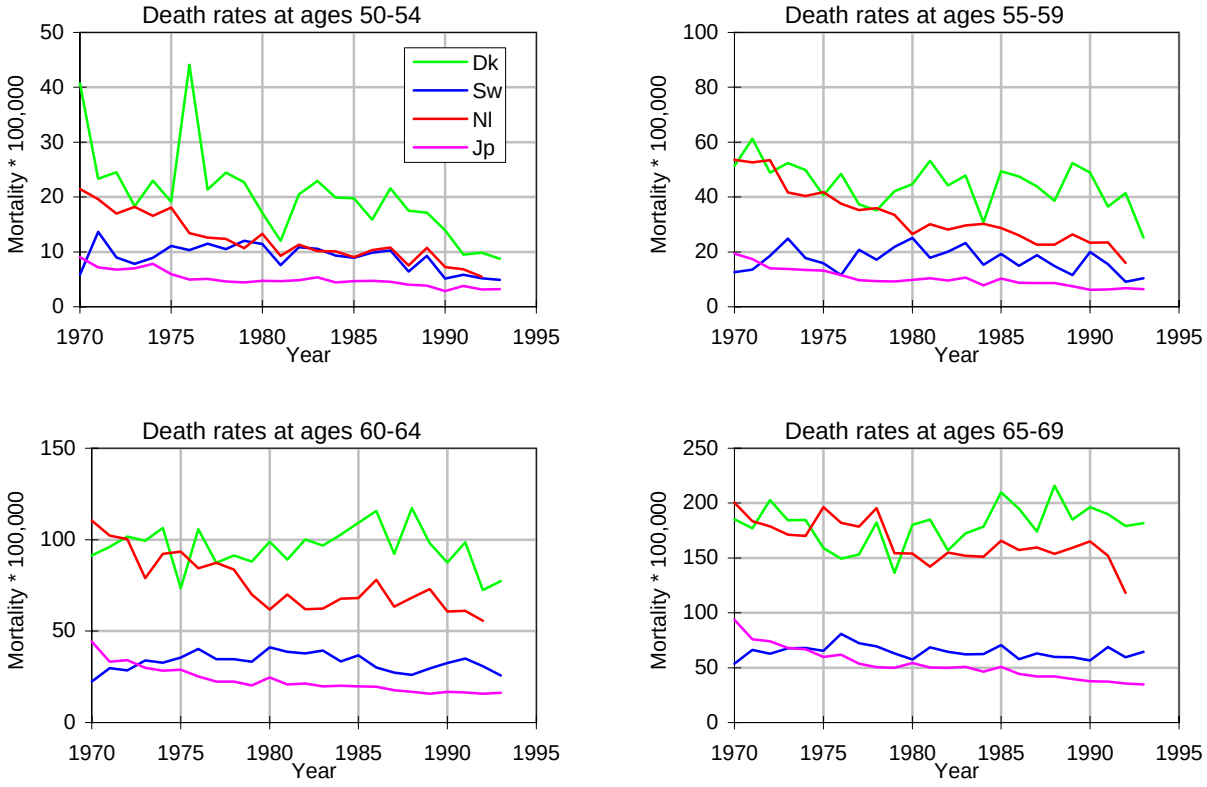


Figure 4.7(a) (cont.)

Bronchitis, emphysema and asthma(15), Males



Cirrhosis of liver(18), Males

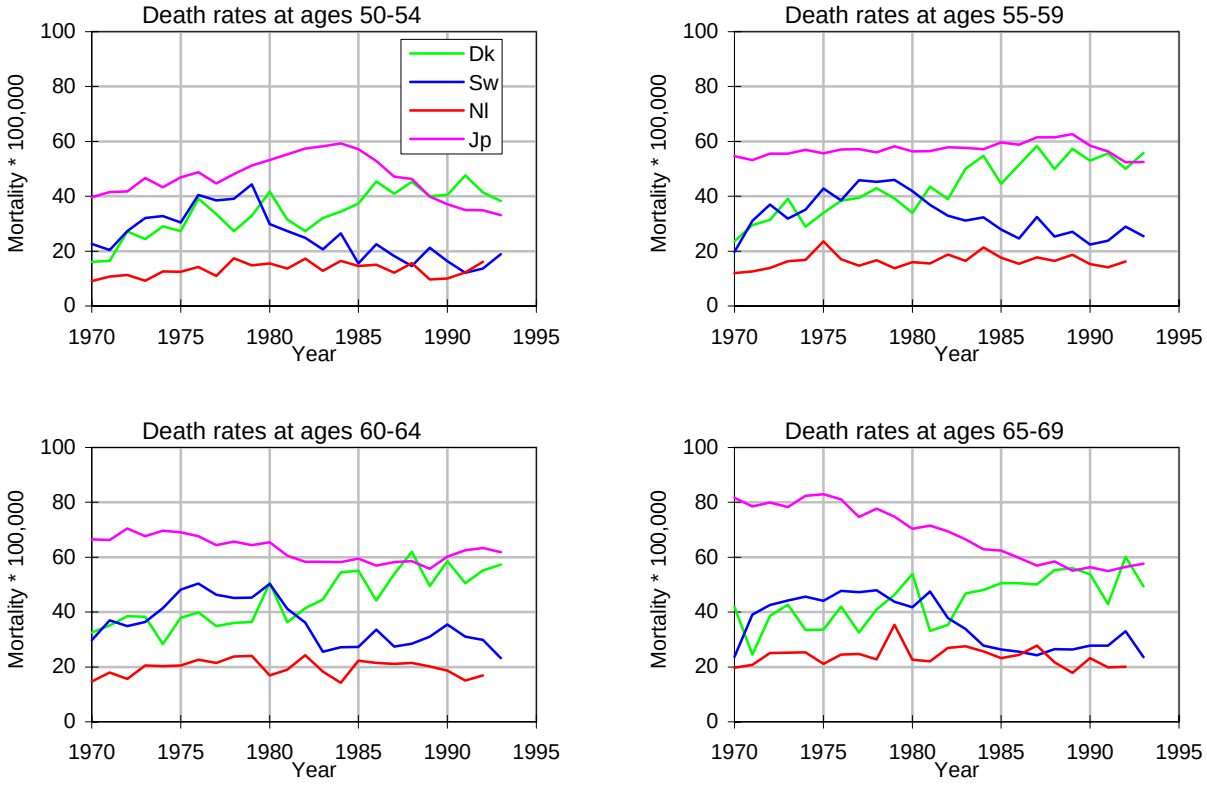
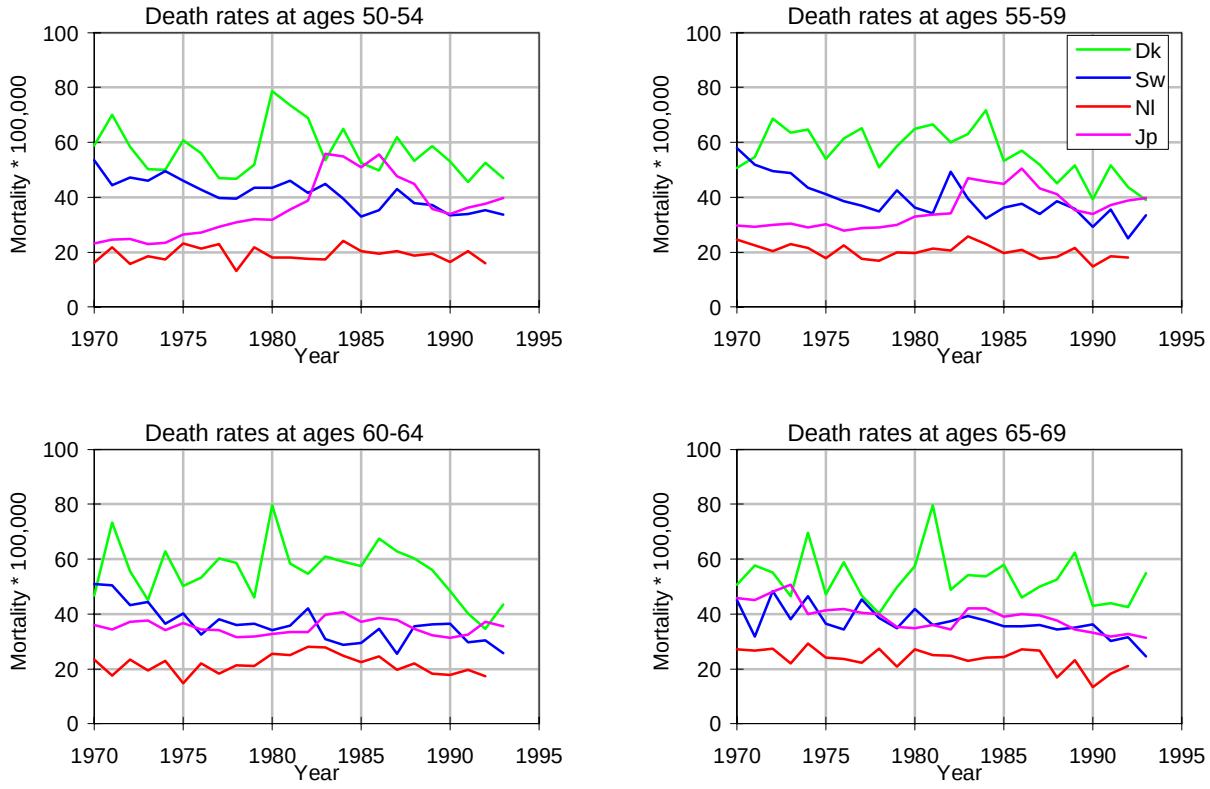


Figure 4.7(a) (cont.)

Suicide and self inflicted injury(21), Males



was remarkably higher than in the Nordic countries in 1970, the level of mortality in Denmark and in Japan was virtually the same in the 1993. Mortality in Sweden at that time was about half of that. In the Netherlands no notable trends in mortality can be observed; it remained constant at low level.

Death rates from suicide (21) have traditionally been higher in Danish males than in other countries. There have been no real improvements here except for some convergence to the levels of Sweden and the Netherlands at ages 55–64 in recent years. It is difficult to judge whether this is the onset of a general trend or some temporary phenomenon, because there is no evidence of a similar decline at ages 50–54 or 65–69.

For the female populations we will discuss the same causes of death as for males, adding only the trends in breast cancer. In the case of lung cancer mortality (2), there is a remarkable gap between the Danish population and other countries. Mortality from lung cancer has been increasing in all countries since 1970 and the rate of increase has been especially large in Europe (8.5% in Denmark and the Netherlands; 6.5% in Sweden) as opposed to Japan (1%). Mortality in Denmark in the early 1970s was appreciably higher than in other countries and this difference has increased in

the 1990s even though the Danish rate of increase was approximately the same as that of Sweden and the Netherlands.

Breast cancer death rates (7) have been gradually increasing in Denmark and the Netherlands since 1970; this is especially noticeable in the higher age groups. The mortality differences from this cause of death are quite small in these countries except for in the last decade, where Danish mortality has been somewhat higher than Dutch mortality at ages 50–59. In contrast, Swedish death rates have been gradually declining since 1970, and so the gap between Swedish and Danish mortality has increased in recent years. Japanese mortality rates seem to be on the rise but the level of Japanese mortality is much lower than in European countries.

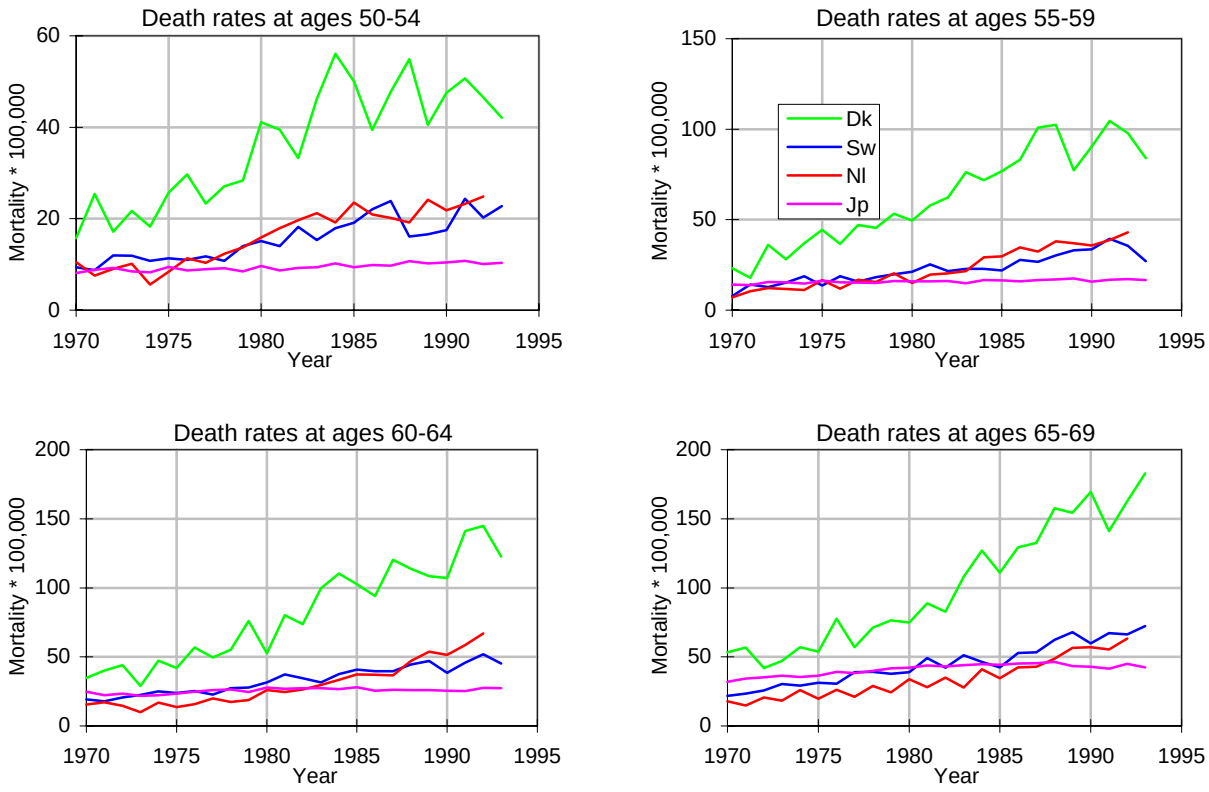
Mortality developments as regards ischaemic heart disease (9) are close to the trends in the male populations, i.e., mortality has been declining in all populations but the level of mortality in Denmark is higher. The main distinction seems lie in the difference in mortality levels in Denmark as compared to other countries. This difference is far more eye-catching for females than for males. This impression could be misleading, however, if we consider the contribution of (9) to the differences in life expectancy since the general level of mortality is significantly higher for males than for females.

Another group of diseases where Danish mortality was considerably higher than in the other countries includes bronchitis, emphysema and asthma (15). Mortality from this cause of death has increased dramatically in Denmark since 1970, when the level of Danish mortality was only slightly higher than in other countries. In contrast, mortality in other countries has either remained constant (60–69) or declined steadily (50–59). This development in mortality rates has led to a considerable excess of Danish mortality in recent years, which is especially remarkable in the age group 60–69. The pattern is very similar to the trends in the lung cancer mortality, where a dramatic increase in Danish mortality can be observed. This finding suggests that there might be some correlation between lung cancer mortality and other diseases of the respiratory system. There may be some common factor contributing to this increase in mortality, such as smoking.

Mortality trends concerning cirrhosis of the liver (18) have been also unfavorable for Denmark, as was the case for males. While mortality in other countries has either declined or remained constant, the Danish rates have increased steadily. Particularly high mortality differences between Denmark and other countries in recent years are to be observed at ages 50–59. Since 1970, Danish death rates at these ages have approximately doubled. In the higher age groups both the increase and the differences between Denmark and other countries are less marked.

Figure 4.7(b) Disadvantageous trends in Danish cause-specific mortality. Females.

Malignant neoplasm of the trachea, bronchus and lungs(2), Females



Malignant neoplasm of breast(7), Females

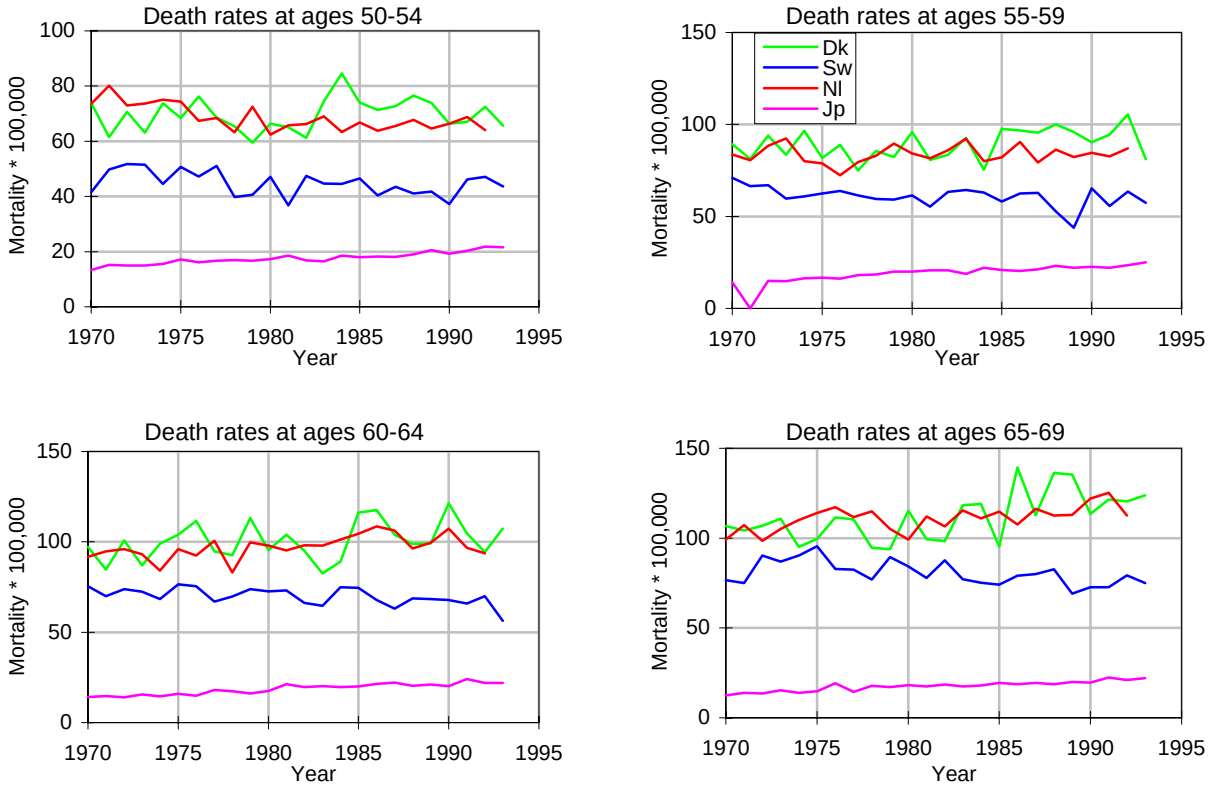
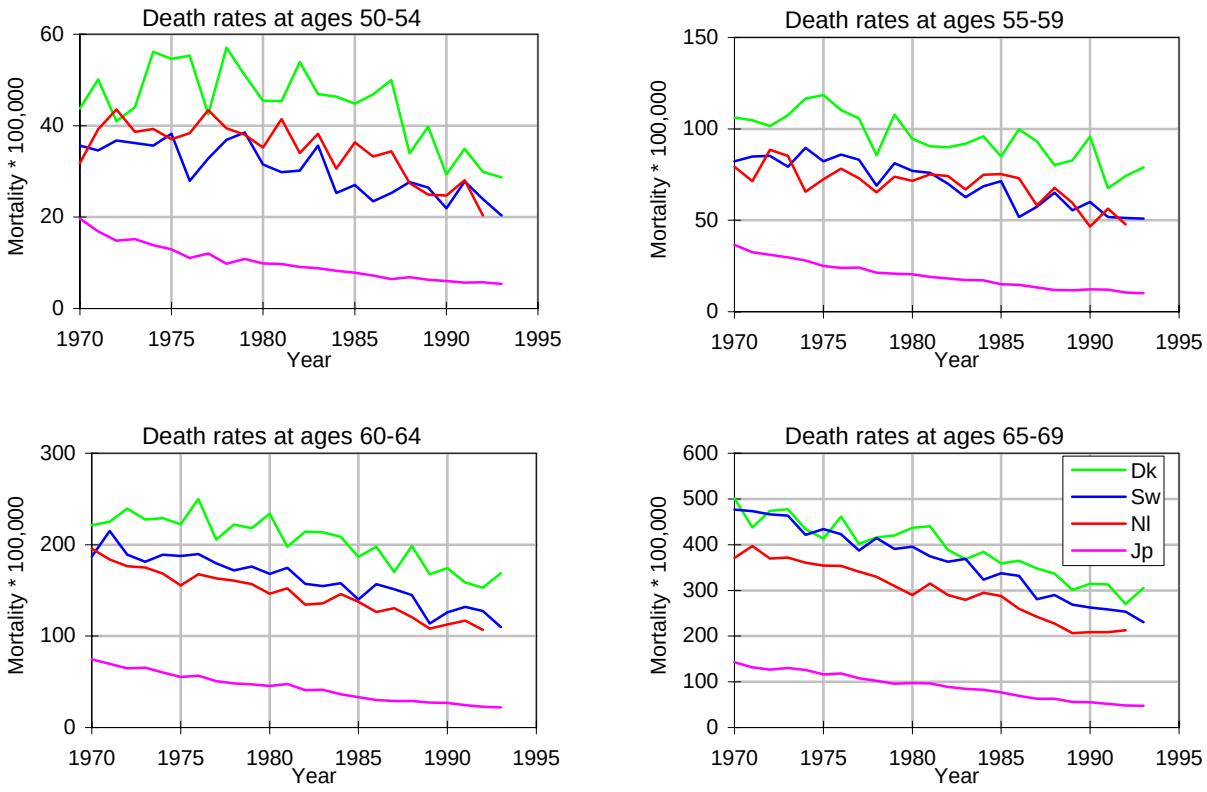


Figure 4.7(b) (cont.)

Ischaemic heart disease(9), Females



Bronchitis, emphysema and asthma(15), Females

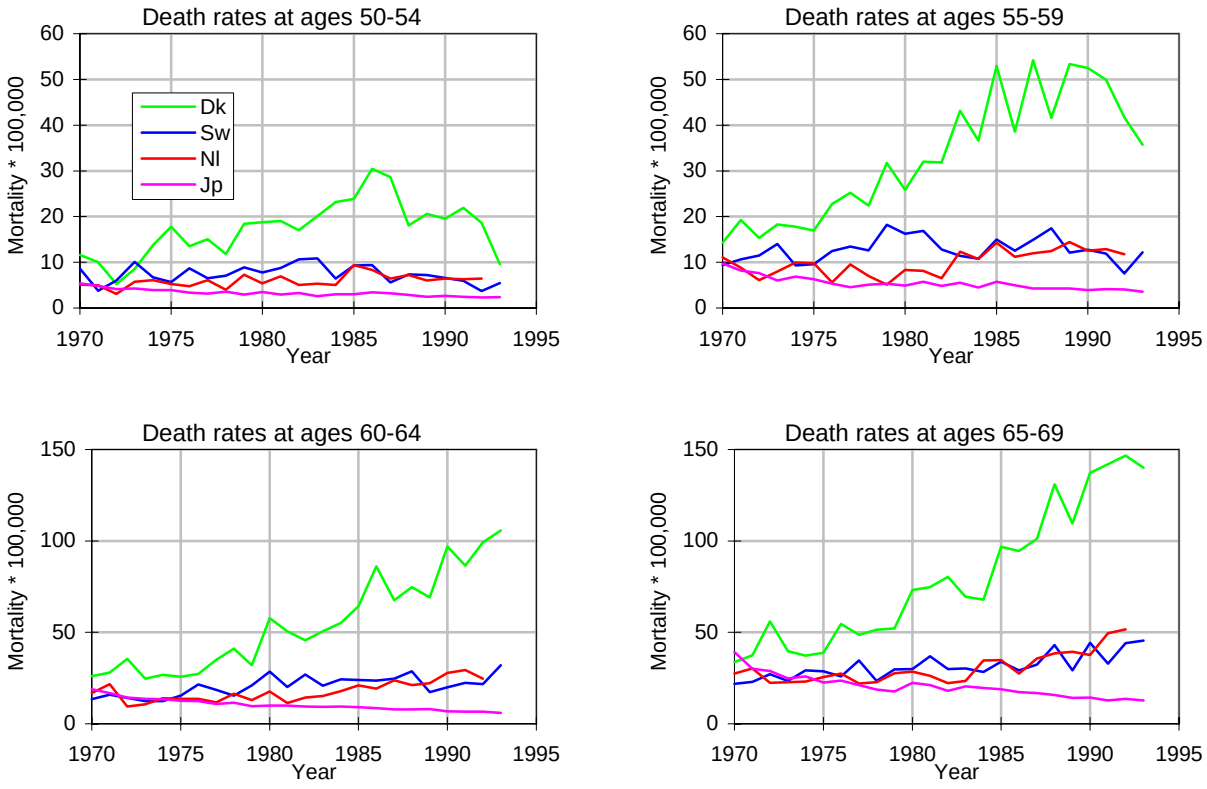
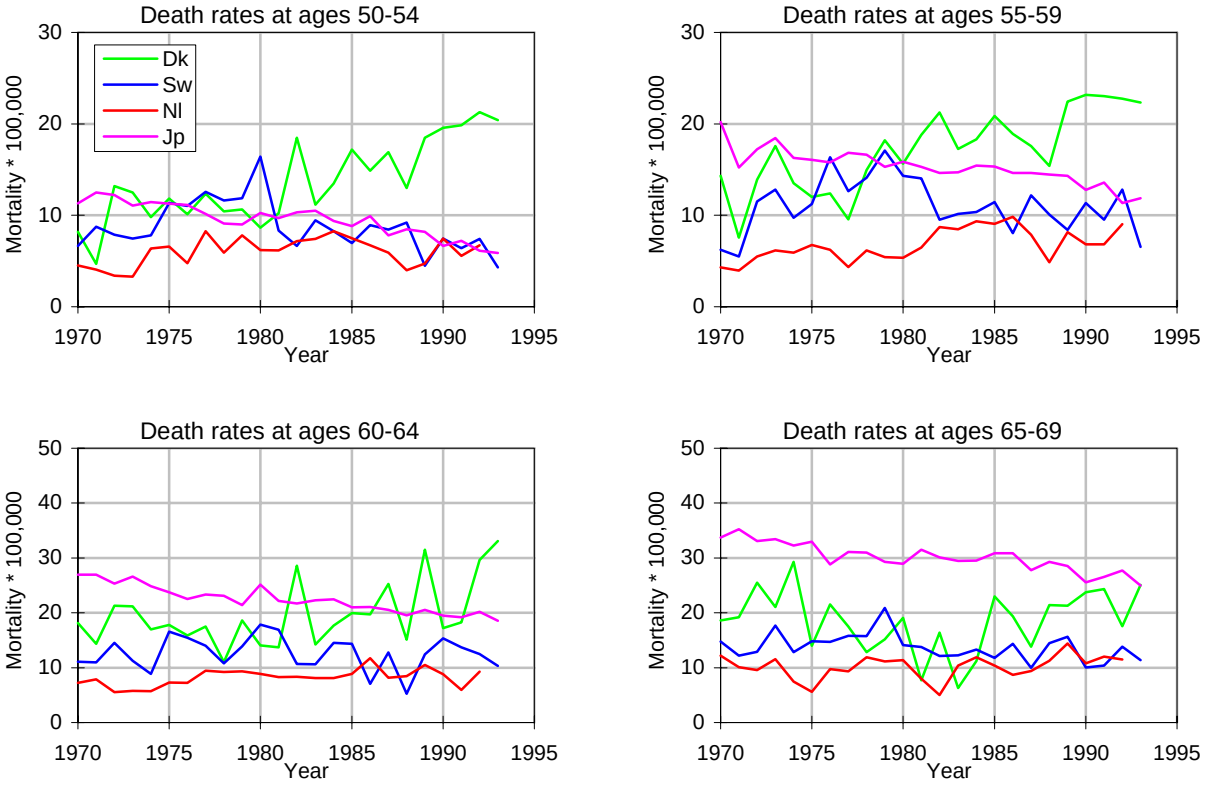
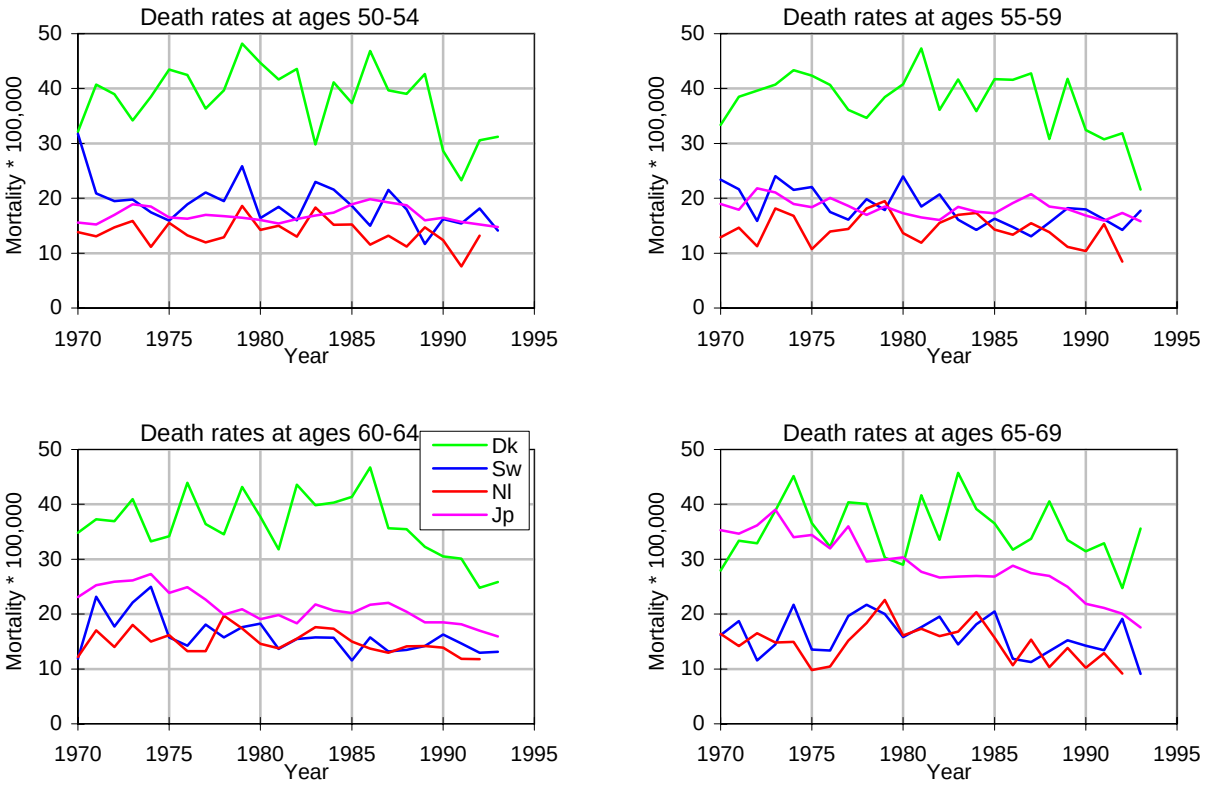


Figure 4.7(b) (cont.)

Cirrhosis of the liver(18), Females



Suicide and self inflicted injury(21), Females



Suicide mortality (21) generally remained at a constant level in all countries from 1970 to 1993. There are only two exceptions. First, mortality at ages 65–69 in Japan declined significantly and reached the level of Sweden and the Netherlands in 1993. Second, there was a notable drop in Danish mortality at ages 60–64 starting in the year 1987. There was a similar decline in lower age groups (50–54 and 55–59), but this was less significant than in the age group 60–64. It is interesting to note that a similar drop in mortality can be observed on the male graph as well, which means that there might be some cohort effect operating in both the male and female populations. On the whole, the pattern of the excess Danish mortality is the same as for males, i.e., mortality levels have not changed very much, while Danish mortality is consistently higher than mortality in other countries. However, the mortality differences are greater in the case of females.

The analysis conducted here permits us to draw some important conclusions. First of all, we must note that the structure of cause-specific mortality in excess Danish mortality is different for the male and female populations. Comparing male mortality rates with those of the Netherlands and Japan, we note that the most significant contribution to excess Danish mortality is added by cardiovascular diseases. Comparing Danish and Swedish rates, on the other hand, we see that cancer is the main factor involved in male mortality differences. In contrast, the results obtained from the analysis of the female populations suggest that the contributions from the different causes of death to the excess of Danish mortality are quite similar for all countries. It is evident from Table 4.5, that the most important contribution to excess female Danish mortality is that of cancer (especially lung and breast cancer).

An examination of trends in cause-specific mortality allowed us to discover those causes of death that contribute to excess Danish mortality. It turns out these causes of death corresponds closely for males and females (except for breast cancer) though their role in explaining the total mortality differences between Denmark and other countries is not the same. The most unfavorable trends involve diseases of the respiratory system: lung cancer, bronchitis, emphysema and asthma. Mortality from breast cancer also exhibits a negative trend, since it has increased in Denmark while it has declined in Sweden. Mortality from cirrhosis of the liver is also a concern, as rising trends have been observed in Denmark alone. This disease is usually linked to the consumption of alcohol, which is significantly higher in Denmark than in Sweden, for example. Finally, I need to mention the importance of mortality differences as regards ischaemic heart disease. Although the trends in Danish mortality have been in concordance with the developments in other countries, I found that the Danes have somehow lagged behind their European counterparts, since the Danish death rate remains consistently on a higher level. Because mortality from this cause of death is considerably

higher than from other diseases, it might provide a main contribution if the differences in life expectancy are analyzed.

4.4 Discussion

It is well known that life expectancy in Denmark has increased significantly since the middle of 19th century. The age-specific mortality changes are less well-known since investigation thereof has been hampered by a lack of data and of convenient visualization tools. Such data are now available and can be obtained from the Danish mortality database located at Odense University. The visualization program to produce demographic contour maps has also been developed (Chapter 5) and included with this PhD thesis. All contour maps presented here were produced with the help of this program.

The first objective of this work was to demonstrate the potential importance of Danish mortality data for demographic research. I focused on the investigation of the Danish mortality surface with special attention given to age-specific mortality changes. Contour maps of Danish mortality and maps of mortality progress allowed us to identify the timing of the demographic transition and the age-specific structure of mortality changes. The results of my analysis suggest that mortality transition in Denmark at the end of the 19th century belonged to the mainstream of transitions in other European countries. In fact, the mortality changes were even more favorable in Denmark than in other countries and Danish life expectancy seems to have been among the highest in Europe around 1910 (Table 4.1). Nonetheless, a comprehensive study of factors behind the Danish mortality transition and the role they played in the observed mortality decline has yet to be carried out.

Until the 1960s mortality declined very rapidly, and life expectancy rose to exceptionally high levels which were unprecedented in Danish history. But mortality progress then decelerated significantly, and the rate of increase in life expectancy fell down to remarkably low levels. Despite stagnation or even a degree increase in mortality in middle ages, life expectancy continued to grow because of the rapid mortality reductions at oldest-old ages and continuing mortality decline in infancy and childhood ages. These mortality developments were unusual when compared with mortality trends observed in other European countries, where gains in life expectancy were appreciably higher. Faced with these developments, the Danish government set up a committee to investigate the slowdown in the increase of expectation of life. The investigation focused mainly on

trends in the standardized mortality rates, social-economic variables and the analysis of differences in life styles. The age-specific differences in Danish survival fell outside the scope of this study.

In order to shed light on age-specific mortality differences, I constructed mortality databases similar to the Danish one for Sweden, the Netherlands and Japan, and estimated the ratio surfaces of Danish mortality to those of other countries. This allowed me to identify the area with excess Danish mortality and to follow the age- and time-specific dynamics of the mortality ratios. The results of this analysis suggest that the area of excess mortality began to form in the late 1960s at the age of 50–60. Over time this spread out to lower and higher age groups, thus making for more striking mortality differences. This pattern of development in the mortality ratios prevailed until the latest years for which data are available, and so far there no favorable tendencies to be observed.

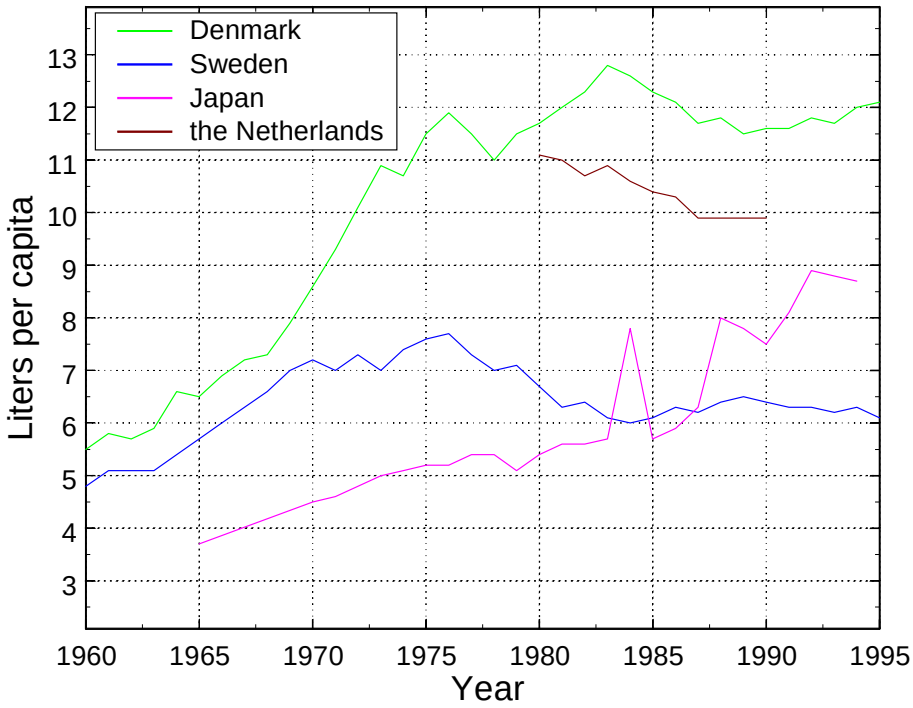
Finally, I analyzed cause-specific mortality in order to explore the relative contribution of the different causes of death to the excess of Danish mortality. I decompose the total excess mortality for the years starting 1985 and for the age groups where the highest mortality differences has been observed (50–69). This analysis does not overlap with or repeat other studies. It provides useful insight into the cause-specific structure of the excess Danish mortality. I have found that the main contribution to excess mortality varies between causes of death in the male populations, while for females the pattern is remarkably similar. Cardiovascular diseases were the most important cause of excess Danish mortality compared with the male populations of the Netherlands and Japan, while in Sweden the most significant differences were observed in cancer mortality. In the case of females the results point without doubt to cancer mortality (especially lung and breast cancer) as the main contributor to excess Danish mortality.

Further research should involve the biostatistical analysis of survival data on risk factors, i.e., smoking, alcohol consumption, etc. It would also be helpful to incorporate social-economic and life-style variables, e.g., GDP, unemployment rates, fat consumption. But there are two reasons why it will not be easy to accomplish such an analysis. The first is that our knowledge about the relationship between mortality and risk factors is not precise and that no analytical model has been developed so far that specifies the influence of risk factors on mortality and takes into account interdependencies between variables and the lagged effects.

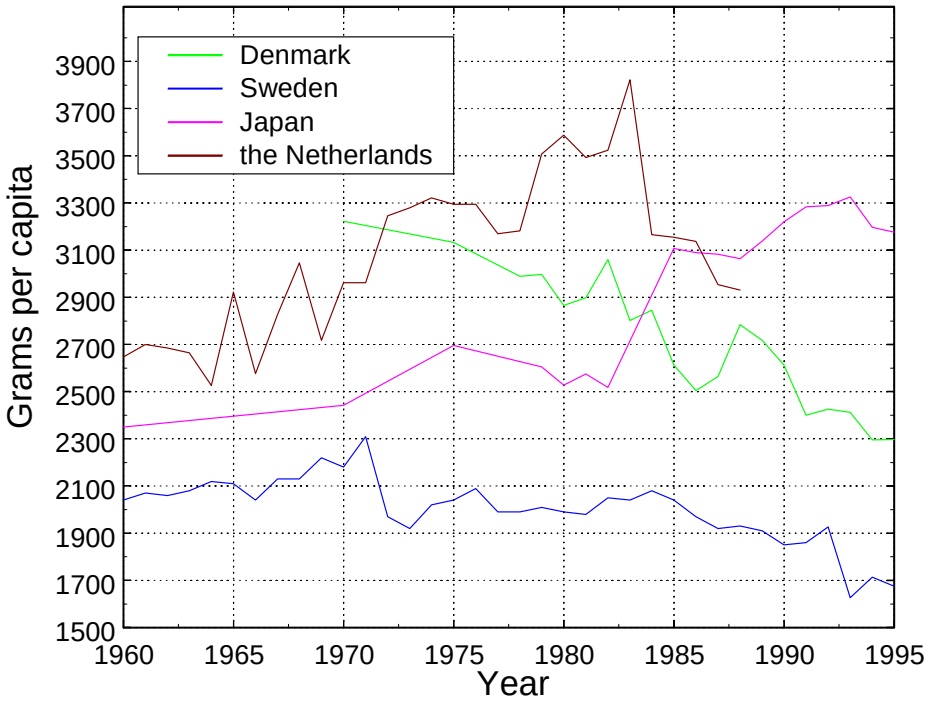
The second reason is the lack of adequate data; the data on social-economic variables are usually only available in relation to the total population. In other words, the age distribution is unknown. As shown above, excess Danish mortality has not been uniform over age, and the highest mortality differences are to be observed at ages 50–70. It could be the case that the effect of a social-economic variable depends on age. It could be harmful in one age group and beneficial in another.

Figure 4.8 Trends in alcohol and tobacco consumption in Denmark, Sweden, the Netherlands and Japan.

(a) Annual consumption of alcohol (population aged 15 and over)



(b) Annual consumption of tobacco (population age 15 and over)



In this case the age structure of this variable must be known in order to account correctly for the effect of this factor.

In addition, a time series should start well back before 1970 when excess Danish mortality first became evident. The effect of a risk factor on mortality could be lagged rather than instantaneous, so the reason for currently observed excess mortality can lie far back in the past. This follows from a pilot study which was done to survey the trends in alcohol and tobacco consumption. The data were taken from the OECD Health database[33] and the trends are shown in Fig. 4.8.

It has been found that there was a sharp increase in the annual consumption of alcohol in Denmark in the period from 1960 to 1975; the number of liters per capita rose from 5.5 in 1960 to 12 in 1975. Consumption has remained at this high level ever since. Alcohol consumption in Sweden, on the other hand, rose from about 5 liters in 1960 to 7.5 liters in 1975, only to drop to the level of 6 liters per capita sometime later. The timing of observed differences in the annual consumption of alcohol corresponds well to the timing of the emergence of excess Danish mortality.

It might be the case that the high levels of alcohol consumption have an immediate effect on the health and mortality of a population. This hypothesis can be tested with data from other countries in which governmental interventions or anti-alcohol campaigns have taken place to reduce the level of alcohol consumption. In Russia, for example, the sale of alcohol was restricted in 1985–1986 and alcohol production was significantly reduced. This resulted in an immediate increase in life expectancy to the highest levels in recent years. With the end of the anti-alcohol campaign, the level of consumption increased again and life expectancy fell.

Tobacco consumption in Denmark have been declining since 1970. In contrast, it increased in Japan, so that in 1983 consumption was at the same level in the two countries. Since then tobacco consumption has been higher in Japan than in Denmark (in 1995 the levels of consumption were 3200 and 2300 grams per capita, respectively). Nonetheless, the gap between Danish and Japanese mortality has continued to grow since 1983. This suggests that the level of tobacco consumption has a lagged effect on survival and can be observed only at some later point in time.

Tobacco consumption in Denmark received a great amount of attention in a recent report from the Danish Ministry of Health (Sundhedsministeriets Middellevetidsudvalg, 1998). According to this report tobacco-related mortality is responsible for a considerable part of the negative development in Danish life expectancy. If tobacco-related mortality were eliminated, life expectancy would rise by three years in Denmark. This supports the finding in this chapter. However, the relative importance of this factor is perhaps exaggerated in this report as compared with other factors, e.g., the consumption of alcohol.

For further research it would be better to concentrate on the investigation of mortality differences between Denmark and Sweden rather than to attempt to include all countries. Since there are well-known similarities between these countries, we can exclude a large number of factors that otherwise might be hypothesized to account for mortality differences. The investigation of differences in the health care system and in preventive intervention should also prove valuable for determining factors behind the higher mortality in Denmark. It seems that even minor differences can have a profound effect on mortality. Such a study should prove to be important not only for Danish society; it should also provide a significant contribution to general mortality research.

CHAPTER 5

Overview of the program Lexis 1.1

5.1 Introduction

Lexis is a graphic program designed to help you create publication-quality contour maps with ease. This software mainly addresses the need for visualization tools in demographic research arising from the analysis of demographic events on the Lexis diagram. However, it is not limited to this area of applications and can be used as a general tool for the visualization of large matrices.

The modern demographer operates with extensive arrays of population statistics collected by the official statistical offices, research institutions and organizations (e.g. Heuser, 1984; Kannisto, 1994; Mamelund and Borgan, 1996; Natale and Bernassola, 1973; Vallin, 1973; Veys, 1983). In most cases, demographic characteristics – e.g. population levels, fertility, morbidity, marriage, divorce or mortality rates – can be plotted in an intelligible and revealing manner because they are usually structured by age, time and cohort. The estimate of the Danish mortality surface (Fig. 4.1), for example, comprises 32,200 death rates which can be portrayed with a single Lexis map from which one can get a general overview of the evolution of Danish mortality. However, this graphic approach has been hampered by the lack of appropriate demographic software. Such a program should be powerful enough to handle demographic problems and yet equipped with a user-friendly interface so as to make it easy to use even for non-experienced computer users. Lexis addresses both issues, which makes it a practical demographic tool.

This program is named after the German demographer Wilhelm Lexis, who in 1875 suggested describing the life course of individuals with the Lexis diagram (Fig. 3.1). The interpretation of this diagram depends on the particular problem one is dealing with. Suppose we have a closed population observed over age and time (e.g. Arthur and Vaupel, 1984). In this case the life course of every individual born in some year z follows the diagonal line called “cohort”. Line BC is interpreted as the number of individuals who survived until the age x in the year y. If x is zero it is the number of births in the year y. Some of the individuals will not survive to the next year y+1 (line CD), and triangle BCD is interpreted as the number of deaths in the cohort z at age x and in the year y. Line CD is the number of individuals who survived until 1 January of year y+1. These

individuals are of the same age in the range from x to x+1. Likewise, the triangle CDG is interpreted as the number of deaths from the same cohort and the same age but in the next year, y+1.

The principal difference between Lexis and other graphic programs is that Lexis permits the plotting of contour maps based on all the principal sets of the Lexis diagram (Hoem, 1976; Keiding, 1990). For example, the age-specific probabilities of dying (cf. e.g. Chiang, 1984) are usually calculated using the data from two adjacent years. In the Lexis diagram these quantities are depicted by the parallelogram ▱. Another frequently-used measure of mortality is the age-specific death rate (cf. e.g. Chiang, 1984), which is the ratio of deaths that occurred in a certain year and age to the total time lived by the population at risk. In this case a single death rate pertains to another principal set, i.e. a Lexis rectangle □.

There are other alternative Lexis sets that also frequently arise in demographic analysis. Vaupel et al. (1998), for example, used Lexis triangles to portray the development of oldest-old mortality in Sweden. In this case the estimates pertain to two Lexis triangles (BCD and ABD in Fig. 3.1) – and the surface of mortality consists of such elements.

The interested reader can find more information on applications of Lexis for demographic research in the monograph by Vaupel et al. (1998), which includes a rich assortment of Lexis maps and an extensive discussion of demographic surfaces[35].

5.2 Program design

5.2.1 Contour map construction

Given a three-dimensional surface z = f(x,y), we can assign an integer to any value of z by means of some scale. If, for example, the scale is a 3x1 vector { -1, 0, 1 }, then the following numbers are assigned to the z values:

- if $z \leq -1$ $\rightarrow$ 0
- if $-1 < z \leq 0$ $\rightarrow$ 1
- if $0 < z \leq 1$ $\rightarrow$ 2
- if $z > 1$ $\rightarrow$ 3

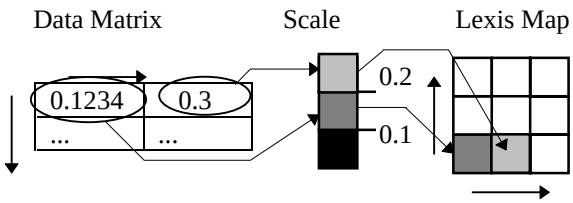
Subsequently, we can assign a color to each integer value, which will be used to paint all elements of z falling between two scale levels. In this example, the resulting contour map will have 4 color

[35] Available online at the MPIDR website: http://www.demogr.mpg.de/Books/PopData/PopData1.htm.

areas because the number of scale values is 3 (=4-1). The same method can be applied to matrices. In this case each element of the matrix will get its own color depending on a scale.

This simple procedure explains the principle of how Lexis works. Fig. 5.1 illustrates the process of translation of the matrix of numeric values (Data Matrix) into the Lexis map. In this example the first element of the matrix (m[1,1]=0.1234) falls between levels 0.1 and 0.2 and it is assigned the color gray as indicated by the scale legend in Fig. 5.1. Finally, the matrix element is painted as a gray rectangle. The second element of the matrix (m[1,2]=0.3) falls above the highest level of the scale and it is painted as a light gray rectangle. The arrows in Fig. 5.1 show the relation between the matrix indices and the orientation of the Lexis map.

Figure 5.1. Translation of data matrix to Lexis map element.



Optionally, the Data Matrix can include a number of missing elements (NaN)[36]. The missing elements are usually used when a particular element of a matrix cannot be computed. For example, if the matrix of death rates is calculated, the denominator (total lived time) can eventually be zero at older ages. In this case it is convenient to set this death rate to a missing value. The missing values are painted with a special color (white by default).

There are 7 principal sets on the Lexis diagram. All of them are supported by Lexis. Table 5.1 shows the correspondence between the Map Type and the Lexis element. If the Map Type is 'Triangle' or 'Left Slope Triangle' it must have double the number of columns in the Data Matrix of other map types. The matrix values for the two Lexis triangles pertaining to the same year and age are retrieved from the two adjacent columns.

If the Map Type is a 'Horizontal Parallelogram' or 'Left Slope Horizontal Parallelogram', the Lexis element extends over two units on the x-axis (e.g. years) and if the map type is 'Vertical

[36] The IEEE arithmetic representation for Not-a-Number (NaN). These result from operations which have undefined numerical results.

Parallelogram' or 'Left Slope Vertical Parallelogram', it extends over two units on the y-axis. In all other cases the Lexis element extends over one unit on both the x- and the y-axis.

Table 5.1 Lexis map types.

Lexis Map Type	Data Matrix		Lexis Map Element
Rectangle	0.1234		
Triangle	0.1234	0.1234	
Left Slope Triangle	0.1234	0.1234	
Horizontal Parallelogram	0.1234		
Left Slope Horizontal Parallelogram	0.1234		
Vertical Parallelogram	0.1234		
Left Slope Vertical Parallelogram	0.1234		

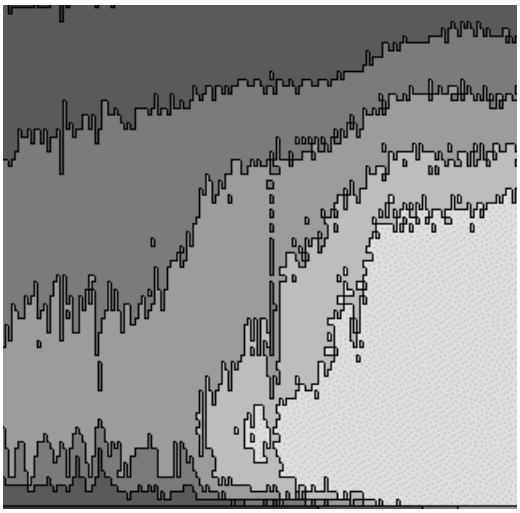
5.2.2 Graphic design

The graphic image visible to the user is a representation of the underlying container of graphic objects which are linked to the data and internal structures (e.g. the data matrix, scale, color tables etc.). The following objects can be present in the graphic container:

- Lexis Map Object
- Plot Frame Object
- Scale Object
- Text Objects
- Rectangle Objects
- Line Objects

The Lexis Map Object (Fig. 5.2) governs the painting of the Lexis map image. The painting algorithm depends on the Lexis Map Type (Table 5.1). In addition, contour lines can be added to the plot and their color and width can be customized. The Lexis map image is always fitted to the dimensions of the plot client area specified in the Plot Frame Object. Lexis does not use any smoothing or interpolation techniques, so the contour map reflects the data you actually have.

Figure 5.2 The example of a Lexis map object.



The Plot Frame (Fig. 5.3) is a graphic object surrounding the plot client area. It includes titles, labels, axis ticks and the grid lines. The properties of all objects belonging to the Plot Frame can be customized. This object also specifies the plot coordinate system and its relation to the physical page. By changing the relation of the object to physical page the printout can be made smaller or larger.

Figure 5.3 The example of Plot frame object.

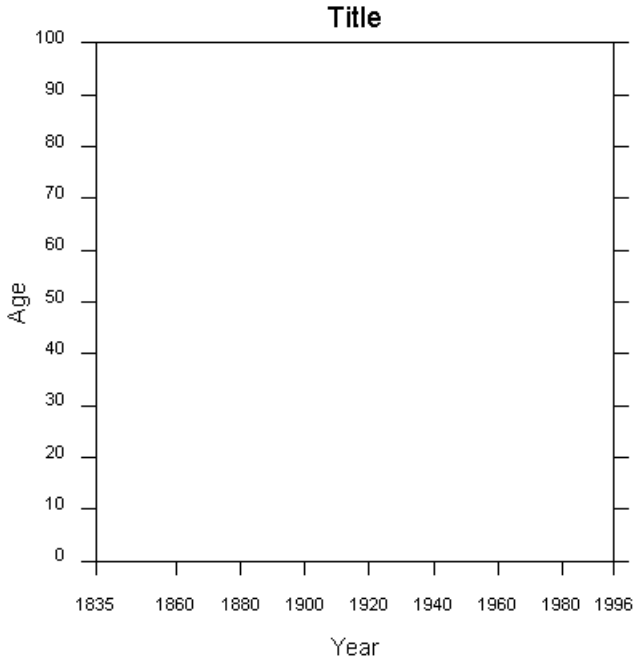
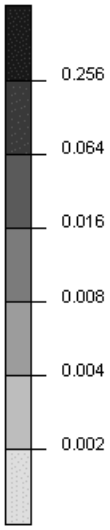


Figure 5.4 The example of Scale object.



The Scale Object (Fig. 5.4) is the graphic representation of the scale vector, which is used for conversion of the Data Matrix into the Lexis map image. The colors shown in the Scale Object are always the same as the Lexis map colors. If the user changes a scale color, the corresponding area of the Lexis map is automatically repainted with the new color. The user can customize position, size, number of levels, level colors and the format used to convert the scale values into the scale object tick labels.

Text, Rectangle and Line Objects are additional annotation elements that can be inserted, deleted or hidden. Their purpose is to customize the appearance of the map. They can be used, for example, to construct custom graphic objects that are not directly supported by Lexis. Properties of all graphics objects can be modified either via the menu system or by clicking on the object.

5.2.3 The Lexis map document

Lexis stores all parameters for displaying a plot in plain text (ASCII) files (Setup File) with the extension LEX. Information in the Setup File is divided into sections and each section contains a group of related items. Consider the following example:

```
[DATA]
MapMatrix=fusper.fmt;
Format=Gauss 3.2 DOS FMT;
```

Information in the section [DATA] is used by Lexis to determine the location (MapMatrix) and format (Format) of the Data Matrix. In this example, the matrix is loaded from the file fusper.fmt, which must be stored in the same folder as the Setup File. The next item (Format) tells Lexis that the matrix is stored in the format used by the program Gauss 3.2[37]. The online help system provided with Lexis contains exact documentation about which sections and items can be included in the Lexis Setup File.

Lexis creates associations with all files having the extension LEX. This means that the Lexis Setup Files are displayed with the Lexis icon in Windows[38] Explorer and can be opened by double clicking on the file name (or by clicking the right mouse button and then ‘Open’ in the shortcut menu).

Lexis can also be used as a command line utility. For example, the command

```
c:\>lexis fusper.lex musper.lex
```

[37] Gauss is a trademark of Aptech Systems, Inc.: www.aptech.com.

[38] Windows is a trademark of the Microsoft Corporation: www.microsoft.com.

will automatically launch Lexis and load two Lexis maps. Computer programs that can execute operating system commands can take advantage of this feature for viewing Lexis maps instantaneously without having to go via the Windows interface. In Gauss, for example, users can run Lexis by

```
» dos lexis fusper.lex
```

and in Matlab[39]

```
» !lexis fusper.lex
```

If you are used to working with the command prompt, it is even more simple to open a Lexis document: simply type in the name of the document and press ENTER:

```
c:>fusper.lex
```

The operating system will find the application (Lexis) associated with this file type (LEX) and open the Lexis map in this program.

Finally, we note that the Lexis Setup Files can be generated in virtually any programming language since they are simply plain text files. This permits the user to handle the extensive problems involved when hundreds of maps need to be created and analyzed in order to get a general overview of a problem.

5.2.4 Map Editor

The Map Editor displays a contour map and allows you to customize the plot appearance. You can use either the menu system or the graphical user interface of Lexis to modify the plot. A description of the most common tasks that can be carried out with the Map Editor is provided below.

Changing the appearance of a Lexis map

You can select the menu command Edit|Map to bring up the Lexis Map Appearance dialog box (Fig. 5.5). Here you can change the type of Lexis map (Table 5.1), add or remove contour lines, and change the color or width of the contour lines.

Changing the Plot Frame

Choose Edit|Plot Frame to bring up the complete Plot Frame dialog box (Fig. 5.6). Here you can

- specify the plot coordinate system;
- specify the page location of the plot;
- add, remove or edit titles, x- and y-labels;

[39] Matlab is a trademark of MathWorks, Inc.: www.mathworks.com.

- customize the x- and y-axes (coordinate range, tick locations, tick labels, etc.);
- add, remove or change the appearance of the grid lines;
- hide or bring to view the entire Plot Frame Object.

The Plot Frame can also be customized with a mouse click. You can point to the plot title, for example, and click the left mouse button. That part of the Plot Frame dialog will be brought up with which you can modify the title of the plot.

Figure 5.5 The Lexis Map Appearance dialog.

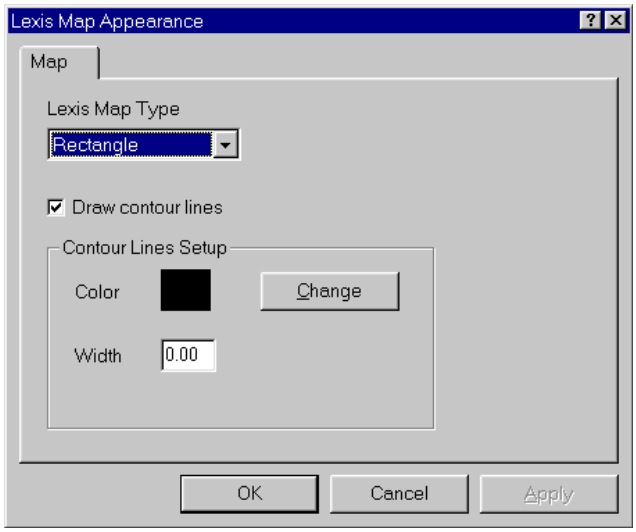
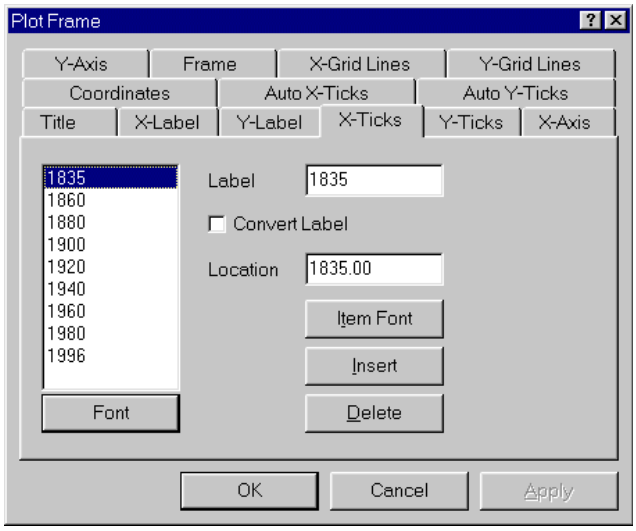


Figure 5.6 The Plot Frame dialog.

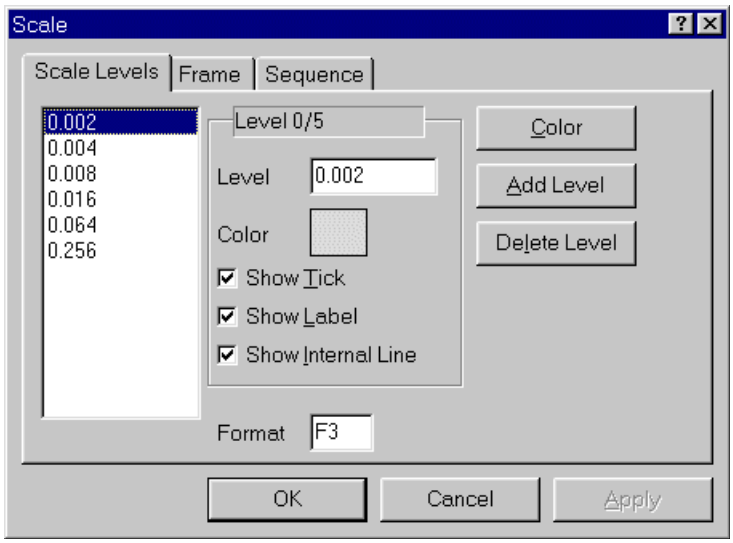


Changing the scale levels

There are three different ways to change the scale levels. The most general is to select Edit|Scale and make the necessary changes in the Scale dialog box (Fig. 5.7). Here you can

- add, delete or modify any scale level;
- show or hide the tick, the tick label and the internal line on the scale legend;
- change the format of conversion of scale values to the tick labels;
- change the page location of the scale;
- automatically setup the scale values from additive or multiplicative sequences;
- hide or bring to view the entire scale legend.

Figure 5.7 The Scale dialog.



You can also bring up the shortcut menu by clicking the right mouse button on the Lexis map. The action available to you now depends on the point at which you click the mouse. You can

- move the upper/lower contour line to the point of the mouse click;
- insert a new contour line at the point of the mouse click;
- delete the clicked map region.

Alternatively, you can use the shortcut menu, which is brought up by clicking the right mouse button on the scale legend itself.

Finally, you can change the scale values by selecting the menu command Edit|Smart Scale. The scale levels will now be automatically computed by Lexis in order to equalize the map areas of different colors. This option can be very useful if the Lexis surface is highly non-linear and the contour lines are clumped together. For example, the use of equally spaced scale values for producing a map of human mortality results in rather uninformative Lexis map (Fig. 5.8(a)). By selecting the menu command Edit|Smart Scale new scale values are computed and the map becomes more informative (Fig. 5.8(b)).

Changing the colors

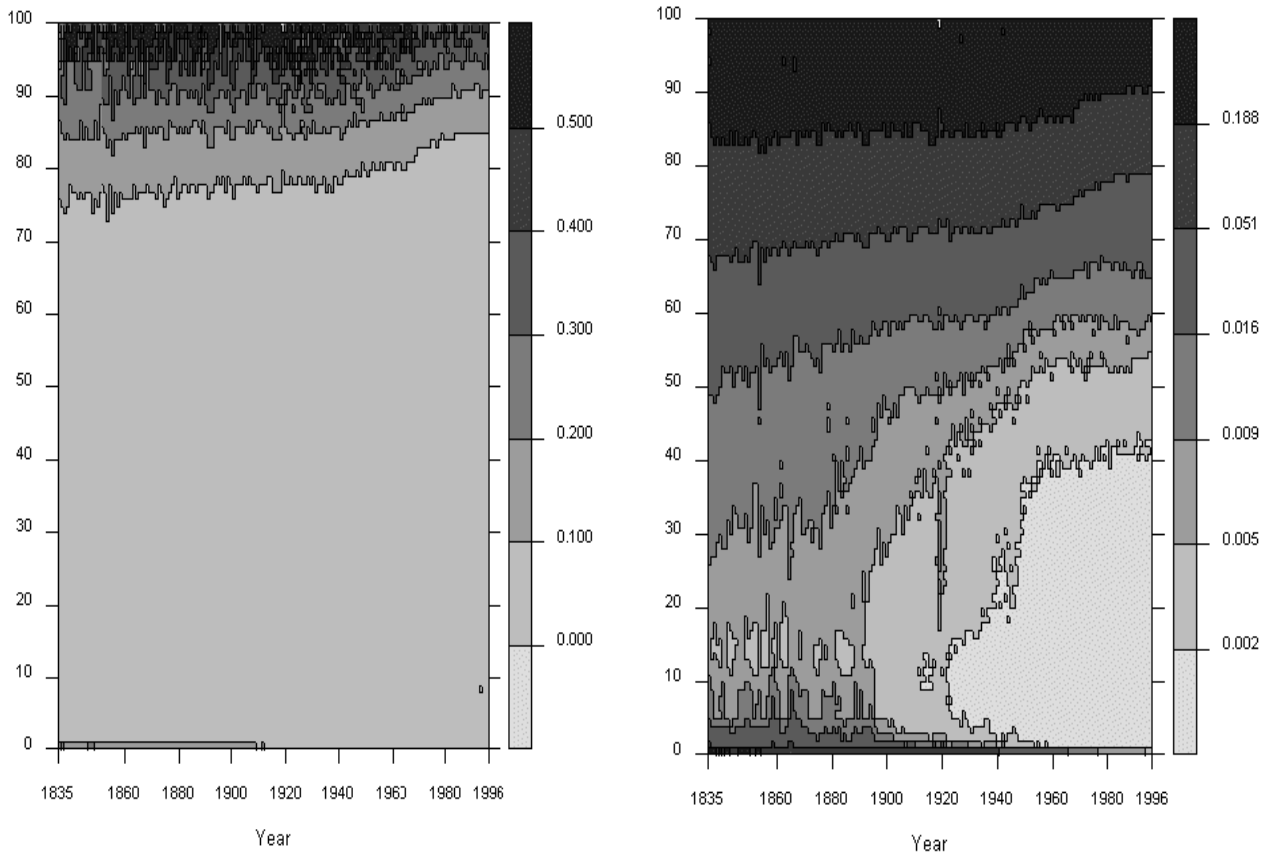
The colors of a Lexis map can be changed either manually with the help of the standard Windows Color dialog box or by loading the predefined color schemes. The options for changing the colors are provided in the Scale dialog box and in the shortcut menus. To change the color with the help of shortcut menu, for example, you have to click with the right mouse button on the scale box with the color you want to change. Then you choose ‘Change Color’ from the shortcut menu. This brings up the Color dialog box, and the first custom color box will be filled with the color of the

corresponding scale box. By changing the color in this box you can change the color of the map region.

Figure 5.8 Illustration of the menu command Edit|Smart scale.

a) A Lexis mortality map with evenly spaced contour lines

b) The same map after selecting the menu command Edit|Smart scale



Alternatively, you can choose Edit|Colors|Color Scheme and select the color from the predefined color schemes. Some color schemes (e.g. 'Geography') are provided only for certain numbers of scale levels whereas other color schemes (e.g. 'Rainbow', 'Random') can be used with any number of scale levels. You can click on the button 'Apply' in the dialog box to change the map colors temporarily. You can then always abandon the changes by clicking on the 'Cancel' button.

Zoom

The Map Editor allows you to zoom a plot area. The zoom factor is not fixed: it depends on the selected rectangle. Lexis will automatically adjust the x- and y-axis zoom factors in order to keep the proportions unchanged. The following steps are required for zooming in on a plot area:

- press and hold the CTRL key;
- click and hold the left mouse button;
- drag the mouse to select a rectangle to be zoomed;
- release the mouse button.

Another way to zoom a plot is by using the shortcut menu:

- press and hold the CTRL key;
- click the right mouse button on the place you want to zoom;
- select the 'Zoom In' option from the shortcut menu.

You can move around the enlarged map image with the arrow buttons on the keyboard or with the buttons on the Map Editor toolbar().

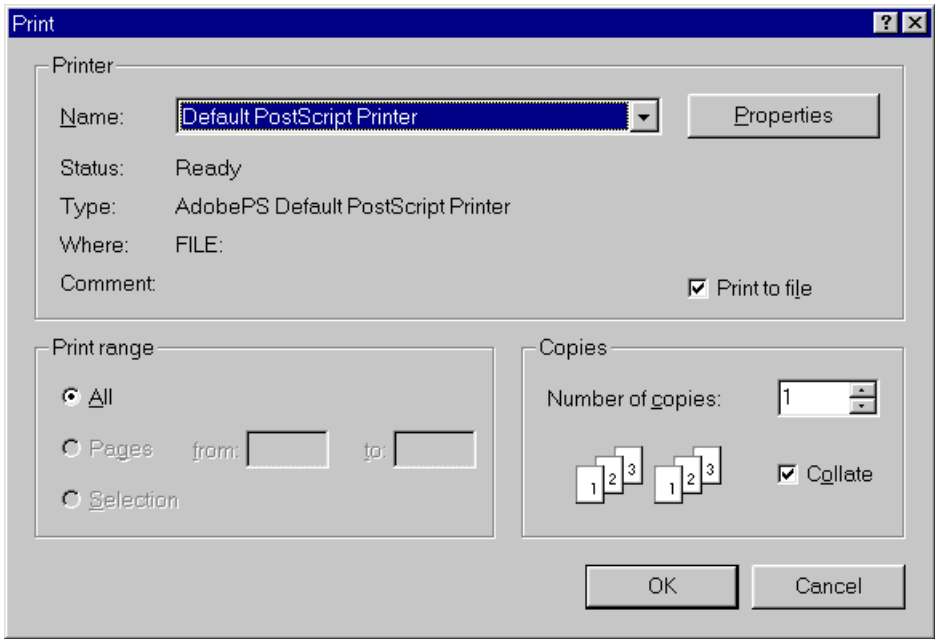
Printing

Lexis uses the metric coordinate system by default, and the location of the Plot Frame Object (Fig. 5.3) on the page is initialized for the A4 paper format. By using the Plot Frame Dialog box (Fig. 5.6) you can change the location of the plot on the page. The paper format is retrieved from the current printer settings, which you can view by choosing the menu option File|Print Setup. Here you can also change the orientation of the page. To facilitate printing Lexis draws a dashed rectangle around the plot area which shows the actual physical page. You can see how well your plot fits the page and make the necessary changes.

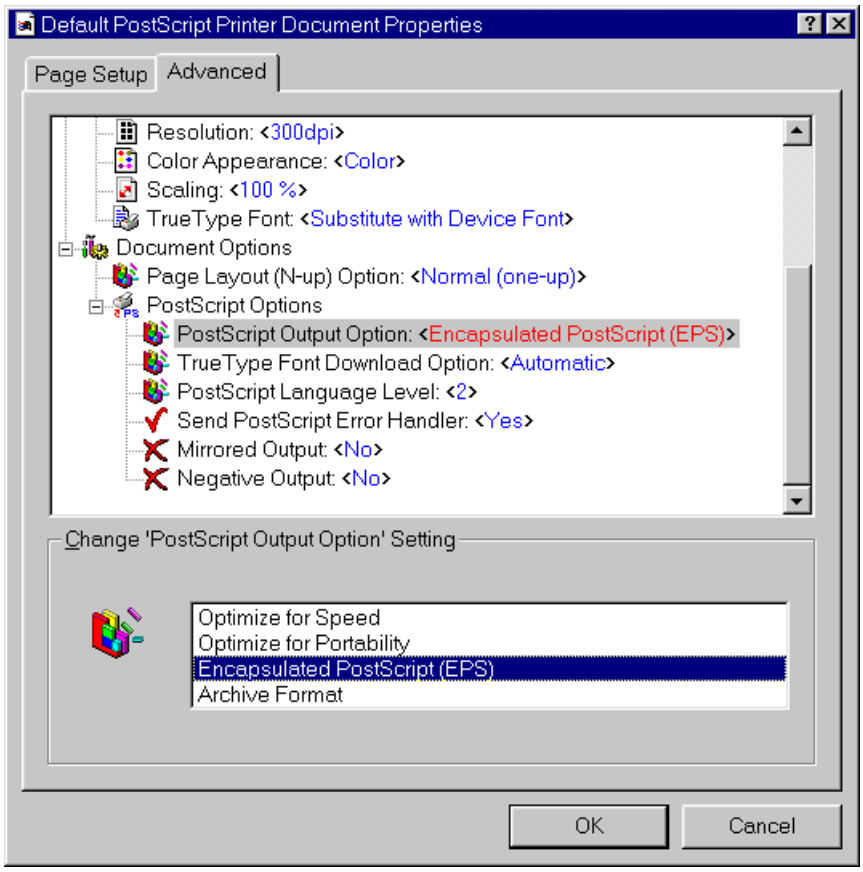
The option File|Print active windows is provided for printing more than one map on a single page. Lexis goes through all active windows and prints the contents of the windows onto one page. In this way a superimposed image is constructed.

It is easy to include a Lexis map in another document such as an MS Word[40] document. All you need to do is to write a printout into a file using a Postscript printer driver. The printer driver can be downloaded free of charge from the Adobe web site (www.adobe.com). After installation of the printer driver you get an additional printer on your system. However, the printer driver will not be connected to a physical device. Instead it will be connected to a file on your computer. In Lexis you need to select File|Print and the AdobePS Default Postscript printer:

[40] MS Word is a trademark of the Microsoft Corporation: www.microsoft.com.



Next, select “Encapsulated PostScript” in the printer properties and print the Lexis map.



A file which contains the Lexis map in Postscript format will be created on your hard disk. To insert the file into your Word document, select the menu command Insert|Picture. Here, you might wish to check the box “Link To File” to include only a link to the file since this can considerably reduce the

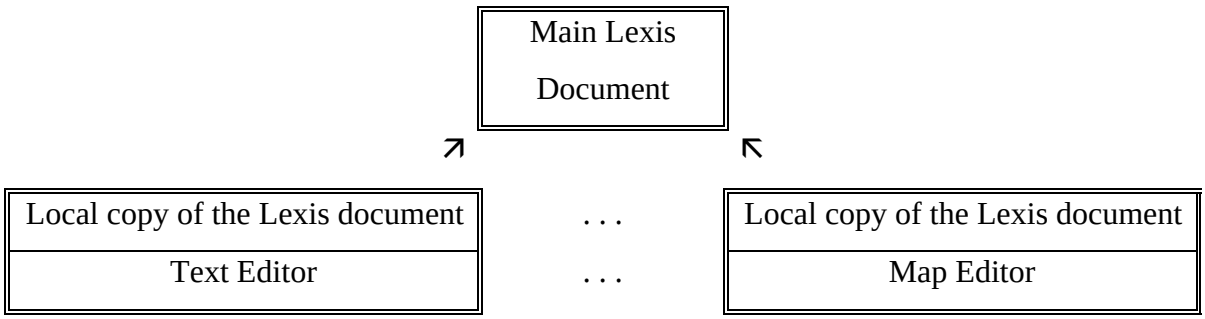
size of the document. Now you are ready to place and scale the Lexis map inside your document. If you print this document the Lexis map will now be printed as well.

5.2.5 Text editor

The Text Editor is included in Lexis for the direct manipulation of the Lexis Setup Files. It is a simple editor comparable to the program NOTEPAD, which is included with the Windows operating system. You can open the Text Editor either by choosing Window|Add View|Text or by clicking on the 'T' button on the toolbar.

The direct manipulation of a Setup File can result in problems with the plot if the syntax is not followed exactly. For this reason direct editing is recommended only for experienced users. A safer way to modify the plot is to use the graphical user interface of Lexis.

The same Lexis document can be opened both in the Text and the Map Editor. Each editor has its own copy of the main Lexis document stored on disk:



If you make any changes in either editor only the local copy of the document will be modified. To store the changes to disk you have to execute File|Save from the menu. As these changes are stored to disk all opened editors will reload the main document in order to update their local copies and repaint their windows. If there are any unsaved changes in the editors they will be lost. For example, you can change the name of the linked Data File in the Text Editor and save the document. If this document was opened in the Map Editor as well, it will reload the Data File and repaint its window reflecting the latest changes in the Lexis document.

The Text Editor also provides context-sensitive help. You can move the cursor to any item and press F1 to get more information about it.

If you open the Lexis document in a word-processor such as MS Word or Word Perfect you must specify that the file be saved in text format. Otherwise it will be saved in the application-specific binary format and Lexis will not be able to open it.

5.3 Graphical user interface (GUI)

The GUI facilitates the interaction between user and Lexis. Most actions performed by Lexis can be carried out with the help of the Lexis GUI, which consists of:

- a menu system;
- standard dialog boxes (Open, Save, Print);
- toolbars;
- mouse interface;
- tabbed dialog boxes;
- zoom;
- drag and drop support.

The menu system, standard dialog boxes and toolbars serve the same purpose as in most Windows-based software. The user can use the menu system for issuing commands and standard dialog boxes for opening, saving and printing the Lexis maps. Toolbars provide quick access to the most frequently-executed commands.

5.3.1 Mouse interface

The left mouse button is associated with default actions which can be carried out by pointing and clicking. Usually, this brings up a dialog box for modifying properties of the underlying graphic object. If, for example, you move the mouse pointer over the scale and click the left mouse button, this brings up the standard Color dialog, which you can use to modify the color associated with this scale level.

The right mouse button provides access to the list of commands that are appropriate in the given context. You can select a command from the shortcut menu or abort the action by pressing ESC. If, for example, you click the right mouse button on the scale image you will have the options of changing, deleting, or inserting the scale level or modifying its color.

5.3.2 Tabbed dialog boxes

The tabbed dialog boxes (cf. eg. Fig. 5.6) provide a safe way to modify the plot. A tabbed dialog includes a number of sub-dialogs which can be accessed by clicking on the associated tab. Although Lexis documents can be modified in any text editor, it is recommended that one use the dialog boxes since they provide an error-free way to modify the properties of the graphic objects and internal data structures. The input from the user is always validated and incorrect input information is rejected.

The dialog boxes can be accessed either through the menu system or by mouse click. All dialog boxes are provided with three buttons: ‘Ok’, ‘Apply’ and ‘Cancel’. The ‘Ok’ button closes the dialog and carries out the action. All changes specified in the dialog will be accepted if the validation succeeds. To discard all changes made in the dialog choose the ‘Cancel’ button or press Esc. The ‘Apply’ button acts much like the ‘Ok’ button except that it does not close the dialog box. You can make changes in the dialog box and then click the ‘Apply’ button to pass new parameters to the editor. The dialog will stay open and you can make more changes if necessary. All changes made by the ‘Apply’ button can still be abandoned by the ‘Cancel’ button. The ‘Apply’ button is disabled unless something in the dialog is modified.

5.3.3 Drag and drop support

You can drag a Lexis Setup File from Windows Explorer and drop it onto a running Lexis window. Lexis automatically loads the file and opens it in the Map Editor. You can select and drop as many files as needed to open them simultaneously.

5.4 Making a new map

In this section I describe the steps that bring you from the raw data to the Lexis map. First of all, select File|New in the menu and enter the file name of the matrix that will be plotted as a Lexis map. At this point you must specify the format of the file in the dialog field Files of type:. After you select the file name and format, click ‘Ok’ to proceed.

Lexis can load Data Files in the following formats:

- Gauss 3.2 DOS FMT files;
- ASCII files.

For more information on Gauss formats see Gauss manuals distributed by Aptech Systems, Inc[41]. Gauss 3.2 format is used by Gauss 3.2 DOS to store matrices of double values. Gauss missing values are supported by Lexis and automatically converted into the Lexis missing values.

To load an ASCII file you must specify the field (column) delimiter and the record (row) delimiter. The fields of the ASCII file will be converted into numeric values. If the sequence of ASCII characters cannot be converted into a numeric value, for example just point ".", this field will be converted into a missing value. ASCII is a widely used format and virtually all applications can

[41] www.aptech.com.

export data in this format. If you are working, for example, in Excel you can save your matrix in CSV (comma delimited) format and load it in Lexis as an ASCII Data File.

It is important to select an appropriate file type in the dialog since Lexis will use it to run the suitable conversion routine. For example, if your data file is stored in text format but you try to open it as a Gauss file, you will receive the error message that Lexis cannot load it because of the wrong format of the input file.

Upon opening the file, Lexis creates a Lexis document with the same file name as specified before but with the extension LEX. All parameters of this document are filled in using the default settings. Later on, the document is loaded into the Map Editor and the contour map is displayed in the window of the editor.

You are now ready to customize the appearance of the map with the graphic user interface. Click on the graphic object you wish to change, fill in the appropriate dialog entries and click 'Ok'.

Finally, execute the File|Save command to store the changes to disk. More details on making a Lexis map from scratch are provided in the online documentation. Here you can also find information on supported formats of input data and on how to prepare your data for loading into Lexis.

5.5 Technical data

Lexis is a 32-bit application for the operating systems Windows 95, 98 and NT 4.0. The software was written in C++ with the addition of assembler code to speed up the matrix operations.
Technical specifications:

- Size of the Data Matrix: - limited by available computer memory
- Maximum number of scale levels: - 65,535
- Maximum number of colors: - 16,777,216
- Maximum number of additional graphic elements: - limited by available computer memory

5.6 Distribution and copyright

Lexis 1.1 is copyrighted by Kirill Andreev and the Max Planck Institute for Demographic Research, Rostock, Germany. It is distributed free of charge from the MPIDR web site

(www.demogr.mpg.de). For more information on using Lexis for demographic research see the monograph by Vaupel et al. (1998), which is also available online from this site.

If you use the program, please acknowledge that it was developed by Kirill Andreev at Odense University and the Max Planck Institute for Demographic Research. You should also cite this PhD thesis. If you would like to be notified about further developments of this project, please send a request to Kirill Andreev (andreev@demogr.mpg.de).

SUMMARY

The abundant statistical data available for Denmark allow us to construct a Danish mortality surface for all ages for the period 1835–1996. I compiled all existing sources of information and then constructed a consistent database on Danish mortality. For the earlier years I applied methods of interpolation and prorating to obtain death distributions by single year of age and to estimate population counts between censuses. To produce population estimates for age 80 and above I applied a modified extinct-cohort method. Finally, I checked the database for errors and compared it with the official Danish life tables.

My mortality database can be used to study age-specific and cohort-specific mortality trends in the Danish population. It allows us to gain insights into the nature of mortality developments that are deeper than those we get by simply using crude mortality indicators such as life expectancy at birth or age-standardized mortality rates. I produced and discussed Lexis maps of Danish mortality, mortality progress, the death distribution, the mortality sex ratio and relative changes in population distribution. Here are the most important findings of this analysis:

- mortality transition in Denmark at the end of the 19th century belonged to the mainstream of transitions in other European countries. In fact, the mortality changes were even more favorable in Denmark than in other countries. Danish life expectancy seems to have been among the highest in Europe around 1910;
- starting in the 1960s mortality progress decelerated significantly, mostly because of stagnation or even a certain degree of increase in mortality in middle ages;
- starting in the 1960s rapid mortality progress is to be observed at age 80 and above;
- the sex ratio of mortality shows a very distinct age- and time-specific pattern. The highest differences between sexes are to be observed at ages 60–80 in the 1980s. The subsequent decline in sex differences in this age group was the main reason for a convergence in life expectancies for the male and female populations of Denmark in recent years;
- until the 1960s there was a general trend of compression of Danish mortality. After that time the decline in oldest-old mortality gained in importance and the proportions of deaths at the oldest ages rose substantially, which reduced the level of compression.

Similar mortality surfaces can be estimated for Sweden, the Netherlands, and Japan, which we can then compare with the mortality surface of Denmark. The most recent decades are of special interest since Danish life expectancy gains have been fairly moderate compared with those of other

developed countries. Age-specific differences in Danish survival were revealed by estimating the mortality ratio surfaces. This allowed me to identify the age- and time-specific areas of excess Danish mortality. Then, I analyzed cause-specific mortality in order to explore the relative contribution of the different causes of death to the excess of Danish mortality. The analysis was performed for 25 causes of death at ages 50–69, where the highest differences in mortality between Denmark and the other countries are found. While the causes of death that contribute most to excess mortality vary for the male populations, the pattern for females is remarkably similar. The most important cause of excess Danish male mortality compared with the Netherlands and Japan was cardiovascular diseases, whereas compared with Sweden it was cancer mortality. In the case of females the results point clearly to cancer mortality (especially lung and breast cancer) as the main contributor to excess Danish mortality.

I also explored time trends in cause-specific mortality rates for four age groups in the period from 1970 to 1993. An examination of trends in cause-specific mortality allowed us to discover those causes of death that contribute to excess Danish mortality. It turns out that they correspond closely for males and females. The most unfavorable trends involve diseases of the respiratory system: lung cancer, bronchitis, emphysema, and asthma. Mortality from breast cancer also exhibits a negative trend, since it has increased in Denmark while it has declined in Sweden. Mortality from cirrhosis of the liver is also a concern, as rising trends have been observed in Denmark alone. The trends in Danish mortality from ischaemic heart disease have been in accordance with developments in other countries. However, the decline in Danish mortality lagged behind that of other European countries, and Danish death rates remained consistently at a higher level. All 200 graphs of the cause-specific mortality trends are provided on the CD-ROM.

The demographic transition has led to the emergence of a new field: demography of the oldest-old. In the years 1990 to 1995 33% of male and 51% of female deaths in Denmark occurred at ages above 80. Unfortunately, the estimation of mortality at such advanced ages is hampered by inaccuracies in the data for national populations. In this thesis I evaluated existing methods of quality checking and developed new ones, as well as developing new methods of estimating the number of survivors of non-extinct cohorts at advanced ages. I applied these methods of quality checking to the Danish population and did not find any serious errors. In addition, I applied these methods to all databases included in the Odense Archive of Population Data on Aging and produced a quality assessment report.

Methods for estimating the number of survivors of non-extinct cohorts were tested on reliable data for Nordic populations and for other countries for earlier periods. The analysis

indicates that a newly developed method (MD) performs best in the most common situation – when no information is known other than death counts. The other methods (survival ratio and Das Gupta's methods) produce less accurate results because of mortality decline at advanced ages. This means that these methods can be used only if additional corrections for mortality decline are made.

During the course of my Ph.D. project I also developed a program called Lexis, which facilitates the creation and presentation of demographic surfaces based on the Lexis diagram. This program is a 32-bit application for Windows NT, Windows 95 and Windows 98. It is being used intensively for demographic research both at the Max Planck Institute for Demographic Research and at other collaborating scientific organizations. The software, which was written in C++, is provided on the accompanying CD-ROM.

DANSK RESUMÉ

Den store mængde statistiske data, som er tilgængelig for Danmark, giver mulighed for at konstruere en overflade af dansk mortalitet for alle aldre i perioden 1835-1996. I denne Ph.d.-afhandling samlede jeg alle de forskelligartede informationskilder og opbyggede en konsistent database over dansk mortalitet. Jeg har produceret og diskuteret Lexis diagrammer over dansk mortalitet, mortalitetsudvikling, aldersspecifik fordeling af dødsfald, kønsfordeling og relative ændringer i befolkningsfordeling. De vigtigste resultater i denne analyse er følgende:

- transitionen af mortaliteten i Danmark i slutningen af det 19. århundrede ligner det store flertal af overgange i andre europæiske lande. Faktisk var mortalitetsændringerne endnu mere gunstige i Danmark end i andre lande. Den forventede levealder i Danmark ser ud til at have været blandt de højeste i Europa omkring 1910;
- begyndende i 1960'erne mindskedes fremgangen betydeligt, hovedsageligt på grund af stagnation eller endog en vis grad af stigning i mortalitet blandt midaldrende;
- ligeledes begyndende i 1960'erne observeres en hastig mortalitetsreduktion ved 80 år og over;
- kønsfordelingen i mortaliteten viser et udpræget alders- og tidsspecifik mønster. De største forskelle mellem køn kan iagttages ved 60 til 80 årsalderen i 1980'erne. Faldet i kønsforskellen i denne aldersgruppe efter 1980'erne var hovedårsagen til en konvergens af den forventede levealder for den mandlige og den kvindelige del af befolkningen i Danmark i de senere år;
- indtil 1960'erne observeres en kompression af mortaliteten i den danske befolkning. Efter den tid tog nedgangen i mortalitet blandt de allerældste over, og andelen af dødsfald blandt de ældste steg betydeligt, hvilket reducerede kompressionsniveauet.

Lignende mortalitetsoverflader kan beregnes for Sverige, Holland og Japan, hvilket gør det muligt for os at sammenligne dem med Danmarks mortalitetsoverflade. Aldersspecifikke forskelle i overlevelse i Danmark blev demonstreret ved at beregne overfladerne af mortalitetsratio. Dette tillod mig at identificere det alders- og tidsspecifikke område af den forhøjede mortalitet i Danmark.

Derefter analyserede jeg den årsagsspecifikke mortalitet for at udforske de forskellige dødsårsagers relative bidrag til den forhøjede mortalitet i Danmark. Analysen blev udført for 25 dødsårsager i alderen 50-69, hvor der findes de største forskelle i mortalitet mellem Danmark og de andre lande. Jeg undersøgte også tidstendenser i de årsagsspecifikke mortalitetsrater for fire aldersgrupper i perioden fra 1970 til 1993. En undersøgelse af tendenser i årsagsspecifik mortalitet tillod os at udpege de dødsårsager, som bidrager til den danske forhøjelse af mortaliteten. Alle 200

grafer over de årsagsspecifikke tendenser i mortalitet er stillet til rådighed på CD-ROM'en.

I denne afhandling evaluerede jeg eksisterende metoder for kvalitetskontrol og udviklede nye metoder, herunder nye metoder til at estimere antallet af overlevende af "non-extinct cohorts" med fremskredne aldre. Desuden anvendte jeg disse metoder på alle databaser inkluderet i "Odense Archive for Population Data on Aging" og udarbejdede en rapport om denne kvalitetsvurdering.

I løbet af mit PhD-projekt udviklede jeg også et program ved navn Lexis, som letter oprettelsen og præsentationen af demografiske overflader baseret på Lexis diagrammet. Softwaren er stillet til rådighed på den medfølgende CD-ROM.

REFERENCES

1. Aarssen, K. and de Haan, L. On the Maximal Life Span of Humans. Mathematical Population Studies. 1994; 4(4):259-281.
2. Andersen, Otto. Dødelighedsforholdene i Danmark 1735-1839. Særtryk Af Nationaløkonomisk Tidsskrift; 1973; Statistisk Institute, Københavens Universitet.
3. Andreev, Kirill, Yashin, A. I., and Vaupel, J. W. The Danish Mortality Database and Mortality Differences in Denmark, Sweden, the Netherlands and Japan. Materials of Symposium i Anvendt Statistik 1997, Danmarks Tekniske Univerisitet. 1997 Jan:189-202.
4. Arthur, Brain W. and Vaupel, James W. Some General Relationships in Population Dynamic. Population Index. 1984 Summer; 50(2):214-226.
5. Bjerregaard, P. and Juel, K. Middellevetid og dødlighed i Danmark. UGESKR Læger. 1993 Dec 13; 155(50).
6. Bjerregaard, P. and Juel, K. Middellevetid og dødelighed. En analyse af dødeligheden i Danmark og nogle europæiske lande, 1950-1990. København, Dansk Institut for Klinisk Epidemiologi: Middellevetidsudvalget; 1994.
7. Caselli, G.; Vallin, J., Vaupel, J., and Yashin, A. Age-Specific Mortality Trends in France and Italy Since 1900: Period and Cohort Effects. European Journal of Population. 1987; 3:33-60.
8. Caselli, G., Vaupel, J., and Yashin, A. Mortality in Italy: Contours of a Century of Evoution. Paper Presented at Session F.8 of IUSSP International Population Conference, Florence 7-12 June, 1985. 1985.
9. Chiang, Chin Long. The Life Table and its applications. Robert E. Krieger Publishing Company, Inc; 1984; ISBN: 0-89874-570-5.
10. Christensen, Kaare; Vaupel, James W.; Holm, Niels V., and Yashin, Anatoli I. Mortality among twins after age 6: fetal origins hypothesis versus twin method. British Medical Journal. 1995 Feb; 310(6977):432-435. ISSN: 0959-8138.
11. Cleveland, William S. The Elements of Graphing Data. Hobart Press, Summit, New Jersey; 1994; ISBN: 0-9634884.
12. Coale, Ansley J. and Caselli, Graziella. Estimation of the Number of Persons at Advanced Ages from the Number of Deaths at Each Age in the Given Year and Adjacent Years. Genus. 1990; LXVI(1):1-23.
13. Coale, Ansley J. and Kisker, Ellen E. Mortality Crossovers: Reality or Bad Data? Population Studies. 1986; 40:389-401.

14. Coale, Ansley J. and Kisker, Ellen E. Defects in Data on Old-Age Mortality in the United States. Asian and Pacific Population Forum. 1990 Spring; 4(1):1-31. ISSN: 0891-2823.
15. Condran, Gretchen A., Himes, Christine L., and Preston, Samuel H. Old-Age Mortality Patterns in Low-Mortality Countries: an Evaluation of Population and Death Data at Advanced, 1950 to the Present. Population Bulletin of the United Nations. 1991; 30:23-59.
16. Curtsinger, J., Fukui, H. , Townsend, D., and Vaupel, J. Demography of genotypes: Failures of the limited life-span paradigm in Drosophila melanogaster. Science. 1992; 28:461-463.
17. Das Gupta, Prithwis. Reconstruction of the Age Distribution of the Extreme Aged in the1980 Census by the Method of Extinct Generations. Washington, D.C. 20233: Population Division U.S. Bureau of the Census; 1990.
18. Dechter, Aimee R. and Preston, S. H. Age misreporting and its effects on adult mortality estimates in Latin America. Population Bulletin of the United Nations. 1991; (31-32):1-16.
19. Dierckx, Paul. Curve and Surface Fitting with Splines. United States: Oxford university Press Inc., New York; 1993; ISBN: 0-19-853441-8.
20. Elo, Irma T. and Preston, Samuel H. Estimating African-American Mortality from Inaccurate Data. Demography. 1994 Aug; 31(3):427-458.
21. Fries, J. Aging, Natural Death, and Compression of Morbidity. The New England Journal of Medicine. 1980 Jul 17; 303(3):130-135.
22. Heligman, L. and Pollard, J. H., The age pattern of mortality, Journal of the Institute of Actuaries. 1980; 107:49-80.
23. Heuser, R. L. (1984), Fertility Tables for Birth Cohorts by Color, United States, 1917-1980, U.S. Department of Health Education, and Welfare, National Center for Health Statistics.
24. Hoem, Jan M. The Statistical theory of demographic rates. A review of current developments. Scandinavian Journal of Statistics. 1976; 3:169-185.
25. Hosmer, David W. and Lemeshow, Stanley. Applied logistic regression. New York: John Wiley & Sons; 1989; ISBN: 0-471-61553-6.
26. Hvidt, Kristian. Flugten til Amerika eller Drivkreafter i masseudvandringen fra Danmark 1868-1914.; 1971.
27. Impagliazzo, John. Deterministic Aspects of Mathematical Demography. : November 1984;
28. Johansen, H. C. and Boje, P. Working Class Housing in Odense 1750-1914. Scandinavian Economic History Review. 1986; 34(2):132-52.
29. Johansen, H. C. The Development of Reporting Systems for Causes of Deaths in Denmark. Unpublised Paper. 1996.

30. Johansen, H. C. Early Danish Parish Registers. Danish Center for Demographic Research. Research Report 3. 1998. ISSN: 1398-4292.
31. Juel, Knud and Sjol, Anette. Decline in Mortality from Heart Disease in Denmark : some Methodological Problems. Journal of Clinical Epidemiology. 1995; 48(4):467-472.
32. Kannisto, V., Lauritsen, J., Thatcher, R., and Vaupel, J. Reductions in Mortality at Advanced Ages: Several Decades of Evidence from 27 Countries. Population and Development Review. 1994 Dec; 20(4):793-810. ISSN: 0098-7921.
33. Kannisto, Väinö. Estimating Current Survivors from Number of Deaths: Some Experience and Unsolved Issues. Research Workshop on Oldest-Old Mortality.: Duke University; 1993 Mar.
34. Kannisto, Väinö. Quality Indicators For Data On Oldest-Old Mortality. Research Workshop on Oldest-Old Mortality: Duke University; 1993 Mar.
35. Kannisto, Väinö. Development of Oldest-Old Mortality, 1950-1990: Evidence from 28 Developed Countries. Odense University: Odense University Press; 1994; ISBN: 87 7838 015 4.
36. Kannisto, Väinö, Christensen, Kaare, and Vaupel, James. No increased mortality in later life for cohorts born during famine. Am J Epidemiol. 1997 Jun; 11(145):987-994.
37. Keiding, Niels. Statistical inference in the Lexis diagram. Phil. Trans. R. Soc. Lond. A (1990). 1990; 332:487-509.
38. Labat, J-C. and Dekneudt, J. Combien y a t-il de centenaires? In I.N.S.E.E. (ed), Les Menages: Mélanges en l'honneur de Jacques Desabie. Paris: Imprimerie Nationale, April 1989.
39. Lancaster, H. O. Expectations of Life : A Study in the Demography, Statistics, and History of World Mortality. New York: Springer; 1990.
40. Legge, Thomas M. Public Health in European Capitals. London; 1896.
41. Lexis, W. Einleitung in die Theorie der Bevölkerungsstatistik. Strassburg: Trübner. 1875; Pp. 5-7; translated to English by N. Keyfitz and printed, with Fig. 1, in Mathematical Demography 1977 (ed. D. Smith & N. Keyfitz). Berlin: Springer.
42. Madsen, Th. and Madsen, S. Diphtheria in Denmark. From 23,695 to 1 case - Post or propter. I. Serum therapy. II. Diphtheria immunization. Dan. Med. Bul. 1956; 3:112-21.
43. Mamelund, Svenn-Erik and Borgan, Jens-Kristian. Cohort and Period Mortality in Norway 1846-1994. Statistics Norway; 1996; ISBN: 82-537-4278-9.
44. Manton, Kenneth G. and Vaupel, James W. Survival after the age of 80 in the United States, Sweden, France, England, and Japan. The New England Journal of Medicine. 1995 Nov 2; 333(18):1232-1235.

45. Matthiessen, Poul C. Some aspects of the demographic transition in Denmark. Copenhagen: Copenhagen University; 1970; ISBN: 87 505 0091 0.
46. McKeown, T. The Modern Rise of Population. London; 1976.
47. McNeil, Donald R.; Trussel, James T.; Turner, John C. Spline Interpolation of Demographic Data. Readings in Population Research Methodology. Volume 1. Basic Tools. Reprinted from Demography 14, 2 (1977). pp. 245-52. 1993.
48. Natale, M., and A. Bernassola (1973), La mortalita per causa nelle regioni italiane, Tavole per contemporanei 1965-66 e per generazioni 1790-1969, Istituto di Demografia, Universita di Roma, n. 25, Roma.
49. Preston, Samuel H., Elo, Irma T., and Stewart, Quincy. Effects of Age Misreporting on Mortality Estimates at Older Ages. Population Aging Research Center, University of Pennsylvania, Working Paper Series No. 98-01. 1997 Sep.
50. Preston, Samuel H., Elo, Irma T., Rosenwaike, Ira, and Hill, Mark. African-American mortality at older ages : results of a matching study. Demography. 1996; 33(2):193-209.
51. Rosenwaike, Ira and Logue, Barbara. Accuracy of death certificate ages for the extreme aged. Demography. 1983; 20(4).
52. Schofield, R. Ed., Reher, D. S. Ed., and Bideau, A. Ed. The Decline of Mortality in Europe. Oxford: Clarendon Press; 1991.
53. Shryock, Henry S., Siegel, Jacob. Selected General Methods. Readings in Population Research Methodology. Volume 1. Basic Tools.; 1993.
54. Smith, E. The Peasant's Home 1760-1875. London; 1876.
55. Sundhedsministeriet. Danskernes dødelighed i 1990'erne. 1. delrapport fra Middellevetidsudvalget. Nyt Nordisk Forlag Arnold Busck A/S; 1998 Dec; ISBN: 87-17-06878-9.
56. Sundhedsministeriets Middellevetidsudvalg. Danmark. Rapport. Komplet 1-14. København: Middellevetidsudvalget; 1993; ISBN: 87-601-4108-5.
57. Tabeau, Ewa, Frans van Poppel, and Willekens, Frans. Mortality in the Netherlands: The Data Base. The Hague; 1994; ISBN: 90-70990-46-6.
58. Thatcher, A. R., Väinö, Kannisto, and Vaupel, James W. The force of mortality at ages 80 to 120. Odense: Odense University Press; 1998; Odense Monographs on Population Aging ; 5.
59. Thatcher, A. Roger. Trends in Numbers and Mortality at High Ages in England and Wales. Population Studies. 1992; 46:411-426.

60. Thatcher, A. Roger. Overview of Methods for Estimating Population Numbers At High Ages from Data on Deaths. Draft; 1993 Feb.
61. Thatcher, A. Roger. The Quality of Data on High Ages in England and Wales. Workshop at Duke University. 1993 Mar 4-1993 Mar 6.
62. Vallin, J. (1973), La mortalité par génération en France, depuis 1899, Travaux et Documents, Cahier n. 63, Press Universitaires de France, Paris.
63. Vaupel, J. W., Zhenglian, W., Andreev, K. F., and Yashin, A. I. Population Data at Glance: Shaded Contour Maps of Demographic Surfaces over Age and Time. Odense University, Denmark: Odense University Press; 1998; ISBN: 87-7838-338-2.
64. Veys, D. (1983), Cohort Survival in Belgium in the Past 150 Years, Catholic University of Leuven, Sociological Research Institute, Leuven, Belgium.
65. Vincent, Paul. La Mortalité des vieillards. Population. 1951; 6(2):181-204.
66. Wilmoth, J. R., Lundström, H. Extreme Longevity in five countries. European Journal of Population. 1996; 12:63-93.
67. Yashin, A. I., Vaupel, J. W., Andreev, K. F., Tan, Q., Iachine, I. A., Carotenuto, L.; De Benedictis, G.; Bonafe, M., Valensin, S., and Franceschi, C. Combining genetic and demographic information in population studies of aging and longevity. Journal of Epidemiology and Biostatistics. 1998; 3(3):289-294.

APPENDIX

1. Appendix Table 2.1 The mortality databases used in data quality checks.
2. Appendix Table 3.1 Raw population data.
3. Appendix Table 3.2 Raw death counts data.
4. Appendix Table 3.3 Earlier publications of Danish population statistics.
5. Appendix Table 3.4 The average deviation between the genuine and interpolated death distributions for the years 1916, 1921-1940.
6. Appendix 4.1 Estimating mortality progress surfaces.
7. Appendix 4.2 Kernel smoothing of Lexis maps.
8. Appendix 4.3 Estimating mortality ratio surfaces.
9. Appendix Table 4.1 List of causes of deaths selected for the analysis of mortality differences.

Appendix Table 2.1 The mortality databases[42] used in data quality checks.

Country	First Year	Last Year	First	Last	Abbreviation
Australia	1965	1991	80	+	AUSTL
Austria	1947	1996	80	+	AUSTR
Belgium	1950	1996	80	+	BELGI
Canada	1950	1996	80	+	CANAD
Chile	1980	1989	80	+	CHILE
Czech Republic	1950	1996	80	101	CSR
Denmark (NSO)[43]	1835	1996	0	+	DENMA
England & Wales	1911	1996	80	+	ENWAL
Estonia	1950	1996	80	+	ESTON
Finland	1878	1996	50	+	FINLA
France	1950	1996	80	+	FRANC
Germany, East	1954	1996	80	+	GERME
Germany, West	1951	1996	80	+	GERMW
Hungary	1950	1991	80	100	HUNGA
Iceland	1947	1995	80	+	ICELA
Ireland	1950	1993	80	+	IRELA
Italy	1952	1994	80	+	ITALY
Japan	1950	1996	80	+	JAPAN
Japan (BMD)[44]	1950	1996	0	+	JAPW
Latvia	1950	1996	80	100	LATVI
Luxembourg	1953	1996	80	100	LUXEM
Netherlands (NSO)[45]	1850	1994	0	+	NLEWA
Netherlands	1950	1997	80	+	NETHE
New Zealand (Maori)	1950	1996	80	+	NZMAO
New Zealand (non-Maori)	1950	1996	80	+	NZNON
Norway (NSO)[46]	1846	1994	0	+	NORJK
Norway	1911	1996	80	+	NORWA
Poland	1971	1997	80	100	POLAN
Portugal	1929	1996	80	+	PORTU
Scotland	1950	1996	80	100	SCOTL
Singapore (Chinese)	1982	1996	80	+	SINGC
Slovakia	1950	1991	80	100	SLR
Slovenia	1983	1996	80	+	SLOVE
Spain	1946	1994	80	+	SPAIN
Sweden (BMD)[44]	1861	1996	0	+	SWWIL
Sweden	1920	1996	80	+	SWEKA
Switzerland	1950	1996	80	+	SWITZ
USA[47]	1962	1990	80	+	USACB

[42] The data belong to the K-T database unless otherwise indicated.

[43] The database was compiled by K. Andreev using available Danish statistical publications.

[44] The data are from Berkeley Mortality Database (BMD). For more information read the online documentation at http://demog.berkeley.edu/wilmoth/mortality.

[45] The data are originally from the Central Statistical Bureau of the Netherlands. The construction of the mortality database was carried out by Tabeau at al. (1994).

[46] The data are originally from the Central Statistical Bureau of Norway. The construction of the mortality database was carried out by Mamelund and Borgan (1996).

[47] The data are provided by Prof.Manton, Duke University (http://cds.duke.edu/). Population estimates are not available for this database.

Appendix Table 3.1 Raw population data.

Period	Age groups	Reference
1801	0-10, 10-20 ... 100+	Befolkningsforholdene i Danmark i det 19 aarhundrede. Census.
1834		
1840	0-1, 1-3, 3-5, 5-10, ... 110+	Census.
1845		
1850	0-1, 1-3, 3-5, 5-7, 7-10, 10-15, ... 100+	Census.
1855	0-1, 1-3, 3-5, 5-6, 6-7, 7-10, 10-14, 14-15 ... 24-25, 25-30 ... 100+	Census.
1860		
1870	0-1, 1-2 ... 100+	Befolkningsforholdene i Danmark i det 19 aarhundrede. Census.
1880		
1890		
1901		
1906-1940	0-1 ... 85+	Estimates from Danmarks Statistik.
1941-1970	0-1 ... 100+	Estimates from Danmarks Statistik.
1971-1975	0-90+	Befolkningens bevægelser. Danmarks Statistik. KBH.
	90-99+	Befolkningen i kommunerne pr 1. Januar.
1976-1991	0-1 ...	Befolkningens bevægelser. Danmarks Statistik.
1992-1993	0-1 ...	Befolkningens bevægelser. Danmarks Statistik.
1994	0-100	Danmarks Statistik. Befolknings bevægelser 1993. Table 103, page 170.
	100+	Provided by A. Skytthe, Odense University. Originally from Danish CPR register.
1995	0-100	Danmarks Statistik. Befolknings bevægelser 1994. Table 109, page 176.
	100+	Provided by V. Kannisto.
1996	0-100	Danmarks Statistik. Befolknings bevægelser 1995. Table 115, page 196.
	100+	Provided by V. Kannisto.

Appendix Table 3.2 Raw death counts data.

Period	Age	Source
1835-1854	0-1, 1-3, 3-5, 5-10 ... 110+	Statistik Tabelværk.
1855-1869	0-1, ... 4-5, 5-10 ... 100+	Statistik Tabelværk. Detailed statistics for the first year of age is available.
1870-1915	0-1 ... 4-5, 5-10 ... 95-100, 100+	Statistik Tabelværk.
1916-1920	0,1,2 .. 100+	Statistik Tabelværk.
1921-1942	0,1,2 .. 100+ by year and cohorts	Statistik Tabelværk. Befolkningens bevægelser.
1943-1995	All ages by year and cohorts	Befolkningens bevægelser. Death counts for ages 100 and above were obtained directly from Danmarks Statistik.

Appendix Table 3.3 Earlier publications of Danish population statistics (reproduced from Befolkningens Bevægelser 1993, Danmarks Statistik).

Statistisk tabelværk (Table Works)		Vielser, fødsler og dødsfald (Marriages, Births and Deaths)					
1801-33:	I, 1	1870-74:	IV A,	1	1906-10:	IV A,	8
1834-39:	I, 6	1875-79:	IV A,	2	1911-15	IV A,	13
1840-44:	I, 10	1880-84:	IV A,	5	1916-20:	IV A,	15
1845-49:	II, 1	1885-89:	IV A,	7	1921-25:	IV A,	17
1850-54:	II, 17, 1. del	1890-94:	IV A,	9	1926-30:	IV A,	19
1855-59:	III, 2	1895-1900:	V A,	2	1931-40:	IV A,	22
1860-64:	III, 12	19. aarh.*	V A,	5	1941-55:	1962:	I
1865-69:	III, 25	1901-05:	V A,	6	1956-69:	1973:	XI

*Befolkningsforholdene i Danmark i det 19. Aarhundrede. Statistisk tabelværk. Femte Række, Litra A, Nr.5, 1905

Statistiske meddelelser				Befolkningens bevægelser			
1931-33:	4, 95,4	1946:	4, 126,6	1958:	1960:2	1970:	1972:7
1934:	4, 97,6	1947:	4, 133,3	1959:	1961:1	1971:	1973:10
1935:	4, 100,4	1948:	4, 138,3	1960:	1962:8	1972:	1974:9
1936:	4, 102,5	1949:	4, 143,4	1961:	1963:5	1973:	1975:9
1937:	4, 106,5	1950:	4, 147,2	1962:	1964:5	1974:	1976:5
1938:	4, 109,3	1951:	4, 150,3	1963:	1965:5	1975:	1977:4
1939:	4, 110,5	1952:	4, 154,2	1964:	1966:4	1976:	1978:1
1940:	4, 111,5	1953:	4, 157,4	1965:	1967:7	1977:	1978:12
1941:	4, 155,5	1954:	4, 161,4	1966:	1968:6	1978:	1980:3
1942:	4, 119,4	1955:	4, 166,3	1967:	1969:1	1979:	1981:1
1943:	4, 120,5	1956:	4, 167,2	1968:	1970:4	1980:	1982:1
1944-45:	4, 125,4	1957:	4, 173,2	1969:	1971:3		

Yearbooks Befolkningens bevægelser

1981 pub. in 1983	1985 pub. in 1987	1989 pub. in 1991	1993 pub. in 1995
1982 pub. in 1984	1986 pub. in 1988	1990 pub. in 1992	1994 pub in 1996
1983 pub. in 1985	1987 pub. in 1989	1991 pub. in 1993	1995 pub in 1997
1984 pub. in 1986	1988 pub. in 1990	1992 pub. in 1994	

Appendix Table 3.4 The average deviation between the genuine and interpolated death distributions for the years 1916, 1921-1940.

Deviation equation	Interpolation scheme					
	Sprague	Beers Ordinary	Beers Modified	Karup-King	Cubic spline	5th order spline
	Males					
(A.3.1)	6.403e-03	6.283e-03	5.998e-03	6.100e-03	**5.627e-03**	5.765e-03
(A.3.2)	6.431e-03	6.247e-03	6.196e-03	6.238e-03	**5.751e-03**	5.870e-03
(A.3.3)	1.097e-06	1.089e-06	1.137e-06	1.139e-06	**1.087e-06**	1.101e-06
(A.3.4)	1.092e-06	1.085e-06	1.137e-06	1.133e-06	**1.084e-06**	1.099e-06
(A.3.5)	6.368e-05	6.274e-05	6.466e-05	6.491e-05	**6.144e-05**	6.239e-05
	Females					
(A.3.1)	5.736e-03	5.455e-03	5.319e-03	5.398e-03	**5.042e-03**	5.226e-03
(A.3.2)	5.774e-03	5.466e-03	5.552e-03	5.636e-03	**5.190e-03**	5.343e-03
(A.3.3)	1.004e-06	**9.944e-07**	1.053e-06	1.071e-06	1.000e-06	1.026e-06
(A.3.4)	1.011e-06	**1.002e-06**	1.066e-06	1.080e-06	1.008e-06	1.034e-06
(A.3.5)	5.624e-05	5.529e-05	5.771e-05	5.839e-05	**5.483e-05**	5.623e-05

Equations used to compute the deviation δ between the original and interpolated death distributions:

$$\delta = \frac{1}{n} \sum_{y=y_{\min}}^{y_{\max}} \sum_{x=5}^{99} \frac{(p_o(x,y) - p_i(x,y))^2}{p_i(x,y)} \qquad \text{(A.3.1)}$$

$$\delta = \frac{1}{n} \sum_{y=y_{\min}}^{y_{\max}} \sum_{x=5}^{99} \frac{(p_o(x,y) - p_i(x,y))^2}{p_o(x,y)} \qquad \text{(A.3.2)}$$

$$\delta = \frac{1}{n} \sum_{y=y_{\min}}^{y_{\max}} \sum_{x=5}^{99} (p_o(x,y) - p_i(x,y))^2 \; p_i(x,y) \tag{A.3.3}$$

$$\delta = \frac{1}{n} \sum_{y=y_{\min}}^{y_{\max}} \sum_{x=5}^{99} (p_o(x,y) - p_i(x,y))^2 \; p_o(x,y) \tag{A.3.4}$$

$$\delta = \frac{1}{n} \sum_{y=y_{\min}}^{y_{\max}} \sum_{x=5}^{99} (p_o(x,y) - p_i(x,y))^2 \tag{A.3.5}$$

where $p_o(x,y)$, $p_i(x,y)$ are proportions of the original and the interpolated death distributions at age x and year y, and n is the total number of years used in summing up.

Appendix 4.1 Estimating mortality progress surfaces.

Let

$$m_{x,y} = \frac{D_{x,y}}{N_{x,y}} \tag{A.4.1}$$

be the central death rate at age x and year y, where $D_{x,y}$ is the death counts and $N_{x,y}$ is the population estimate in the middle of year y. In order to estimate mortality progress I select death rates in k preceding years and k following years and at the same age x. I also assume that mortality increases exponentially during the period $[y-k, y+k]$:

$$\ln m_{x,y} = \ln m_x + \rho_{x,y} Y \tag{A.4.2}$$

where Y is the current year, $\rho_{x,y}$ is mortality progress at age x and year y (this parameter will be negative if mortality is declining and positive if it is increasing) and m_x is the death rate at $Y = 0$ (for estimation purposes it is recommended that the variable Y be normalized by subtracting the current year y).

Parameter estimates are obtained by maximizing the Poisson loglikelihood function:

$$L = \sum_{Y=y-k}^{Y=y+k} D_Y (\ln m_x + \rho_{x,y} Y) - N_Y e^{\ln m_x + \rho_{x,y} Y} \tag{A.4.3}$$

Hypothesis $\rho_{x,y} = 0$ can be tested with the likelihood ratio test.

Appendix 4.2 Kernel smoothing of Lexis maps.

Let $m_{i,j}$ be an element of the matrix used to produce a Lexis map in which i is the row index and j is the column index. Usually i denotes the current age and j is the current year but this is not required. The $m_{i,j}$ itself can be any demographic indicator such as central death rate, population level, mortality ratio, etc. In this method the value of $m_{i,j}$ is replaced by the weighted average of the $(2k+1)^2$ values in the $2k+1$ square of points:

$$m^*_{i,j} = \sum_{x=i-k}^{x=i+k} \sum_{y=j-k}^{y=j+k} w_{x,y} m_{x,y} \qquad \text{(A.4.4)}$$

The weights $w_{i,j}$ can be computed by selecting a bivariate kernel function $K(x, y)$. Using this kernel function we can select any size k of the smoothing matrix and compute the weighting matrix $w_{i,j}$:

$$w_{i,j} = \int_{i^*}^{i^*+h} \int_{j^*}^{j^*+h} K(x, y)dxdy \qquad \text{(A.4.5)}$$

where $h = \frac{2}{2k+1}$, $i^* = (i-1)h - 1$, $j^* = (j-1)h - 1$ and $i, j \in [1,2,..2k+1]$.

A convenient choice would be the bivariate Epanechnikov kernel $K(x, y) = 0.75^2(1-x^2)(1-y^2)$. In this case the 3x3 smoothing matrix is as follows $w_{i,j} = \begin{matrix} 0.06722 & 0.12483 & 0.06722 \\ 0.12483 & 0.23182 & 0.12483 \\ 0.06722 & 0.12483 & 0.06722 \end{matrix}$.

Appendix 4.3 Estimating mortality ratio surfaces.

Let $m_{x,y}$ and $m^*_{x,y}$ be the central death rates in the first population and in the second population, respectively (see Appendix 4.1). Let $w_{i,j}$ be the weighting matrix generated by some kernel $K(x, y)$ (see Appendix 4.2). We can employ Poisson regression to estimate the ratio $r_{x,y}$ of two mortality surfaces at age x and in the year y:

$$\ln m_{x,y} = \beta_0 + \beta_1 X \qquad \text{(A.4.6)}$$

where X is the dummy variable equal to 0 for the first population and 1 for the second one. Parameter estimates of an analytic form can easily be found:

$$\hat{\beta}_0 = \ln \frac{\sum_{i,j} w_{i,j} D_{i,j}}{\sum_{i,j} w_{i,j} N_{i,j}} \qquad \text{(A.4.7)}$$

$$\hat{\beta}_1 = \ln \frac{\sum_{i,j} w_{i,j} D^*_{i,j}}{\sum_{i,j} w_{i,j} N^*_{i,j}} - \hat{\beta}_0 \qquad \text{(A.4.8)}$$

Finally, $r_{x,y}$ is computed as $r_{x,y} = e^{\hat{\beta}_1}$. In addition, a likelihood ratio test can be performed in order to test hypothesis $\beta_1 = 0$.

A convenient choice for the kernel functions would be the Epanechnikov kernel (Appendix 4.2) and single-year-age kernel ($w_{i,j}$ is equal to 1 for $i = x$ and $j = y$, and 0 otherwise). In latter case $r_{x,y}$ is simply the ratio of the corresponding central death rates $r_{x,y} = \frac{m_{x,y}}{m^*_{x,y}}$.

Appendix Table 4.1 List of causes of deaths selected for the analysis of mortality differences.

Cause	Cause of death	ICD9 BTL	ICD8 A-List	ICD 7 A-List
	Chapter I. Infective and parasitic diseases.			
1	Infective and parasitic diseases	B01x-07x	A1-44	A1-43
	Chapter II. Malignant neoplasm.			
2	Malignant neoplasm of trachea, bronchus and lungs	B101	A51	A50
3	Malignant neoplasm of prostate	B124	A57	A54
4	Malignant neoplasm of intestine, except rectum	B092-093	A48	A47
5	Malignant neoplasm of stomach	B091	A47	A46
6	Malignant neoplasm of rectum and rectosigmoid junction	B094	A49	A48
7	Malignant neoplasm of breast	B113	A54	A51
8	Residual neoplasm	B08x-17x	A45-61	A44-60
	Chapter III. Cardiovascular Diseases.			
9	Ischaemic heart disease	B27	A83	A81
10	Other forms of heart disease	B28	A84	A82
11	Cerebrovascular disease	B29	A85	A70
12	Disease of arteries, arterioles and capillaries	B300-302	A86	A85
13	Residual cardiovascular diseases	B25x-30x	A80-88	A79-86
	Chapter IV. Diseases of respiratory system.			
14	Pneumonia	B321	A91-92	A89-91
15	Bronchitis, emphysema and asthma	B323-325	A93	A93
16	Influenza	B322	A90	A88
17	Residual respiratory diseases	B31x-32x	A89-96	A87-97
	Chapter V. Diseases of digestive system.			
18	Cirrhosis of liver	B347	A102	A105
19	Peptic ulcer	B341	A98	A99-100
20	Residual diseases of digestive system	B33x-34x	A97-A104	A98-107
	Chapter VI. Accidents, poisonings, and violence (E).			
21	Suicide and self inflicted injury	B54	A147	A148
22	Motor vehicle accidents	B471	A138	A138
23	Accidental falls	B50	A141	A141
24	Residual accidents	B47x-56x	A138-150	A138-150
	Chapter VII. Residual group of diseases.			
25	Residual group of diseases			