Roland Rau
Christina Bohk-Ewald
Magdalena M. Muszyńska
James W. Vaupel

Visualizing Mortality Dynamics in the Lexis Diagram

Springer Open

The Springer Series on Demographic Methods and Population Analysis

Volume 44

Series Editor
Kenneth C. Land, Duke University

In recent decades, there has been a rapid development of demographic models and methods and an explosive growth in the range of applications of population analysis. This series seeks to provide a publication outlet both for high-quality textual and expository books on modern techniques of demographic analysis and for works that present exemplary applications of such techniques to various aspects of population analysis.

Topics appropriate for the series include:

- General demographic methods
- Techniques of standardization
- Life table models and methods
- Multistate and multiregional life tables, analyses and projections
- Demographic aspects of biostatistics and epidemiology
- Stable population theory and its extensions
- Methods of indirect estimation
- Stochastic population models
- Event history analysis, duration analysis, and hazard regression models
- Demographic projection methods and population forecasts
- Techniques of applied demographic analysis, regional and local population estimates and projections
- Methods of estimation and projection for business and health care applications
- Methods and estimates for unique populations such as schools and students

Volumes in the series are of interest to researchers, professionals, and students in demography, sociology, economics, statistics, geography and regional science, public health and health care management, epidemiology, biostatistics, actuarial science, business, and related fields.

More information about this series at http://www.springer.com/series/6449

Roland Rau • Christina Bohk-Ewald
Magdalena M. Muszyńska • James W. Vaupel

Visualizing Mortality Dynamics in the Lexis Diagram

Springer Open

Roland Rau
Faculty of Economic and Social Sciences
University of Rostock
Rostock, Germany

Christina Bohk-Ewald
Max Planck Institute for Demographic
    Research
Rostock, Germany

Magdalena M. Muszyńska
Collegium of Economic Analysis
Warsaw School of Economics
Warsaw, Poland

James W. Vaupel
Max Planck Institute for Demographic
    Research
Rostock, Germany

Department of Public Health
University of Southern Denmark
Odense C, Denmark

Duke University
USA

Acknowledgments

Contents

List of Figures

List of Tables

Chapter 1
Introduction: Why Do We Visualize Data and What Is This Book About?

> *Restate my assumptions:*
> *One, mathematics is the language of nature.*
> *Two, everything around us can be represented and understood through numbers.*
> *Three, if you graph the numbers of any system, patterns emerge.*
>
> Sean Gullette as Maximilian Cohen in the movie π (1998).

The goal of this book is simple: We would like to show how mortality dynamics can be visualized in the so-called Lexis diagram. To appeal to as many potential readers as possible, we do not require any specialist knowledge. This approach may be disappointing: Demographers may have liked more information about the mathematical underpinnings of population dynamics on the Lexis surface as demonstrated, for instance, by Arthur and Vaupel in 1984. Statisticians would have probably preferred more information about the underlying smoothing methods that were used. Epidemiologists likewise might miss discussions about the etiology of diseases. Sociologists would have probably expected that our results were more embedded into theoretical frameworks....

We are aware of those potential shortcomings but believe that the current format can, nevertheless, provide interesting insights into mortality dynamics, and we hope our book can serve as a starting point to visualize data on the Lexis plane for those who have not used those techniques yet.

Visualizing data has become increasingly popular in recent years.[1] But why do we visualize data at all? Countless books on *how* to visualize data — often with a specific software tool in mind — are published every year. Maybe it seems to be too

---

[1]This trend is probably best demonstrated by visualizing the popularity of the term "visualizing data" over time, for instance, via Google's *Ngram viewer*. Google Books Ngram Viewer displays the relative frequency of a search term in a corpus of books during a given time frame. Please see, for example: https://books.google.com/ngrams/graph?content=visualizing+data+&year_start=1960&year_end=2008

obvious, but only a few of those publications address the question of *why* one should visualize data at all. According to the ones covering the topic, the purpose of data visualization can be narrowed down to three reasons (e.g., Tukey 1977; Schumann and Müller 2000; Few 2014):

1. *Exploration:* John Tukey stresses that exploratory data analysis "can never be the whole story, but nothing else can serve as the foundation—as the first step" (Tukey 1977, p. 3). He uses the expression of "graphical detective work" by trying to uncover as many important details about the underlying data as possible. If one explores data only with preconceived notions and theories, it is likely that essential characteristics remain undiscovered.
2. *Confirmation:* It could be argued that the mere exploration of data without any hypotheses is a misguided endeavor. Exploration needs to be firmly distinguished from confirmatory analysis, though. While the exploration is comparable to the work of the police, this step can be seen as the task of a judge or the jury. Both are important to advance science, the first step is to gather the facts whereas the second step is of judgmental nature: Can the "facts" be interpreted to support the theory? Or do certain findings exclude some hypotheses? In this sense, confirmatory analysis represents the core of scientific progress in Popper's sense, namely by falsifying theories.
3. *Presentation:* Presenting and communicating the findings from the data analysis to the reader, or more appropriately, to the viewer, represents the third pillar of why data visualization is important. Mixing up confirmatory analysis with the presentation of the findings is probably one of the root causes for poor scientific communication. It is a common occurrence at scientific conferences that researchers use the same graphical tools to present their results to others as they used to obtain their findings in the first place. As pointed out by Schumann and Müller (2000, p. 6), this step requires careful thought that third parties are able to understand the findings without any unnecessary difficulties.

Maps and diagrams were already known in ancient Egypt but also communicating scientific results via visualization is at least 400 years old when Galileo Galilei (1613) and others published their observations of sunspots and other celestial bodies (Friendly 2008). But why is data visualization only becoming increasingly popular during the last 15–20 years? We argue that the key reason is the trend towards virtually ubiquitous access to electronic computing resources, enabling more and more people to participate in this endeavor. One could call it even a democratization of computing. In our opinion we can distinguish three key developments that played a crucial role since the 1980s and especially the 1990s. They are not listed in order of importance nor can they be considered in isolation from each other.

Hardware: The introduction of the predecessor of all modern PCs, the IBM personal computer, in 1981 as well as of microcomputers (e.g., the "C-64") in the same era triggered a shift away from the so-called minicomputers of the 1970s[2] to

---

[2]As noted at https://en.wikipedia.org/wiki/Minicomputer#cite_note-Smith_1970-4 (last accessed on 13 June 2017), the New York Times wrote in 1970 that minicomputers were computers that cost less than US-$ 25,000.

computers that could be purchased by households of average income. The speed of the processors was too slow and the size of computer memory was too small to process data as conveniently as we can nowadays, though. The first PC had an upper limit for working memory (RAM) of 256 kB, that is about 0.000778% of the first author's current desktop workstation. If we disregard developments in cache technology, parallel processing, etc., the pure clock speed of processors is now three orders of magnitude higher than in the early 1980s. Only 20 years ago, the typical size of total RAM was about as large as the size of a *single* digital photo today. But even if there was enough RAM and sufficient clock speed of the CPU, data storage was another limiting factor. The first hard disk with a capacity of more than one gigabyte was introduced in 1980 and cost at least US-$ 97,000.[3] One thousand times the storage capacity is available now at less than US-$ 100. This trend allowed the collection of massive data sets. To illustrate current capabilities: If we were interested in creating a data set, which contains about 1000 alphabetic characters (more than enough for the name, birth date and current residence) of any person alive, we would have to invest less than US-$ 400.[4] But, once again, even if we had the affordable computer storage of today, communicating results graphically was hindered by the low resolution combined with relatively few colors of early graphics standards such as CGA and EGA. Only with the introduction and the extension of the VGA standard, high resolution displays have become feasible.

Software: Having hardware in terms of processing speed, working memory and hard disk capacity to process graphics coincided with a revolution in software in the 1990s: Similar to the introduction of home computers that gave access to almost everyone, the emergence of *free software*, also called open source software, allowed anyone to use software without the costs and other restrictions often imposed by software products. Examples for this development can be found in the area of

- general programming languages (e.g., Python, Perl) as well as
- languages tailored or at least particularly suited for statistical programming and data analysis. The invention of the S language, started in the 1970s, was instrumental.[5] The most prominent example today is probably R (Ihaka 1998), but also other languages such as the now almost completely abandoned XLISP-STAT (de Leeuw 2005) facilitate(d) the visualization of data.[6]
- Lastly, in the area of efficient data storage, especially with the advent of "big data". Although it might be one of the most abused buzzwords currently, data

---

[3]See: https://www-03.ibm.com/ibm/history/exhibits/storage/storage_3380.html, last accessed on 13 June 2017.

[4]Assuming a world population of less than eight billion, a price for a 2TB hard disk of less than US-$S 100 and one byte per alphabetic letter.

[5]Please see Appendix A in Chambers (2008) for some notes on the history of S.

[6]It should be mentioned, though, that Matlab (Mathworks 2017), which is not published under a free/open-source license, was and is also key for the analysis and visualization of data.

sets in the gigabyte and terabyte range, partly in non-rectangular formats, have become ubiquitous. Those data can be handled by relational and non-relational database systems that are also available under free and open source licenses (e.g., SQLite, MySQL, Postgresql, Cassandra).

Connectivity: While the internet existed already for more than 20 years, the introduction and rising popularity of the world wide web (WWW) was a catalyst for the exchange of information via electronic networks. This technology allows now billions of people on earth to have almost instant access to data. The speed of the internet connection, which is crucial for the exchange of information such as downloading large data sets, has also increased by at least two orders of magnitude since the middle of the 1990s when 56 kbit/s modems were the standard.

Chapter 2
The Lexis Diagram

To look at 20,000 numbers and draw out their meaning
is a major research enterprise in itself.
Yet on the methods used in [Vaupel et al. (1985a)]
all that information is contained in a single contour map.

Nathan Keyfitz in his foreword of Vaupel et al. (1985a).

Any dynamics in vital events such as births and deaths involve change over calendar time, age, and/or cohort. The so-called Lexis diagram represents the ideal canvas to illustrate such dynamics. The Lexis Diagram as we use it today consists of a Cartesian coordinate system where calendar time ("period") is depicted on the *x*-axis and age on the *y*-axis (see Fig. 2.1 on page 6).[1] We added horizontal and vertical reference lines to facilitate orientation.

Birth cohorts move in such a diagram along the 45° line since a person is 1 year later 1 year older. Expressed differently: The current age of a person can be calculated if we subtract the birth date from the current calendar date. We used the example of three eminent demographers of the twentieth century in Fig. 2.1 to illustrate this relationship: William Brass, Ansley Coale, and Nathan Keyfitz. To be able to follow the cohorts on the 45° line, we made sure in Fig. 2.1—as well as in all other figures in this monograph—that the aspect ratio maps the length of one calendar year to exactly one age year.

Of course, we are not restricted to depict individuals on the Lexis plane. The standard approach is, indeed, to use population level data. It is obvious that we can not draw lines for every individual in that case. Colors are used instead to indicate the same value for the chosen statistic. While most figures in the remaining chapters show (smoothed) age-specific mortality or its time derivative, we opted to illustrate the basic approach of Lexis surface maps by depicting the population size of the

---

[1]It should be noted that the Lexis diagram can be considered to represent an example of "Stigler's law of eponymy" that states "No scientific discovery is named after its original discoverer." Please see Vandeschrick (2001) for a discussion about the problem of calling the diagram used in this book a "Lexis diagram".
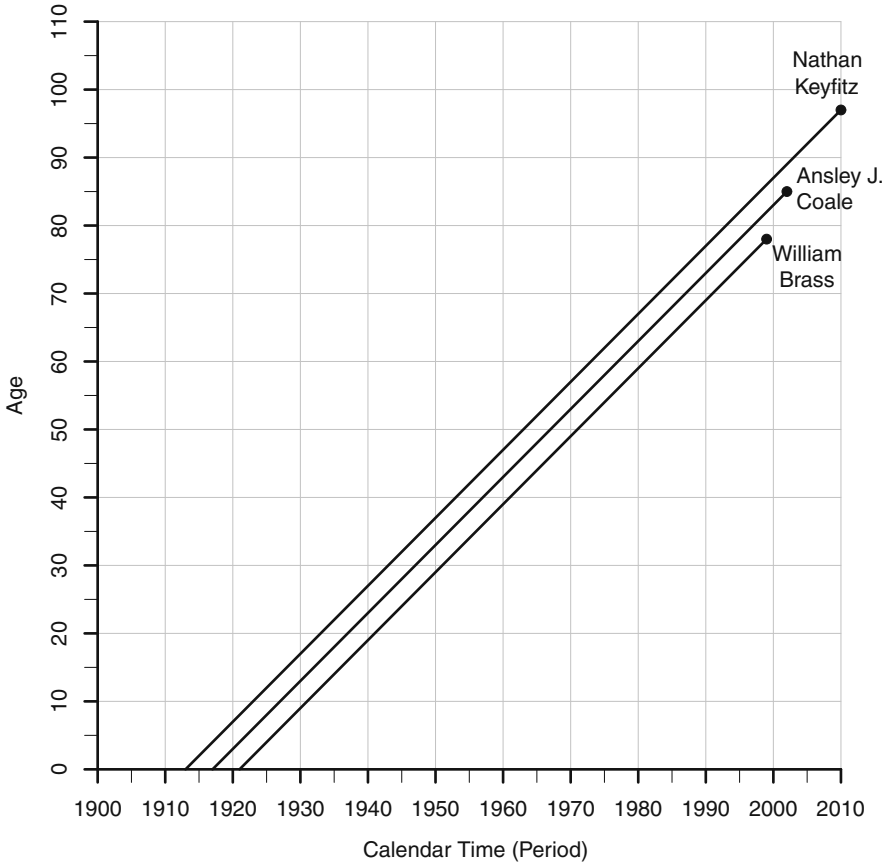
Fig. 2.1 An example of a Lexis diagram with individual life lines for William Brass, Ansley J.Coale, and Nathan Keyfitz

United States for women and men combined from 1900 until 2010 for ages 0–110 in Fig. 2.2. Thus, we have $111 \times 111 = 12{,}321$ individual datapoints. They are less than the 20,000 mentioned by Keyfitz in Vaupel et al. (1985a) but considerably more than the median number of entries in data matrices for statistical graphics found by Tufte (2003) in various scientific and non-scientific publications. Tufte—who was described as the "da Vinci of Data" by The New York Times (Deborah 1998)—states in a related book (Tufte 2001, p. 166): "Data graphics should often be based on large rather than small data matrices and have a high rather low data density. More information is better than less information, especially when the marginal costs of handling and interpreting additional information are low, as they are for most graphics."

In our Lexis maps we employed a color scheme reminiscent of geographic maps where green colors indicate lower values and brown colors are used for high "altitudes". Analogously to standard maps, we added contour lines to emphasize areas of equal elevation, which translates to the same number of people in our figure.
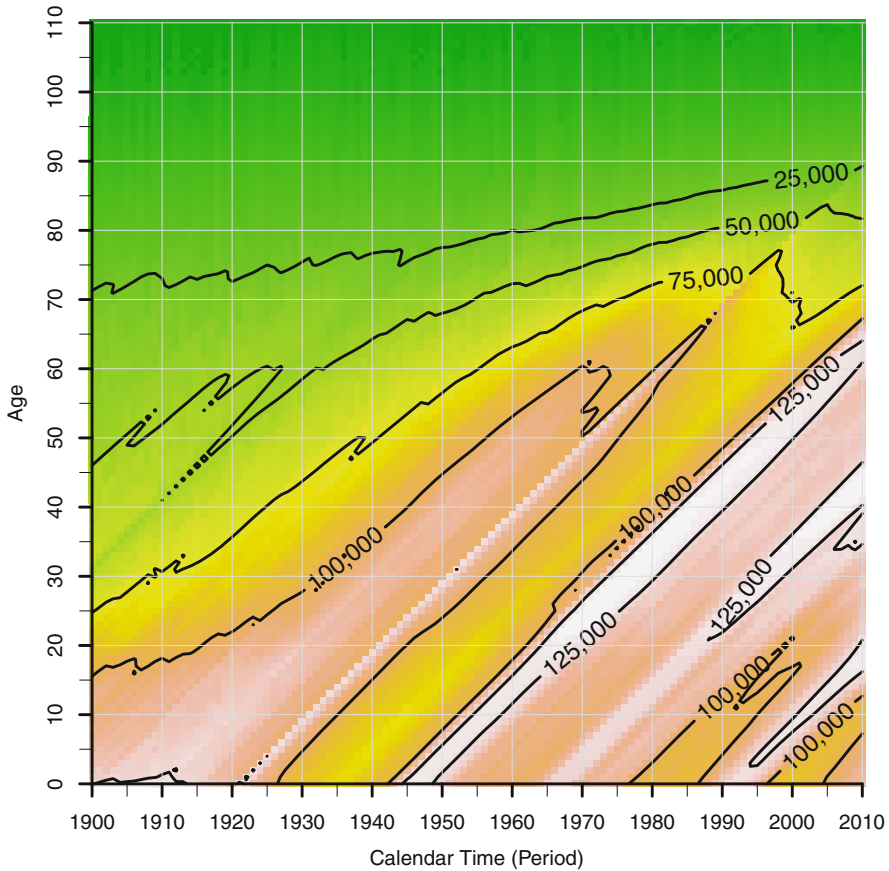
Fig. 2.2 An example of a Lexis surface depicting the population size of the United States by calendar year and age (Source: Own illustration based on data from the Human Mortality Database 2017)

Depicting mortality, fertility or other population characteristics in the Lexis diagram provides a useful framework to analyze data for the presence of age-, period-, and cohort- ("APC") effects. The major problem of standard statistical approaches (e.g., regression analysis) in this area is the so-called *Identification Problem*, which refers to the perfect correlation of age plus cohort equaling period. Various methods have been introduced (e.g., constraining the parameters in a regression setting) but "there is no magic solution" (Wilmoth 2006, p. 235).[2] With our surface maps, we suggest instead a graphical approach that can be used for questions such as "[w]hether mortality improvements takes place by cohorts or by periods" (Keyfitz in Vaupel et al. 1985a, p. ix).

---

[2]Please refer to this article also for a systematic overview of APC models used in demographic research.
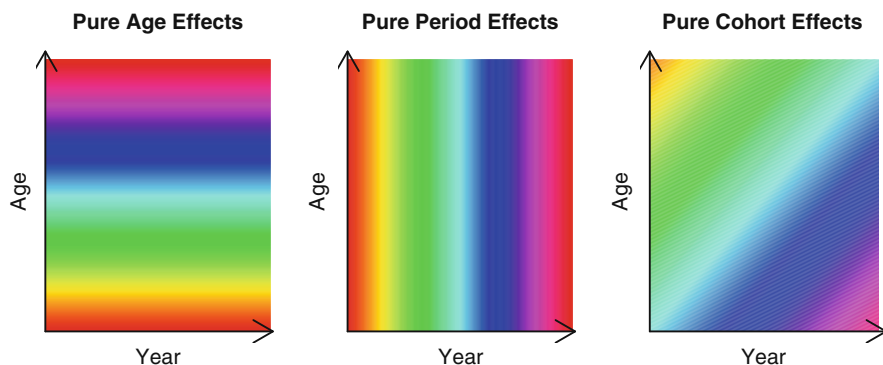
Fig. 2.3 "Ideal" age-, period-, and cohort-effects on the Lexis surface

Figure 2.3 gives an overview how age-, period-, and cohort effects would ideally look like on the Lexis surface. The same color indicates the same value in the variable of interest (e.g., death rates). The left panel represents "pure" age effects. That means that the only variation in the variable of interest takes place across the age dimension, regardless of calendar year or cohort. The panel in the middle denotes "pure" period effects, i.e., the same values are measured at all ages but they differ along the calendar time/period dimension ("Year"). Finally, the panel on the right illustrates how a surface map would like if (birth) cohorts alone were driving the development in the variable of interest. The same color along the 45° line shows that each cohort has their own characteristic value of the variable of interest, which does not change throughout their life course. Obviously, those are idealized and simplified representations. We expect to find rather interactions of these three forces than such "pure" effects. Furthermore, we should acknowledge the biggest drawback of our method: In contrast to other methods of APC analysis, our visual approach does not attribute any numerical value to each of those effects. Hence, one can neither compare various effects with each other nor is it possible to conduct significance tests that are typical of regression analyses and other standard methods in statistics.

We are not the first to illustrate demographic phenomena in three dimensions, i.e., *either* on the Lexis plane using colors to indicate the third dimension *or* by wireframe plots. An interesting overview of the history of such "Frequency Surfaces and Isofrequency Lines" is given in Caselli and Vallin (2006). They cite the example of Luigi Perozzo's depiction of the change in the Swedish age pyramid in 1880, based on a diagram by Gerard Van Den Berg (1860), as one of those earliest examples. We have reproduced Perozzo's diagram in Fig. 2.4. About 60 years later, Pierre Delaporte used such wireframes to depict French mortality (1938) and contour lines for European mortality (1942).

An explicit case of using such plots to separate age-, period-, and cohort-effects from each other can be found in Thomas Pullum's article on US fertility published in 1980. A few years later, the population program at the International Institute
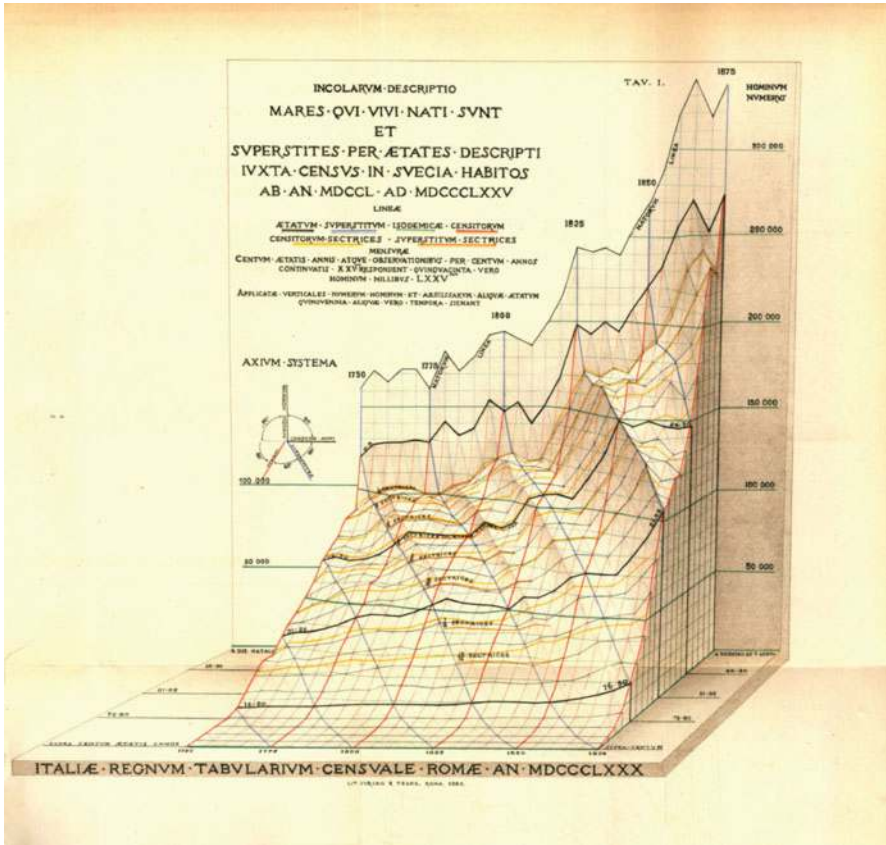
Fig. 2.4 Change in the Swedish age pyramid as depicted by Luigi Perozzo in 1880 (Source: Timothy Riffe, with kind permission)

for Applied Systems Analysis (IIASA) in Laxenburg in Austria turned out to be an incubator for advancing the display of population dynamics on the Lexis plane in the 1980s. Vaupel, Yashin, Caselli, and others introduced colored/shaded contour maps to depict, for example, population size, mortality, or birth rates (e.g., Vaupel et al. 1985a,b, 1987; Caselli et al. 1985; Gambill and Vaupel 1985). The "democratization" effort described in the introductory chapter was also mirrored in the late 1990s for Lexis surfaces: Kirill Andreev developed not only the user-friendly software Lexis to analyze demographic trends in Denmark and other highly developed countries (Vaupel et al. 1997; Andreev 2002). He also shared it freely with anyone interested.[3] Despite being a milestone for the creation of Lexis surface

---

[3]While writing his Master's thesis, the first author of this monograph received the Lexis software from Kirill Andreev simply via email in early 2000.

maps, almost no one is using it anymore. The aforementioned specialized languages such as Matlab (Mathworks 2017) or R (R Development Core Team 2015) have become the favorite tools nowadays along with Python (van Rossum 1995). With the exception of the reproduction of Perozzo's plot all figures in this monograph were created with R as we will explain in Sect. 3.2 and in the appendix, starting on page 161.

Chapter 3
Data and Software

3.1 Data

3.1.1 Human Mortality Database

Most of our analyses are based on data from the Human Mortality Database
("HMD", 2017), which can be freely accessed after registration at http://www.
mortality.org. The database is a collaborative project of research teams from the
Department of Demography at the University of California, Berkeley (USA) and the
Max Planck Institute for Demographic Research in Rostock (Germany). It contains
aggregate mortality statistics such as death counts, population estimates, exposure
to risk estimates, life tables as well as some other statistics of more than 35 countries
(see Table 3.1). Further distinctions into sub populations are possible for some
countries such as Germany (East and West Germany), the United Kingdom (England
and Wales, Northern Ireland, Scotland) or New Zealand (Maori, Non-Maori). The
database has its focus on highly developed countries.

Since its launch in 2002, the HMD has become the gold standard for the aggre-
gate level (demographic) analysis of mortality. Apart from the diligent collection of
data, its widespread adoption can mainly be attributed to two reasons: (1) Rigorous
quality checks are conducted before new data are added to the database. (2) The
biggest asset of the HMD is that it does not simply publish processed data. Instead,
the HMD estimates life tables and other statistics itself using raw data, applying
the same set of methods. Thus, any differences over time or across region can not
be attributed to different methodologies, for instance, how the life table was closed
(HMD 2007).

As some life tables in the HMD are smoothed at ages 80 and higher, we did
not rely on life tables estimates at all but used exclusively the death counts and
the corresponding exposures from the HMD on a 1-calendar-year by 1-age-year
grid to estimate death rates. Most of our analyses deal with mortality developments

R. Rau et al., *Visualizing Mortality Dynamics in the Lexis Diagram*,
The Springer Series on Demographic Methods and Population Analysis 44,
DOI 10.1007/978-3-319-64820-0_3

Table 3.1 Countries covered in the Human Mortality Database and data coverage after 1950 on January 10th, 2017, when the most recent update of data was conducted for the present monograph

Country	Deaths	Data Coverage
Australia	6,956,698	
Austria	5,575,409	
Belarus	5,739,008	
Belgium	7,216,189	
Bulgaria	5,663,709	
Canada	11,056,710	
Chile	1,121,837	
Czech Republic	7,366,321	
Denmark	3,385,967	
Estonia	933,627	
Finland	3,025,080	
France	34,953,421	
Germany	20,660,306	
Germany−East	11,839,606	
Germany−West	40,080,602	
Greece	3,314,273	
Hungary	8,264,444	
Iceland	100,400	
Ireland	2,107,181	
Israel	1,132,519	
Italy	33,542,107	
Japan	54,561,300	
Latvia	1,667,906	
Lithuania	1,953,998	
Luxembourg	217,818	
Netherlands	7,242,595	
New Zealand	1,610,320	
New Zealand: Maori	94,938	
New Zealand: Non−Maori	1,368,240	
Norway	2,563,597	
Poland	18,949,583	
Portugal	6,296,406	
Russia	90,757,552	
Slovakia	2,988,888	
Slovenia	610,127	
Spain	20,721,634	
Sweden	5,590,479	
Switzerland	3,773,553	
Taiwan	4,961,642	
UK	40,029,614	
UK, England & Wales	35,120,479	
UK, Northern Ireland	995,217	
UK, Scotland	3,913,918	
Ukraine	31,897,144	
USA	133,314,965	

1950   1960   1970   1980   1990   2000   2010

since 1950. We selected this threshold year because of the availability of more data compared to earlier time periods. Furthermore, it also marks the beginning of a new era: Most gains in life expectancy are nowadays due to survival improvements among the elderly (Christensen et al. 2009), a development, which was virtually non-existent before the middle of the twentieth century. Kannisto (1994), for instance, estimated that the onset of sustained decline in old-age mortality occurred for women in Switzerland, Belgium and Sweden in 1956.

As shown in Table 3.1 total deaths range from barely 100,000 (Iceland) to more than 130 million in the United States. We analyzed all countries; the only exceptions are Chile and the Maori population of New Zealand due to problematic data quality (Jdanov et al. 2008) and the low number of years covered (Chile). Nevertheless, we did not include those figures for all countries and both sexes as it would have resulted in a monograph consisting of hundreds of additional pages. We typically restricted ourselves, instead, to a few examples that feature interesting characteristics.

3.1.2 Cause-Specific Death Counts in the United States

The National Center for Health Statistics of the United States provides a unique collection: Individual death counts by sex, age at death, year of death, cause of death, and many more characteristics can be freely downloaded from its web page. The data are available since 1968 in annual files. Additionally, the website of the National Bureau of Economic Research (NBER) provides data since 1959, which we used in our analyses. The last year in our analysis is 2014. With the exception of 1972, when only a 50% sample was taken, each file contains all deaths in the United States. In the analysis by cause of death in later chapters of this volume, we simply multiplied the number of deaths for a given age, sex, and cause in the year 1972 by a factor of 2.

Causes of death are coded by the so-called "International Classification of Diseases" (ICD). Since its introduction in the late nineteenth century, the system has been revised at irregular intervals (Meslé 2006). The tenth revision is currently used. During the first years of our analysis, ICD-7 was used. ICD-8 was in effect in the United States between 1968 and 1978, followed by ICD-9 from 1979 until 1998.

Obtaining consistent time series of causes of death across ICD revisions requires meticulous work and care (e.g., Meslé and Vallin 1996; Pechholdová 2009). We therefore decided to use only very broad categories for causes of death and followed primarily the coding of Janssen et al. (2003) and of Meslé and Vallin (2006a). Both papers include an appendix with detailed ICD codes across the four revisions required in our analysis.

Table 3.2 is split into two halves. The upper panel provides the ICD codes we used to extract the causes of death, whereas the lower panel lists the number of deaths in absolute and relative terms for the selected causes by sex.

Our database consists of more than 118 million deaths. Although we have selected very few causes, they account for about three quarters of all deaths (Category 13 "Other" is 23.75%). A bit more than 44% of all deaths classified as originating from circulatory diseases. In that category, heart diseases are about one third of all deaths for women and men alike. The almost 10 million deaths from cerebrovascular diseases between 1959 and 2014 represent about eight percent of all deaths. The most common cerebrovascular disease is stroke. Malignant neoplasms ("cancer") are the second largest chapter in the ICD. Regardless of sex of the decedent, about one in every fifth death belongs to that category. We

Table 3.2 ICD codes and counts (absolute and relative) for females, males, and both sexes combined selected causes of death, 1959–2014

Nr.	Cause	ICD codes			
		ICD-7	ICD-8	ICD-9	ICD-10
	Years in use:	1959–1967	1968–1978	1979–1998	1999–2014
(1)	All causes	—	—	—	—
(2)	Circulatory dis.	300–334, 400–468	390–458	390–459	I00–I99
(3)	Heart	400–447	390–429	390–429	I00–I52
(4)	Cerebrovasc.	300–334,	430–434, 436–438	430–434, 436–438	I60–I69
(5)	Other	All (2) not in (3) or (4)			
(6)	Cancers	140–239	140–239	140–239	C00–D48
(7)	Breast	170	174	174, 175	C50
(8)	Lung	162, 163	162	162	C33, C34
(9)	Colorectum	153, 154	153, 154	153, 154	C18–C21
(10)	Other	All (6) not in (7), (8), or (9)			
(11)	Resp. diseases	470–527	460–519	460–519	J00–J99
(12)	Motor vehicle acc.	E810–E825	E810–E819	E810–E819	V00–V89
(13)	Other	All (1) not in (2)–(12)			

		Number of cases					
		Total		Female		Male	
Nr.	Cause	Counts	%	Counts	%	Counts	%
(1)	All causes	118,678,283	(100.00)	56,432,184	(100.00)	62,246,099	(100.00)
(2)	Circulatory dis.	52,668,448	(44.38)	25,985,900	(46.05)	26,682,548	(42.87)
(3)	Heart	40,342,012	(33.99)	19,072,073	(33.80)	21,269,939	(34.17)
(4)	Cerebrovasc.	9,381,071	(7.90)	5,430,076	(9.62)	3,950,995	(6.35)
(5)	Other	2,945,365	(2.48)	1,483,751	(2.63)	1,461,614	(2.35)
(6)	Cancers	25,722,893	(21.67)	12,096,049	(21.43)	13,626,844	(21.89)
(7)	Breast	2,067,878	(1.74)	2,050,192	(3.63)	17,686	(0.03)
(8)	Lung	6,393,007	(5.39)	2,260,023	(4.00)	4,132,984	(6.64)
(9)	Colorectum	2,884,519	(2.43)	1,458,772	(2.59)	1,425,747	(2.29)
(10)	Other	14,377,489	(12.11)	6,327,062	(11.21)	8,050,427	(12.93)
(11)	Resp. diseases	9,566,798	(8.06)	4,457,141	(7.90)	5,109,657	(8.21)
(12)	Motor vehicle acc.	2,538,449	(2.14)	742,599	(1.32)	1,795,850	(2.89)
(13)	Other	28,181,695	(23.75)	13,150,495	(23.30)	15,031,200	(24.15)

selected three prominent cancer sites: Breast, lung and colorectum. Please note that while there are many more deaths from breast cancer for women, also more than 17,000 men died from it during the 56 years of our observation period. Respiratory diseases are with approximately 8% of all deaths slightly more common than cerebrovascular diseases. Although it is not a major cause of death (2%), we also

included information about motor vehicle accidents since it turned out to be an interesting case study for seasonality in deaths, which we analyze in Chap. 9.

3.1.3   SEER Cancer Register Data 1973–2011

The Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute of the United States allows researchers access to longitudinal data on the individual level about the incidence of cancer and includes also information about the survival of patients. The data coverage—the SEER data start in 1973—and the large size of data, combined with the ease of access, make the SEER data an ideal instrument for the analysis of cancer survival by age over calendar time. We were using data that were released in April 2014 with a follow-up cutoff date of December 31, 2011 (Surveillance, Epidemiology, and End Results (SEER) Program 2014). The SEER data do *not* cover all cancer diagnoses of the United States. It is a collection of data from several registries. With the exception of Seattle (Puget Sound) and Metropolitan Atlanta that started in 1974 and 1975, respectively, we only used registers that covered the whole time span from 1973 until the end of 2011. Although we use less data than we could have, we thought that a heterogeneous set of registers would have induced problems for the analysis over time. The registers included in our analysis were: San Francisco-Oakland SMSA, Connecticut, Metropolitan Detroit, Hawaii, Iowa, New Mexico, Utah as well as Seattle and Metropolitan Atlanta.

   In our analysis of cancer survival in Chap. 10, starting on page 123, we selected five cancer sites: Breast cancer; cancer of the lung and bronchus; cancer of the colon, rectum, and anus; pancreatic cancer; prostate cancer. As shown in Table 3.3, those five cancer sites constitute about 55% of all cancer diagnoses for women as well as for men out of the 4.5 million cases recorded during our observation period. The largest categories are by far breast cancer for women (30.44%) and prostate cancer for men (25.79%). The absolute and relative frequencies of the other cancer sites as well as their respective ICD codes can be inspected from Table 3.3. While ICD-8 was in use at the beginning of the observation period in 1973 and cancer cases are typically coded by the ICD-O standard, all ICD codes were converted to ICD-10 by SEER.

3.2   Software

All analyses have been conducted and all figures have been produced using R (Version 3.2.3), a free software environment for statistical computing and graphics (R Development Core Team 2015). The surface maps were created by the `image()` function and contour lines were added with the `contour()` function. To facilitate the creation of surface maps of rates of mortality improvement for other researchers,

Table 3.3 ICD-10 codes and incidence counts (absolute and relative) by cancer site of females, males, and both sexes combined in the SEER Data, 1973–2011

Cancer site		ICD-10 Code	Incidence Total	
			Counts	in %
(1)	All	C00–D48	4,524,099	(100.00)
(2)	Breast	C50	713,376	(15.77)
(3)	Bronchus and lung	C34	557,901	(12.33)
(4)	Colon, rectum, and anus	C18-C21	520,456	(11.50)
(5)	Pancreas	C25	103,152	(2.28)
(6)	Prostate	C61	566,311	(12.52)
(7)	Rest	All (1) not in (2)–(6)	2,062,903	(45.60)

Cancer site		Incidence Female		Male	
		Counts	in %	Counts	in %
(1)	All	2,328,116	(100.00)	2,195,983	(100.00)
(2)	Breast	708,696	(30.44)	4,680	(0.21)
(3)	Bronchus and lung	224,927	(9.66)	332,974	(15.16)
(4)	Colon, rectum, and anus	257,406	(11.06)	263,050	(11.98)
(5)	Pancreas	51,712	(2.22)	51,440	(2.34)
(6)	Prostate	N/A	(N/A)	566,311	(25.79)
(7)	Rest	1,085,375	(46.62)	977,528	(44.51)

an R package called `ROMIplot` has been created and uploaded to CRAN, the general archive of R packages. Installation and usage of this package are explained in Appendix "Software: R package `ROMIplot`" (p. 161).

Chapter 4
Surface Plots of Observed Death Rates

4.1 From Death Counts to Death Rates

The basic units of any mortality analysis are death counts. In most scientific disciplines those counts are expressed as rates by dividing them by a unit of time. Examples are heart rates counting beats per minute or becquerel measuring the radioactive decay of nuclei per second. Things are more complicated when death counts are analyzed: For instance, 30,140 people died at age 80 in Germany in 2000. The corresponding number of Austria is 2,765 (HMD, 2017). Inferring that the risk of dying is more than ten-fold higher in Germany than in Austria is obviously wrong. Death rates are—as all demographic rates—therefore standardized dividing the counts by the corresponding number of life-years lived (see, for example Chap. 1.4 in Preston et al. 2001). The latter are often called "exposures" and are typically approximated by an estimate of the mid-year population. In the example above, the death rates at age $x = 80$ in year $t = 2000$, usually denoted as $m(x, t)$ would correspond to:

$$\text{Austria}: \quad m(x, t) = \frac{D(x, t)}{N(x, t)} = \frac{2765}{42{,}070.77} = 0.06572259$$

$$\text{Germany}: \quad m(x, t) = \frac{D(x, t)}{N(x, t)} = \frac{30140}{444{,}400.81} = 0.06782166$$

with death counts and exposures denoted as $D(x, t)$ and $N(x, t)$, respectively. Hence, mortality is still higher in Germany than in Austria but only by about three per cent and not by an order of magnitude. Death rates at those single ages x, that are used

exclusively in this book, are often a good approximation for the continuous force of mortality at the middle of that age $\mu(x + 0.5)$ (Thatcher et al. 1998). Nearly all of the analyses contained in this volume are based on such death rates.

4.2   Results

The raw surface plots on the following pages depict the observed death rates for women and men in a few selected countries. Death rates were estimated for single ages and single years from 1950 until the last available year in the Human Mortality Database, in most cases 2014 (see Chap. 3). Our color scheme ranges from blue to green to red. To facilitate interpreting the plots, we added contour lines for various levels of mortality similar to the ones for elevation on topographic maps. The levels of 1 death per 10 person-years lived, per 100 person-years lived, per 1,000 person-years lived, and per 10,000 person-years lived have been printed as bold lines as visual cues not because of any implicit distinct meaning apart from the digit preference.

Generally speaking, we do not think that raw surface plots are the best option to visualize mortality dynamics. That is why we only depict a few countries here. One of the main problems is that the observed rates suffer from random fluctuations. At young ages because death rates are so low; at older ages because there are so few people left. Thus, the numerator for the observed death rates is relatively small in the first case whereas the denominator is relatively small in the latter case.

What we can observe for Australian women and men in Figs. 4.1 and 4.2 is representative for many countries in the Human Mortality Database[1]: Most contour lines tend to move upwards over time. This indicates that the same level of mortality is being observed at higher and higher ages. Or, expressed differently, mortality is continuously decreasing at almost any given age. Switzerland and Spain in Figs. 4.3, 4.4, 4.5 and 4.6 are further examples of this general trend. It seems to be noteworthy that the late 1990s seems to be an important era for major improvements in mortality among young males.

We can already observe here the unfortunate mortality developments that took place in Russia (Figs. 4.7–4.8) as well as in many other eastern European countries

---

[1]See Figs. A.1–A.6 in the appendix for corresponding plots for France, England and Wales, and Norway.
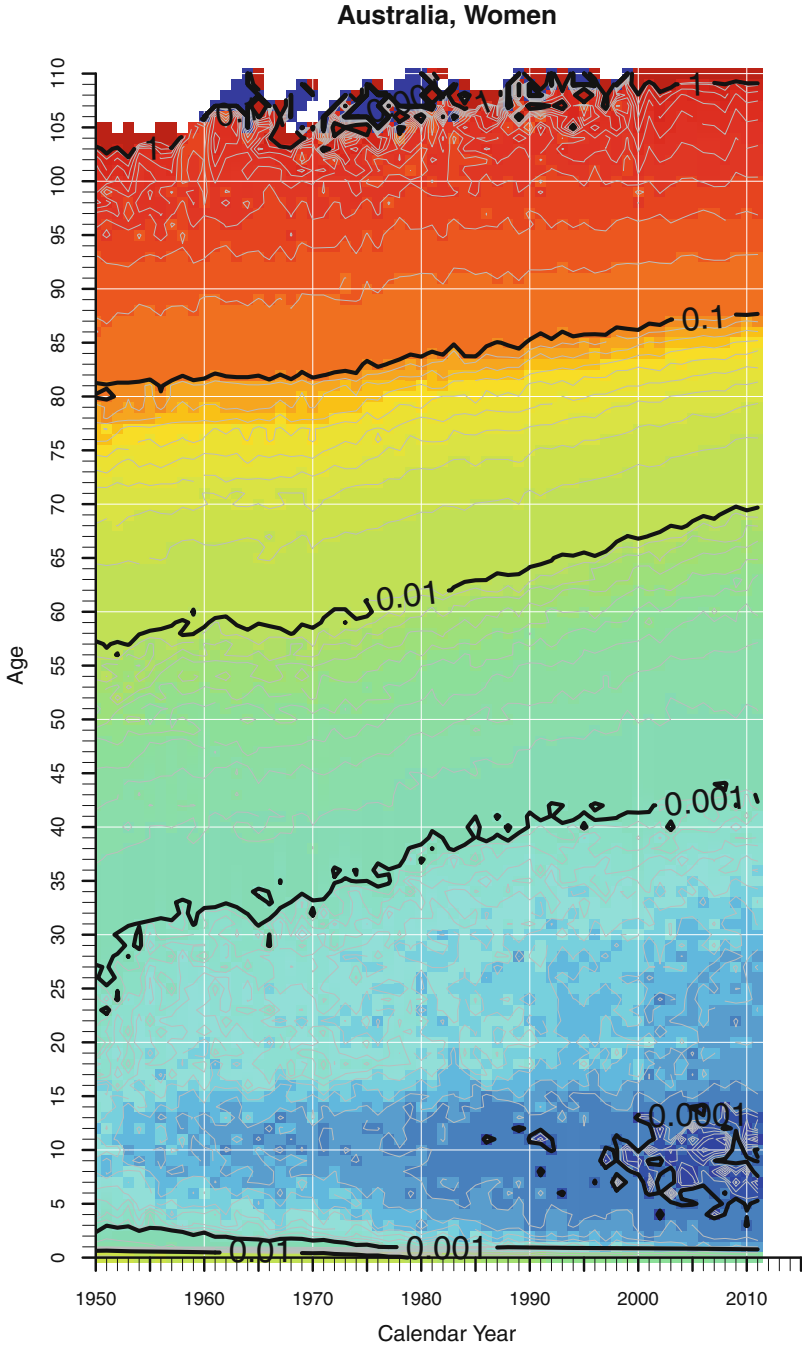
Fig. 4.1 "Raw" death rates for women in Australia, 1950–2011 (Data source: Human Mortality Database)
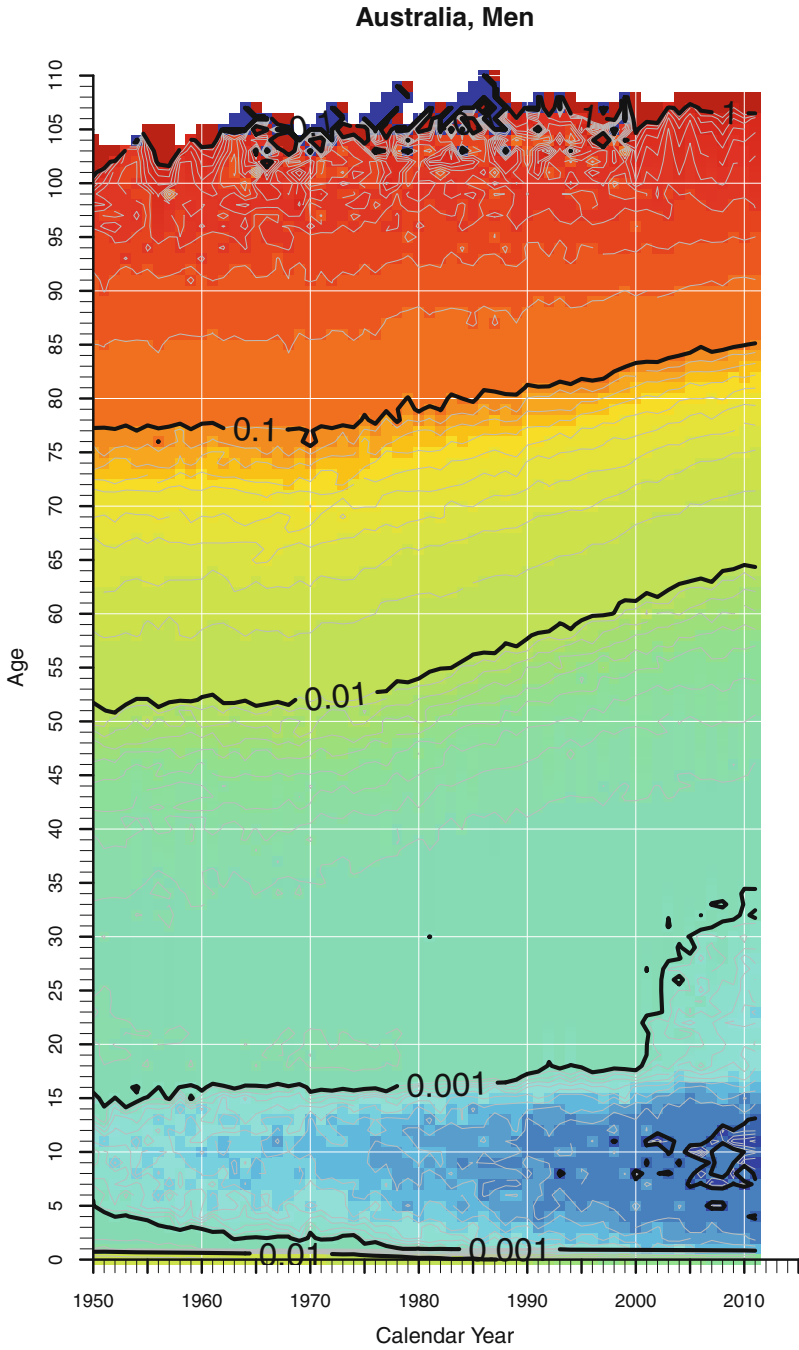
Australia, Men

Fig. 4.2 "Raw" death rates for men in Australia, 1950–2011 (Data source: Human Mortality Database)

Spain, Women
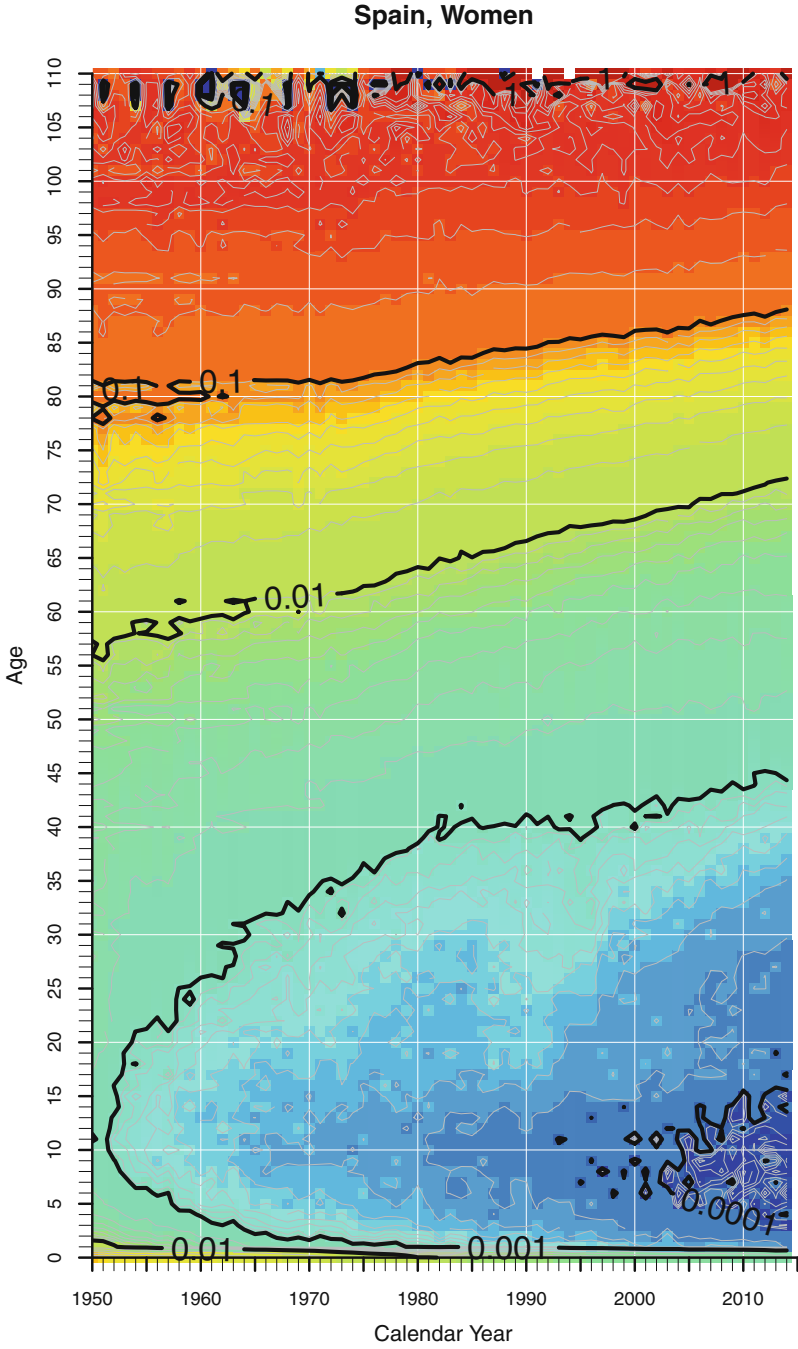
Fig. 4.3 "Raw" death rates for women in Spain, 1950–2014 (Data source: Human Mortality Database)

Spain, Men
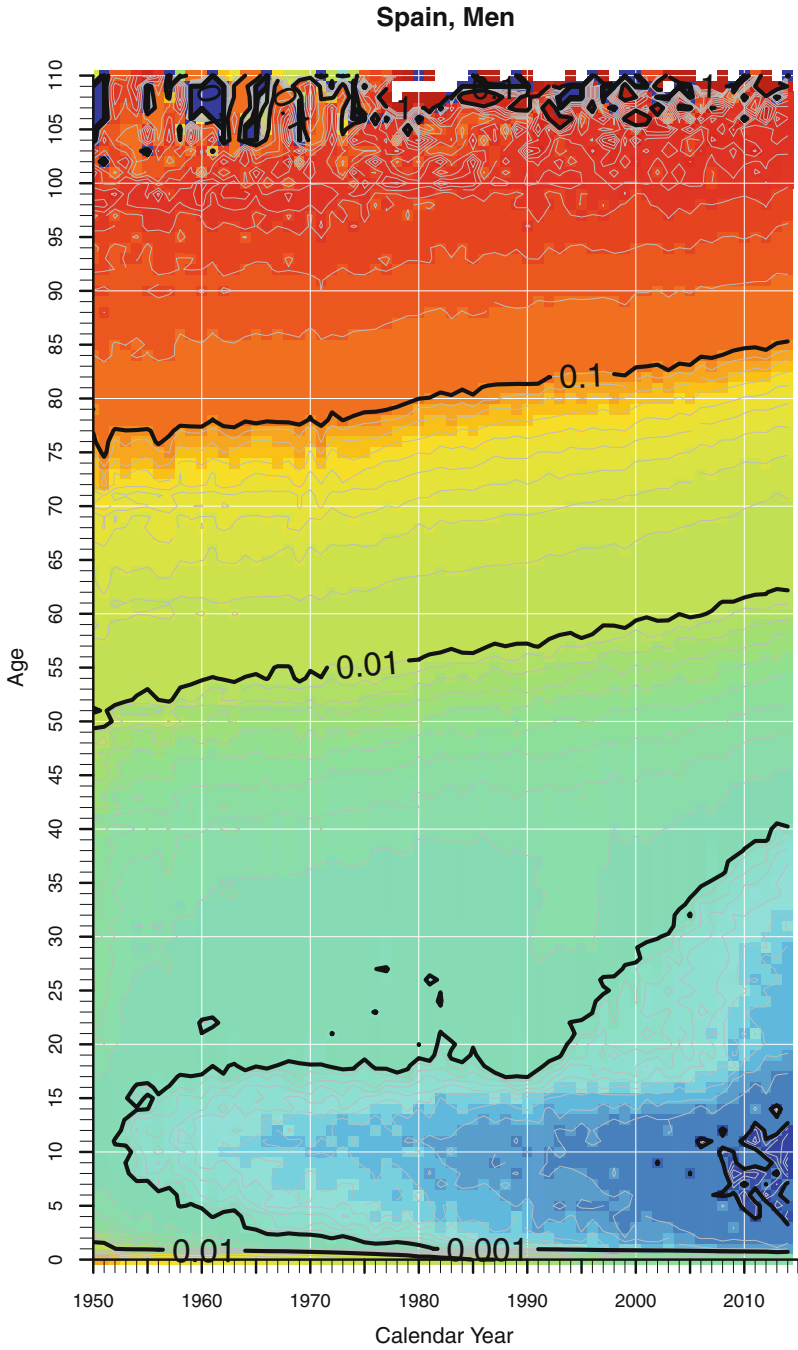
Fig. 4.4 "Raw" death rates for men in Spain, 1950–2014 (Data source: Human Mortality Database)
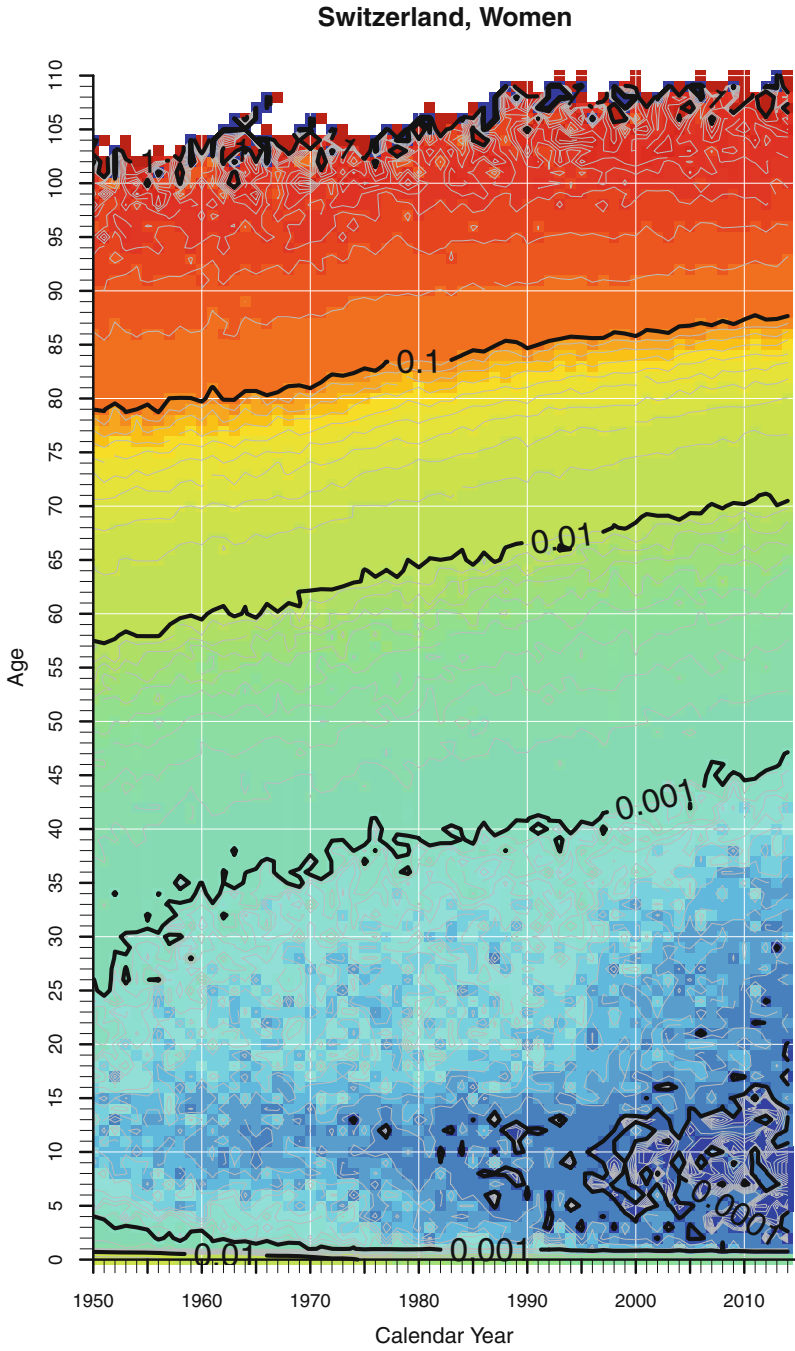
Fig. 4.5 "Raw" death rates for women in Switzerland, 1950–2014 (Data source: Human Mortality Database)
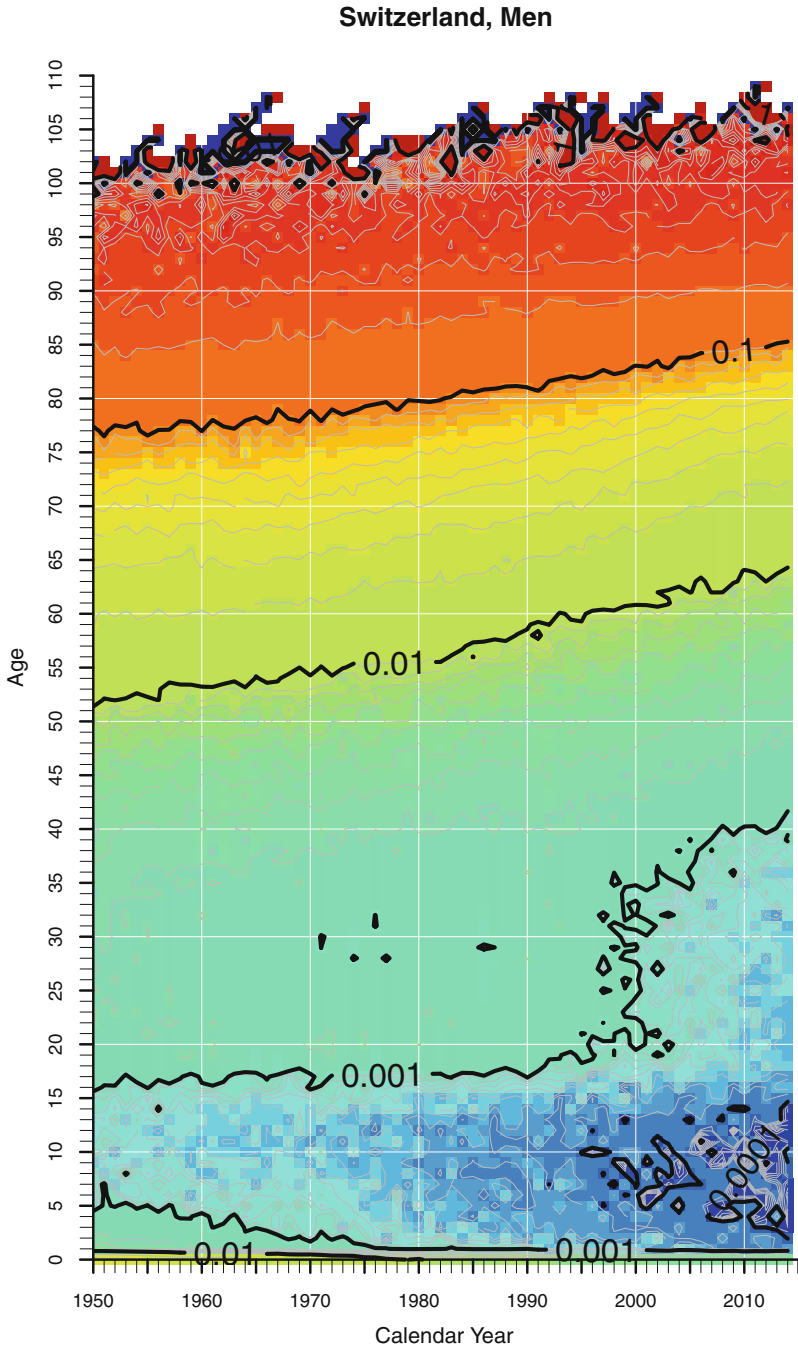
Fig. 4.6 "Raw" death rates for men in Switzerland, 1950–2014 (Data source: Human Mortality Database)
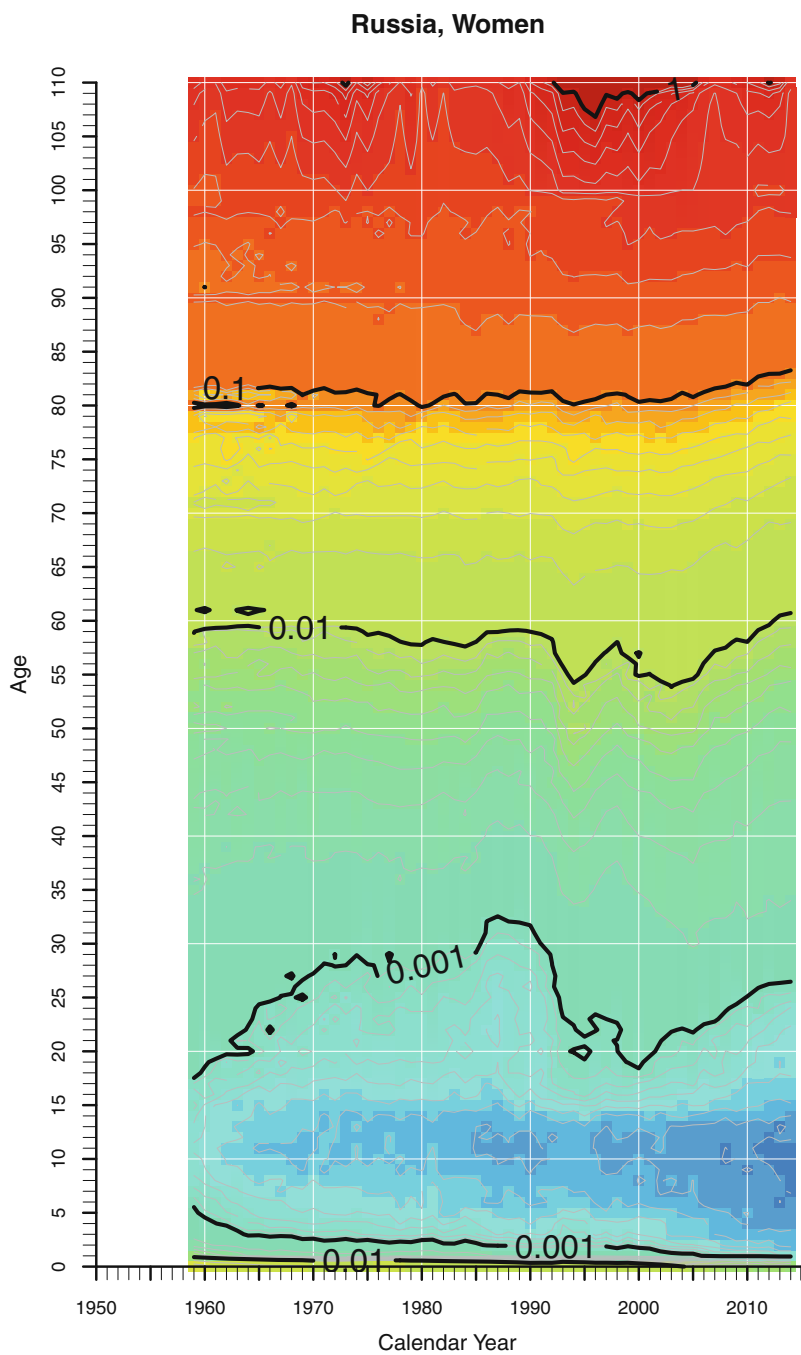
Russia, Women

Fig. 4.7 "Raw" death rates for women in Russia, 1959–2014 (Data source: Human Mortality Database)

Russia, Men
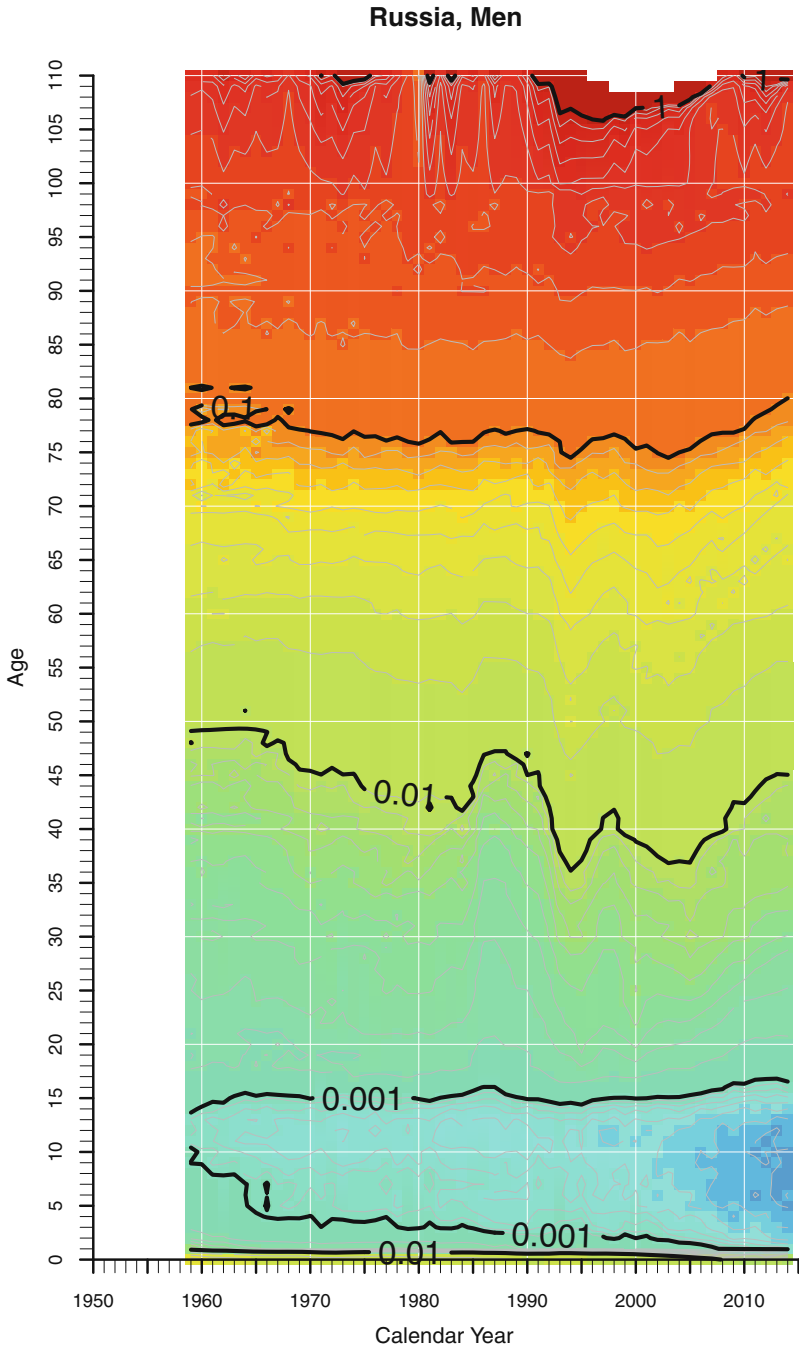
Fig. 4.8 "Raw" death rates for men in Russia, 1959–2014 (Data source: Human Mortality Database)

(not shown here) that have been distinct from the rest of Europe: Irregular trends, especially among males, and even increasing mortality as depicted by the downward contour lines have been rather the rule than the exception between the 1960s and the early 2000s.

Chapter 5
Surface Plots of Smoothed Mortality Data

5.1   From Raw Death Rates to Smoothed Death Rates

We have seen in the previous chapter, that "raw" death rates can suffer from considerable random fluctuations. Assuming that data quality is not an issue, this noise can be caused by (1) very few numbers of deaths (numerator), by (2) very few persons exposed to the risk of dying (denominator) or by (3) small populations in general. Problem (1) typically occurs at young ages. We selected age 15 in France in Panel (a) of Fig. 5.1. Despite a large population in general, deaths occur—thankfully—relatively rarely at that age. (2) The opposite is true at advanced ages as shown in the middle panel of the same figure. Very few people are still alive at age 95 in Italy, although it is a large population having relatively high life expectancy. Problems (1) and (2) occur in countries with tens of millions of people only at young and old ages. The smaller the population size, the more ages are affected. Panel (c) illustrates issue (3) using Danish data. The mortality trajectory in highly developed countries is rather smooth around age 80. In countries with just a few millions of people, considerable random fluctuations can be even observed there. Please note that more than five million people live in Denmark. Hence, the challenge becomes even bigger in smaller countries such as the Baltic states, Luxembourg or, especially, in Iceland.

We decided therefore to smooth the data. Myriads of methods exist to smooth data. While the pattern over age can be appropriately captured by parametric models, the trajectory over time differs considerably between ages and countries. Our decision was therefore to use a non-parametric smoothing approach. We selected the so-called *P*-spline approach, originally developed by Eilers and Marx (1996), adapted to the analysis of mortality by Currie et al. (2004) and further refined by Camarda (2008). The author, Carlo Giovanni Camarda, also provides the R extension package "MortalitySmooth" (Camarda 2012), which makes it easy and straightforward to apply the method. At its core, the model assumes
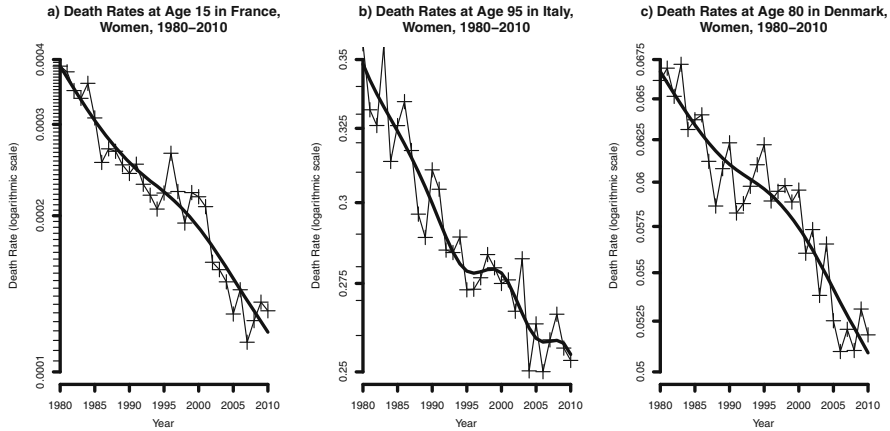
Fig. 5.1 The necessity to smooth raw death rates. Using data for France, Italy, and Denmark, panel (**a**), (**b**) and (**c**) illustrate three sources of random fluctuations: few numbers in the numerator (Panel (**a**) for age 15), few numbers in the denominator (Panel (**b**) for age 95) or small population sizes in general (Panel (**c**) for age 80) (Data source: Human Mortality Database)

Poisson distributed death counts with the (log-)exposures as an offset to account for changing population sizes over time and/or age. The method uses *B*-splines as regression bases. Whereas the number and position of the basis functions is crucial for standard smoothing with *B*-splines, the *P*-spline approach uses "too many" bases, which would normally result in overfitting. The *P* in the name of the method refers to the penalization of adjacent regression coefficients that differ too much from each other. Further technical details about the basis functions, the order of the differences, the penalty term λ, etc. are extensively discussed in the aforementioned references. The bold solid black lines in each panel of Fig. 5.1 depict the data smoothed with *P*-splines for the three given ages over time. One can easily recognize that the selected smoothing method is flexible enough to model irregular developments but is not prone to overfit the data.

The univariate time series of Fig. 5.1 is synthetic. Only cartoon characters such as Bart Simpson or Eric Cartman can retain their age over time. In reality, each individual is 1 year later 1 year older. Therefore we smoothed the data simultaneously over age and time using the function `Mort2Dsmooth` of Camarda's package "MortalitySmooth" (2012).

Raw death rates for Estonian women aged 60–80 years from 1980 to 2000 are illustrated in the left panel of Fig. 5.2 as a three-dimensional mortality surface. The general shape of increasing mortality over age can easily be observed. The right panel, featuring smoothed data, also shows the decline in mortality at higher ages over time, which is difficult to track down in the presence of noise in the data.
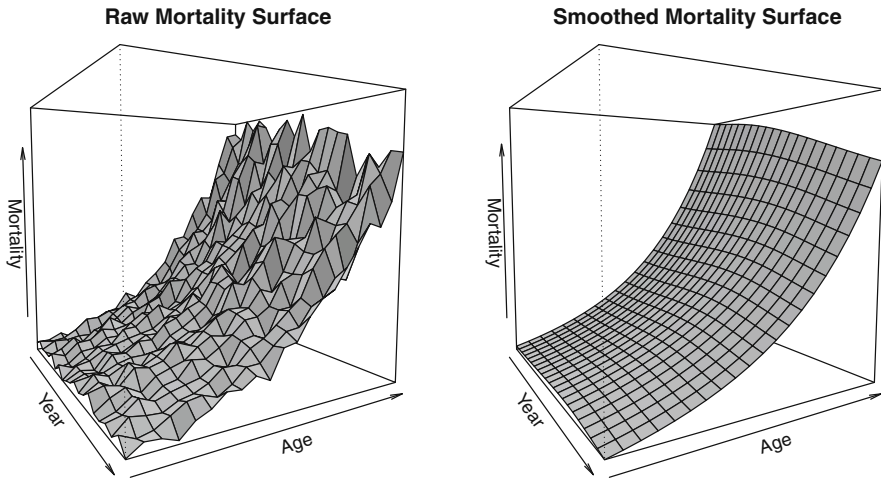
Fig. 5.2 3D plot of raw and smoothed death rates of Estonian women aged 60–80 years in 1980–2000 (Data source: Human Mortality Database)

The selected three-dimensional perspective plot appears appealing at first sight. The choice of angle and elevation is somehow arbitrary, though, and allows to accentuate certain features and suppress others. Since we often want to use the mortality surface for exploratory purposes, we have to give equal exposure to each unit. Therefore, we projected the three-dimensional data on the two-dimensional Lexis-plane, denoting the level of mortality by different colors (see Fig. 5.3 as an example).

Comparable to topographic maps, we added contour lines to depict the same levels of mortality. The general upward tendency of the contour lines indicate that the same level of mortality is shifting to higher and higher ages. Thus, for a given age mortality is decreasing, resulting in an increase in life expectancy.

5.2 Results

Figures 5.4, 5.5, 5.6, 5.7, 5.8, 5.9, 5.10, and 5.11 depict the same set of countries as Figs. 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, and 4.8 in Chap. 4 for a proper comparison between "raw" rates and smoothed rates.[1] The smoothed surface maps make the major trends in the data more pronounced such as almost parallel straight upward

---

[1]The appendix contains therefore also maps of smoothed death rates for France, England & Wales, and Norway. They can be found in Figs. A.7, A.8, A.9, A.10, A.11, and A.12.
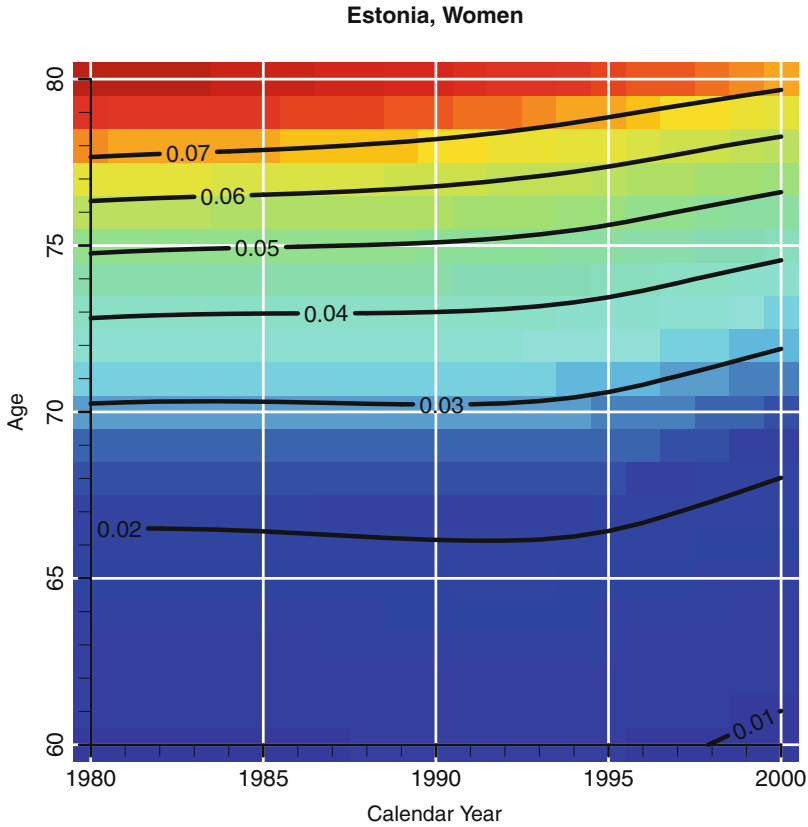
Estonia, Women

Fig. 5.3 Death rates of Estonian women aged 60–80 years in 1980–2000 as an example of smoothed death rates on the Lexis plane (Data source: Human Mortality Database)

lines in Australia, Spain, and Switzerland or the sudden survival improvements in survival among young Spanish men, starting in about 1990. Also large random fluctuations due to very few deaths as we have seen in the plot of raw death rates among children in Switzerland (Figs. 4.5 and 4.6) are removed by the smoothing procedure. While smoothing intrinsically involves some dampening of sudden changes in trends, the automatic procedure to find the optimal penalizing λs still
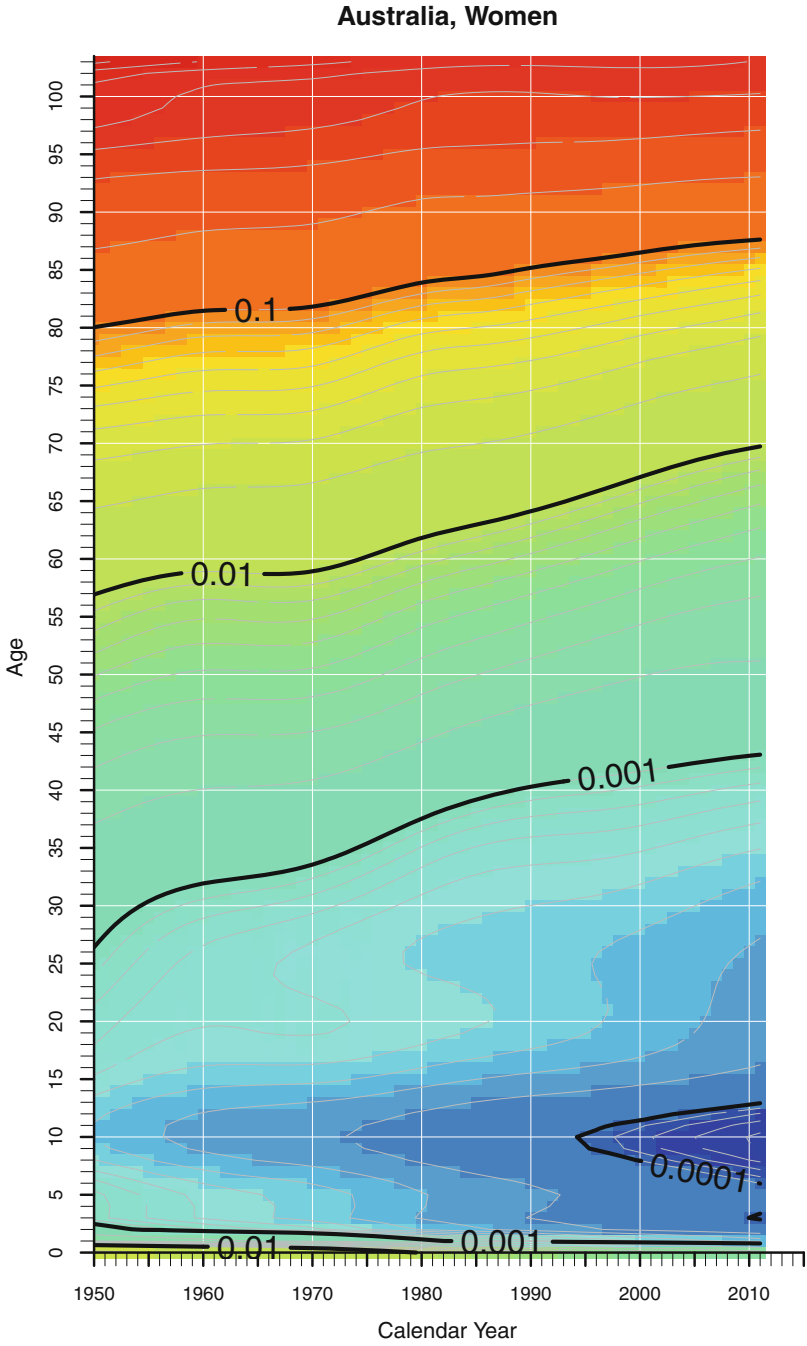
Fig. 5.4 Smoothed death rates for women in Australia, 1950–2011 (Data source: Human Mortality Database)
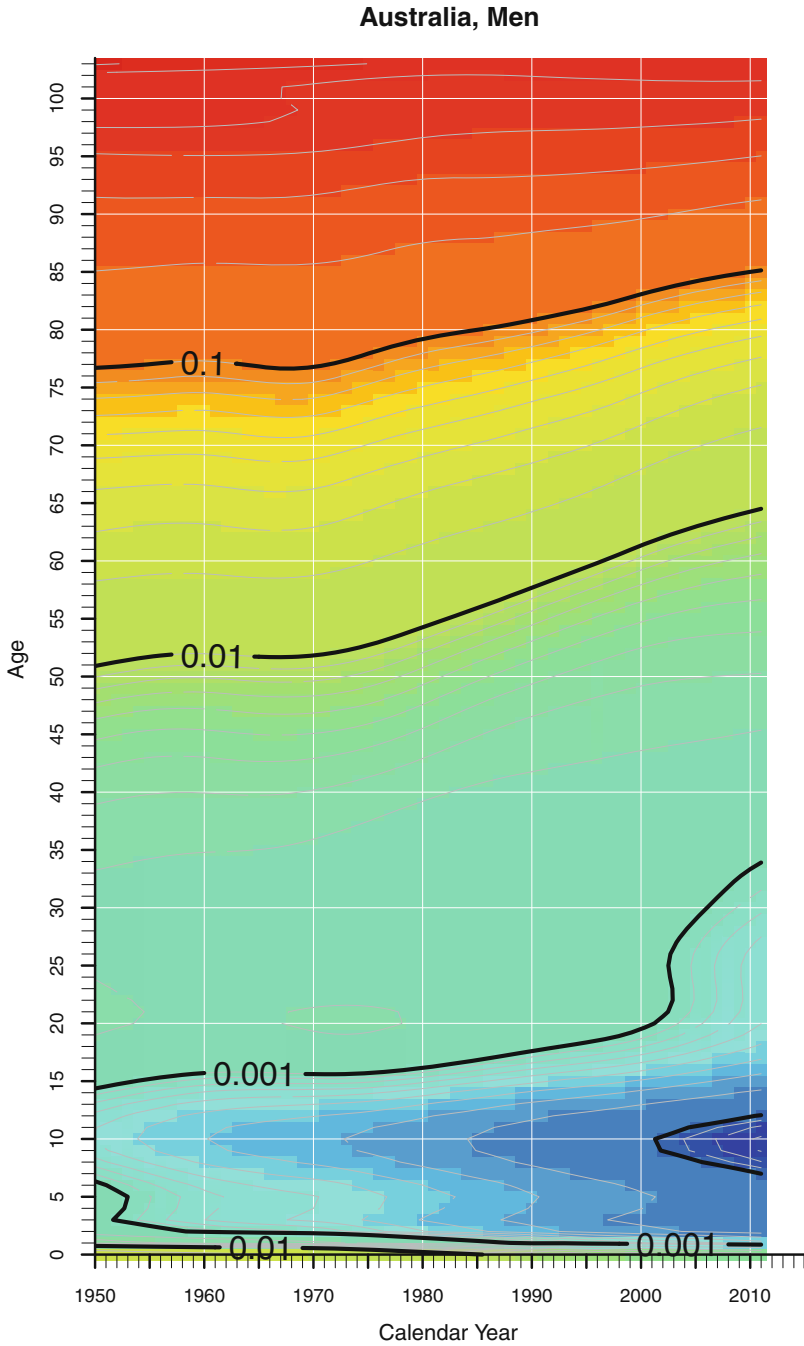
Fig. 5.5 Smoothed death rates for men in Australia, 1950–2011 (Data source: Human Mortality Database)
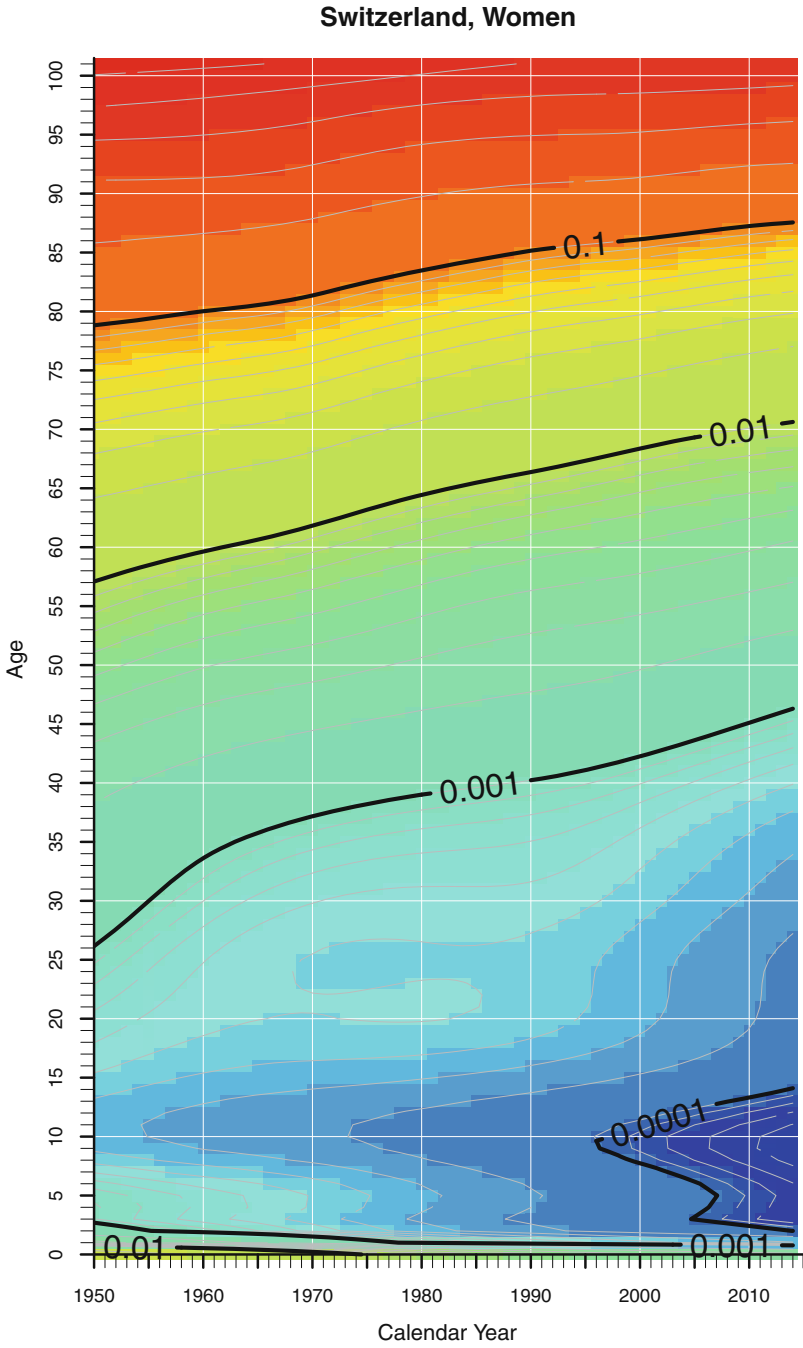
Fig. 5.6 Smoothed death rates for women in Switzerland, 1950–2014 (Data source: Human Mortality Database)
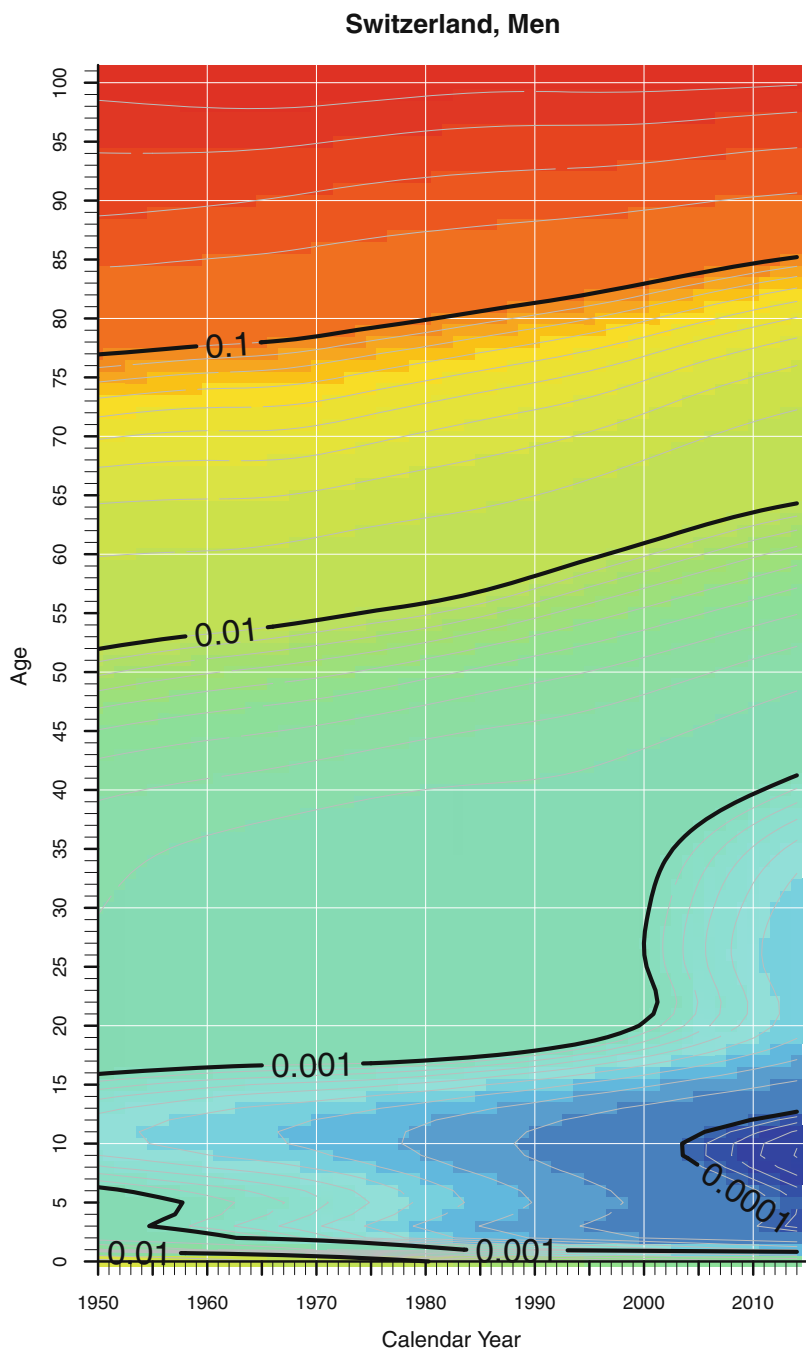
Fig. 5.7 Smoothed death rates for men in Switzerland, 1950–2014 (Data source: Human Mortality Database)

Spain, Women
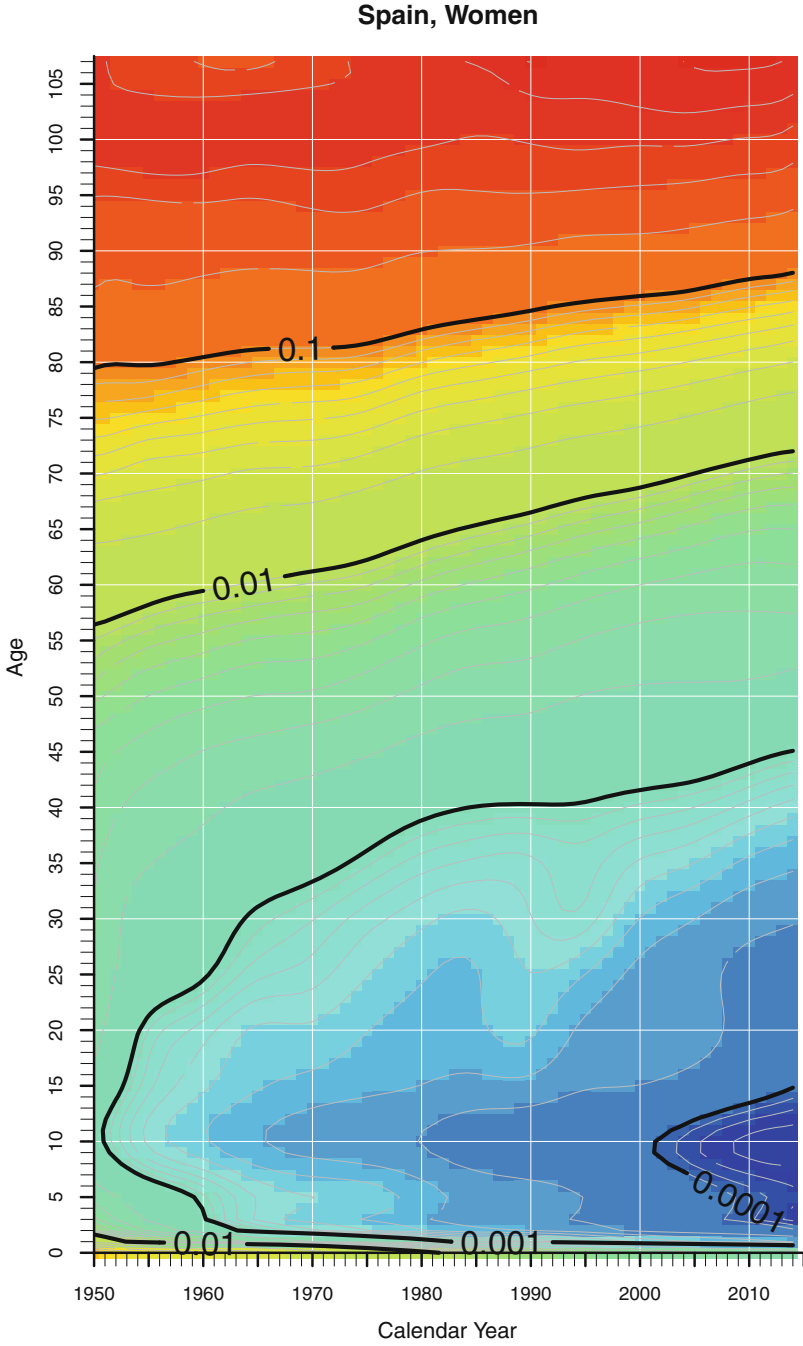
Fig. 5.8 Smoothed death rates for women in Spain, 1950–2014 (Data source: Human Mortality Database)

Spain, Men
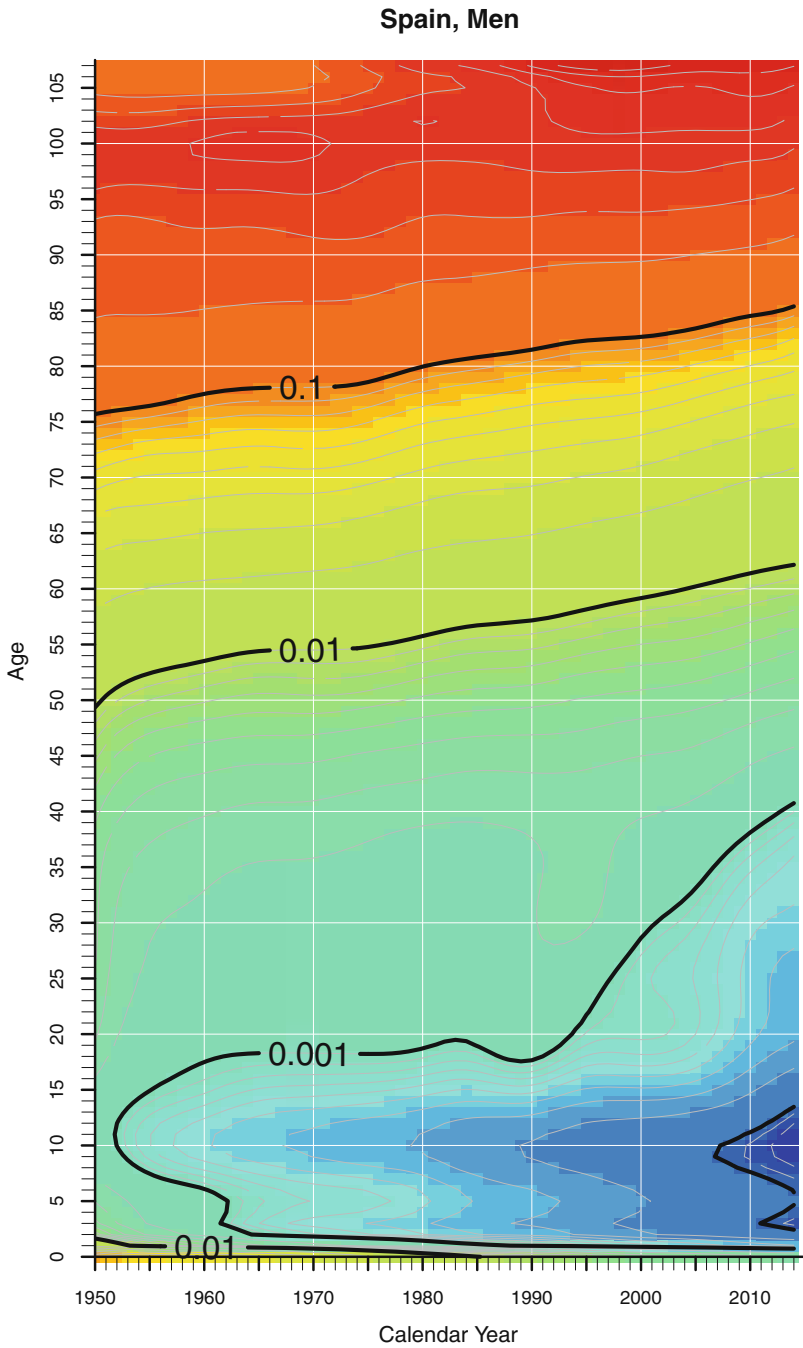
Fig. 5.9 Smoothed death rates for men in Spain, 1950–2014 (Data source: Human Mortality Database)
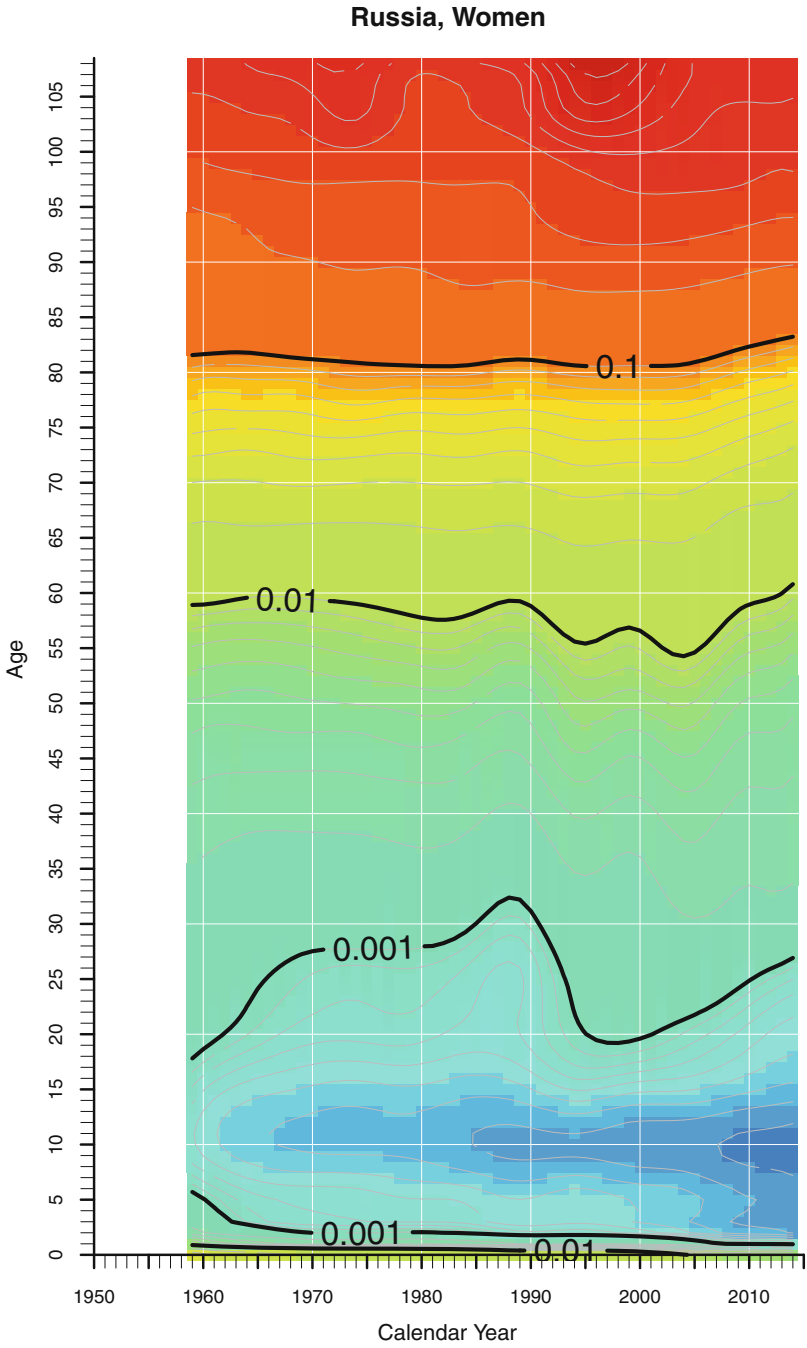
Russia, Women

Fig. 5.10 Smoothed death rates for women in Russia, 1959–2014 (Data source: Human Mortality Database)
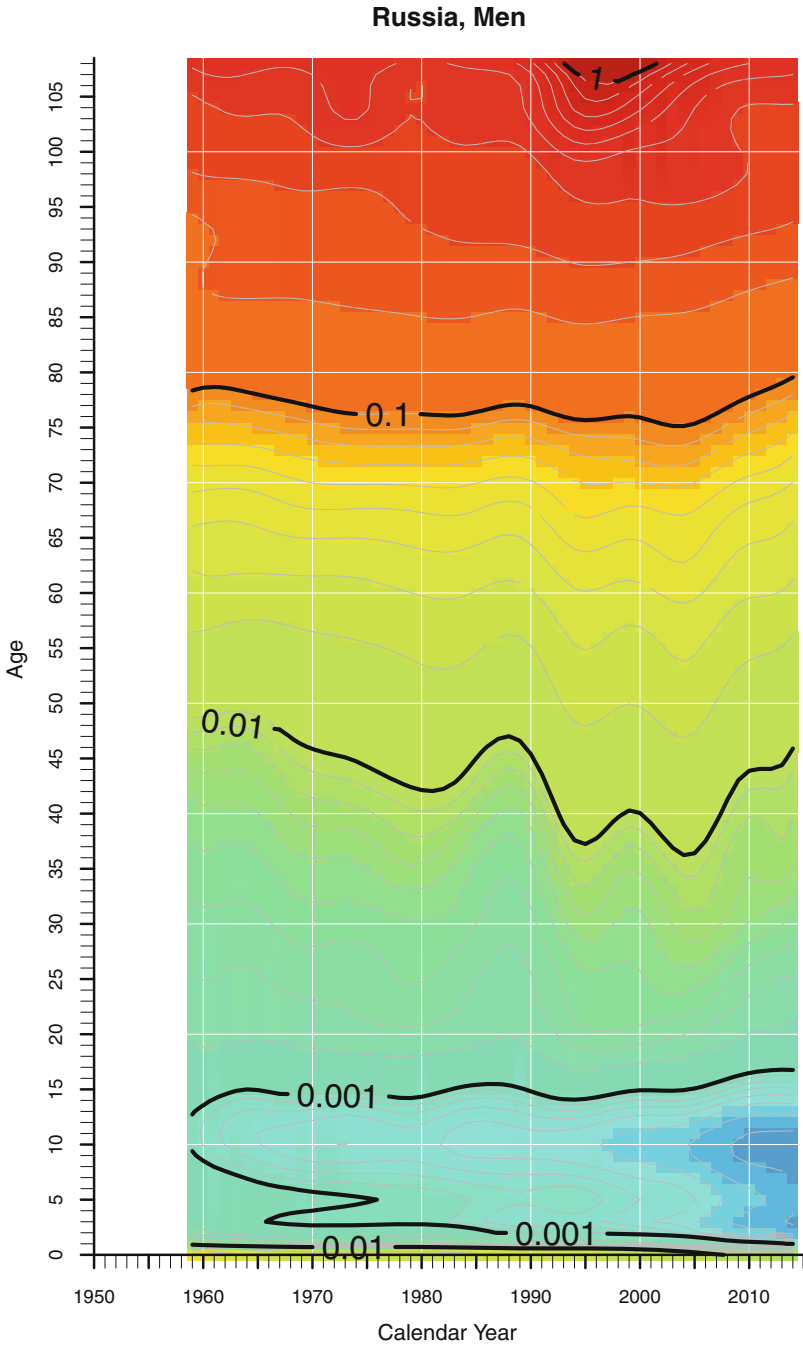
Fig. 5.11 Smoothed death rates for men in Russia, 1959–2014 (Data source: Human Mortality Database)

feature, for instance, the mortality crises among Russian men during the 1980s and 1990s. We do not want to go into further detail here as these smoothed surface maps serve as the major building blocks for the surface maps of rates of mortality improvement, which are the focus of our book and are presented in the next chapter.

Chapter 6
Surface Plots of Rates of Mortality Improvement

6.1 From Smoothed Death Rates to Rates of Mortality Improvement

The colors and contour lines in Fig. 5.3 suggest also a change in pace over time: Each level of mortality seems to change its slope in the early 1990s. We argue that those trend changes are better illustrated with "rates of mortality improvement", which we labeled "ROMIS", than with (smoothed) surface maps of mortality. Given death rates at age x in year t, $m(x, t)$, we defined the rates of mortality improvement, ρ, by assuming a constant rate of change within the period of comparison. In this monograph, we only used annual changes. Hence:

$$\rho(x, t) = -\log_e \left(\frac{m(x, t + 1)}{m(x, t)} \right)$$

It is simply a reformulation of the standard equation for growth with a constant rate r: $P(t) = P(0)e^{rt}$ (e.g., Keyfitz 1977). The minus sign ensures to have positive numbers for survival improvements. We expressed the respective values for ρ in percent. It is comparable to Kannisto et al. (1994) who used a discrete version of the growth equation and aggregated several ages and years.

Figure 6.1 illustrates those ROMIS again with data for Estonian women. To provide a more comprehensive overview, we expanded the age range as well as calendar time. No change or negligible changes ($-0.5\% \leq \rho \leq 0.5\%$) are depicted in white. Slight improvements ($0.5\% < \rho \leq 2.0\%$) are shown in three shades of blue, larger improvements in green colors ($2.0\% < \rho \leq 4.0\%$) and very strong improvements ($\rho > 4.0\%$) in red colors and yellow. If mortality increased, i.e., the survival conditions worsened, we used darker shades of gray for larger mortality increases. Please note that an annual change of $\rho = 0.035 = 3.5\%$ cuts mortality
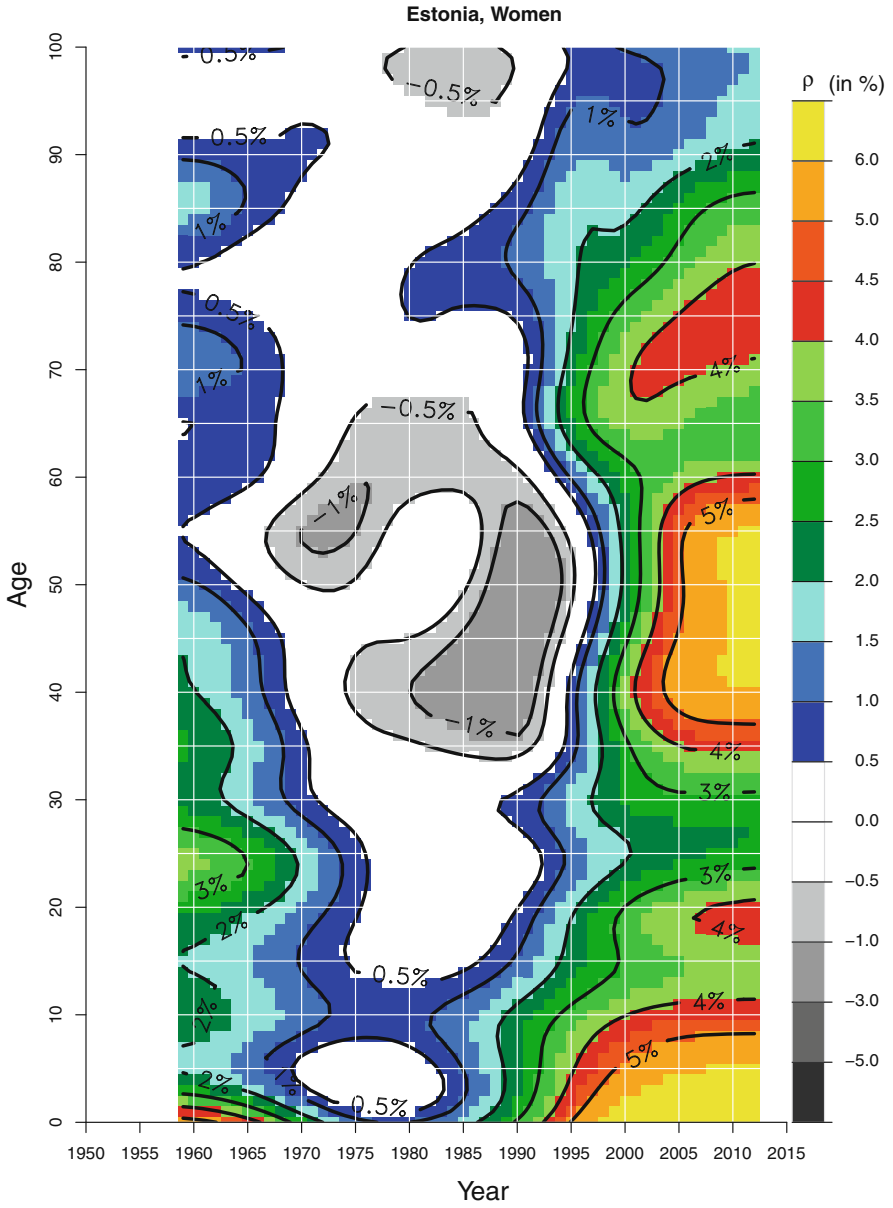
Fig. 6.1 Example of rates of mortality improvement on the Lexis plane: Estonian women aged 0 to 100 years in 1959–2012 (Data source: Human Mortality Database)

in half in less than 20 years.[1] But even at $\rho = 2\%$, which we listed at the threshold from moderate to strong improvements, it requires less than 35 years for a reduction by 50%.

How can we interpret Fig. 6.1, which could be mistaken for a piece of modern art at a first glance? The main shapes appear to be vertical. This implies that mortality changes affected virtually all age groups at the same moment in time—classical period effects. We can also see that white and gray are the dominant colors for females in Estonia for the 1970s and the 1980s. Thus, mortality remained more or less constant during those two decades. During the 1980s at ages 35–60, we can even spot some dark gray areas that correspond to increasing levels of mortality. We can witness a trend reversal approximately in 1990. Within a couple of years, Estonian women at almost all ages experienced remarkable survival improvements. The colors illustrate that mortality dropped by more than 4% for several years at some ages. At such a rapid pace, it takes about 10 years to cut mortality by a third.

6.2 Results

Figures 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, 6.10, 6.11, 6.12, 6.13, 6.14, 6.15, 6.16, 6.17, 6.18, 6.19, 6.20, and 6.21 (pages 46–65) depict Lexis diagrams of rates of mortality improvements ("ROMIs"), which are the time derivative of age-specific death rates. We argue that those maps are better able to illustrate mortality dynamics than the commonly used "heat maps" of mortality. We plotted our first ROMIs on the Lexis surface about 10 years ago (Rau et al. 2008). In the meantime, those plots have become more commonplace, especially among actuaries, to visualize mortality dynamics. Our method can be considered as a descriptive tool. It is able to detect the predominant dynamics of mortality (or of any other phenomenon measured on the Lexis surface). We think that those "ROMI"-maps provide better insights into mortality dynamics than standard surface maps but are equally intuitively understandable.

During the 1950s, the first years of our observation period, survival improved tremendously especially for infants, children, and young adults. The most remarkable declines in mortality were recorded for Japanese females (Fig. 6.14, page 58). After the end of World War II, life expectancy in Japan was below the average of western European countries. According to data from the Human Mortality Database, life expectancy for Japanese females rose from 60.9 years in 1950 to 72.3 in 1963. Thus, life expectancy increased by almost 1 year within each calendar year during that time span! But also France (Fig. 6.9, p. 53), Italy (Fig. 6.13, p. 57), England & Wales (Fig. 6.7, p. 51) or the United States (Fig. 6.21, p. 65), to name only a few, gained several years of life due to mortality declines at younger ages.

---

[1] $0.5m(x, t) = m(x, t)e^{\rho\,t}; 0.5 = e^{\rho\,t}; \log_e(0.5) = \rho t; \log_e(0.5)/\rho; \log_e(0.5)/0.035 = -19.80421.$
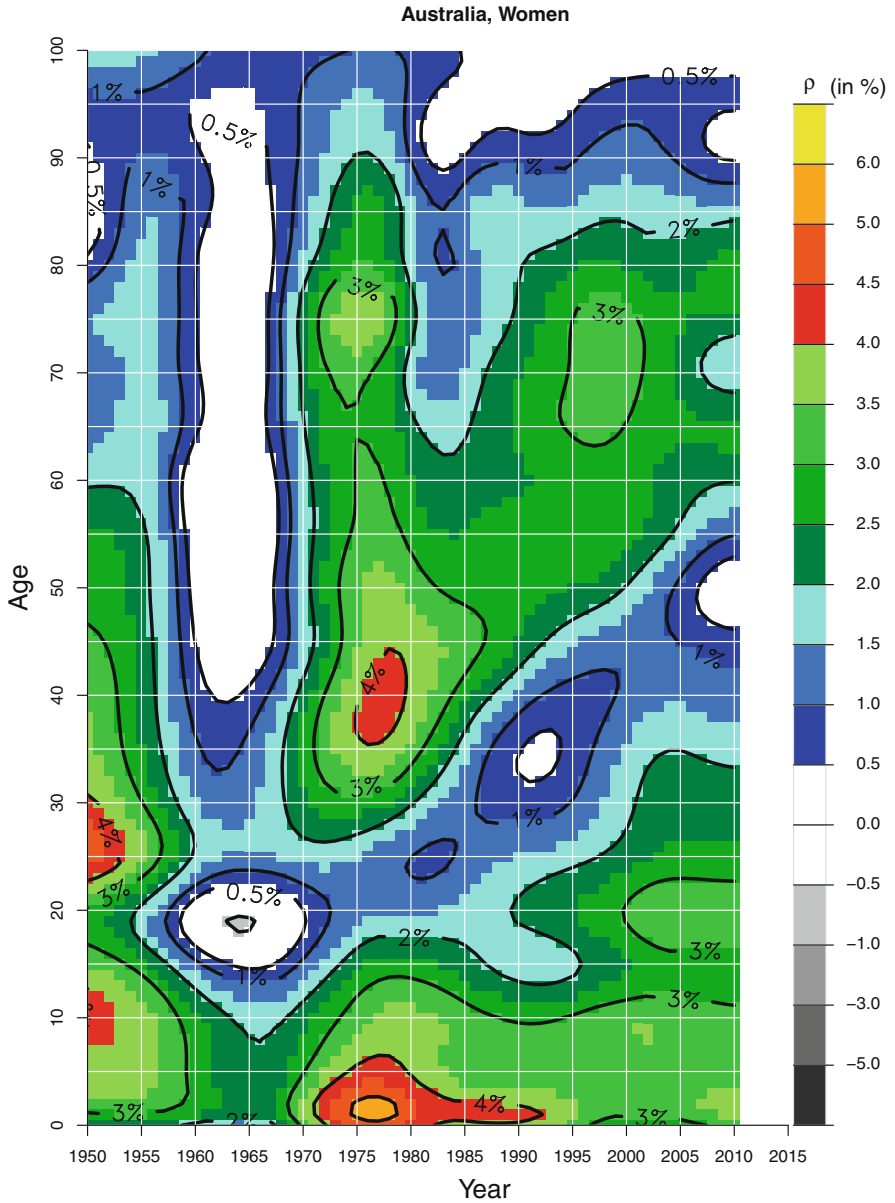
Fig. 6.2 Rates of mortality improvement for women in Australia, 1950–2010 (Data source: Human Mortality Database)
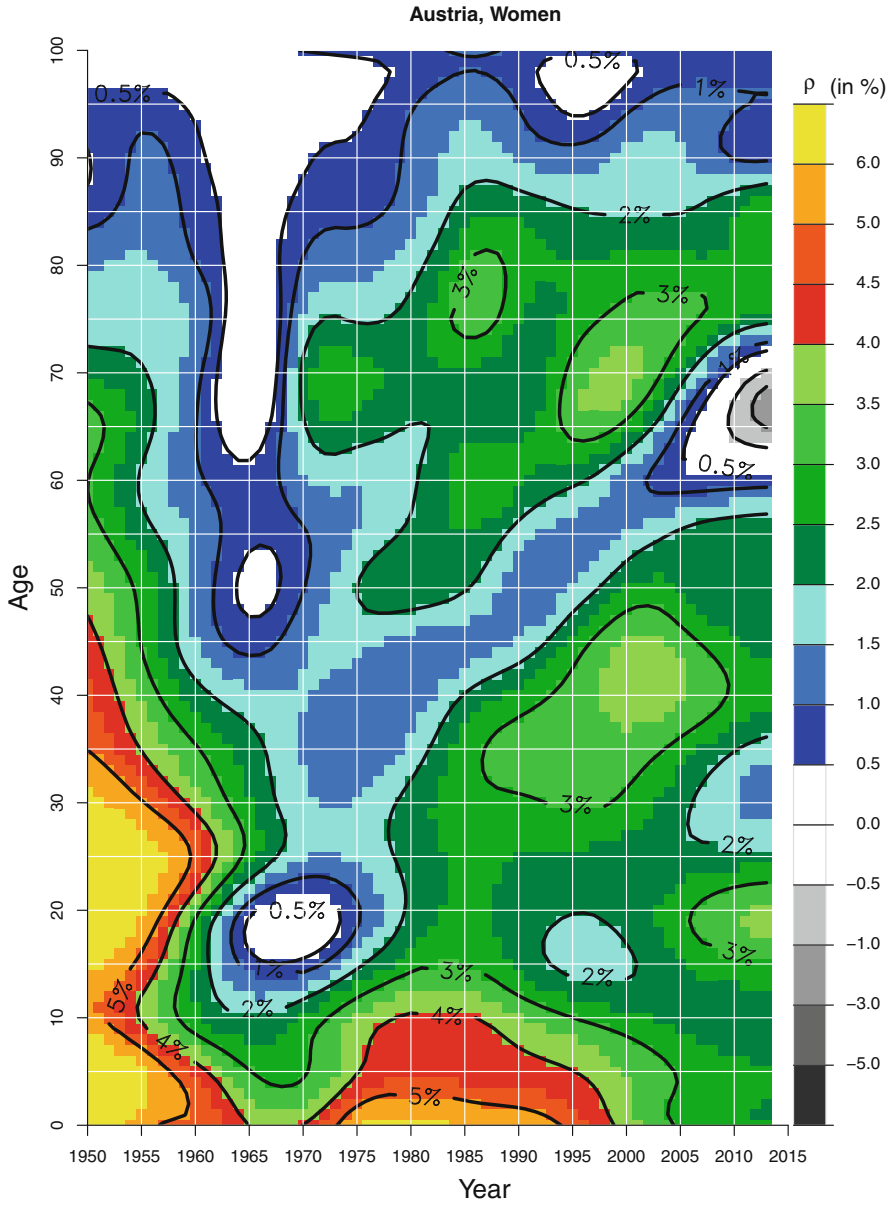
Fig. 6.3 Rates of mortality improvement for women in Austria, 1950–2013 (Data source: Human Mortality Database)
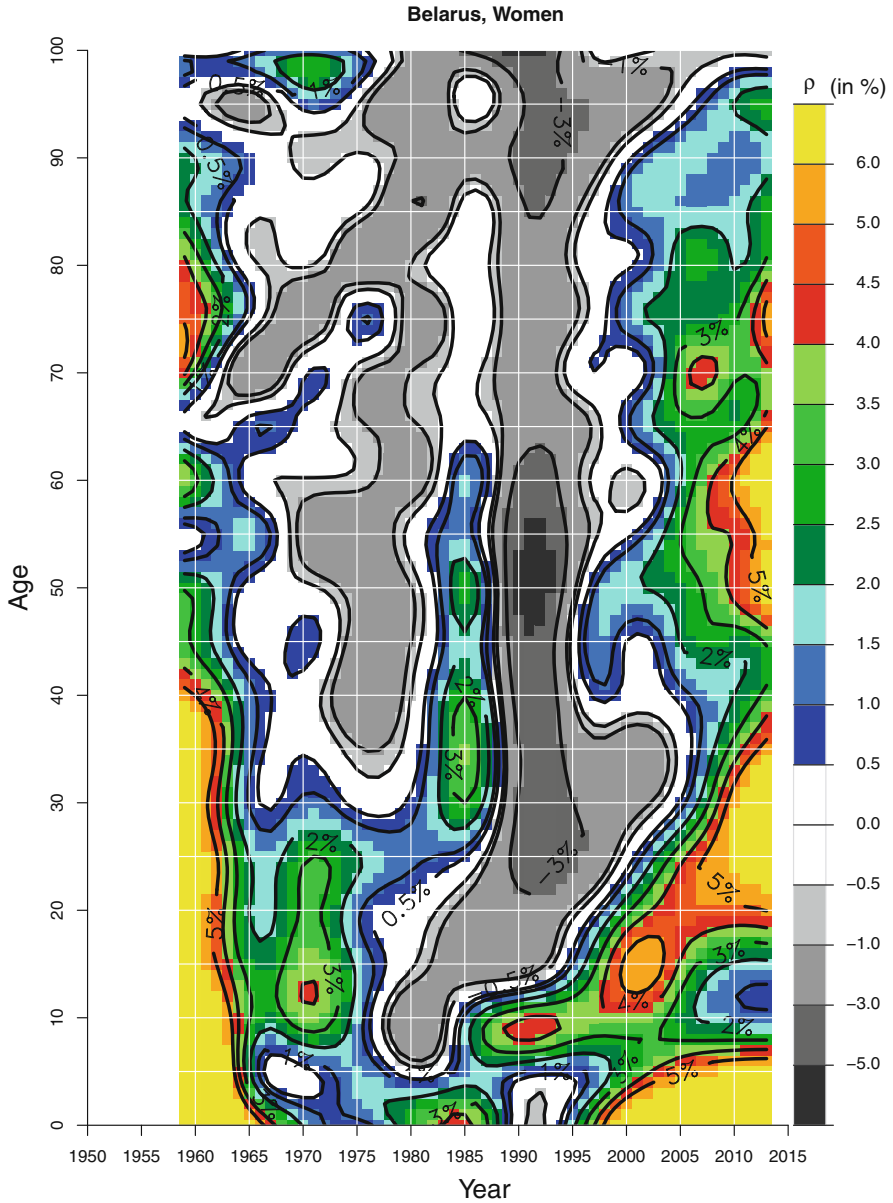
Fig. 6.4 Rates of mortality improvement for women in Belarus, 1950–2013 (Data source: Human Mortality Database)
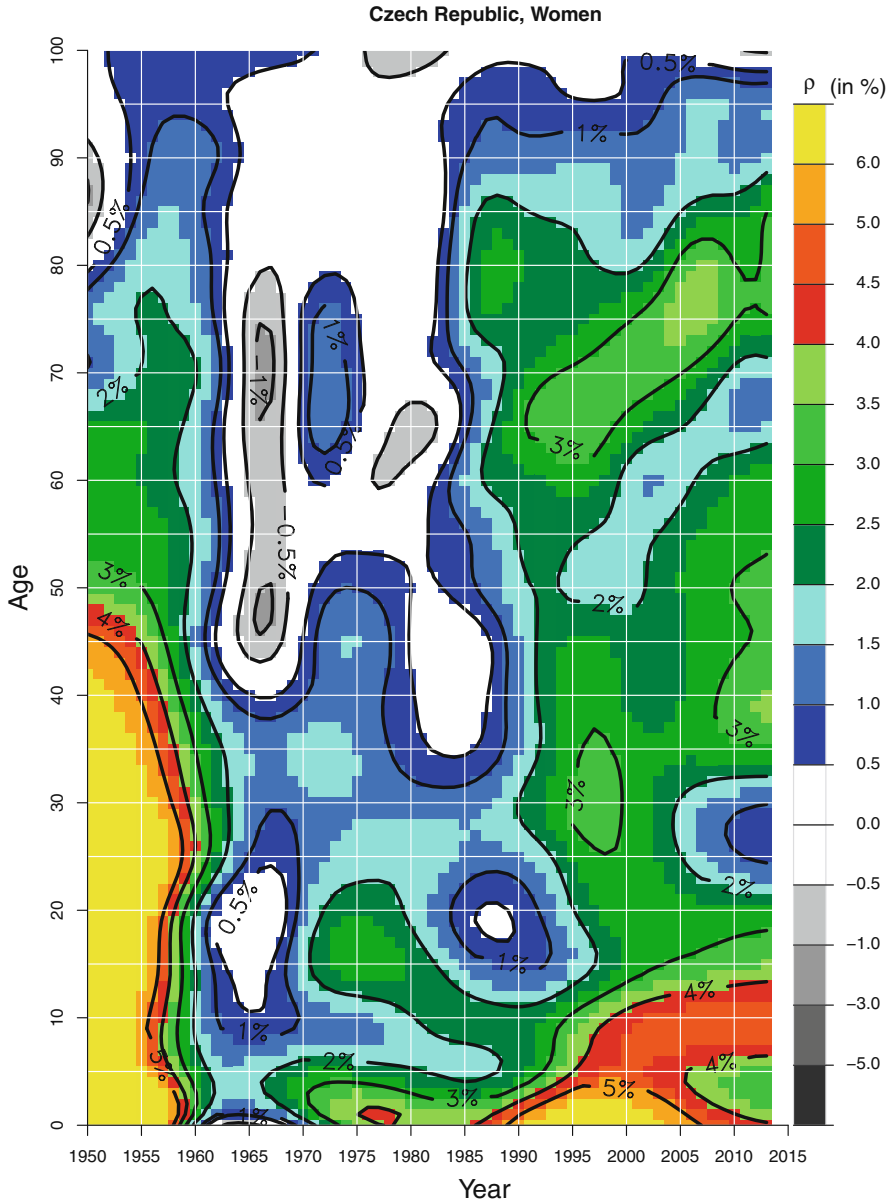
Fig. 6.5 Rates of mortality improvement for women in Czech Republic, 1950–2013 (Data source: Human Mortality Database)
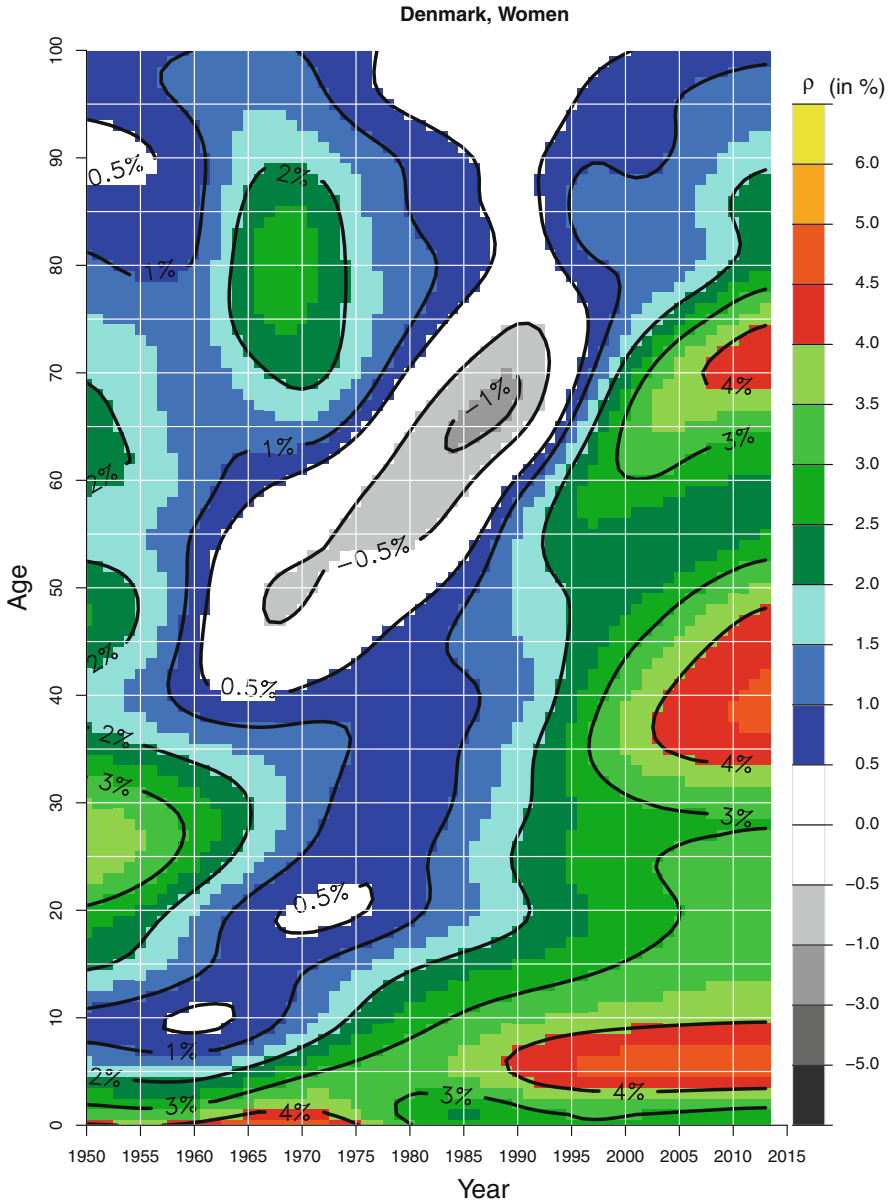
Fig. 6.6 Rates of mortality improvement for women in Denmark, 1950–2013 (Data source: Human Mortality Database)
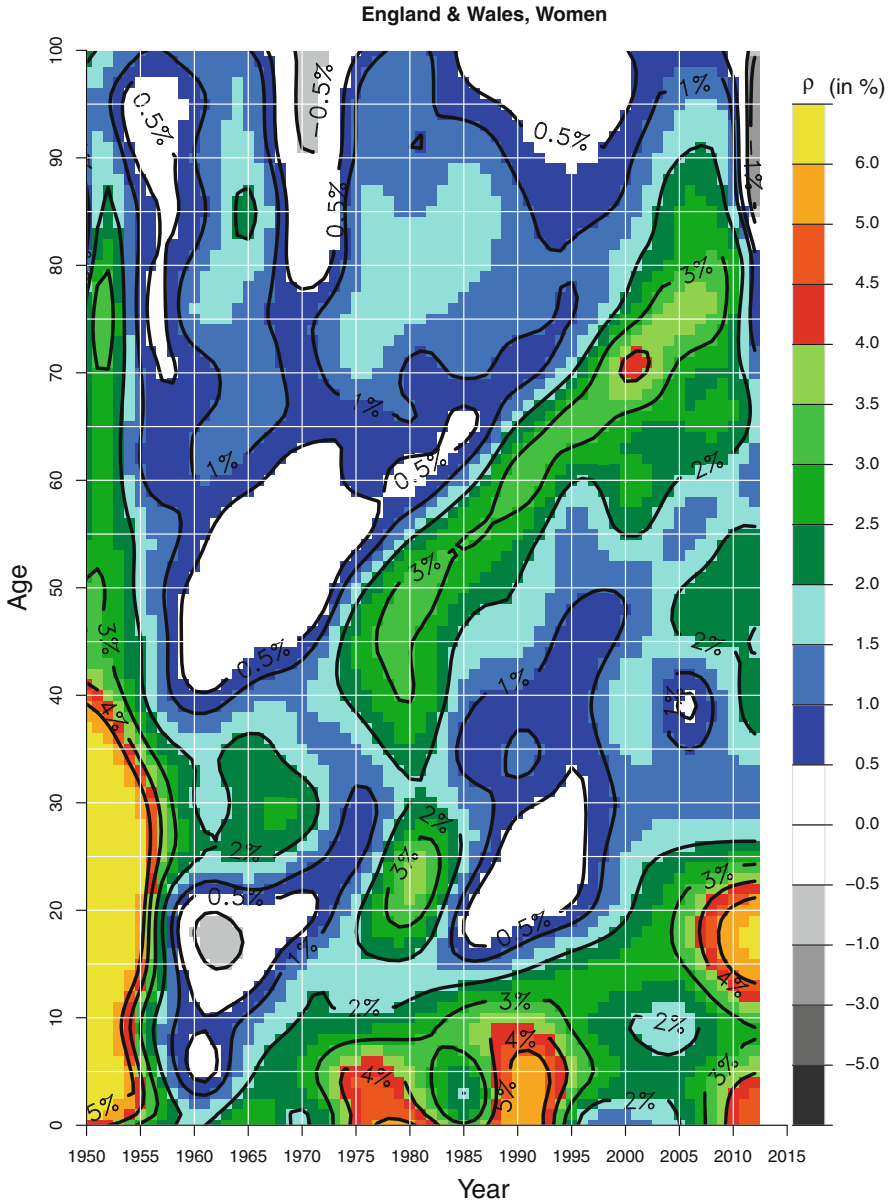
Fig. 6.7 Rates of mortality improvement for women in England & Wales, 1950–2012 (Data source: Human Mortality Database)
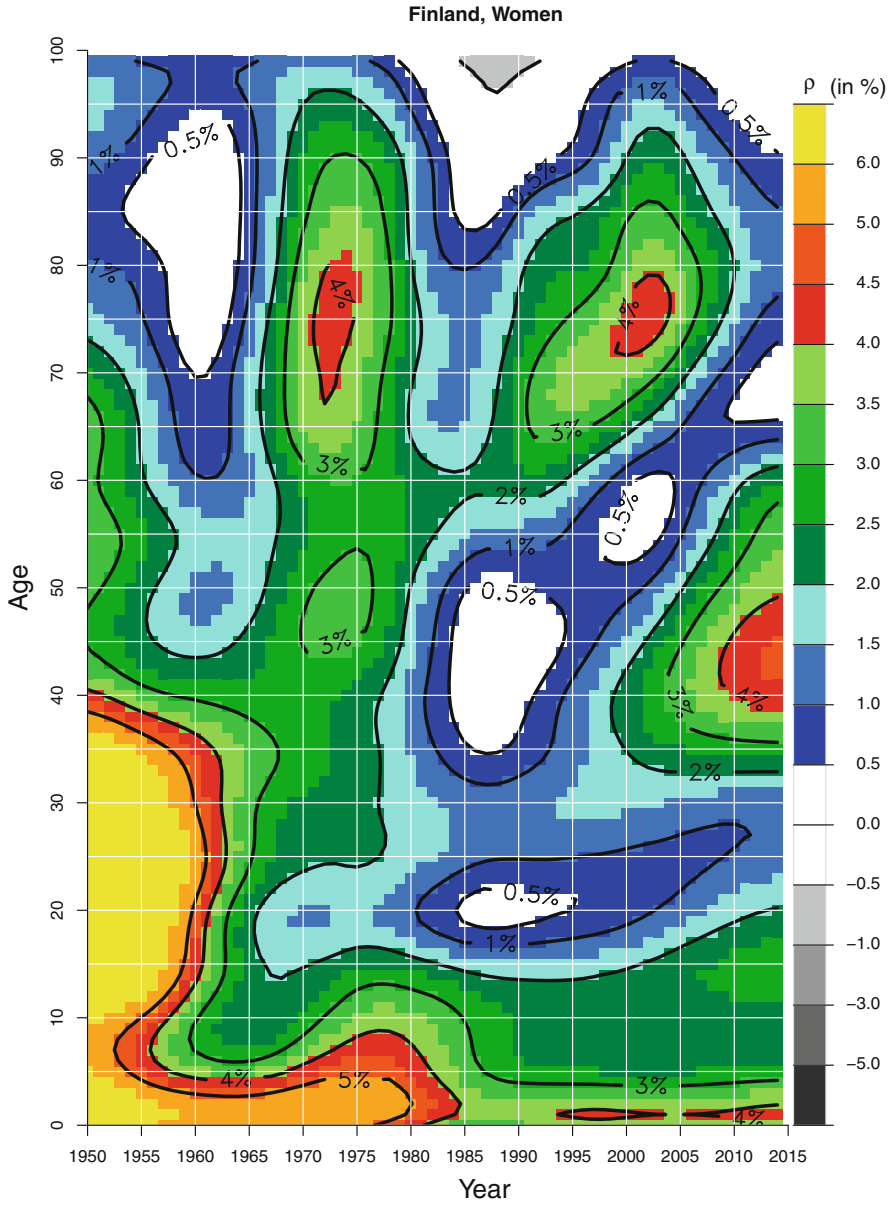
Fig. 6.8 Rates of mortality improvement for women in Finland, 1950–2014 (Data source: Human Mortality Database)
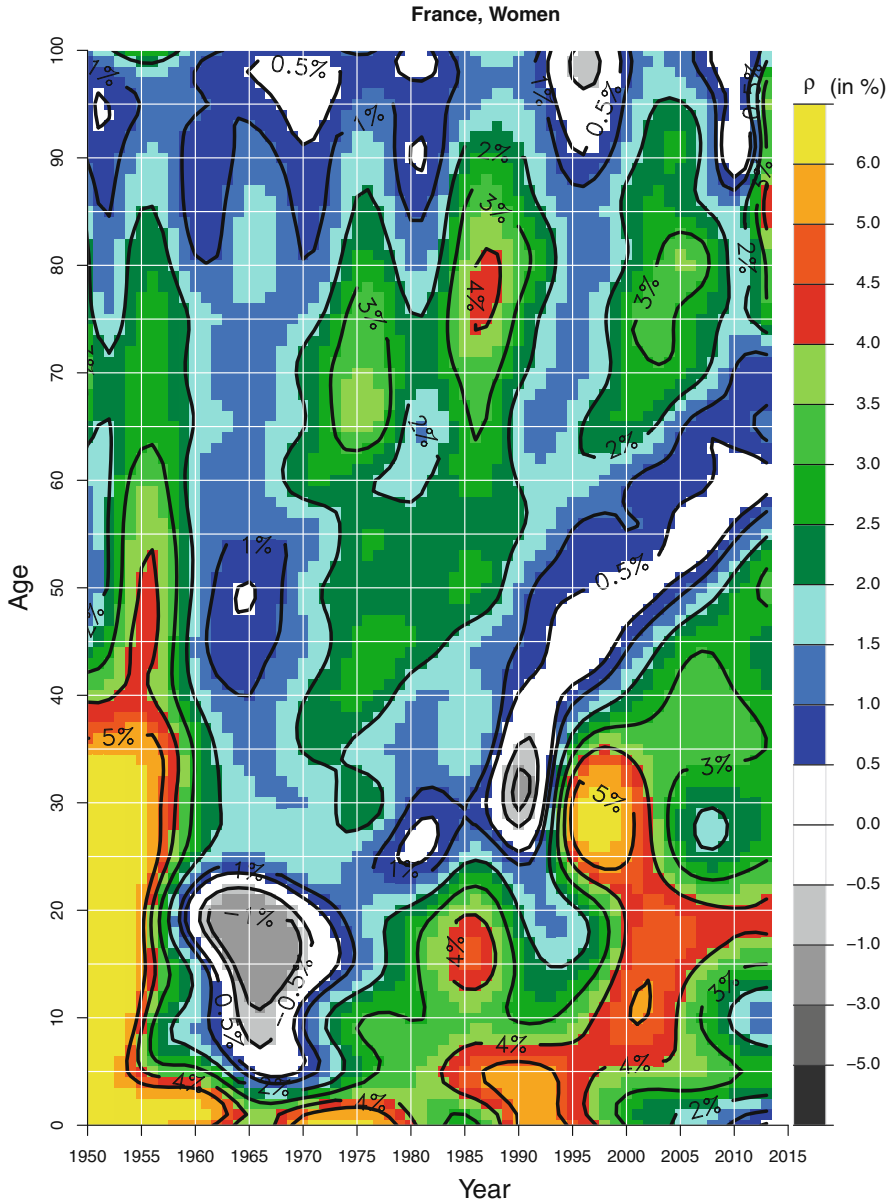
Fig. 6.9 Rates of mortality improvement for women in France, 1950–2013 (Data source: Human Mortality Database)
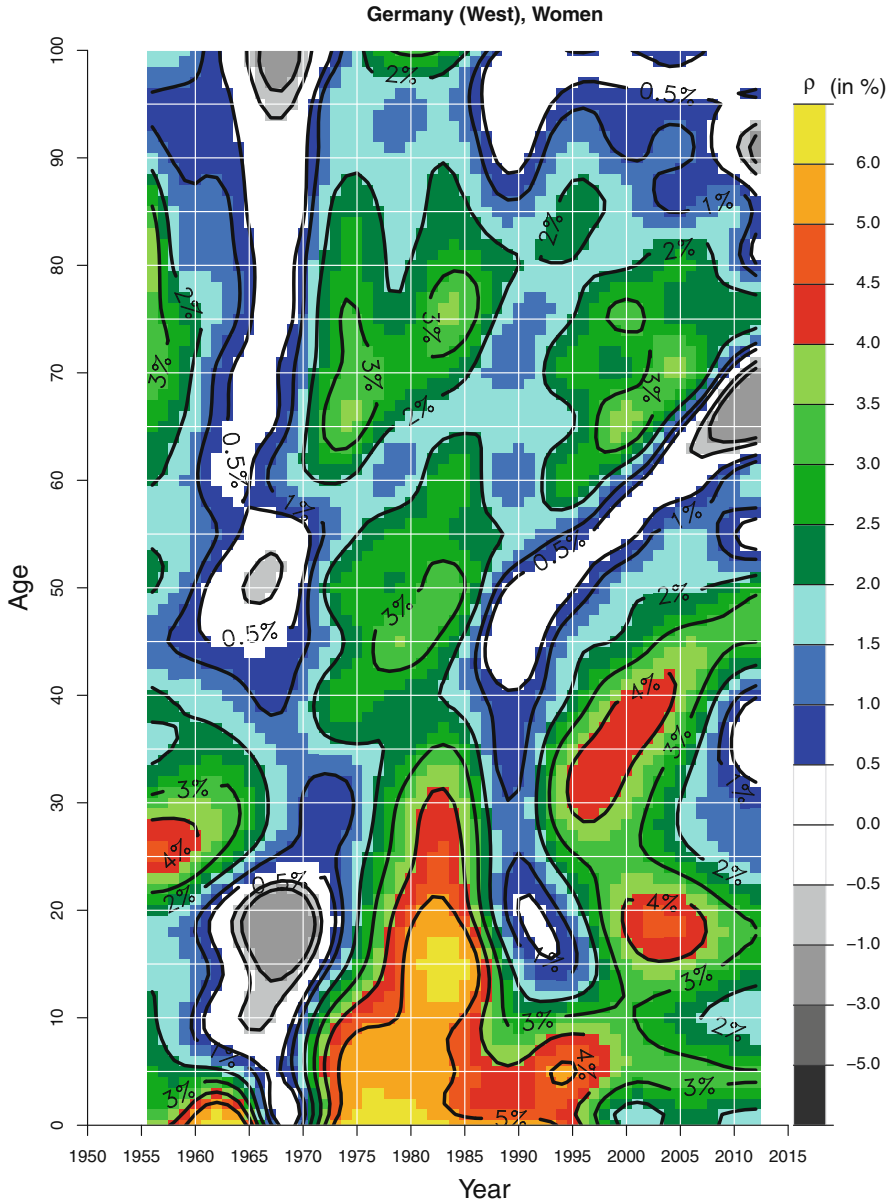
Fig. 6.10 Rates of mortality improvement for women in western Germany, 1956–2012 (Data source: Human Mortality Database)
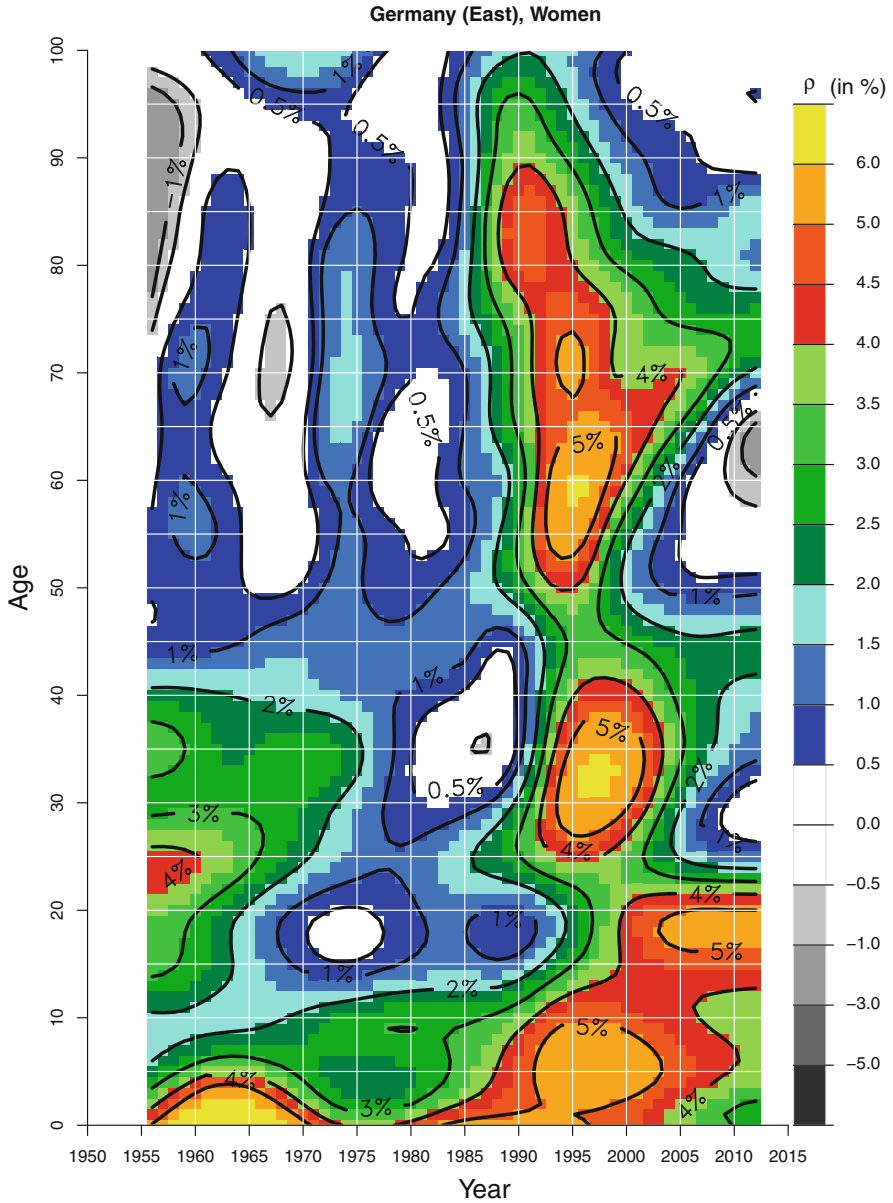
Fig. 6.11 Rates of mortality improvement for women in eastern Germany, 1956–2012 (Data source: Human Mortality Database)
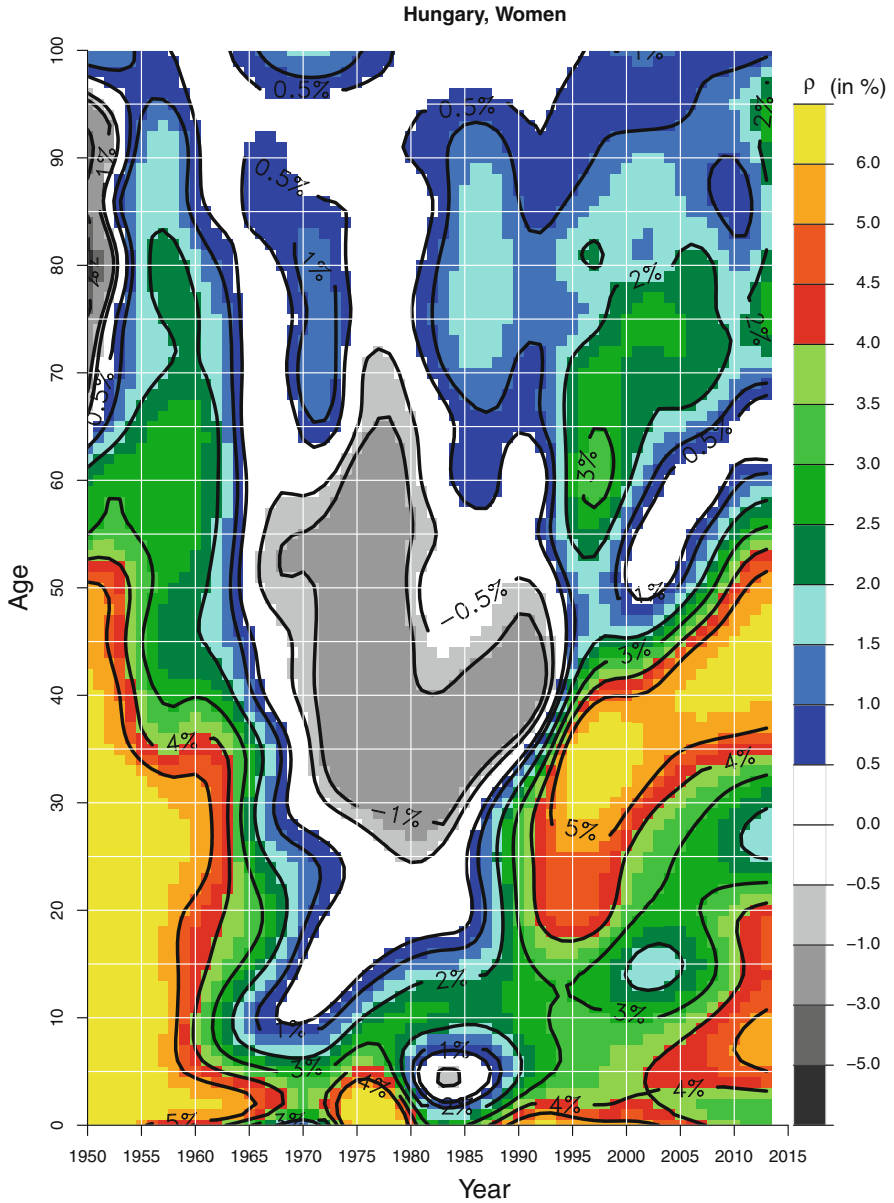
Fig. 6.12 Rates of mortality improvement for women in Hungary, 1950–2013 (Data source: Human Mortality Database)
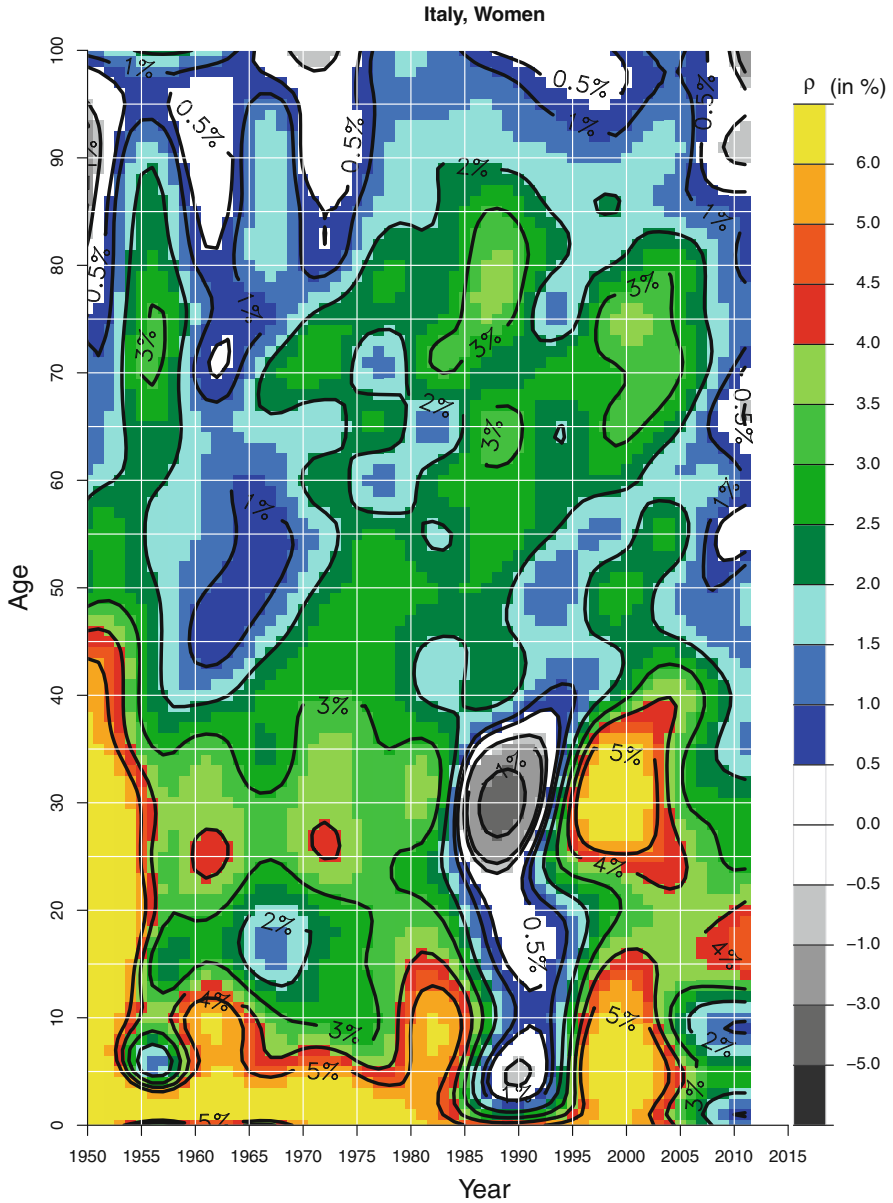
Fig. 6.13 Rates of mortality improvement for women in Italy, 1950–2011 (Data source: Human Mortality Database)
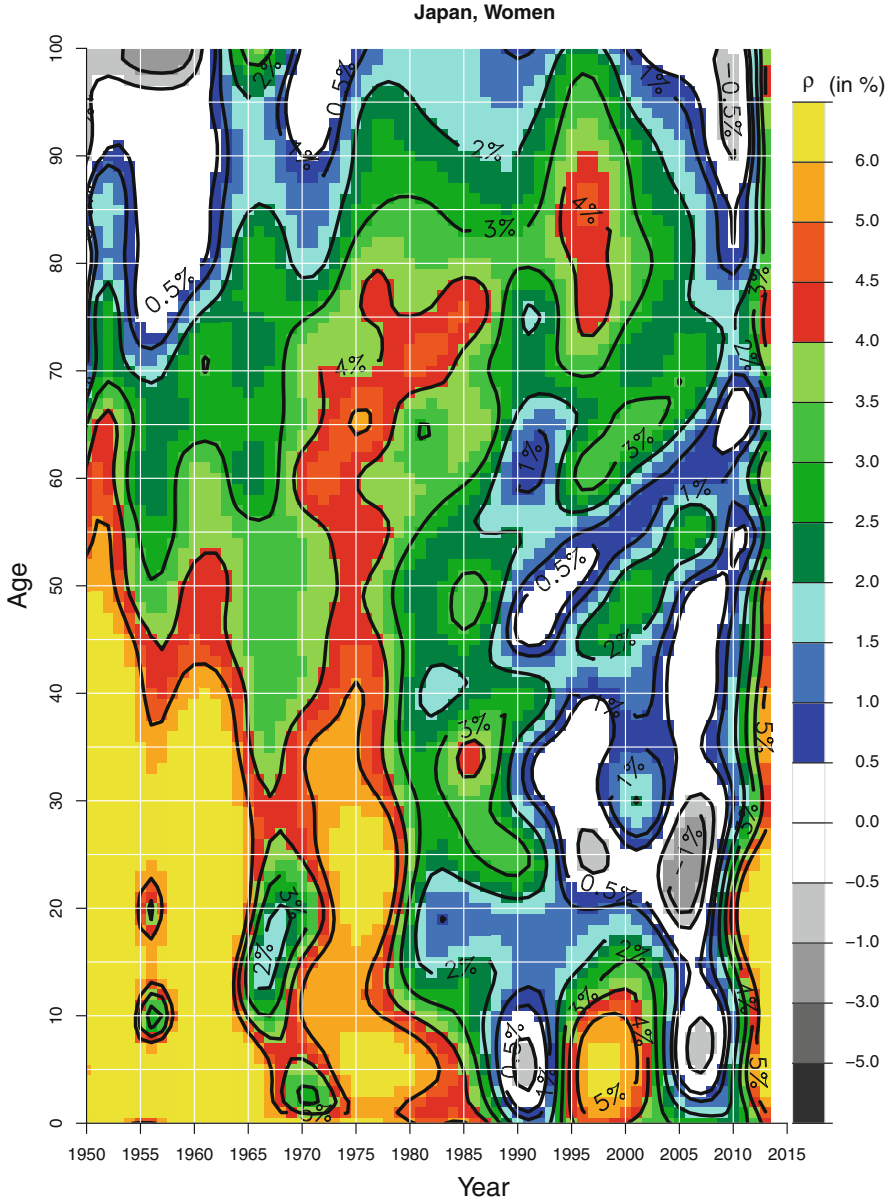
Fig. 6.14 Rates of mortality improvement for women in Japan, 1950–2013 (Data source: Human Mortality Database)
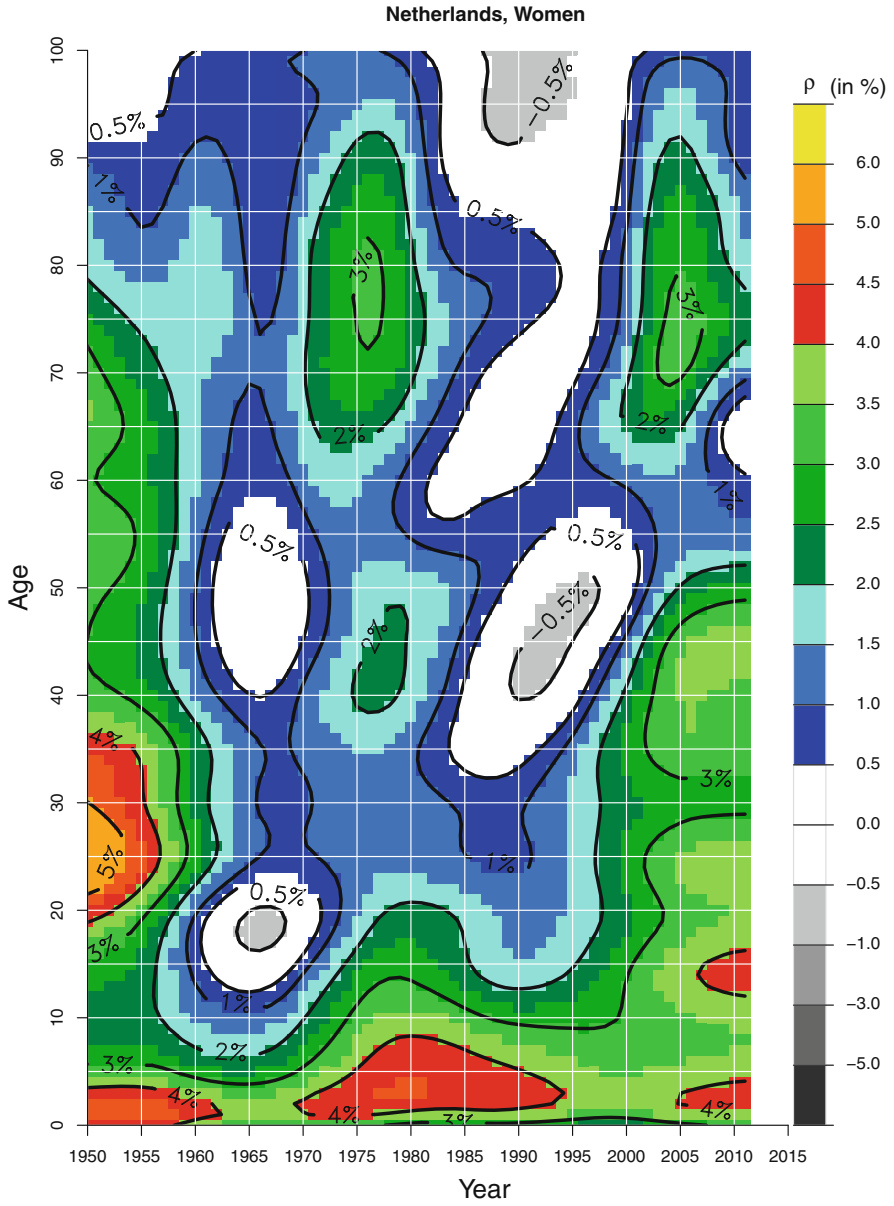
Fig. 6.15 Rates of mortality improvement for women in Netherlands, 1950–2011 (Data source: Human Mortality Database)
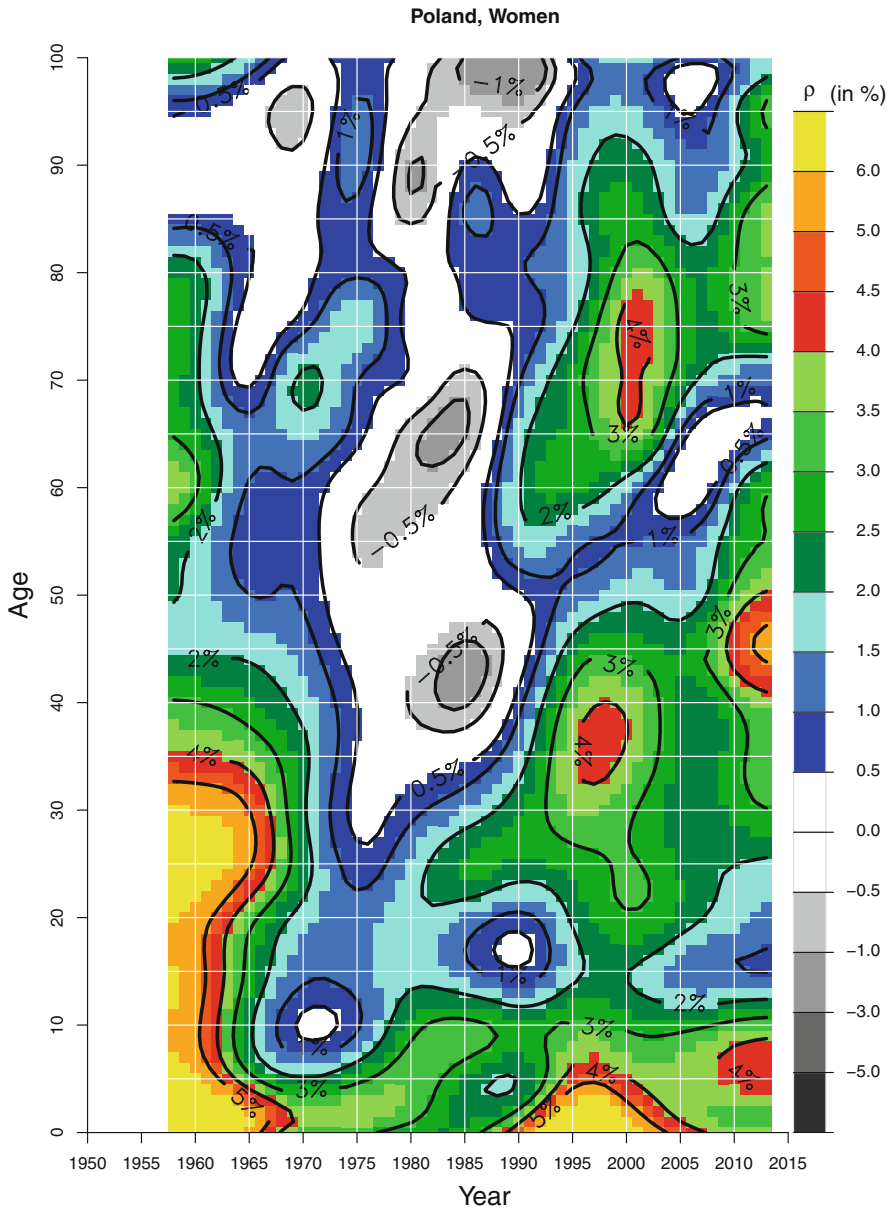
Fig. 6.16 Rates of mortality improvement for women in Poland, 1958–2013 (Data source: Human Mortality Database)
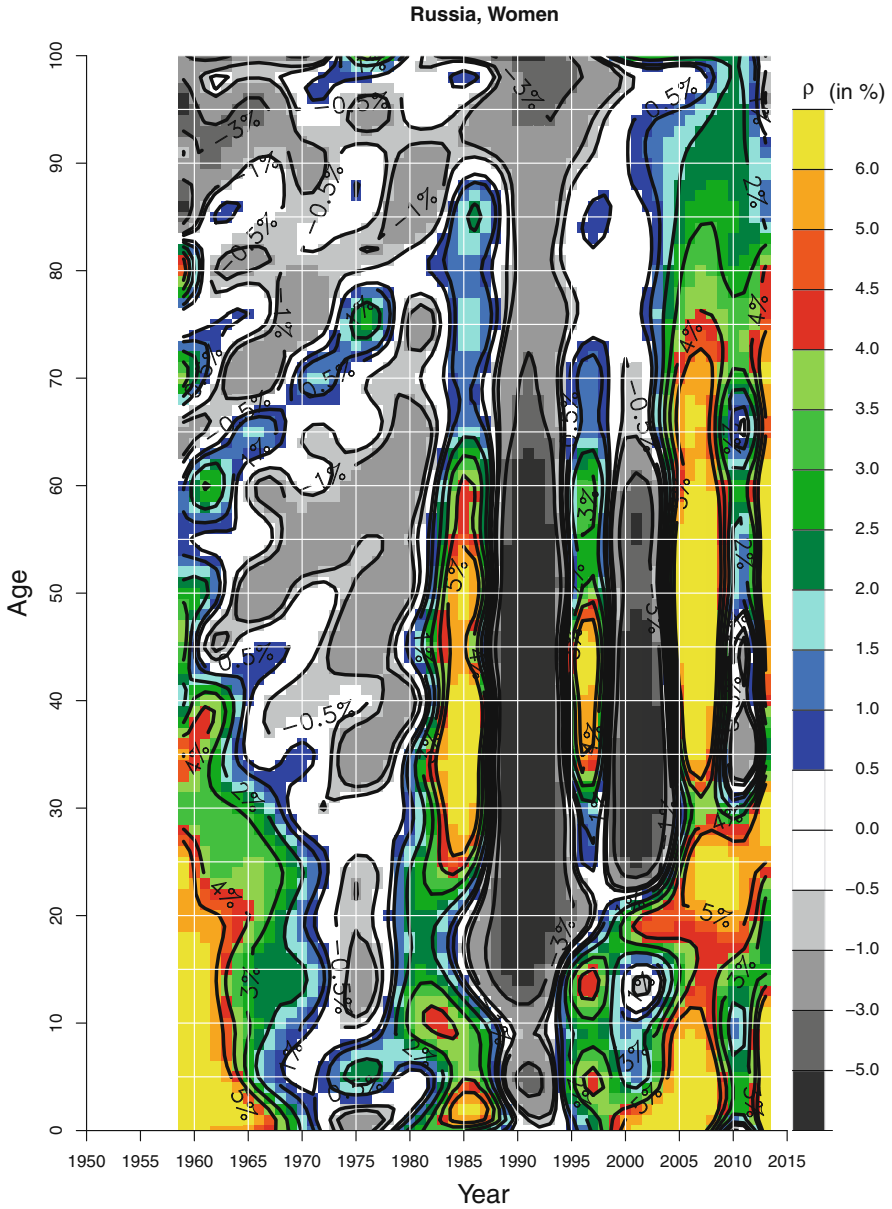
Fig. 6.17 Rates of mortality improvement for women in Russia, 1959–2013 (Data source: Human Mortality Database)
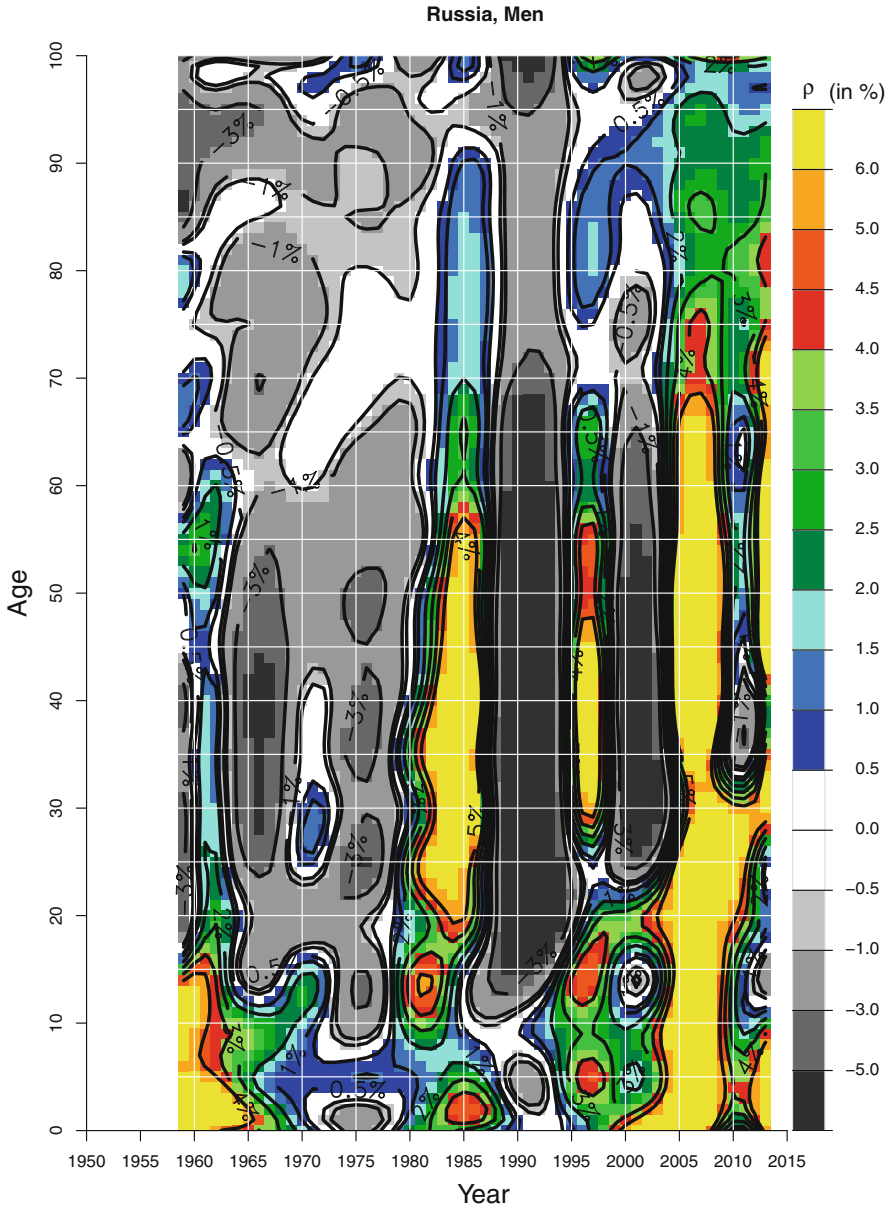
Fig. 6.18 Rates of mortality improvement for men in Russia, 1959–2013 (Data source: Human Mortality Database)
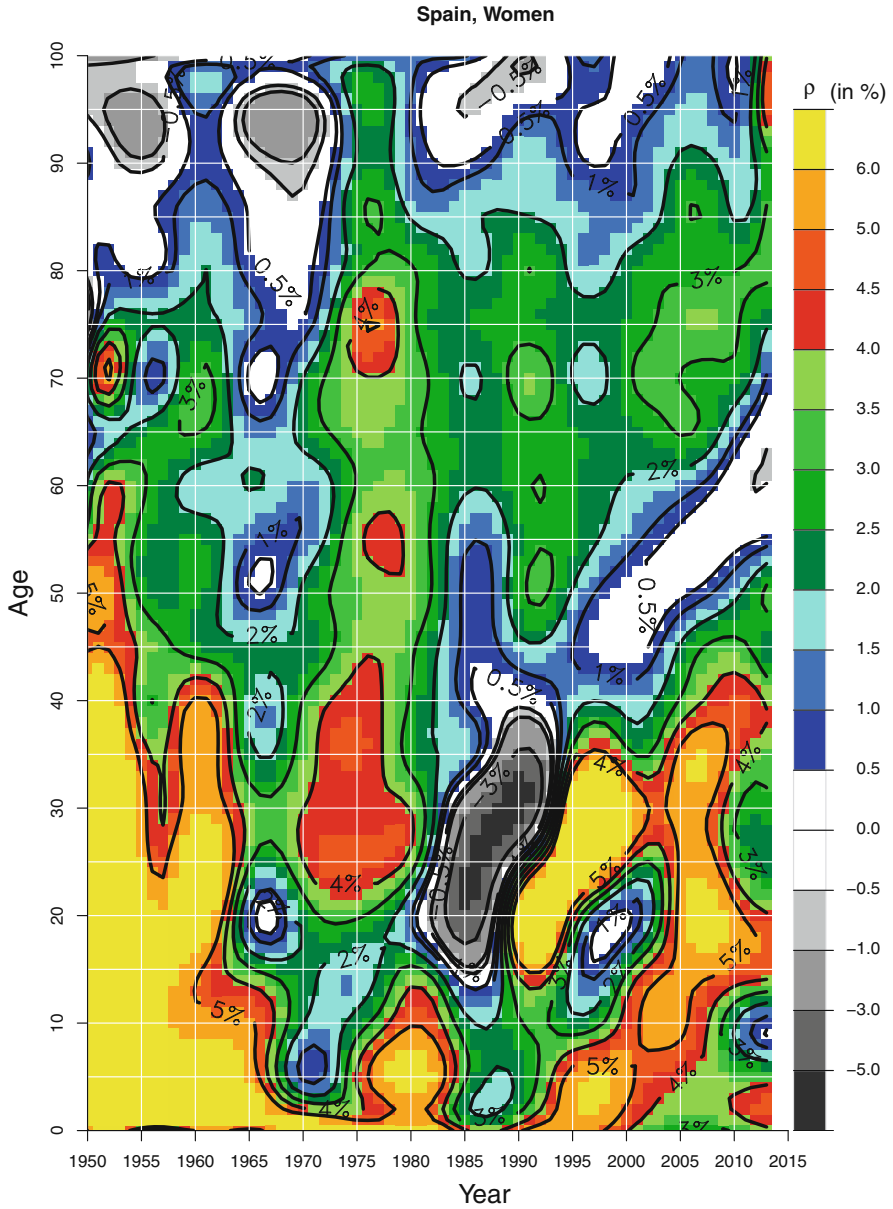
Fig. 6.19 Rates of mortality improvement for women in Spain, 1950–2013 (Data source: Human Mortality Database)
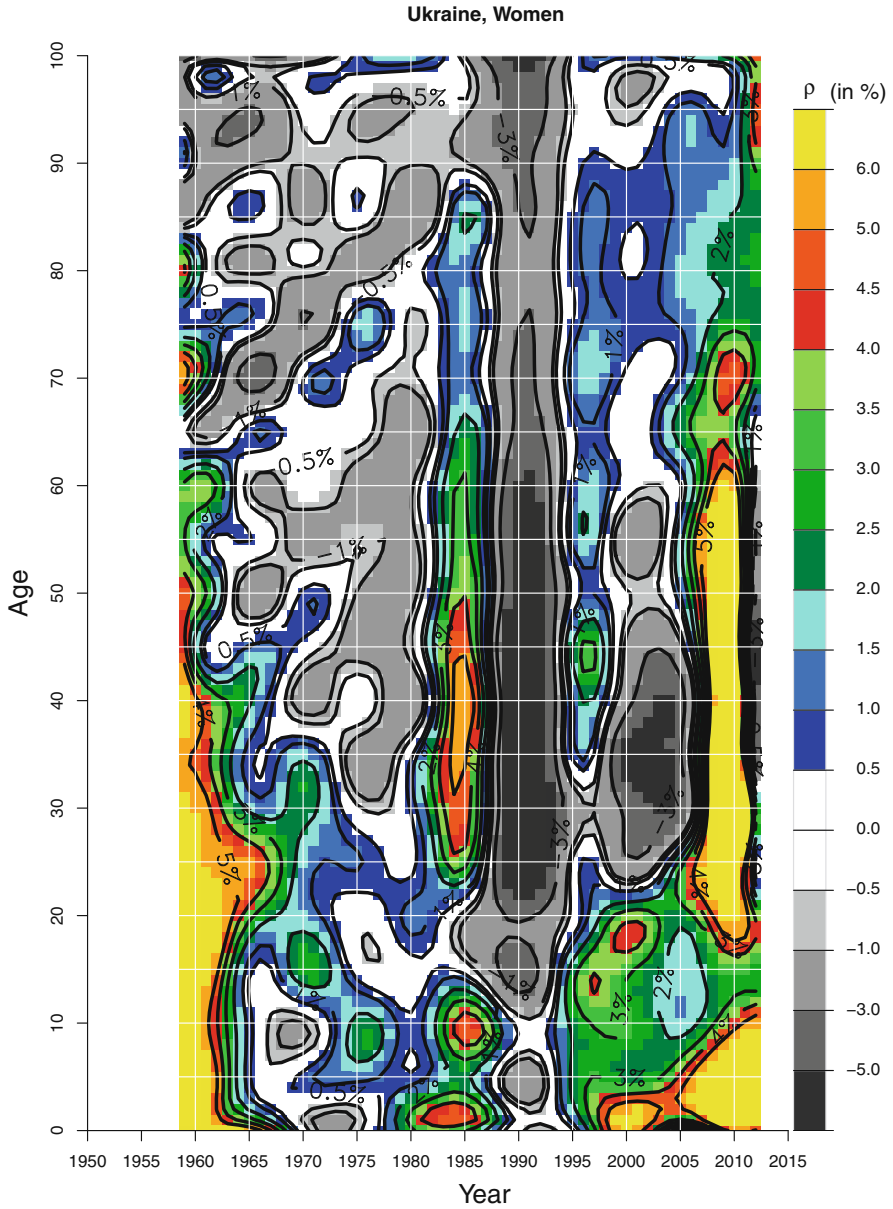
Fig. 6.20 Rates of mortality improvement for women in Ukraine, 1959–2012 (Data source: Human Mortality Database)
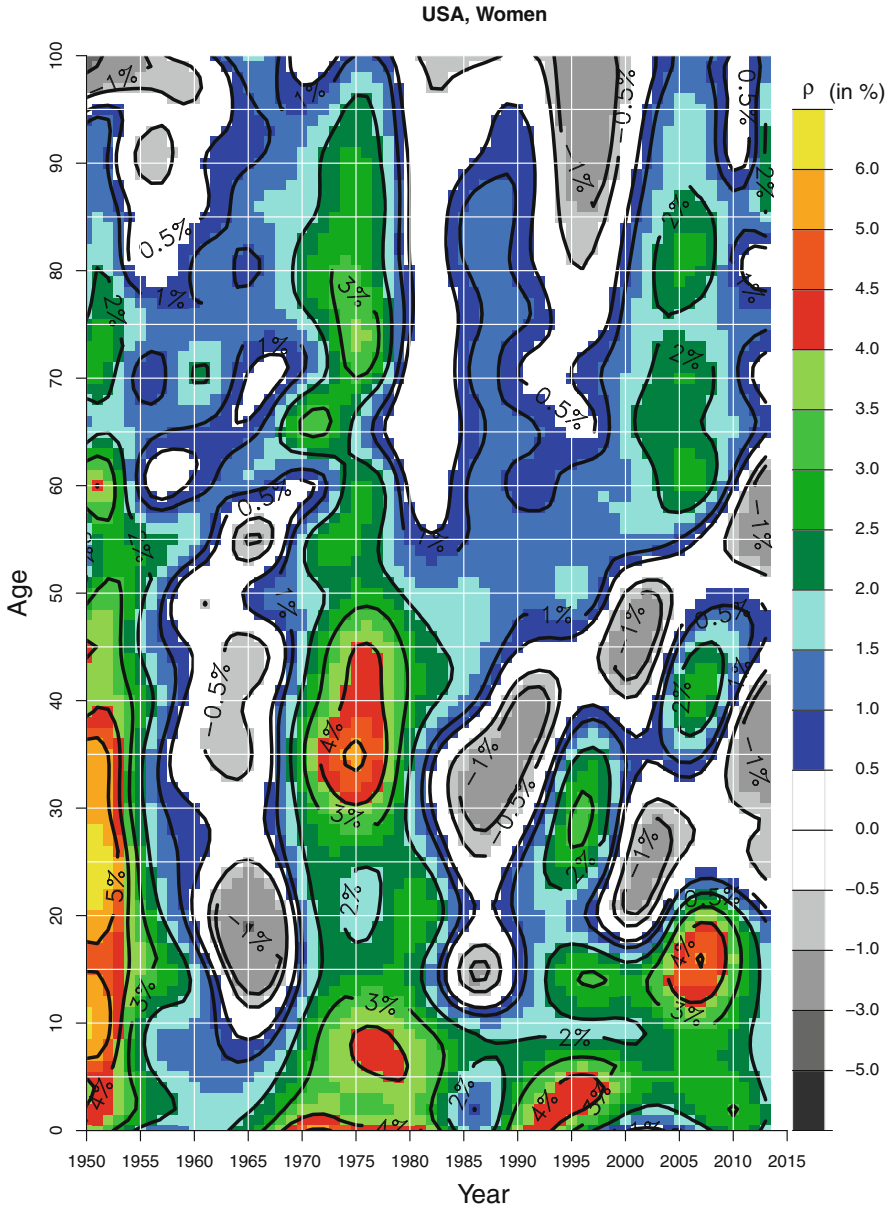
Fig. 6.21 Rates of mortality improvement for women in USA, 1950–2013 (Data source: Human Mortality Database)

Another vertical pattern, suggesting a period effect, can be observed in many countries during the 1970s. Among the countries presented here, Australia (p. 46), Finland (p. 52), western Germany (p. 54), Spain (p. 63) and the United States (p. 65) belong to that group, for instance. We can only speculate that the so-called "cardiovascular revolution" (Meslé and Vallin 2006b) played an important role. It was during the 1970s that medical procedures such as bypass surgery, pace makers to treat cardiovascular diseases were introduced to larger parts of the population. But it was not only the treatment but also the prevention of cardiovascular diseases by drugs such as beta blockers that received a major boost during that time frame.

Many countries that benefited from that period effect during the 1970s exhibit a pattern that resembles a cohort effect in the years thereafter for persons aged approximately 40–80 in the 1970s. It could be argued that those green and red colors along the 45° line that last into the 2000s could be interpreted as a protective effect for those cohorts that benefited first from the new treatment and prevention methods during the 1970s. Please note that this does not imply that subsequent cohorts did not benefit from the advances of the 1970s. This would have resulted in gray cohorts areas. Instead we typically encounter positive developments, just at a smaller scale than the ones of the initial cohorts. This pattern is most visible for Japan (p. 58), Spain (p. 63), Finland (p. 52) and Australia (p. 46), and—to a lesser degree—in France (p. 53) and western Germany (p. 54).

This period effect followed by a cohort effect is not a universal finding, however. Even among western European countries, we detect some outliers. The most prominent example is probably the case of Danish women (p. 50). While the past 20 years or so have shown moderate to strong survival improvements across most of the age range as indicated by the green and red colors, there is one issue that sets Denmark apart from other countries: A cohort effect from the 1960s that lasted well into the early 1990s with stagnating survival, shown in white, or even increasing mortality as suggested by the gray shades. It has been now conclusively shown that Danish women born between the two world wars and their relatively high smoking prevalence are at the root of this cohort effect (e.g., Jacobsen et al. 2002, 2004, 2006; Lindahl-Jacobsen et al. 2016). This cohort effect coincides with relatively minor life expectancy gains among Danish women during that period. Also the United States (p. 65) features a strange pattern. It will be investigated further when we analyze rates of mortality improvement for selected causes of death in Chap. 7.

Similar to the Danish situation, modest life expectancy gains or even losses during the 1970s and 1980s have also been observed in several eastern European countries. But it has not been caused by a cohort effect as the vertical shapes for Hungary (p. 56), the Czech Republic (p. 49), Poland (p. 60) or the former GDR (p. 55) indicate a clear period effect. It can be rather expected that those countries could not (yet) reap the benefits of the cardiovascular revolution that many western countries experienced during that time period. This is supported by the subsequent strong period effects in many of those countries. The most prominent example is probably the former GDR/eastern Germany. When Germany re-unified, there was a difference of almost 3 years among women for life expectancy at birth. Just 15 years later, the difference virtually disappeared for females between the two parts of

Germany. Germany's Federal Health Reporting database (www.gbe-bund.de) can be queried to show that mortality of the circulatory system declined by 47% in eastern Germany between 1990 and 2005.

The most turbulent mortality history during the last 60 years has been probably experienced by Russia and other former Soviet republics (see Figs. 6.4, 6.17, and 6.20 on pages 48, 61, and 64). Since the 1960s, those countries (or then parts of the USSR) have seen sudden changes in mortality spikes and subsequent survival improvements. Those were typically period effects as the vertical patterns in those figures indicate. While we have only focused on mortality of women, we included the case of Russian men in Fig. 6.18 on page 62. There were a few years featuring survival improvements for instance during the mid 1980s, coinciding with Gorbachev's anti-alcohol campaign (Leon et al. 1997), life expectancy of Russian men declined by more than 5 years between 1965 and 2000. France Meslé (2004) points out in her decomposition analysis, that the majority of life years lost was due to increasing mortality from circulatory diseases and violent deaths. Those are precisely the causes, which are mainly responsible for the increase in life expectancy during the first decade of the 2000s: "Our analyses have shown that the recent improvements in life expectancy have mainly been driven by reductions in mortality from circulatory diseases and external causes" (Shkolnikov et al. 2013, p. 930).

The last few years of our observation period provide a mixed result. Life expectancy continued to increase for Russian men, primarily caused by annual survival improvements of more than 3% at ages 70 and above. Mortality declined modestly between ages 40 and 70. And there are some ages between 35 and 40 where mortality increased slightly again. But it is too early to determine whether we see another trend reversal.

Chapter 7
Surface Plots of Rates of Mortality Improvement for Selected Causes of Death in the United States

The current chapter shows how surface maps of rates of mortality improvement can also be used to analyze causes of death. This might enable researchers to gain better insights into the underlying mortality dynamics than merely looking at the Lexis surface of rates of improvement for all-cause mortality. We selected the United States for two reasons:

- Data on deaths are available as public use files since 1959 (National Center for Health Statistics 1959–2015; National Bureau of Economic Research 1959–2015). Information is included not only on age at death and sex of each deceased individual but also on cause of death and many other variables. See Chap. 3, starting on page 11, for further details about the more than 118 mio. deaths contained in the data.
- The pattern of the rates of mortality improvement for women in the United States looked different than in any of the other countries (see Fig. 6.21 on page 65). Since the late 1970s/early 1980s, the US has not experienced any prolonged period of survival improvements. Indeed, the United States gained less years of life than most other western countries during the latter part of the twentieth century. As a consequence the National Institute on Aging in the United States "requested that the National Research Council (NRC) launch a major investigation to clarify patterns in the levels and trends in international differences in life expectancy above age 50" (Crimmins et al. 2011, p. 2).

We used again the same techniques and color schemes as in Chap. 6. To avoid any spurious conclusions due to small numbers of deaths, we excluded deaths above age 95 and below age 20.
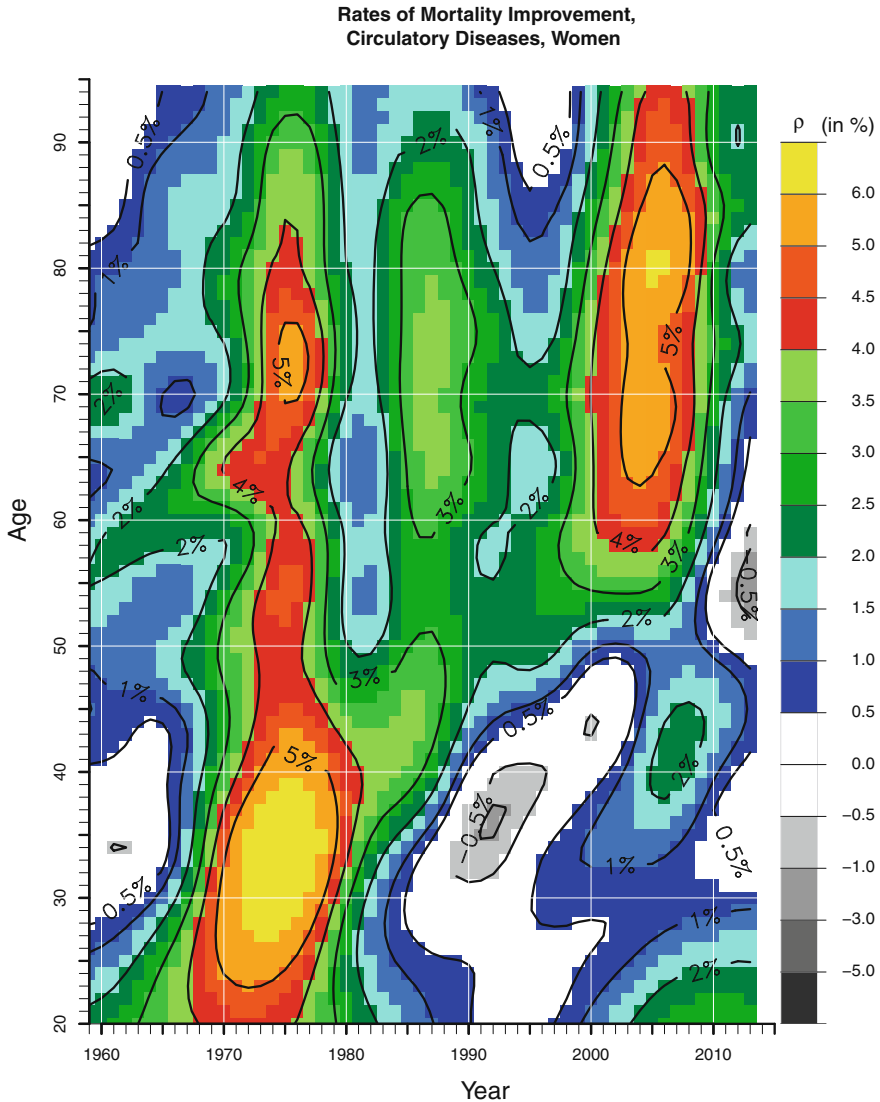
Fig. 7.1 Rates of mortality improvement for all circulatory diseases for women in the United States aged 20–95 between 1959 and 2013 (Data source: Human Mortality Database, National Center for Health Statistics, and National Bureau of Economic Research)

More than 50 mio. deaths—corresponding to almost 45% of all deaths—can be attributed to diseases of the circulatory system. The ROMI plot for mortality due to these causes is depicted in Fig. 7.1. Heart diseases (Fig. 7.2), e.g., myocardial infarction, and cerebrovascular diseases (Fig. 7.3) such as stroke constitute about
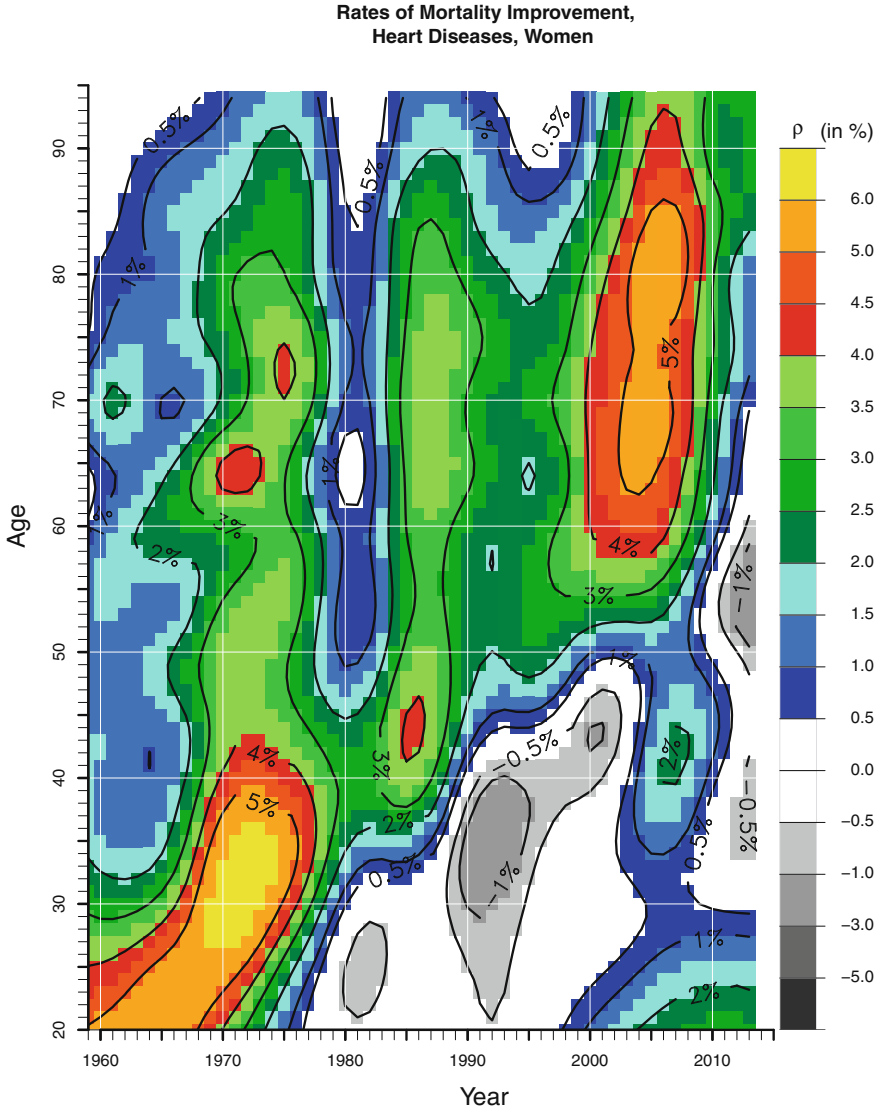
Fig. 7.2 Rates of mortality improvement for heart diseases for women in the United States aged 20–95 between 1959 and 2013 (Data source: Human Mortality Database, National Center for Health Statistics, and National Bureau of Economic Research)

Fig. 7.3 Rates of mortality improvement for cerebrovascular diseases for women in the United States aged 20–95 between 1959 and 2013 (Data source: Human Mortality Database, National Center for Health Statistics, and National Bureau of Economic Research)

95% of all deaths from circulatory diseases. We can draw at least two conclusions from those figures:

- Circulatory diseases can not be the reason why (female) life expectancy in the United States barely increased during the last two decades of the twentieth century. We see major annual declines (three percent and more) in mortality due to these causes.
- The pattern found for mortality from heart diseases and cerebrovascular diseases as well as from the composite picture of all circulatory diseases resembles the pattern we found in Chap. 6 for rates of mortality improvement from all causes in many countries such as Spain, Japan, or Italy. At that time we were only able to speculate that the "cardiovascular revolution" was the primary reason for the observed pattern. While Figs. 7.1, 7.2, and 7.3 are no definite proof, we can feel more certain about our suggestions.

So if circulatory diseases were the main reason for life expectancy gains in many European countries during the 1980s and 1990s, why did life expectancy in the United States not increase in a similar manner since mortality from heart diseases, stroke and similar causes also declined remarkably in the US?

If circulatory diseases can be excluded, we turned our attention to malignant neoplasms ("cancers"). They are responsible for more than one in five deaths. Among the various cancer sites, we decided to look at three major sub-categories: colorectal, breast and lung cancer (Figs. 7.5, 7.6, 7.7, and 7.8) in addition to mortality from all cancers (Fig. 7.4).

Deaths from any kind of cancer for women (Fig. 7.4) show a mixed pattern: Below age 50 we can detect a continuous trend of improving survival conditions throughout most of our observation period. Lower mortality from cancer extends also to higher and higher ages after the mid-1980s (Fig. 7.4). Those survival improvements that show some characteristics of a cohort effect could be influenced by declining mortality from colorectal cancers as suggested by Fig. 7.5. Also breast cancer (Fig. 7.6) displays steady improvements albeit starting only in the 1990s. The main cause for the poor development of female life expectancy during the late twentieth century is probably lung cancer. Among the authors of this book, Fig. 7.7 on page 77 is the strongest cohort effect they have encountered when analyzing rates of mortality improvement by cause of death. Also men (Fig. 7.8, p. 78) feature such a strong cohort effect. The pattern for males is located further left on the Lexis map, i.e., earlier in calendar time, supporting the idea of the "'cigarette diffusion' explanation [...] that convergence in male and female smoking is the byproduct of a female lag in the process of cigarette adoption, diffusion, and abatement" (e.g., Pampel 2001, p. 388). Furthermore, our figures on lung cancer, in conjunction with the detrimental effects shown in Fig. 7.9 for respiratory diseases, are in line with
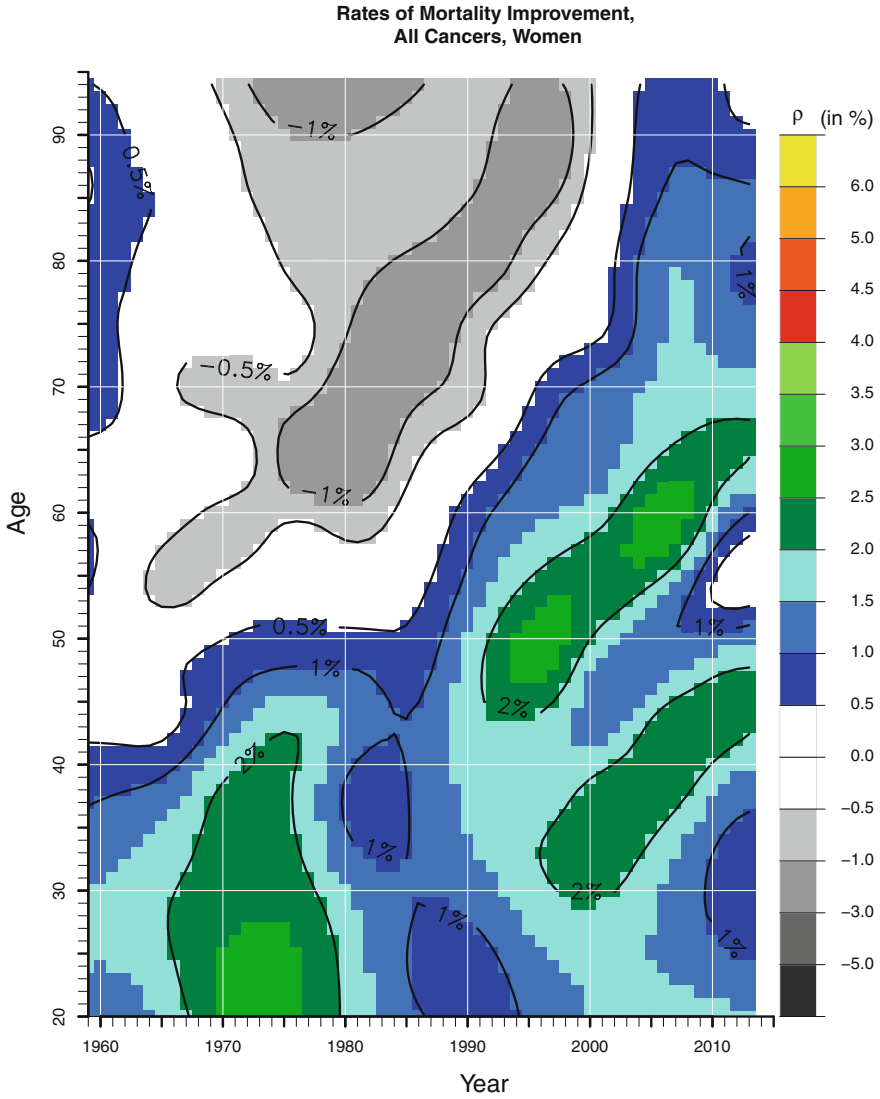
Fig. 7.4 Rates of mortality improvement for malignant neoplasms for women in the United States aged 20–95 between 1959 and 2013 (Data source: Human Mortality Database, National Center for Health Statistics, and National Bureau of Economic Research)
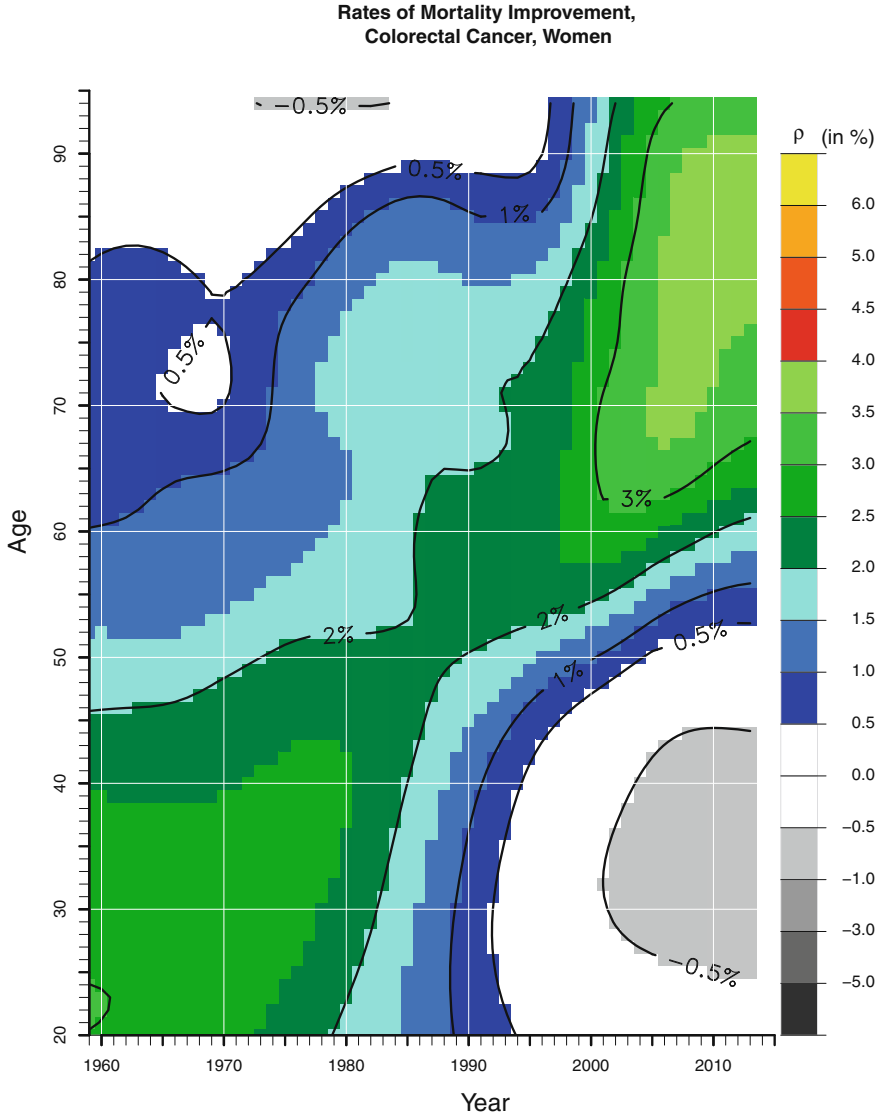
Fig. 7.5 Rates of mortality improvement for colorectal cancer for women in the United States aged 20–95 between 1959 and 2013 (Data source: Human Mortality Database, National Center for Health Statistics, and National Bureau of Economic Research)
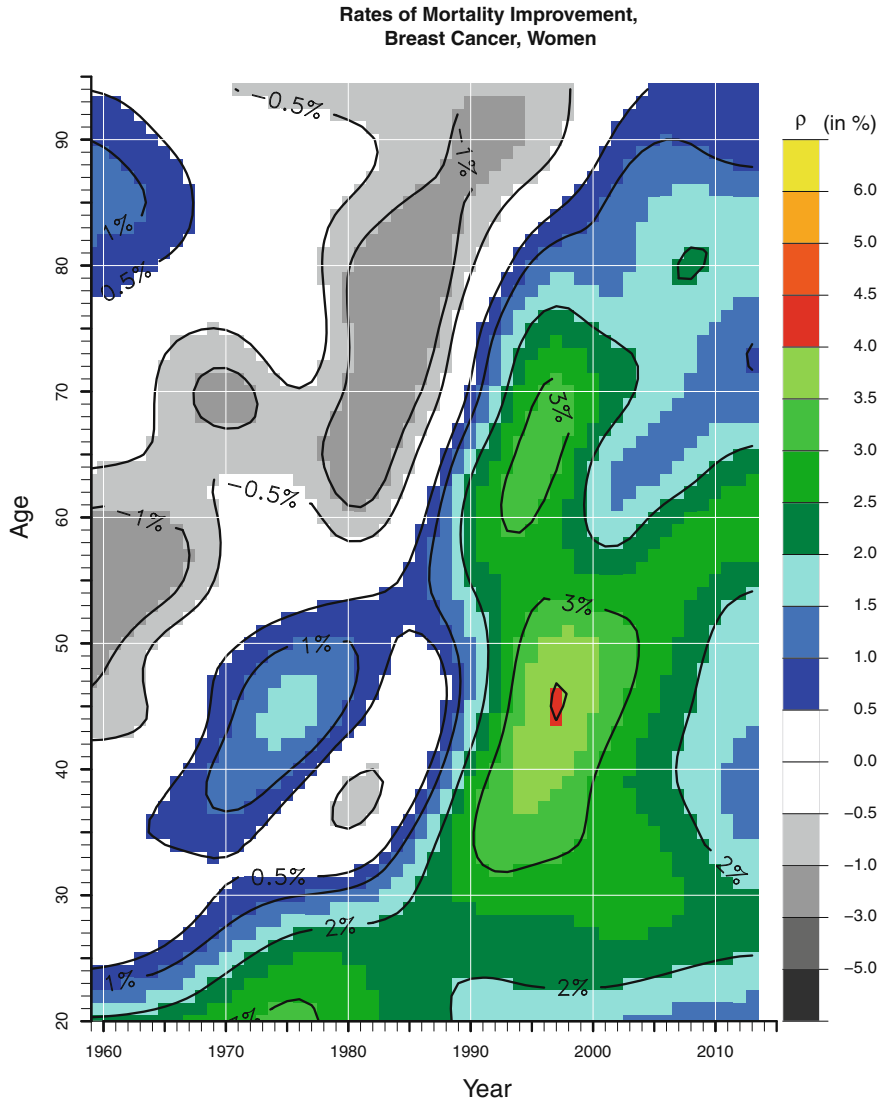
Fig. 7.6 Rates of mortality improvement for breast cancer for women in the United States aged 20–95 between 1959 and 2013 (Data source: Human Mortality Database, National Center for Health Statistics, and National Bureau of Economic Research)
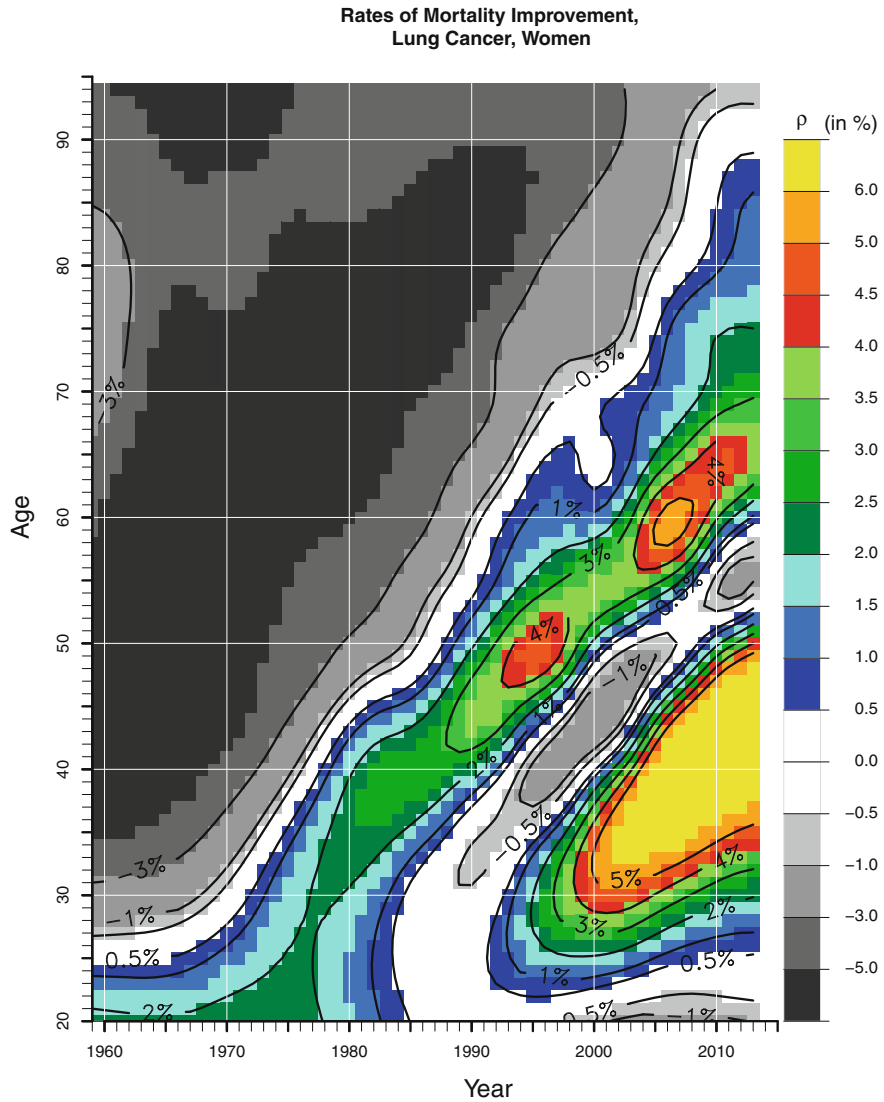
Fig. 7.7 Rates of mortality improvement for lung cancer for women in the United States aged 20–95 between 1959 and 2013 (Data source: Human Mortality Database, National Center for Health Statistics, and National Bureau of Economic Research)
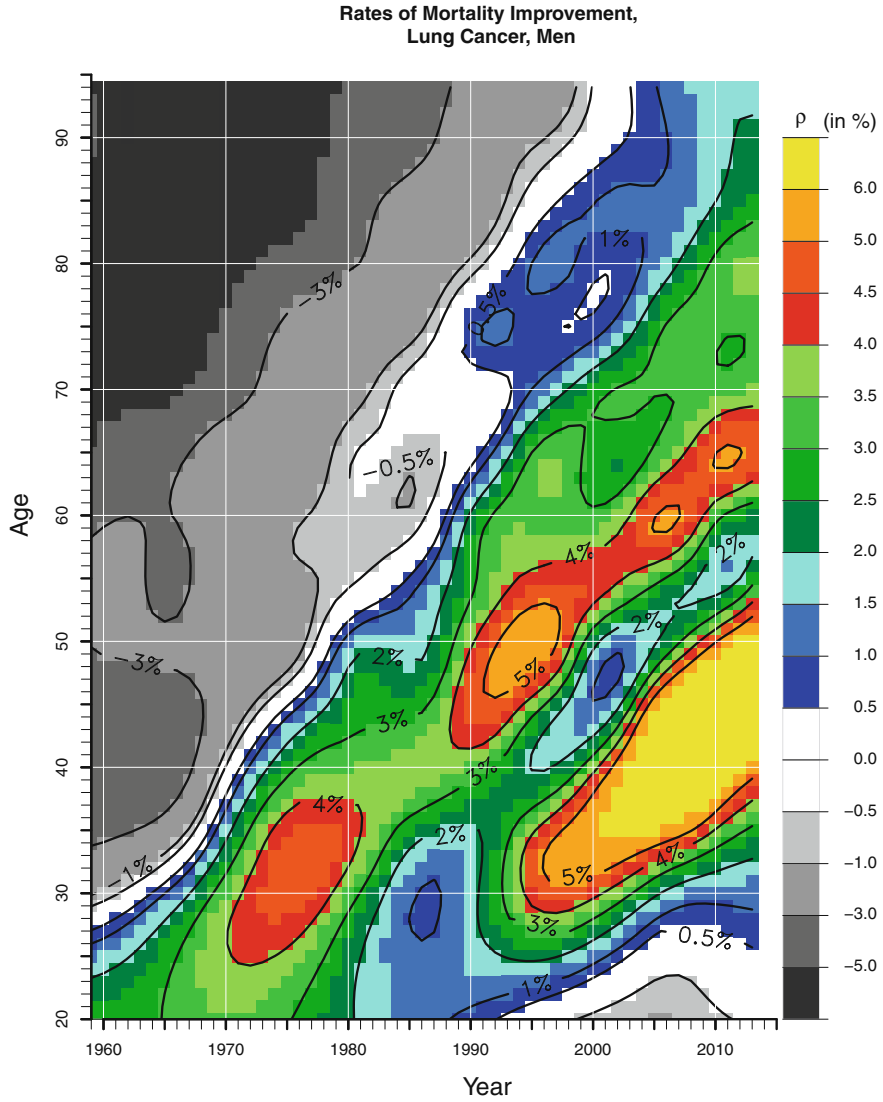
Fig. 7.8 Rates of mortality improvement for lung cancer for men in the United States aged 20–95 between 1959 and 2013 (Data source: Human Mortality Database, National Center for Health Statistics, and National Bureau of Economic Research)
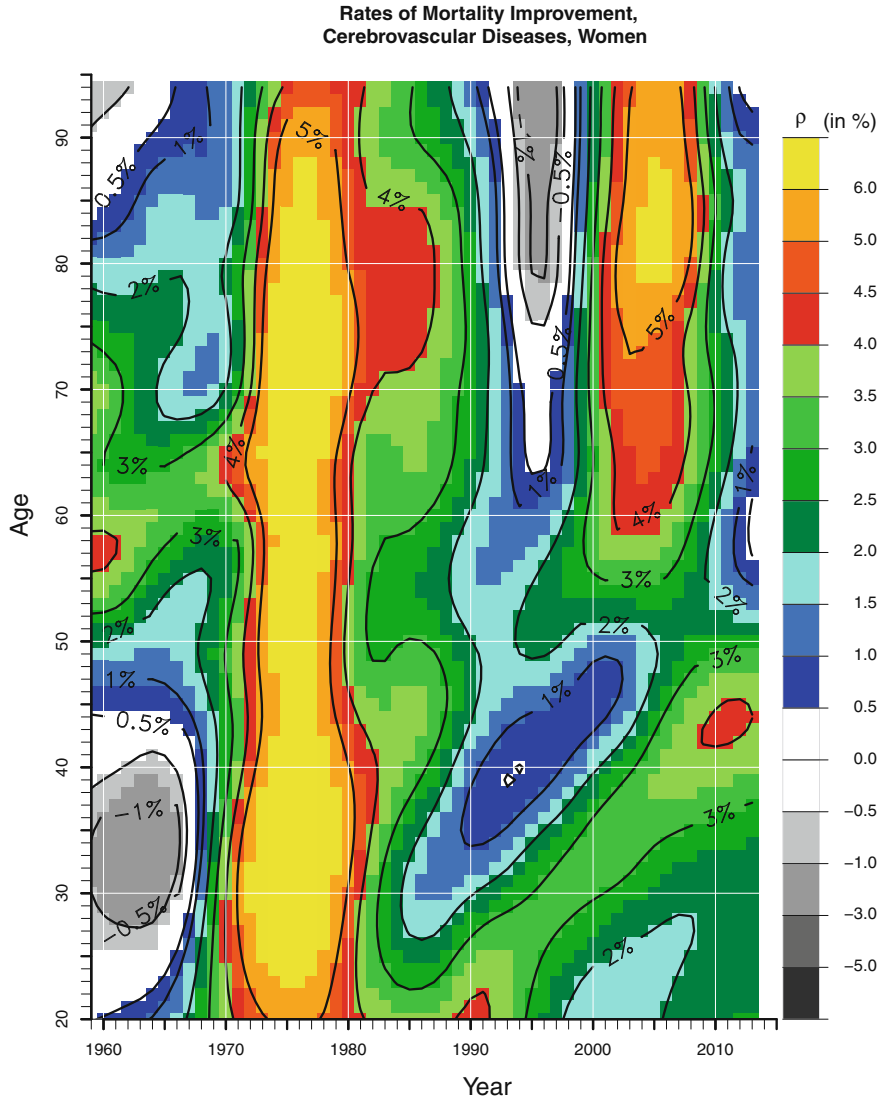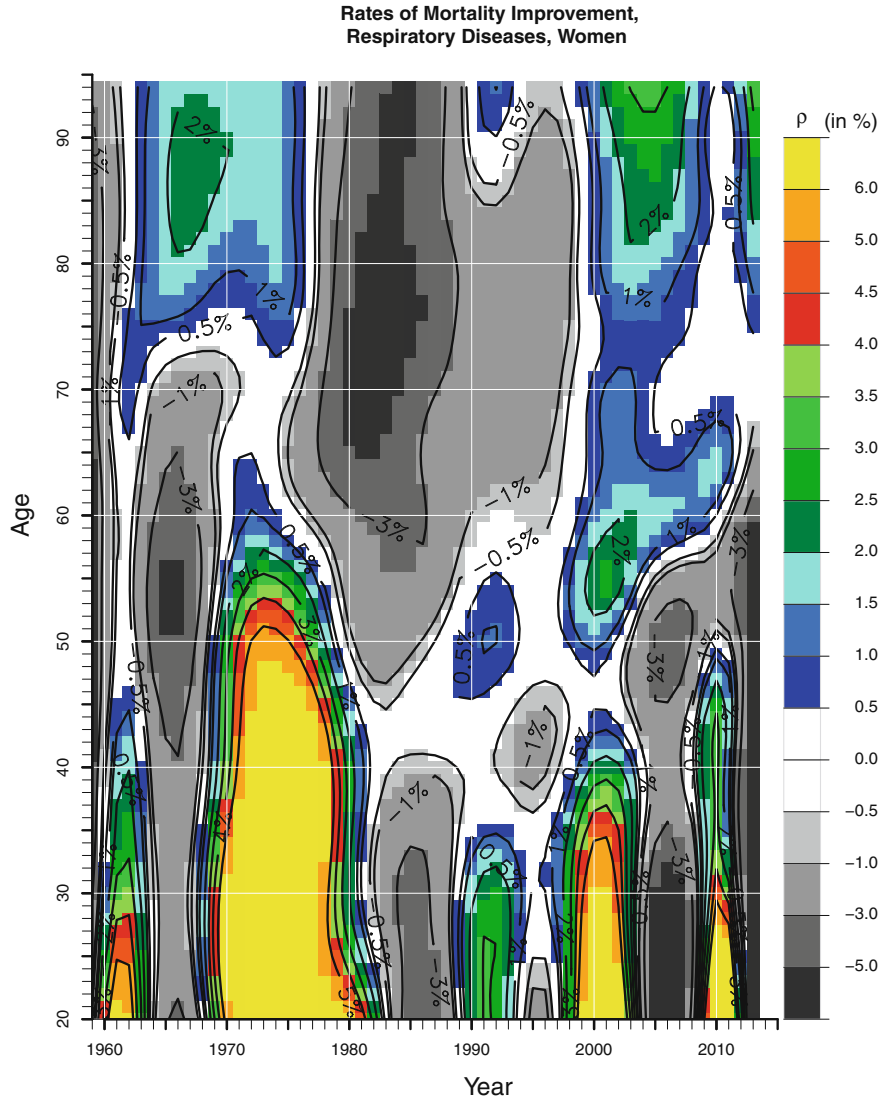
Fig. 7.9 Rates of mortality improvement for respiratory diseases for women in the United States aged 20–95 between 1959 and 2013 (Data source: Human Mortality Database, National Center for Health Statistics, and National Bureau of Economic Research)

Wang and Preston (2009, p. 398) who argue that "[b]ecause of changes in smoking behavior that have already occurred or that can be reliably projected, American mortality is likely to fall more rapidly than is commonly anticipated."

Chapter 8
Surface Plots of Age-Specific Contributions to the Increase in Life Expectancy

8.1 How to Estimate Age-Specific Contributions to the Change in Life Expectancy

Different perspectives can provide different insights into mortality dynamics. Chapters 6 and 7 investigated the relative change of death rates over time. The same "ROMI" does not necessarily translate to the same change of life expectancy, though, neither over time nor at different ages: A large reduction of infant mortality in the past had a major impact on life expectancy whereas the same proportional reduction would affect life expectancy only slightly since infant mortality is already (and thankfully) at a very low level. Analogously, the same rate of mortality improvement at the same time may have considerably different effects on life expectancy. For instance, an annual mortality decline by x per cent at age 80 has a much larger impact on life expectancy than a decline by x per cent at age 100.

We decided therefore to estimate the age-specific contributions to the change in life expectancy. Among the various methods available—see Canudas-Romo (2003) for an overview—we applied the approach of Arriaga (1984) using the exposition of Preston et al. (2001, pp. 64–65). Having data for single ages available, allowed us to further simplify the notation. With the conventional life table l_x for the life table survivors at age x, L_x for the number of life years lived at age x, and T_x for the number of life years lived at age x and above, we can estimate Δ_x, the contribution of mortality at age x to differences in life expectancy between two points (or between any two life tables), denoted by superscripts [1] and [2] as:

$$\Delta_x = \frac{l_x^1}{l_0^1} \cdot \left(\frac{L_x^2}{l_x^2} - \frac{L_x^1}{l_x^1} \right) + \frac{T_{x+1}^2}{l_0^1} \left(\frac{l_x^1}{l_x^2} - \frac{l_{x+1}^1}{l_{x+1}^2} \right)$$

The age-specific contribution to the difference in life expectancy, Δ_x, consists of two parts. The first part (until the + sign) estimates the *direct* effect, i.e., the

change in life expectancy only due to the change in mortality at this given age x. The second part is the sum of an indirect effect and an interaction effect (Preston et al. 2001, p. 64). A geometric explanation might help to understand what is meant by this second component as it often appears to be confusing: Life expectancy can be interpreted as the area under the survival curve. In case of a decline in mortality at age x at time point t the survival curve at this age is higher than at time point $t - 1$. This is meant by the direct effect. For the sake of simplicity, let's assume that mortality only changed at age x. Nevertheless, the survival curve will be also higher at age $x + 1$: A survival curve where the survival at age $x + 1$ was at the same level as at $t - 1$ would require an increase in mortality. This "wake" of a change in mortality at one age affecting the survival function at subsequent ages is estimated by the second component.

We followed exactly the same procedure as in Chap. 5 to obtain the required death rates: Raw death rates, based on death counts and corresponding exposure times from the Human Mortality Database (HMD, 2017), were smoothed assuming Poisson distributed death counts using Camarda's `MortalitySmooth` package (Camarda 2012, 2015). The life table functions l_x, L_x, and T_x were estimated using the approach outlined in Chapter 3 of Preston et al. (2001). The values for a_x, the mean duration lived at age x by those who died at age x, were taken from the HMD.

While any kind of difference in calendar time could be used, we decided to estimate the age-specific contributions within an interval of ten years. I.e., we compared 1960 to 1950, 1961 to 1951, ... The years on the x-axis in Figs. 8.1, 8.2, 8.3, 8.4, 8.5, 8.6, 8.7, 8.8, 8.9, 8.10, 8.11, 8.12, and 8.13 refer to the *latter* time point. Thus, the values at any age x in year 1980 denote the contribution of changing mortality at age x between 1970 and 1980. The choice of a ten year difference is, of course, arbitrary but it allowed us also to express the contribution in "meaningful" units: We used days and weeks and—in exceptional cases of substantial improvements or deterioriation in survival—months. The surface maps were plotted using a terrain color scheme: Green indicates moderate contributions to life expectancy. When the color turns to brown, that age alone contributed at least one week to the increase in life expectancy during the decade of observation. Very bright brown areas depict contributions of one month or more. Blue colors denote negative contributions. Just like deeper shades of blue suggests lower depths below sea level on geographic maps, they indicate here changes in age-specific mortality that bring life expectancy down.

Again, we have not included the whole set of countries from the HMD but rather a subset of countries with rather peculiar features, which we already pointed at in previous chapters.
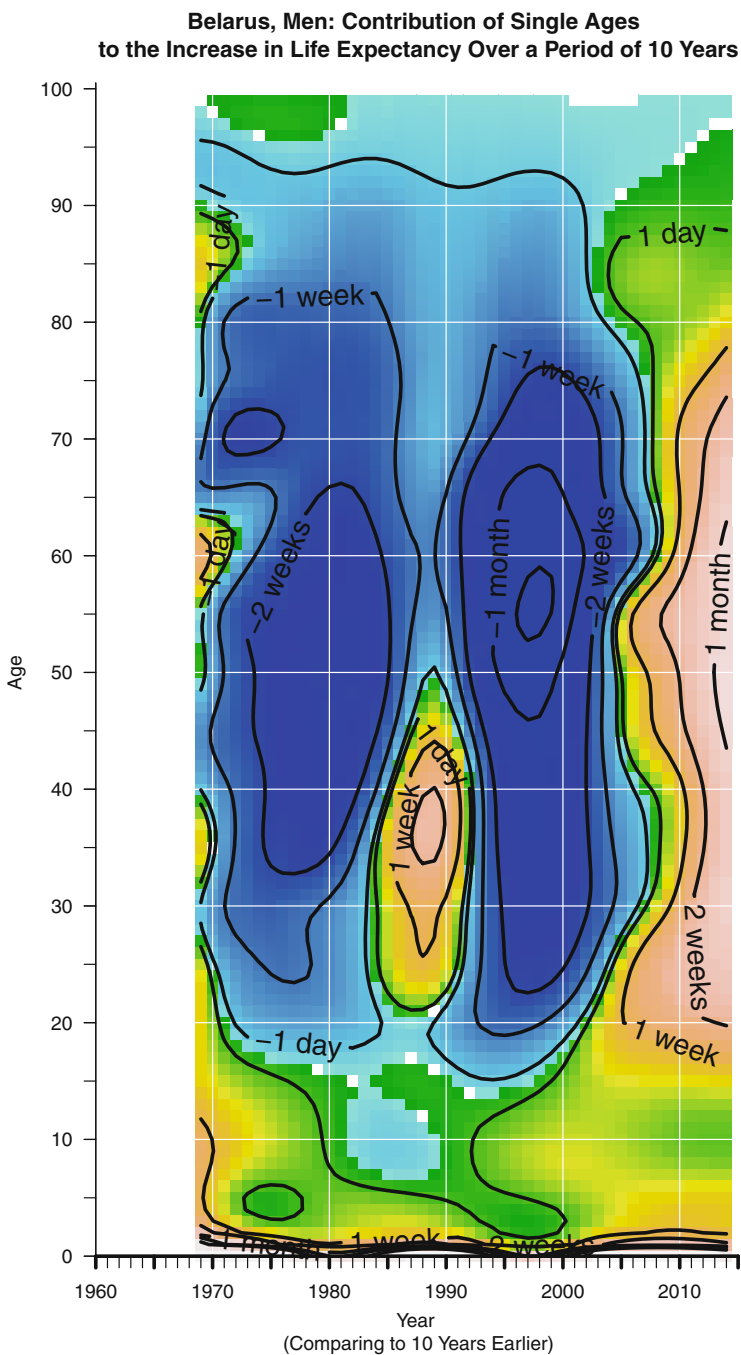
Fig. 8.1 Age-specific contributions to the increase in life expectancy among men during the past 10 years in Belarus, 1969–2014 (Data source: Human Mortality Database)
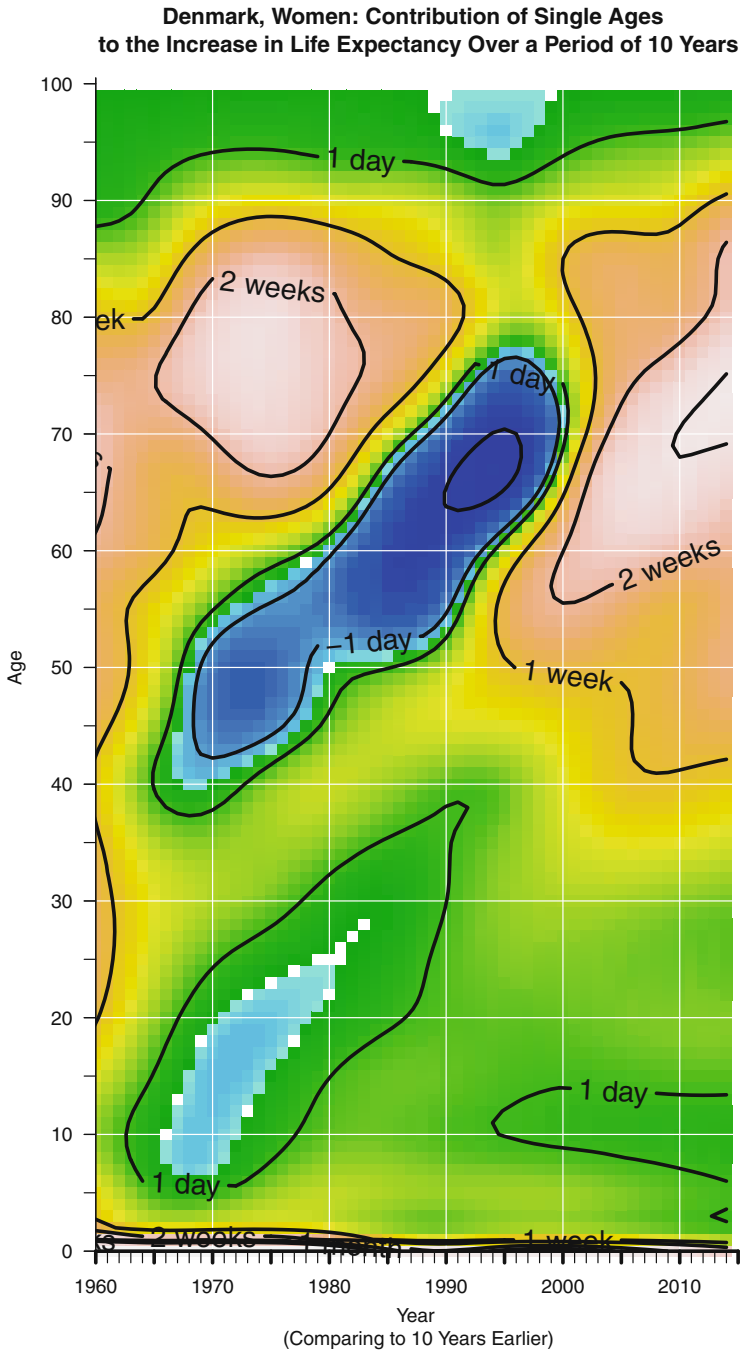
Fig. 8.2 Age-specific contributions to the increase in life expectancy among women during the past 10 years in Denmark, 1960–2014 (Data source: Human Mortality Database)
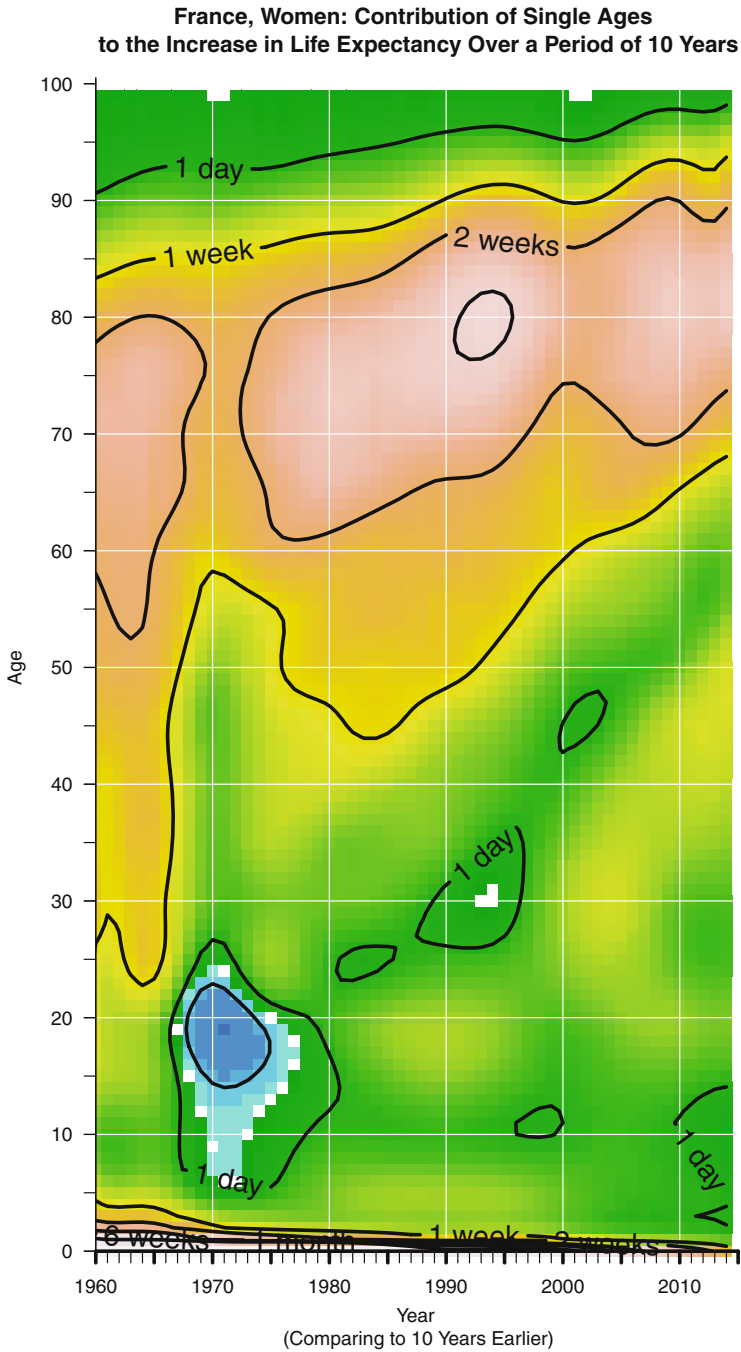
Fig. 8.3 Age-specific contributions to the increase in life expectancy among women during the past 10 years in France, 1960–2014 (Data source: Human Mortality Database)
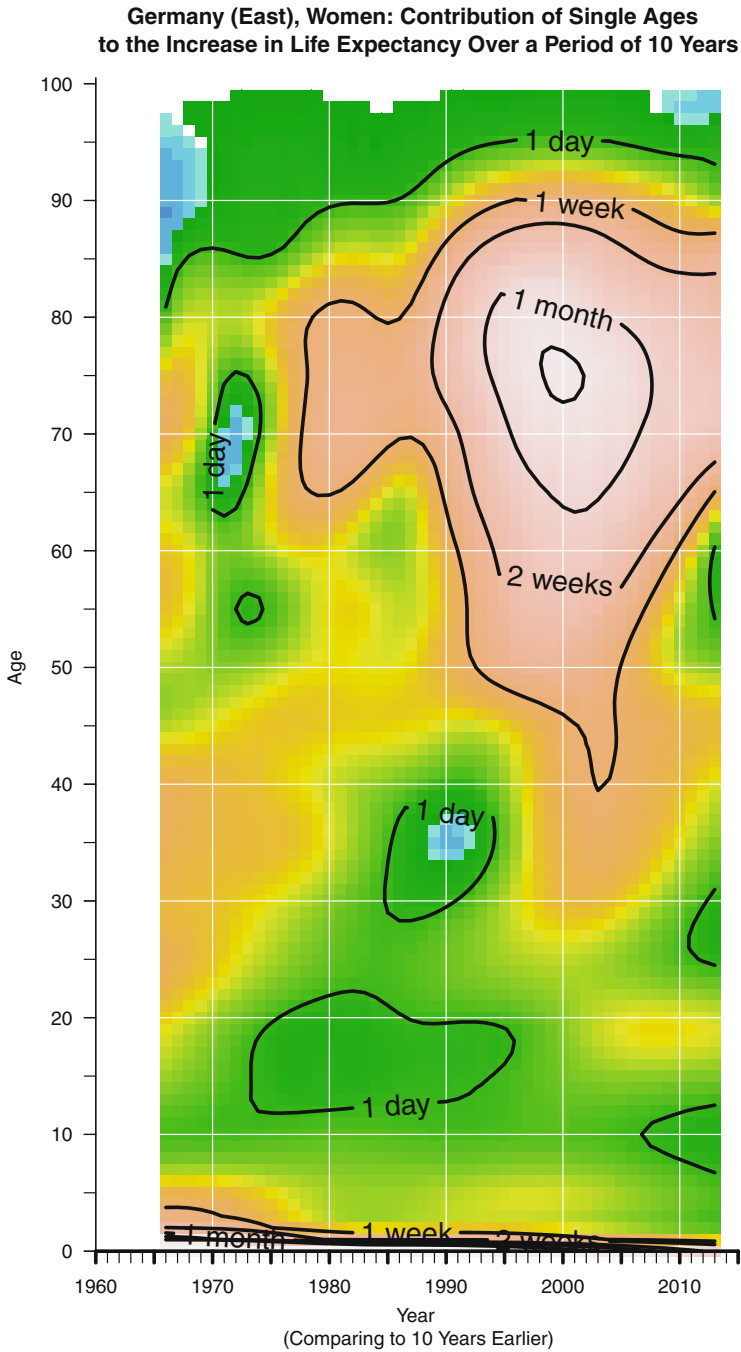
Fig. 8.4 Age-specific contributions to the increase in life expectancy among women during the past 10 years in Germany (East), 1966–2013 (Data source: Human Mortality Database)
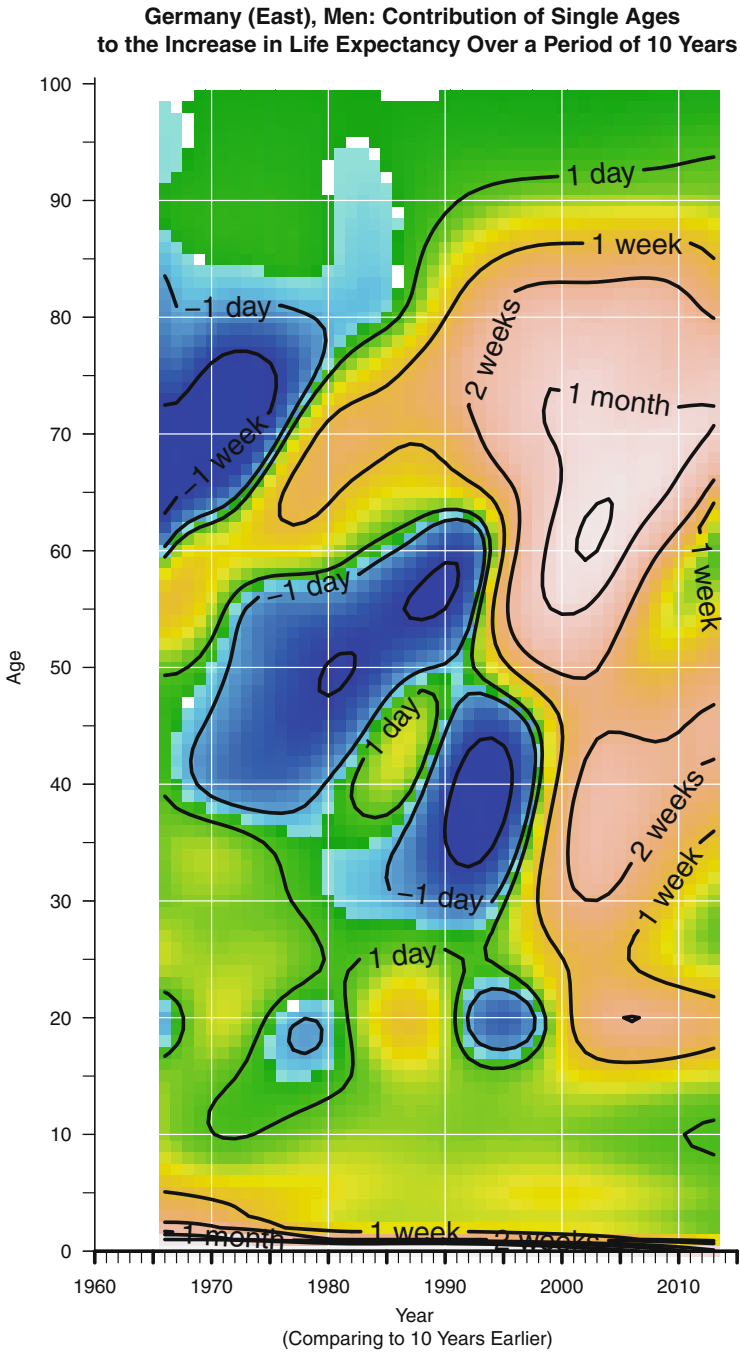
Fig. 8.5 Age-specific contributions to the increase in life expectancy among men during the past 10 years in Germany (East), 1966–2013 (Data source: Human Mortality Database)
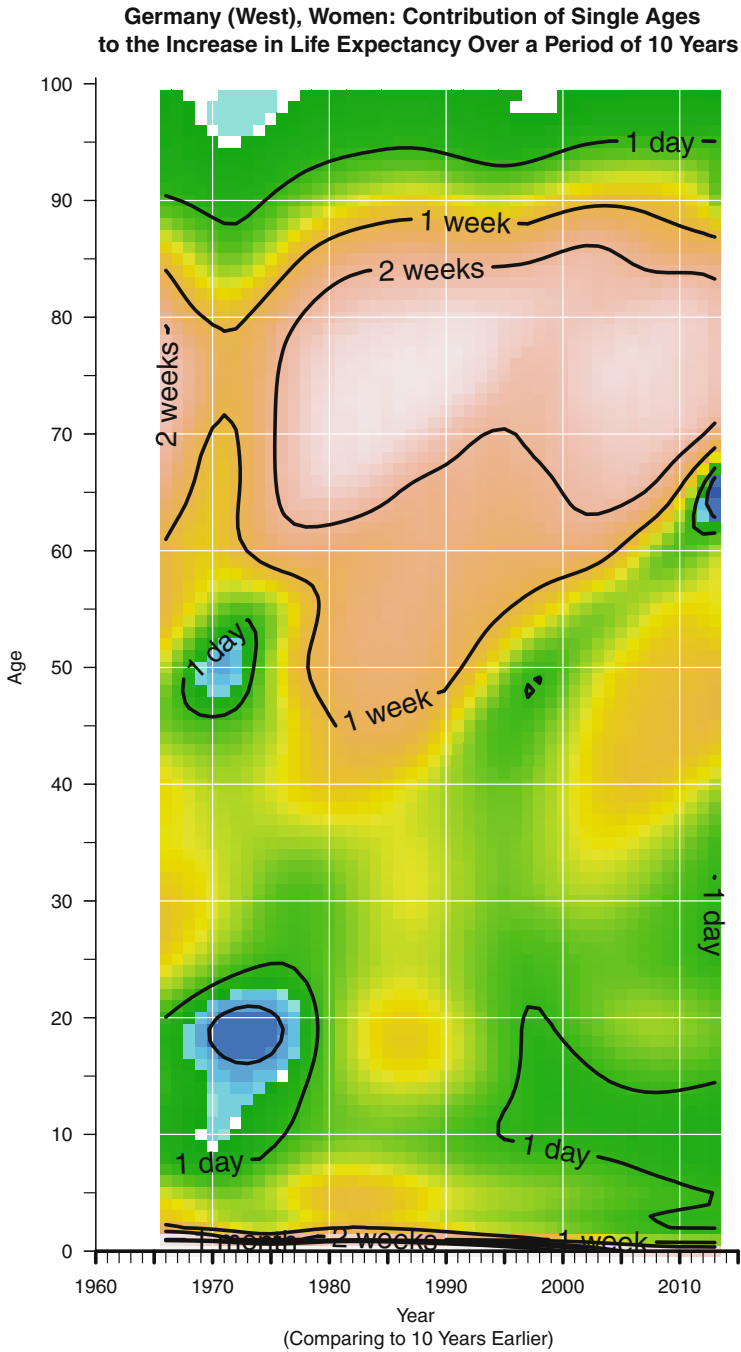
Fig. 8.6 Age-specific contributions to the increase in life expectancy among women during the past 10 years in Germany (West), 1966–2013 (Data source: Human Mortality Database)
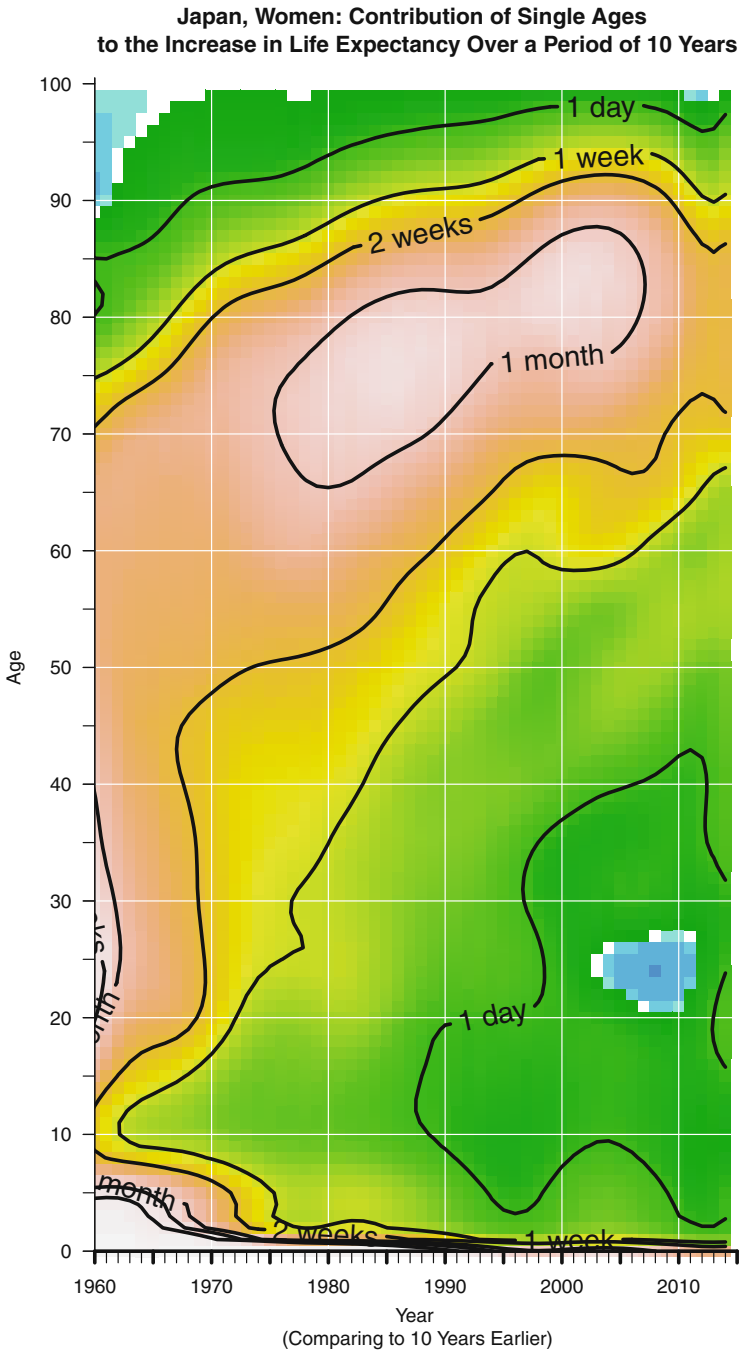
Fig. 8.7 Age-specific contributions to the increase in life expectancy among women during the past 10 years in Japan, 1960–2014 (Data source: Human Mortality Database)
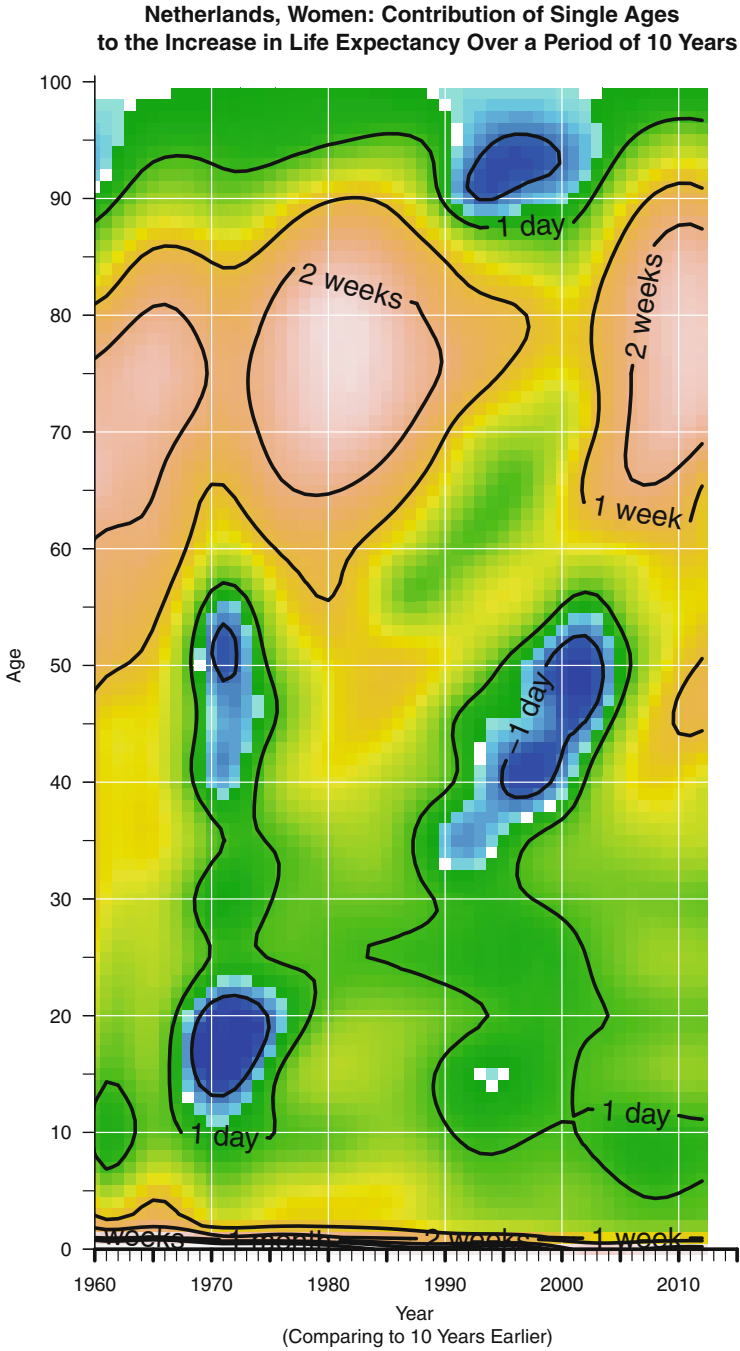
Fig. 8.8 Age-specific contributions to the increase in life expectancy among women during the past 10 years in Netherlands, 1960–2012 (Data source: Human Mortality Database)
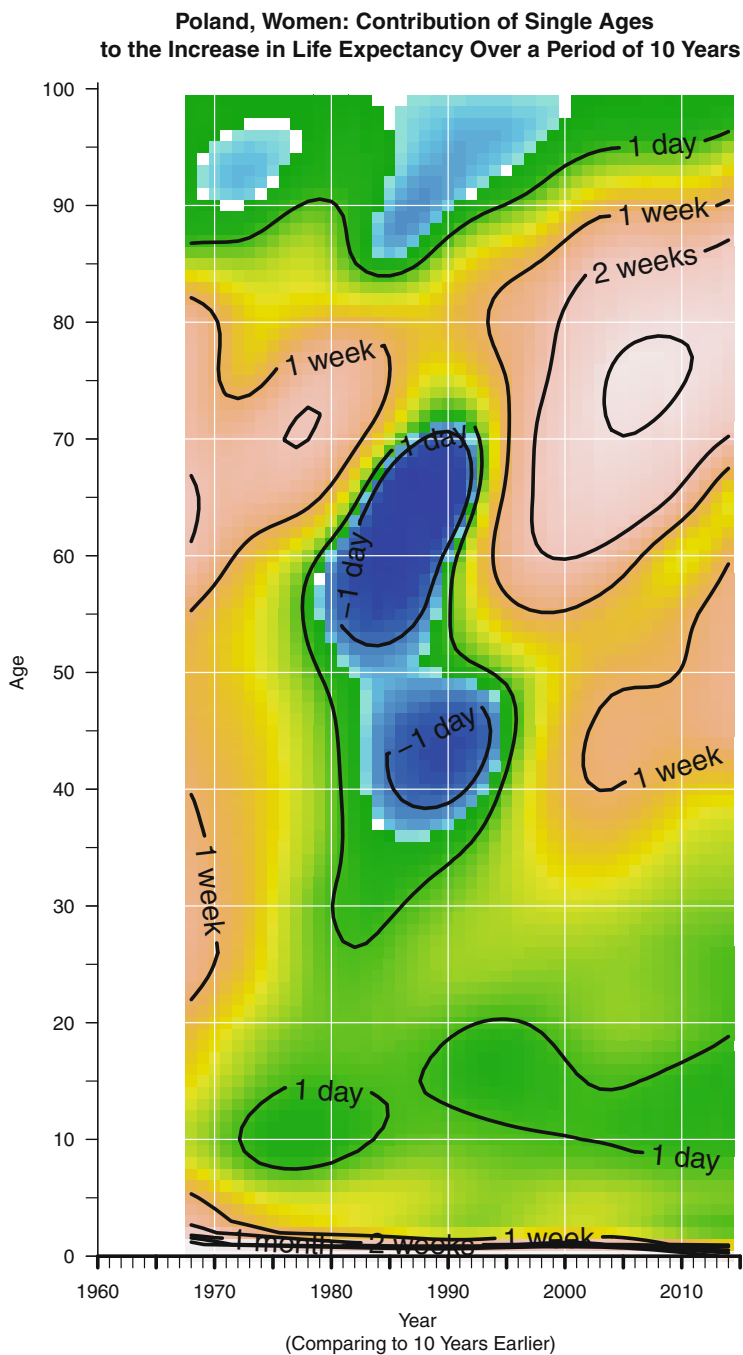
Fig. 8.9 Age-specific contributions to the increase in life expectancy among women during the past 10 years in Poland, 1968–2014 (Data source: Human Mortality Database)
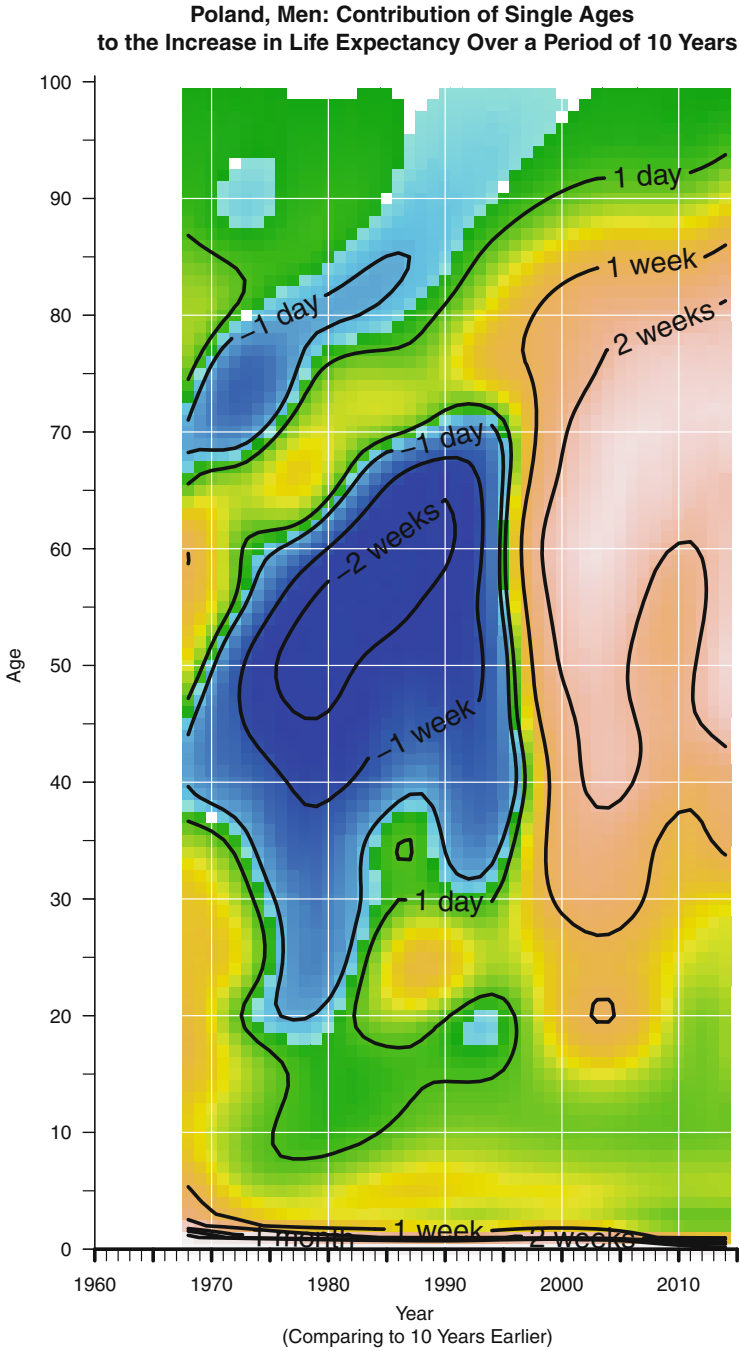
Fig. 8.10 Age-specific contributions to the increase in life expectancy among men during the past 10 years in Poland, 1968–2014 (Data source: Human Mortality Database)
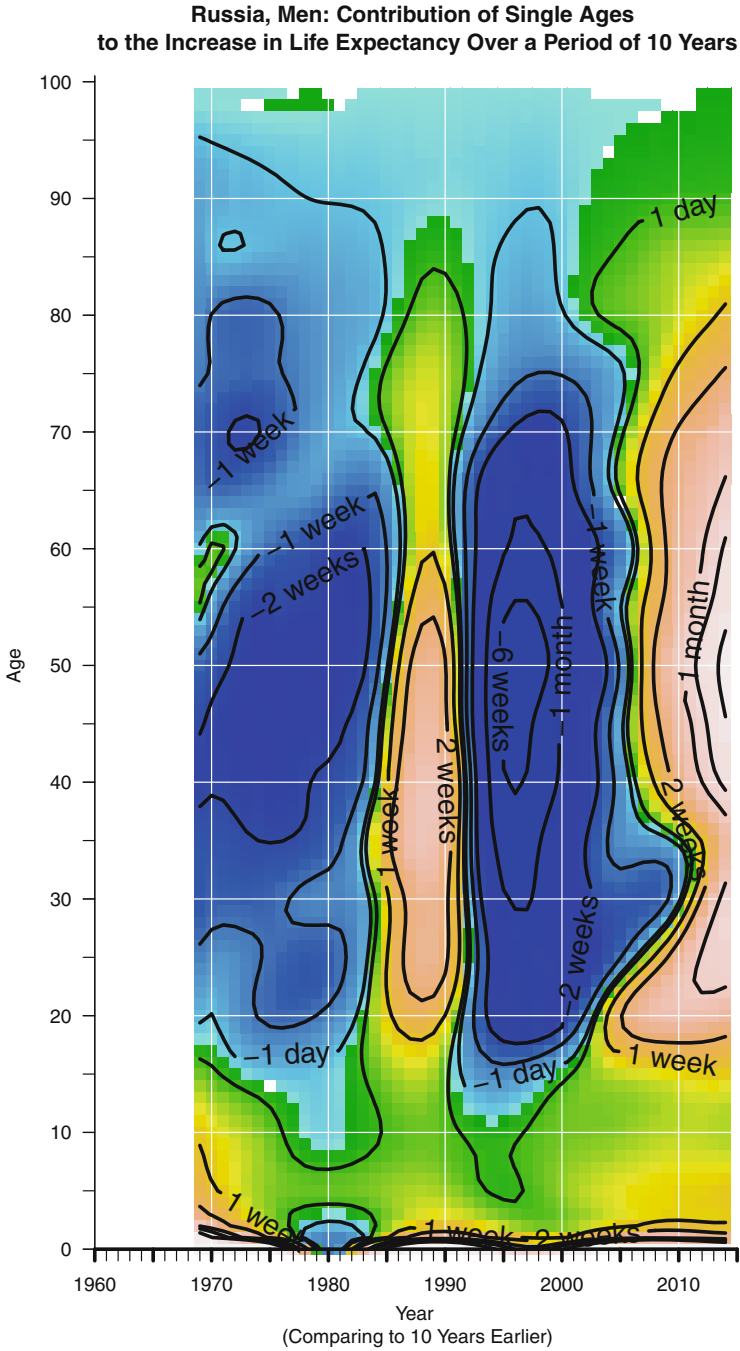
Fig. 8.11 Age-specific contributions to the increase in life expectancy among men during the past 10 years in Russia, 1969–2014 (Data source: Human Mortality Database)

**Sweden, Women: Contribution of Single Ages
to the Increase in Life Expectancy Over a Period of 10 Years**
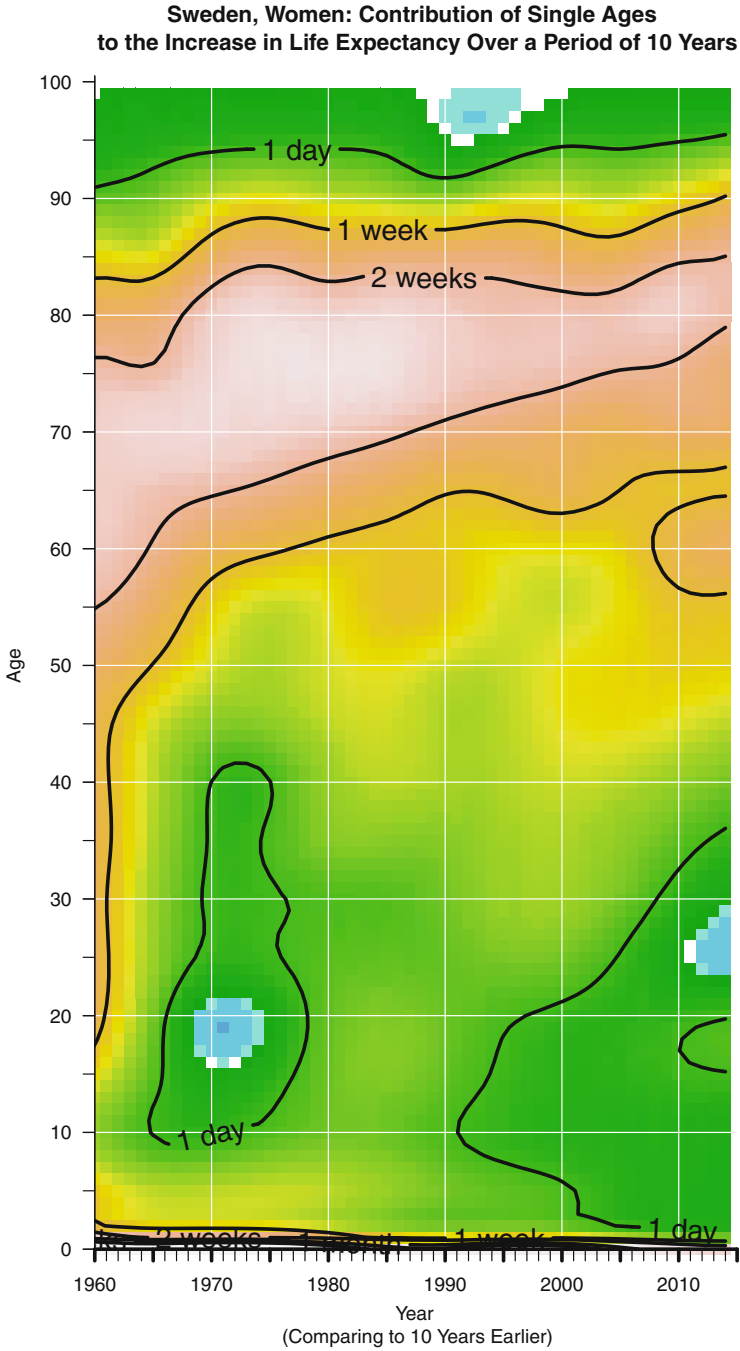
Fig. 8.12 Age-specific contributions to the increase in life expectancy among women during the past 10 years in Sweden, 1960–2014 (Data source: Human Mortality Database)

**USA, Women: Contribution of Single Ages
to the Increase in Life Expectancy Over a Period of 10 Years**
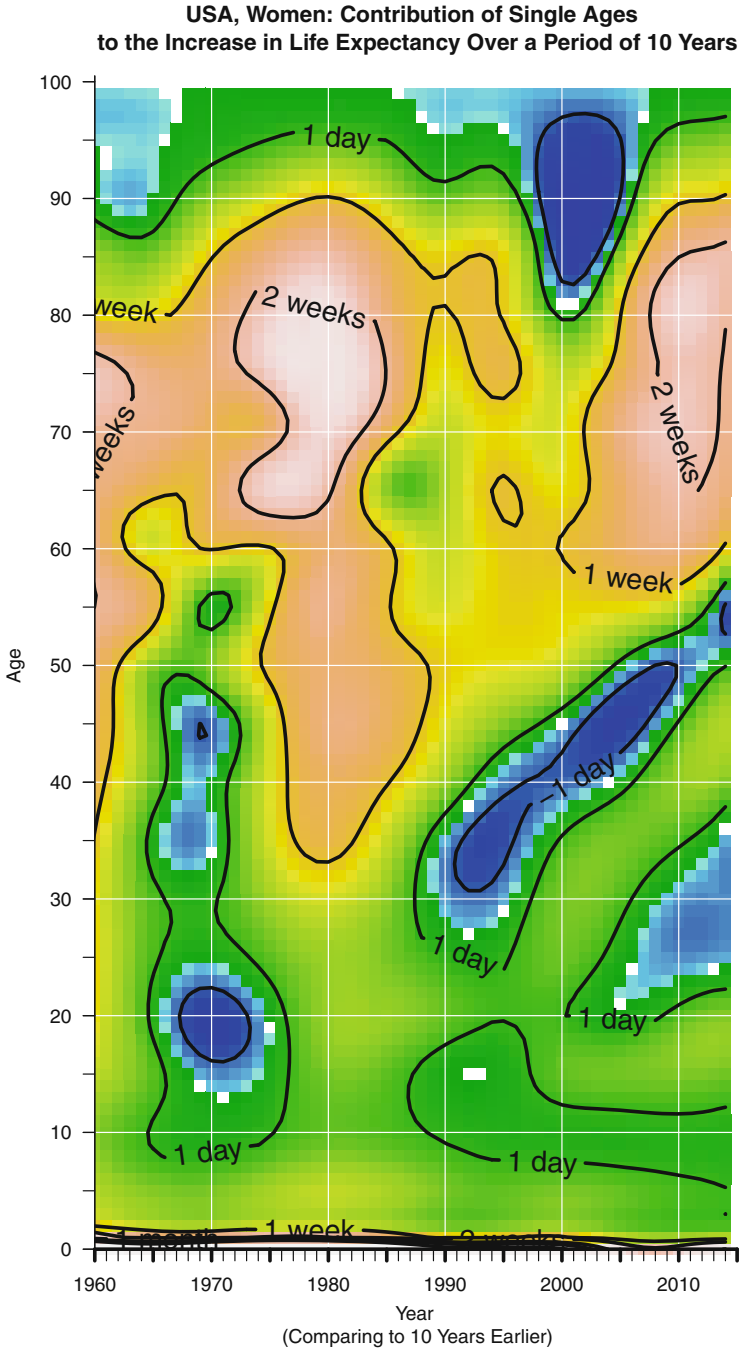
Fig. 8.13 Age-specific contributions to the increase in life expectancy among women during the past 10 years in USA, 1960–2014 (Data source: Human Mortality Database)

8.2   Results

Figures 8.1 and 8.11 for Belarus and Russia, respectively, reiterate our findings of strong period effects from Chap. 6. Although our focus is mainly on mortality dynamics of women, we selected data for men here on purpose since the decline in life expectancy and fluctuations over time were more pronounced for males than for females (e.g., Meslé 2004).

The vertical ROMI patterns in Fig. 6.18 (Chap. 6, p. 62) suggested that all ages between 15 and 75 were affected by the strong positive and negative period effects in Russia. Figure 8.11 in the present chapter, though, allows us to narrow down the age-range if we are interested in the contribution to changes in life expectancy. Compared to ten years earlier, changing mortality of men aged between 20 and 50 years appears to be the main contributor to the increase in life expectancy during the 1980s, fueled at least partly by Gorbachev's anti-alcohol campaign (Leon et al. 1997). As we have already seen in Chap. 6, the end of the Soviet Union in the early 1990s induced a major rise in mortality in Russia and other successor states. It seems almost impossible that mortality increased as much at ages 50 to 65 in Belarus (Fig. 8.1) that some single ages depressed life expectancy by one month or more within a ten-year interval. Even more astonishing are the results for Russia (Fig. 8.11) where the change in mortality at single ages between 40 and 55 caused a decline of life expectancy of six weeks and more.

The end of socialism/communism in eastern Europe in the early 1990s was less of a problem for Poland, though, serving as an example of a country from the former Warsaw Pact (see Fig. 8.9 for women and Fig. 8.10 for men). Whereas mortality also increased for men at working ages throughout the 1970s and 1980s, it took only a few years after the fall of the iron curtain, to see exactly the same kind of ages contributing two weeks or more to gains in life expectancy throughout a decade—a development, which appears to be still ongoing. It took even less time for Polish women to benefit from the regime change than for their male peers. The increase in life expectancy almost immediately after 1989/1990 was primarily triggered by survival improvements among women aged 60–85 years.

A similar picture as for Poland emerges for females and males from the former "GDR" (Figs. 8.4 and 8.5): Stagnating or even increasing mortality throughout the 1970s and 1980s among men at working ages does not immediately disappear with the end of the political regime. Indeed, mortality even increased slightly for males aged about 30–50 years. Marc Luy (2004, p. 133) showed that "[t]his effect can be attributed almost exclusively to diseases of the digestive system (mainly due to diseases of the liver) and the cause of death chapter 'injury, poisoning and certain other consequences of external causes' (mainly resulting from traffic accidents)." The group that was the fastest to adapt to the new situation were German women from the former eastern part. Faster than Polish women or men from eastern Germany, improvements in survival started immediately in 1990. Declining mortality where single ages contributed at least two weeks to the increase in life expectancy within ten years were representative of the first two decades after

Germany's reunification. Women aged 65 to 80 contributed even one month or more from the mid-1990s to the mid-2000s.

A contribution of two weeks or more of single ages to the increase in life expectancy was already common among women in the former western part of Germany since the 1970s (Fig. 8.6). In fact, the peak of one month and more for women aged 65 to 80 could be interpreted as an indicator for the catching-up period of the "cardiovascular revolution" that already started in the 1970s in the former FRG and many other western countries—see, for instance, the figures for French and Swedish women in Figs. 8.3 and 8.12.

Sweden was actually one of the first countries with a sustained decline in old-age mortality as pointed out by Kannisto (1994). As reflected by the narrowing bands of two weeks and more in Fig. 8.12, contributions of older ages to the increase in life expectancy have been smaller than in some other "vanguard" countries such as France or Japan. Drefahl et al. (2014) demonstrate that different trends for mortality from circulatory diseases were the main reason that Sweden is "losing ground".

Once again we can detect a clear cohort effect in Denmark (Fig. 8.2) for the women born between the two world wars. While the blue colors indicate worsening survival, the detrimental effects were just a few days at most for single ages, much less than what we observed for Belarus or Russia (Figs. 8.1 and 8.11). As we have shown in previous chapters, the United States (Fig. 8.13) also deviated negatively from the international trend observed in many western countries. We can expect that the seemingly interrupted pattern between 1980 and 2000 can be attributed to the severe effects of lung cancer, which we demonstrated in Chap. 7.

Another country where life expectancy improvements were not as high as anticipated during the 1980s and 1990s were the Netherlands (Fig. 8.8). The typical explanation of the smoking epidemic and lung cancer does not hold here, though. Peters (2015, p. 185), for example, argues that "[t]he internationally deviating Dutch trends over the past three decades are not explained by changes in the impact of smoking. Accounting for the impact of smoking revealed simultaneous trend breaks in mortality decline of Dutch men and women around 2002. These breaks occurred most likely due to sudden changes in healthcare expenditures that explained about half of the acceleration in life expectancy during 2000–2009."

Chapter 9
Seasonality of Causes of Death

9.1 Decomposing Seasonal Data

The majority of deaths in most countries can be attributed to causes that feature a distinct seasonal pattern. Figure 9.1 depicts the relative monthly frequencies of nine selected causes of death in the United States for women and men combined for the years 1959–2014. The reported number of counts in parentheses in the title of each panel is the actual number of deaths. To control for varying lengths of months, the monthly columns in each histogram have been adjusted for a uniform length (30 days). The horizontal reference lines denote the expected value of a uniform distribution (=1/12).

The typical distribution follows a sinusoidal pattern with highest mortality in winter and relatively few cases in the summer. Primarily, those are circulatory diseases (e.g., heart diseases, cerebrovascular diseases)—as shown in the first row of Fig. 9.1—and respiratory diseases such as chronic obstructive pulmonary disease ("COPD"), pneumonia or influenza (Eurowinter Group 1997, 2000; Mackenbach et al. 1992; Kunst et al. 1990; Rau 2007; Yen et al. 2000; Seretakis et al. 1997), which are displayed in three horizontal panels in the middle of Fig. 9.1.

If diseases, and ultimately, mortality occur seasonally, it has been argued that "an environmental factor has to be considered in the etiology of that disease" (Marrero 1983, p. 275).[1] The main environmental factor to trigger higher mortality during winter for circulatory diseases and respiratory diseases—the rows on top of

---

[1]It should be noted that the impact of environmental factors on diseases and deaths is as not a finding of the latter part of the twentieth century but is well known for more than 2000 years. In about 400BC Hippocrates started his treatise "On Airs, Waters, and Places" with the following words: "Whoever wishes to investigate medicine properly, should proceed thus: in the first place to consider the seasons of the year, and what effects each of them produces for they are not at all alike, but differ much from themselves in regard to their changes."
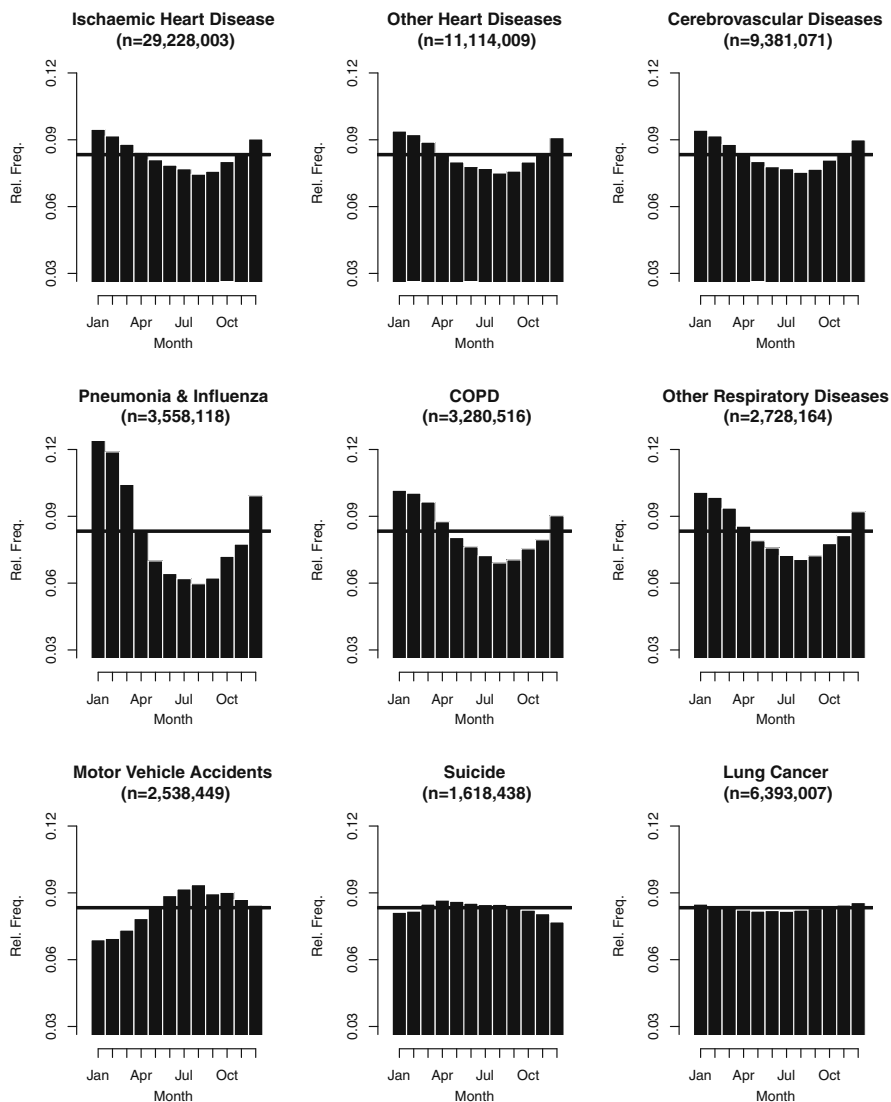
Fig. 9.1 Seasonality of selected causes of death in the United States for both sexes combined for the years 1959–2014. The counts reported in each panel denote the actual numbers of death. The relative frequencies in each histogram are adjusted for a uniform length of 30 days per month (Data source: National Center for Health Statistics and National Bureau of Economic Research)

Fig. 9.1—is well understood: temperature. Cold temperatures constrict the blood vessels and change the composition of the blood; furthermore, low temperatures facilitate the survival of bacteria in droplets and increase the risk for pulmonary infections (Eurowinter Group 1997, 2000; Huynen et al. 2001; Rau 2007).

The patterns observed in the three panels at the bottom of Fig. 9.1 deviate from the ones for circulatory and respiratory diseases above. Motor vehicle accidents do not peak in winter but around July and August. Many people assume that the reason for the peak in all-cause mortality is due to suicides in winter. The middle panel at the bottom of Fig. 9.1 illustrates why this assumption is wrong for three reasons: (1) The seasonal pattern is less pronounced for suicide than for other causes. (2) If one can speak of a seasonal pattern at all, the peak occurs definitely not during winter. (3) The 30,000 observed deaths are less than 1.5% of all deaths; not enough to shape the pattern for all causes. Lung cancer, whose impact on mortality in the United States was discussed in previous chapters, is—like many malignant neoplasms—an example of no or only negligible seasonality.

Figure 9.1 displays an aggregated picture of monthly deaths. In our analysis we want to investigate, however, whether the seasonal pattern for selected causes of death differs by age as well as whether the seasonal pattern changed over calendar time. The multiplicative model[2] suggested by Eilers et al. (2008) to decompose seasonal data allows such an analysis. The model is, at its core, another application of smoothing data via *P*-splines (Eilers and Marx 1996) as in Chap. 5. It is rather flexible since it allows the estimation not only of counts but also of rates. Exposures are then included as log offsets if the latter is desired, similar to Camarda's approach (2012, 2015) employed in Chaps. 5, 6, and 7. We use the model in its most simple form: The model is estimating counts assuming an annual unimodal pattern in the data. Not allowing for bimodal patterns or even higher frequencies should not induce any problems in our analysis since the causes in which we are interested in feature clear patterns with one peak and one trough (see Fig. 9.1).

We model the expected value of death counts y over age a and time t, $\mu_{ta} = E(y_{ta})$, to be Poisson distributed using a log-link function

$$\log(\mu_{ta}) = v_{ta} + f_{ta}\cos(\omega t) + g_{ta}\sin(\omega t)$$

with $\omega = 2\pi/p$, where p is the period. In our case of monthly values $p = 12$. Further technical details are given in Eilers et al. (2008).

The estimation yields three smooth matrices/surfaces, v_{ta} for the trend as well as the smooth cosine and sine surfaces f_{ta} and g_{ta}. The trend surface captures any major changes in the overall pattern that could be caused by varying population sizes, survival improvements, competing risks …. We are mainly not interested in this trend surface nor in the the actual sine and cosine surfaces. The two latter surfaces allow us, however, to obtain an estimate for the amplitude and the phase over age and time via simple trigonometric functions. The latter denotes the location of the annual peak of the death counts and is expressed in the difference in days from the 1st of January; i.e., a value of 30 corresponds to late January whereas -30 indicates that mortality is highest in the beginning of December.

---

[2]Since the logarithm of death counts is modeled, it actually becomes an additive model.

9.2   Results

Our data and results are displayed in five panels for each selected cause. On the first page for each cause, we show the observed ("raw") monthly numbers of death by calendar time and single age (adjusted for a duration of 30 days) in the upper panel. The panel below plots the fit of the model, i.e., the combined pattern of the trend and the sine and the cosine surfaces, which is equivalent to the observed counts minus the (raw) residuals; see, for example, Fig. 9.2 on page 103 for mortality from all causes combined for women. Our main interest is displayed on the second page for each cause. The top panel shows the estimated trend surface v_{at}. In the case of seasonality of all-cause mortality among US women (Fig. 9.3), we can see that the number of deaths from that category increases with age and reaches its "hotspot" for octogenarians before the numbers of death decline again. As the trend surface plots the seasonally-adjusted *density* of deaths, the lower number of deaths for nonagenarians are the consequence of less people being alive rather than a decline in the risk of dying. Even without the additional seasonal component, up to 3,500 women died at a single age during a single month. The height of "excess mortality" is depicted by the amplitude in the middle panel. Higher ages correspond not only to higher mortality; the colors and the contour lines suggest that mortality differences between winter and summer also become larger at higher ages. Increasing seasonality with age has already been described by Adolphe Quetelet in 1838 and is typically also found in more contemporary populations (Feinstein 2002; McDowall 1981; Rau and Doblhammer 2003; Rau 2007). Over time we can not really discern a clear trend. It seems rather that deaths for 70-year-old women in the US are about 10% higher during the peak season and about 15% higher for 90-year-old women than on average during a year. If we multiply the seasonal estimate of a given age and calendar time (e.g., 1.1) with the corresponding square of the trend surface (e.g., 1,500 deaths), we obtain the fitted value (e.g., 1,650 deaths) shown in the lower panel on the previous page. When the peak season occurs in a year is illustrated in the lower panel. The colors indicate a value slightly below 30. Hence, deaths occur most often in the end of January, regardless of age or calendar year.

The corresponding plots for men are depicted in Figs. 9.4 and 9.5. While male mortality is higher than female mortality at any age—at least in highly developed countries, the seasonal characteristics are rather similar between the two sexes: The proportion of excess deaths during winter varies between 5% at age 50 and 15% at age 90 with no apparent period effect. Also the part of the year when deaths peak among men occurs at the end of January. Those *seasonal* mortality similarities between women and men are not only present for all-cause mortality but also for most causes of death. That is why we restricted ourselves to show only the results for women but they apply equally to men. We show the results for men only in the case of motor vehicle accident because much less women die of that cause.

The largest subcategory analyzed by us in this chapter is death from heart diseases (see Figs. 9.6 and 9.7, pp. 107–108). Up to 1,300 deaths were recorded at a single age during a single month of a given year. As we can infer from the
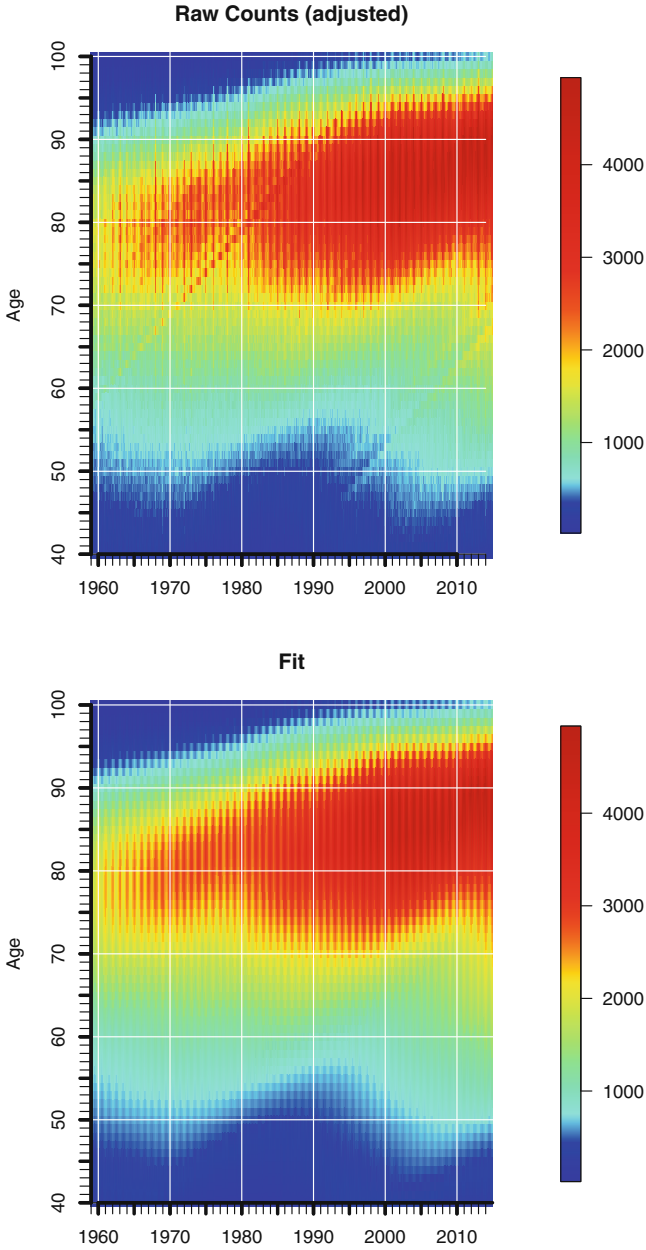
Fig. 9.2 Seasonality of mortality from all causes in the United States, 1959–2014, women, raw counts (adjusted for length of month) and fitted model (Data source: Human Mortality Database)

Fig. 9.3 Seasonality of
mortality from all causes in
the United States, 1959–2014,
women, estimated trend
surface (*top panel*), amplitude
(*middle panel*), and phase
(*bottom panel*) (Data source:
Human Mortality Database)
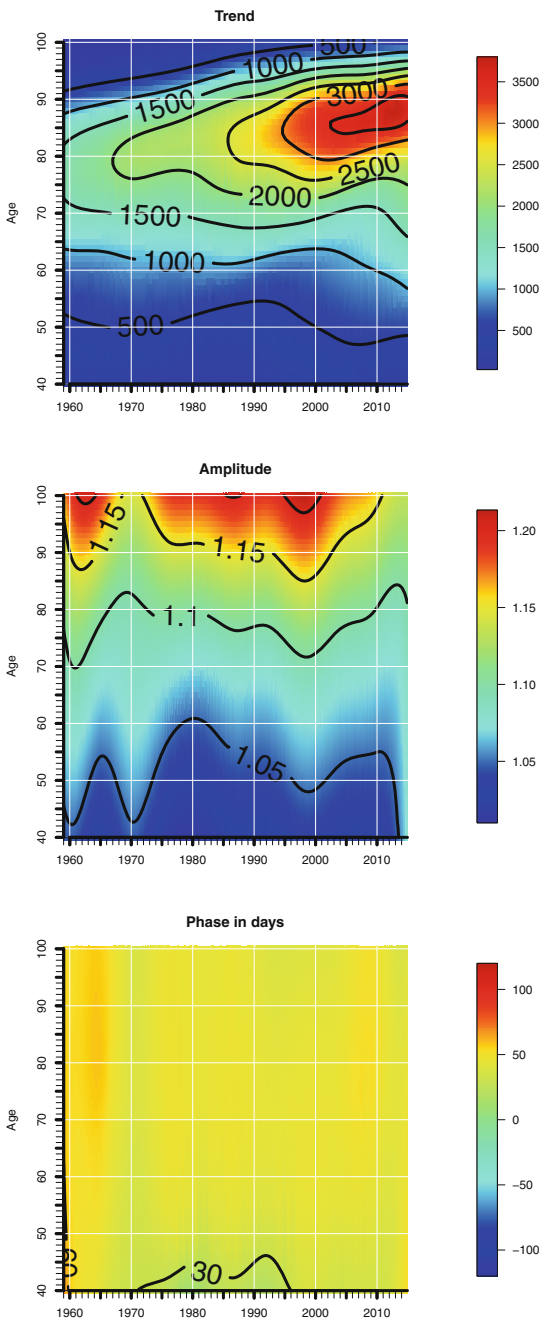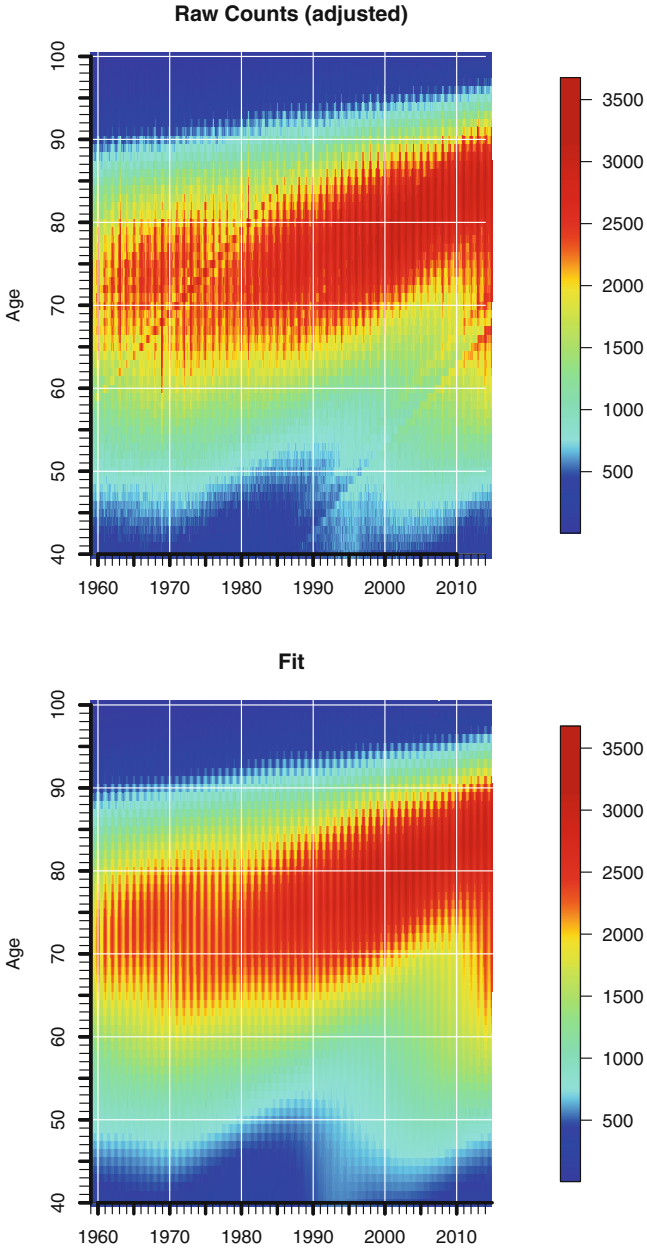
Fig. 9.4 Seasonality of mortality from all causes in the United States, 1959–2014, men, raw counts (adjusted for length of month) and fitted model (Data source: Human Mortality Database)

Fig. 9.5 Seasonality of
mortality from all causes in
the United States, 1959–2014,
men, estimated trend surface
(*top panel*), amplitude
(*middle panel*), and phase
(*bottom panel*) (Data source:
Human Mortality Database)

Fig. 9.6 Seasonality of mortality from heart diseases in the United States, 1959–2014, women, raw counts (adjusted for length of month) and fitted model (Data source: Human Mortality Database)

Fig. 9.7 Seasonality of
mortality from heart diseases
in the united states,
1959–2014, women,
estimated trend surface (*top
panel*), amplitude (*middle
panel*), and phase (*bottom
panel*) (Data source: Human
Mortality Database)
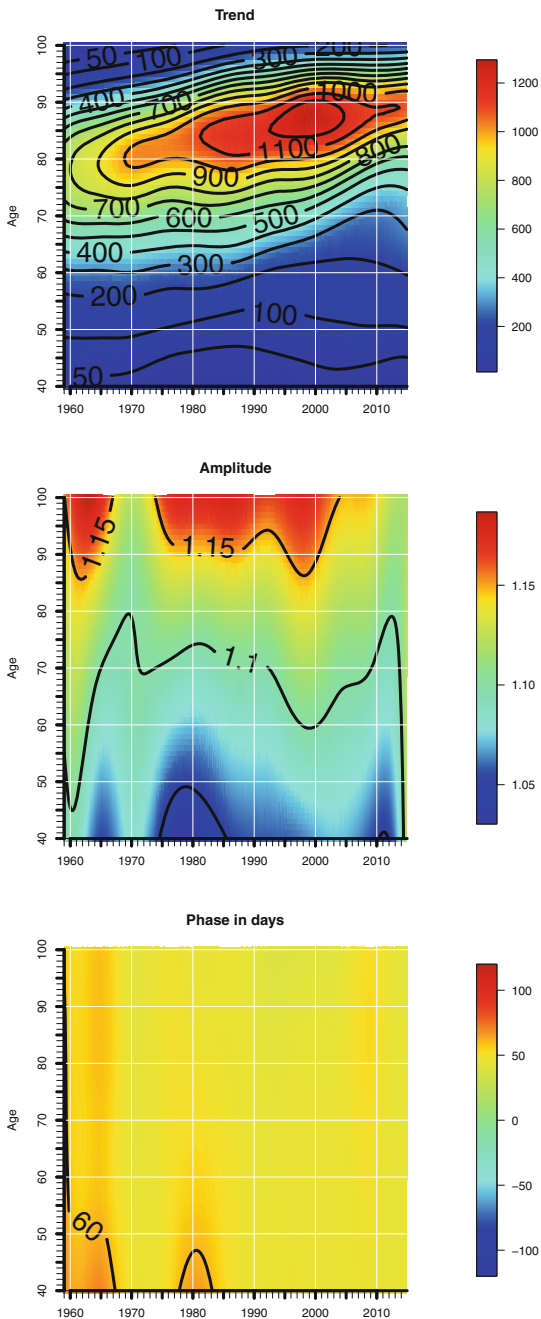
seasonal decomposition, this is the outcome of about 10 to 15% of excess deaths during the peak season. Also here we can not detect any period effects. In contrast to all-cause mortality with its peak at the end of January, deaths from heart diseases are highest at the end of February since the colors indicate a value of slightly below 60.

Most deaths from circulatory diseases can be attributed either to heart diseases or to cerebrovascular diseases. We analyzed the seasonal pattern of the latter category for women in Figs. 9.8 and 9.9 for men on pages 112–113. Comparable to heart diseases, the corridor with the largest number of deaths is moving to higher ages; the actual numbers are much smaller than for the other category, though. The extent of the seasonal pattern is remarkably similar to heart diseases. The amplitude is elevated again by about 10% around age 70 with larger fluctuations at higher ages and smaller fluctuations at younger ages. A clear trend over time is again not visible. Cerebrovascular diseases peak a bit earlier than heart diseases as suggested by the lower panels of Figs. 9.9 and 9.11. The highest number of deaths can be typically observed before the 30th day of the year, i.e., sometime between the middle and the end of January.

The Eurowinter group investigated the impact of cold temperatures on mortality about 20 years ago (e.g., Eurowinter Group 1997). They looked at ischaemic heart disease, cerebrovascular diseases, and respiratory diseases. As those three categories are mainly responsible for the seasonal pattern, we also analyzed the pattern for respiratory diseases, please see Figs. 9.12 and 9.13 on pages 114 & 115. The observed number of deaths is a bit higher than for cerebrovascular diseases. The seasonal decomposition on the second page shows that this is primarily the outcome of large seasonal fluctuations. Even the highest values in the trend surface on top are smaller than the corresponding values for cerebrovascular diseases. Excess deaths are, however, not only 10 to 15% higher in winter than throughout the year in general. The middle panel clearly illustrates that deaths from diseases such as pneumonia, influenza, COPD, etc. are at least 30% higher during peak season, which occurs at the end of February as the plot for the phase at the bottom illustrates. In contrast to the previously discussed two groups of circulatory diseases, the darker shades of blue during more recent years in the plot of the amplitude for respiratory diseases suggest that seasonal fluctuations became smaller over time.

Although motor vehicle accidents are by no means a major cause of death category, we decided nevertheless to include it. In the worst case 200 people of a given age died during a single month. The raw counts and fitted counts in Fig. 9.14 and the trend surface in Fig. 9.15 demonstrate that the period with the highest numbers of deaths is (thankfully) over. It occurred during the 1970s and 1980s to men aged around 20 years. The same plots show also that those men, born between 1950 and about 1965 suffer from a higher number of deaths also at higher ages. Since we are not looking at mortality per se but at death counts, this cohort effect is not necessarily the outcome of higher mortality; it could also be caused by the high number of births during those years ("baby boomers"). It is interesting to note, however, that we can also here detect a pattern on the 45° line for the seasonal amplitude and for the phase, which should be unaffected by

Fig. 9.8 Seasonality of mortality from cerebrovascular diseases in the United States, 1959–2014, women, raw counts (adjusted for length of month) and fitted model (Data source: Human Mortality Database)
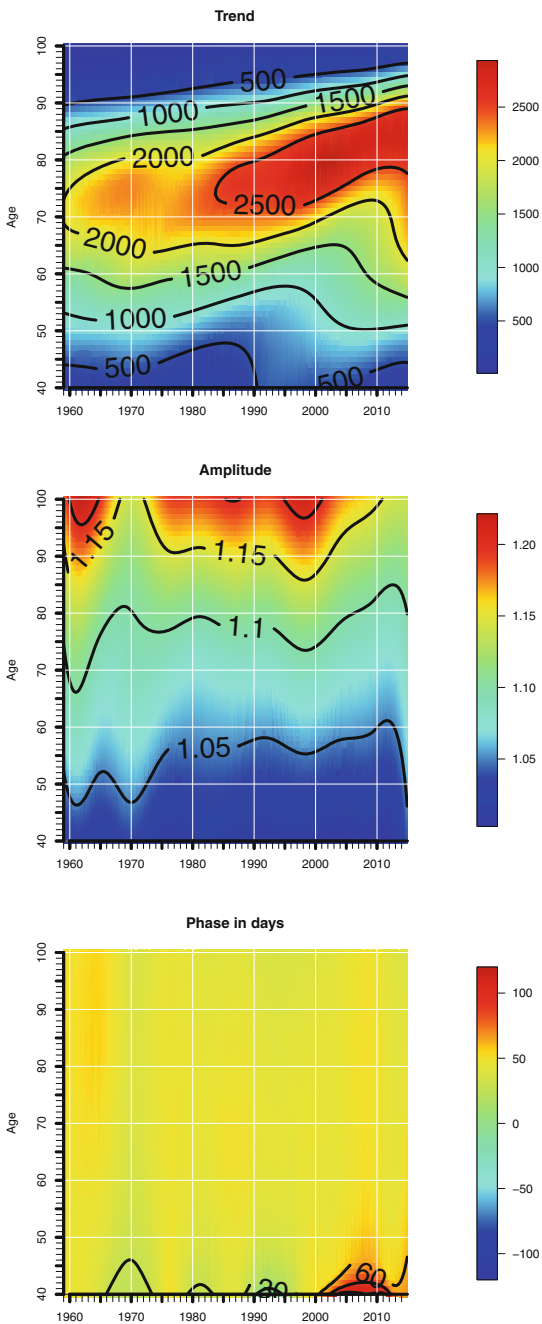
Fig. 9.9 Seasonality of mortality from cerebrovascular diseases in the United States, 1959–2014, women, estimated trend surface (*top panel*), amplitude (*middle panel*), and phase (*bottom panel*) (Data source: Human Mortality Database)
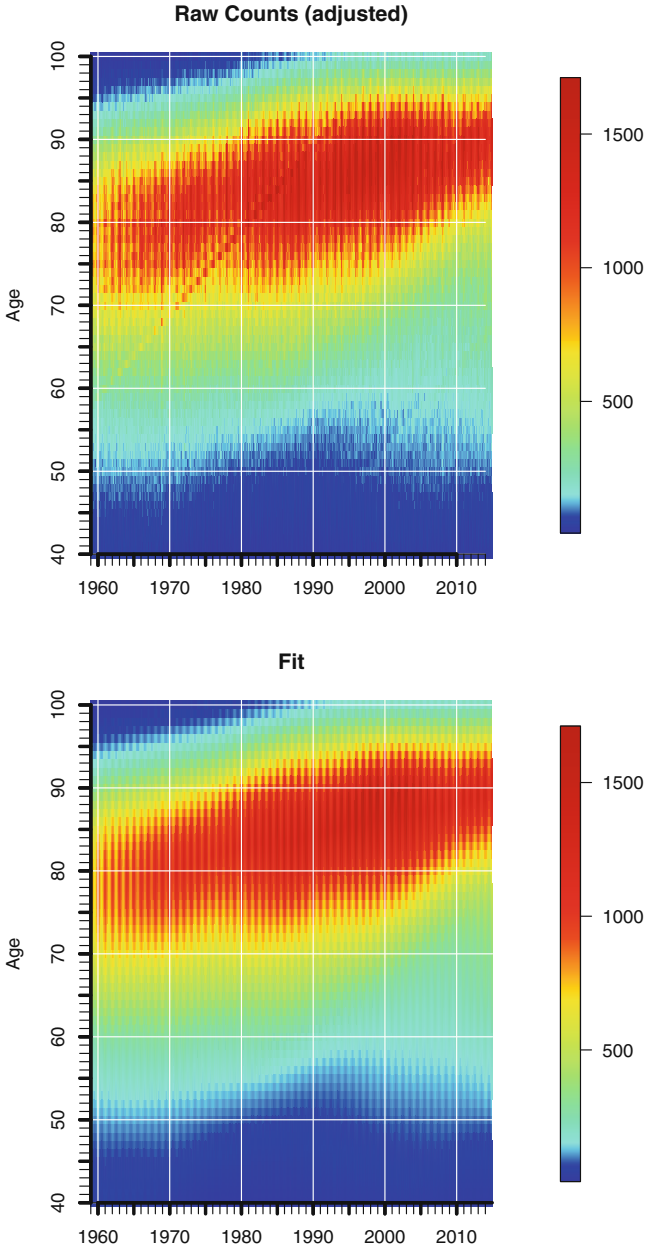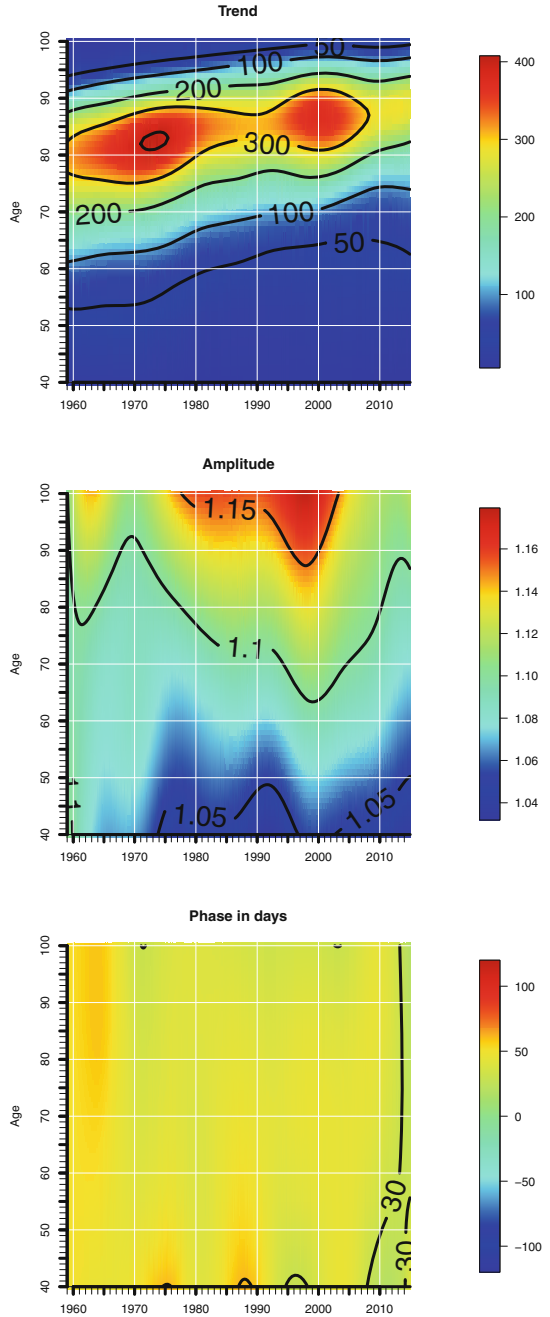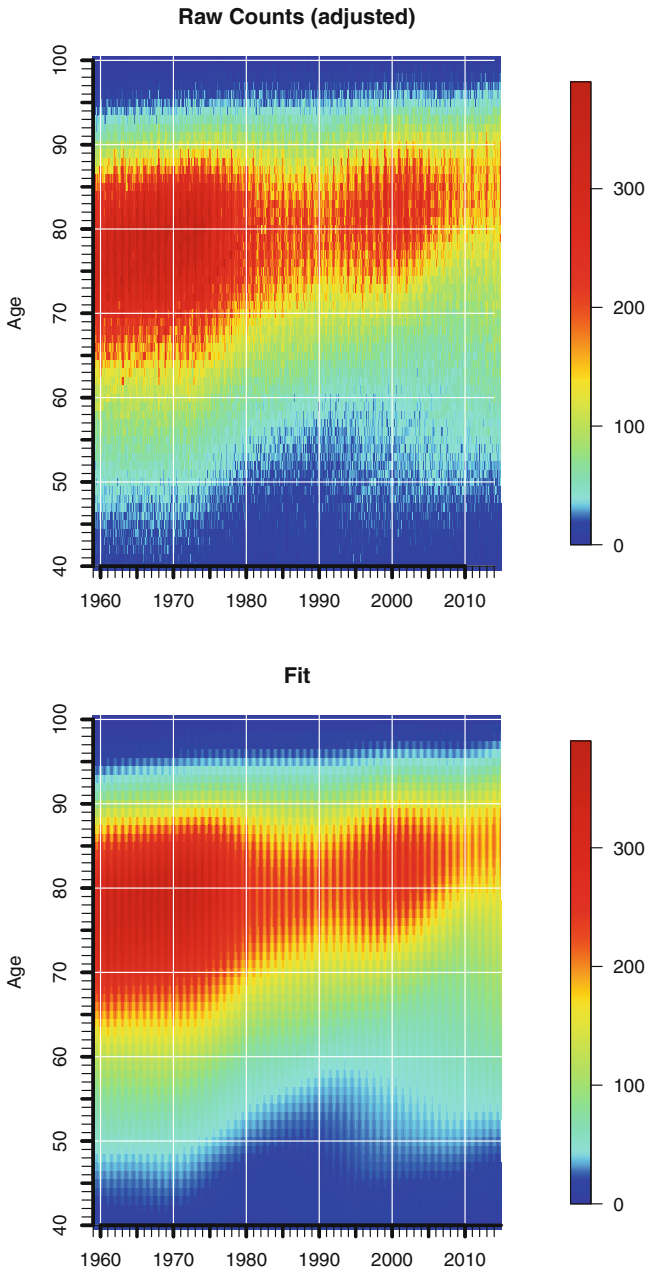
Fig. 9.10 Seasonality of mortality from cerebrovascular diseases in the United States, 1959–2014, men, raw counts (adjusted for length of month) and fitted model (Data source: Human Mortality Database)

Fig. 9.11 Seasonality of
mortality from
cerebrovascular diseases in
the United States, 1959–2014,
men, estimated trend surface
(*top panel*), amplitude
(*middle panel*), and phase
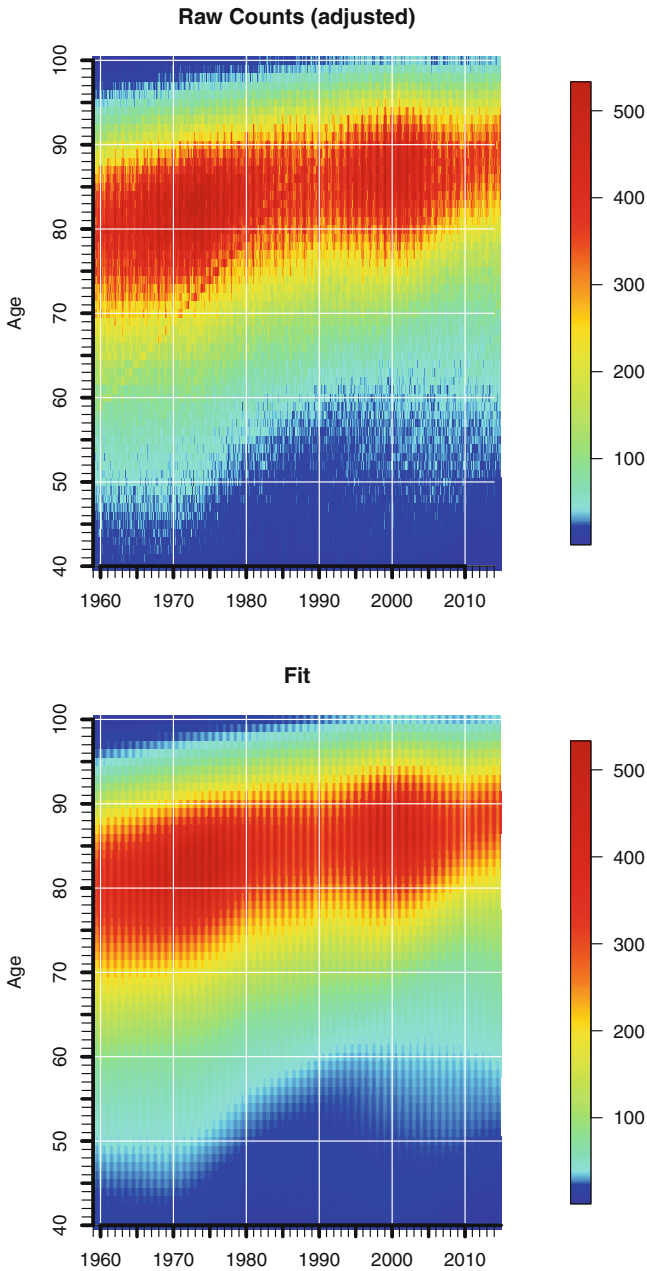(*bottom panel*) (Data source:
Human Mortality Database)
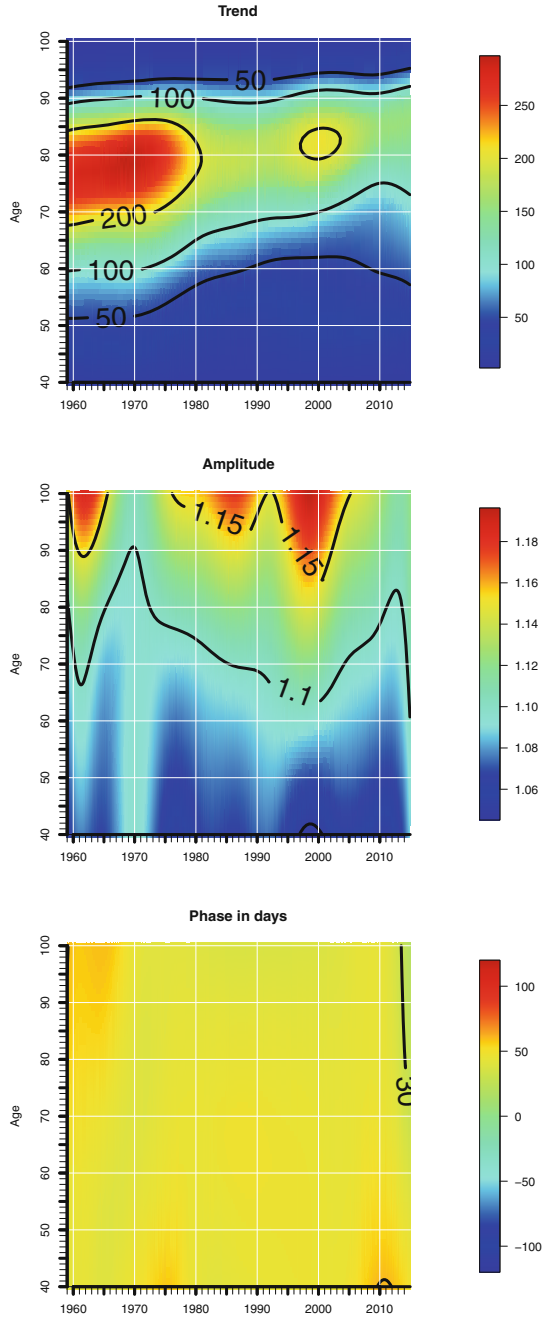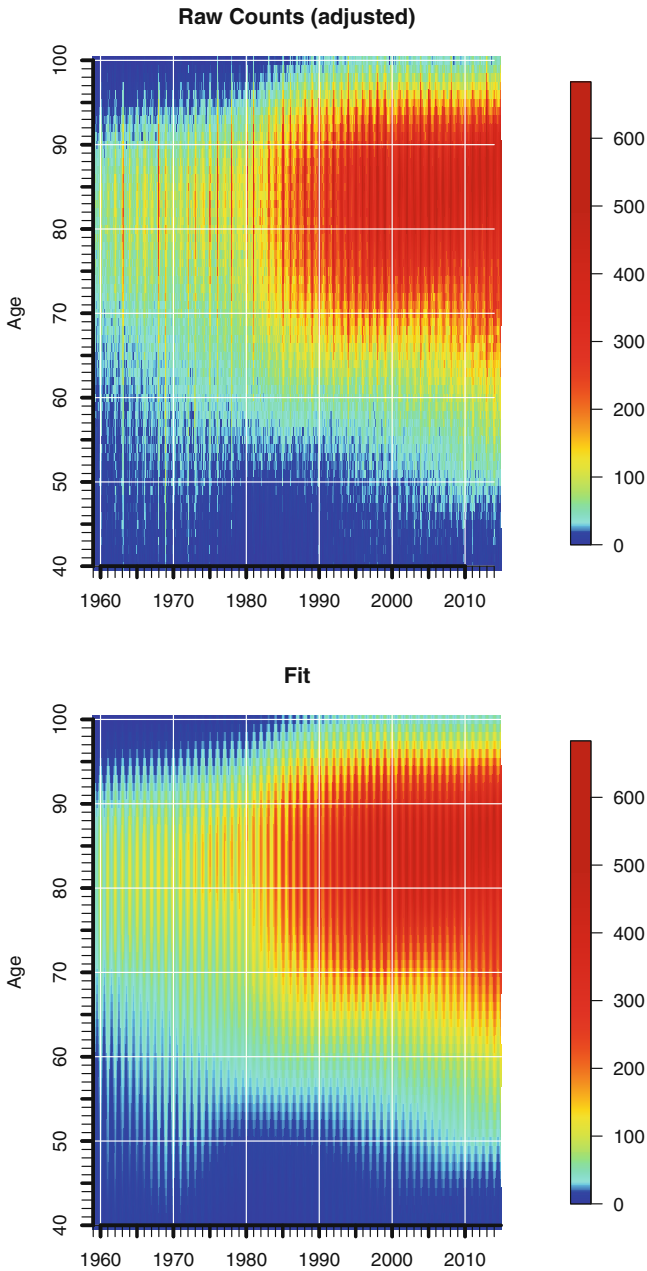
Fig. 9.12 Seasonality of mortality from respiratory diseases in the United States, 1959–2014, women, raw counts (adjusted for length of month) and fitted model (Data source: Human Mortality Database)

Fig. 9.13 Seasonality of mortality from respiratory diseases in the United States, 1959–2014, women, estimated trend surface (*top panel*), amplitude (*middle panel*), and phase (*bottom panel*) (Data source: Human Mortality Database)

Raw Counts (adjusted)

Fit
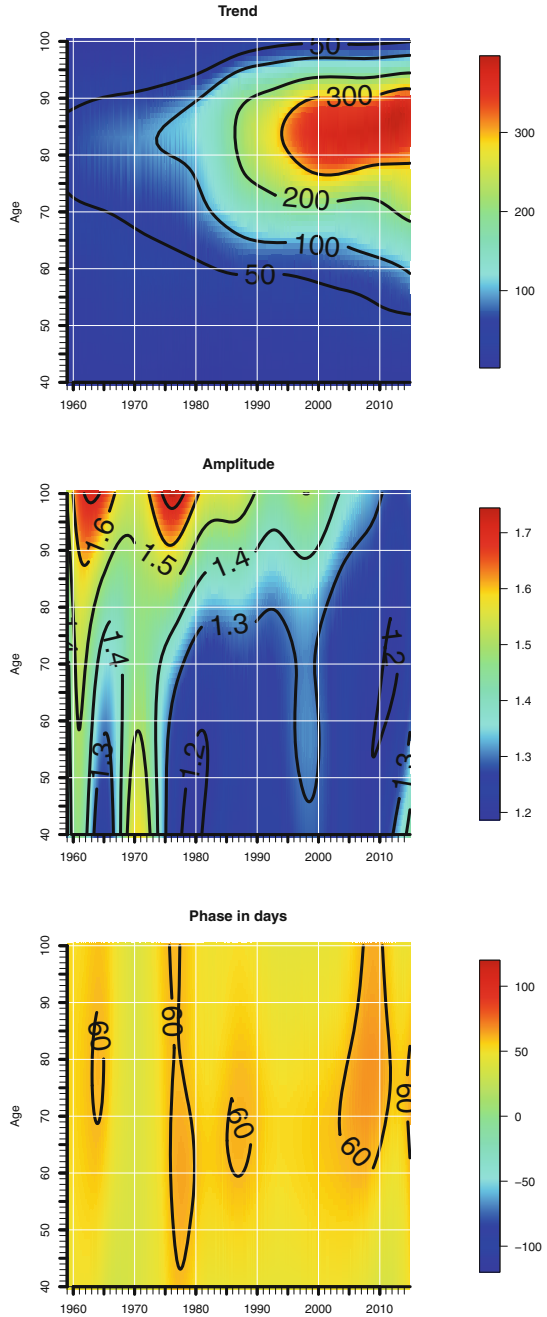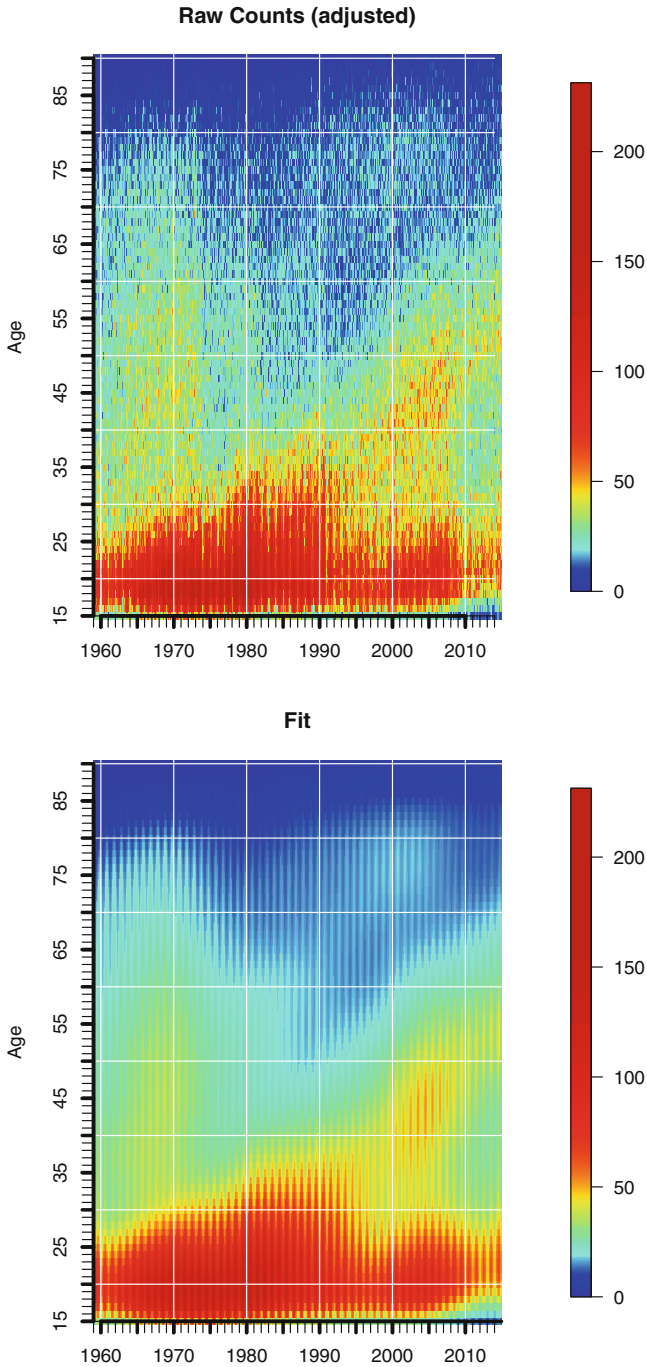
Fig. 9.14 Seasonality of motor vehicle accidents in the United States, 1959–2014, men, raw counts (adjusted for length of month) and fitted model (Data source: Human Mortality Database)

Fig. 9.15 Seasonality of
motor vehicle accidents in the
United States, 1959–2014,
men, estimated trend surface
(*top panel*), amplitude
(*middle panel*), and phase
(*bottom panel*) (Data source:
Human Mortality Database)

Raw Counts (adjusted)

Fit
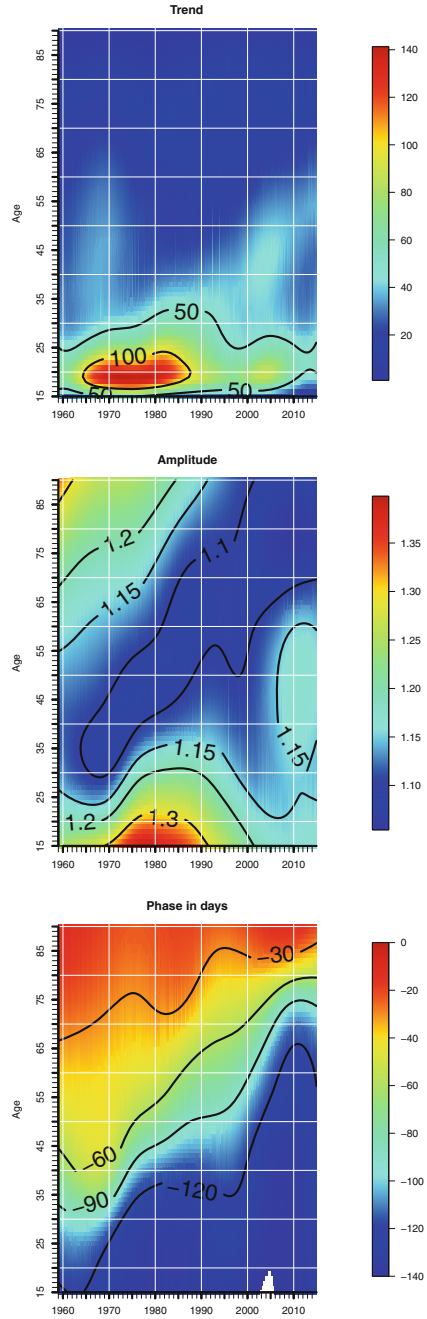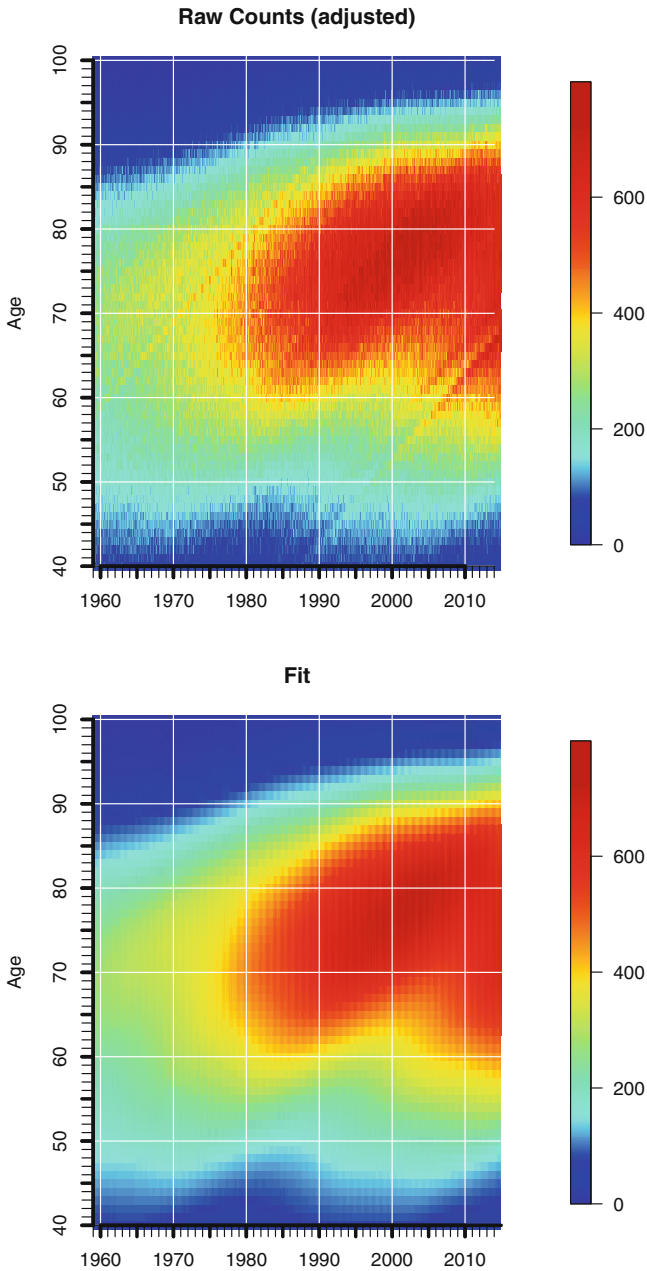
Fig. 9.16 Seasonality of mortality from all cancers in the United States, 1959–2014, women, raw counts (adjusted for length of month) and fitted model (Data source: Human Mortality Database)

Fig. 9.17 Seasonality of mortality from all cancers in the United States, 1959–2014, women, estimated trend surface (*top panel*), amplitude (*middle panel*), and phase (*bottom panel*) (Data source: Human Mortality Database)
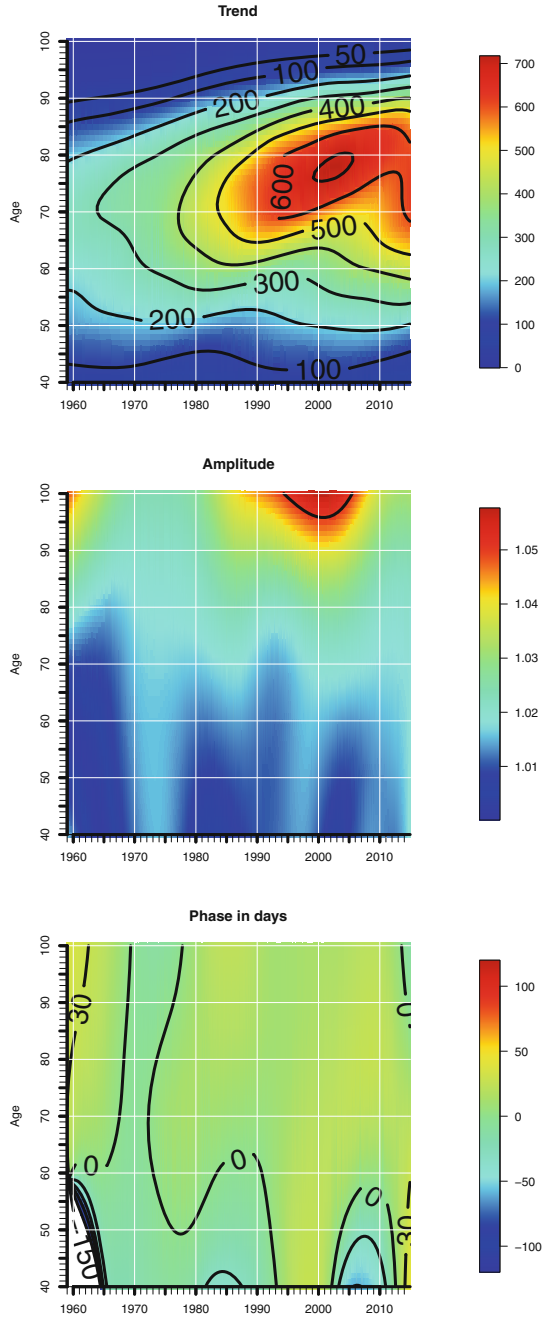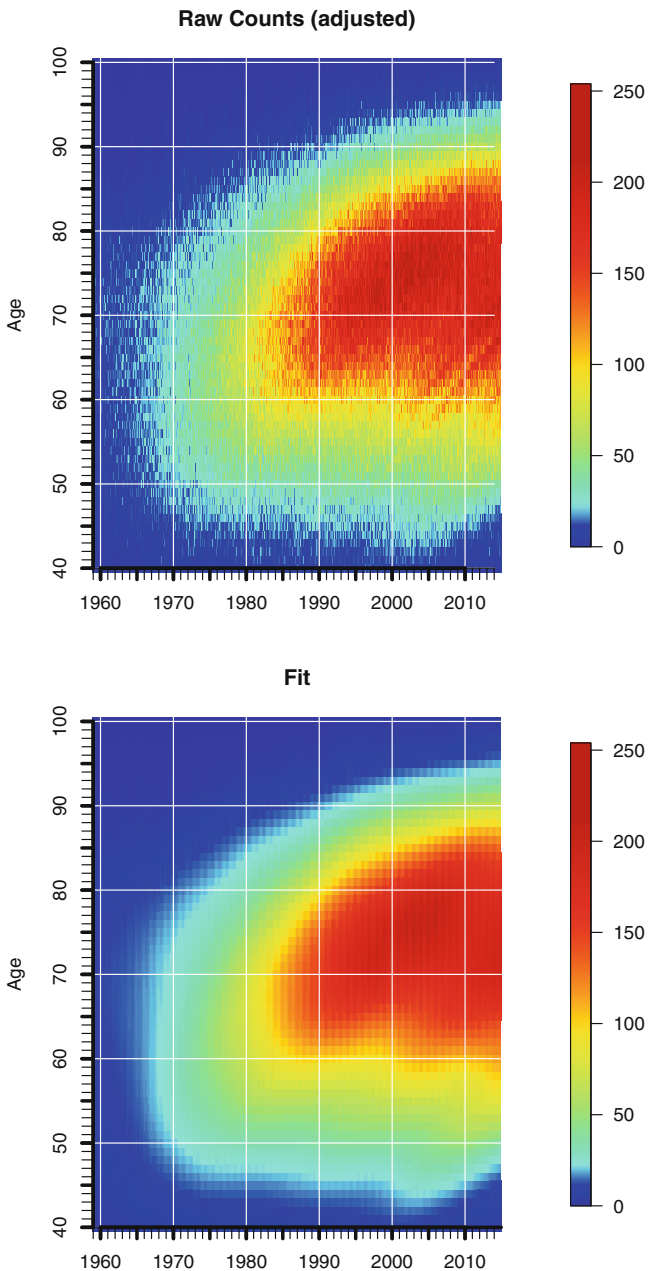
Fig. 9.18 Seasonality of mortality from lung cancer in the United States, 1959–2014, women, raw counts (adjusted for length of month) and fitted model (Data source: Human Mortality Database)

Fig. 9.19 Seasonality of mortality from lung cancer in the United States, 1959–2014, women, estimated trend surface (*top panel*), amplitude (*middle panel*), and phase (*bottom panel*) (Data source: Human Mortality Database)
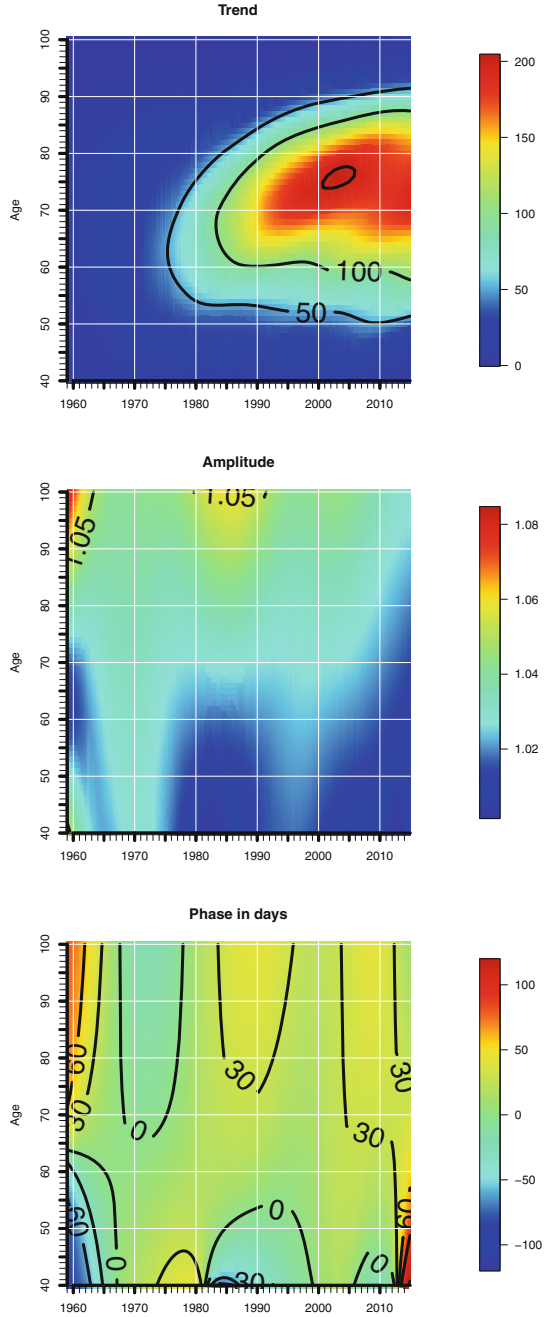
a larger population at risk since the trend surface accounts for it. The panel in the middle of Fig. 9.15 shows lowest seasonality for the birth cohort born before the baby boomers mentioned above. Also the change of the period when most deaths from motor vehicle accidents occur throughout a year features a cohort pattern. Whereas deaths from car accidents and similar causes peaked late in fall for older cohorts, the highest number of deaths for baby boomers and later generations are recorded at least 120 days before the 1st of January, which corresponds to August of a year.

We want to conclude this chapter by showing that cancers in general (see Figs. 9.16 and 9.17 on pages 118–119) and lung cancer (see Figs. 9.18 and 9.19 on pages 120–121) are examples of non-seasonal diseases. Clearly the fluctuations throughout a year are barely noticable as the middle panels of Figs. 9.17 and 9.19 illustrate.

Chapter 10
Surface Plots for Cancer Survival

10.1 Introduction and Overview: The Impact of Cancer on Mortality in the United States

With 23.4% or 614,348 out of 2,626,418 deaths in the United States, heart diseases remained the leading cause of the death in the United States in 2014 (CDC/NCHS 2015). Hence, heart diseases contributed most to the age-standardized crude death rate in that year. The absolute level of mortality from heart diseases and other circulatory diseases diminished remarkably during recent decades as we show in Fig. 10.1. To avoid spurious results from the changing age composition of the population, we used the population of the year 2000 to age-standardize the rates. During the observed 60 years, mortality—as measured by the age-standardized crude death rate—dropped steadily for women as well as for men. This trend of declining mortality from circulatory diseases and rather stagnant cancer mortality may result in a reversal of the leading group of causes of death in the near future when more people might die of malignant neoplasms than of heart diseases or stroke.

The converging trajectories of these two major causes of death can be also presented from the perspective of cause-elimination life tables (results not shown here; see, for instance Preston et al. (2001) or Kintner (2004) for the methodology): If circulatory diseases had been non-existent, life expectancy at birth would have been 11 years higher in the 1960s. This gap decreased to about 4 years during the most recent years (3.62 for women, 4.16 for men), whereas the impact of eradicating cancer remained relatively stationary over time for malignant neoplasms.

The proportion of deaths from cancer in relation to all causes varies considerably by age as well as over time as we show in Fig. 10.2. The marginal distribution over age is bimodal. A local peak is reached at childhood ages with the main contributing cancers being leukemia and lymphoma as pointed out in Moore and Hurvitz (2009).
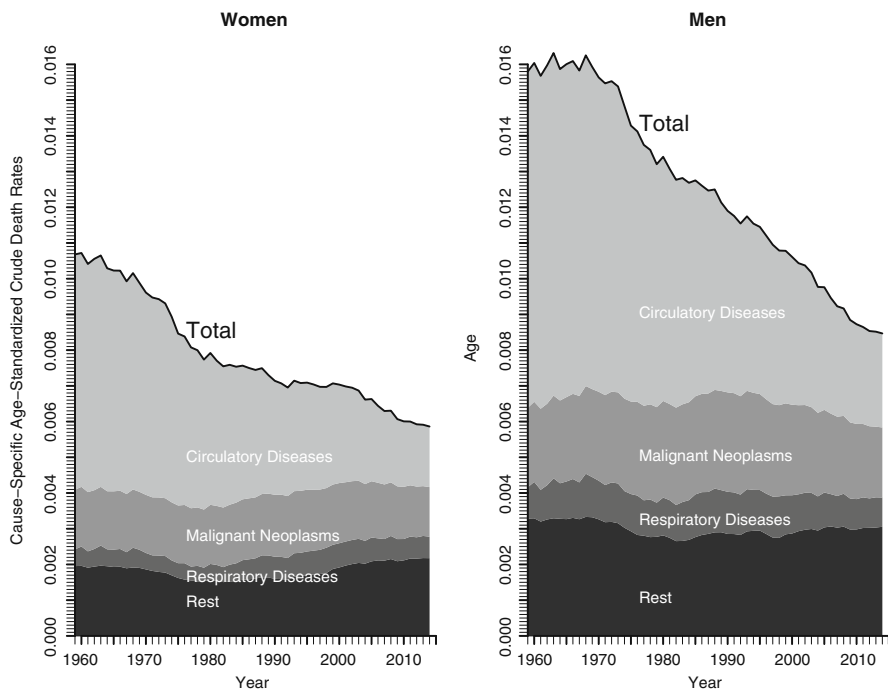
Fig. 10.1 Age-standardized crude death rates by cause and sex (*left panel*: women; *right panel*: men) in the United States from 1959–2014 (Data source: Own estimation based on data from the Human Mortality Database and the National Center for Health Statistics. The population of the year 2000 was used as the standardization population)

The age when the global peak is reached depends on the sex. 40% or more of all deaths of women around age 50 can be attributed to cancers whereas the largest proportion among men is reached between ages 60 and 70.

10.2   Dynamics of Cancer Survival by Cancer Site

People are usually not healthy and then die suddenly of a chronic, non-communicable disease such as cancer. In a very simplified manner, we can regard this as a two-step process: (1) People are healthy and then are diagnosed with a certain chronic disease x. (2) People who are diagnosed with disease x die of x or of another disease. The SEER data allow us to investigate developments for both steps. We can look at incidence data for the first step and see how incidence has changed over time by age. This might allow us to make inferences about the successes and failures of cancer *prevention*. We focus, however, on the second step: Analyzing survival from the moment of diagnosis to death. Thus, our focus is rather on the successes and failures of cancer *treatment*.
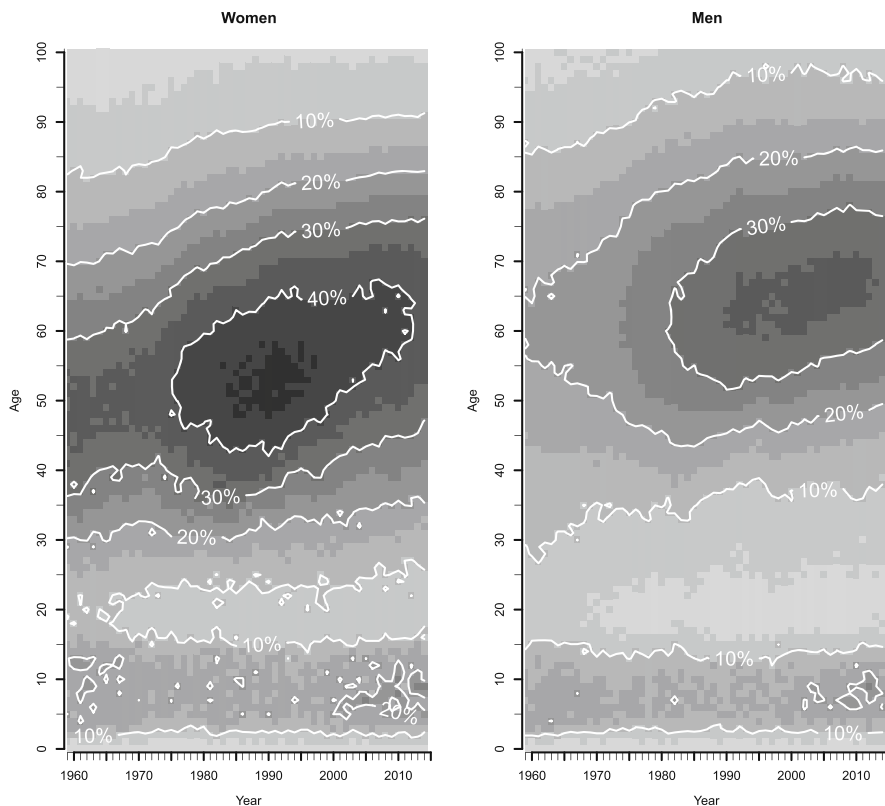
Fig. 10.2 Proportion of deaths from cancer in relation to all causes (*left panel*: women; *right panel*: men) in the United States at ages 0–100 from 1959–2010 (Data source: National Center for Health Statistics)

We decided to base our analysis on the five year survival rate. According to the National Cancer Institute (2017) it is the "percentage of people in a study or treatment group who are alive 5 years after they were diagnosed with or started treatment for a disease, such as cancer."[1] We use three different operationalizations of five-year survival:

1. For each cancer site and sex we estimate by single calendar year and single age how many persons are still alive 60 months after diagnosis. Thus, the first approach measures the survival chances *in general* of someone who was diagnosed with a specific cancer.

2. Obviously, the first operationalization is highly dependent on age: someone aged 95 years has much lower survival chances in general than someone aged 45 years with the same diagnosis. The interest is often not on survival/mortality in general but on mortality due to the diagnosed disease. We therefore estimated also the

---

[1]Since it is a percentage/proportion, we wonder why the term "rate" has become so commonly used.

probability that someone will not die of the diagnosed cancer within 5 years. This second operationalization is sometimes called "corrected survival rate", "net survival" or "disease-specific survival" (Parkin and Hakulinen 1991, p. 167). We use the last term.

3. While the first two approaches describe the risk of dying of any cause (1) or of the diagnosed cancer (2), the third approach compares the survival chances of the diagnosed individuals with the general population. The ratio of observed survival to expected survival is called "relative survival" and can be traced back to Berkson and Gage (1952). Relative survival is "defined as the observed survival of the cancer patients divided by the expected survival of a comparable group from the general population, free from the cancer under study" (Talbäck and Dickman 2011, p. 2626). The observed survival rate for relative survival corresponds to our first approach, i.e., the probability of surviving from all causes of death. The most common methods to estimate relative survival (e.g., Ederer I, Ederer II, Hakulinen) differ with regard to the estimation of expected survival, though (Cho et al. 2011). As shown by Rutherford et al. (2012, p. 20), "[t]aking age into account [...] removes most of the differences between the methods." Since we analyze by single ages and single calendar years, the choice of method to estimate expected survival is less of a problem. We estimated expected survival with life table data from the Human Mortality Database (2017): Expected five year survival for 55 year old women in the year 2000 was the probability to survive age 55 in the year 2000 multiplied by the probability to survive age 56 in the year 2001, ... multiplied by the probability to survive age 59 in the year 2004. Using the general population instead of the general population free from cancer violates the definition of relative survival. It has been done and justified, however, since the inception of the method (please see Appendix Note 2 of Berkson and Gage (1952) or Ederer et al. (1961)). Also recent papers such as Talbäck and Dickman (2011, p. 2626 and Table 2) argue "that the bias is sufficiently small to be ignorable for most applications." Not accounting for the inclusion of cancer patient mortality becomes a problem only for the oldest subjects and follow-up times of 10 years or more. We would also argue that our estimates for five-year survival are sufficiently close to the official estimates. For example, SEER estimates relative survival of women diagnosed with breast cancer aged 50–64 years to be 90.1% during the period 2007–2013.[2] Our results for the most recent 3 years of our analysis varied between 90.05% and 91.08%.

The three approaches are featured in a panel each of Fig. 10.3 for breast cancer. We restricted our analysis of breast cancer to women although men can die from it as well. Our estimates for single year and age for breast cancer as well as for all other cancer sites have been smoothed, again using *P*-Splines as outlined in Chap. 5 Eilers and Marx (1996); Camarda (2012, 2015).

---

[2]See   https://seer.cancer.gov/explorer/application.php?site=55&data_type=4&stat_type=5&compareBy=sex&series=race&chk_sex_3=3&chk_race_1=1&chk_age_range_141=141&chk_age_range_160=160&chk_stage_101=101&advopt_precision=1&showDataFor=age_range_160_and_stage_101.
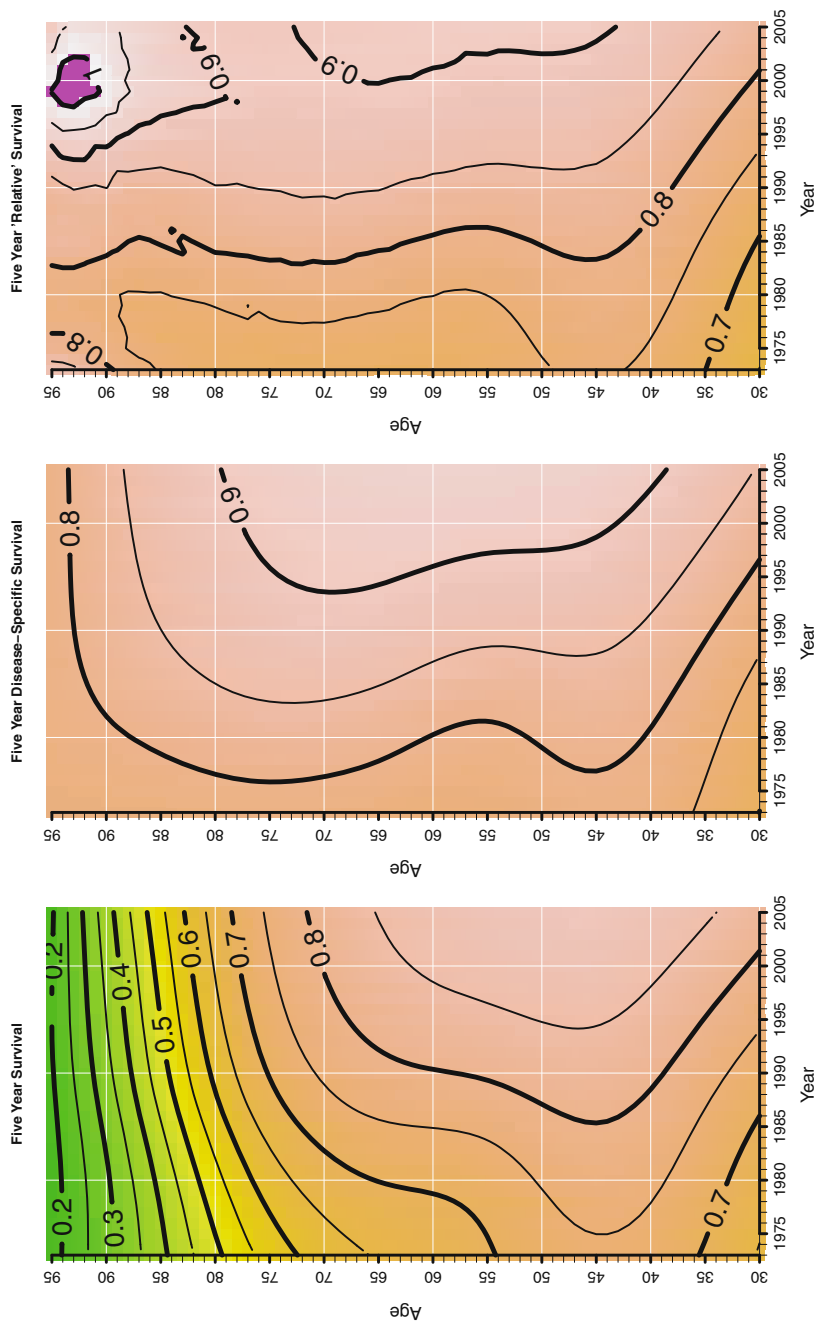
Fig. 10.3 Five year survival for breast cancer at ages 30–95 from 1973–2005. *Left panel*: Probability to survive for 5 years after diagnosed with breast cancer (any cause). *Middle panel*: Probability of not dying from breast cancer within 5 years after diagnosis. *Right panel*: Five year survival of women diagnosed with breast cancer in relation to five year survival of women in the general population (Data Source: SEER and Human Mortality Database)

The panel on the left denotes the probability to survive for another 5 years after being diagnosed with breast cancer, regardless of the actual cause of death. The figure exhibits an obvious age gradient: Values of 30% or less at ages above 90 are the consequence that the women are not only at an elevated risk of dying from breast cancer. Other causes, most notably circulatory diseases, further reduce the chances to survive for five more years. Consequently, the upward trend of the contour lines can not be interpreted as progress made against the lethality of breast cancer. Still it provides the answer to the question "How likely is it that I survive for another five years?" for someone who got diagnosed with breast cancer.

While the left panel takes all "exit" possibilities into account, the panel in the middle looks only at death from breast cancer. As a consequence, one minus the depicted value equals the probability to die from breast cancer within 5 years after diagnosis. The rather vertical lines from about age 40 to about age 80 indicate that the chance of surviving breast cancer for at least 5 years has increased over time. For instance, the probability for 60-year-old women who got diagnosed with breast cancer in 1980 to survive 5 years was 80%; the equivalent value in 2000 was higher than 90%. To express it even more positively: The risk of dying was cut in half within less than 20 years (1980: $1 - 0.8 = 20\%$; $\sim 1995 : 1 - 0.9 = 10\%$)!

The panel on the right of Fig. 10.3 shows "relative survival", i.e., it illustrates the relative survival disadvantage of those diagnosed with breast cancer in relation to the general population. A level of one would indicate that there was no difference in the chance to survive for five more years between someone with a cancer diagnosis and the general population. Unfortunately—but also not surprisingly—women with breast cancer have lower survival chances than the general population. We can detect, however, progress over time. The excess risk is less than 10% in recent years for women with breast cancer in comparison to the general population (contour line of 0.9) whereas it was about 30% just 25 years earlier. It is important to point out that the increasing values of the vertical lines suggest a clear period effect: Progress against breast cancer was faster than progress in survival in general, regardless of the age when the woman was diagnosed.

It is theoretically possible to observe relative survival estimates that are higher than one. For instance, it could be the outcome of a selection effect: Persons that take advantage of screening programs and other early preventive measures are possibly leading rather healthy lifestyles. If those persons are diagnosed with a cancer that is virtually non-lethal, their survival advantage of their health behavior might be stronger than the additional mortality risk of the malignant neoplasm. Hence, it can not be concluded that getting diagnosed with a certain cancer could actually improve survival chances. We would argue, though, that the small area at ages 90–95 in 2000 is not the outcome of such a selection effect. Instead, we assume that it is the outcome of random data fluctuations due to small numbers of persons getting diagnosed. For example, 46 women at age 93 were diagnosed with breast cancer in 2000.

The corresponding estimates for colorectal cancer are depicted in Figs. 10.4 and 10.5 for women and men, respectively (pages 129 & 130). Both sexes feature comparable estimates. The dynamics are somehow reminiscent of breast cancer

Fig. 10.4 Five year survival for colorectal cancer at ages 30–95 from 1973–2005. *Left panel*: Probability to survive for 5 years after diagnosed with colorectal cancer (any cause). *Middle panel*: Probability of not dying from colorectal cancer within 5 years after diagnosis. *Right panel*: Five year survival of women diagnosed with colorectal cancer in relation to five year survival of women in the general population (Data Source: SEER and Human Mortality Database)
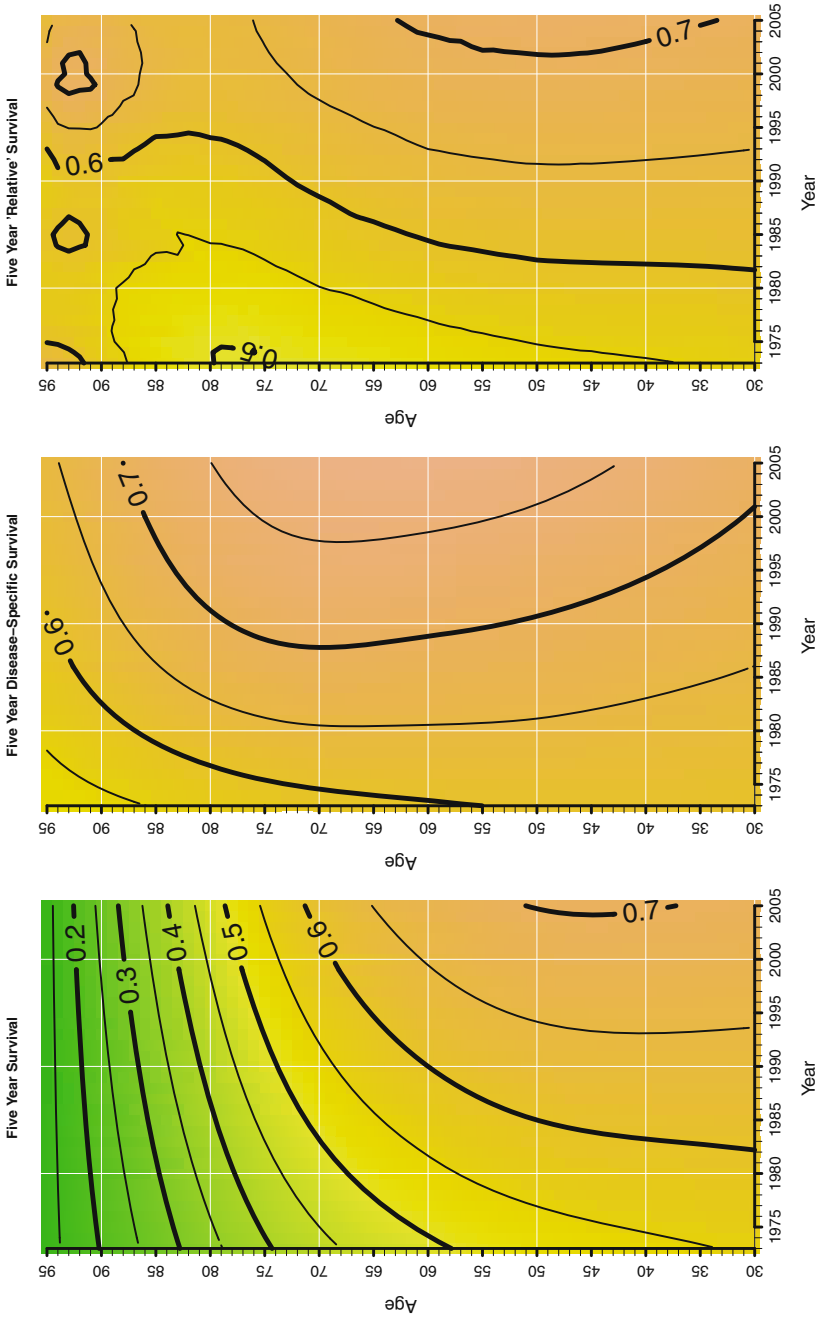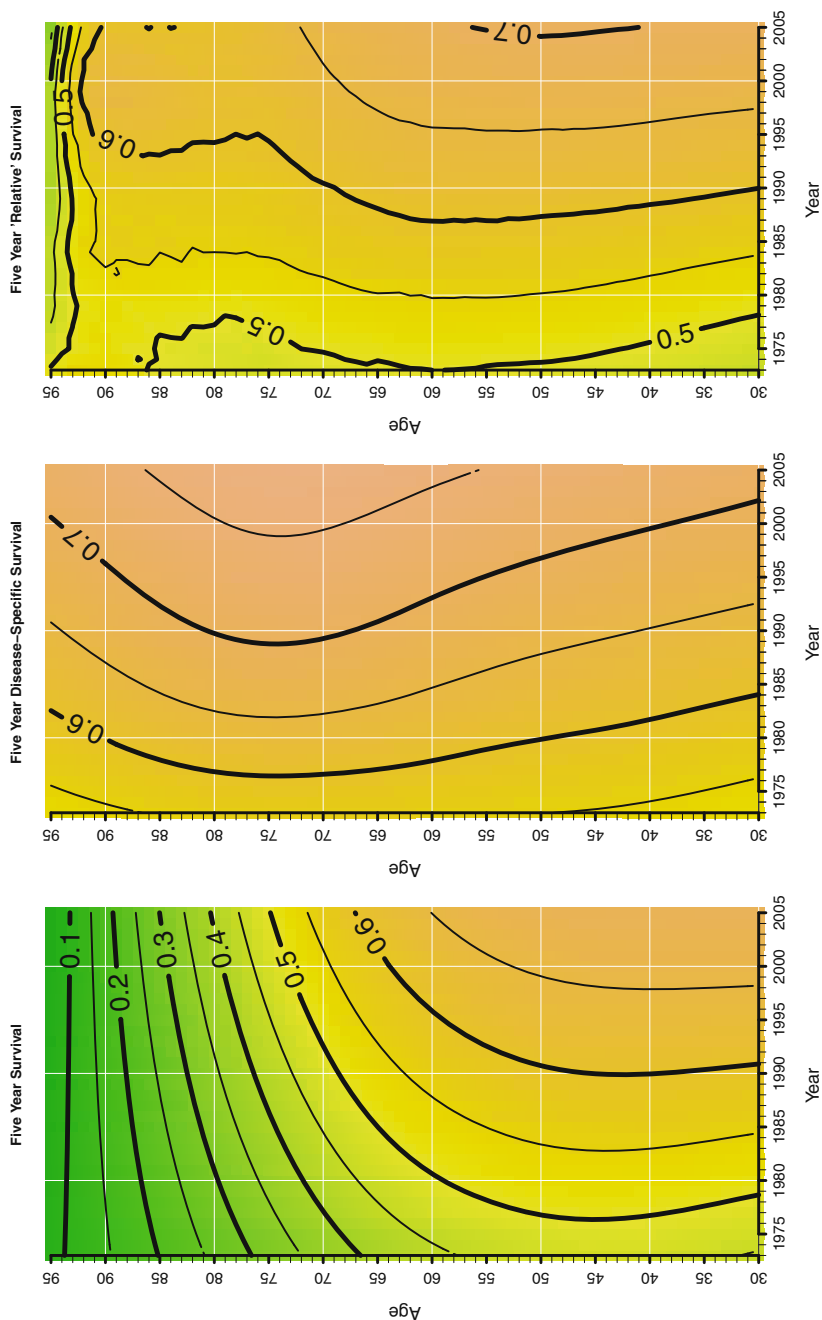
Fig. 10.5 Five year survival for colorectal cancer at ages 30–95 from 1973–2005. *Left panel*: Probability to survive for 5 years after diagnosed with colorectal cancer (any cause). *Middle panel*: Probability of not dying from colorectal cancer within 5 years after diagnosis. *Right panel*: Five year survival of men diagnosed with colorectal cancer in relation to five year survival of men in the general population (Data Source: SEER and Human Mortality Database)

albeit on a lower survival level: The chances to survive for another 5 years (left panels) above age 80 tend to follow a horizontal trend over time. This could be caused by at least two factors: Either there was no progress over time or that competing causes at those advanced ages are more important. There was, indeed, progress over time as shown by the panels in the middle of both figures. But despite all this progress, relative survival is still at least 30% lower than in the general population (right panels).

The dominance of shades of green in Figs. 10.6 and 10.7 illustrate that survival chances are much worse for lung cancer than for breast or colorectal cancer. The chances to survive for another 5 years after being diagnosed with cancer are less than 30%. Even at very advanced ages, relative survival is very low. On average it is about 80% lower in comparison to the general population.

Pancreatic cancer, as shown in Fig. 10.8 for women and men, belongs to the cancer sites with the worst survival chances. Living for another 5 years after diagnosis is extremely unlikely with a proportion of survivors of less than 10%. It is therefore not surprising that relative survival is also very low.

The last cancer site we investigated was prostate cancer (see Fig. 10.9). In terms of survival it can be found at the other side of the spectrum of pancreatic cancer. The vertical, numerically increasing, contour lines in the panel for relative survival provide evidence for a clear period effect: Relative survival became more common at all ages at a pace that was faster than improvements in survival in the general population. The most recent estimates show values of relative survival of more than 95%.

Differences in survival do not only exist between cancer sites. An important factor is also the stage when the cancer is diagnosed first. The data used in this study provide stage information for[3]

- "in situ"—a noninvasive neoplasm
- "localized"—an invasive neoplasm confined entirely to the organ of origin
- "regional"—a neoplasm that can not only be found in the organ of origin
- "distant"—a neoplasm that has spread to parts of the body remote from the primary tumor site.

We only present an example for colorectal cancer, contrasting the survival chances of persons where a localized tumor was detected with those with a distant malignant neoplasm. Figure 10.10 present the results for women; the corresponding plots for males are contained in Fig. 10.11. Both six-panel plots provide clear evidence that early detection of colorectal cancer is, literally, a matter of life

---

[3]Further details can be found in the field description of variable "SEER Historic Stage A" in the SEER research data record description.
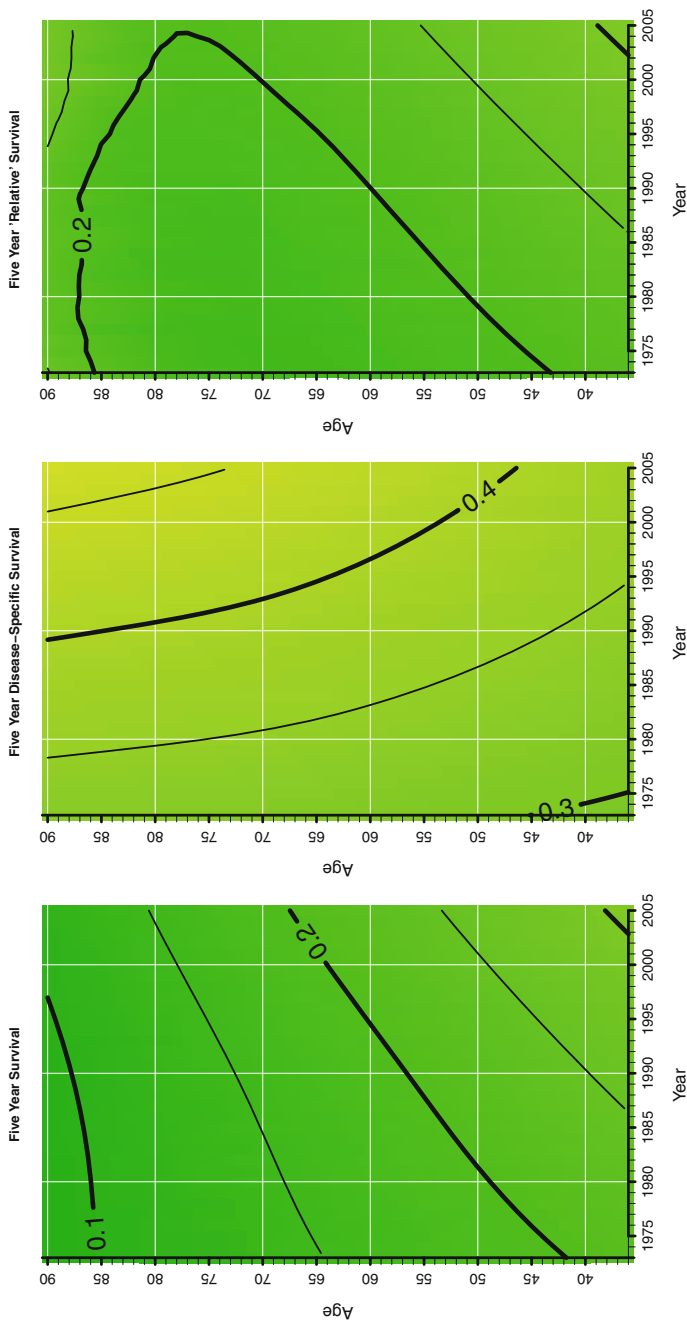
Fig. 10.6 Five year survival for lung cancer at ages 36–95 from 1973–2005. *Left panel*: Probability to survive for 5 years after diagnosed with lung cancer (any cause). *Middle panel*: Probability of not dying from lung cancer within 5 years after diagnosis. *Right panel*: Five year survival of women diagnosed with lung cancer in relation to five year survival of women in the general population (Data Source: SEER and Human Mortality Database)
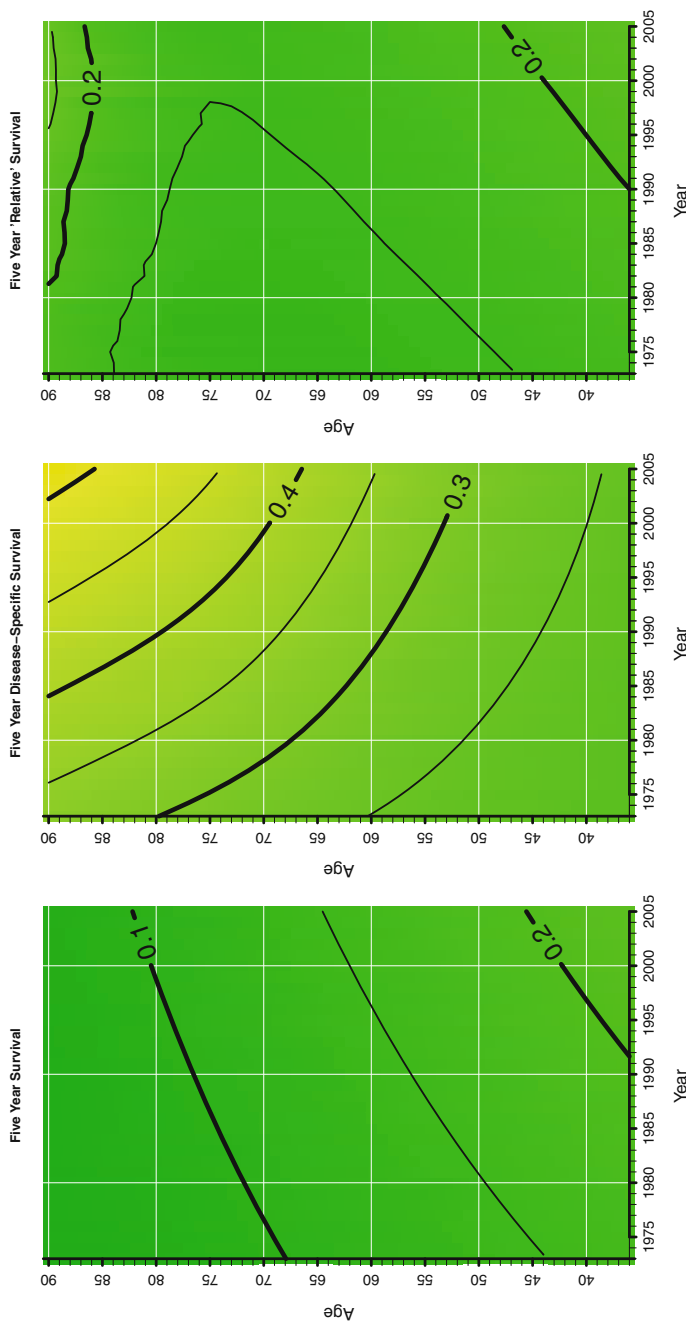
Fig. 10.7 Five year survival for lung cancer at ages 36–95 from 1973–2005. *Left panel*: Probability to survive for 5 years after diagnosed with lung cancer (any cause). *Middle panel*: Probability of not dying from lung cancer within 5 years after diagnosis. *Right panel*: Five year survival of men diagnosed with lung cancer in relation to five year survival of men in the general population (Data Source: SEER and Human Mortality Database)
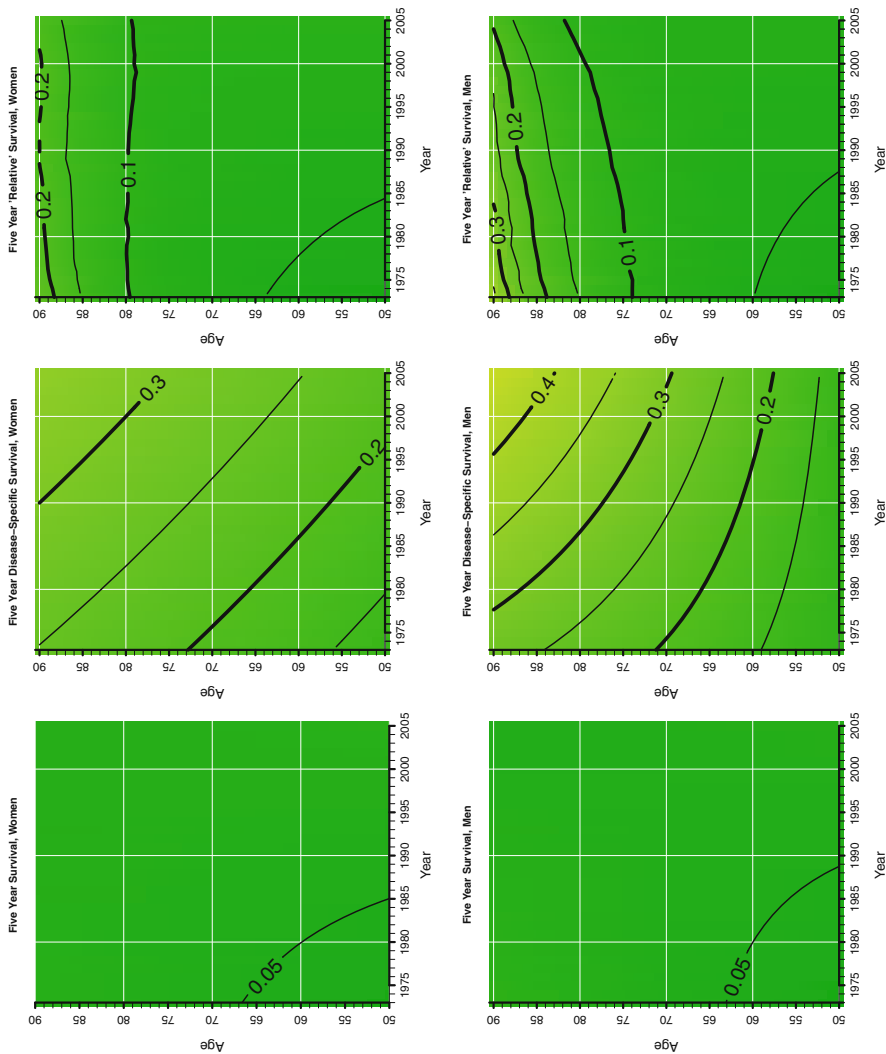
Fig. 10.8 Five year survival for pancreatic cancer at ages 50–90 from 1973–2005. *Left column*: women; *right panel*: men. *Upper panels*: Probability to survive for 5 years after diagnosed with pancreatic cancer (any cause). *Middle panels:* Probability of not dying from pancreatic cancer within 5 years after diagnosis. *Lower panels*: Five year survival of women or men diagnosed with pancreatic cancer in relation to five year survival of women or men in the general population (Data Source: SEER and Human Mortality Database)
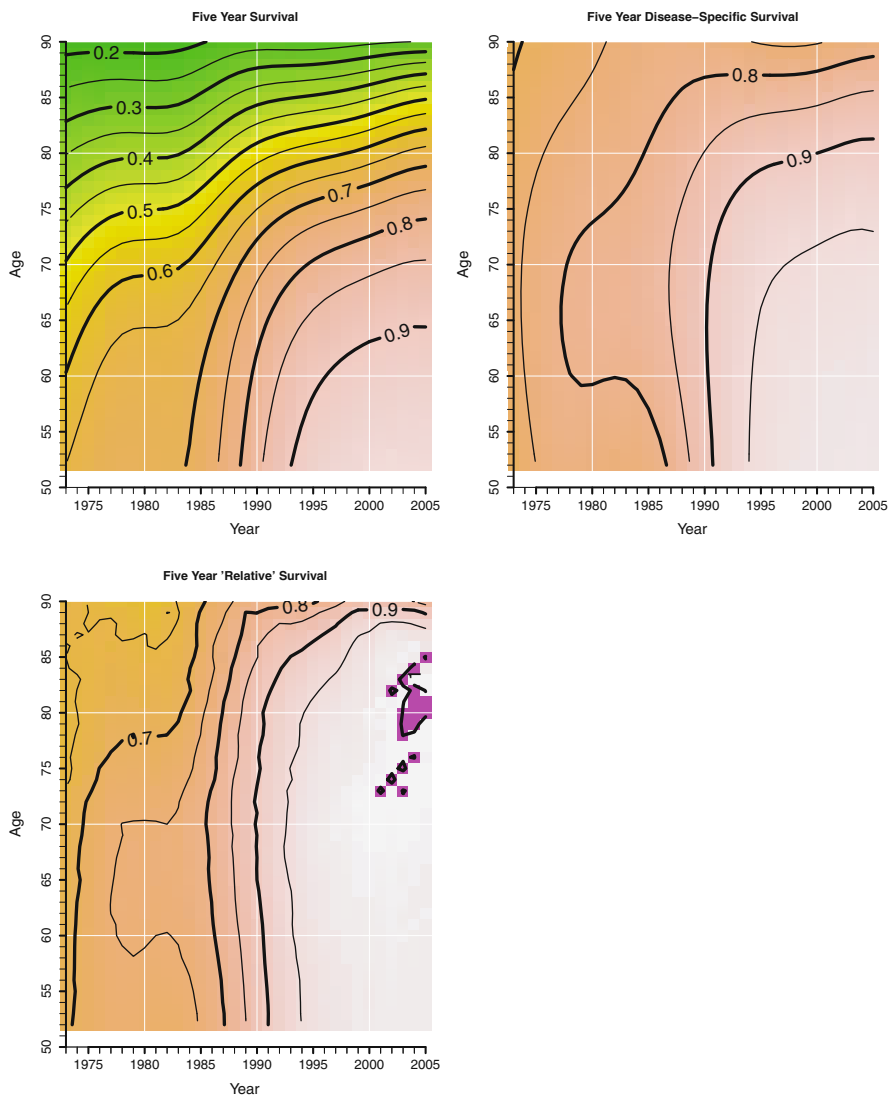
Fig. 10.9 Five year survival for prostate cancer at ages 52–90 from 1973–2005. *Upper left panel*: Probability to survive for 5 years after diagnosed with prostate cancer (any cause). *Upper right panel*: Probability of not dying from prostate cancer within 5 years after diagnosis. *Lower left panel*: Five year survival of men diagnosed with prostate cancer in relation to five year survival of men in the general population (Data Source: SEER and Human Mortality Database)
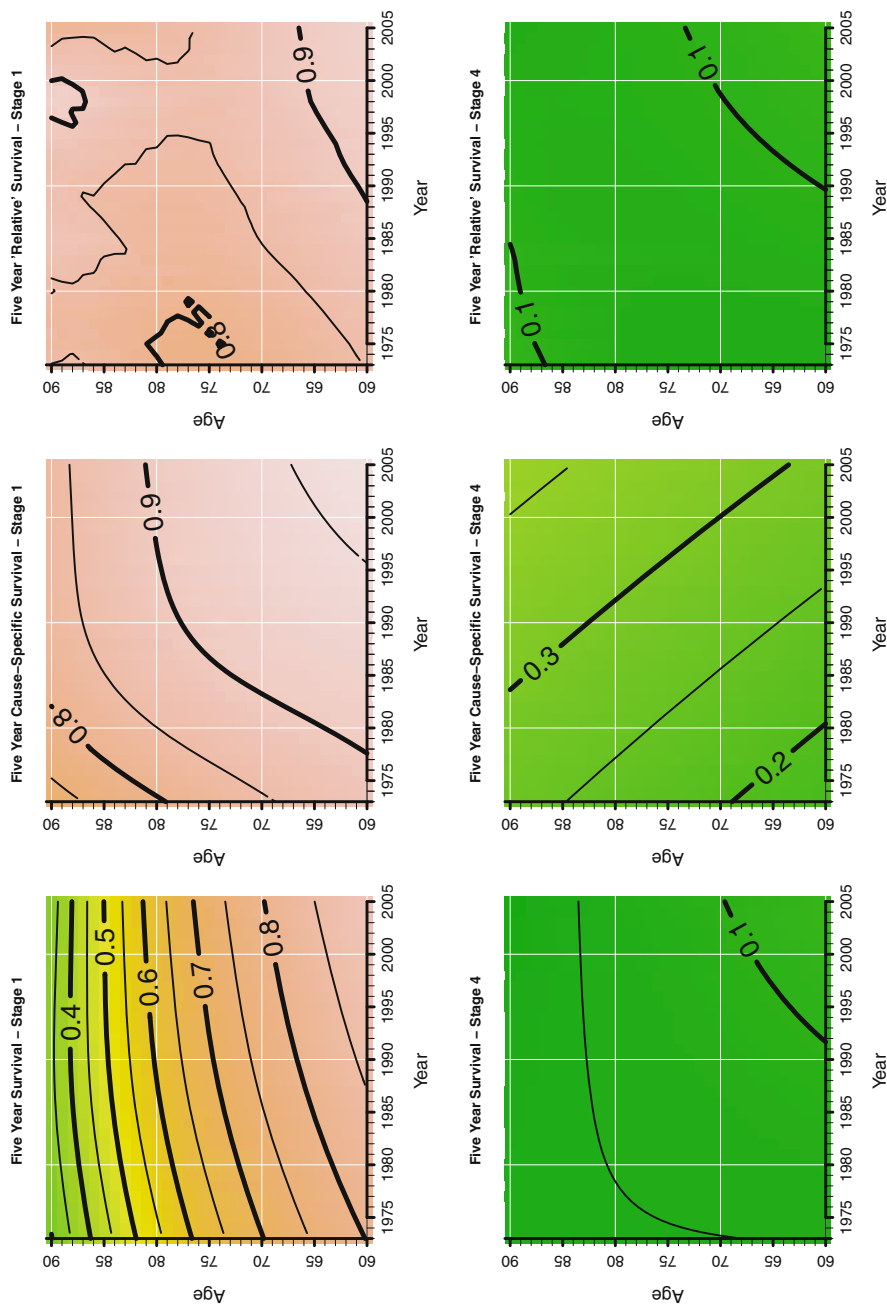
Fig. 10.10 Five year survival for colorectal cancer at ages 60–95 from 1973–2005 by stage. *Upper row*: Stage 1, localized cancer. *Lower row*: Stage 4, distant cancer. *Left panels*: Probability to survive for 5 years after diagnosed with colorectal cancer (any cause). *Middle panels*: Probability of not dying from colorectal cancer within 5 years after diagnosis. *Right panels*: Five year survival of women diagnosed with colorectal cancer in relation to five year survival of women in the general population (Data Source: SEER and Human Mortality Database)
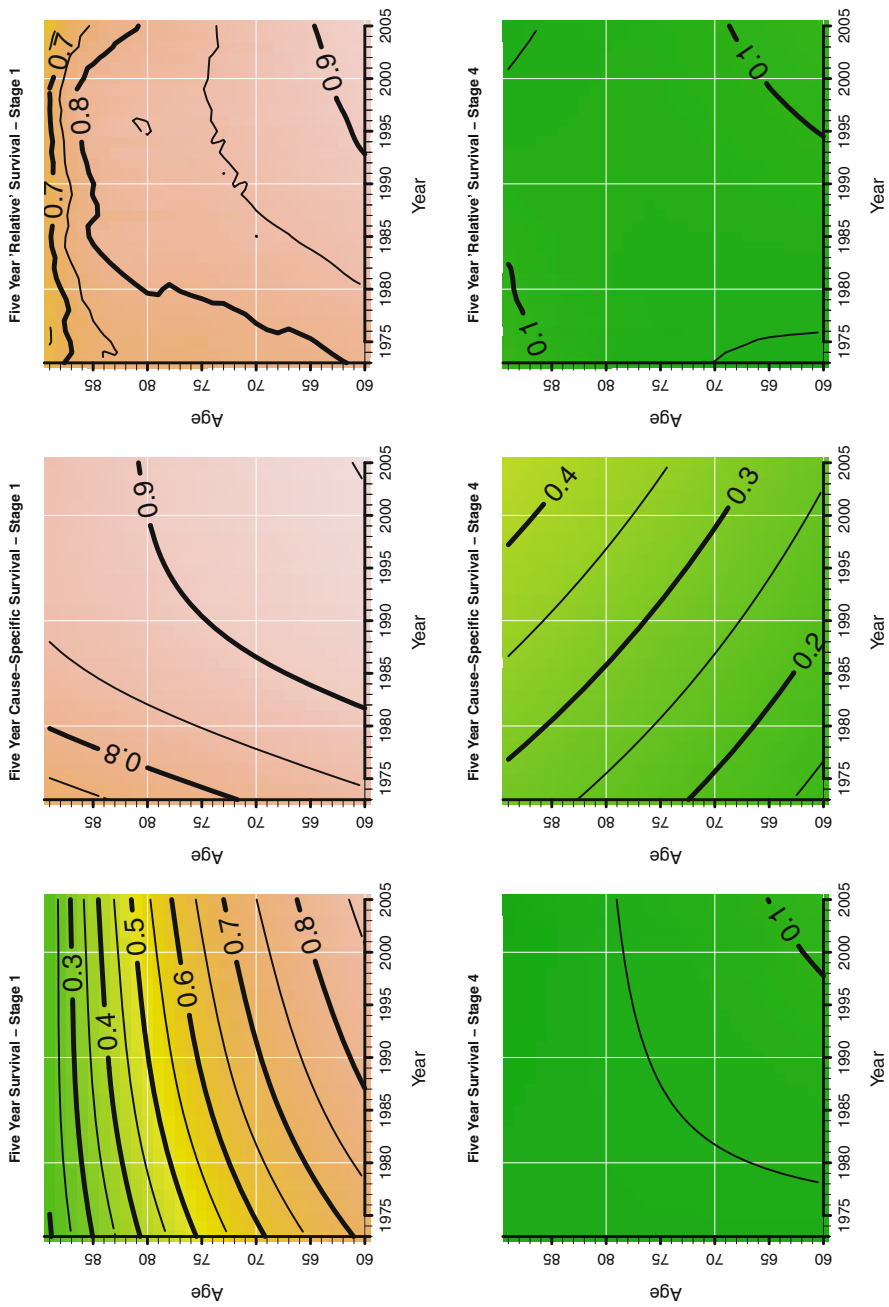
Fig. 10.11 Five year survival for colorectal cancer at ages 60–89 from 1973–2005 by stage. *Upper row*: Stage 1, localized cancer. *Lower row*: Stage 4, distant cancer. *Left panels*: Probability to survive for 5 years after diagnosed with colorectal cancer (any cause). *Middle panels*: Probability of not dying from colorectal cancer within 5 years after diagnosis. *Right panels*: Five year survival of men diagnosed with colorectal cancer in relation to five year survival of men in the general population (Data Source: SEER and Human Mortality Database)

and death: Relative survival is about ten to 20% lower than in the general population when being diagnosed at an early stage (upper three panels in each figure). This excess mortality is pale beside cancer that has already metastasized when being diagnosed (lower three panels in each figure): Only 10% as many people survive the next 10 years as in the general population.

Chapter 11
Summary and Outlook

The goal of this monograph was to show a variety of possibilities to visualize mortality dynamics on the Lexis plane. While we provided examples of raw and smoothed mortality surfaces, our focus was on visualizing rates of mortality improvement ("ROMIs"), i.e., the derivative of age-specific death rates with respect to time. We provided ROMI examples for national populations covered by the Human Mortality Database as well as for selected causes of death in the United States. These "ROMI-plots" were quite instructive to detect period and cohort effects. We also illustrated how changes in age-specific mortality contribute to a gain (or loss) in life expectancy. In Chap. 9 we decomposed seasonal data for causes of death to investigate whether the seasonal pattern, measured via the amplitude and the peak moment ("phase"), has changed over or age. The previous chapter dealt with survival chances of persons who were diagnosed with cancer.

Despite the large number of figures, our list is obviously not exhaustive; here we want to provide a few more two ideas how the Lexis diagram can be used to illustrate not only mortality dynamics.

Figure 11.1 adapts our ROMI approach to fertility. The top panel contains a surface map of age-specific fertility rates in the eastern part of Germany. Birth counts and corresponding exposures by single year of age and calendar time were downloaded from the Human Fertility Database (2017). The estimates were (again) generated with Camarda's R package for smoothing surfaces with P-splines
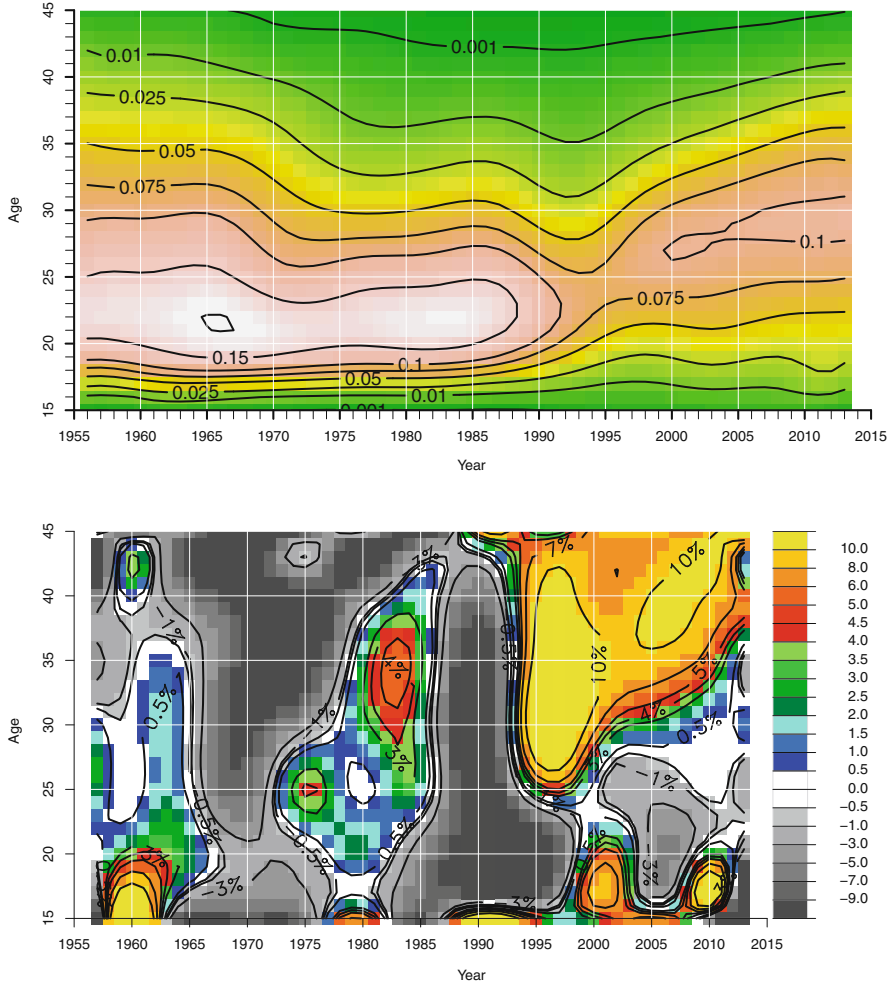
Fig. 11.1 Age-specific fertility in eastern Germany from 1956 until 2013. *Top panel*: Smoothed surface map. *Bottom panel*: Surface map of rates of fertility improvement (Data source: Human Fertility Database)

(Camarda 2012, 2015). The lower panel shows rates of fertility improvement, where improvement means an increase in fertility. Thus, it is the opposite definition of mortality where a decline in mortality was interpreted as an improvement. It is already apparent in the upper panel that fertility dropped considerably a few years after reunification. In 1993 and 1994, the so-called total fertility rate dropped to

0.78 children per woman. This development corresponds to the dark, almost black, vertical area around 1990.[1]

The subsequent recovery of the total fertility rate can be traced back to a strong cohort effect as illustrated by the red, orange, and yellow triangular area starting in about 1995 at ages 25 and above. It is equally interesting that age-specific fertility of younger women (aged about 20–24) has not gone back to pre-reunification levels but continues to decrease. The figure also shows that a similar development of a sudden decline and recovery was experienced already in during 1960s and 1970s.

The last example we want to provide is for the third main parameter in demography: Migration. In Fig. 11.2 we depicted the smoothed age-profile of immigrations of men in Sweden. We estimated for each year from 1968 to 2016 the relative frequencies of each single age. The corresponding plot based on unsmoothed data is included in the appendix in Fig. A.14. We selected this plot because it shows that the age schedule of immigration movements is rather time-invariant—despite the increase of immigrants in recent years coming to Sweden. Male migrants during the past 50 years were typically 20 to 30 years old when they arrived in Sweden. There was virtually not a single year, when more than 1% at a single age of men, i.e., approximately the expected value of a uniform distribution over age, coming to Sweden were older than 45 years.

Using the Lexis diagram is not restricted to depict dynamics of populations. In principle any phenomenon that can be classified by age and calendar time could be illustrated. One example could be unemployment. We would argue that a plot created analogously to the ROMI-plots could easily reveal how labor market reforms may affect various age-groups differently.

We also would like to point out that a plot in the Lexis diagram is not the answer to any question related to population dynamics; for example, maps might be more suitable for spatial analyses or circular plots for migration flows as popularized among demographers by Abel and Sander (2014).

---

[1]The reader might be surprised that the dark gray areas start already in the late 1980s and may attribute it to the impact of smoothing. Please note that fertility started to decline at several ages already before re-unification in 1990. Thus, the gray areas that show up in the late 1980s can not be traced back completely to the impact of smoothing. Please see Fig. A.13 in the appendix for the corresponding surface maps based on *unsmoothed* age-specific fertility rates.
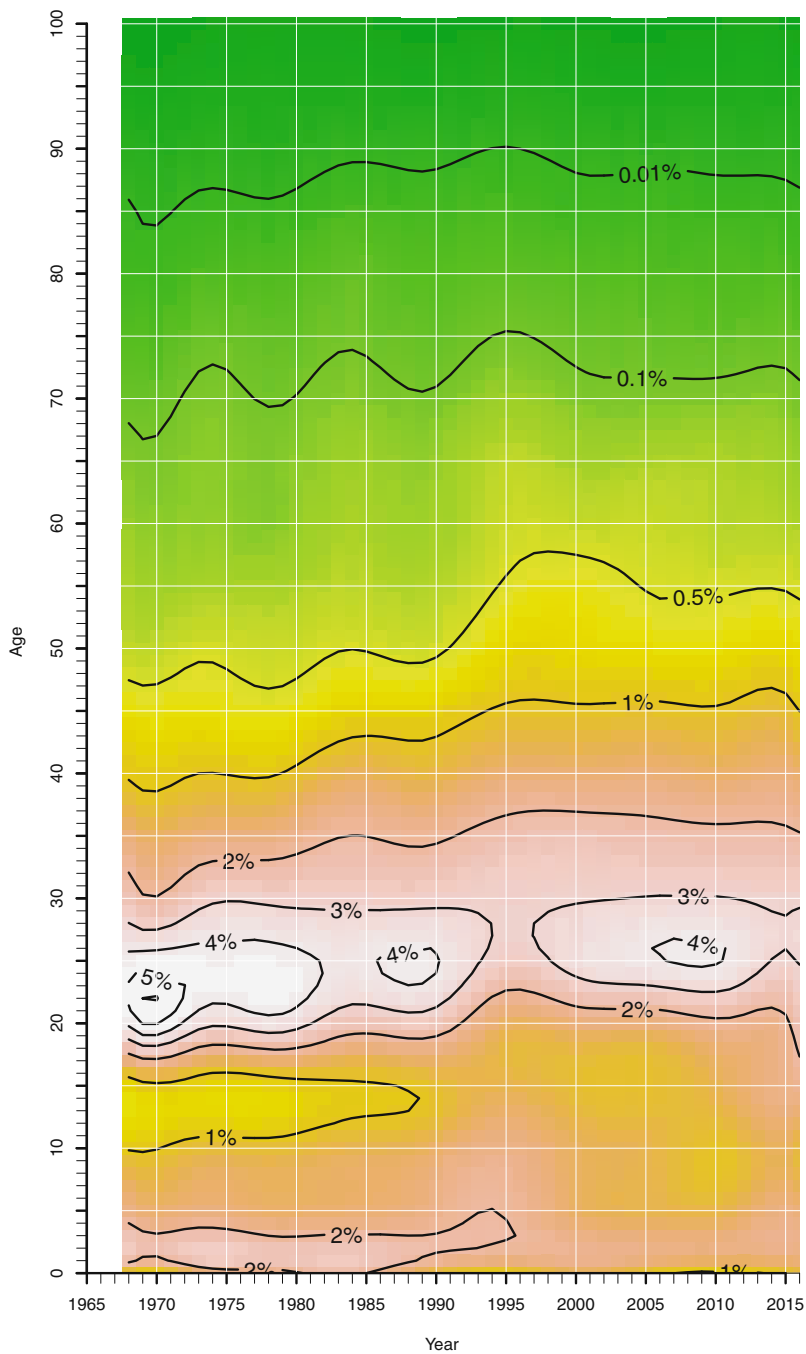
Fig. 11.2 (Smoothed) Age-profile (relative frequencies) of immigrations of men to Sweden, 1968–2016 (Data source: Statistics Sweden)

In the introductory chapter we wrote that the main reasons to visualize data can be summarized to be exploration, confirmation, and presentation. We assume that more exploratory analyses will be conducted in coming years using dynamic graphics as their generation is nowadays greatly facilitated by platforms such as `node.js`, for instance. Nevertheless, we remain confident that plots as the ones contained in this monograph will continue to serve as important tools in all three areas.

Additional Figures

Fig. A.1 "Raw" death rates for women in France, 1950–2014 (Data source: Human Mortality Database)
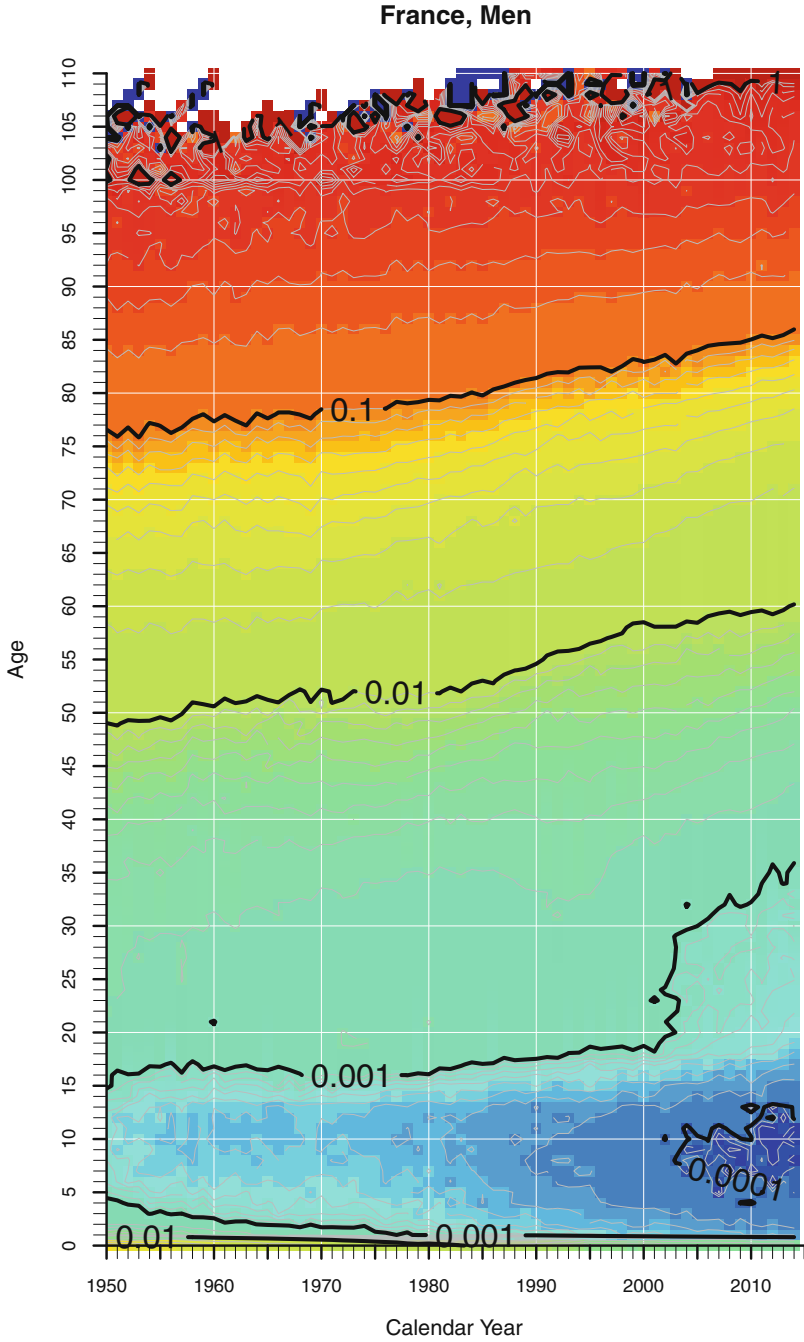
Fig. A.2 "Raw" death rates for men in France, 1950–2014 (Data source: Human Mortality Database)

England & Wales, Women
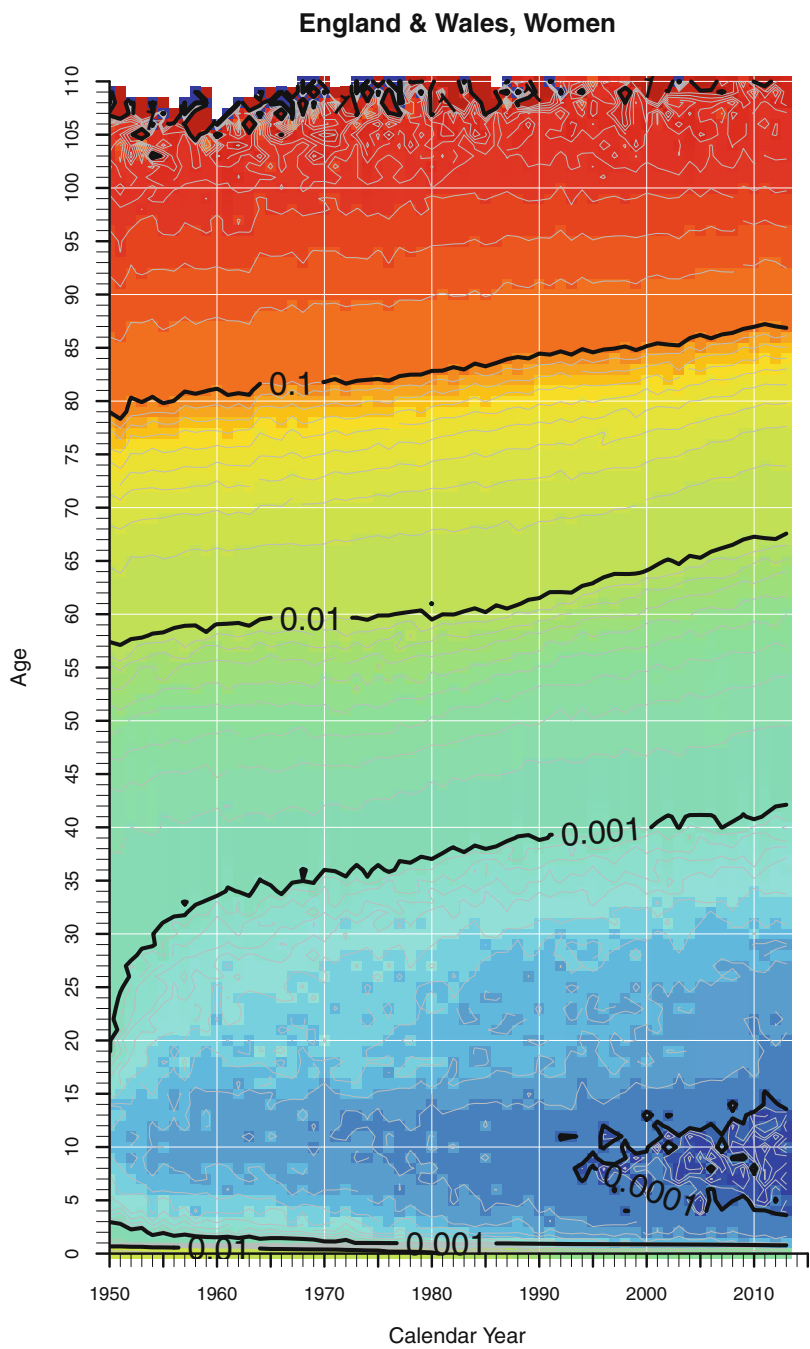
Fig. A.3 "Raw" death rates for women in England & Wales, 1950–2014 (Data source: Human Mortality Database)
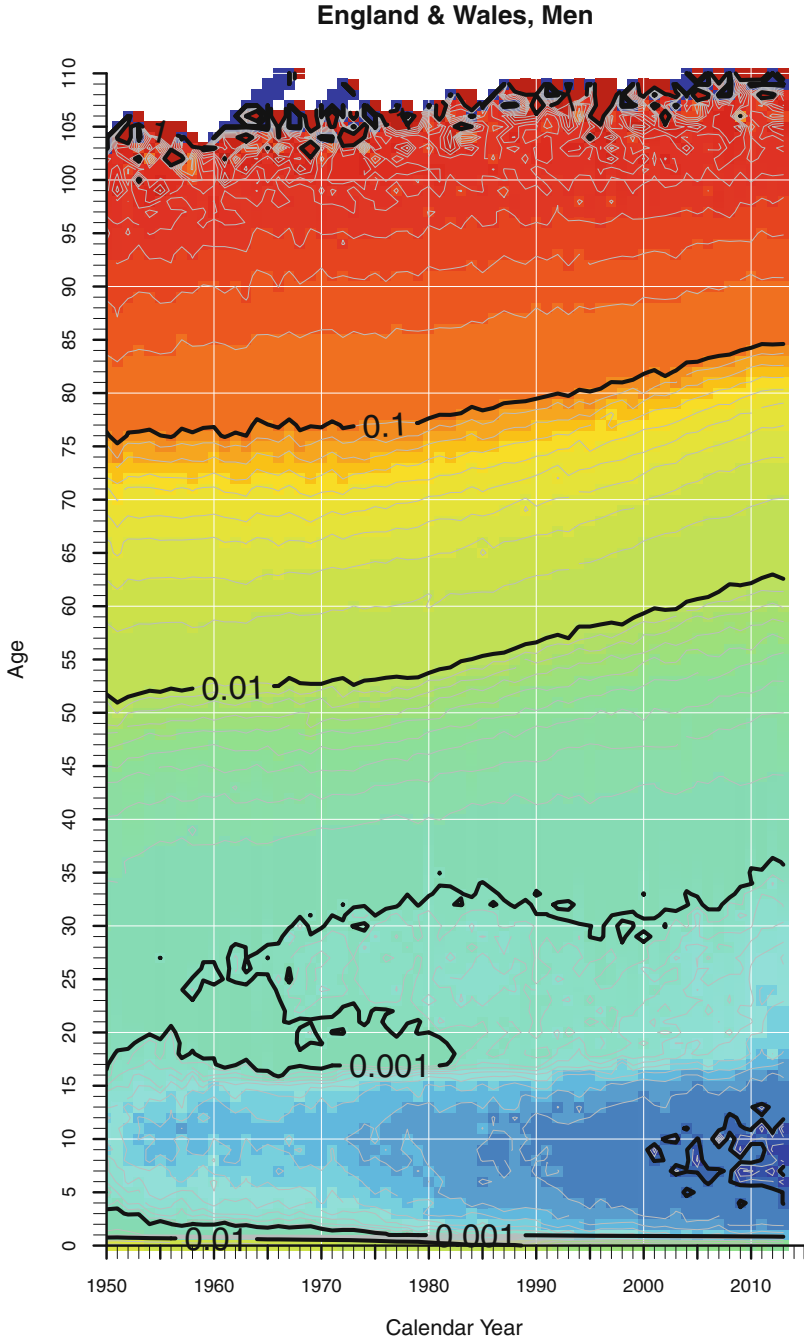
Fig. A.4 "Raw" death rates for men in England & Wales, 1950–2013 (Data source: Human Mortality Database)
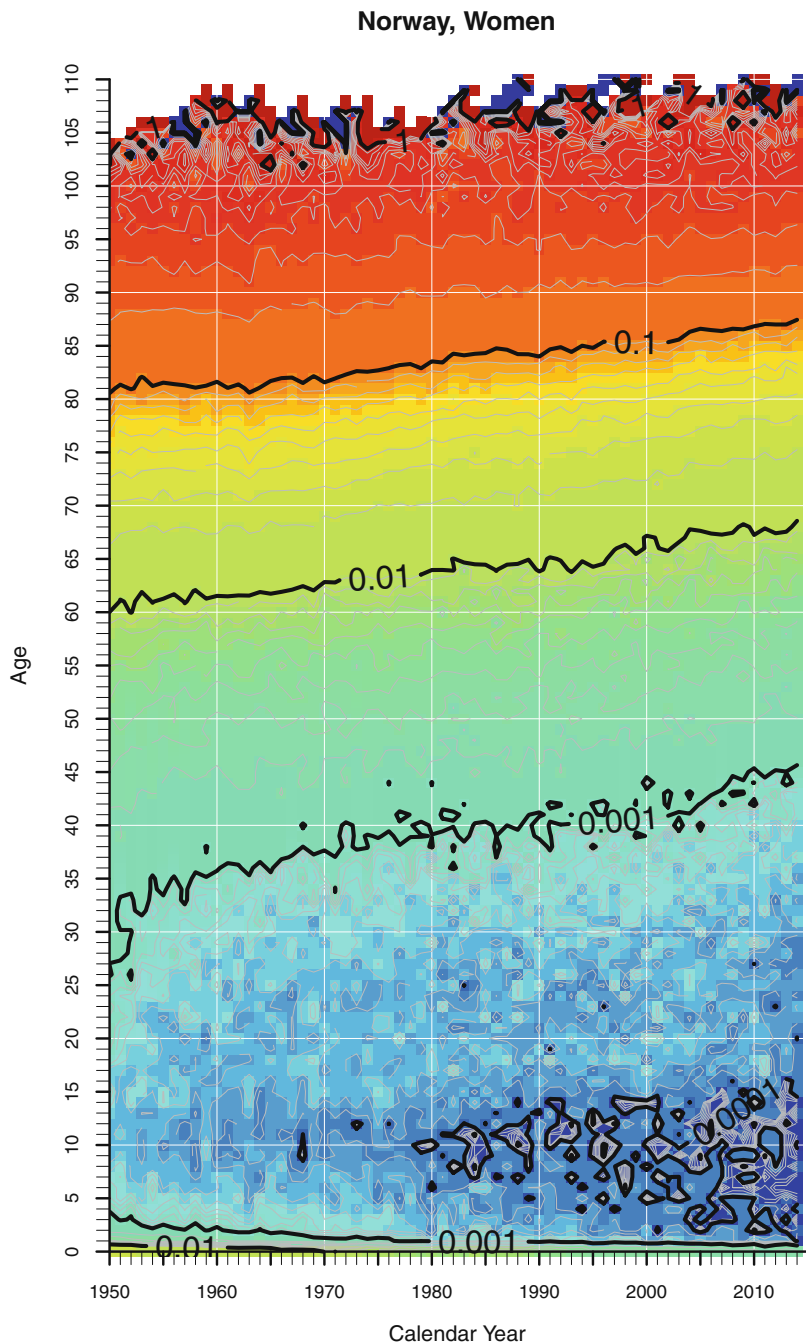
Fig. A.5 "Raw" death rates for women in Norway, 1950–2013 (Data source: Human Mortality Database)
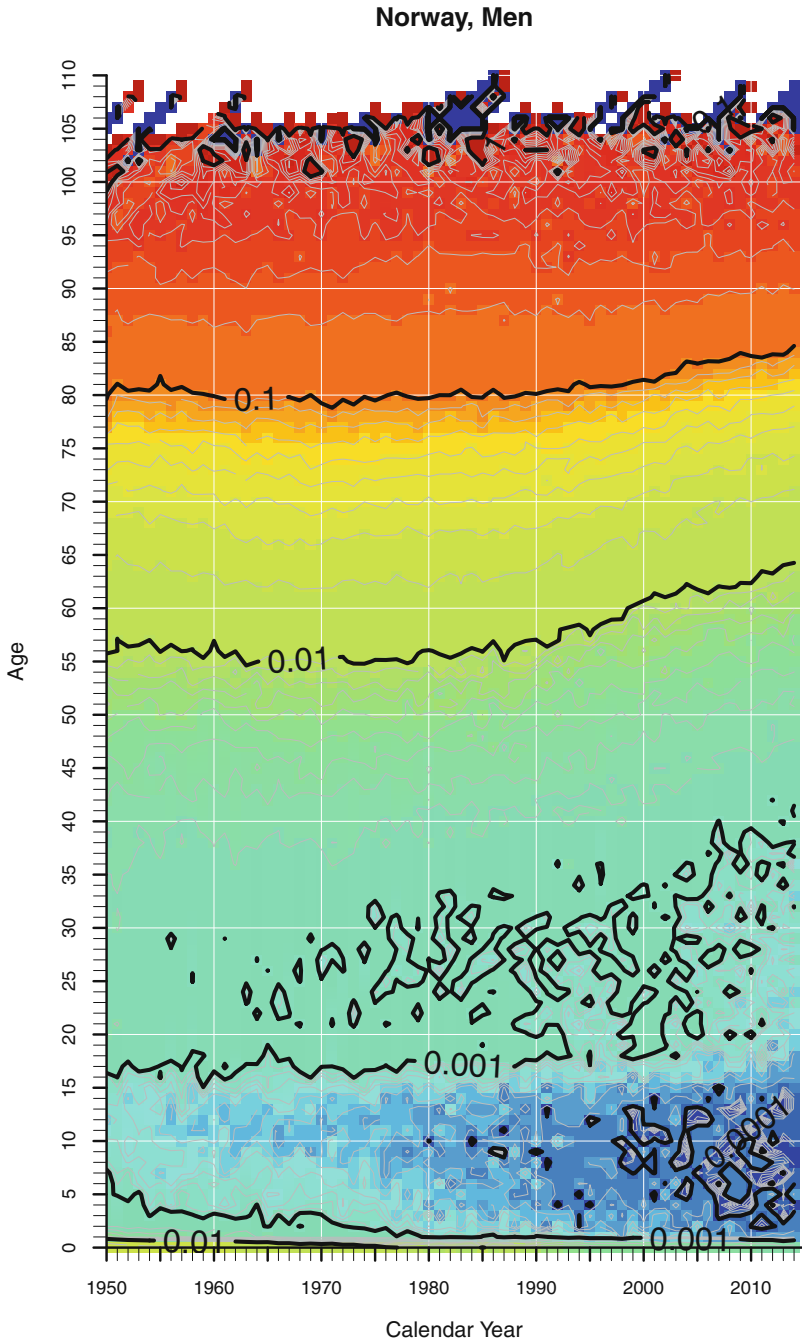
Fig. A.6 "Raw" death rates for men in Norway, 1950–2014 (Data source: Human Mortality Database)

France, Women
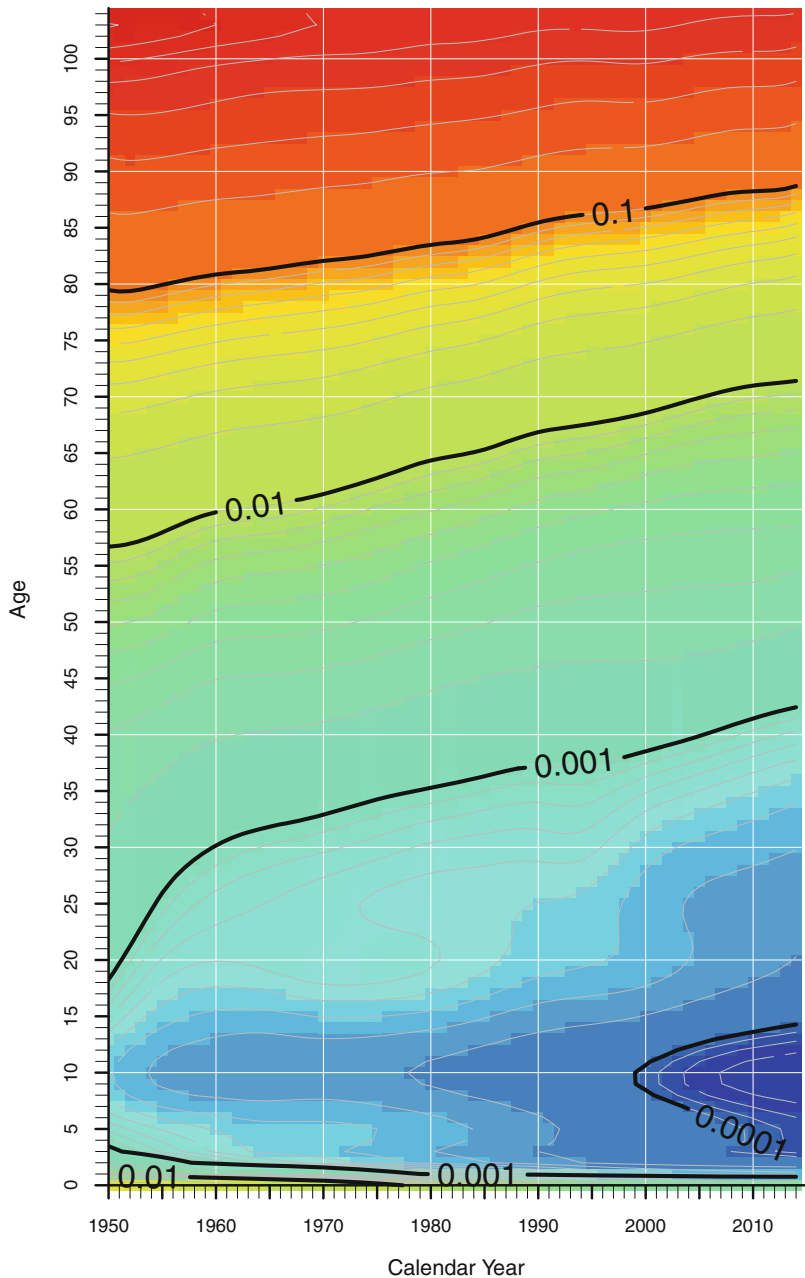
Fig. A.7 Smoothed death rates for women in France, 1950–2014 (Data source: Human Mortality Database)

France, Men
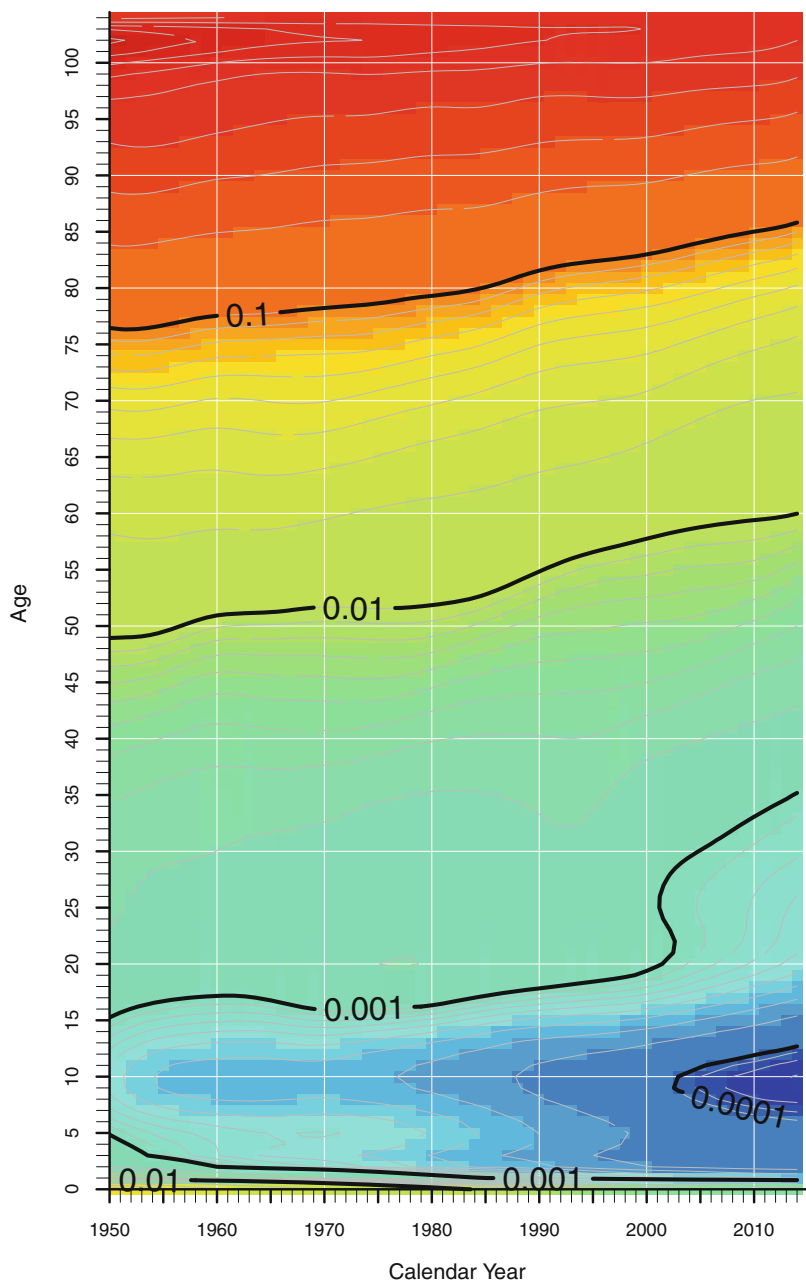
Fig. A.8 Smoothed death rates for men in France, 1950–2014 (Data source: Human Mortality Database)

England & Wales, Women
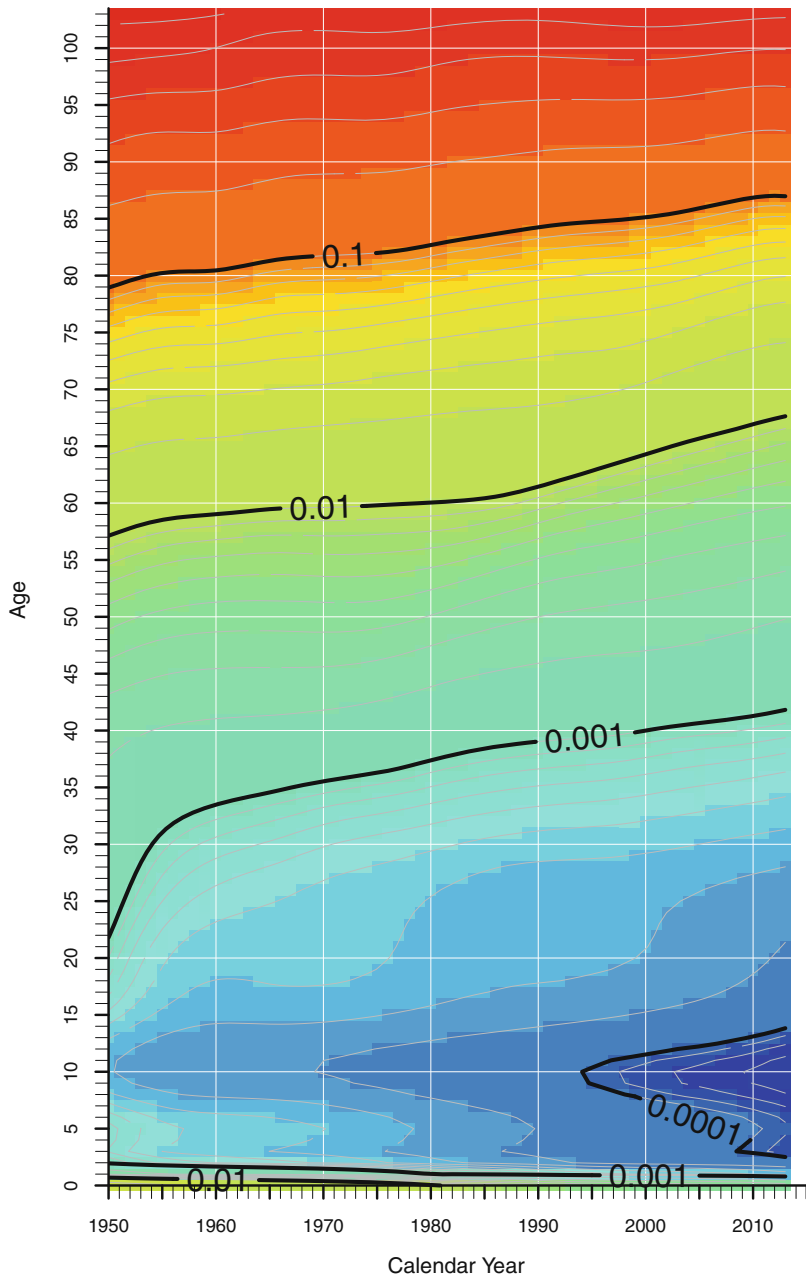
Fig. A.9 Smoothed death rates for women in England & Wales, 1950–2014 (Data source: Human Mortality Database)
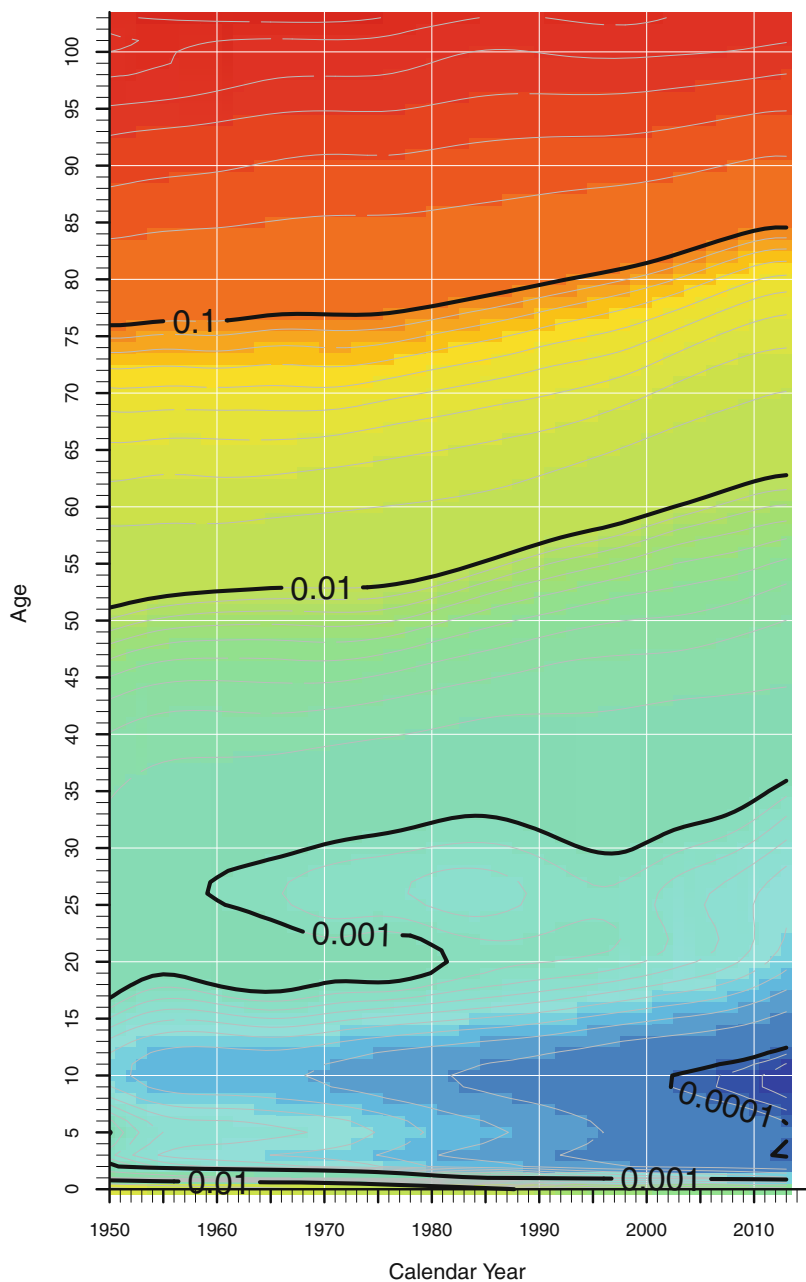
Fig. A.10 Smoothed death rates for men in England & Wales, 1950–2013 (Data source: Human Mortality Database)
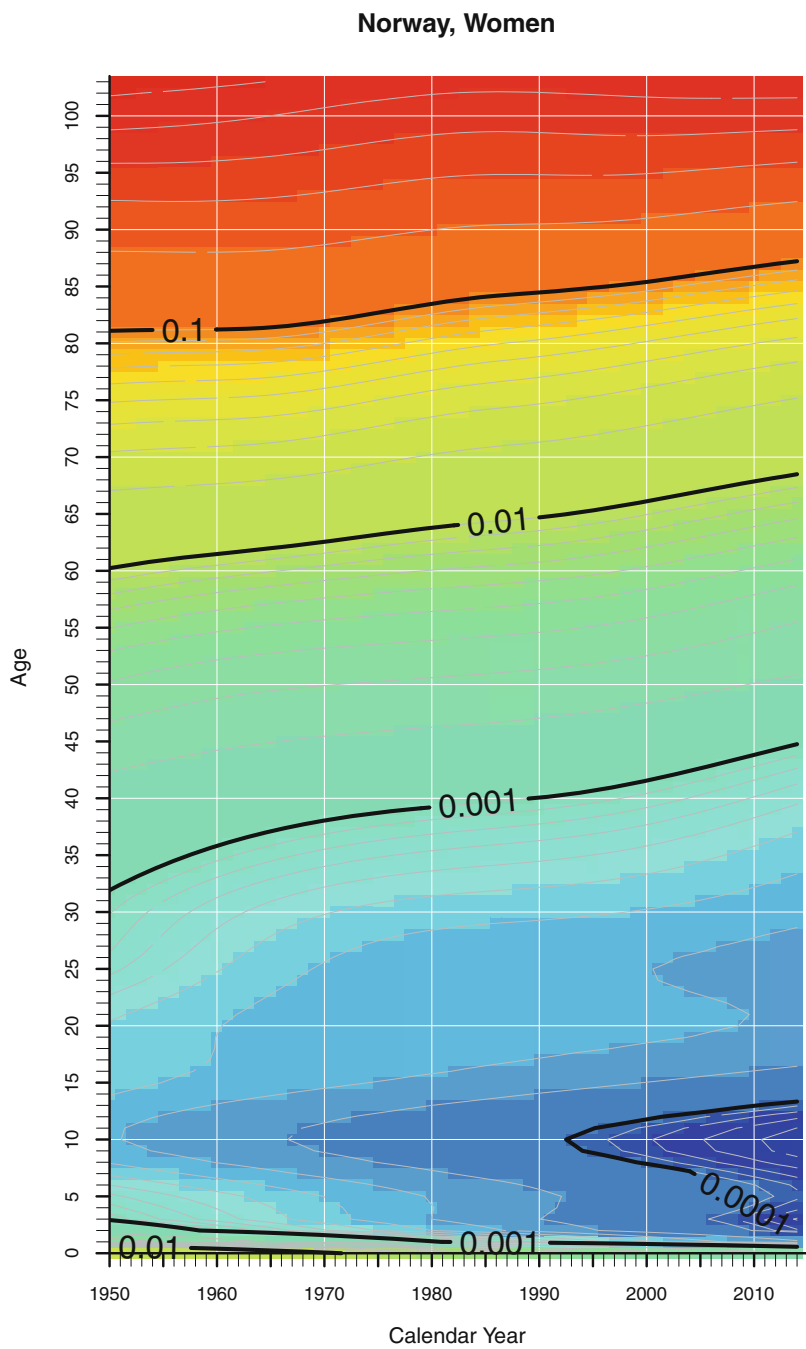
Fig. A.11 Smoothed death rates for women in Norway, 1950–2013 (Data source: Human Mortality Database)
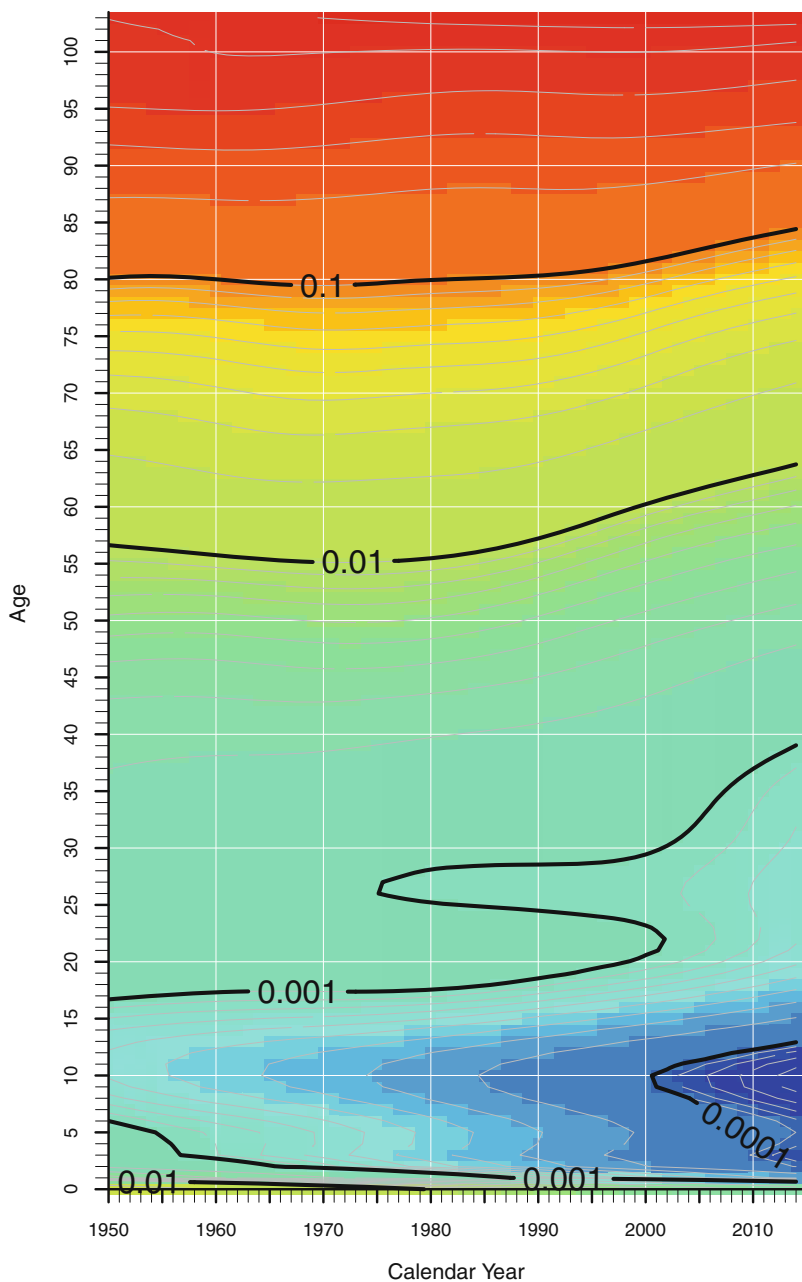
Fig. A.12 Smoothed death rates for men in Norway, 1950–2014 (Data source: Human Mortality Database)
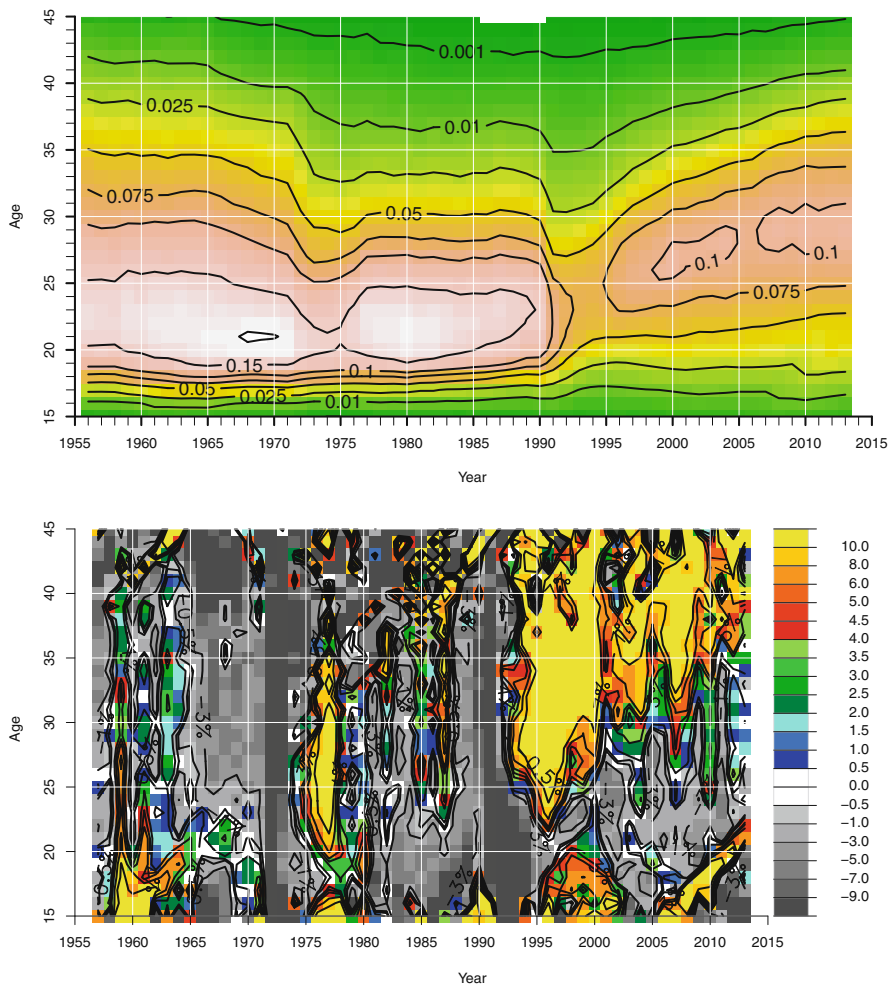
Fig. A.13 Age-specific fertility in eastern Germany from 1956 until 2013. *Top panel*: Unsmoothed surface map. *Bottom panel*: Surface map of rates of fertility improvement based on unsmoothed age-specific fertility rates (Data source: Human Fertility Database)
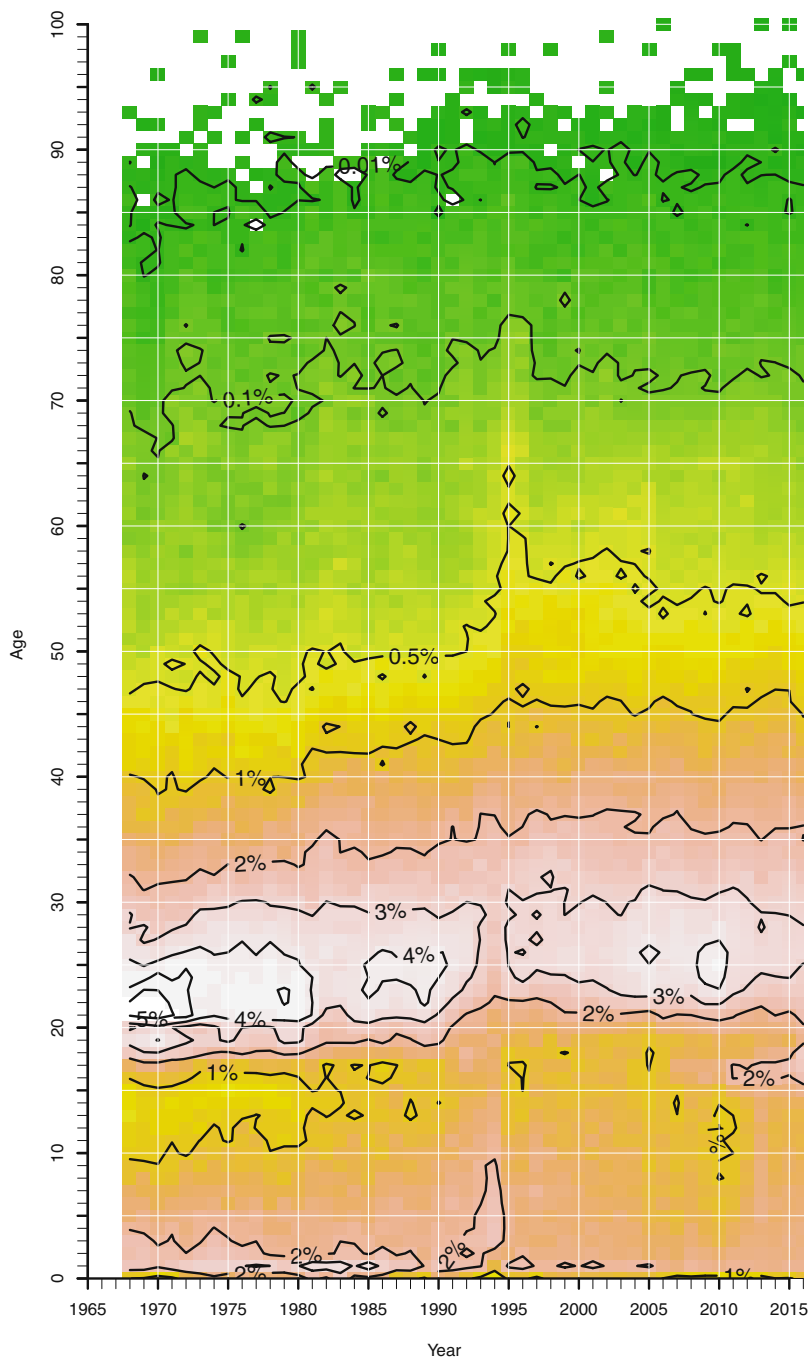
Fig. A.14 (Unsmoothed) Age-profile (relative frequencies) of immigrations of men to Sweden, 1968–2016 (Data source: Statistics Sweden)
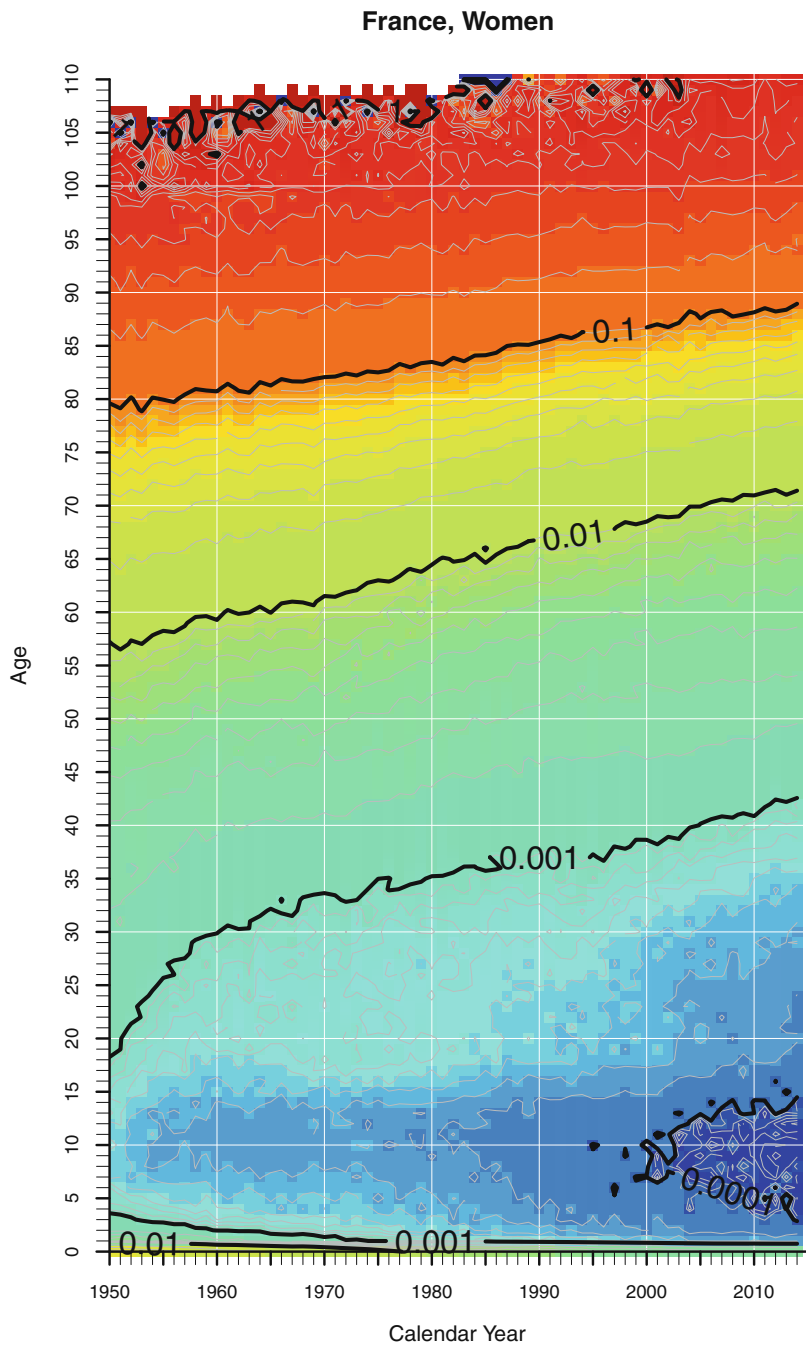
Software: R package `ROMIplot`

A.1 Background, Installation and Requirements

All figures in this monograph have been created using R (R Development Core Team 2015), a free software environment for statistical computing and graphics. The first author of this monograph has written an extension package for R to facilitate the creation of plots of rates of mortality improvement for others (Rau and Riffe 2015). The current version of the package includes code written by Tim Riffe to read data from the Human Mortality Database.

The package is called `ROMIplot` and can be downloaded from any CRAN mirror, the central repository of all R packages, in the canonical way:

```
install.packages("ROMIplot")
```

It needs to be downloaded only once but has to be activated whenever it is needed in an R session via:

```
library(ROMIplot)
```

Apart from the `base` system and the packages `utils`, `graphics`, and `grDevices`—which are all included in any standard distribution of R—package `ROMIplot` has two dependencies, i.e., it requires two additional packages to function properly:

- `MortalitySmooth` to smooth mortality data (Camarda 2015, 2012).
- `RCurl` is an interface to the `libcurl` library that enables accessing data on the internet (Lang and the CRAN team 2015). Package `ROMIplot` uses it to download data from the Human Mortality Database.

A.2   Functions

A.2.1   readHMDformat()

The function `readHMDformat()` requires four input parameters.

- `CNTRY`: A character string denoting the country for which the data should be downloaded. It is specified as an abbreviated name, which follows in most cases the ISO 3166-1 alpha-3 standard. For instance data for Austria can be obtained by setting `CNTRY="AUT"`. There are exceptions if major territorial changes occurred or if subpopulations are available. Examples are `CNTRY="DEUTE"` for data from the (territory of the) former German Democratic Republic or `CNTRY="GBR_NIR"` for data from Northern Ireland. An overview of all possible values for `CNTRY` is given in Table A.1 on page 163.
- `username`: Obtaining data from the Human Mortality Database requires free registration. This is the username to access the data. Typically, this is the email address that was used to register supplied as a character string such as `username="my.name@my.email.com"`.
- `password`: Obtaining data from the Human Mortality Database requires free registration. This is the username to access the data. Typically, this is a 10 digit number supplied as a character string, such as `password="1234567890"`.
- `fixup`: A Boolean value. If set to `TRUE` the obtained data are already "cleaned". For instance, age data from the Human Mortality Database do not include anymore non-numeric values such as "110+". The default setting of this parameter is `TRUE`.

The function returns a list consisting of two data frames.

- `@deaths`: A data frame with deaths by single calendar year (`Year`) and single age year (`Age`) for women (`Female`), men (`Male`) and both of them combined (`Total`).
- `@exposures`: A data frame with exposure-to-risk estimates by single calendar year (`Year`) and single age year (`Age`) for women (`Female`), men (`Male`) and both of them combined (`Total`).

A.2.2   create.Lexis.matrix()

The function `create.Lexis.matrix()` requires six input parameters.

- `HMD.dataset`: A data frame in the format as obtained using the function `readHMDformat()`
- `Sex`: A character string, set by default to `Female` to select data for women. Other possible values are `Male` and `Total`

Table A.1 Abbreviations ("`CNTRY`") and their corresponding country names in the Human Mortality Database

CNTRY	Country	CNTRY	Country	CNTRY	Country
AUS	Australia	FRACNP	France Civil.	POL	Poland
AUT	Austria		National Pop.	PRT	Portugal
BEL	Belgium	HUN	Hungary	RUS	Russia
BGR	Bulgaria	IRL	Ireland	SVK	Slovakia
BLR	Belarus	ISL	Iceland	SVN	Slovenia
CAN	Canada	ISR	Israel	SWE	Sweden
CHL	Chile	ITA	Italy	TWN	Taiwan
CHE	Switzerland	JPN	Japan	UKR	Ukraine
CZE	Czech Republic	LTU	Lithuania	GBR_NP	United Kingdom
DEUTNP	Germany	LUX	Luxembourg		Total Pop.
	National Pop.	LVA	Latvia	GBRTENW	Engl. & Wales
DEUTE	Germany – East	NLD	Netherlands		Total Pop.
DEUTW	Germany – West	NOR	Norway	GBRCENW	Engl. & Wales
DNK	Denmark	NZL_NP	New Zealand		Civil. Pop.
ESP	Spain		National Pop.	GBR_SCO	Scotland
EST	Estonia	NZL_MA	New Zealand	GBR_NIR	N. Ireland
FIN	Finland		Maori	USA	United States
FRATNP	France Total	NZL_NM	New Zealand		
	National Pop.		Non-Maori		

- `minage`: An integer which denotes the *youngest* ages to be included, set by default to 50.
- `maxage`: An integer which denotes the *oldest* ages to be included, set by default to 100.
- `minyear`: An integer which denotes the *earliest* calendar year to be included, set by default to 1950.
- `maxyear`: An integer which denotes the *latest* calendar year to be included, set by default to 2011.

The function returns a matrix that contains the combined number of deaths or exposures for a given combination of calendar year and age. Row names denote ages from `minage` to `maxage`; column names denote calendar years from `minyear` to `maxyear`.

A.2.3 ROMI.plot()

The function `create.Lexis.matrix()` requires up to four input parameters.

- `Dx`: A matrix of death counts, expected to be in the format as prepared by function `create.Lexis.matrix`.

- `Nx`: A matrix of exposure to risk estimates corresponding to argument `Dx`, expected to be in the format as prepared by function `create.Lexis.matrix`.
- `mx`: A matrix of death rates. If death counts and their corresponding exposure estimates are not available, it is possible to provide instead a matrix of death rates. Please note that argument `smooth` requires death counts and exposure estimates. Thus, if only death rates `mx` are available, they can only be used "raw" or have to be already "smoothed".
- `smooth`: A boolean value, set to `TRUE` by default. If set to TRUE, the data are smoothed using *P*-Splines. Please note that this smoothing approach models death counts as a Poisson process using exposure to risk estimates as a (log) offset. Hence, smoothing can not be performed if only death rates are provided (argument `mx`). *P*-Spline smoothing has been introduced by Eilers and Marx (1996) and was extended to two dimensions for mortality by Currie et al. (2004). We use the user-friendly implementation by Camarda (2012, 2015).

Based on the matrix of (smoothed) death rates, function `ROMI.plot()` estimates a matrix of rates of mortality improvement, $\rho_{x,t}$, applying and re-arranging the standard equation for continuous population growth. Since we estimate the rates annually, $t = 1$:

$$m_{x,t+1} = m_{x,t}\, e^{-\rho_{x,t}t}; \quad \rho_{x,t} = -\log_e\left(\frac{m_{x,t+1}}{m_{x,t}}\right)$$

In most applications the returned matrix is not the main interest of the researcher but the plot that is produced as a side effect. Please note that we used the same color scheme as in the present volume. But this is only a suggestion. Our package is free software. Thus, anyone should feel invited to modify and possibly also improve our package as only the most fundamental elements have been included in the current version.

References

Abel, G. J., & Sander, N. (2014). Quantifying global international migration flows. *Science, 343*(6178), 1520–1522.

Andreev, K. F. (2002). *Evolution of the Danish population from 1835 to 2000* (Odense Monographs on Population Aging, Vol. 9), Odense: University Press of Southern Denmark.

Arriaga, E. E. (1984). Measuring and explaining the change in life expectancies. *Demography, 21*(1), 83–96.

Arthur, W. B., & Vaupel, J. W. (1984). Some general relationships in population dynamics. *Population Index, 50*(2), 214–226.

Berkson, J., & Gage, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association, 47*(259), 501–515.

Camarda, C. G. (2008). *Smoothing methods for the analysis of mortality development*. PhD thesis, Universidad Carlos III de Madrid.

Camarda, C. G. (2012). MortalitySmooth: An R package for smoothing Poisson counts with *P*-splines. *Journal of Statistical Software, 50*(1), 1–24.

Camarda, C. G. (2015). *Smoothing and forecasting Poisson counts with P-splines*. http://CRAN. R-project.org/package=MortalitySmooth, R package version 2.3.4.

Canudas-Romo, V. (2003). *Decomposition methods in demography*. PhD thesis, Rijksuniversiteit Groningen, Groningen, NL.

Caselli, G., & Vallin, J. (2006). Frequency surfaces and isofrequency lines. In G. Caselli, J. Vallin, & G. Wunsch (Eds.), *Demography. Analysis and synthesis* (Vol. I, Chap. 7, pp. 69–77). Amsterdam: Elsevier.

Caselli, G., Vaupel, J. W., & Yashin, A. I. (1985). Mortality in Italy: Contours of a century of evolution. *Genus, 41*(1–2), 39–55.

CDC/NCHS. (2015). Mortality in the United States, 2014. *NCHS Data Brief, Number 229*, Dec 2015. Available online as Supplementary Material at: https://www.cdc.gov/nchs/data/ databriefs/db229_table.pdf#1.

Chambers, J. (2008). *Software for data analysis*. New York: Springer.

Cho, H., Howlader, N., Mariotto, A. B., & Cronin, K. A. (2011). *Estimating relative survival for cancer patients from the SEER program using expected rates based on Ederer I versus Ederer II method*. Tech. Rep. 2011-01, National Cancer Institute.

Christensen, K., Doblhammer, G., Rau, R., & Vaupel, J. (2009). Ageing populations: The challenges ahead. *The Lancet, 374*(9696), 1196–1208.

Crimmins, E. M., Preston, S. H., & Cohen, B. (Eds.). (2011). *Explaining divergent levels of longevity in high-income countries*. Washington, DC: The National Academies Press.

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

Deborah, S. (1998). *The da Vinci of Data*. The New York Times, 30 March 1998.

Delaporte, P. (1938). Évolution de la mortalité française depuis un siècle. *Journal de la société de statistique de Paris, 79*, 181–206.

Delaporte, P. (1942). Évolution de la mortalité en Europe depuis l'origine des statisques. *Journal de la société de statistique de Paris, 83*, 183–203.

Drefahl, S., Ahlbom, A., & Modig, K. (2014). Losing ground — Swedish life expectancy in a comparative perspective. *PLOS ONE, 9*(2), e88357.

Ederer, F., Axtell, L. M., & Cutler, S. J. (1961). *The relative survival rate: A statistical methodology* (Chap. 6, pp. 101–121). National Cancer Institute Monograph, National Cancer Institute.

Eilers, P. H. C., & Marx, B. D. (1996). Flexible Smoothing with *B*-splines and Penalties. *Statistical Science, 11*(2), 89–102.

Eilers, P. H. C., Gampe, J., Marx, B. D., & Rau, R. (2008). Modulation models for seasonal incidence tables. *Statistics in Medicine, 27*(17), 3430–3441.

EurowinterGroup. (1997). Cold exposure and winter mortality from ischaemic heart disease, cerebrovascular disease, respiratory disease, and all causes in warm and cold regions of Europe. *Lancet, 349*, 1341–1346.

EurowinterGroup. (2000). Winter mortality in relation to climate. *International Journal of Circumpolar Health, 59*, 154–159.

Feinstein, C. A. (2002). Seasonality of deaths in the U.S. by age and cause. *Demographic Research, 6*, 469–486.

Few, S. (2014). Why do we visualize quantitative data? Available Online At: http://www.perceptualedge.com/blog/?p=1897, Visual Business Intelligence. A blog by Stephen Few. Last verification: 14 June 2017.

Friendly, M. (2008). A brief history of data visualization. In C. H. Chen, W. K. Härdle, & A. Unwin (Eds.), *Handbook of data visualization* (Springer Handbooks of Computational Statistics, Chap. I, pp. 15–56). Berlin/Heidelberg: Springer.

Galilei, G. (1613). *Istoria e dimostrazioni intorno alle macchie solari*. Rome.

Gambill, B. A., & Vaupel, J. W. (1985). *The LEXIS program for creating shaded contour maps of demographic surfaces*. Tech. Rep. RR–85–94. International Institute for Applied Systems Analysis (IIASA), Laxenburg, A.

Hippocrates. (400BC). *On airs, waters, and places*. Translated by Francis Adams. Available online at: http://classics.mit.edu/Hippocrates/airwatpl.html.

Huynen, M. M., Martens, P., Schram, D., Weijenberg, M. P., & Kunst, A. E. (2001). The impact of heat waves and cold spells on mortality rates in the Dutch population. *Environmental Health Perspectives, 109*, 463–470.

Ihaka, R. (1998). *R: Past and future history*. Available online at https://cran.r-project.org/doc/html/interface98-paper/paper.html.

Jacobsen, R., Keiding, N., & Lynge, E. (2002). Long term mortality trends behind low life expectancy of Danish women. *Journal of Epidemiology and Community Health, 56*, 205–208.

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

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

Janssen, F., Nusselder, W. J., Looman, C., Mackenbach, J. P., & Kunst, A. E. (2003). Stagnation in mortality decline among elders in the Netherlands. *Gerontologist, 43*(5), 722–734.

Jdanov, D. A., Jasilionis, D., Soroko, E. L., Rau, R., & Vaupel, J. W. (2008). *Beyond the Kannisto-Thatcher database on old age mortality: An assessment of data quality at advanced ages*. Working paper MPIDR Working Paper WP-20083-013, Max Planck Institute for Demographic Research, Rostock, Germany.

Kannisto, V. (1994). *Development of oldest-old mortality, 1950–1990: Evidence from 28 developed countries* (Monographs on Population Aging, Vol. 1). Odense: Odense University Press.

Kannisto, V., Lauritsen, J., Thatcher, A. R., & Vaupel, J. W. (1994). Reductions in mortality at advanced ages: Several decades of evidence from 27 countries. *Population and Development Review, 20*, 793–810.

Keyfitz, N. (1977). *Applied mathematical demography*. New York: John Wiley & Sons.

Kintner, H. J. (2004). The life table. In J. B. Siegel & D. A. Swanson (Eds.), *The methods and materials of demography* (2nd ed., Chap. 13, pp. 301–340). San Diego: Elsevier.

Kunst, A., Looman, C., & Mackenbach, J. (1990). The decline in winter excess mortality in the Netherlands. *International Journal of Epidemiology, 20*, 971–977.

Lang, D. T., & the CRAN team. (2015). *RCurl: General Network (HTTP/FTP/. . .) Client Interface for R*. http://CRAN.R-project.org/package=RCurl, R package version 1.95-4.7.

de Leeuw, J. (2005). On abandoning XLISP-STAT. *Journal of Statistical Software, 13*(7), 1–5.

Leon, D. A., Chenet, L., Shkolnikov, V. M., Zakharov, S., Shapiro, J., Rakhmanova, G., Vassin, S., & McKee, M. (1997). Huge variation in Russian mortality rates 1984–1994: Artefact, alcohol, or what? *The Lancet, 350*(9075), 383–388. https://doi.org/10.1016/S0140-6736(97)03360-6, http://www.sciencedirect.com/science/article/pii/S0140673697033606.

Lindahl-Jacobsen, R., Rau, R., Jeune, B., Canudas-Romo, V., Lenart, A., Christensen, K., & Vaupel, J. W. (2016). Rise, stagnation, and rise of Danish women's life expectancy. *Proceedings of the National Academy of Sciences, 113*(15), 4015–4020. doi:10.1073/pnas.1602783113, http://www.pnas.org/content/113/15/4015.abstract, http://www.pnas.org/content/113/15/4015.full.pdf.

Luy, M. (2004). Mortality differences between Western and Eastern Germany before and after Reunification. A macro and micro level analysis of developments and responsible factors. *Genus, 60*(3/4), 99–141.

Mackenbach, J., Kunst, A., & Looman, C. (1992). Seasonal variation in mortality in the Netherlands. *Journal of Epidemiology and Community Health, 46*, 261–265.

Marrero, O. (1983). The performance of several statistical tests for seasonality in monthly data. *Journal of Computational Statistics and Simulation, 17*, 275–296.

Mathworks. (2017). *Matlab*. Available at www.mathworks.com.

Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria). (2017). *Human fertility database*. Available at http://www.humanfertility.org.

McDowall, M. (1981). Long term trends in seasonal mortality. *Population Trends, 26*, 16–19.

Meslé, F. (2004). Mortality in Central and Eastern Europe: Long-term trends and recent upturns. *Demographic Research Special Collection, 2*, 45–70.

Meslé, F. (2006). Medical causes of death. In G. Caselli, J. Vallin, & G. Wunsch (Eds.), *Demography. Analysis and synthesis* (Vol. II, Chap. 42, pp. 29–44). Amsterdam: Elsevier.

Meslé, F., & Vallin, J. (1996). Reconstructing long-term series of causes of death. *Historical Methods, 29*, 72–87.

Meslé, F., & Vallin, J. (2006a). Diverging trends in female old-age mortality: The United States and the Netherlands versus France and Japan. *Population and Development Review, 32*, 123–145.

Meslé, F., & Vallin, J. (2006b). The health transition: Trends and prospects. In G. Caselli, J. Vallin, & G. Wunsch (Eds.), *Demography. Analysis and synthesis* (Vol. II, Chap. 57, pp. 247–259). Amsterdam: Elsevier.

Moore, T. B., & Hurvitz, C. G. (2009). Cancers in childhood. In D. A. Casciato & M. C. Territo (Eds.), *Manual of clinical oncology* (Chap. 18, p. 397). Philadelphia: Lippincott Williams & Wilkins.

National Bureau of Economic Research. (1959–2015). *Mortality data — Vital statistics NCHS' multiple cause of death data, 1959–2015*. http://www.nber.org/data/vital-statistics-mortality-data-multiple-cause-of-death.html.

National Cancer Institute. (2017). *NCI dictionary of cancer terms*. Available Online at https://www.cancer.gov/publications/dictionaries/cancer-terms.

National Center for Health Statistics. (1959–2015). *Mortality multiple cause files*. Available online at https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm#Mortality_Multiple.

Pampel, F. C. (2001). Cigarette diffusion and sex differences in smoking. *Journal of Health and Social Behavior, 42*, 388–404.

Parkin, D., & Hakulinen, T. (1991). Analysis of survival. In O. Jensen, D. Parkin, R. MacLennan, C. Muir, & R. Skeet (Eds.), *Cancer registration: Principles and methods* (Chap. 12, pp. 159–176). No. 95 in IARC Scientific Publication, International Agency for Research on Cancer (IARC).

Pechholdová, M. (2009). Results and observations from the reconstruction of continuous time series of mortality by cause of death: Case of West Germany, 1968–1997. *Demographic Research, 21*, 535–568.

Peters, F. (2015). *Deviating trends in Dutch life expectancy*. PhD thesis, Erasmus University Rotterdam, Rotterdam, NL.

Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography. Measuring and modeling population processes*. Oxford, UK: Blackwell Publishers.

Pullum, T. W. (1980). Separating age, period, and cohort effects in white US fertility, 1920–1970. *Social Science Research, 9*(3), 225–244.

Quetelet, A. (1838). De l'influence des saisons sur la mortalité aux différens ages dans la Belgique. M. Hayez, Bruxelles, B.

R Development Core Team. (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org, ISBN:3-900051-07-0.

Rau, R. (2007). *Seasonality in human mortality. A demographic approach* (Demographic research monographs). Heidelberg: Springer.

Rau, R., & Doblhammer, G. (2003). Seasonal mortality in Denmark. The role of sex and age. *Demographic Research, 9*, 197–222.

Rau, R., & Riffe, T. (2015). *ROMIplot: Plots surfaces of rates of mortality improvement*. R package version 1.0.

Rau, R., Jasilionis, D., Soroko, E. L., & Vaupel, J. W. (2008). Continued reductions in mortality at advanced ages. *Population & Development Review, 34*(4), 747–768.

van Rossum, G. (1995). *Python reference manual*. Tech. rep., CWI (Centre for Mathematics and Computer Science), Amsterdam, The Netherlands.

Rutherford, M. J., Dickman, P. W., & Lambert, P. C. (2012). Comparison of methods for calculating relative survival in population-based studies. *Cancer Epidemiology, 36*(1), 16–21.

Schumann, H., & Müller, W. (2000). *Visualisierung. Grundlagen und allgemeine Methoden*. Berlin/Heidelberg: Springer.

Seretakis, D., Lagiou, P., Lipworth, L., Signorello, L. B., Rothman, K. J., & Trichopoulos, D. (1997). Changing seasonality of mortality from coronary heart disease. *Journal of the American Medical Association, 278*, 1012–1014.

Shkolnikov, V. M., Andreev, E. M., McKee, M., & Leon, D. A. (2013). Components and possible determinants of decrease in Russian mortality in 2004–2010. *Demographic Research, 28*(32), 917–950.

Surveillance, Epidemiology, and End Results (SEER) Program. (2014). *Research data (1973–2011)*. Available online at www.seer.cancer.gov. National Cancer Institute, DCCPS, Surveillance Research Program, released April 2014, based on the November 2013 submission.

Talbäck, M., & Dickman, P. W. (2011). Estimating expected survival probabilities for relative survival analysis—Exploring the impact of including cancer patient mortality in the calculations. *European Journal of Cancer, 47*(17), 2626–2632.

Thatcher, R. A., Kannisto, V., & Vaupel, J. W. (1998). *The force of mortality at ages 80 to 120* (Monographs on population aging, Vol. 3). Odense: Odense University Press.

Tufte, E. R. (2001). *The visual display of quantitative data* (2nd ed.). Cheshire: Graphics Press.

Tufte, E. R. (2003). *The cognitive style of PowerPoint*. Cheshire: Graphics Press.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

University of California, Berkeley (USA), and Max Planck Institute for Demographic Research, Rostock, (Germany). (2007). *Methods protocol for the human mortality database*. Available at http://www.mortality.org/Public/Docs/MethodsProtocol.pdf.

University of California, Berkeley (USA), and Max Planck Institute for Demographic Research, Rostock, (Germany). (2017). *Human mortality database*. Available at http://www.mortality.org.

Van Den Berg, G. J. (1860). Befolknings statistik in: Underdaniga berattelse for dren 1856–1860, ny folijd II, 3. Stockholm, Statistika Central-Byrans.

Vandeschrick, C. (2001). The Lexis diagram, a misnomer. *Demographic Research, 4*(3), 97–124. doi:10.4054/DemRes.2001.4.3, http://www.demographic-research.org/volumes/vol4/3/.

Vaupel, J. W., Gambill, B. A., & Yashin, A. I. (1985a). *Contour maps of population surfaces*. Tech. Rep. RR–85–47, International Institute for Applied Systems Analysis (IIASA), Laxenburg, A.

Vaupel, J. W., Gambill, B. A., & Yashin, A. I., Bernstein, A. J. (1985b). *Contour maps of demographic surfaces*. Tech. Rep. RR–85–33, International Institute for Applied Systems Analysis (IIASA), Laxenburg, A.

Vaupel, J. W., Gambill, B. A., & Yashin, A. I. (1987). *Thousands of data at a glance: Shaded contour maps of demographic surfaces*. Tech. Rep. RR–87–16, International Institute for Applied Systems Analysis (IIASA), Laxenburg, A.

Vaupel, J. W., Zhenglian, W., Andreev, K. F., & Yashin, A. I. (1997). *Population data at a glance: Shaded contour maps of demographic surfaces over age and time* (Odense Monographs on Population Aging, Vol. 4). Odense: University Press of Southern Denmark.

Wang, H., & Preston, S. H. (2009). Forecasting United States mortality using cohort smoking histories. *Proceedings of the National Academy of Sciences, 106*(2), 393–398. doi:10.1073/pnas.0811809106, http://www.pnas.org/content/106/2/393.abstract, http://www.pnas.org/content/106/2/393.full.pdf.

Wilmoth, J. R. (2006). Age-period-cohort models in demography. In G. Caselli, J. Vallin, & G. Wunsch (Eds.), *Demography. Analysis and synthesis* (Vol. I, Chap. 18, pp. 227–236). Elsevier, Amsterdam.

Yen, C. J., Chao, C. L., Sung, F. C., Chen, W. J., Liau, C. S., & Lee, Y. T. (2000). Seasonal effects on cardiovascular mortality in older patients. *Age & Ageing, 29*, 186–187.