LA SAPIENZA, UNIVERSITÀ DI ROMA

DOCTORAL THESIS

Insight on mortality compression: a cause-of-death analysis of variability in human longevity in Italy

Author:
Andrea GAMBONI

Supervisor:
Elisabetta BARBI

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in

Demography
School of Statistical Science

October 31, 2017

LA SAPIENZA, UNIVERSITÀ DI ROMA

Abstract

Demography
School of Statistical Science

Doctor of Philosophy

Insight on mortality compression: a cause-of-death analysis of variability in human longevity in Italy

by Andrea GAMBONI

As a result of the mortality transition, humans today live much longer than in the past and also experience a lesser uncertainty about the eventual timing of their death. This lesser uncertainty is the consequence of the compression of the age-at-death distribution: deaths are more and more concentrated in a narrower age range as time passes by.

In this dissertation, I investigate on changes in length of life variability in Italy. Thanks to cause-specific data from 1980 to 2013, I am able to assess the role played by various causes of death on the observed mortality compression, trying to give an epidemiological explanation of trends in lifespan variability and gender differences in lifespan variability. In particular, I focus my attention on two distinct aspects of this phenomenon, that is, the changes in the distribution of deaths over a broad age range and the changes in the distribution of deaths within the old age range.

In order to analyze the whole 1980-2013 period, cause-specific time series were reconstructed because of the changes brought by the introduction of the tenth International Classification of Diseases revision (ICD-10) in 2003.

Contents

Chapter 1

Introduction

In his unfinished opera "Pensées", the French mathematician and philosopher Blaise Pascal states: "*The human being is only a reed, the most feeble in nature; but this is a thinking reed [...] The human being becomes still nobler than that which kills him, because he knows that he is dying, and the advantage that the universe has over him. The universe, it does not have a clue.*" (Pascal, 1670). Awareness of dying has accompanied human beings all over their history, and it is exactly this awareness that, according to Pascal, makes them nobler than anything else.

Along with being conscious of death, human beings experienced the fear of dying. It is not surprising, thus, that speculations on the concept of death can be found in many philosophers of any era. From a mathematical point of view, however, the first noticeable attempt in studying how death occurs in a population has only been made in the XVIIth century by John Graunt. Graunt, who can be probably considered as the first demographer in history, systematically collected mortality data for the London's population and used them to produce life tables (Graunt, 1662). From that moment on, the mathematical study of mortality increased more and more, and it is nowadays of fundamental importance in many disciplines such as demography, economy, biology and medicine.

Until relatively few decades ago, humans experienced short average lifespan, with many not living past the first year of life, and a great uncertainty about the timing of deaths (Blacklow, 2007; Edwards and Tuljapurkar 2005; Fries, 1980; Myers and Manton, 1984; Oeppen and Vaupel, 2002; Vallin and Mesle, 2001). In demography, life expectancy (i.e. the average length of life) is the most used indicator to describe mortality condition of a population. In my opinion, however, it is not less important monitoring also how much the length of life varies among members of a population. Life ends for everybody but not for everybody lasts the same amount of time, and since time is all we have, uncertainty in the timing of death could be probably considered as the most important inequality for humans. That is the reason why this latter concept is the main object of this dissertation. Throughout the dissertation, I will refer to it using different terms, but they all have the same meaning: uncertainty in the timing of death, lifespan variability, lifespan inequality, length of life variability, variability of the age-at-death distribution.

From the end of the XVIIIth century to the beginning of the XIXth century, mortality conditions started to improve noticeably in European countries. Such improvements were mainly driven by progress against infectious diseases and the consequent reduction of many forms of premature mortality, in particular infant and childhood mortality (Caselli et al., 2002;

Country	e_0 (1900)	e_0 (2013)	S_{10} (1900)	S_{10} (2013)
Belgium	46.5	81.1	25.7	13.3
Denmark	51.8	80.3	23.7	12.8
Finland	41.7	80.9	31.3	13.6
France	45.0	82.0	25.6	13.9
Italy	41.8	82.7	31.1	12.3
Netherlands	48.8	81.3	25.2	12.5
Norway	53.5	81.7	23.7	12.8
Spain	41.3	82.5	30.5	12.7
Sweden	52.2	81.9	24.8	12.5
England and Wales	46.3	81.4	25.6	13.1

TABLE 1.1: Life expectancy at birth, e_0, and standard deviation of ages at death above age 10, S_{10}, in ten European countries in 1900 and 2013. Total population. Data source: Human Mortality Database

Corsini and Viazzo, 1997). Thenceforth, mortality continued to decline and nowadays we enjoy longer and less uncertain lifespan.

Table 1.1 gives an idea of the remarkable improvements in either lengthening of average lifespan or decline of lifespan variability. It displays life expectancy at birth, e_0, and standard deviation of ages at death above age 10, S_{10}, values registered in ten European countries in 1900 and 2013 for the total population. In slightly more than a century, average lifespan has been growing up outstandingly: in 1900, among the considered countries, e_0 was in the range of 41.8 years (Italy) and 53.5 years (Norway) while in 2013 it exceeds 80 years in all cases with the highest value of 82.7 reported in Italy. Improvements in reducing the uncertainty in the timing of death, here summarized by S_{10}, are not less outstanding. From the beginning of the XIXth century to nowadays lifespan inequality has been halved, or more, in all countries.

So, even if we are still bound by the chains of mortality, thanks to the progress made, the time of death has considerably been postponed and it is much less aleatory. As a consequence of such improvements, human beings have also changed their way of approaching death. While in the past the death was seen as something not expected, traumatic and random, death is nowadays seen as something related to the unavoidable consumption of human body and the biological limits of human beings.

Starting from the 1980s, researchers have made several efforts to deepen the knowledge of changes in lifespan variability. Currently, an open question is whether, after decades of improvements, lifespan variability will decline further or not. Some countries, especially Japan, are indeed showing in the most recent years a slow down in the pace of decline or even constant trends (Bergeron-Boucher et al., 2015; Canudas Romo, 2008; Cheung et al., 2007; Zurieck, 2010). Answering this question is quite complicated. As it will be shown in the Chapter devoted to the literature review, the

main issue is that variability of the age-at-death distribution can be defined in many ways and, therefore, many are the indicators that can be used to measure it. As a consequence, it is difficult to give a unique answer to the question.

Despite a growing number of studies, the topic of variability in length of life has almost always been tackled using all-cause mortality data. Only very recently, few studies have tried to document lifespan inequality trends using a cause-specific approach (Horiuchi et al., 2012; Nau and Firebaugh, 2014). Thus, the knowledge on the role played by causes of death in shaping the variability of the age-at-death distribution is still very poor. Using a cause-of-death perspective, this dissertation intends to fill this lack of information and provide new epidemiological understanding about the greater and greater certainty of our lifespans. The all-cause age-at-death distribution, d_x, can just be seen as what results summing up the various cause-specific age-at-death distribution, $d_{x,k}$. Figure 1.1 helps in visualizing such decomposition. It reports for Italian males and females in 1980 and 2013, the age-at-death distribution of diseases of the circulatory, respiratory and digestive system, neoplasms, external causes and other causes of death (i.e. the remaining causes grouped all together.) These are the single components of the all-cause age-at-death distribution and thus, looking at their changes permits to get new insights on the evolution of lifespan variability over the course of time. In particular, taking advantage of Italian data on causes of death from 1980 to 2013 provided by the Italian national institute of statistics (ISTAT), this dissertation aims to revisit lifespan variability trends in Italy in the light of information coming from cause-specific analysis.

Before presenting my results, I provide a review of literature of all main topics covered in the dissertation (Chapter 2) and details on the data, measures and methods used (Chapter 3). The literature review, apart from exploring the current state of scientific knowledge in lifespan variability, also offers a revision of the demographic and epidemiological transition in Italy and introduces the issue of harmonizing different International Classification of Diseases (ICD). Indeed, since Italian data on causes of death have been recorded using the ninth ICD revision (ICD-9) from 1980 to 2002 and using the tenth revision (ICD-10) from 2003 to 2013, to get consistent results the statistical disruption introduced by the new classification must be fixed. With regard to data and methodological issues presented in Chapter 3, I give particular attention to the reasons behind the choice of standard deviation of ages at death above age 10, S_{10}, and standard deviation of ages at death above the mode, *SD(M+)*, as indicators to monitor lifespan variability in Italy and I highlight how recent advances in sensitivity analysis and decomposition techniques unable to explore the effects of changes in age and cause-specific mortality rates on measures of lifespan inequality.

Chapter 4 is organized into two main sections. In the first, after having identified and analyzed the statistical disruption in cause-specific time series introduced by the adoption of ICD-10, I present the new reconstructed and coherent time series for 17 causes of death. The reconstruction represents the fundamental step for any further analysis reported in the dissertation. In fact, only having in hand the new coherent time-series, it would be possible to study the evolution of lifespan variability in Italy from 1980 using a cause-specific perspective. The second section offers a revisitation

FIGURE 1.1: Cause-specific age-at-death distributions. Calculations are based on smoothed data.

of mortality by cause of death in Italy in the light of reconstruction results. The information resulting from this analysis, giving an overview of cause-specific mortality in Italy, will be helpful, in later chapters of the thesis, to better comprehend lifespan variability dynamics. In this sense, of particular interest is the outcome coming from the investigation on the mortality improvement's pattern which shows how mortality is not declining uniformly either across ages or causes of death.

In Chapter 5 and Chapter 6 I tackle the main goal of my work. As mentioned above, variability in length of life can be defined in different ways and according to the definition various indicators have been developed. In Chapter 5, using S_{10} to measure lifespan variability, I am interested in looking at how the shape of the age-at-death distribution is changing as a whole. S_{10} perfectly fits this aim, taking into account a vast age range [1]. On the other hand, being in the longevity extension era, it is also essential monitoring how lifespans are distributed among the old. For this reason, in Chapter 6 lifespan variability is measured using the standard deviation of ages at death above the mode, *SD(M+)*, focusing the attention on the right-hand tail of the distribution. Thus, Chapter 5 and Chapter 6 consider two distinct features of variability in length of life, but they both have in common the cause-of-death approach. Stressing the role played by different causes of death reveals a new understanding of trends in lifespan variability and also offers an epidemiological explanation about the current gender gap in favor of females.

Finally, in Chapter 7, I draw my conclusion, discussing the implications of my results, summarizing the limitations of my study and outlining avenues for further research.

[1] Excluding ages until 10 has some advantages as it will be shown in Chapter 3.

Chapter 2

Literature Review

My review of literature focuses on three main topics: 1) variability in length of life 2) demographic transition and related epidemiological transition in Italy 3) the issue of harmonizing different International Classification of Diseases (ICD).

The first part of the review offers an overview about lifespan variability which represents the primary subject of this dissertation. Particular attention will be given to: how the concept of variability in length of life has been interpreted in demographic studies and its relationship with mortality models; which indicators have been developed to measures lifespan variability; the implications for humans being of changing in lifespan variability over the course of time; how the study of variability in length of life has been related to causes of death. As specified in the introductive chapter, my analysis of lifespan variability focuses on the Italian situation using a cause-of-death approach. Therefore, the target of the second part of the literature review is the demographic and epidemiological transition in Italy. Finally, since two different ICD are involved in the data used in this dissertation, the third part illustrates how different ways of classifying diseases over the course of time have introduced statistical disruptions in the death's time series and the solutions developed to solve this problem.

2.1 Variability in length of life

Mortality transition's majors benefits for humans being are basically two: a longer life on average and a greater certainty about the timing of death. In fact, life expectancy has been steadily increasing in the last two centuries (Oeppen and Vaupel, 2002) and the present variability of age at death is much lower than the variability of age at death before the mortality transition (Edwards and Tuljapurkar, 2005; Kannisto, 2000; Wilmoth and Horiuchi, 1999). As stated above, in this dissertation I intend to study the latter benefit: less uncertainty in the timing of death.

Variations in length of life can be regarded as one of the most important inequalities for humans. Decreasing trends in the variability of age at death, known as mortality compression in the demographic literature, have created a lot of interest among researchers in this field (Eakin and Witten, 1995; Fries, 1980; Myers and Manton 1984; Rothenberg et al., 1991). Experiencing greater certainty about the eventual timing of death empowered humans with a greater sense of control over their own lives. Moreover, measurement of transformations in lifespan inequality enable governments and policymakers to ensure sustainability of social security and health-care

systems (Ouellette and Bourbeau, 2011). The implication of changes in certainty of timing of death are numerous and regard not only demography but also psychology and economy as we will see later.

In developed countries, mortality compression was mainly driven by the huge decline in death rates in infant, childhood and young adult ages during the first part of the XXth century. In the second half, lifespan inequality continued to go down thanks to the considerable decline of adult-age mortality (Edwards, 2009 ; Kannisto et al. 1994). In these last decades, however, the decrease in lifespan variability was slower than at the beginning of XXth century and a new scenario of shifting mortality (a situation of constant variability with the age-at-death distribution shifting to the right retaining its shape) has been hypothesized by some scholars (Boongaarts, 2005; Canudas-Romo, 2008).

2.1.1 Measures of mortality compression

Kannisto defines mortality compression *as occurring when a given proportion of deaths take place in a shorter age interval than before* (Kannisto, 2000). Using this definition, mortality compression can be documented examining changes in the age-at-death distribution, d_x, of either period or cohort life tables. An alternative and entirely complimentary way of examining trends in mortality compression, it is to look at changes in the survivorship distribution, l_x, across period or cohort life tables. From this perspective, trends in mortality compression are assessed by measuring the degree of rectangularization of the survival curve. The survival curve, indeed, becomes increasingly rectangular over time as more members of the cohort survive to older ages. These two ways of analysis are displayed in Figure 2.1 which represents the age at death distribution, d_x, and the survivorship distribution, l_x, for the French population in three different points of time: 1910, 1960 and 2010. In both cases, the decline in lifespan variability over time is evident and can be detected by a simple graphical inspection.

To precisely assess the degree of mortality compression (and/or the rectangularization of the survival curve) several measures have been developed in the last decades. Cheung et al. have done a nice review of these measures dividing the existing indicators into seven categories (Cheung et al. 2005). Table 2.1 provides the classification of Cheung et al., including also $e^\dagger$ (life expectancy lost due to premature death) and *SDM* (Standard deviation around the mode) which are not reported in the original paper.

Even if mortality compression has been broadly studied, there has not been any agreement on how to measure it. Moreover, the results obtained for different measures do not always show a perfect consensus on the magnitude of lifespan variability decline and the timing of its emergence. In their analysis of ten measures of mortality compression, Wilmoth and Horiuchi, have pointed out that exists a high degree of correlation among these measures but, in spite of this, the discordance in their trends is often undeniable (Wilmoth and Horiuchi, 1999).

Trends in four measures ($e^\dagger$, H[1], S_{10} and *IQR*) in Japan are shown in Figure 2.2. The measures offer different interpretation of the history of mortality compression in this country from 1954 and 2012. For example, according

[1] H's values multiplied by 100 to make them comparable with other indicators.

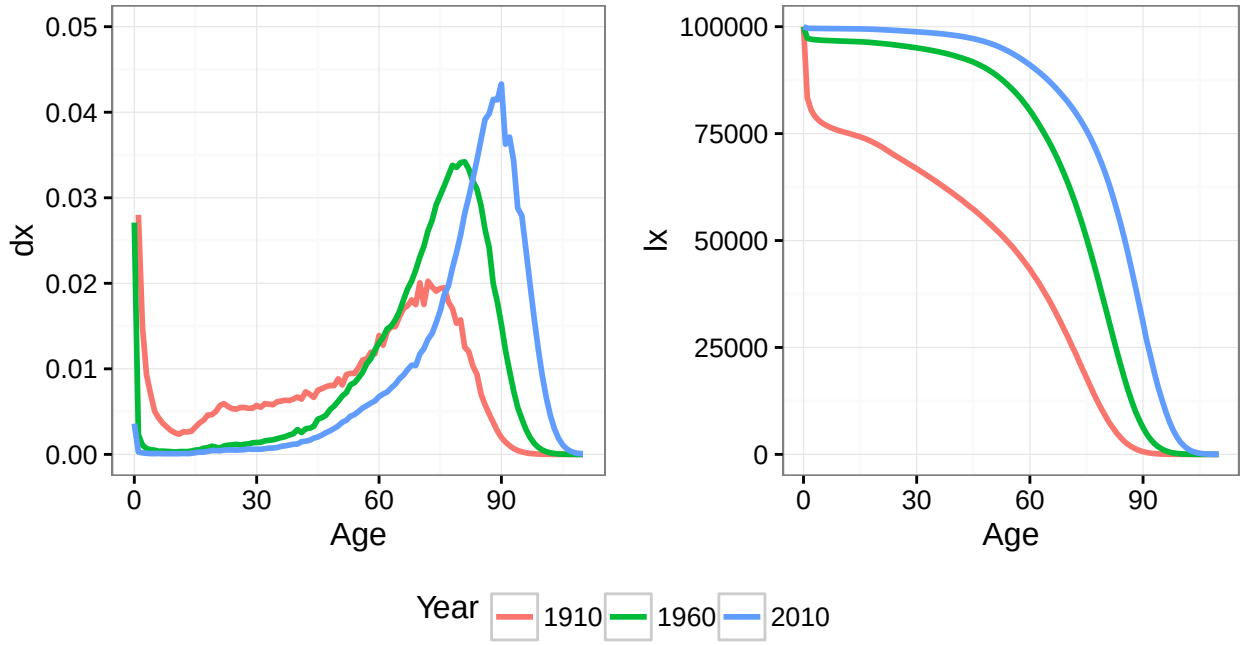


FIGURE 2.1: Death distribution, d_x, and survivorship distribution, l_x, in France. Total population, period data. Data source: HMD.

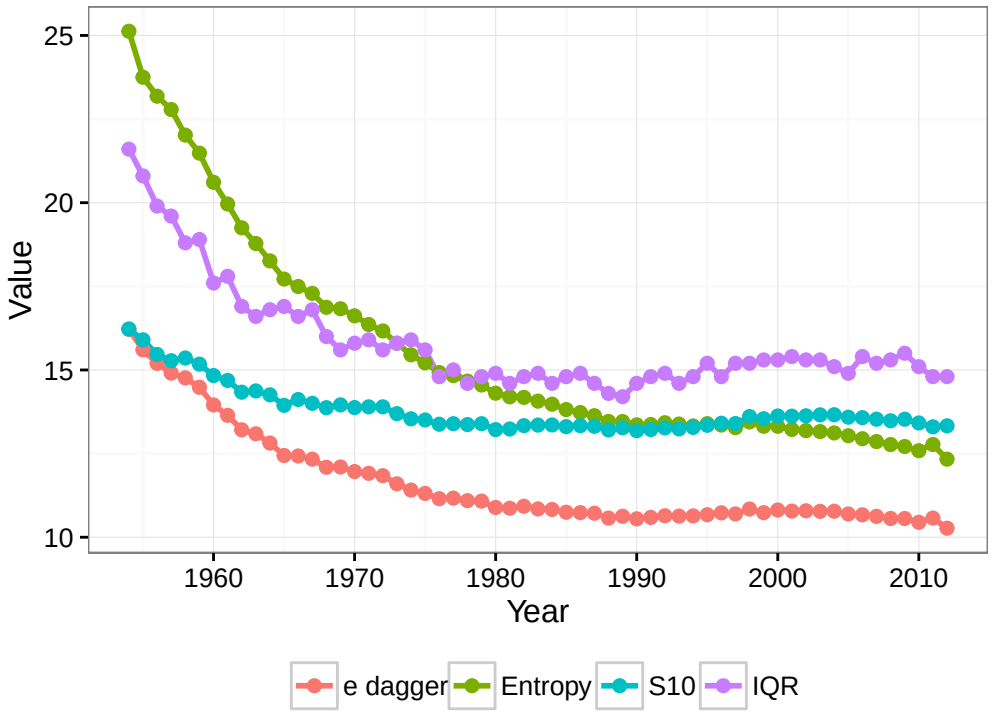


FIGURE 2.2: Trends of various measures of variability in length of life for Japan. Total population. Data source: HMD

Classification	**Measure**
Central longevity indicator	Life expectancy (e_0) Median age at death *(Me)* Mode age at death *(M)*
Horizontalization indicator	Degree of horizontalization (β)
Concentration and/or Verticalization indicator	Standard deviation of ages at death (S_x) Standard deviation above the mode *(SD(M+))* Mean square deviation around the modal age at death *(SDM)* Standard deviation above the third quartile Interquartile range *(IQR)* C-Family (C_{10}, C_{50}, C_{90}) Prolate Index Entropy Keyfitz's *(H)* Life expectancy lost due to premature death ($e^{\dagger}$) Life expectancy at median lifespan and third quartile Fatest decline and/or highest proportion of deaths Degree of verticalization (θ and θ^{*})
Rectangularization indicator	Fixed rectangle Moving rectangle and/or index of rectangularity *R* Person-years ratio *(PR)* Person-years differential *(PD)*
Maximum Longevity indicator	Life endurancy Maximum lifespan *(MLS)* Length of the outer tail of longevity *M + 4SDM+*
Mapping indicators	Percentile
Other indicators	Coefficient of variation *(CV)* Numerator of Keyfitz's *(N)* Gini coefficient

TABLE 2.1: Classification of existing indicators proposed to monitor the rectangularization of the survival curve and/or the compression of mortality. Taken from Cheung et al. (2005). $e^{\dagger}$ and *SDM* are not reported in the original paper.

to H, lifespan variability in Japan is still declining, while the other three indicators show a situation of shifting mortality (even a slight expansion in the case of *IQR*). Also, the degree of compression varies measures by measures: in the first twenty years of the series, for instance, all indicators have a decreasing trend but the magnitude of the decline is much more evident for $e^{\dagger}$, H and *IQR* than S_{10}. What is clear is that differences arise because of differences in the age range covered, differences in the central indicator used in the calculation of variability and whether or not the measurement is attached to fixed percentiles of the death distribution (Zurieck, 2010). Thus, each measure captures a different kind of variability and consequently each of them highlights a different aspect of lifespan inequality. It simply doesn't exist a perfect indicator, it is the researcher that has to choose the indicator that best fits his/her aims. In this dissertation I rely my analysis on two measures: S_{10} and *SD(M+)*. The first measure is used to study lifespan variability over (almost) the entire age range, while the second focus the attention on old-age lifespan variability that, in the era of longevity extension, has become increasingly important. The reasons for adopting these two indicators and the methodological details will be laid out in Chapter 3.

2.1.2 What do we know of lifespan variability?

The study of mortality of mortality compression has its roots in the field of biology (Comfort, 1979; Pearl, 1940; Upton, 1977). After these pioneering works, it was the article published by Fries *Aging, natural death and the compression of morbity* that has stimulated a lot of scholars to focus their attention on that topic (Fries, 1980). Fries' theory states that premature mortality would eventually be eradicated and so the age-at-death distribution would then follow a normal distribution centered at age 85 years and with a standard deviation of 4 years (two third of deaths occurring between 81 to 89). Thus, for Fries, the maximum feasible longevity is fixed and mortality will be more and more compressed until humans reach such fixed biological limit that will eventually result in an ideal age-at-death distribution. Fries forecasted a maximum average age at death of 85 years by extrapolating linear trends in life expectancy at different ages and considering the point of intersection of these trend lines. The concept of normal life duration was not, however, was not introduced by Fries. More than a century early, indeed, Lexis gave his view of the distribution of deaths as consisting of three parts (Lexis, 1878): an high number of deaths within the first year of life (infant mortality), deaths that occur at young ages (premature mortality) and deaths centered around the modal age at death accounting for senescent mortality (normal deaths). A graphical representation of Lexis' view is given in the top panel Figure 2.3.

Since Fries' paper, a lot of researchers have obtained results disproving his theory (Eakin and Witten, 1995; Fries, 1980; Myers and Manton 1984; Rothenberg et al., 1991). The predicted maximum mean age at death of 85 years has been exceeded in many developed countries, especially by females, and the related variability of age-at-death is still much higher than what Fries hypothesized. The fallacy of these two predictions is displayed in the bottom panel of Figure 2.3, in which Fries' deaths distribution is compared to Japanese females distribution (2012) that can be regarded as the

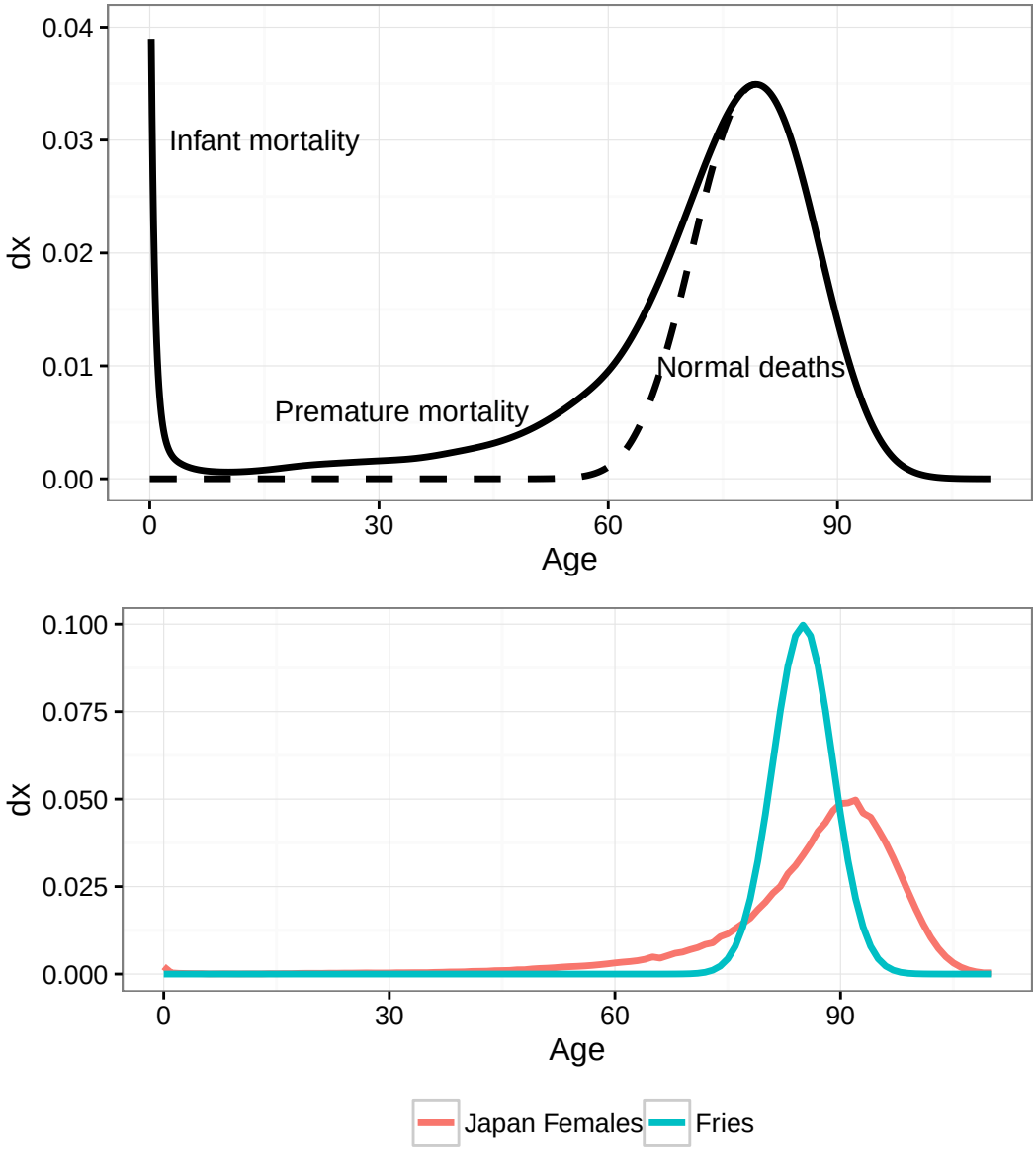


FIGURE 2.3: Top panel: Lexis' distribution of deaths. Bottom panel: Fries' ideal age-at-death distribution compared to Japanese females age-at-death distribution in 2012.

most evoluted one. Also, the assumption that humans are reaching a biologically fixed limit to lifespan has been challenged by many researchers. Kannisto has shown that mortality rates above age 80 have been falling in Western countries and that the pace of decline has raised in last decades (Kannisto et al. 1994). Cheung and Robine demonstrate that the maximum reported age at death has been increasing in Japan and that such increase is not only the consequence of a greater number of living people which make it more likely to observe higher age at death than in the past (Cheung and Robine, 2007). Wilmoth and Lundstrom show that the upper tail of the age-at-death distribution has moved steadily higher over a period of 130 years in Sweden and over shorter periods in other developed countries (Wilmoth and Lundstrom, 1996).

The study of mortality compression is undoubtedly much more complicated than the study of life expectancy. In fact, while a decline in mortality rates at any age acts incrementing life expectancy, the effect of mortality decline on variability measures is not so straightforward and varies by age: avoiding deaths can either increase or decrease lifespan variability. For one of the measures presented previously, $e^{\dagger}$, Zhang and Vaupel have demonstrated that exists an age such that avoiding deaths before that age reduces inequality and avoiding deaths after that age increases inequality (Zhang and Vaupel, 2009). These scenarios are illustrated in Figure 2.4. From the top panels one can see that the effect of reducing mortality rates, m_x, at young ages is to decrease the variability of the density function, d_x , as deaths are pushed out further on the tail of the age-at-death distribution. From the bottom panels, instead, one can see that the effect of reducing mortality rates at old ages is to increase the variability of the density function as deaths of the right-hand tail of the age-at-death distribution are more spread out and concentrated at later ages. Thus, the way in which the age pattern of mortality change is of fundamental importance for the interpretation of lifespan inequality.

The definition of a young-old threshold age, before which mortality decline decreases lifespan variability and after which mortality decline increases lifespan variability, has also been developed by Gillespie et al. with regard to the variance of ages at death above age x, V_x (Gillespie et al., 2014). Investigating on what would be the change due to a proportional mortality reduction at an age a greater than x, the authors have found the rate of change of V_x with respect to the force of mortality μ_x and defined it as the sensitivity of V_x. An example of such sensitivity analysis and the eventual finding of a young-old threshold age is displayed in Figure 2.5, which shows the sensitivity of V_{10} for French females in 1950 and 2010.

In this time interval, the threshold age has increased by 13 years (from 63 to 76) but this shift to the right it is not the only difference between the two curves. While in 1950 V_{10} was really sensitive to mortality reduction at young ages, this is not the case in 2010, since mortality at young ages has reached such a low level that any supplementary decrease cannot impact strongly lifespan variability. Nowadays, it is mainly the decrease in middle-adult mortality that has the capacity to further lower variability in age-at-death.

Another important finding that relates the age pattern of mortality change and the compression of deaths, specifically concerns the standard deviation of ages at death above the mode, *SD(M+)*, i.e. old age lifespan variability.

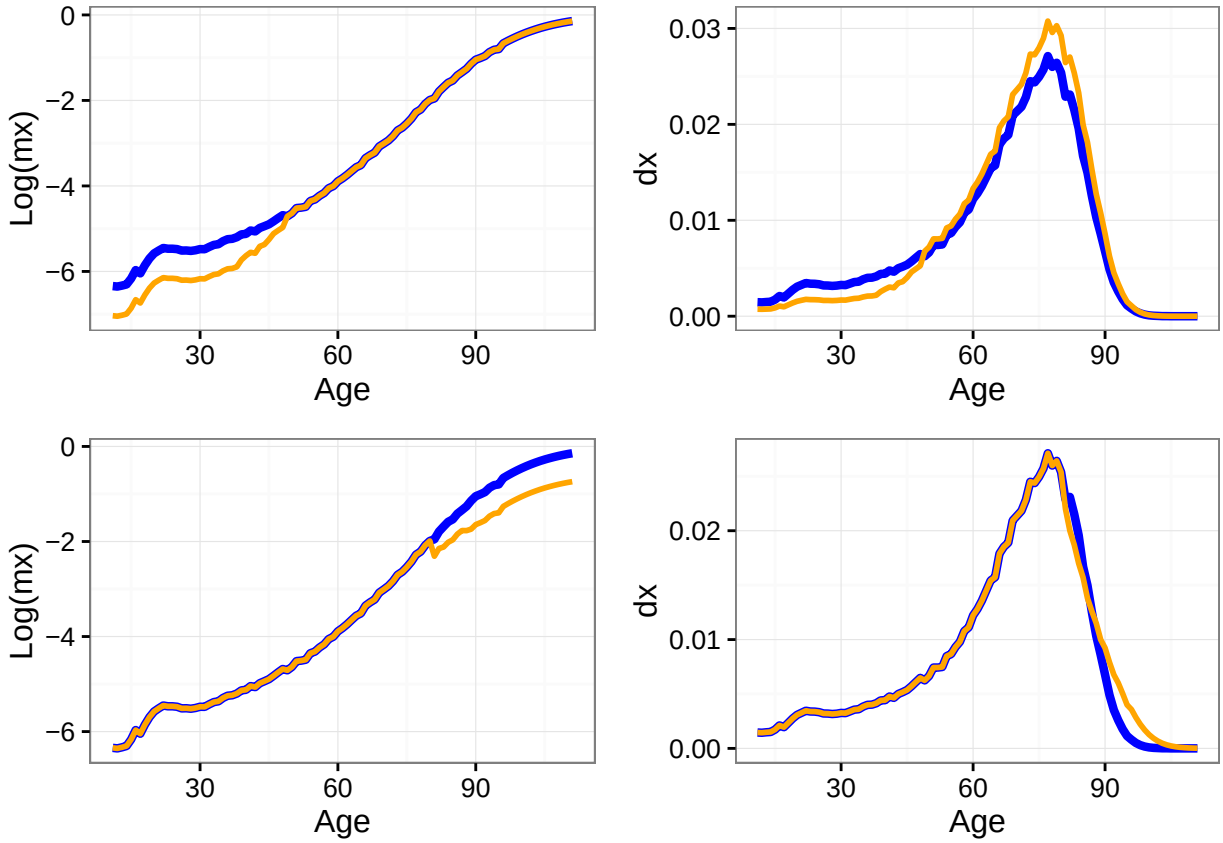


FIGURE 2.4: Hypothetical effects of changes in age-specific mortality on variability of age at death. Topleft panel: reduction in m_x at young ages. Topright panel: decrease in variability of age at death due to reduction in m_x at young ages. Bottomleft panel: reduction in m_x at old ages. Bottomright panel: increase in variability of age at death due to reduction in m_x at old ages. Initial m_x and d_x taken from Denmark 1935, males population. Data source: HMD

Fitting a simple version of the logistic model with only two parameters (as known as Kannisto model), to death rates at age 70 and 90, Thatcher et al. have investigated on the conditions under which compression of deaths above the mode will occur (Thatcher et al., 2010). The model is as follow:

$$\mu(x) = \frac{ae^{bx}}{1 + ae^{bx}} \tag{2.1}$$

Where $\mu(x)$ is the force of mortality at age x, a is the parameter that indicates the level of mortality and b the parameter for the rate of increase in mortality, i.e. the rate of aging.

The authors discovered that the modal age at death, M, depends on both parameters, while $SD(M+)$ depends uniquely on the parameter b:

$$M = \frac{ln(b)}{b} - \frac{ln(a)}{b} \tag{2.2}$$

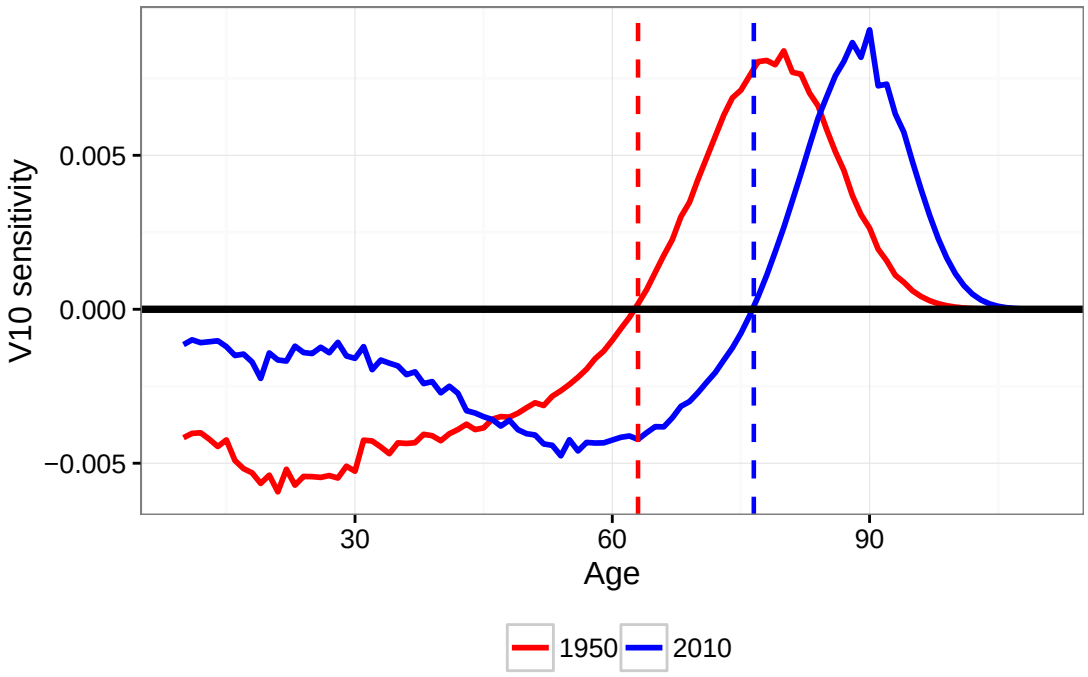


FIGURE 2.5: Sensitivity of V_{10} to proportional mortality decrease. French females in 1950 and 2010. Data source: HMD.

$$SD(M+) = \sqrt{\sum_{x=M}^{\infty} x^2 (\frac{1+b}{1+be^{bx}})^{1/b} \frac{be^{bx}}{1+be^{bx}}} \tag{2.3}$$

Mortality compression occurs if the logit of the death rate at age 70 falls faster than the logit of the death rate at age 90. To achieve shifting mortality, where the modal age increases while variability of age at death above the mode remains constant, the logits of the death rates at age 70 and 90 must fall at the same rate so that the parameter retains the same value over time. These two situations are displayed in Figure 2.6.

Thatcher et al. have fitted their model for six countries (England&Wales, France, Japan, Italy, Switzerland and Sweden) for each sex separately. In all cases, the b at the latest date was higher than the b at the earliest date, so the predicted standard deviation above the mode *SD(M+)* was lower at the later date. Thus, over the period as a whole, there had been compression above the mode in all six countries, for both males and females (Thatcher et al., 2010).

The conclusion of Thatcher et al., differs from one of Bongaarts that used a 3-parameter logistic model, which add to the model of equation 2.1 the Makeham's constant, γ, to measure background mortality (Boongaarts, 2005). Fitting the model to ages 25 to 109 Boongaarts found that the rate of increase in mortality, measured by the parameter b, has remained nearly constant from 1950 to 2000 and as a consequence the age-at-death distribution will shift to the right retaining its shape without showing any compression. The reason for the difference between these two findings lies in the difference between the ages range for which the two model fit data well.

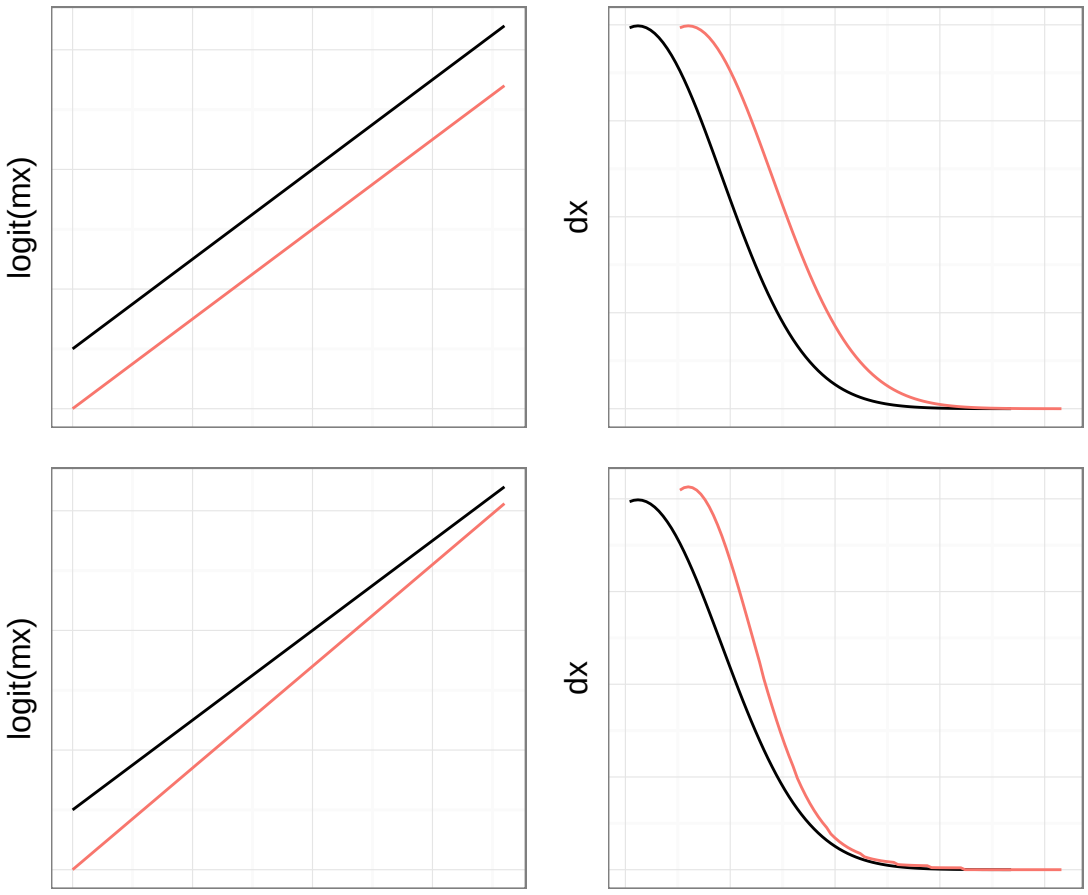


FIGURE 2.6: Graphical representation of shifting mortality (top panels) and compression of mortality (bottom panels) above the mode, M, using the simplified version of the logistic model.

Fitting the model for a wider age range, as Bongaarts did since he used the 3-parameter version of the model, which is thought to include background mortality, leads to an estimate of b constant over time, whereas, if the same model is fitted only at old ages, as in the case of Thatcher at al. which used a model appropriate for old ages, the slope parameter tend to increase over time. After a few years, Bongaarts himself arrived at these conclusions showing that if his model is fitted over shorter age ranges, the estimates of b are not all the same (Boongaarts, 2009).

In the context of old age mortality compression, *SD(M+)* has been adopted as a key indicator also in many other studies, especially by Cheung and her colleagues (Cheung et al. 2005, 2008, 2009; Cheung and Robine 2007; Robine et al. 2006; Canudas-Romo, 2010; Ouellette and Bourbeau, 2011). These studies show that trends in *SD(M+)* have been very similar in developed countries with basically no important changes in this measure until 1950. After that point a decline, along with some fluctuations, has been observed, meaning that compression of old-age mortality has been going on in the last decades. The recent decline in *SD(M+)* is particularly intriguing because it *SD(M+)* can be interpreted as the fact that the increase in human longevity is meeting some resistance.

The adoption of *SD(M+)* as a lifespan inequality measure is largely related to the increasing importance given to the modal age at death as a longevity indicator (Canudas-Romo, 2008; Cheung and Robine, 2007; Kannisto, 2001). Indeed, even if life expectancy at birth is still the most used index to measure the lifespan of a population, since nowadays in developed countries the majority of deaths occur at old ages, focusing on life expectancy can disregard some important information. On the contrary, the modal age at death is determined by adult mortality only and, in many cases, can be a better indicator in studies of senescence and longevity. Moreover, the number of life table deaths at the mode, *d(M)*, has also an important meaning in the contest of lifespan variability. An increase in *d(M)*, indeed, is related to a compression of the age-at-death distribution. To achieve shifting mortality, the distribution has not only to shift to the right (increase in *M*) but also to retain the same shape (stability of *d(M)*). For the Gompertz, Logistic (3-parameter version) and Weibull model, Robine et al. have shown as the number of deaths at the modal age at death is related to the rate of aging parameter *b*, by the following relationships (Robine et al., 2006):

$$d(M) = be^{-1+(a/b)} \approx b/e \quad Gompertz \tag{2.4}$$

$$d(M) = \frac{b}{[1+(b/g)]^{1+g/b}} \approx b/e \quad Logistic \tag{2.5}$$

$$d(M) = \frac{b}{M} e^{-b/(b+1)} \approx b/eM \quad Weibull \tag{2.6}$$

Thus *b* not only determines *SDM+* in the Logistic model, but it also has a relationship with *d(M)* in the most popular mortality models: as the rate of aging increases, the number of deaths occurring at the modal age at death increases too.

2.1.3 Lifespan variability among groups

Measures of lifespan inequality can largely vary, within the same country, among different groups. Gender, socioeconomic status (SES) and race may have an influence not only on life expectancy but also on measures of lifespan variability.

Nowadays, in most of developed and not developed countries, females are experiencing lower mortality rates than males at all ages and, consequently, a higher life expectancy (Edwars and Tujapurkar, 2005). Until early XXth century, however, males enjoyed better mortality conditions than females in childhood and early adulthood principally because of an advantage in mortality related to infectious diseases. Currently, females' higher average lifespan in developed countries also reflects into a lower variability of age at death. Edwards and Tuljapurkar find that S_{10} among females is lower than for both sexes combined of about 1 year (Edwards and Tuljapurkar, 2005). The extreme case is represented by France, where female S_{10} is 1.5 years lower than for both sexes combined, and by Denmark where this difference it is only of 0.5 year. These results are supported by those of Pampel, who reports evidence of lighter smoking among French females, heavier smoking among Danish females, and sex differentials in mortality

that appear to reflect those behavioral differences (Pampel, 2002). Females' lower lifespan variability is also found by using other inequality indicators: the Theil's index (Van Raalte, 2011), the C-family (Kannisto, 2000) and SDM+ (Thatcher et al., 2010).

Also with regard to SES there is a long tradition of studies exploring how socioeconomic differences have an impact on life expectancy. Recently, researchers have shown that lower SES is not only associated with shorter lifespan but also with a greater uncertainty about the timing of death (Brown et al., 2012; Edwards and Tuljapurkar, 2005; Shkolnikov et al. 2003; Van Raalte et al. 2011).

Using longitudinal data on individuals in the U.S. National Longitudinal Mortality Study (NLMS) Edwards and Tuljapurkar have explored the relationship between SES (measured by income and education) and the variability of the age-at-death distribution (measured by S_{10}). Using income as a proxy of SES, individuals were sorted into two groups, those in the first quintile of household income and the rest, while when education is used as a proxy of SES the two groups are represented by those having a high school diploma and those having a lower education level. The results show that using income or education doesn't alter the outcome which is similar in both cases: adults in lower socioeconomic strata not only suffer shorter life expectancy, they also endure greater variability. The gap in average lifespan between individuals in the first income quintile and those in the top four-fifths is 5.5 years, while the difference in standard deviations is nearly 2.5 years. Similarly, high school graduates live an average of 5 years longer than their less educated counterparts, while enjoying a standard deviation that is 2 years lower. These results for the United States are confirmed by the research of Brown et al. (2012). Based on the Health and Retirement Study and the National Health Interview Survey Linked Mortality File they found a strong educational gradient in both longevity and mortality compression and that mortality is more compressed within educational groups among women than men. An important difference with respect to the research of Edwards and Tuljapurkar is that the results refer to old-age mortality compression. Indeed, the indicators used to assess lifespan variability, in this case, are *SMD+* and *d(M)*. Similar outcomes have been found in 10 European countries by Van Raalte et al. (2011). They used harmonized census-based mortality data to construct life tables by sex and educational level, identified by three groups: low (less than secondary education), medium (complete secondary education) and high (tertiary education). The lifespan variability indicator, used to assess whether mortality compression has a relationship with the education level, is S_{35}, the standard deviation of ages at death above age 35. Moreover, the authors also decomposed differences between educational groups in lifespan inequality by age and cause of death. The results indicate, clearly, that the age-at-death distribution is more spread out among the lower educated in every country, and especially among men and in Eastern Europe. About causes of death, the greater lifespan variability in lower educated groups was mainly driven by conditions causing death at younger ages, such as external causes of death and neoplasms.

Finally, a couple of studies have also documented, within the same country, different levels of mortality compression among different racial groups. All of them pertain the United States, where racial differential in

life expectancy are well documented and are viewed as an important indicator of socioeconomic inequality (Edwards and Tuljapurkar 2005; Go et al. 1995; Lynch et al. 2003).

Go et al. (1995) studied race specific California vital statistics data for 1970, 1980 and 1990 finding that, in each period, mortality was more compressed among whites than it was among other racial groups. Lynch et al. (2003), using U.S. vital statistics data between 1970 and 1992, analyzed differences in lifespan inequality between whites and blacks finding a higher level of age-at-death variability among the latter group. A result confirmed by Edwards and Tuljapurkar (2005), in a study of U.S. vital statistics data from the late 1960s to the early 1990s.

2.1.4 Lifespan variability and causes of death

Despite mortality compression has been widely documented, its study has been mainly focused on all-cause mortality. With regard to the role played by specific cause of death into the decline of variability of the age-at-death distribution the knowledge is extremely poor. A few attempts in this sense have been recently made using French data (Zurieck, 2010; Horiuchi et al., 2012) and US data (Nau and Firebaugh, 2014), plus the previously mentioned study of Van Raalte et al. (2011) in which information on the relationship between SES, causes of death and mortality compression are combined.

Zurieck has documented trends in variability of age at death, specifically in S_{10}, and causes of death from 1925 to 1999. Decomposing differences in sex-specific S_{10} in the periods 1925-29, 1960-64 and 1995-99 she analyses the contributions of differences in age and cause-specific mortality between the sexes. During the first period, 1925-29, female S_{10} was higher than male S_{10} by 0.79 years: female disadvantage in infectious disease and maternal mortality at younger ages and female advantage in mortality for all causes at older ages led to a more disperse death distribution for females relative to males. The results for the two following period, 1960-64 and 1995-99, reveal similar patterns of age and cause-specific contributions which are in contrast to the 1925-29 period. Female's S_{10} is now lower than male's S_{10} and the difference seems to have been primarily driven by female advantage in external related mortality at younger ages. Moreover, in the 1995-99 period, the effect of female mortality advantage in neoplasm and heart disease in the middle adult ages on the sex gap in S_{10} becomes more pronounced.

Horiuchi et al., have instead analyzed effects of cause-specific death rates on the rise of all-cause b parameter of the logistic model used by Thatcher et al. and previously mentioned. They consider two possible reasons for the increase of all-cause b. First, if it is difficult to slow or delay senescence, death rates from high-b causes of death may decline more slowly than death rates from low-b causes of death. This will increase the proportion of deaths in old age from high-b causes of death, raising the value of all-cause b (level effect hypothesis). Second, if mortality rises from various CODs, including many of both high-b and low-b causes of death, become steeper, all-cause b will increase (slope effect hypothesis). Using French data for the period 1979-1994, they found strong evidence supporting the second hypothesis.

Nau and Firebaugh analysis', starting from the consideration that, in the United States, lifespans are more variable for blacks than for whites, have tried to explain this difference using a cause-of-death approach. They have decomposed the black-white difference in V_0 into allocation (blacks are more likely than whites to die of causes that tend to strike the young as well as the old), spread (blacks and whites might die of the same causes at the same rates, but yet the variability in age at death could be greater for any or all causes among blacks), timing (the age distributions of specific causes are centered on different mean ages for blacks and whites) and joint effect (captures the part of the racial difference in lifespan variance that is attributable to simultaneous differences in incidence and in cause-specific variances).

In their study, the all-cause spread component accounts for about 87% of the disparity, indicating that lifespans are more variable for blacks largely because age at death varies more for blacks than for whites among those who succumb to the same cause. The all-cause allocation component is about 12%, indicating that only about 12% of the disparity in lifespan variance would persist if blacks and whites differed only with regard to cause-specific death rates. The all-cause timing component is even smaller and is negative (-4.7%), indicating that lifespans would vary less for blacks than for whites if blacks and whites differed only with respect to variance in the average age at death across causes. The all-cause joint component is also small, about 5%. The cause-specific spread components show that age at death in the United States varies more for blacks than for whites for almost all causes of death. As one would expect, the most common causes of death contribute the most to the spread component. Heart disease, for example, is the biggest contributor, with a spread component of 29%. The profile is very different for the cause-specific allocation components. Relatively rare causes of death AIDS, homicide, suicide, and accidental poisoning contribute much more to the allocation component than do much more common causes of death, such as heart disease, cancer and cerebrovascular diseases. Moreover, unlike the cause-specific spread components (all of which are either positive or negligibly negative), the cause-specific allocation components largely offset one another. Homicide's large allocation effect (38%), for example, is almost entirely offset by the combined allocation effects of suicide (-22%) and accidental poisoning (-15%). The timing and joint components contribute very little to the disparity in lifespan variance. Their contributions are minimal because virtually all of their cause-specific components are negligible. In the case of timing, accidental poisoning has the largest effect by far (-14.9%).

2.1.5 Implications of changes in lifespan variability

The changes in mortality observed over the course of the last two centuries must have had an impact on other areas of human life besides death. A first question that has to be answered is whether or not humans are aware of their own mortality risk. If human behavior is sensitive to changes in lifespan variability, indeed, persons have to be some consciousness of it.

The information coming from studies on mortality risk perception have given mixed outcomes. Empirical evidence have shown that objective risk

measurement, experts' risk estimate and people's perceptions differ (Sunstein, 2002). In judging their own mortality risk, people are heavily influenced by their own experience (e.g. death of a parent or spouse or a friend) and by how they perceive the risk. Also the gender seems to be an important factor: females perceive health and environmental risks as greater than males do (Brown and Cotton, 2003; Dosman et al., 2001; Liu and Hsieh, 1995; Lundborg and Andersson, 2006; Lundborg and Lindgren, 2004; Savage, 1993). Hurd and McGarry, using data from the Health and Retirement Survey, have discovered that the self-assessed survival probabilities of those who survived between two waves this longitudinal survey were considerably higher in comparison to those who died during the interval between the two surveys, showing a good predictive power of respondents' self-assessments of their own survival probabilities (Hurd and McGarry, 2002). Similar results have been found by Mirowsky, who using data from the 1995 Aging, Status and the Sense of Control survey and comparing them to actuarial tables for the United States, found that on average subjective and actual life expectancies agree, even though younger persons fail to account for expected improvements in survival over time (Mirowsky, 1999). High income and high education level seem to be two characteristics which reduce individual perception of risk (Dosman et al., 2001; Savage, 1993). The reason could lie in the fact that wealthier and better-educated people may have the opportunity, through money and/or consciousness, to actually expose themselves to less risk. Sunstein has also found an optimistic bias in how people perceive voluntary risks which are judged to be controllable by personal action (Sunstein, 2002). An attempt to disentangle the preference for greater certainty in timing of death from the desire for greater longevity has been done by Edwards (Edwards, 2009). He proposed a model which, based on the ratio of the expected utilities from greater certainty and greater longevity, theoretically permits to determine the amount of life expectancy one would be willing to give up in order to be more certain about the timing of death. Edwards' finding shows that an average person would renounce to 0.5 year of average life span for a 1 year reduction in variability of age at death.

Another aspect of the possible consequences of the decline in variability of age at death surely concerns psychological implications. It may be that human being are less scared of death when there is greater certainty on the time of its occurrence: death has seen as something natural and expected among old people and therefore less traumatic. But, on the contrary, it may also be that such certainty could increase fear of death because deaths that occur outside of the natural order are perceived as more awful. This second reasoning is supported by a study of attitudes towards death among impoverished persons in Brazil suffering high rates of mortality in infancy and childhood, which shows that mothers are relatively indifferent to the deaths of their children (Scheper-Hughes, 1992). Another study, reporting evidence from the past, suggests that maternal attitudes regarding infant death changed considerably from the mid XVIIIth to early XXth century in the United States (Dye and Smith, 1986). While earlier, Americans mother held a deterministic attitude toward infant deaths, in the more recent era this gave way to a recognition of the role of human agency, which in turn put pressure on mothers to ensure their children's survival.

Thus, the psychological implications of the decline in variability of age

at death are still vague and not so much studied and, to a certain extent, the same can be said for the economic implications. For instance, Edwards and Tuljapurkar state that the effect of this greater certainty on age at retirement is still unclear (Edwards and Tuljapurkar, 2005). Persons might retire earlier if life span is more uncertain if they view retirement as a benefit for years of work. However, they might also be inspired to work longer in order to accumulate savings because they are more uncertain about the transition to ill-health. Likewise, it is difficult to determine how uncertainty about timing of death will affect saving behaviors because it is not simple to understand whether bequests are intentional or circumstantial. About the public pension system, on the contrary, the situation seems to be clearer: higher is the uncertainty in timing of death at the population level higher is also the cost for a public pension system. The costs of supporting those who live longer, indeed, are not balanced by earlier deaths because income is collected by the dependents of those who die early. Also, those who die earlier have contributed less to the system and those who live longer collect greater benefits as the poor have a survival disadvantage (Edwards and Tuljapurkar, 2005). Apart from the pension system, an important question regarding the exceptional growth of the elderly population brought by mortality compression (and the rising in life expectancy) is how much it will cost to support this aging population. A key factor to answer this query is whether or not mortality compression will be accompanied by morbidity compression.

The tremendous growth of the elderly population implied by the decline of variability in age at death, also poses a demographic question about the population growth. If it is true that at the individual level fertility is going down, this does not mean that the number of births decreases at population level since more individuals are surviving to their childbearing ages as deaths are compressed into older ages. Demographically speaking, another implication of the decreasing lifespan inequality is that the risk of being orphaned for children is much lower than in the past as well as married persons are much less likely to be widowed during their reproductive years in comparison to previous generations. As a result persons can rely more heavily on the nuclear family. But, if on one hand mortality compression seems to ensure a solid family structure, on the other greater certainty in timing of death and longer life expectancy have also probably brought an increase in the number of divorces, a phenomenon that weakens the nuclear family unity. Persons unhappy in their marriage may be more likely to seek divorce if they expect their partner to live a long time and if their own survival prospects are such that they have many years to seek out a new relationship.

2.2 An overview of the morbidity and mortality transition in Italy

The *demographic transition* theory was firstly developed by Thompson who observed a transition from high birth and death rates to low birth and death rates in developed societies which passed from a pre-industrial to an industrialized economic system (Thompson, 1929). The theory was then developed more formally by Notestein who noted that the mortality decline

preceded the fertility decline producing, as a consequence, a fast population growth (Notestein, 1945). Such transition started, in each of the European countries, in different periods. England is regarded as the first in which the transition began, at the end of the XVIIIth century, followed by France, Norway and Sweden after a few decades while In Italy the transition started only in the late XIXth century. Improvements in food supply, water supply, hygiene and sanitation are regarded as the major force driving the transition in its first stage, allowing a dramatic reduction of mortality, especially infant and childhood mortality. The consequent fertility reduction also started for a variety of factors such as an increase in the status and education of women, the diffusion of contraception, a reduction in the value of children work and the urbanization.

A theoretical explanation of the changes in mortality observed over the course of time, using an epidemiological approach, was firstly given by Omran (Omran, 1971). Omran's *epidemiological transition* theory consists of three different stages:

- *The age of pestilence and famine*. In this phase mortality, is high and fluctuating with life expectancy around 30 years. Infectious diseases represent the first cause of death and the fluctuations are a consequence of the cyclical famine.
- *The age of receding pandemics*. This is the stage of the transition, life expectancy rises thanks to the infectious diseases reduction.
- *The age of degenerative and man-made diseases*. After a long period of decreasing, mortality levels off. Much of deaths are due to degenerative diseases (cardiovascular diseases, neoplasms, diabetes mellitus) and man-made diseases (suicide, alcoholism, traffic accidents, pollution).

Omran's theory is based on the idea of a transition from a stable regime of high mortality to another stable regime of low mortality. This theory, highly descriptive, is strictly connected to the historical period in which it was conceived and for this reason, at that time, it was able to explain the terrific mortality reduction and its posterior stabilization observed during the 60s of the XXth century. In the following decades, however, the hypothesis of a new stable mortality regime was contradicted: death rates started to fall again thanks to a considerable reduction of cardiovascular diseases. A new theory, thus, called *health transition* theory, was proposed in 1991 by Frenk (Frenk et al., 1991). It adds other two stages to the three proposed by Omran:

- Consistent decrease of the cardiovascular diseases due to the recent medical knowledge improvements.
- A possible further drop of death rates caused by the reduction of neoplasms and senescence-related mortality

Focusing the attention on what happened in Italy, Figure 2.7 displays life expectancy trends for females, males and total population from 1872 to 2013. As mentioned before, in Italy the transition to a regime of low mortality started later than in England or other northern European countries and this can be clearly noted by the fact that in 1872 the average lifespan was

equal to 30.2 years for females and 29.2 years for males, values in line with the pre-transition stage described by Omran.

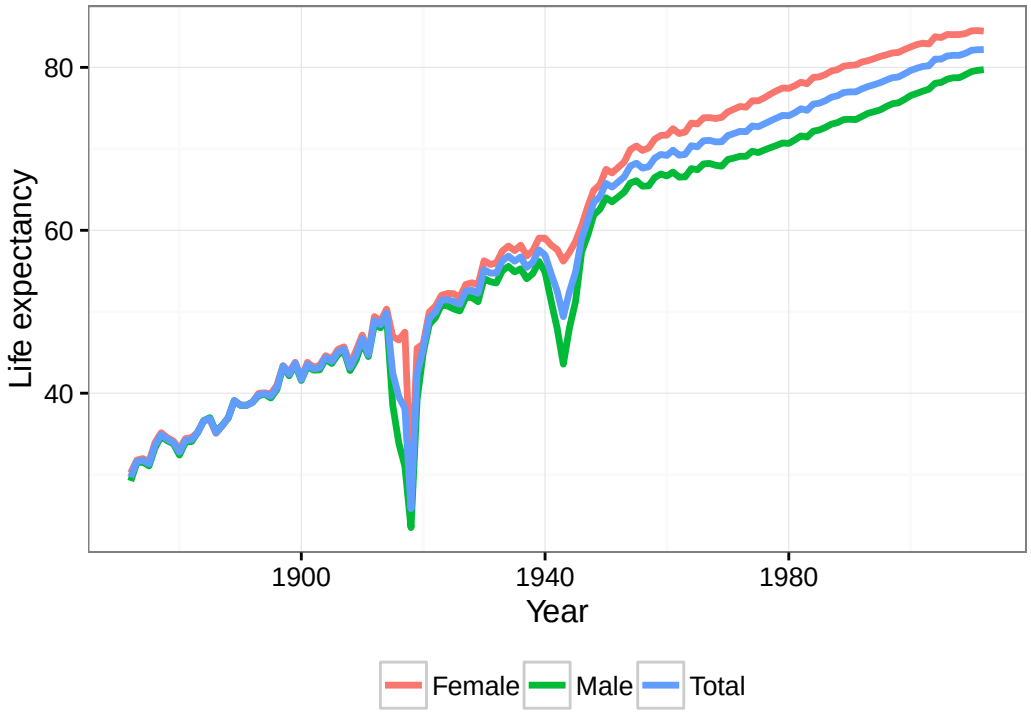


FIGURE 2.7: Life expectancy trends for females, males an total population in Italy from 1872 to 2013. Source: HMD

From that moment life expectancy started a steady and almost linear rise, interrupted only during World War I and World War II. In particular, during the 1914-1919 period, the effects of World War I were further aggravated by the outbreak of the Spanish Flu epidemic (Caselli and Egidi, 1991). Nowadays, females are enjoying a mean lifespan of 84.4 years while males of 79.7 years. The gender gap became clear only after World War II and reached its peak in 1979 when the difference between the sexes was of 6.75 years.

Such exceptional improvements in life expectancy can be better assessed looking at Figure 2.8, which displays the surface of age-specific death probabilities of Italian males and females over the period from 1872 to 2013. The figure extends the one of Barbi et al. in which the death probabilities are reported over the period from 1887 to 1994 (Barbi et al., 2000). The two World Wars impact on mortality can be easily identified on the maps by the elongated rays of high mortality crossing all ages. For men, these effects are visible also some years after the wars among the cohorts that were born or grew up during the war years (Barbi et al., 2000).

Omran's outlook about a new stabilization of mortality after decades of improvements is detected for males, who experienced no substantial decline in the probability of dying, especially at ages above 60, until the 70s of the XXth century. It was this leveling off, also observed in other developed countries, that led to the consideration, then proved as wrong, that all the attainable decline of mortality had been reached and further improvements

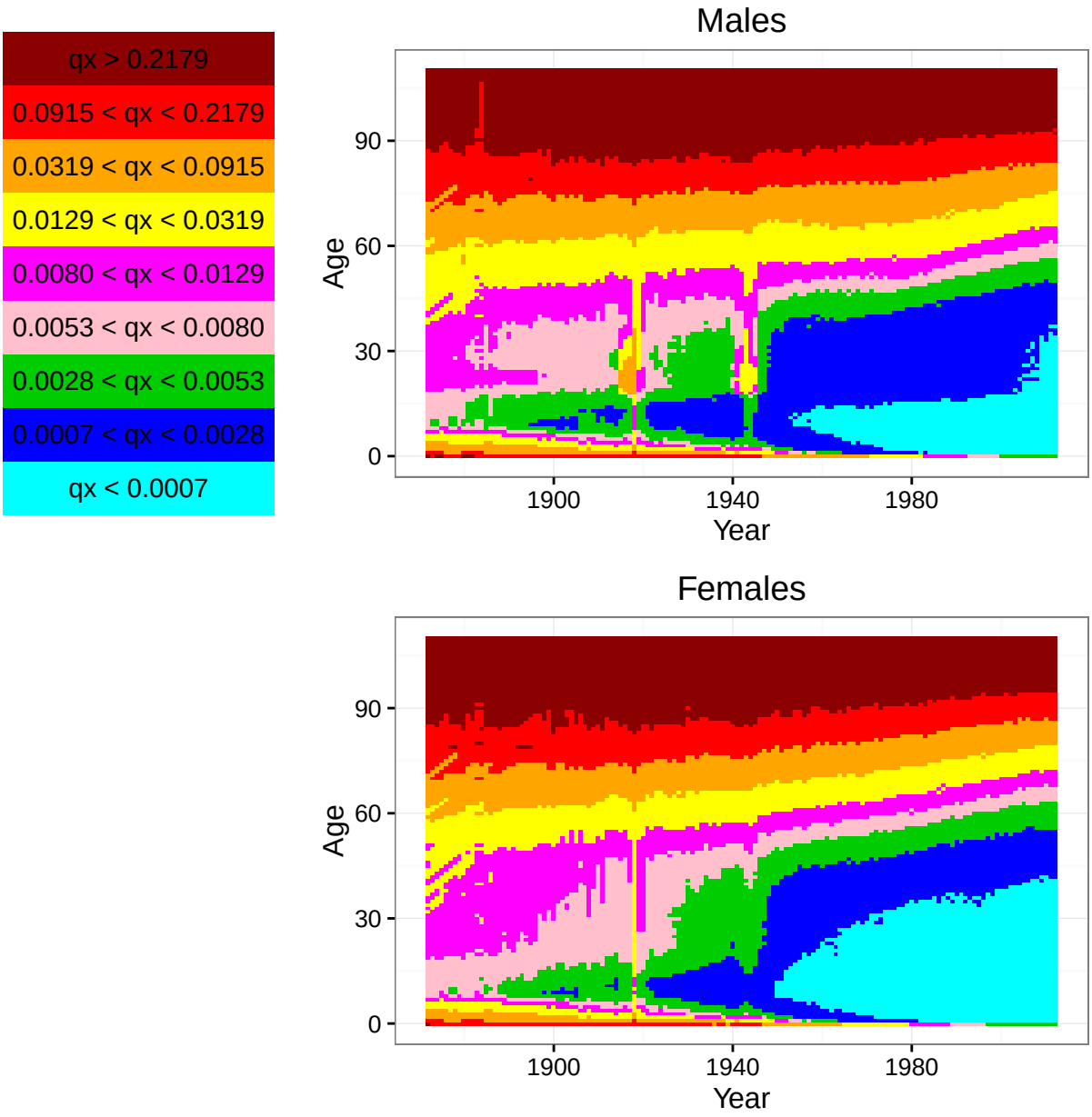


FIGURE 2.8: Probability of dying for Italian males and females from 1872 to 2013. Adapted from Barbi et al., 2000.

were impossible. This stage of stabilization was not reported by females for whom the probability of dying steadily went down from the end of the XIXth century. Beginning from the mid-1970s, a new stage of mortality decline took off, even for the elderly, initially at a slower pace for men than for women, while in the last decade the situation seems to be reversed.

A deeper insight into the mortality transition is given breaking down the gain in life expectancy according to underlying cause of death and age, as done by Caselli and Egidi who have interpreted this rising trend in the light of the health transition started in Italy at the end of the XIXth century (Caselli and Egidi, 1991; Caselli, 1991). In their study they assess that out of the 33 and 39 years gained by males and females respectively from 1885 to 1986, 25 years are due to the fall in mortality from infectious diseases, and more specifically to tuberculosis, pneumonia and influenza. Those causes strongly reduced their lethality in particular in early infancy as a consequence of improvements in living conditions, nutrition, hygiene and advancement made in medical treatment and prevention. On the other hand, in the period considered by the study, the contribution of diseases of the circulatory system to the gain in life expectancy has been less than 1 year for men and less than 2 years for women, thank principally to the decline in mortality among the elderly. The situation is even reversed for malignant neoplasms which had a negative contribution to life expectancy of about 1.5 years for males and 0.5 years for females, although in the last two decades analyzed some improvements were registered for under 45's men.

Barbi et al. analyze the passage to a more modern mortality regime from another point of view, looking at the first leading group of causes of death from 1895 to 1993 for both sexes, from birth to age 98 (Barbi et al., 2000). In this analysis all the stages of the transition come up precisely: infectious diseases were predominant, especially at young ages, until the 1940s, to then leave space to a more modern regime in which external causes for the young, neoplasms for adults and circulatory diseases for the elderly are the most relevant causes of death. Especially for females, the predominant role of circulatory diseases started to decrease, from the 1960s, in favor of cancer in particular between age 40 and 60. In Chapter 4 I will extend the study of the leading group of causes of death of Barbi et al., first by adding other 20 years to the series and second by using a deeper level of analysis in terms of causes for deaths since 2003.

2.3 Reconstructing coherent series of causes of death

The main purpose of this dissertation is studying mortality compression in Italy using a cause-of-death approach. In order to accomplish this task I used data on causes of death in Italy from 1980 to 2013. As always when working with causes of death the main problem is represented by the fact that such causes are registered according to different classifications. Since 1900, indeed, the International Classification of Diseases (ICD) has undergone ten revisions. These revisions are necessary because of changes in medical knowledge, changes in disease profiles and changes in how diseases are perceived and understood (Meslé, 2008). In my case two revisions are involved, ICD-9 and ICD-10. ICD-10 was introduced in Italy in 2003 and consequently discontinuities in the statical series of mortality by cause

of death were produced. Details on the reconstruction of the Italian series will be given in Chapter 3, here I summarize the history of International Classification of Diseases and I review the statistical method developed to deal with the issue of harmonizing different classifications.

2.3.1 International Classification of Diseases

The botanist and medical doctor Francois Bossier de Laroix(1706-1777) is credit as the first person who tried to develop a classification of diseases (Bossier de Laroix, 1763). Until the mid-XIXth century, however, the most used classification of diseases, ideated by William Cullen (1710-1790), was the "*Synopsis nosologiae methodicae*" (Cullen, 1785). Cullen's classification was taken as a model by William Farr (1807-1883) who tried to improve it and decisively push toward the adoption of a unique classification: "*The advantages of a uniform statistical nomenclature, however imperfect, are so obvious, that is surprising no attention has been paid to its enforcement in Bills of Mortality. Each disease has, in many instances, been denoted by three or four terms, and each term has been applied to as many different diseases: vague, inconvenient names have been employed, or complications have been registered instead of primary diseases. The nomenclature is of as much importance in this department of inquiry as weights and measures in the physical sciences, and should be settled without delay.*" (Farr, 1839).

The utility of a uniform classification of causes of death was recognized only in 1853 during the First International meeting of Statistics that took place in Bruxelles. In the second meeting, two years later, William Farr and Marc D'Espine (1806-1860) presented two different classifications that were based on different principles. Farr's classification was divided into five groups: epidemic diseases, general diseases, local diseases differentiated by the anatomic location, development disorders and diseases caused by violence. D'Espine, instead classified the diseases according to their nature (gouty, erpetic, hematic etc). The board of the meeting decided to adopt a classification resulting from the union of the two. Despite this classification was never used at the international level it constituted the basis on which ICD-1 would be shaped a few decades later. Another important figure in the development of a unique classification was Jacques Bertillon (1851-1922) who in 1891 presented in Wien a new classification of diseases that was adopted two years later, in 1893, by the International Statistical Institute of Chicago. Bertillon's classification was based on the principle, proposed by Farr, of the distinction between general diseases and local diseases differentiated by the anatomic location. This list consisted of three main categories, the first formed by 44 chapters, the second by 99 and the third by 66. In a few years Bertillon's classification started to be used in many countries and in 1898 the American Public Health Association recommended its adoption. Finally, in 1900, the first conference for the revision of the international classification of diseases was organized in Paris. Starting from Bertillon's classification, experts from 26 countries compiled ICD-1.

A second revision took place 9 years later, in 1909, also this time in Paris and lead to ICD-2. The three following revision didn't bring any really important change. It was in 1948, with ICD-6, that some important provisions were introduced. First of all, it was approved the international medical certificate of causes of death; secondly it was accepted that the underlying

cause of death was the origin of the process leading to death; thirdly there were defined the rules for the selection of the underlying cause of death. Moreover the World Health Organization (WHO) delineate a series of regulamentation on compiling statistics of mortality and morbidity that, all the countries adopting ICD-6, were recommended to follow. This collaboration between the statistical offices of each country and WHO surely represents an important turning point in the development of the international homogeneity in mortality and morbidity statistics.

The seventh revision took place one more time in Paris, in 1955, while the eighth in Geneva 10 years later. Either ICD-7 or ICD-8 didn't change considerably the structure of ICD-6. During the ninth revision, instead, some relevant modifications were introduced. The level of detail increased remarkably from 2862 listed diseases to 5614 and a new codification system was introduced. Moreover, some technical adjustments were made to make ICD-9 more flexible and easier to use, facilitating the production of statistics and indicators.

The tenth revision was finished in 1992, even if, due to its complexity, its introduction worldwide was really slow. This complexity comes, principally, from an impressing level of detail: the 5600 items of ICD-9 were replaced by more than 10000 items and the codification system switches from numeric to alphanumeric (the 22 main Chapters of ICD-10 and their codes are reported in Table 2.2). To each disease, indeed, are associated two letters and four numbers (four-digit level of detail). More than other revisions, ICD-10 made numerous changes to the way deaths are classified by cause. Chronic viral hepatitis, for instance, which was classified as digestive diseases in ICD-9, is now classified as infectious diseases, whereas other or unspecified forms of chronic hepatitis continue to come under digestive diseases (Meslé, 2008). Another important difference with ICD-9 is the changes to the rules governing the selection of the underlying cause of death. In the death certificate, indeed, more than one disease is often reported but only one has to be selected as underlying cause of death. From a reconstruction of coherent time series point of view, the introduction of these new rules is considerably important since it brings, more than ever, statistical disruptions in the time series of causes of death.

The eleventh revision is planned for 2018.

Chapter	Title	Code
I	Certain infectious and parasitic diseases	A00–B99
II	Neoplasms	C00–D48
III	Diseases of the blood	D50–D89
IV	Endocrine, nutritional and metabolic diseases	E00–E90
V	Mental and behavioural disorders	F00–F99
VI	Diseases of the nervous system	G00–G99
VII	Diseases of the eye and adnexa	H00–H59
VIII	Diseases of the ear and mastoid process	H60–H95
IX	Diseases of the circulatory system	I00–I99
X	Diseases of the respiratory system	J00–J99
XI	Diseases of the digestive system	K00–K93
XII	Diseases of the skin and subcutaneous tissue	L00–L99
XIII	Diseases of the musculoskeletal system and connective tissue	M00–M99
XIV	Diseases of the genitourinary system	N00–N99
XV	Pregnancy, childbirth and the puerperium	O00–O99
XVI	Certain conditions originating in the perinatal period	P00–P96
XVII	Congenital malformations, deformations and chromosomal abnormalities	Q00–Q99
XVIII	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	R00–R99
XIX	Injury, poisoning and certain other consequences of external causes	S00–T98
XX	External causes of morbidity and mortality	V01–Y98
XXI	Factors influencing health status and contact with health services	Z00–Z99
XXII	Codes for special purposes	U00–U99

TABLE 2.2: ICD-10 main chapters and codes.

2.3.2 Reconstruction method

The issue of harmonizing different ICD classifications to get coherent time series of deaths by cause has been popularized by two researchers, Jacques Vallin and France Meslé. Their reconstruction of French deaths from 1925 to 1978 was the first in the history of demography and epidemiology (Vallin and Meslé, 1988). Until that moment any long-term statistical analysis of causes of death was impossible as successive change in the ICD resulted in disruptions in the series. Nowadays, thanks to the method developed by Vallin and Meslé, reconstructions for many countries, especially of the former Soviet Union, are available (Shkolnikov et al. 1996; Meslé and Vallin, 2003; Meslé and Vallin, 2008).

In many countries, the major difficulty in reconstructing coherent time series is caused the absence of a cross-classification between successive ICD version which would allow determining the exact transfer from old items to new ones making possible to calculate transition coefficients and to redistribute deaths according to the most recent classification in use. Such cross-classification, also known as bridge-coding, is available for USA, Canada, UK, Sweden and Norway, but it was never used for the reconstruction. This is probably due to the fact that it regards just a too small sample of deaths that don't permit a robust and statistically significant reconstruction. Also in Italy a cross-classification between ICD-9 and ICD-10 is available. In this case the sample is extraordinary big (78% of total deaths). However, any reconstruction was performed. I will use such bridge-coding data to reconstruct the Italian series from 1980 to 2013[2]. However, it is worth reviewing Vallin and Meslé method since the reconstruction procedure starting from the cross-classification is similar, to a certain extent, to their technique.

The method consists of three phases: the construction of a dual correspondence table linking the items of the two revisions, the definition of fundamental association of items gathering the same medical diagnosis at both revisions and ensuring a statistical continuity and finally the elaboration of a transition table indicating how to redistribute the deaths (Meslé and Vallin, 1996).

- The correspondence table

  For each item of the old revision, all the items of the new revision that have in common with it one or more conditions have to be found. Conversely, for each item of the new revision, all the items of the old revision that have in common with it one or more conditions have to be found. At the end of this step two correspondence tables are produced (Table 2.3 and 2.4 respectively). For instance, in Table 2.3 one can see that the condition in item 0.000 of the old classification, *Classical cholera*, is the same as that found in item 001.0 of the new classification, *Cholera due to Vibrio cholerae*. Thus, in this case the most recent revision didn't introduce any change. Instead, to match the condition in item 412.0 of the old classification, *Chronic ischemic heart disease with hypertension*, seven items of the new classification have to be picked up, meaning that serious disruptions are produced. On the other hand, the same reasoning can be done for items of the new classification with items of the old one, as shown in Table 2.4.

[2] All the limitation of the procedure are explained in Chapter 3.

Old classification		New classification	
Item	Name	Item	Name
000.0	Classical cholera	001.0	Cholera due to Vibrio cholerae
⋮	⋮	⋮	⋮
162.1	Malignant neoplasms of bronchus and lung	162.2	Malignant neoplasms of main bronchus
		162.2	Malignant neoplasms of upper lobe, bronchus or lung
		162.3	Malignant neoplasms of middle lobe, bronchus or lung
		162.4	Malignant neoplasms of lower lobe, bronchus or lung
		162.8	Other malignant neoplasms of bronchus and lung
⋮	⋮	⋮	⋮
411.0	Other acute forms of ischemic heart disease with hypertension	411	Other acute form of ischemic heart disease
411.9	Other acute forms of ischemic heart disease without hypertension	411	Other acute form of ischemic heart disease
412.0	Chronic ischemic heart disease with hypertension	411	Other acute form of ischemic heart disease
		412	Old myocardial infarction
		414.0	Coronary atherosclerosis
		414.1	Aneurysm of heart
		414.8	Other forms of ischemic heart disease
		414.9	Chronic ischemic heart disease unspecified
		429.2	Cardiovascular disease unspecified
⋮	⋮	⋮	⋮

TABLE 2.3: Selected items of the correspondence table from the old to the new classification. Adapted from Mesle, 1996.
* In the article the two classifications are ICD-8 and ICD-9.

New classification		Old classification	
Item	Name	Item	Name
001.0	Cholera due to Vibrio cholerae	000.0	Classical cholera
⋮	⋮	⋮	⋮
162.2	Malignant neoplasms of main bronchus	162.1	Malignant neoplasms of bronchus and lung
⋮	⋮	⋮	⋮
411	Other acute form of ischemic heart disease	411.0	Other acute forms of ischemic heart disease with hypertension
		411.9	Other acute forms of ischemic heart disease without hypertension
		412.0	Chronic ischemic heart disease with hypertension
		412.9	Chronic ischemic heart disease without hypertension
⋮	⋮	⋮	⋮

TABLE 2.4: Selected items of the correspondence table from the new to the old classification. Adapted from Mesle, 1996.
* In the article the two classifications are ICD-8 and ICD-9.

- Fundamental associations of items

  To better illustrate the modifications introduced by the new classification the fundamental association of items has to be built. These associations are constructed in a stepwise manner. The second correspondence table (Table 2.4) gives for the first item of the new classification, the item(s) of the old one that matches it. The following step is to turn back to the first correspondence table (Table 2.,3) to check whether this/these item(s) of the old classification incorporate(s) conditions found in other items of the new classification. If this is not the case the fundamental association is complete, otherwise it is needed to go back to the second correspondence table to check whether these other items of the new classification incorporates, in turn, conditions belonging to other items of the old classification and so on, until the fundamental association is complete. Then, the same procedure has to be repeated for all other items not involved in previous associations. In the end, four different kinds of association will result:

  a) Only one item of each revision is involved

  b) The item of the old classification is split up into various items of the new classification

Old classification	New classification				Total deaths
	Cause A	Cause B	Cause C	Cause D	
Cause 1	22	18	30	0	70
Cause 2	0	13	2	41	56
Cause 3	342	37	54	33	466

TABLE 2.5: Cross-classification of an hypothetic fundamental association of item of type d)

Old classification	New classification				Sum
	Cause A	Cause B	Cause C	Cause D	
Cause 1	0.31	0.26	0.43	0	1
Cause 2	0	0.23	0.04	0.73	1
Cause 3	0.73	0.08	0.12	0.07	1

TABLE 2.6: Transition coefficients derived from Table 2.5

c) The item of the new classification is split up into various items of the old classification

d) This is the most complex case: the changes introduced by the new revision consist of splitting up of certain item of the old revision into several elements and the reconstruction of several items of the new revision by assembling these elements differently.

- The transition table

  The fundamental associations of items allow the calculation of transition coefficients to pass from the old to the new classification. This step basically consists of producing a cross-classification between two successive revisions manually and then from such cross-classification finally get the transition coefficients. Depending on the type of the association, different procedure are required.

  In case of association of type a) the solution is very simple, 100% of the old classification item goes to the new classification item. In case of association of type b), the deaths classified at the old revision item are redistributed among the new revision items proportionally to the distribution of the year in which the new classification was introduced. In case c) 100% of each item of the old classification goes to the item of the new classification. The only difficulties are represented by associations of type d). Following the same approach used for association of type b), so redistributing deaths proportionally, a lot of cases have only one solution. If, instead, different solutions are possible, a pragmatic decision has to be taken (Meslé and Vallin, 1996).

  Table 2.5 displays the final cross-classification of a hypothetic fundamental association of item of type d). Three causes of the old classification are linked to four causes of the new classification. At the end of the redistribution procedure, 22 of the 70 deaths belonging to "Cause 1" according to the old classification, now are transferred to "Cause

A" of the new classification, 18 to "Cause B", 30 to "Cause C" and 0 to "Cause D". The same reasoning can be done for "Cause 2" and "Cause 3". To get the transition coefficients contained in Table 2.6, every cell is then divided by the corresponding number of total deaths. Now, assuming that the coefficients obtained for one year are acceptable for the entire period covered by the ICD version under consideration, it is possible to reconstruct the time series according to the most recent classification in use. Transition coefficient are applied by sex and age groups, so for instance, the number of deaths of "Cause A" from age *x* to age *x+n* in year *i*, ${}_nd{A}_x^i$, is computed as follow:

$$ {}_nd{A}_x^i = {}_nd{1}_x^i * 0.31 + {}_nd{2}_x^i * 0.00 + {}_nd{3}_x^i * 0.73 \tag{2.7} $$

Where ${}_nd{1}_x^i$, ${}_nd{2}_x^i$ and ${}_nd{3}_x^i$ are the number of deaths, from age *x* to age *x+n* in year *i*, of "Cause 1", "Cause 2" and "Cause 3" respectively. When the reconstruction is over the results have to be checked statistically to detect if there is still any remaining disruption in the series. This procedure, done by cause, age group and sex, usually consist of assessing by eyes the presence of any break, looking at the graph of the series of the total number of deaths. Recently, however, Camarda and Pechholdova have developed an R package to statistically assess the presence of disruptions in cause-specific mortality series thanks to which the long visual inspection procedure can be avoided (Camarda and Pechholdova, 2014). The method relies on a (back)forecasting of the trend from the revision year to the first year considered in the study.

Chapter 3

Data and Methods

In this Chapter, I provide information about the data and the methods used to carry out my investigation on lifespan variability in Italy through causes of death.

3.1 Data

The analysis contained in any Chapter of this thesis, rely on two datasets: the Human Mortality Database (HMD) and the Italian cause-specific database by age and sex from 1980 to 2013.

The first can be downloaded from the website www.mortality.org and was helpful for any long-term and all-cause mortality analysis of Italy and other countries; the second was provided to me by the Italian national institute of statistics (ISTAT) and represents the main source of the dissertation, giving information on cause-specific mortality in Italy. Apart from these two datasets, useful information on Italian bridge-coding data are taken from the ISTAT report *"Analisi del bridge coding ICD-9 ICD-10 per le statistiche di mortalità per causa in Italia"* (Frova et al. 2010). Such information will be helpful in the reconstruction of coherent Italian time series of causes of death.

Below, all data's features and problems are discussed.

3.1.1 Human Mortality Database

The Human Mortality Database is probably the most dominant collection of high-quality data on all-cause mortality available on-line. It contains, for 38 countries, information on:

- Annual live births count by sex.
- Annual deaths count by sex and age.
- Deaths by Lexis triangles
- Annual estimates of population size on January 1st.
- Annual estimates of the population exposed to the risk of death
- Life tables for period and cohort data.

Data availability and the time interval covered vary country by country. Sweden has the longest data series (1751-2014), Chile the shortest (1992-2005). In the case of Italy, period data time series starts in 1872 and ends in 2013, for all the above-listed information. With regard to cohort data,

exposure to risk and death rates are available from 1794 to 1982, while life tables from 1872 to 1921.

A file named *Background and documentation*, containing detailed advice and metadata, is provided for each country. The file includes information on the source of the data, on data coverage, completeness and issues. In Italy the data sources are official vital statistics, census counts and population estimates published by ISTAT. Issues are present for 1872-1905 due to problems with data quality. For most of these early years, indeed, the original death counts were available only in five-year age groups with an open age interval at age 75+. Thus, life tables by single year of age are based on estimates (HMD, 2016). The decision I took was not to consider this time interval, beginning the all-cause mortality analysis in 1906. This criterion is also applied to all other countries which appear throughout the dissertation.

The HMD was launched in 2002 by researchers from the department of Demography at the University of California (Berkeley, USA) and the Max Planck Institute for Demographic Research (Rostock, Germany). In recent years also the French Institute for Demographic Studies (INED) has supported the development of the database. The scope and principles are clearly stated in the overview page of the website: *"The main goal of the Human Mortality Database is to document the longevity revolution of the modern era and to facilitate research into its causes and consequences. As much as possible, we have followed four guiding principles in creating this database: comparability, flexibility, accessibility, reproducibility"*. The database is constantly and punctually updated as well as the number of countries included has been rising over time. However, since the database is limited by design to populations where death registration and census data are virtually complete (HMD, 2016), it is mainly representative of the experience of more developed countries because they fulfill the strict data quality standards required by the creators.

3.1.2 Causes of death in Italy and bridge-coding information

The core of the thesis, represented by a cause-specific approach to the issue of lifespan variability in Italy, is developed using the dataset on causes of death in Italy provided by ISTAT.

The dataset covers the period from 1980 to 2013. For each year the dataset provides deaths count by:

- Cause-of-death at the most detailed level possible (4 digit level)
- Age (single year of age, without any open-ended category)
- Sex

The tenth revision of the International Classification of Diseases (ICD-10) was introduced in Italy in 2003. Therefore deaths from 1980 to 2002 are classified according to ICD-9 while deaths from 2003 to 2013 according to ICD-10. The problem of missing data is almost nonexistent, only for 11 deaths the age at death was not known. In such cases, the decision was to take over the unknown values with the modal age at death (so the most probable value) of the year in which deaths were registered.

The main problem to solve it is represented by the disruptions in the series caused by the introduction of ICD-10 in 2003. In Chapter 2, where a comprehensive literature review of the topic was given, we have seen that in the presence of bridge-coding data (a cross-classification between successive ICD version) it would be possible to determine the exact transfer from old items to new ones and calculate coefficients of redistribution. To complete the reconstruction it would be enough to assume that the coefficient obtained for one year are acceptable for the entire period covered by the ICD version under consideration (Meslé and Vallin, 1988). Fortunately, bridge-coding data between ICD-9 and ICD-10 in Italy exists. In 2003, indeed, ISTAT had classified a sample of causes of death according to both revisions. This sample is large, 454'897 deaths out of 580'200 (78% of total deaths) and includes all the 2'124 deaths happened in the first year of life in 2003. Unfortunately, such bridge-coding data are not available at the moment. However, in the publication of Frova, mentioned in paragraph 3.1, a significant number of information on the Italian bridge-coding are given (Frova et al. 2010). Such information are sufficient to perform a reconstruction of the series, even if not at a very detailed level. Indeed, the data taken from the ISTAT report on bridge coding allows only the reconstruction of seventeen groups of causes. In the publication, in fact, the information are not reported at the finest level of detail possible, making unrealizable the reconstruction for a bigger number of causes. The mentioned seventeen groups of causes for which the reconstruction is performed are reported in Table 3.1. These almost correspond to the ICD-10 main chapters reported in Table 2.2. The difference is that some chapters are grouped into only one. For instance, while in ICD-10 diseases of the nervous system, diseases of the eye and adnexa and diseases of the ear and mastoid process constitute three different chapters, they form only one chapter in the classification used for the reconstruction.

The decision to still use bridge-coding data, even if they allow the reconstruction only of a limited number of causes, comes from the consideration that the transfer of items between ICD-9 and ICD-10 is more complex than ever and bridge-coding data offers a statistically really valid solution to the problem. Of course reconstructing the time-series only for seventeen groups of causes brings some limitations to the analysis at the epidemiological level. For example, it would be very interesting to distinguish between "Ischaemic diseases" and "Cerebrovascular diseases" inside the "Diseases of the circulatory system" group or to split the "Neoplasms" group into several subsets ("Malignant neoplasms of bronchus and lung", "Malignant neoplasms of pancreas", "Malignant neoplasms of breasts" etc). However, even if such more specific analysis cannot be done for the whole period 1980 to 2013, they are feasible from 2003 to 2013, in the ICD-10 period, where no problems of disruptions caused by different revisions are involved. So, although for a shorter time interval, more detailed analysis at the cause-of-death level will be carried out. To do that it will be used the intermediate classification (104 causes) proposed by the Human Cause-of-Death Database (HCOD, 2016). The complete classification is reported in the Appendix.

More in detail, the macro data on bridge-coding, taken from the publication of Frova et al., consists of two correspondence tables, one for deaths occurred in the first year of life and one for deaths above age 1, in which

Cause	ICD-10 code	ICD-9 code
Certain infectious and parasitic diseases	A00–B99	001-139, 279.1
Neoplasms	C00–D48	140-239
Diseases of the blood	D50–D89	279-281
Endocrine, nutritional and metabolic diseases	E00–E90	240-278
Mental and behavioural disorders	F00–F99	290-319
Diseases of the nervous system	G00–H95	320-389
Diseases of the circulatory system	I00–I99	390-459
Diseases of the respiratory system	J00–J99	460-519
Diseases of the digestive system	K00–K93	520-579
Diseases of the skin and subcutaneous tissue	L00–L99	680-709
Diseases of the musculoskeletal system and connective tissue	M00–M99	710-739
Diseases of the genitourinary system	N00–N99	580-629
Pregnancy, childbirth and the puerperium	O00–O99	630-676
Certain conditions originating in the perinatal period	P00–P96	760-779
Congenital malformations, deformations and chromosomal abnormalities	Q00–Q99	740-759
Ill-defined causes	R00–R99	780-799
External causes of mortality	V01–Y98	E800-E999

TABLE 3.1: The seventeen group of causes for which the reconstruction is performed.

deaths are cross-classified, between ICD-9 and ICD-10, for seventeen group of causes. As already said, the sample for which the double coding was performed includes the totality of deaths at age 0. Thus, the correspondence table in the first year of life represents exactly the match between the two ICD revisions. This is not the case for the correspondence table of deaths above age 1 because it doesn't contain all deaths but just a sample of them. Anyway, given the considerable largeness of the sample the accuracy of the results is ensured. Having two correspondence tables, that split deaths into below and above age 1, it's an advantage for the reconstruction. It permits to compute coefficients of redistribution separately for these two age groups improving the reliability of the reconstruction since deaths at age 0 are really particular in terms of causes of death and their cross-classification it is likely to be very different from the one of other age groups.

Table 3.2 displays the correspondence table of deaths above age 1 for a subset of the seventeen causes for which the cross-classification is available (the two complete transition matrices are reported in the Appendix). In the first row it is possible to see that, in the sample of double classified deaths, 3713 were assigned to infectious diseases according to ICD-9. Out of this 3713 deaths, 3192 were also assigned to infectious diseases according to ICD-10, while 101 to neoplasms, 4 to external causes and so on. On the other hand, the first column can be read in a specular way: 4635 deaths were assigned to infectious diseases according to ICD-10 and out of these 4635 deaths, 3192 were also assigned to infectious diseases according to

ICD-10 → ICD-9 ↓	Infectious diseases	Neoplasms	...	External causes	Total-9
Infectious diseases	3192	101	...	4	3713
Neoplasms	97	127626	...	24	128509
⋮	⋮	⋮	⋱	⋮	⋮
External causes	13	137	...	5466	6163
Total-10	4635	130959	...	5911	452773

TABLE 3.2: Transition matrix of deaths above age 1 in Italy, 2003. Subset of the seventeen causes for which the cross-classification is performed. Adapted from Frova et al., 2010.

ICD-9, while 97 to neoplasms, 13 to external causes and so on. This correspondence table, and the one of deaths below age 1, will be used to compute coefficients of redistribution and eventually reconstruct the time series. A full explanation of the method will be given soon in Paragraph 3.2.2.

3.2 Methods

In this section I provide a broad overview of the methods on which I rely my studies about length of life variability. Some details on the methodological strategies performed, however, can also be found in the substantive Chapters of the dissertation.

3.2.1 Measures of lifespan variability

In Chapter 2 I have given a formal definition of lifespan variability and I have listed a series of measures which have been used to assess mortality compression and/or the rectangularization of the survival curve. Among all of them I decided to rely my analysis on S_{10} and *SD(M+)*. Before laying out my reasoning for adopting them I give a formal definition of these two indicators.

Let x be the age at death, d_x the number of deaths that occur at age x, a_x the average time lived from age x to $x+1$ for those dying in the interval, M_{10} the average age at death for those who survive to age ten (e_{10}+10), M the modal age at death and w the oldest age group. The standard deviation of ages at death above age 10, S_{10}, and the standard deviation above the

modal age at death, *SD(M+)*, are defined as follow:

$$S_{10} = \sqrt{\frac{\sum_{x=10}^{w}(x + a_x - M_{10})^2 d_x}{\sum_{x=10}^{w} d_x}} \tag{3.1}$$

$$SD(M+) = \sqrt{\frac{\sum_{x=M}^{w}(x + a_x - M)^2 * d_x}{\sum_{x=M}^{w} d_x}} \tag{3.2}$$

S_{10} has been popularized by the work of Edwards and Tuljapurkar (Edwards and Tuljapurkar, 2005; Edwards, 2009b; Edwards, 2009a; Tuljapurkar and Edwards, 2009) while *SD(M+)* was firstly used by Kannisto (Kannisto, 2000; Kannisto, 2001) and more recently especially by Cheung and Robine (Cheung and Robine, 2007; Cheung et al., 2009). The formulas can, of course, be adapted to causes of death just by using the cause-specific density function instead of the all-cause density function and the cause-specific M_{10} instead of the all-cause M_{10}.

As seen in Paragraph 2.1.1, over the course of last decades the study of lifespan inequality has produced a lot of indicators to analyze this phenomenon. All these measures capture a different kind of variability and it is the researcher that has to choose the one that best fits his/her aims. I, first of all, decided to use two indicators because I intend to study two distinct aspects: lifespan variability over the whole age range and lifespan variability over a little age range, the one of very old. This decision comes from the consideration that, in the era of longevity extension, has become increasingly important to monitor the evolution of lifespan variability at old age but, at the same time, it is also important to look at how the shape of the age-at-death distribution is changing in its whole.

A possible alternative to *SD(M+)* could be the standard deviation above the third quartile or the mean square deviation around the modal age at death or the standard deviation of ages at death after a given age such as 65 or 70. *SD(M+)*, however, it is better linked to parametric mortality models and its relationship with the rate of aging make possible a deeper interpretation of its evolution over time. Moreover, the fact that *SD(M+)* is (obviously) related to *M* is also an advantage. Indeed, as seen in Paragraph 2.1.2, the modal age at death, although not a measure of compression, can be very useful when studying lifespan variability. It quantifies how much the age-at-death distribution is shifting to the right and, as seen before, the number of deaths at the mode, *d(M)*, can be related to lifespan variability in various parametric mortality models.

About the choice of S_{10} an initial consideration has to be made. Variability indicators that take into account the entire age range, as known as unconditional indicators, are heavily influenced by infant and child mortality. The truncation at age 10 eliminates this problem, that instead is still present in measures like *IQR*, *SDM* and $e^{\dagger}$. Moreover, truncation also produces a distribution that is relatively normal making it unnecessary to employ percentile based measures like *IQR* and C_{50}, which would be more

advantageous if the distributions were heavily skewed (Edwards and Tuljapurkar, 2005). Finally, the standard deviation is invariant on an additive scale, i.e. if every death is postponed by 5 years the standard deviation remains the same. On the contrary measures like the Gini index or the Theil index are invariant on a proportional scale, i.e. if every lifespan is doubled these measures stay the same. While some would argue that invariance on a proportional scale is preferable in the case of income inequality, Edwards argues that demographers are more interested in additive differences in life span thus a measure of inequality that is invariant on an additive scale is more desirable (Edwards, 2009b).

3.2.2 Putting together ICD-9 and ICD-10

As explained in Paragraph 3.1.2, in the ISTAT database deaths are assigned to a certain cause according to two different ICD revisions, ICD-9 and ICD-10. ICD-10 was introduced in 2003, therefore the aim is to redistribute deaths from 1980 to 2002, that were classified in line with ICD-9, according to the most recent revision, ICD-10. The reconstruction of the series represents the basis of any further study of lifespan variability in Italy by cause of death. Before analyzing data, indeed, it necessary to ensure the coherence of their information and so to perform the reconstruction of the time-series of causes of death.

The reconstruction carried out in this dissertation relies, as already said, on the macro information available in the ISTAT report *"Analisi del bridge coding ICD-9 ICD-10 per le statistiche di mortalità per causa in Italia"* about the bridge-coding performed in Italy in 2003 (Frova et al., 2010). In Paragraph 2.3.2, I have given an overview of the reconstruction method developed by Vallin and Meslé. This method consists of three steps: the construction of a dual correspondence table linking the items of the two revisions, the definition of fundamental association of items gathering the same medical diagnosis at both revisions and ensuring a statistical continuity and finally the elaboration of a transition table indicating how to redistribute the deaths. Thanks to the macro information on Italian bridge-coding, it is possible to skip the first two steps and directly elaborate the transition table. As said in paragraph 3.1.2, these information consist of two correspondence tables, i.e. the cross-classification of deaths between ICD-9 and ICD-10 for seventeen groups of causes of death. One correspondence table regards deaths occurred in the first year of life while the other one deaths above age 1. This latter table is then used to get one tailored for men. In fact, since males' death cannot be caused by "Pregnancy, childbirth and puerperium" the horizontal and vertical cells corresponding to this cause of death, have to be converted to 0.

To compute the transition coefficient and reconstruct the series, each cell of the correspondence table is divided by the correspondent total row, just as in the Vallin and Meslé method. This leads to the tuning of other two matrices which contains the transition coefficients at age 0 and at age greater then 0 separaely (these two matrices can be found in the Appendix). The advantage of having specific coefficient for age 0 is that, since at this age the deaths cross-classification it is likely to be very different from the one of other age groups, it ensures a higher reliability when the reconstruction is performed.

The coefficient are then applied for one-year age group and for each sex separately. Of course, age 0 has is own transition coefficient while to any other age group are applied the same coefficients. So, for instance, the number of deaths due to infectious diseases at age 0, in year *i* (*i=1980, . . . , 2002*), is computed as follow:

$$\begin{aligned} {}_id_0^{Inf} =_i d_0^{Inf} * 0.767 +_i d_0^{Neo} * 0 +_i d_0^{Blo} * 0 +_i d_0^{ENM} * 0.00+ \\ +_i d_0^{Men} * 0 +_i d_0^{Ner} * 0 +_i d_0^{Cir} * 0 +_i d_0^{Res} * 0+ \\ +_i d_0^{Dig} * 0 +_i d_0^{Ski} * 0 +_i d_0^{Mus} * 0 +_i d_0^{Gen} * 0+ \\ +_i d_0^{Per} * 0.200 +_i d_0^{Con} * 0.033 +_i d_0^{Ill} * 0+ \\ +_i d_0^{Ext} * 0 \end{aligned} \tag{3.3}$$

While at any other age *x*:

$$\begin{aligned} {}_id_x^{Inf} =_i d_x^{Inf} * 0.860 +_i d_x^{Neo} * 0.027 +_i d_x^{Blo} * 0.013 +_i d_x^{ENM} * 0.006+ \\ +_i d_x^{Men} * 0.004 +_i d_x^{Ner} * 0.008 +_i d_x^{Cir} * 0.020 +_i d_x^{Res} * 0.013+ \\ +_i d_x^{Dig} * 0.029 +_i d_x^{Ski} * 0.007 +_i d_x^{Mus} * 0.004 +_i d_x^{Gen} * 0.005+ \\ +_i d_x^{Pre} * 0 +_i d_x^{Con} * 0.001 +_i d_x^{Ill} * 0.001+ \\ +_i d_x^{Ext} * 0.001 \end{aligned} \tag{3.4}$$

The assumption underlying the reconstruction is, thus, that the coefficients are valid for the entire period (1980-2002) covered by the ICD revision under consideration.

Once coefficients are applied to all age groups and causes the reconstruction is accomplished. The work, however, is not finished yet. As mentioned in the literature review chapter, the results obtained have to be checked to detect if any important disruption is still present. Consequently, to assess the goodness of the reconstruction, the presence any potential disruption is checked by eyes, looking at the graph of the series of the expected total number of deaths for each cause, age group and sex. Although rough, this method it is still considered by the expert the better one (Meslé, 2010). Luckily, since the number of reconstructed causes is only 17, the number of graphs to be checked is still reasonable. In case of particular indecision, the mathematical method (see Paragraph 2.3.2), developed by Camarda and Pechholdova, to detect statistically significant discontinuity, was also used to support the final decision (Camarda and Pechholdova, 2014).

In order to get the expected total number of deaths and assess the presence of any disruption devoid of changes in age structure of the population, let $\mathbf{D} = (d_{ij})$ be the matrix of deaths at age *i*=1, . . . ,110 and year *y*=1980, . . . , 2013 and $\mathbf{E} = (e_{ij})$ be the matrix of exposure to risk over the same dimensions. The matrix of cause specific death rates can be easily computed as follow:

$$\mathbf{M} = (m_{ij} = \frac{d_{ij}}{e_{ij}}) \tag{3.5}$$

The standardized death rates are given by:

$$\mathbf{M}^s = diag(\mathbf{p}) * \mathbf{M} \tag{3.6}$$

Where $\mathbf{p} = (p_i) : \sum_{i=1}^{110} p_i = 1$ is the standard age structure of the population (in this case the age structure of the Italian population in 2000). Then, to get the expected number of deaths over age and year, M^s is multiplied by a factor *k* which represent the average of the population size (in the period covered by the analysis) in a specific age group:

$$\mathbf{D}_i^s = k_i * \mathbf{M}^s \tag{3.7}$$

After the visual inspection on the expected number of deaths for each cause, age group and sex a few a-posteriori corrections have been made, especially for young ages. Then, the new transition coefficients, obtained after these corrections, are applied again in the same way as before to get, finally, the reconstructed series of causes of death.

3.2.3 Smoothing densities: P-Splines

After completing the reconstruction, data are ready to be analyzed. To study cause-specific lifespan variability, I assumed that causes of death were mutually exclusive and exhaustive, using the principles of the multiple-decrement life-table (Preston et al., 2000).

In this context, the all-cause force of mortality, $\mu(x)$, is defined as the sum of cause-specific force of mortality, $\mu_k(x)$:

$$\mu(x) = \lim_{\Delta x \to 0} \frac{P(x < X < x + \Delta x | X > x)}{\Delta x} \tag{3.8}$$

$$\mu_k(x) = \lim_{\Delta x \to 0} \frac{P(x < X < x + \Delta x, k = k | X > x)}{\Delta x} \quad k = 1 \dots K \tag{3.9}$$

$$\mu(x) = \mu_1(x) + \mu_2(x) + \cdots + \mu_k(x) + \cdots + \mu_K(x) \tag{3.10}$$

The cause-specific density function (i.e the cause-specific age-at-death distribution), $d_k(x)$, thus, can be written as follow:

$$d_k(x) = \mu_k(x) l(x) \quad k = 1 \dots K \tag{3.11}$$

Where $l(x)$ is the all-cause survivorship function.

These equations are presented in a continous setting whereas the data are in a discrete setting (one year age interval). Assuming a constant cause-specific force of mortality over each one year age interval, $\mu_k(x)$ is estimated by the central death rate, $m_k(x) = D_k(x)/e(x)$, where $D_k(x)$ is the number of deaths of cause *k* at age *x* and $e(x)$ is the number of persons exposed to the risk of dying at age *x*.

Once calculated cause-specific density function, $d_k(x)$, and all the other usual function of a multi-decrement life table, the two indicators used to assess cause-specific mortality compression in this dissertation, S_{10} and $SD(M+)$, can be easily calculated from equation 3.1 and 3.2 respectively. Very often, indeed, these two measures are computed taking directly the

age-at-death distribution extracted from life tables (multi-decrement life tables in this case). Another option, many times used in literature, it is to fit the function using a parametric model. These two choices, however, present both some problematic. The first use an age-at-death distribution that tends to be erratic, at least in some points (this is much more evident for cause-specific age-at-death distributions because the number of deaths is lower). The age at death, in fact, can be seen as a random process and irregularities in the distribution are likely to be caused by chance. Therefore, the computation of S_{10} and $SD(M+)$, or any other compression indicator, based on life table densities (again, multi-decrement life tables densities in case of cause-specific approach) could be not very accurate. The use of a parametric model allows avoiding this erratic age-at-death distribution problem. Parametric models, however, are sometimes too rigid and/or do not fit the data well. A solution to these issues is offered by the adoption of non-parametric models. As parametric models do, non-parametric models overcome the issue of irregularities in the age-at-death distribution since they smooth the cause-specific density function, $d_k(x)$, but they also have the advantage not to impose any predetermined structure on the data. Thus, thanks to this flexibility, the problematics related to the use of parametric models mentioned above, is overcome too.

For these reasons I decided to use a non-parametric approach, whose outcomes are exploited for the construction of continuous life tables and multi-decrement life tables and then for analyzing the changes in the shape of all-cause and cause-specific density functions. Indeed, S_{10} and $SD(M+)$, and the modal age at death, *M*, will all be computed based on the smoothed density. The estimation of the modal age at death is particularly interesting. As mentioned above, the age-at-death distribution of life table tends to be erratic, and this is particularly true in the area surrounding the mode. In the past various methodologies have been proposed to estimate the modal age at death (Pearson, 1902; Kannisto, 2001; Canudas-Romo, 2008; Cheung, Robine and Caselli, 2008; Thatcher et al. 2010). One of the most used is Kannisto's quadratic procedure, which consists of fitting a quadratic model using life table deaths at ages *M*, *M-1* and *M+1*. A problem with this approach is that *M* doesn't always emerge clearly and so there are various age candidates which of course lead to different estimates of the modal age at death. The smoothing provides a solution to this problem and the ability to refine the monitoring of changes in modal age at death over time (Ouellette and Bourbeau, 2011). The estimation of *M* will be implemented just finding the (unique) maximum of smoothed density functions:

$$\hat{M} = \max_{\forall x \in X} d(x) \quad X = 10, \ldots, 110 \tag{3.12}$$

The non-parametric approach adopted in this dissertation is called P-splines which has been already used in demography in the latest years (Camarda, 2008; Ouellette and Bourbeau, 2011; Diaconu et al., 2016). Among all non-parametric models available, the choice of P-splines is justified by its good properties in the context of mortality. First of all, this method has not boundary effects and thus it performs well at the border of the domain over which smoothing operates (very old age at death in this case). Secondly, even if in the domain there are some zeros - there are ages at which

no deaths occur - P-splines works finely as well. Moreover, P-splines is simple to use, program, understand and previous studies have already demonstrated that it performs well for smoothing anomalously distributed data such as mortality data (Currie et al., 2004; Camarda, 2008; Ouelette and Bourbeau 2011).

In this dissertation, P-splines is applied to data from age 10, because infant and child mortality present unique features that would require the use of a smoothing method suited for ill-posed data (Ouelette and Bourbeau, 2011). This is not a problem, however, because both indicators used to assess mortality compression do not consider deaths before age 10.

P-splines can deal with aggregated data which are assumed to be Poisson distributed. Now, to smooth mortality data with P-splines, let $\boldsymbol{d_k}$ denote the vector of observed death counts, $\boldsymbol{e_k}$ the vector of exposure to risk and $\boldsymbol{\mu_k}$ the vector of force of mortality, for a given cause of death k. The response variable, $\boldsymbol{d_k}$, is assumed to follow a Poisson distribution $\boldsymbol{d_k} \sim$ P($\boldsymbol{e_k} * \boldsymbol{\mu_k}$), so the expected value is the product of exposure to risk and force of mortality. It is possible to model the Poisson data introducing the linear predictor $\boldsymbol{\eta_k}$:

$$\boldsymbol{\eta_k} = \mathbf{ln}(\boldsymbol{E}|\boldsymbol{d_k}|) \tag{3.13}$$

$$\boldsymbol{\eta_k} = \mathbf{ln}(\boldsymbol{e_k} * \boldsymbol{\mu_k}) = \mathbf{ln}(\boldsymbol{e_k}) + \mathbf{ln}(\boldsymbol{\mu_k}) = \mathbf{ln}(\boldsymbol{e_k}) + \boldsymbol{Ba} \tag{3.14}$$

The P-splines approach is then used to estimate the unknown parameters that through a linear combination model $\boldsymbol{\eta_k}$. $\boldsymbol{B}$ is the B-spline basis matrix and $\boldsymbol{a}$ is the vector of respective regression parameters to estimate:

$$\hat{\boldsymbol{\eta}}_{\boldsymbol{k}} = \mathbf{ln}(\boldsymbol{e_k}) + \boldsymbol{B}\hat{\boldsymbol{a}} \tag{3.15}$$

Taking the exponential of equation 3.15 gives smoothed death counts. Those are, however, not standardized by age because influenced by exposure to risk and as a consequence not comparable. To get age-standardized death counts, one must extract the smoothed forces of mortality from the smoothed death counts. Given equation 3.15, a smoothed trend for the force of mortality is readily obtained as (Ouellette and Bourbeau, 2011):

$$\hat{\boldsymbol{\mu}}_{\boldsymbol{k}} = \boldsymbol{exp}(\boldsymbol{B}\hat{\boldsymbol{a}}) \tag{3.16}$$

The penalized log-likelihood function to maximize in order to find $\hat{\boldsymbol{a}}$ is:

$$l_k^* = l(\mathbf{a}; \mathbf{B}; \mathbf{d}) - \frac{1}{2}\lambda \mathbf{a}^{'}\mathbf{D}^{'}\mathbf{D}\mathbf{a} \tag{3.17}$$

Where λ represent the scalar smoothing parmeters and $\boldsymbol{D}$ the difference matrix.

The non-parametric approach of P-splines has been implemented through the R package "Mortality Smooth" (Camarda, 2012). Figure 3.1 displays, as an example, a comparison between the life table age-at-death distribution and the smoothed age-at-death distribution of infectious diseases in Italy in 2008. As expected, the two curves are very similar, but the smoothed

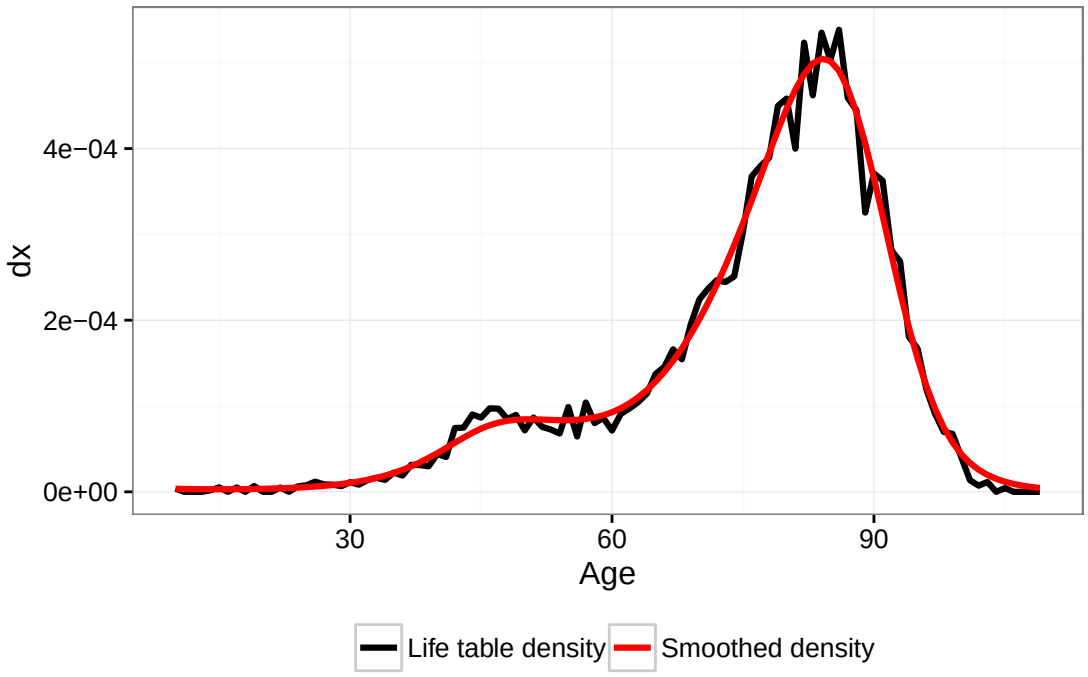


FIGURE 3.1: Comparison between life-table and smoothed cause-specific density. Infectious diseases, 2008. Total population.

one has the additional advantage of eliminating chance, and so irregularities, from the distribution leading to better estimates of variability measures and M.

Not all the analysis conducted in this dissertation, however, will rely on smoothed densities. Indeed, the investigation on old age mortality compression, here measured by $SD(M+)$, will be largely carried out using the simplified logistic model with two parameters. Therefore in Chapter 6, where the topic of old age lifespan variability is treated, $SD(M+)$ will be not only estimated according to smoothed densities but also according to Kannisto model. The model, already briefly shown in Chapter 2, will be presented more in detail in Paragraph 3.2.6.

3.2.4 Decomposition techniques

Decomposition techniques are often used in demography to disentangle the single components of a phenomenon. In this dissertation, to have a better comprehension of lifespan variability in Italy, I decompose both trends in S_{10} over time and differences in S_{10} between sexes. In particular, the results of the decomposition will quantify either the contribution of changes in age and cause-specific mortality rates, $m_{x,k}$, from 1980 to 2013 to change in S_{10} over the same period or the contribution of differences in age and cause-specific mortality rates between sexes to the difference in S_{10} between sexes at fixed point in time (1980 and 2013).

Among the several decomposition techniques available, I rely my analysis on a method recently developed by Horiuchi et al. that can be applied

to any dependent variable and its covariates, as long as the former is a differentiable function of the latter (Horiuchi et al., 2008). So, here the interest is in determining how the total change in the dependent variable S_{10} is decomposed into the contribution of changes of the covariates, i.e. age and specific cause-of-death mortality rates. First of all, it is necessary to express S_{10} in terms of $m_{x,k}$ (Zurieck, 2010):

$$S_{10} = \sqrt{\frac{\sum_{x=10}^{w}(x + a_x - M_{10})^2 d_x}{\sum_{x=10}^{w} d_x}} \tag{3.18}$$

$$= \sqrt{\frac{\sum_{x=10}^{w}(x + a_x - M_{10})^2 l_x m_x}{\sum_{x=10}^{w} l_x m_x}} \tag{3.19}$$

$$= \sqrt{\frac{\sum_{x=10}^{w}(x + a_x - \sum_{x=10}^{w} x(exp(-\sum_0^x m_x))m_x)^2 (exp(-\sum_0^x m_x))m_x}{\sum_{x=10}^{w}(exp(-\sum_0^x m_x))m_x}} \tag{3.20}$$

$$= \sqrt{\frac{\sum_{x=10}^{w}(x+a_x - \sum_{x=10}^{w} x(exp(-\sum_{a=0}^{x}\sum_{k=1}^{j} m_{a,k}))\sum_{k=1}^{j} m_{x,k})^2 (exp(-\sum_{a=0}^{x}\sum_{k=1}^{j} m_{a,k}))\sum_{k=1}^{j} m_{x,k})}{\sum_{x=10}^{w}(exp(-\sum_{a=0}^{x}\sum_{k=1}^{j} m_{a,k}))\sum_{k=1}^{j} m_{x,k}}} \tag{3.21}$$

In general, the method holds that for $y = f(x_1, x_2, \ldots, x_n)$ the change in y from time 1 to time 2 (or between males to females) can be expressed as:

$$y(2) - y(1) = \sum_{i=1}^{n} c_i \tag{3.22}$$

$$c_i = \int_{X_{i1}}^{X_{i2}} \frac{\delta y}{\delta x_i}\frac{dx_i}{dt}dt \tag{3.23}$$

So, the change in the dependent variable y is the sum of the contributions, c_i, of each covariate, x_i. X_{i1} and X_{i2} are the values of x_i at time 1 and time 2 respectively (or the values of x_i for males and females respectively).

For the purposes of this dissertation, I want to take partial derivatives of S_{10} with respect to $m_{x,k}$ to get the contribution of changes in age and cause-specific mortality rates, $m_{x,k}$, to changes in S_{10}:

$$S_{10}(2) - S_{10}(1) = \sum_{x=1}^{w}\sum_{k=1}^{j} c_{x,k} \tag{3.24}$$

$$c_{x,k} = \int_{m_{x,k,2}}^{X_{x,k,1}} \frac{\delta S_{10}}{\delta m_{x,k}}\frac{dm_{x,k}}{dt}dt \tag{3.25}$$

The same decomposition techniques have been used to test the level effect hypothesis and slope effect hypothesis, in Chapter 6, in the context of

old-age lifespan variability. More details about this decomposition will be given in Paragraph 6.2.1.

The method has been implemented through the R package "DecompHoriuchi" (Riffe, 2011).

3.2.5 Sensitivity analysis

As said in Chapter 2, understanding why lifespan variability (the second moment of the density function d_x) decrease or increase is much more complex than understanding why life expectancy (the first moment of the density function d_x) decrease or increase. A decline (rise) in mortality rates at any age, indeed, acts incrementing (declining) the average length of life, while this is not always the case for variability measures: the effects of mortality decline or rise on lifespan variability varies by age. In particular, for $e^{\dagger}$ (life expectancy lost due to premature death) and V_x (Variance of age at death after age *x*), has been demonstrated that exists an age before which avoiding deaths reduces inequality and after which avoiding deaths increases inequality (Gillespie et. al, 2014; Zhang and Vaupel, 2009).

An important tool to understand how changes in mortality rates affect variability of age at death is sensitivity analysis. In the field of demography, sensitivity analysis has been mainly developed by Caswell and his collaborators and allows quantifying, for each age, the potential effects of proportional mortality increase/decrease on a certain lifespan variability measure (Caswell, 1978; Caswell 2006; Caswell, 2008; Caswell, 2009; Caswell, 2010; Caswell and Ouellette, 2015; Van Raalte and Caswell, 2012.). The method rilies on matrix calculation and it is formulated in terms of a Markov chain [1]. Let **P** be the transition matrix for the Markov chain:

$$\mathbf{P} = \begin{bmatrix} \mathbf{U} & \mathbf{0} \\ \mathbf{M} & I \end{bmatrix} \tag{3.26}$$

Where **U** is a matrix with survival probabilities on the subdiagonal and zeros elsewhere, describing transitions among the transient states in the Markov chain; **M** is a diagonal matrix in which deaths, the absorbing state of the Markov chain, are classified by age class, **0** is a matrix of zeros and **I** is the identity matrix.

The statistical properties of longevity, i.e. the time to absorption in the Markov chain, can be computed directly from **P**. For instance, the mean time spent in age class *i*, conditional on starting in age class *j* is given by the *(i, j)* entry of the matrix:

$$\mathbf{N} = (\mathbf{I} - \mathbf{U})^{-1} \tag{3.27}$$

The mean time to absorption, $\boldsymbol{\eta}$, can be computed by the column sums of **N**:

$$\boldsymbol{\eta}^T = \mathbf{e}^T\mathbf{N} \tag{3.28}$$

[1] An exhaustive explanation of the method can be found in Caswell, 2008; Caswell, 2009; Caswell, 2010.

where **e** is a vector of ones. Instead, the vector of variances in longevity conditional upon survival to age class *i*, **v**, is given by:

$$\mathbf{v}^T = \mathbf{e}^T\mathbf{N}(2\mathbf{N} - \mathbf{I}) - \boldsymbol{\eta}^T \circ \boldsymbol{\eta}^T \tag{3.29}$$

where $\circ$ denotes the element by element product (as knows as Hadamard product).

The standard deviation in longevity conditional upon survival to age class *i*, **s**, is simply the square root of **v**:

$$\mathbf{s} = \sqrt{\mathbf{v}} \tag{3.30}$$

Here, the aim is to compute the sensitivity of S_{10} to a proportional drop in mortality rates. This is obtained by taking the derivative of the standard deviation in longevity upon survival to age class 10 vector with respect to the vector of underlying age-specific mortality rates, $\boldsymbol{\theta}$. In formula:

$$\frac{\delta \mathbf{s}}{\delta \boldsymbol{\theta}^T} = \frac{1}{2} diag(\mathbf{s})^{-1} \frac{\delta \mathbf{v}}{\delta \boldsymbol{\theta}^T} \tag{3.31}$$

In this way it will be possible to figure out what would be the contribution of each age to a change in S_{10} if mortality decreased proportionally. As a consequence, the results of sensitivity analysis will also permit to identify a threshold age, T_{10}, before which mortality decline decreases S_{10} and after which mortality decline increases S_{10}.

3.2.6 Kannisto model

The first noticeable attempt to model mortality was made by Benjamin Gompertz in 1825, who described the force of mortality as an exponential function, i.e. mortality increases exponentially with age (Gompertz, 1825). In 1860, Makeham added an age-independent component to the Gompertz model to capture those deaths which are not related to aging (Makeham, 1860). The Makeham model describes adult mortality accurately until age 80 but it is not able to capture the late-life mortality deceleration since it has been shown in many populations that the force of mortality increases at a decreasing rate at very old ages. To solve this problem, logistic models have been proposed to model mortality since the presence of the denominator bends the curve downwards at high ages. They have been published in different forms, but all can be converted into the following expression:

$$\mu(x) = \frac{k * ae^{bx}}{1 + ae^{bx}} + c \tag{3.32}$$

To investigate on old-age variability I adopt a simplified version of the logistic model with 2 parameters instead of 4, known as Kannisto model. In this model, indeed, the k parameter is equal to 1 and the c parameter, representing background mortality introduced by Makeham, is set to 0 since his value is insignificant above age 70. So, the function modeling mortality in the Kannisto model, already reported in Chapter 2, is as follow:

$$\mu(x) = \frac{ae^{bx}}{1 + ae^{bx}} \tag{3.33}$$

Using the logit function the model can be expressed as:

$$logit(\mu(x)) = ln(a) + bx \tag{3.34}$$

Meaning that the force of mortality will lie on a straight line - the logit line-, represented by equation 3.30. The fact that, in modern population, mortality above age 70 fits the logit line well has been shown in many studies (Cheung and Robine, 2007; Kannisto, 1992; Thatcher, 1998, Thatcher et al., 2010).

What is appealing of this model is the property of linking the rate of aging and mortality compression. According to the model, indeed, at old ages $SD(M+)$ is uniquely determined by the rate of increase in mortality, i.e. the rate of aging, which is measured by the parameter *b*: as *b* increases $SD(M+)$ decreases. So, mortality compression occurs if the logit of the death rate at younger ages falls faster than the logit of the death rate at older ages or, in other words, if the slope of the logit line becomes steeper as time passes by.

The other parameter of the model, *a*, which indicates the level of mortality - the intercept of the logit line-, doesn't influence $SD(M+)$. Indeed, even if age-specific death rates depend on both *a* and *b*, compression depends only on *b*. The exact relationship between *b* and $SD(M+)$ it is as follow:

$$SD(M+) = \sqrt{\sum_{x=M}^{\infty} x^2 (\frac{1+b}{1+be^{bx}})^{1/b} \frac{be^{bx}}{1+be^{bx}}}^{2} \tag{3.35}$$

As it will be possible to see in Chapter 6, the steepness of the logit line's slope has increased over time in both males and females in Italy over the course of the considered period (1980-2013), even though the increase has been more marked for males. As a consequence, the uncertainty of deaths above the mode has been reduced. Results about the compression of $SD(M+)$ and the role played by the various causes in increasing the *b* parameter are presented in Chapter 6.

To apply the model, the parameter *b* has to be estimated. As shown by Thatcher et al., since *b* is the slope of the logit line, it is sufficient to know the value of $\mu(x)$ at any two ages, x_1 and x_2, where $x_2 > x_1$. The slope of the line between these ages is then given by (Thatcher et al., 2010):

$$b = \frac{logit(\mu(x_2)) - logit(\mu(x_1))}{x_2 - x_1} \tag{3.36}$$

To estimate *b* from lifetable, the force of mortality $\mu(x)$ is approximated by the central death rate m_x:

$$b = \frac{logit(m(x_2)) - logit(m(x_1))}{x_2 - x_1} \tag{3.37}$$

How to choose ages x_1 and x_2? The study of Thatcher et al. picks 70 as x_1 and 90 as x_2. Their choice has been made considering the following criterions. First, x_1 has to be high enough for the model to fit data well: the background mortality component, here set to 0, may have some effects at younger old ages. But, on the other hand, x_1 has to be under the

[2] For a proof of the equation see Thatcher et al., 2010

modal age at death. Secondly, the age range between x_1 and x_2 should be wide enough because a wider difference will make the standard error of b smaller. However, it is also desirable for x_2 not to be an extremely old age because the small number of deaths may notably increase the standard error of logit(m_2), and in turn, that of b.

Considering these issues, ages 70 and 90 seems to be the best possible choice and are also used in this dissertation to apply the Kannisto model. Age 70, indeed, is high enough in order to not consider background mortality but also low enough to be under all-cause and cause-specific modal age at death. To choose x_2 I compared the results of ages 70-85, 70-90 and 70-95. The parameter trends were shown to be less erratic for age 70-90 than 70-85 and 70-95, and so age 90 was chosen.

Chapter 4

Causes of death in Italy after reconstructing coherent time-series

In this Chapter I intend to show how the reconstruction has changed causes of death series in Italy, i.e. how deaths from 1980 to 2002, originally classified according to ICD-9, are reclassified according to the most recent revision, ICD-10. Then, having in hand coherent time-series, I will present elaborations on cause-specific mortality in Italy over the course of the period considered in this dissertation (1980-2103), with the aim of reviewing the state of the health transition in Italy.

4.1 Reconstructing the 1980-2013 time series

As seen in Paragraph 2.3.2, the introduction of a new ICD revision brings disruptions in the causes time-series for many reasons. Here these discontinuities will be quantified and analyzed to reconstruct coherent causes of death series for the years 1980-2013, making possible, in later chapters, a logical and consistent analysis of the length of life variability through causes of death.

4.1.1 The changes introduced by ICD-10

The introduction of ICD-10 represented, surely, a moment of big innovation in the history of the International classification of diseases. Such innovation comes from the impressing increase in the level of detail (from 5600 items of ICD-9 to more than 10000 items of ICD-10) and the changes in the rules governing the selection of the underlying cause of death.

Tables 4.1 gives a first idea of the magnitude of changes introduced by the new classification in the cause-specific time-series. It reports, for each of the 17 causes, the percentage of deaths that are assigned to the same group according to both the new and the old ICD classification. In total, the 93.81% of deaths remained in the same cause of death group. This great value is primarily driven by the fact that for the two major causes of death, namely neoplasms and diseases of the circulatory system, which represent more than two third of the total deaths, the percentage of deaths staying in the group it is considerably high. This is especially true for neoplasms for which the percentage is slightly below 100%.

On the other hand, the percentage of deaths that are assigned to the same group according to both ICD-9 and ICD-10 it is noticeably lower for

diseases of the blood (65.25%), mental and behavioral disorders (60.67%) and congenital malformations (70,34%).

Cause	**Percentage of deaths remaining in the group**
Certain infectious and parasitic diseases	85.93
Neoplasms	99.31
Diseases of the blood	65.25
Endocrine, nutritional and metabolic diseases	93.71
Mental and behavioural disorders	60.67
Diseases of the nervous system	91.80
Diseases of the circulatory system	94.40
Diseases of the respiratory system	87.54
Diseases of the digestive system	93.12
Diseases of the skin and subcutaneous tissue	87.46
Diseases of the musculoskeletal system and connective tissue	89.06
Diseases of the genitourinary system	91.34
Pregnancy, childbirth and the puerperium	100.00
Certain conditions originating in the perinatal period	100.00
Congenital malformations, deformations and chromosomal abnormalities	70.34
Ill-defined causes	91.55
External causes of mortality	88.71
Total	93.81

TABLE 4.1: Percentage of deaths assigned to the same group according to both ICD-9 and ICD-10.

Another, and complementary, question arising after the reconstruction it is how deaths that don't stay in the same cause of death group are then redistributed among other groups. In this respect, Table 4.2 reports, for each of the 17 groups of causes, the corresponding group where the majority of the deaths (in percentage) move to passing from ICD-9 to ICD-10.

In 9 cases the diseases of the circulatory system are the cause of death which receives more deaths than the others. The fact is certainly not surprising, the diseases of the circulatory system are, especially in the era of longevity extension, frequently involved in the process of death even if not always codified as underlying cause of death but as contributing cause of death (Desesquelles et al., 2010; Desesquelles et al., 2016). As seen in Paragraph 2.3.1, the changes to the rules governing the selection of the underlying cause of death between ICD-9 and ICD-10 played a major role in bringing statistical disruptions. Thus, due to the new rules, in many cases where the diseases of the circulatory system were previously selected as contributing cause are now selected as underlying.

The second cause in terms of number of deaths, neoplasms, are instead the group where the majority of deaths move to only in one case, one of the diseases of the blood. Another fact to be noted is the mutual exchange of deaths between infectious diseases and the diseases of the digestive system. Finally, since for pregnancy and conditions of the perinatal period the

Cause	Cause where the majority of deaths move to
Certain infectious and parasitic diseases	Diseases of the digestive system
Neoplasms	Diseases of the circolatory system
Diseases of the blood	Neoplasms
Endocrine, nutritional and metabolic diseases	Diseases of the circolatory system
Mental and behavioural disorders	Diseases of the circolatory system
Diseases of the nervous system	Diseases of the circolatory system
Diseases of the circolatory system	Diseases of the respiratory system
Diseases of the respiratory system	Diseases of the circolatory system
Diseases of the digestive system	Certain infectious and parasitic diseases
Diseases of the skin and subcutaneous tissue	Diseases of the circolatory system
Diseases of the muscoloskeletal system	Diseases of the nervous system
Diseases of the genitourinary system	Diseases of the circolatory system
Pregnancy, childbirth and the puerperium	-
Certain conditions originating in the perinatal period	-
Congenital malformations, deformations and chromosomal abnormalities	Diseases of the circolatory system
Ill-defined causes	Infectious diseases
External causes of mortality	Diseases of the circolatory system

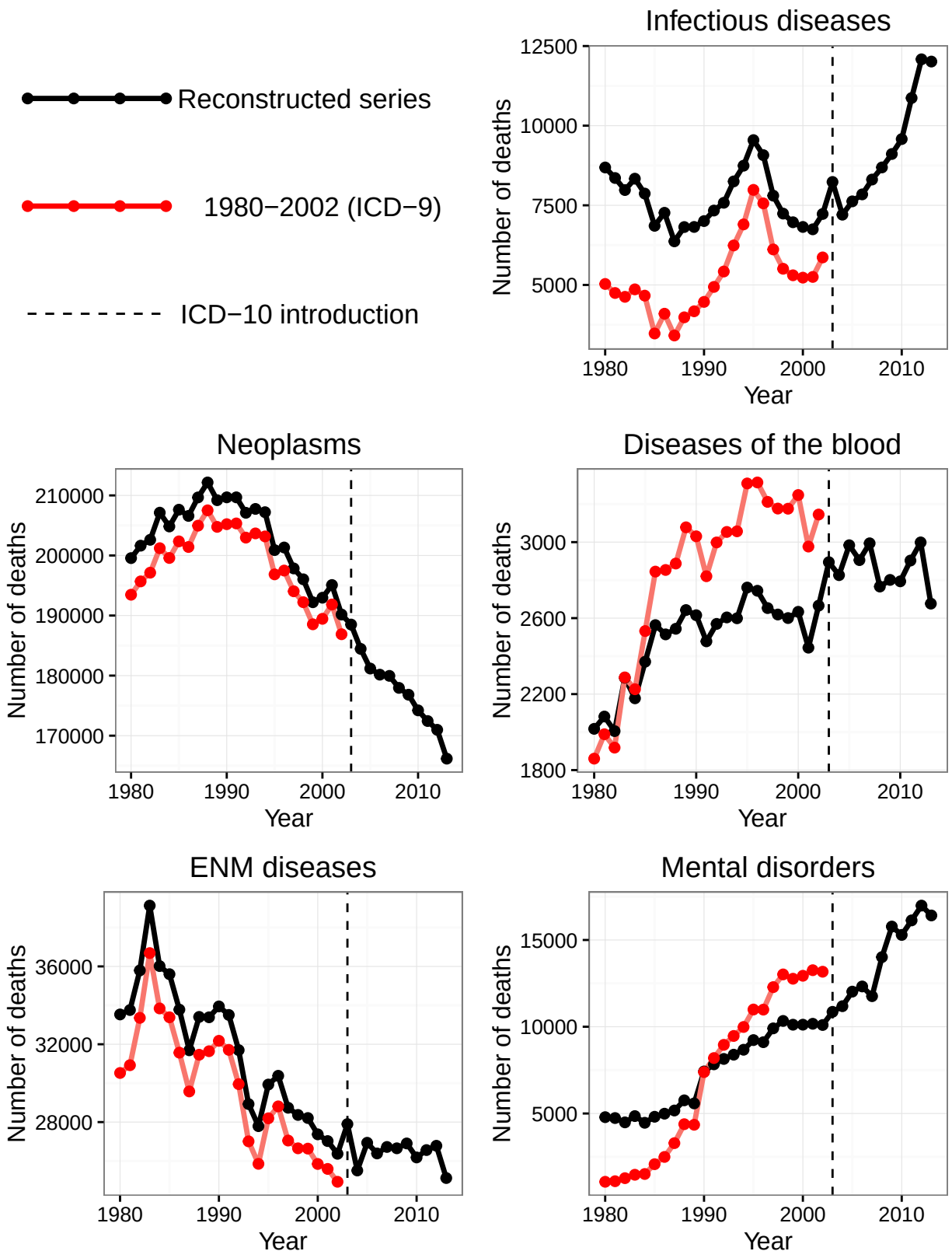
TABLE 4.2: Redistribution of deaths that don't stay in the same cause of death group passing from ICD-9 to ICD-10. For each of the 17 groups of causes it is shown the corresponding group where the majority of the deaths move to.

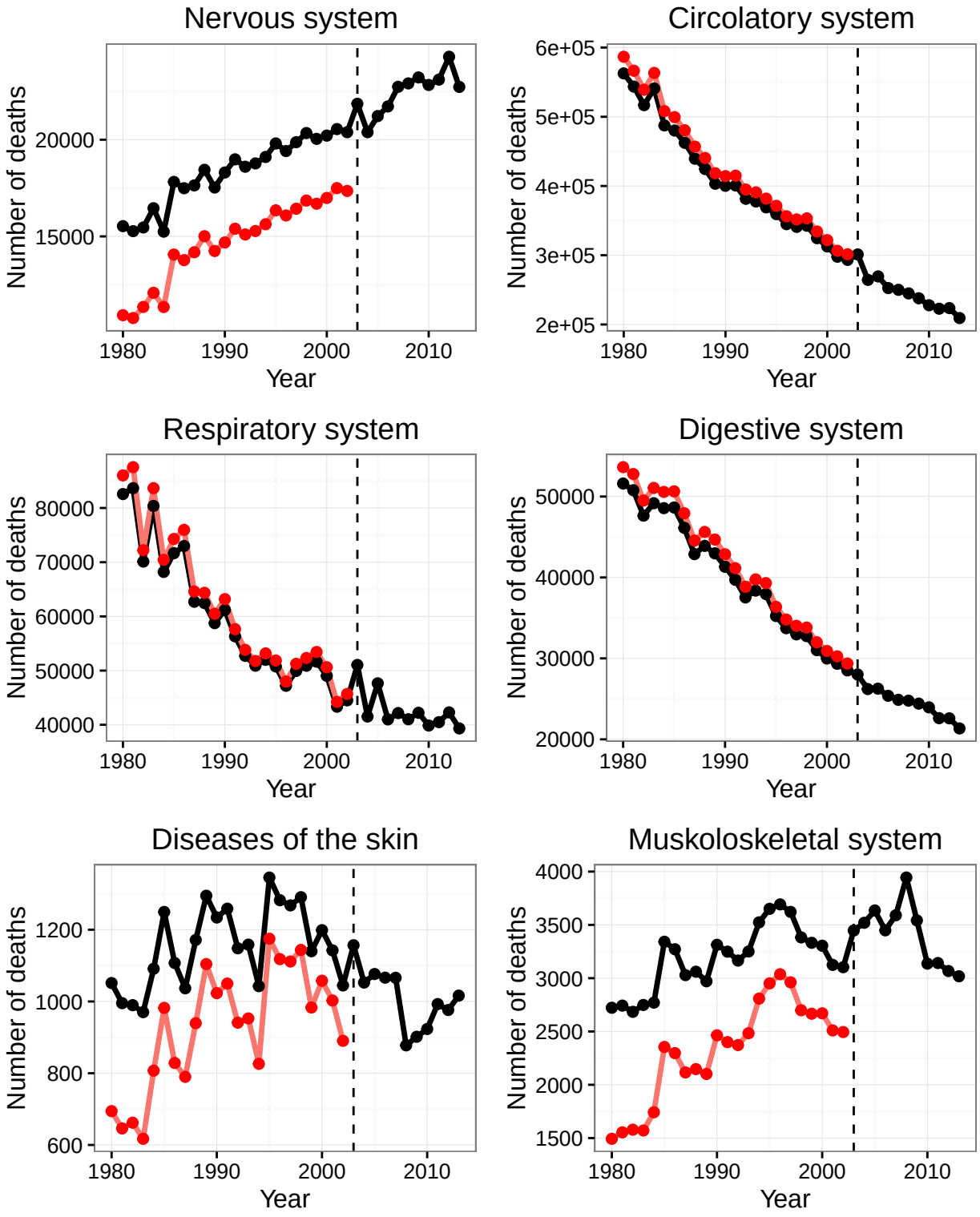
percentage of deaths remaining in the group is 100%, as seen in Table 4.1, they haven't a corresponding match in this type of analysis.

4.1.2 New coherent time-series

A more comprehensive view of how the reconstruction has changed the profile of each cause of death it is offered by Figure 4.1. It displays the standardized number of death, i.e. the expected number of deaths devoid of changes in the population age structure, that are computed as shown in Paragraph 3.2.3, for each of the 17 causes of death either before or after the reconstruction. Thus, the graphs allow visualizing how specific causes of death time-series have changed after redistributing the 1980-2002 deaths according to ICD-10. The results are shown at the level of total population since there aren't particular differences between the sexes.

The sign (whether the reconstruction has increased or decreased the number of deaths of a given disease) and the intensity of the changes largely vary cause by cause. Moreover, it is feasible, even if unlikely, that the sign could be reversed over the time-period covered, i.e. the reconstruction can increase the number of deaths in some years and decrease the number of deaths in some other years. This is possible because, although the same coefficients are applied for any years, the exchange between causes, in terms of number of deaths, it is not predetermined a-priori: cause x will always take the same percentage of deaths from cause y and not the same number.


Reconstructed series
1980−2002 (ICD−9)
ICD−10 introduction
Infectious diseases
Neoplasms
Diseases of the blood
ENM diseases
Mental disorders
Number of deaths
Year
1980
1990
2000
2010
12500
10000
7500
5000
210000
200000
190000
180000
170000
3000
2600
2200
1800
36000
32000
28000
15000


Nervous system
Number of deaths
20000
15000
1980
1990
2000
2010
Year
Circolatory system
Number of deaths
6e+05
5e+05
4e+05
3e+05
2e+05
1980
1990
2000
2010
Year
Respiratory system
Number of deaths
80000
70000
60000
50000
40000
1980
1990
2000
2010
Year
Digestive system
Number of deaths
50000
40000
30000
20000
1980
1990
2000
2010
Year
Diseases of the skin
Number of deaths
1200
1000
800
600
1980
1990
2000
2010
Year
Muskoloskeletal system
Number of deaths
4000
3500
3000
2500
2000
1500
1980
1990
2000
2010
Year

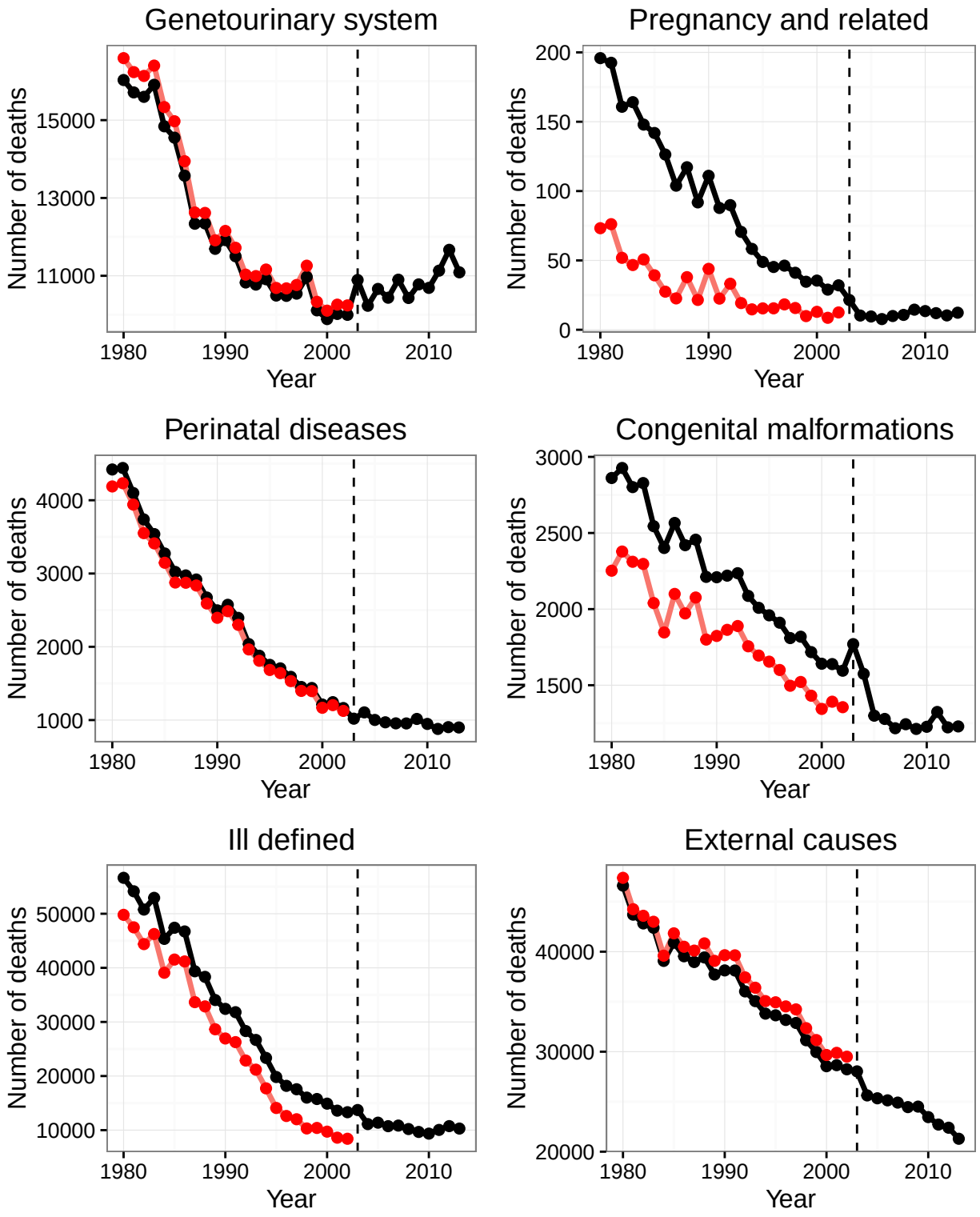


FIGURE 4.1: Reconstructed and original time-series of causes of death in Italy. Total population.

Finally, since the figures show the number of standardized deaths, they also permit an evaluation of the trends over time of each cause. For instance the number of standardized deaths due to diseases of the digestive system has constantly been declining while it is exactly the opposite for the diseases of the nervous system. This point, however, will not be touched in this Paragraph (but in Paragraph 4.2.1) since here the only aim is to show how the reconstruction has modified the number of deaths attributable to each cause of death.

The time-series of the three major causes of death, namely neoplasms, diseases of the circulatory system and diseases of the respiratory system, don't present a dramatic change after performing the reconstruction. This is especially true for the diseases of the respiratory system for which the red and black lines overlap almost perfectly, while deaths attributable to neoplasms according to ICD-10 are slightly greater than those of ICD-9 and the opposite occurs for the diseases of the circulatory system. A similar pattern of homogeneity between the two ICD revisions is also followed by diseases of the digestive system, diseases of the genitourinary system, perinatal diseases and, even if to a lesser extent, by endocrine, nutritional and metabolic diseases, ill-defined and external causes. For the rest of the causes of death considered in this dissertation, instead, the disruptions produced by the ICD-10 introduction are more relevant. In particular, the number of deaths due to infectious diseases, diseases of the nervous system, diseases of the skin, diseases of the muskoloskeletal system, pregnancy and related and congenital malformations are considerably underestimated by ICD-9 compared to the one of ICD-10. The situation is instead reversed for diseases of the blood, while mental and behavioral disorders present a mixed situation over time.

4.2 Cause-specific mortality in Italy from 1980 to 2013

In the previous section I have quantified and analyzed the statistical disruptions in the causes of death time-series due to the introduction of the new ICD classification and I have shown how the reconstruction has led then to new, and coherent, time-series for each of the 17 causes of death considered in this dissertation. Now, having in hand these new reconstructed series, I intend to give an overview of mortality, by causes of death, in Italy from 1980 to 2013 giving new insights on the state of the health transition.

Specifically, this section is divided into three parts: 1) Figure out how the percentage of deaths attributable to each cause of death has changed over time in this period; 2) Investigate on the leading cause of death by age and year for the two sexes using Lexis' surface; 3) Investigate on temporal dynamics of mortality rates over age and time by estimating surfaces of mortality improvements. The results will be helpful in understanding if mortality declined uniformly or not across causes of death, so revealing if there is only one pattern of mortality change or more. Definition of mortality improvements and the related methodology to carry out the analysis are reported in detail in Paragraph 4.2.3.

4.2.1 Density broken down by cause of death

The first overview of mortality in Italy is given by looking at the density function broken down by cause of death. Since the area of a density function has to sum up to 1, breaking it down by cause of death and summing up the area of the cause-specific densities, gets back the percentage of deaths attributable to each cause of death.

This view is displayed in Figure 4.2 and 4.3, in which life table densities in 1980-84 and 2010-13 are broken down by causes of death for men and women and in Table 4.3 and 4.4 which give support to the figure reporting exactly the percentage of deaths attributable to each of the 17 group of causes, in seven points of time, again for the two sexes separately.

What has to be primarily noted is the tendency of neoplasms and diseases of the circulatory system, the groups which collect the majority of deaths. In 1980-84 the 42.97% of deaths among males and the 52.74% among females were due diseases of the circulatory system. In the following decades the percentage of deaths attributed to this cause declined constantly and it was equal to 36.05% for males and 43.26% for females in 2010-13. On the other hand, neoplasms weight on total mortality increased by 6.46% for males (from 24.83% to 31.29%) and 3.97% for females (from 18.08% to 22.05%). So diseases of the circulatory system are still the cause provoking the majority of deaths but their difference with respect to neoplasms is reduced, especially among males. Indeed, despite the population aging and the consequent higher number of total deaths in 2010-13 than 1980-84, the number of deaths attributed to circulatory diseases in the population went down from 250'730 to 221'508.

The postponement of mortality and the consequent population aging reflect the rise of three groups of causes highly related with age: infectious diseases, mental and behavioral disorders and diseases of the nervous system. The first, which in the pre-transition era was the main cause of death during the childhood, now in the era of longevity extension, has become typical of old ages. In fact, at old ages, when death is frequently due to a mix of causes, infectious diseases have increased their lethality (Dorn and Moriyama, 1964; Desesquelles et al, 2012). The rise in the percentage of deaths attributable to mental and behavioral disorders is really noticeable: it is almost four times higher for males (from 0.55% to 2.12%) and even ten times higher for females (from 0.37% to 3.89%). The increase is less pronounced for the diseases of the nervous system, but not for that less important. This group of causes, which contains diseases highly correlated with age like Parkinson and Alzheimer, has augmented either for men (from 1.56% to 3.44%) or women (from 1.57% to 4.34%). External causes, which, as it will be possible to see in later Chapters, is a group of big interest in the context of lifespan variability, have decreased their percentages, visibly for males (from 5.27% to 4.19%) and slightly for females (from 3.54% to 3.20%). About the remaining groups of causes, to be noted is the doubling of diseases of the blood and diseases of the skin and the reduction of ill-defined causes and diseases of the digestive system.

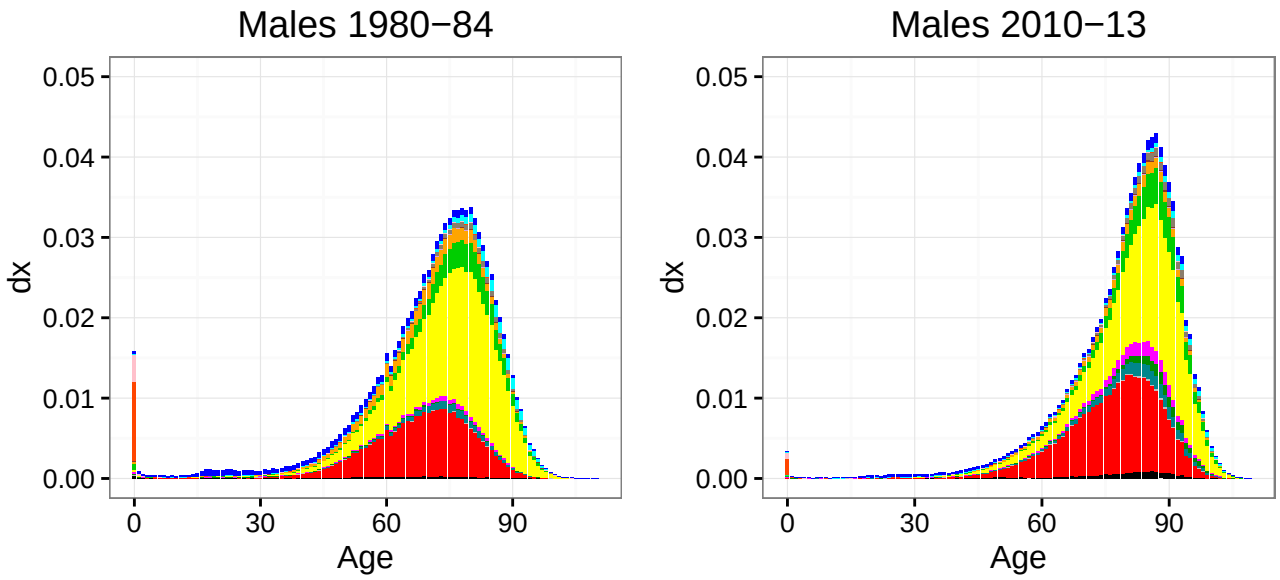


FIGURE 4.2: Males 1980-84 (left panel) and 2010-13 (right panel) life tables density function broken down by cause of death, Italy

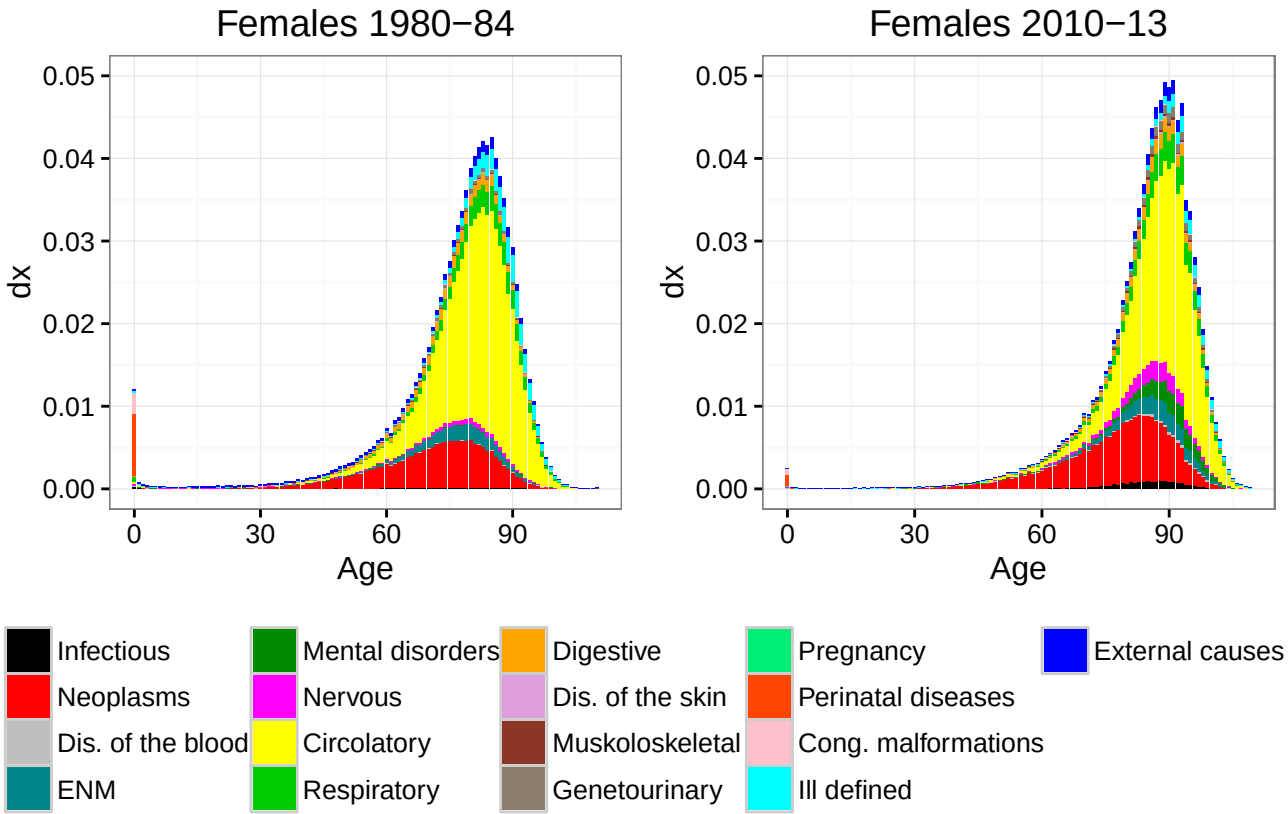


FIGURE 4.3: Females 1980-84 (left panel) and 2010-13 (right panel) life tables density function broken down by cause of death, Italy

Cause	1980-84	1985-89	1990-94	1995-99	2000-04	2005-09	2010-13
Certain infectious and parasitic diseases	0.98	0.89	1.21	1.23	1.05	1.33	1.79
Neoplasms	24.83	27.70	29.42	29.89	30.98	31.57	31.29
Diseases of the blood	0.21	0.27	0.29	0.33	0.35	0.40	0.42
Endocrine nutritional and metabolic diseases	2.51	2.60	2.73	2.98	3.13	3.62	3.91
Mental and behavioural disorders	0.55	0.61	0.91	1.09	1.18	1.57	2.12
Diseases of the nervous system	1.56	1.86	2.07	2.33	2.64	3.19	3.44
Diseases of the circulatory system	42.97	40.77	39.97	39.82	39.14	37.27	36.05
Diseases of the respiratory system	8.10	8.05	7.61	7.77	8.08	8.45	8.65
Diseases of the digestive system	6.14	5.70	5.04	4.44	4.13	3.87	3.71
Diseases of the skin and subcutaneous tissue	0.06	0.08	0.09	0.11	0.11	0.11	0.13
Diseases of the musculoskeletal system	0.24	0.28	0.31	0.34	0.33	0.35	0.33
Diseases of the genitourinary system	1.72	1.50	1.40	1.43	1.52	1.85	2.05
Pregnancy, childbirth and the puerperium	-	-	-	-	-	-	-
Certain conditions originating in the perinatal period	0.89	0.65	0.51	0.34	0.25	0.22	0.21
Congenital malformations, deformations and chromosomal abnormalities	0.54	0.47	0.42	0.36	0.32	0.21	0.22
Ill defined causes	3.04	3.09	2.55	1.94	1.74	1.50	1.53
External causes of mortality	5.27	5.03	4.99	4.70	4.54	4.44	4.19
Total	100	100	100	100	100	100	100

TABLE 4.3: Percentage of deaths attributable to each of the 17 groups of causes at different point in time: 1980-84, 1985-89, 1990-94, 1995-99, 2000-04, 2005-09, 2010-13. Males.

Cause	1980-84	1985-89	1990-94	1995-99	2000-04	2005-09	2010-13
Certain infectious and parasitic diseases	0.68	0.62	0.69	0.82	0.90	1.22	1.80
Neoplasms	18.08	19.77	20.96	21.08	21.92	22.12	22.05
Diseases of the blood	0.22	0.28	0.32	0.37	0.42	0.56	0.59
Endocrine nutritional and metabolic diseases	4.51	4.62	4.56	4.46	4.53	4.85	4.91
Mental and behavioural disorders	0.37	0.51	0.98	1.41	1.85	2.91	3.89
Diseases of the nervous system	1.57	2.02	2.39	2.84	3.33	4.01	4.34
Diseases of the circulatory system	52.74	50.44	49.66	49.33	48.16	45.56	43.26
Diseases of the respiratory system	5.97	5.76	5.38	5.60	5.91	5.95	6.30
Diseases of the digestive system	4.09	4.25	4.25	4.08	3.93	3.85	3.72
Diseases of the skin and subcutaneous tissue	0.12	0.16	0.18	0.22	0.23	0.23	0.24
Diseases of the musculoskeletal system	0.30	0.40	0.47	0.58	0.61	0.82	0.73
Diseases of the genitourinary system	1.22	1.24	1.27	1.36	1.51	1.76	1.99
Pregnancy, childbirth and the puerperium	0.03	0.02	0.02	0.01	0.01	0.01	0.01
Certain conditions originating in the perinatal period	0.71	0.54	0.41	0.30	0.22	0.19	0.17
Congenital malformations, deformations and chromosomal abnormalities	0.44	0.38	0.35	0.31	0.29	0.20	0.20
Ill defined causes	5.05	4.85	3.94	2.65	2.37	2.28	2.45
External causes of mortality	3.54	3.68	3.68	3.60	3.39	3.32	3.20
Total	100	100	100	100	100	100	100

TABLE 4.4: Percentage of deaths attributable to each of the 17 groups of causes at different point in time: 1980-84, 1985-89, 1990-94, 1995-99, 2000-04, 2005-09, 2010-13. Females.

4.2.2 Leading cause of death

Another point of view to look at cause-specific evolution of mortality is offered by the investigation on the leading cause of death (i.e. the cause of death which provokes the higher number of deaths) by age and over time. The results of this three-dimensional study are shown in Figures 4.4, adding 20 years of analysis to the one conducted by Barbi et al. and summarized in Paragraph 2.2 (Barbi et al., 2000).

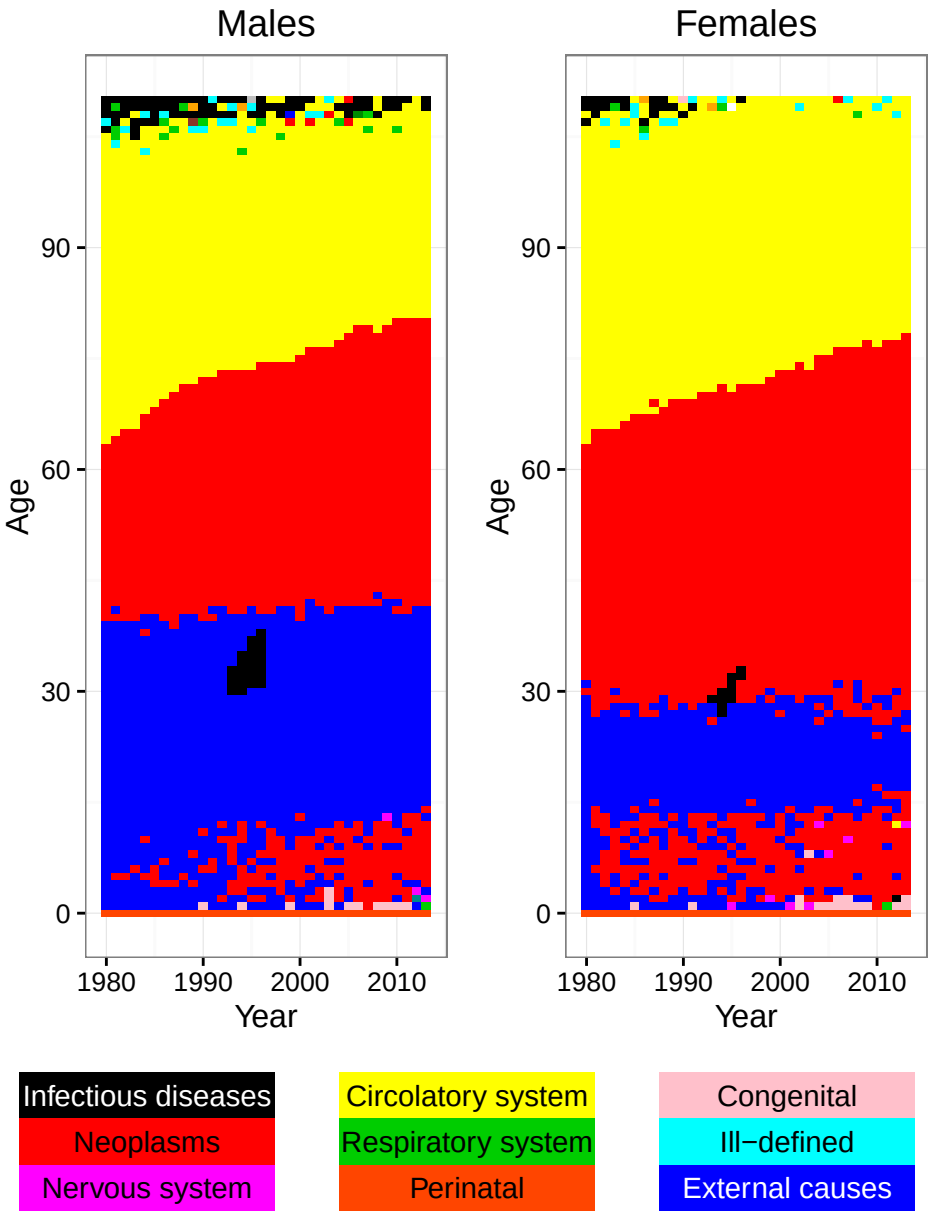


FIGURE 4.4: Leading cause of death from 1980 to 2013 for males (left) and females (right) using the list of 17 group of causes.

Reviewing almost a century (from 1895 to 1993), Barby et al. have shown how the passage to a more modern mortality regime is clearly visible also looking at the trend of the leading cause of death. What happened in the

next 20 years it is here analyzed. Males and females graphs look like similar at first sight, but some interesting distinctions emerge. At young ages external causes of death (which in this case means transport accident as we will see later) are much more predominant for men than women. Among males, external causes are the leading cause of death at any age between 15 and 35 over the whole period (with the only exception of infectious diseases from 1993 to 1996) while from 2 to 14 they are often reported as leading cause of death in the first decade (1980-89) to then leave space to the emergence of cancer. Among females, instead, external causes are the leading cause of death in a smaller range, basically between 18 and 28, and in some cases during the childhood. Between 28 and 35, whereas for men external causes are predominant, for women neoplasms are the leading cause of death. Consequently, the red central area, indicating neoplasms as leading cause, is wider for females. In 1980 cancers were the leading cause from 32 to 63 for women and from 39 to 63 for men. In the following decades, for both sexes, the lower boundary of this range has remained almost constant, while the upper boundary has raised noticeably (at the beginning at a faster pace for men) reaching age 80 for males and 78 for females. The area at the top of the figure, so the one representing very old ages, is characterized by the circulatory diseases as leading cause. The upper boundary just described for neoplasms represent the lower boundary of circulatory diseases. Consequently while in the 1980s this cause was the leading one also at young-old ages (the sixties) in more recent years is not the case anymore. Among centenarians, and especially for males, other two causes appear, infectious diseases and ill-defined. At extremely old ages death is due to a mixture of causes which act simultaneously and the two mentioned above seem to be the prevailing. A final note must be made for the black areas in the mid-1990s, around age 30, representing the emergence of AIDS' epidemic. Of course that area represents the whole group of infectious diseases and not only AIDS because using the list of 17 causes the level of detail is not so fine.

To have a deeper epidemiological understanding, Figure 4.5 displays the results obtained using the HCD intermediate list (104 causes). Of course, since the reconstruction has not been performed at such level of detail, only the period 2003-2013 can be analyzed.

If using the 17 causes' list male's and female's graph looked like pretty similar, now there are only some analogies and many differences. Until age 12 (males) and 15 (females), the situation is not uniform, with several causes, especially different kind of neoplasms (nervous system, haematopoietic system and other malignant neoplasms) and transport accident which alternate throughout the considered period. At young ages, the area that using the less detailed list pertained to external causes, it is now, not surprisingly, dominated by transport accident. Also in this case, the area is wider among males for whom transport accident is the leading cause until age 40, while for females until around age 30. To be noted is the emergence of suicide, starting from 2009, for men between age 30 and 40.

After the transport accident related deaths, males mortality is dominated by ischaemic heart disease between age 40 and 55 and then by neoplasms of bronchus and lung whose upper boundary has increased from 70 in 2003 to 76 in 2013, while the lower has remained almost constant. For females, those ages are instead uniformly characterized by only one cause,

neoplasm of breast. At old ages ischaemic heart disease return to be the leading cause of death among males, although at very old ages the pattern is more mixed as also other heart diseases and cerebrovascular disease emerge. This last cause is much more predominant for females, especially between age 75 and 95.

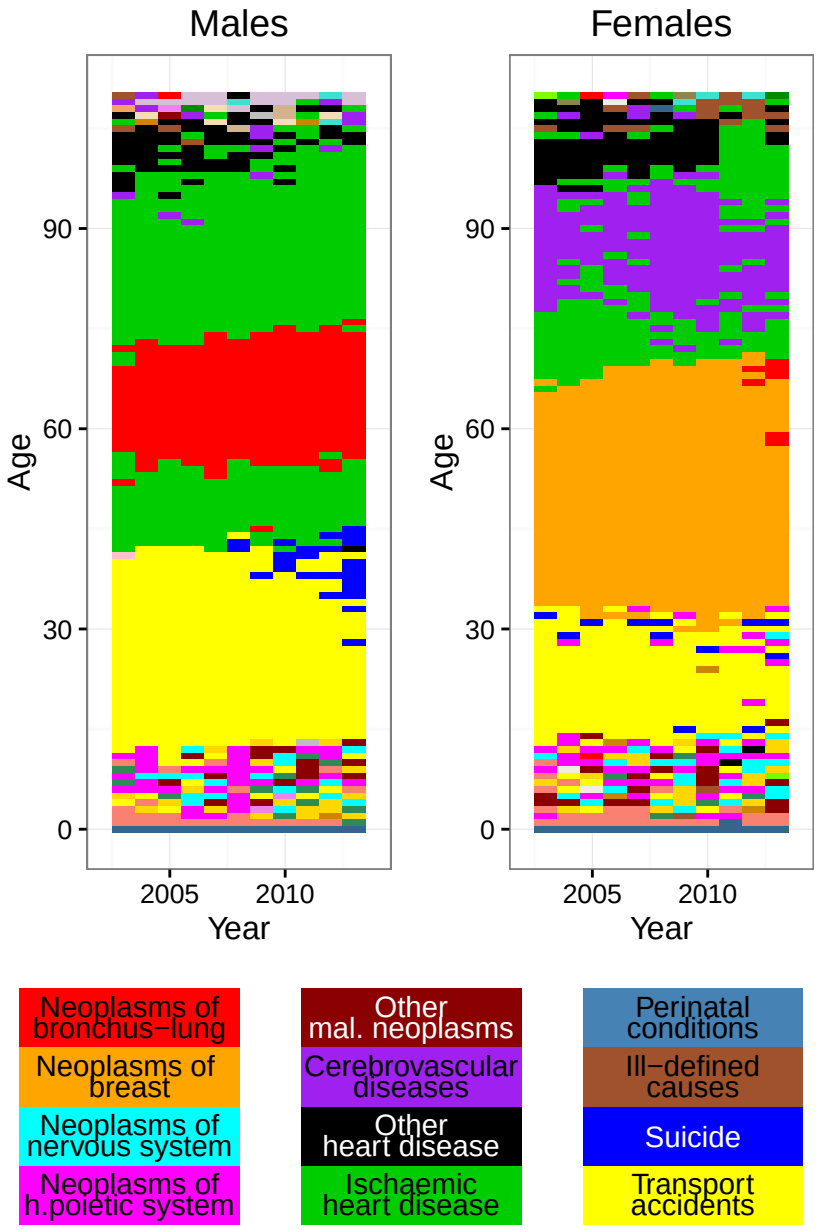


FIGURE 4.5: Leading cause of death from 2003 to 2013 for males (left) and females (right) using the HCD intermediate list of 104 causes.

4.2.3 Rates Of Mortality Improvement (ROMI)

The previous analysis were very useful for depicting Italian cause-of-death mortality but, to get a complete and informative portrait, the force of mortality must also be considered. Looking at the evolution of age and cause-specific death rates, $m_{x,k}$, enable to understand which pattern of decline mortality is following and if this pattern is similar or not among causes of death.

Here the investigation on mortality's rates dynamics over age and time is conducted by estimating surfaces of mortality improvements. For each cause *k*, mortality improvements are assessed through the Rate Of Mortality Improvements (ROMI) over age *x* and time *t*, $r_{x,k,t}$, which is computed as follow:

$$r_{x,k,t} = -(\frac{m_{x,k,t+1}}{m_{x,k,t}} - 1) \tag{4.1}$$

So, ROMI tells us, in which percentage, a given age-specific death rate has grown or declined with respect to the previous year and therefore it can be regarded as the time derivative of age-specific death rates. Equation 4.1 is a transformation of the equation to estimate the rate of growth and the minus sign ensures that reductions in mortality result in positive values (Andreev and Vaupel, 2005; Bokh and Rau, 2017; Keyfitz, 1977).

In my opinion, using ROMIs gives a deeper understanding of the age pattern of mortality change than using age-specific death rates themselves. To clarify this statement, Figure 4.6 displays on the top panel males and females death rates on the Lexis surface and on the bottom panel males and females ROMIs values on the Lexis surface as well. In the first case areas with the same color imply the same level of mortality (from black, the highest, to light green, the lowest) while in the latter case they imply the same level of mortality change (the various shades of blue indicate a worsening in mortality while the various shades of magenta indicate an improvement in mortality). Both pictures highlight that mortality conditions have been improving over the considered time period. In the top panels these improvements are pointed out by the fact that a given color reach higher ages as time passes by and in the bottom panels by the evident predominance of magenta areas with respect to blue areas. However, the intensity of the changes in mortality, and the specific ages at which they occur, are better visualized using ROMIs. For instance, the bad impact of AIDS epidemic between the 1980s and 1990s on mortality at young ages appear much more clearly through using ROMIs than simple death rates. Moreover, while the current level of death rates determines actual values of life expectancy and variability of age at death, it is the age-specific rate of change that determines their evolution into the future (Bokh and Rau, 2017; Kannisto et al., 1994; Vaupel, 1997). Therefore, using ROMIs, mortality dynamics and their consequences on the most important mortality indicators are better captured than using death rates.

ROMIs computation is based on smoothed death rates, obtained by the P-spline method (Currie at al., 2004; Eilers and Marx, 1996). In this case the smoothing procedure is more important than ever. Indeed, computing $r_{x,k,t}$ using the raw estimate of death rates, $m_{x,k,t}$, would lead to a confusing scenario. With regard to this issue, Figure 4.7 shows ROMIs

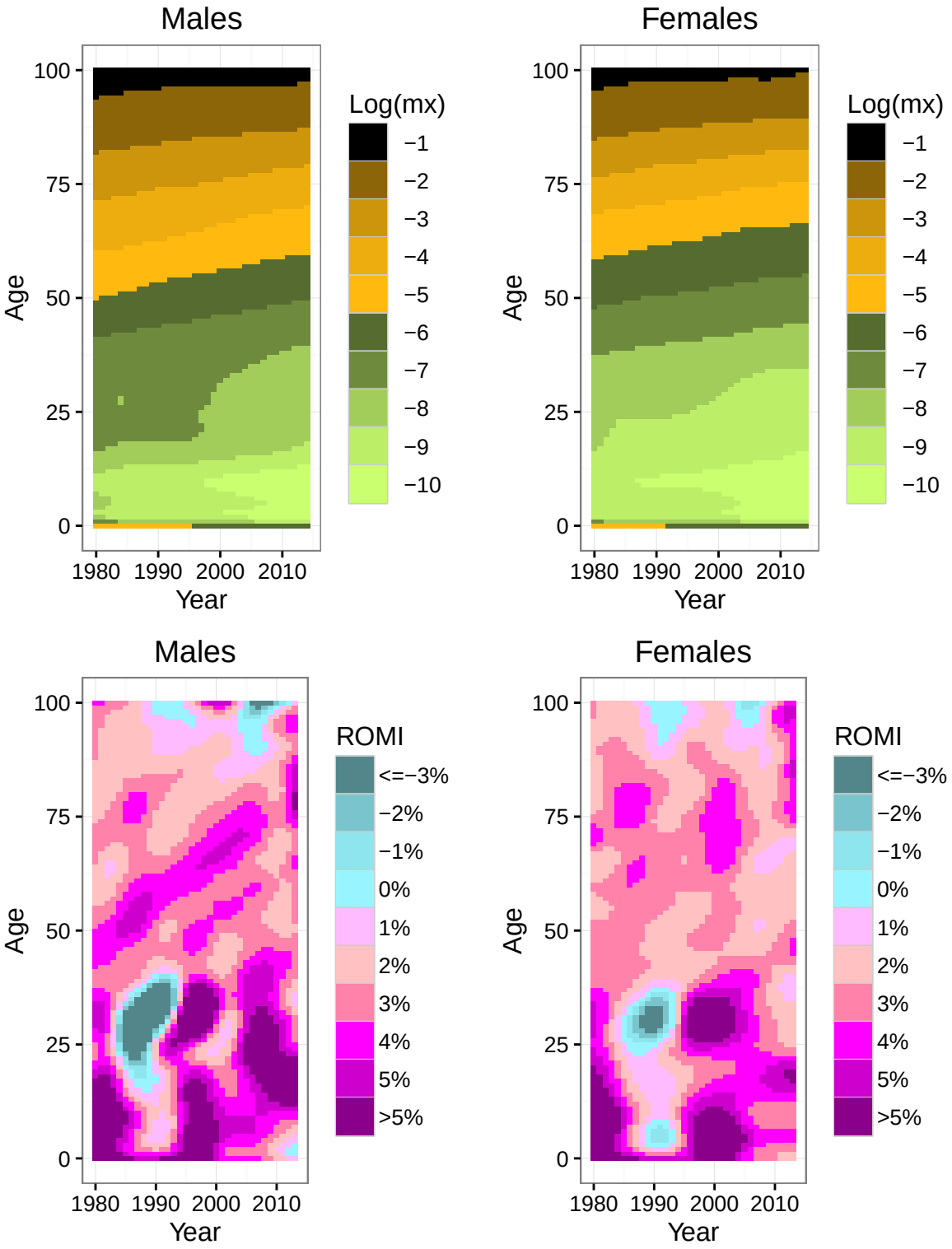


FIGURE 4.6: Top panel: smoothed death rates for males and females in Italy from 1980 to 2013. Bottom panel: ROMI based on smoothed death rates for males and females in Italy from 1980 to 2013.

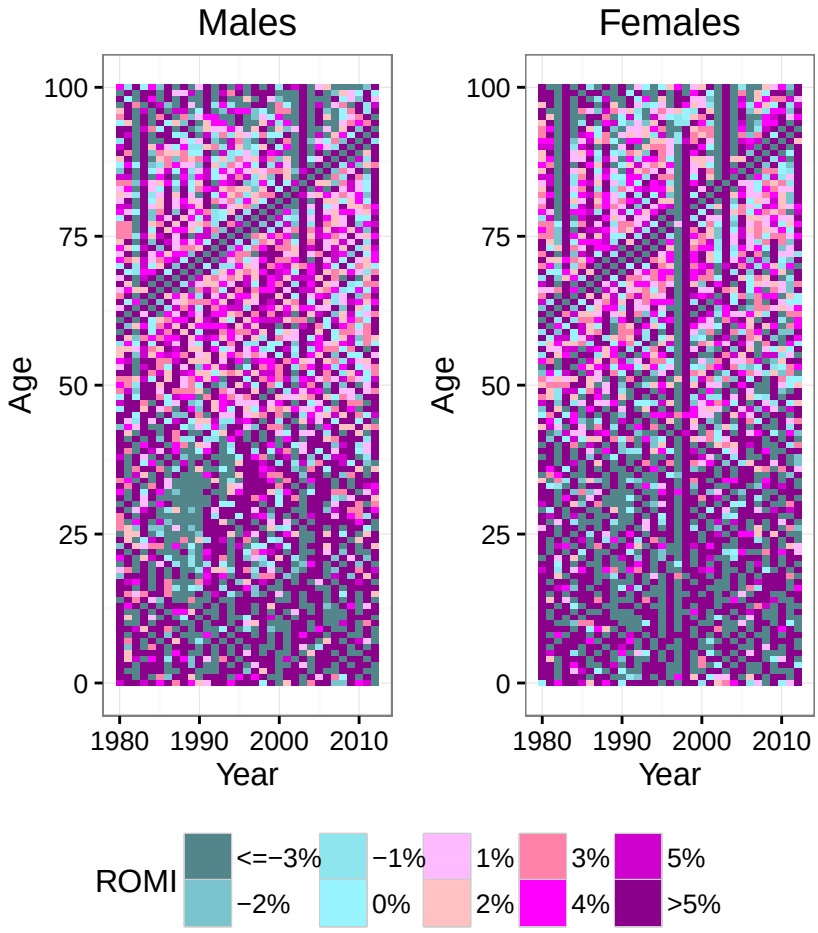


FIGURE 4.7: ROMI based on not smoothed death rates for males and females in Italy from 1980 to 2013

values obtained by using not smoothed death rates. A comparison with the bottom panel of Figure 4.6, whereas ROMIs values were obtained by smoothed death rates, clearly reveals the importance of the smoothing procedure. Indeed, due to the high variability of raw estimate of death rates, a clear understanding of the mortality improvements pattern is not perceptible. By smoothing $m_{x,k,t}$, the distortion caused by random noise processes is removed, leading to what can be regarded as the "true" Lexis surface of mortality improvements, as shown by bottom panels of Figure 4.6.

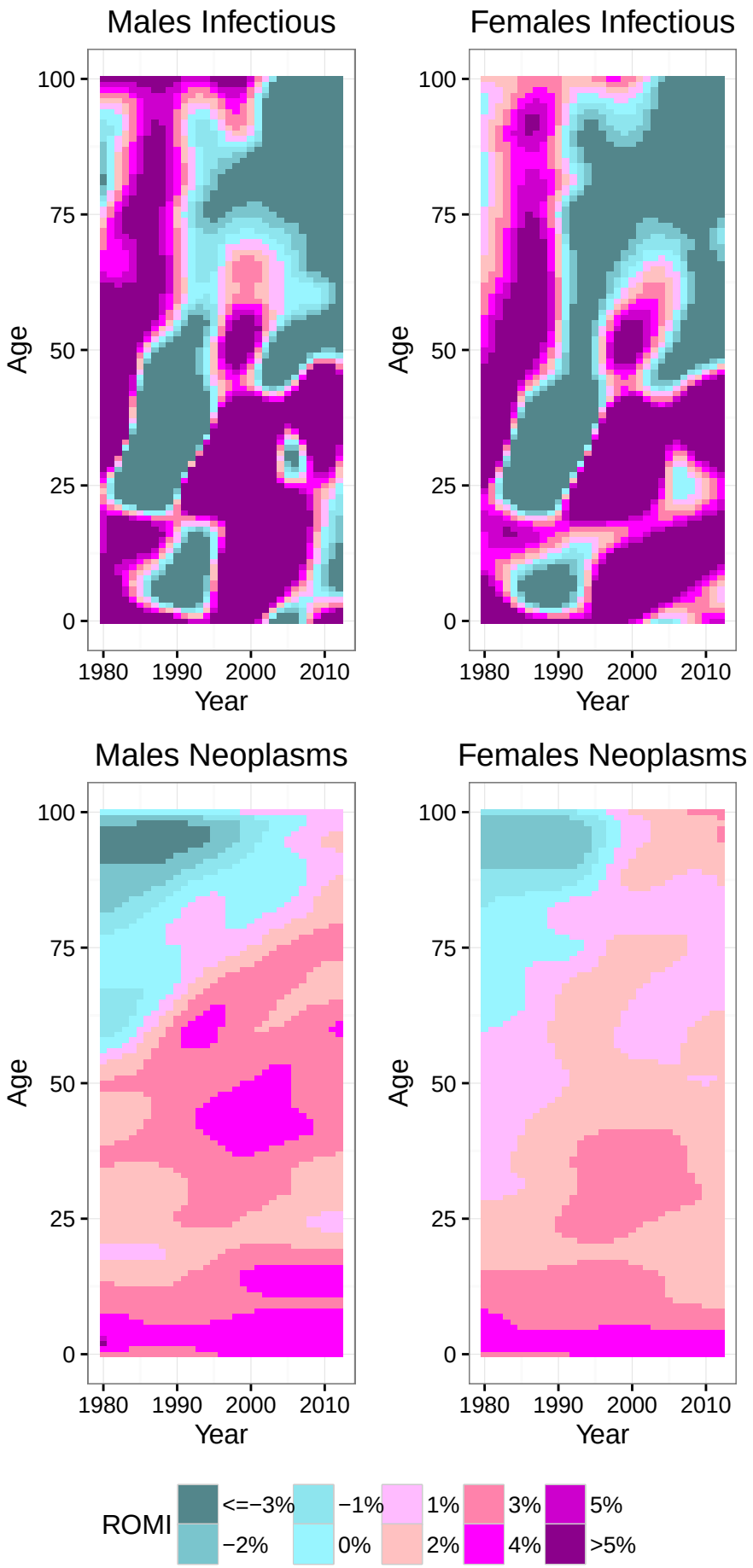
Cause-specific Lexis surfaces of mortality improvements are reported in Figure 4.8. In particular, the Figure displays results of infectious diseases, neoplasms, diseases of the nervous, circulatory, respiratory and digestive system and external causes of death. Remaining causes of death are reported in the Appendix. Before commenting cause-specific results, however, it is worth going back into ROMIs outcome for all-cause mortality.

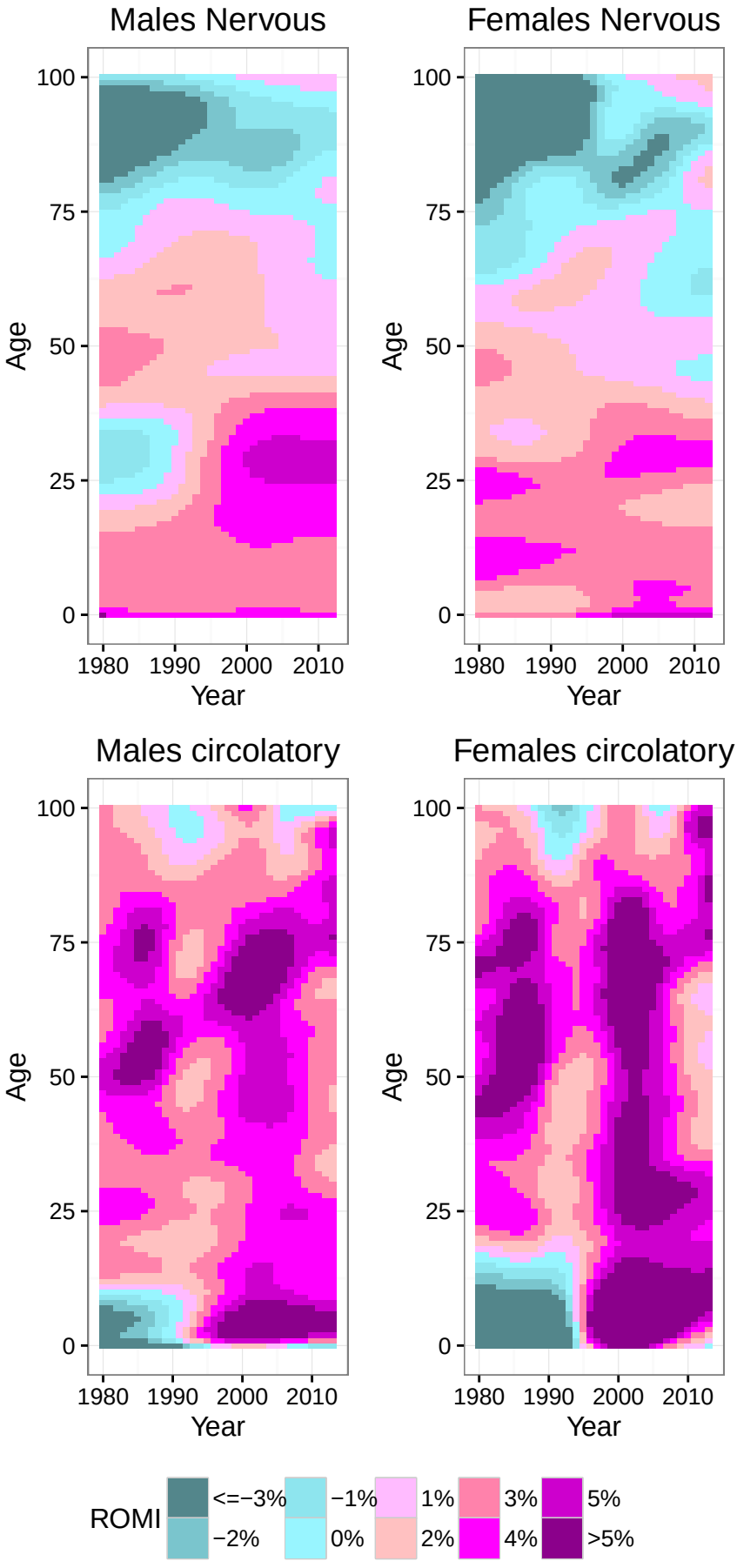
As already said, the various shades of blue indicate a worsening in mortality while the various shades of magenta indicate an improvement in mortality. The deeper is the color the stronger are rates of worsening or improvement. In particular the deepest blue is associated with an increase in death rates greater than 3% while the deepest magenta is associated to a decline in death rates greater than 5%. Generally speaking, both males and

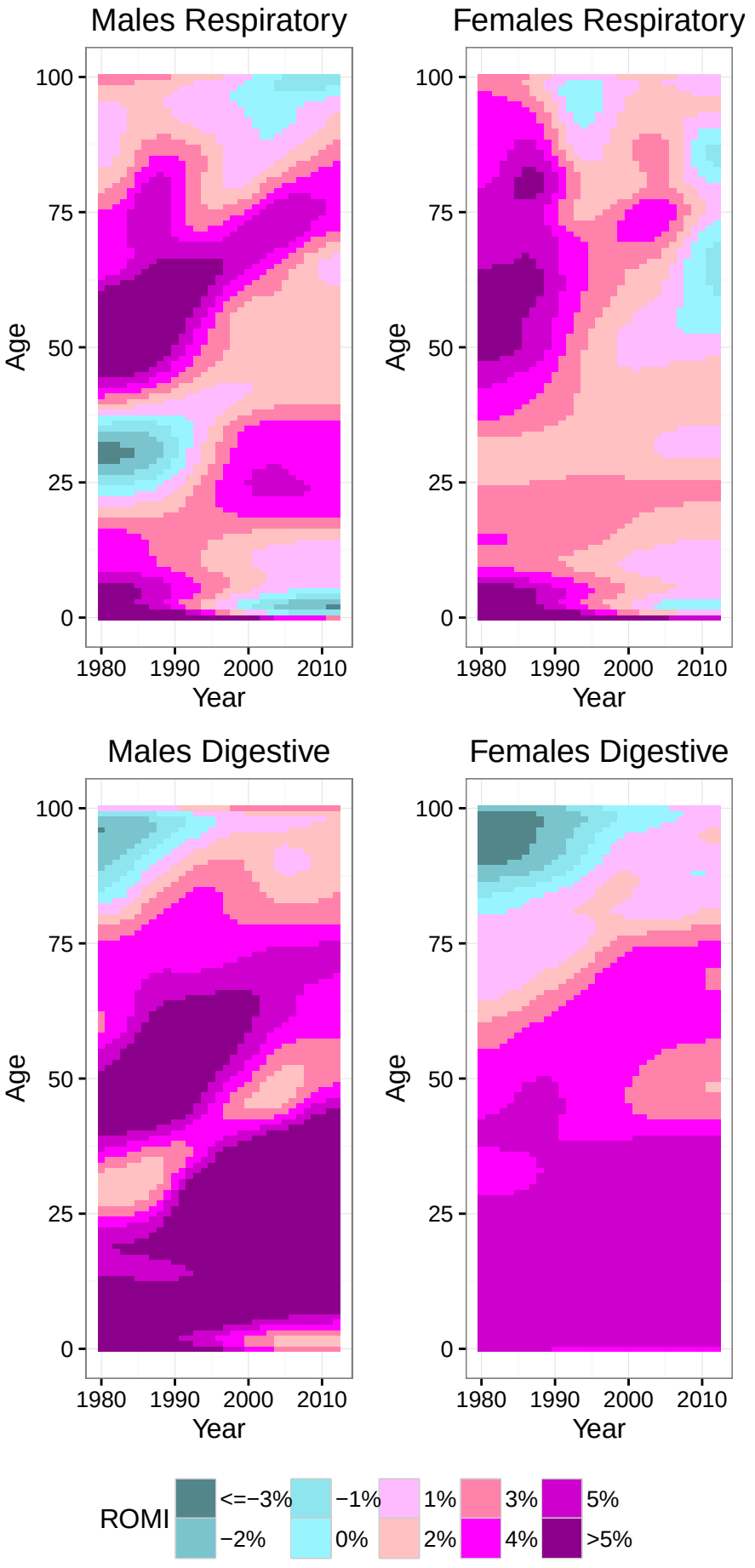
females mortality was decreasing at all ages and the highest rate of decline are registered at ages 0-35, while at adult and in particular old ages the decline was more contained (between 2% and 4%). A noticeable change of age-specific pattern of mortality occurred at young ages between the middle 1980s and early 1990s: a rapid increase in death rates took place both among males and females due to the well known AIDS epidemic. Mortality conditions got worse especially for males who experienced, in these years, a considerable rise in death rates from age 20 to 40. The epidemic's impact on females, even if really relevant, was more contained as already noted by previous studies (Geddes et al., 1994; Longo et al., 2008). From the middle 1990s, once the epidemic was over, death rates between age 20 and 40 started to decrease rapidly, recovering the disadvantage accumulated in the previous years. Another pattern to note concern the cohort effect in mortality improvements rates, more marked for males than females, of those in their forties in 1980. Indeed, there is an easily identifiable diagonal line, from age 40-50 in the 1980s to 70-80 in the 2000s, where the rates of improvement were about 4% and 5% for males and 3% and 4% for females. It would be possible to speculate whether these cohort effects show an exceptional mortality pattern of a particular group of cohorts or it is caused by period and age factors. However, this question will not be taken here.

Cause-specific results are rather heterogeneous: the age pattern of mortality change among causes varies considerably. Lets commenting first the outcome of the two major causes of death, neoplasms and diseases of the circulatory system, which, as pointed out in Paragraph 4.2.2, are responsible for about two-third of total deaths. In both males and females, mortality improvements are more marked for diseases of the circulatory system than for neoplasms. Indeed, a wide proportion of circulatory diseases Lexis' surface displays a rate of decrease in death rates of 4% or greater while for neoplasms the majority of the improvements are between 1% and 3%. In particular, circulatory's diseases rate of decrease has been very pronounced, especially for females, from the middle 1990s to the late 2000s and for the cohort in their forties and fifties in 1980. From the late 2000s, however, between age 35 and 70 a deceleration in the rate of mortality improvements is detected in both sexes. Such deceleration may be a sign of the economic crisis impact on health. Prevention, indeed, often plays a major role in avoiding diseases of the circulatory system related deaths and with the economic crisis it may be that, especially people belonging to lower class, have renounced to a proper prevention. With regard to neoplasms, the major decline in death rates has been registered at young ages (0-15) in both sexes and middle ages (40-60) for males. Another important difference between these two groups of causes concern ROMIs' values at old ages: while at those ages mortality has been going down for circulatory diseases all over the considered period (apart from few years in the middle 1990s at very old ages for females), the same cannot be said for neoplasms. Indeed, neoplasms' Lexis surface shows an increase in death rates at old ages (and also at young-adult ages in the 1980s) which stopped in middle 1990s for females and only in late 2000s for males. The rate of increase was especially pronounced at ages above 85 during the 1980s. This is particularly true for males who registered an increase greater than 3%.

Coming to minor causes of death, infectious' diseases outcome is particularly intriguing. As expected, AIDS's epidemic led to a deep worsening


Males Infectious
Females Infectious
Males Neoplasms
Females Neoplasms
Age
Year
0
25
50
75
100
1980
1990
2000
2010
ROMI
<=−3%
−2%
−1%
0%
1%
2%
3%
4%
5%
>5%


Males Nervous
Females Nervous
Males circolatory
Females circolatory
Age
Year
100
75
50
25
0
1980
1990
2000
2010
ROMI
<=−3%
−2%
−1%
0%
1%
2%
3%
4%
5%
>5%


Males Respiratory
Females Respiratory
Males Digestive
Females Digestive
Age
Year
100
75
50
25
0
1980
1990
2000
2010
ROMI
<=−3%
−2%
−1%
0%
1%
2%
3%
4%
5%
>5%

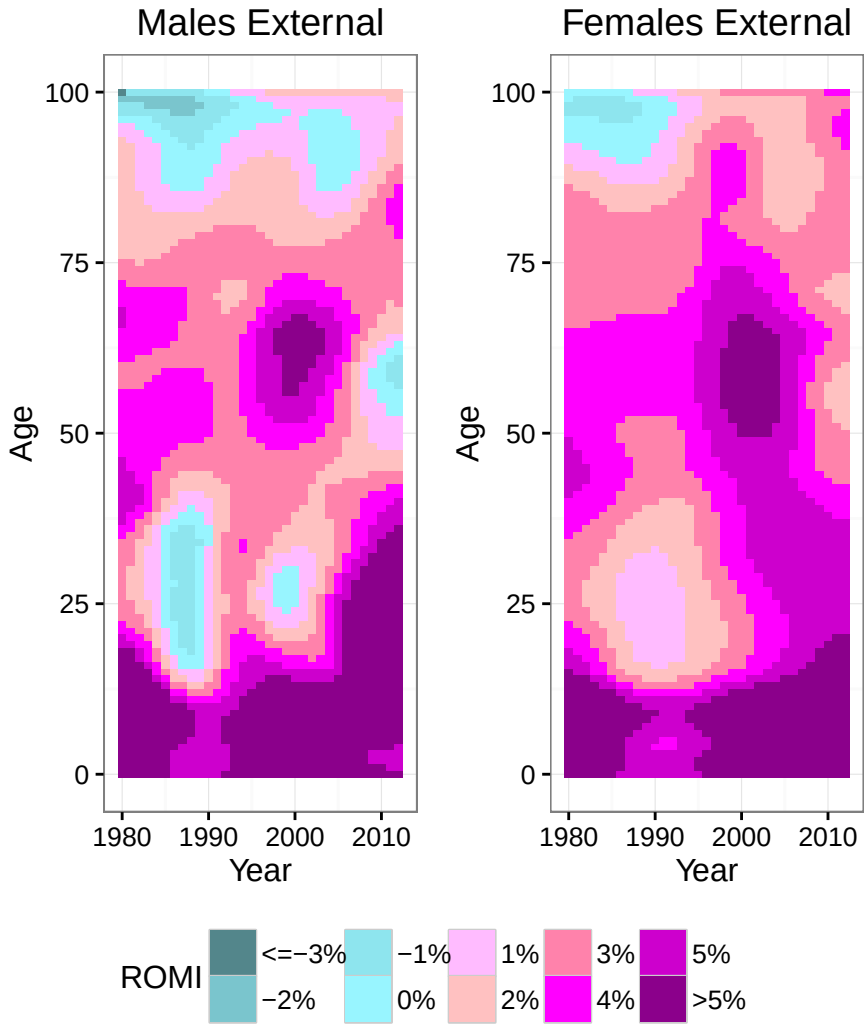


FIGURE 4.8: Cause-specific surface of mortality improvements based on smoothed death rates in Italy from 1980 to 2013.

of mortality at young and middle ages between 1980s and middle 1990s. The impact was so strong that, as seen before, it was clearly visible also at the all-cause mortality level. After the years of AIDS' epidemic, mortality started decreasing again at young and middle ages and the rate of such decline has been really considerable, almost always greater than 5%. At old ages, for both sexes, infectious diseases mortality change pattern may be divided into two distinct phases: until early 1990s, death rates declined considerably while after early 1990s the situation is completely reversed since death rates started increasing at a pace often greater than 3%. This result is in line with what found in Paragraph 4.2.2 about the rise in the percentage of deaths due to infectious diseases.

An increase in old-age death rates is also found for diseases of the nervous system for both males and females. Again, this confirms the results found in Paragraph 4.2.2 about the rise in the percentage of deaths due to diseases of the nervous system. The outcome might be related to the postponement of mortality and the emergence of senescence-related diseases.

The group of diseases of the nervous system, indeed, include several diseases (e.g. Parkinson and Alzheimer) highly correlated with age and very typical of the era of longevity extension.

ROMIs values for diseases of the respiratory system show a substantial (often greater than 3%) and generalized (in terms of ages) rate of decreases for both males and females. As in the case of diseases of the circulatory system is possible to note a cohort effect for those in their fifties and sixties in 1980 whose rate of decrease was particularly high.

Also diseases of digestive system display a huge decrease in death rates, apart from an increase at old ages in the 1980s for males and the 1980s and 1990s for females. The decline has been particularly strong for males, especially at young and adult ages.

It is finally worth noting the considerable rate of decrease of external causes of death at young ages. Therefore, despite transport accidents are still the leading cause of death at young ages, noticeable improvement has been made. As we will see later (Paragraph 5.3.1) such decline played a major role in reducing lifespan variability, especially for males, over the considered period.

Chapter 5

Understanding gender differences in lifespan variability through causes of death

Reconstruction of coherent cause of death time-series, presented in Chapter 4, is the starting point for the epidemiological analysis of lifespan variability in Italy from 1980 to 2013 which represents the main goal of this dissertation.

This Chapter has four main aims: 1) Figure out gender and cause-specific differences in the distribution of lifespans 2) Understand how the age pattern of mortality change interact with S_{10} and its evolution over time 3) Determine, for each sex, which causes of death are producing the observed compression of mortality in the period under study 4) Establish which causes of death are responsible for the variability gender gap in favor of females.

The first goal is achieved through the computation of all-cause and cause-specific mortality trends in *M* and S_{10} for each sex. I anticipate that neoplasms (and congenital malformations) are the only group of causes which report a higher variability among females than males. For this reason, and also for the great importance of this group of causes, a whole paragraph will be dedicated to a focus on neoplasms. Conducting a detailed analysis at the cause of death level in the period 2003-2013, it will be possible to figure out if the greater variability registered by females is attributable to some specific types of neoplasms or if it is more generalized.

Sensitivity analysis is then exploited to get insight about aim 2. As stated in Paragraph 3.2.5, sensitivity analysis allows understanding which ages, after a proportional reduction of death rates, would produce a decrease or an increase in S_{10} and the intensity of this potential decrease or increase. Moreover, it allows identifying a threshold age T_{10}, that is the age at which the contribution to a change in S_{10}, due to a reduction in mortality, is reversed from negative to positive. Thanks to the information obtained through sensitivity analysis, it will be then possible to break down S_{10} trends into the effects of mortality change below and above the previously determined young-old threshold, understanding how these two contrasting force have eventually produced the observed lifespan variability.

Finally, decomposition methods are used to reach goals 3 and 4. The results will quantify:

a) the contribution of changes in age and cause-specific mortality rates,

$m_{x,k}$, from 1980 to 2013 to change in S_{10} over the same period. The outcomes allow to exactly identify, for each sex, the age and specific-cause profile determining changes in variability of age-at-death in the period under analysis.

b) the contribution of differences in age and cause-specific mortality rates, $m_{x,k}$, between sexes to the difference in S_{10} between sexes at fixed point in time (1980 and 2013). In this case the results give a cause-of-death explanation of females advantage in lifespan variability.

5.1 Changing in position and shape of the age-at-death distribution: *M* and $\mathbf{S}_{10}$

Since mortality transition started, in any developed country the age-at-death distribution has greatly changed its position and shape. The improvements in mortality conditions, indeed, led to either a shift on the right or a compression of the distribution.

Figure 5.1 illustrates how the age-at-death distribution has evolved from 1906 to 2013 for the Italian population.

In this section, the shift, or in other words longevity extension, is summarized by the change in the modal age at death *M*, while the compression by the standard deviation of age at death above age ten S_{10}.

As stated in Chapter 2, evidence suggests that life expectancy at birth underestimate progress made in longevity extension when principally driven by improvements in the survival of the elderly and that life expectancy at some middle or early old ages, such as 50 or 65, strongly depend on an arbitrary selection of the age limit, and hence on an arbitrary definition of 'old' (Diaconu, 2016; Canudas-Romo, 2010; Kannisto, 2001). Thus in the era where the extension of human life is primarily due to the reduction of old-age mortality, the adoption of *M*, which is determined by adult mortality only, as a longevity indicator, seems to be preferable.

With regard to the standard deviation of age at death above age ten S_{10}, its choice was largely described in Paragraph 3.2.1 and principally related to the advantage of truncation at age 10.

5.1.1 All-cause long time trends in *M* and $\mathbf{S}_{10}$

As just seen thanks to Figure 5.1, mortality changes in Italy in the last century were impressive. Thus, before going to the analysis of cause-specific age-at-death distributions, I briefly summarize long-term (1906-2013) changes occurred in the Italian age-at-death distribution with regard to all-cause mortality. The estimation of *M* and S_{10} is based on smoothed density resulting from P-Splines, whose advantages have been mentioned in Chapter 3.

Figure 5.2 and 5.3 display, respectively, *M* and S_{10} values for males and females over the course of more than a century. About the modal age at death, *M*, has to be noted that, during the years of World War I and II, the *adult* modal age at death is reported. In these periods, indeed, the mode of the distribution was around 20 years. Since here we are interested in the shift on the right of the old age bell-shaped part of d_x, only the *adult* modal age at death is taken into consideration. On the left panels are displayed trends from 1906 to 2013 while on the right panels only the trends from

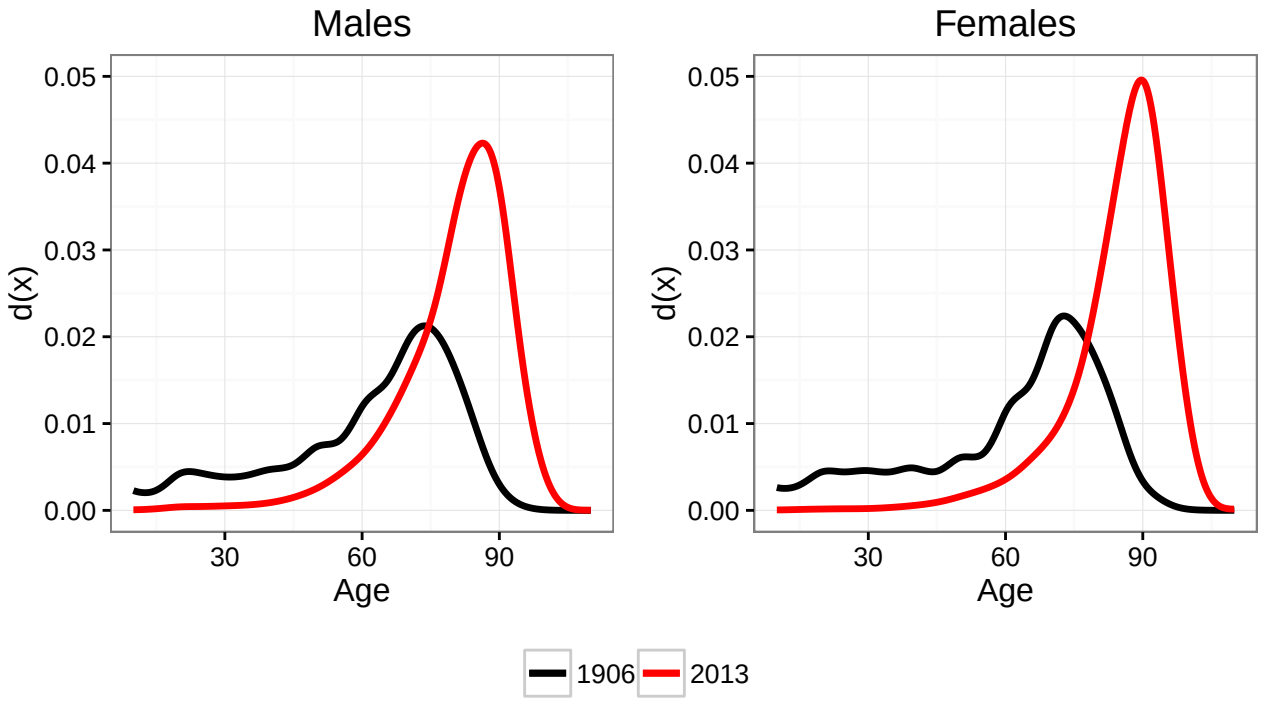


FIGURE 5.1: Smoothed d_x function for males and females in 1906 and 2013.

1980, to get a better visual inspection of the evolution of the two measures in the period for which cause-of-death data are available.

From 1906 to 2013 the mode has risen by 12.07 years for males (73.82 - 85.89) and by 16.23 years for females (72.93 - 89.30). Until the end of World War II the two trends almost overlap each other but then, from the 1950s, they started to separate. While females mode continued its linear rise at a pace of about 1.6 years per decade, the males mode reached a plateau for more than 20 years. It is during this period that the gender gap in the mode was generated. It would have been interesting, having cause-specific data for this period, to give an epidemiological interpretation of that fact. The most reliable hypothesis is that smoking-related diseases had a major role in generating the gap. From the early 1980s, however, males modal age at death started increasing again. It rose at a noticeable pace of about 2.5 years in the 1980s decade, to then slow down its increase from the 1990s. Trends in M for both sexes keep increasing in recent years, demonstrating how progress in longevity extension is still going on.

The great shift of the age-at-death distribution just described was also accompanied by a compression of mortality. S_{10} decreased by 6.43 years for males (19.10 - 12.67) and by 8.28 years for females (19.79 - 11.51). As for the mode, lifespan variability in the first four decades of the XXth century was almost equal between the sexes with females even showing a slightly higher variability in age-at-death. Only in the second half of the XXth century the two trends started to separate, from that moment S_{10} females values have been constantly lower than males values. A special mention has to be done for the two Wolrd War periods. The wars clearly had a worse impact on males than females about uncertainty about the timing of death. This is especially true in the case of World War II, with lifespan variability rising

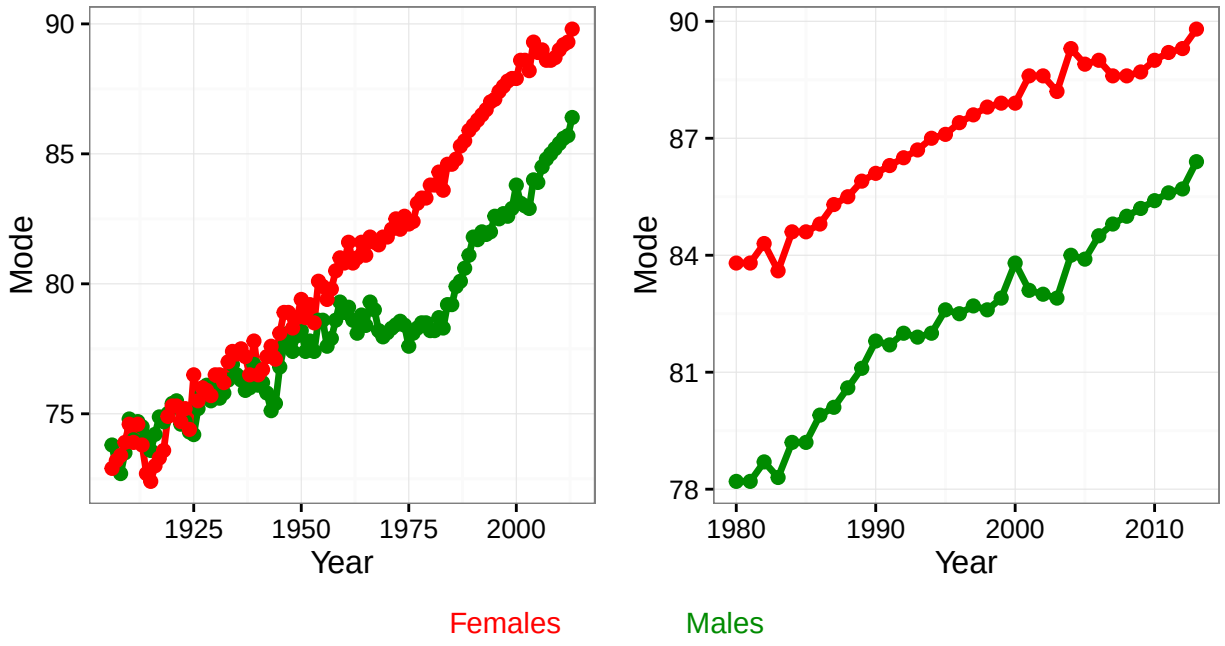


FIGURE 5.2: Modal age at death in Italy, 1906-2013 (left panel), 1980-2013 (right panel).

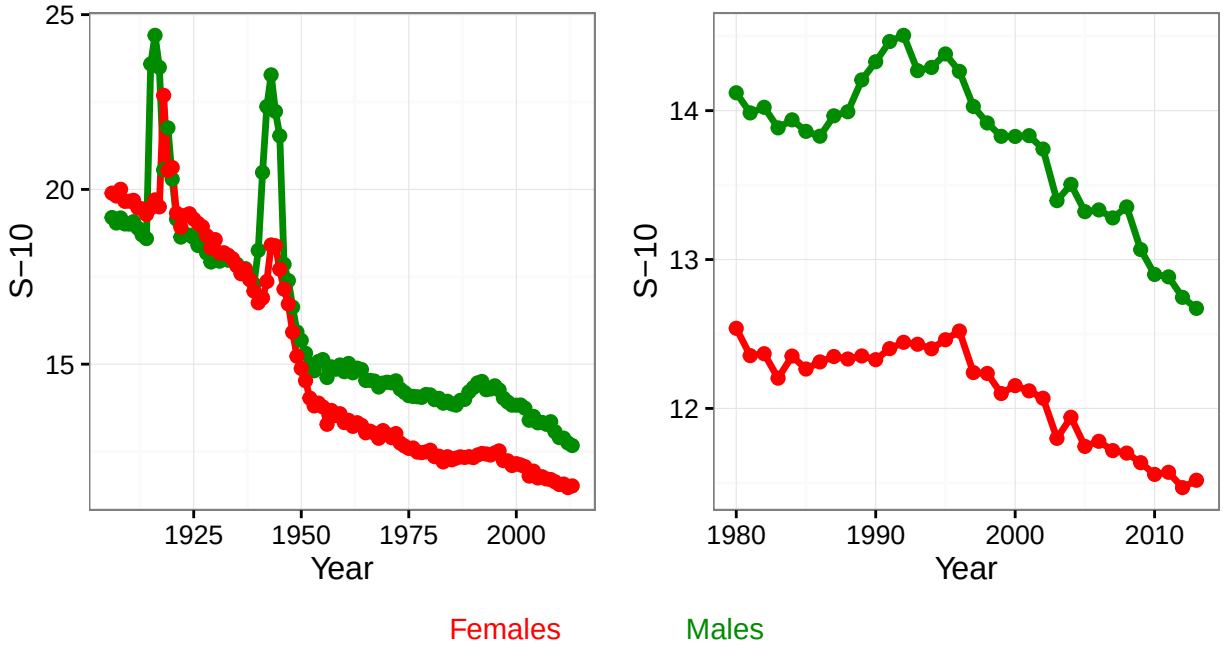


FIGURE 5.3: Standard deviation of age at death above age ten in Italy, 1906-2013 (left panel), 1980-2013 (right panel).

tremendously for men an only in a limited manner for females. During the World War I period, instead, the difference in the rose of S_{10} between the sexes was less accentuated. These, indeed, are also the years of Spanish Flu which, on the contrary of the war, indiscriminately hit both sexes.

The hypothesis of a shifting mortality scenario does not fit the Italian case yet. Even if at a lower pace mortality compression is still occurring and lifespans are becoming more and more concentrated in a narrower age interval. Only during the 1990s, especially for males, S_{10} slightly increased for a few years (an increase mainly due to AIDS as we will see soon). The gender gap reached its peak in 1992 when the difference between the sexes was of 2.06 years. From that moment males have been partially recovered the disadvantage and nowadays the difference is only of 1.16 years.

In the next paragraphs, thanks to data on causes of death, it will be possible to give an epidemiological interpretation of the all-cause mortality compression, as measured by S_{10}, in the period from 1980 to 2013.

5.1.2 Cause-of-death trends in *M* and $\mathbf{S}_{10}$

What basically emerged from the previous section is that in Italy longevity extension is still going on (the modal age at death is rising) and that, despite a slow down in the pace of compression, the duration of life is becoming less and less uncertain. But, how this two processes of longevity extension and compression differ from cause of death to cause of death?

Progress in cause-specific longevity extension is displayed in Figure 5.4. First, cause-specific modal age-at-death estimates differ greatly in terms of level. In both sexes, the lower limit is represented by neoplasms which report the lowest *M* and the upper limit by ill-defined causes which report the highest *M* while other causes are concentrated in a range of about 5 years. Second, although this variation in terms of level, all causes followed a steady upward trend since 1980. Moreover the increase in cause-specific modal age at death estimates occurred at a strikingly similar pace for most causes (a pace greater for males than females reflecting the all-cause *M* evolution seen in the previous paragraph). Among males the average increase was 9 years, mental disorders registered the greater rise (12.9 years) and ill-defined causes the lesser rise (3.9 years). Among females, instead, *M* increased 7 years on average and diseases of the blood registered the major improvements (10.2 years) while ill-defined causes the smaller (4.5 years). Third, the higher all-cause *M* registered among females reflects also into a general higher cause-specific *M*. For each cause the modal age at death is indeed higher, all over the considered period, for females than males.

Going into the cause of death detail, as previously said, neoplasms are the cause-of-death with smaller *M*. Thus, their age-at-death distribution is the one that reaches the peak earlier than any other, highlighting how this group of causes is not really typical of extremely old-ages. On the other hand, it is not surprising to find ill-defined causes as the one with higher modal age at death since, at extremely old-ages, death is often due to several factors and no specific cause can be reported as underlying. The remaining causes are quite close each other in terms of *M*. Among them, diseases of the circulatory and respiratory system, external causes and diseases of the skin are the one that reached major progress in longevity extension, while infectious diseases, diseases of the digestive system and endocrine,

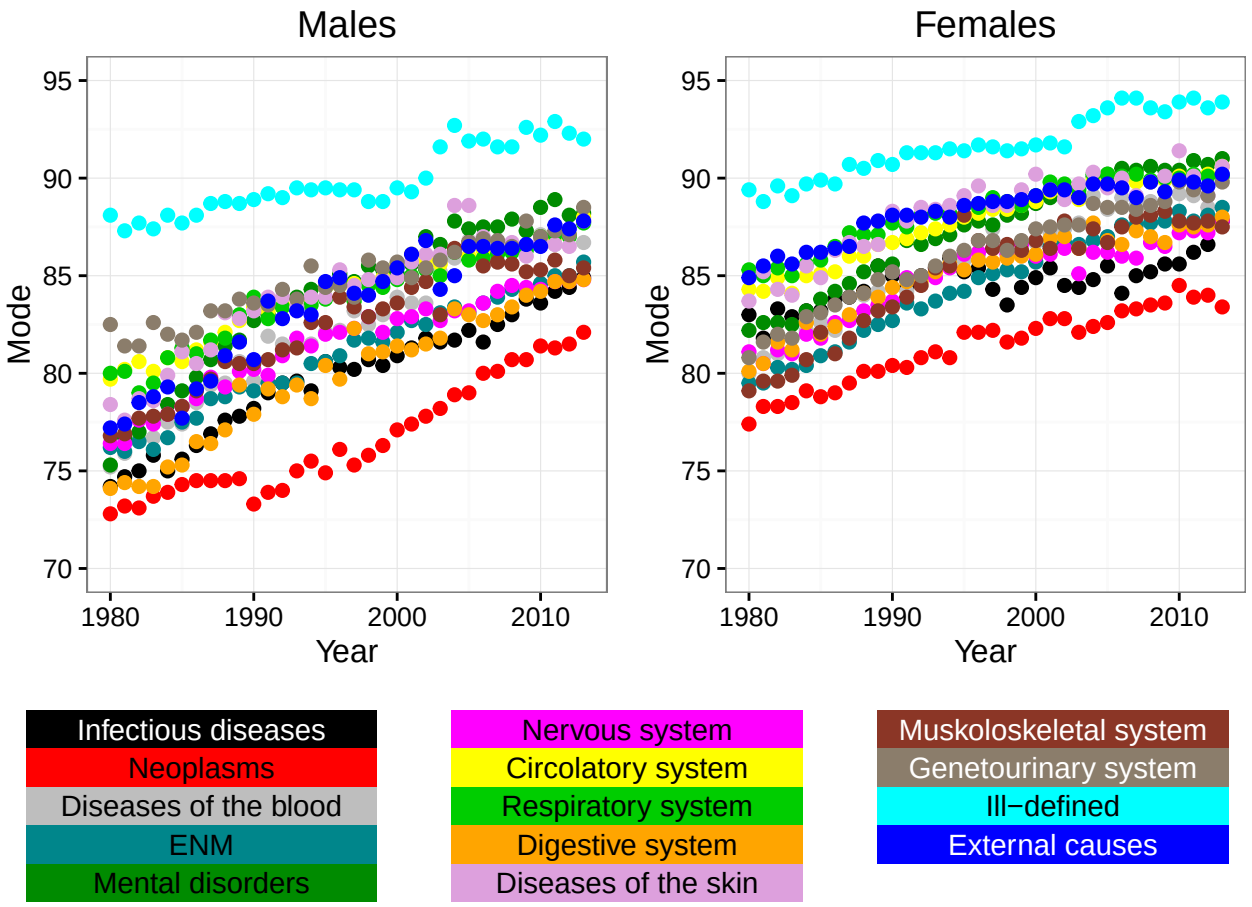
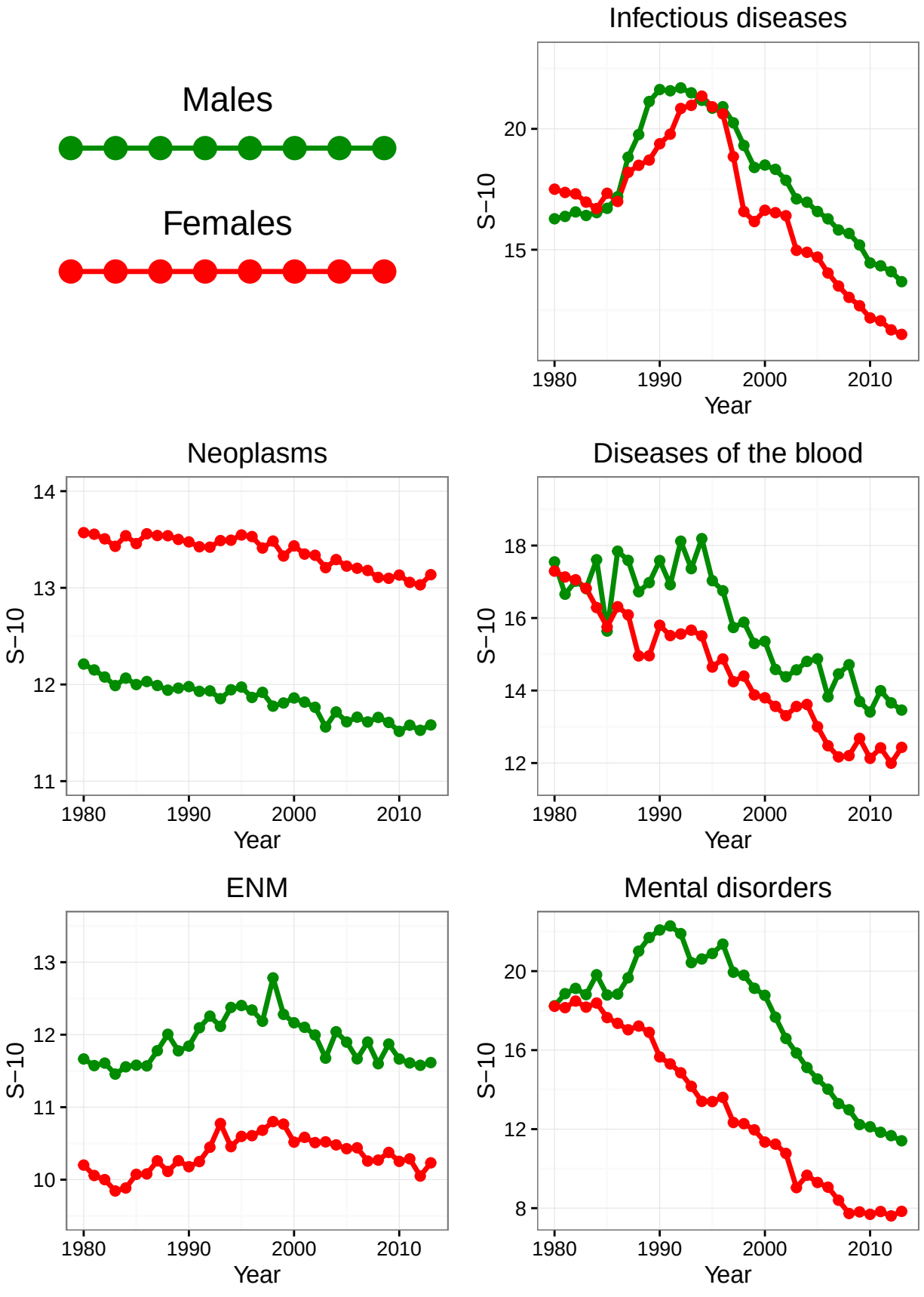


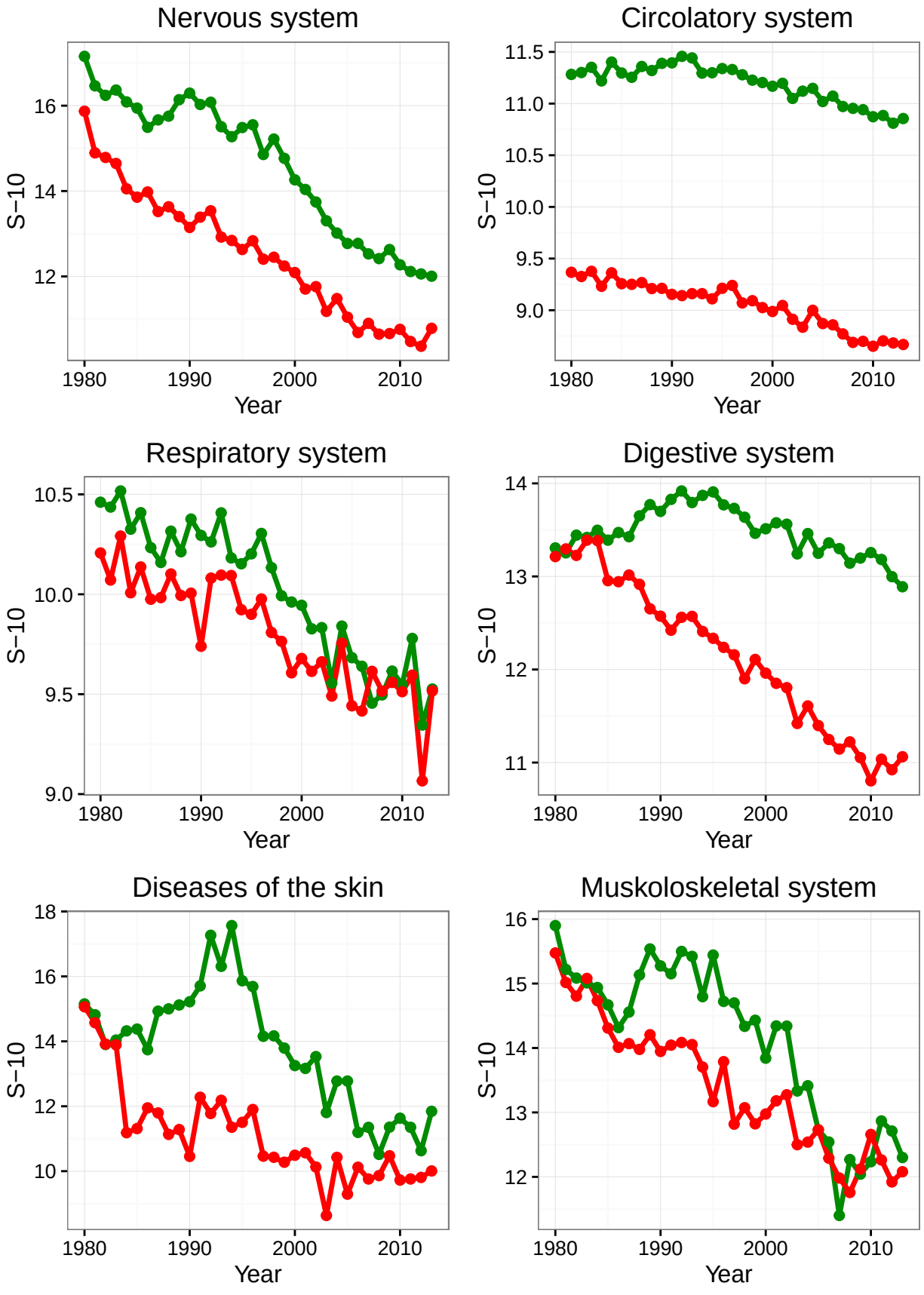
FIGURE 5.4: Cause-specific trend in the modal age at death M from 1980 to 2013 in Italy. Males (left panel) and females (right panel).

nutritional and metabolic diseases the minor progress. Hence, it is clear that the shift to the right of the age-at-death distribution concerns not only all-cause mortality but also cause-specific mortality since longevity has been extended for all causes.

Regarding lifespan variability, the situation is less homogeneous. Figure 5.5 reports the standard deviation of age at death above age ten, S_{10}, for the various causes of death here considered, from 1980 to 2013 and for each sex separately.

The pattern emerging from an overall view could be summarized as follow: a) for most causes (12 out of 15) compression of death is occurring. Among those with declining lifespan variability, the pace of the compression vary greatly. Only endocrine, nutritional and metabolic diseases, congenital malformations and ill-defined causes show a value of S_{10} higher in 2013 than in 1980. b) S_{10} cause of death levels differ noticeably. Diseases of the circulatory and respiratory system report the most compressed age at death distribution while external causes and congenital malformations the most variable one. The difference between these extremes is huge. c) lifespan are less uncertain among females for 13 causes out of 15. In some cases the gender gap is considerable (ENM diseases, mental disorders, nervous system, circulatory system, digestive system, ill-defined and


Males
Females
Infectious diseases
Neoplasms
Diseases of the blood
ENM
Mental disorders
S−10
Year
1980
1990
2000
2010


Nervous system
Circolatory system
Respiratory system
Digestive system
Diseases of the skin
Muskoloskeletal system
S−10
Year
1980
1990
2000
2010

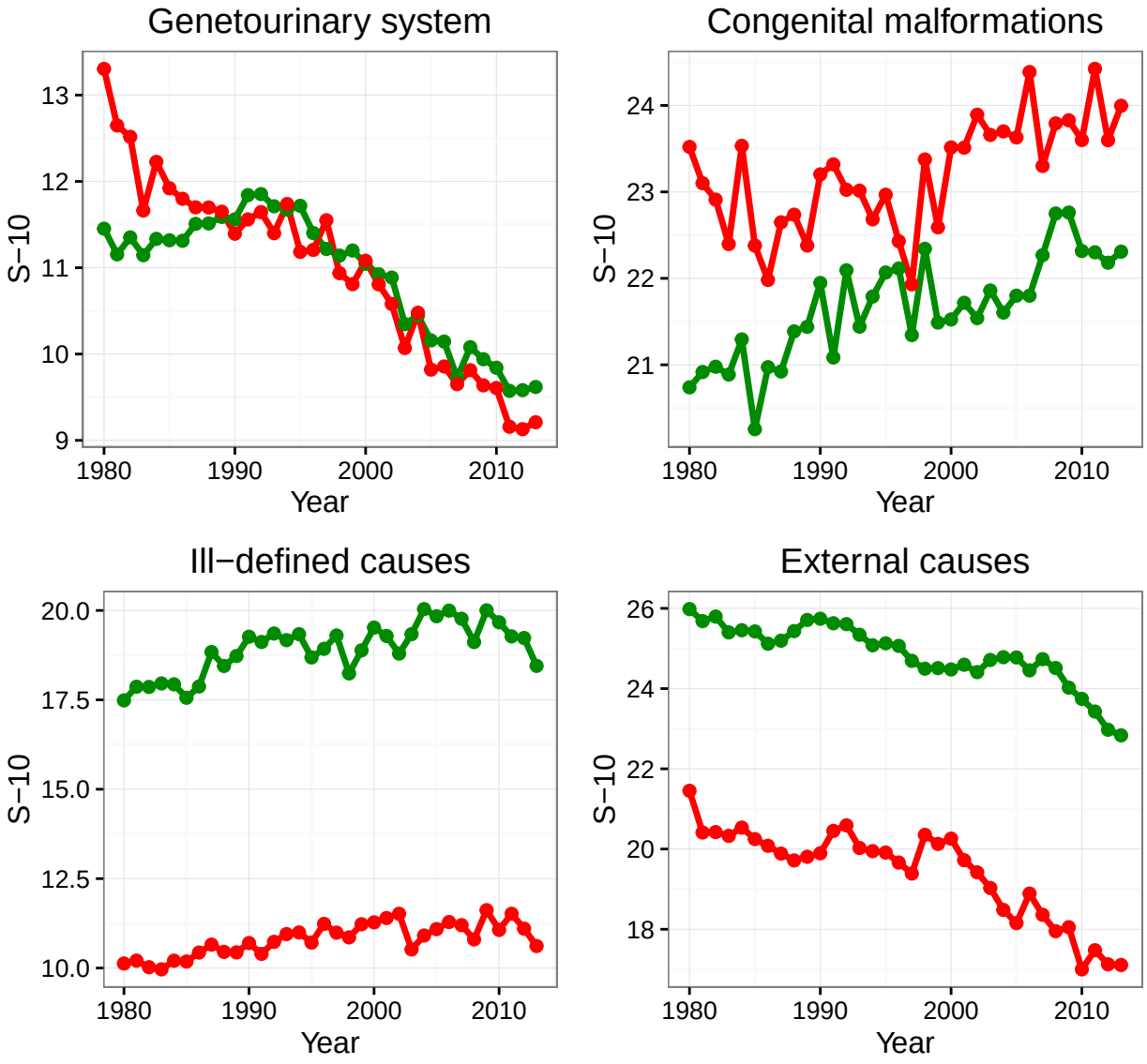


FIGURE 5.5: Standard deviation of ages at death above age ten, S_{10}, for 15 groups of causes of death in Italy from 1980 to 2013.

external causes) while in other less pronounced (infectious diseases, diseases of the blood, diseases of the skin) or close to zero (respiratory system, muskoloskeletal system, genitourinary system). Neoplasms and congenital malformations are the only exceptions, reporting a reversed situation in which lifespan is less uncertain among males. The intriguing finding for neoplasms, as said in the introduction, will be investigated in depth in the next paragraph. For the moment it is worth noting that males advantage is quite important. On average, during the considered period, the difference in terms of lifespan variability was 1.55 years.

Now, turning the analysis into the cause-of-death perspective, other peculiarities emerge. The consequence of AIDS epidemic, which causes thousands of deaths at young ages, led to a considerable increase in infectious

diseases lifespan variability from the middle 1980s to early 1990s (more pronounced for males). Unsurprising those are also the years in which S_{10} all-cause values jumped up as seen in Figure 5.3. Passed the epidemic, S_{10} started to decrease at a remarkable pace either for males or females. During the years of AIDS epidemic an upward jump of lifespan variability, even if much more contained compared to the one reported by infectious diseases, was registered also for other causes of death, especially in males trends (diseases of the blood, ENM diseases, mental disorders, diseases of the digestive system, diseases of the skin, diseases of the muskoloskeletal system, diseases of the genitourinary system). AIDS impact on the whole health it is well known, being correlated to some type of neoplasms, problems of the respiratory and digestive system and diseases of the skin (Greenspan et al., 1996; Jung et al., 1998; Tschachler et al., 1996; WHO, 2007). It may be, thus, that even if AIDS was not registered as underlying cause it had played a role in the process leading to death. If so, during the epidemic also other causes than infectious diseases may experience an increase in lifespan variability. This hypothesis would be tested using multiple cause-of-death data which allow to verify the presence of AIDS as contributory cause of death and not just as underlying. Without multiple cause-of-death data, it is hard to clarify this issue. However, looking back at information coming from surfaces of mortality improvements shown in Paragraph 4.2.3, it appears that many diseases related to AIDS report an increase in death rates at young ages during the 1980s and early 1990s. It may be a clue that AIDS epidemic played a role not only in rising infectious diseases variability but also other causes of death variability.

The two major causes of death, neoplasms and circulatory diseases, reveal a similar pattern of slow compression. Almost a scenario of shifting mortality characterize them: their age-at-death distributions are shifting to the right (M is rising) but almost retaining the same shapes. On the other hand these two causes present differences. The first regarding variability level: S_{10} results higher for neoplasms than circulatory diseases. This confirms as neoplasms still affect young-adult mortality while circulatory diseases are more prevalent at older ages. The second, already mentioned, regards the gender gap which sees males having an advantage in neoplasms and females in diseases of the circulatory system.

As circulatory diseases, also diseases of the respiratory system register low variability level. In this case, though, the compression is a bit more marked and the gender gap less pronounced. Even more marked is the compression registered for diseases of the blood, mental disorders, diseases of the nervous system, diseases of the genitourinary system and external causes. This last group is also the one with higher variability (along with congenital malformations). The decrease in S_{10} for external causes was of a bit less than 4 years and a bit more than 4 years for males and females respectively. Despite this gain, their variability levels are still high, around 22 years for males and 18 years for females. Such great uncertainty in length of life is the consequence of traffic accidents, suicide and homicide which, as seen in Paragraph 4.2.2, play a big role at young ages.

A better understanding of how a given cause of death influence the all-cause S_{10} trend is offered by Figure 5.6. The Figure displays what would be the all-cause S_{10} trend if a specific cause of death was removed. The results are shown only for the causes having a major impact on S_{10} (infectious

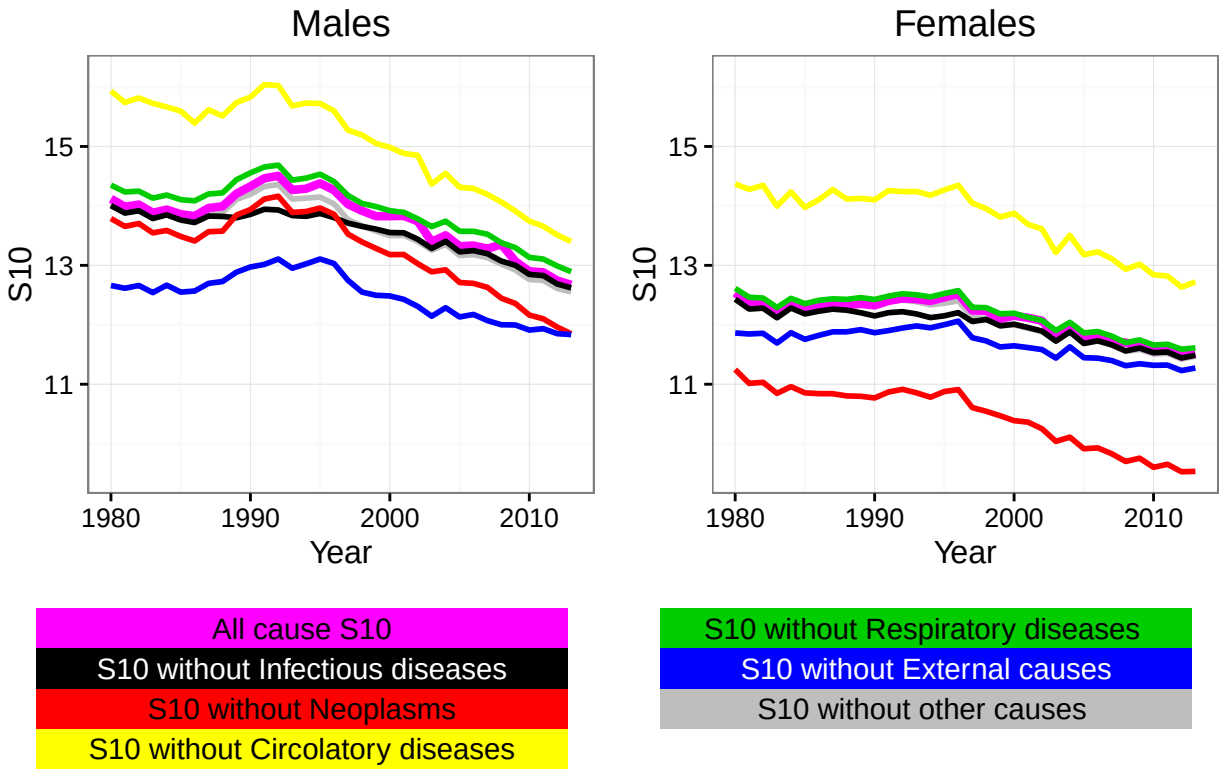


FIGURE 5.6: S_{10} trends for males (left panel) and females (right panel) eliminating a specific cause of death from the computation.

diseases, neoplasms, circulatory diseases, respiratory diseases and external causes) while the other are grouped all together.

Removing circulatory diseases from the computation, the standard deviation of ages at death above age 10 increases greatly either for males or females. It emerges, thus, how this group of causes plays a major role in shaping the all-cause age-at-death distribution and consequently its variability. Also eliminating respiratory diseases S_{10} rises, but this time the rise is almost negligible. On the other hand, the effect of removing neoplasms from calculations, it is of lowering S_{10}. The impact, however, it is much more contained for males than for females. One more time it comes up that neoplasms lifespans are more variable for women than men and how such difference it is considerable. As expected removing external causes of death lead to a decrease in S_{10}, and as expected the decrement is more pronounced for males, who, as seen before, report higher variability for this group of causes than females. As a consequence, a hypothetical elimination of external causes would implicate a drastic reduction of the lifespan inequality gender gap (a better quantification of such reduction will be given in Paragraph 5.3.2). Finally, it is interesting to look at the role played by infectious diseases. Not surprisingly, the effect is to lower S_{10} for both males and females. What is more interesting is the trend between late the 1980s and middle 1990s. Any of the causes examined until now produced an increase or a decrease in S_{10}, but the shape of the trend was pretty much similar to the one reported by all-cause S_{10}. In the case of infectious diseases, instead, apart from S_{10} reduction, it is also possible to note a modification

of the trend's shape between the late 1980s and middle 1990s. As stated earlier, these years were characterized by an increase in lifespan variability, but removing infectious diseases this increase almost disappears. It means that most of S_{10} rise during those years was attributable to infectious diseases and in particular to AIDS epidemic, confirming what hypothesized above.

5.1.3 A focus on neoplasms

One of the most intriguing findings of the previous paragraph surely concerns the higher variability of neoplasms age-at-death distribution among females than males. Why is it so? Is it attributable to a specific type of neoplasms? As mentioned in Paragraph 3.1.2, a deeper level of analysis is feasible from 2003 to 2013, i.e. the ICD-10 period, and will permit to answer the question.

A first reasonable hypothesis to explain the finding is the following: the females higher lifespan variability is not generally true for any type of neoplasm, but attributable to neoplasm of breast. As seen in Figure 4.5, indeed, neoplasms of breast are a) typical of young-adult ages, consequently leading to spread out of the age-at-death distribution b) much more frequent among females, for whom it is the leading cause of death from age 30 to 60.

This hypothesis can be tested simply, computing neoplasms S_{10} removing neoplasm of breast. If true, the variability gender gap would disappear, if false it would persist. The answer is given by Figure 5.7, which shows, for both sexes, neoplasms S_{10} trends and neoplasms S_{10} trends removing neoplasm of breast.

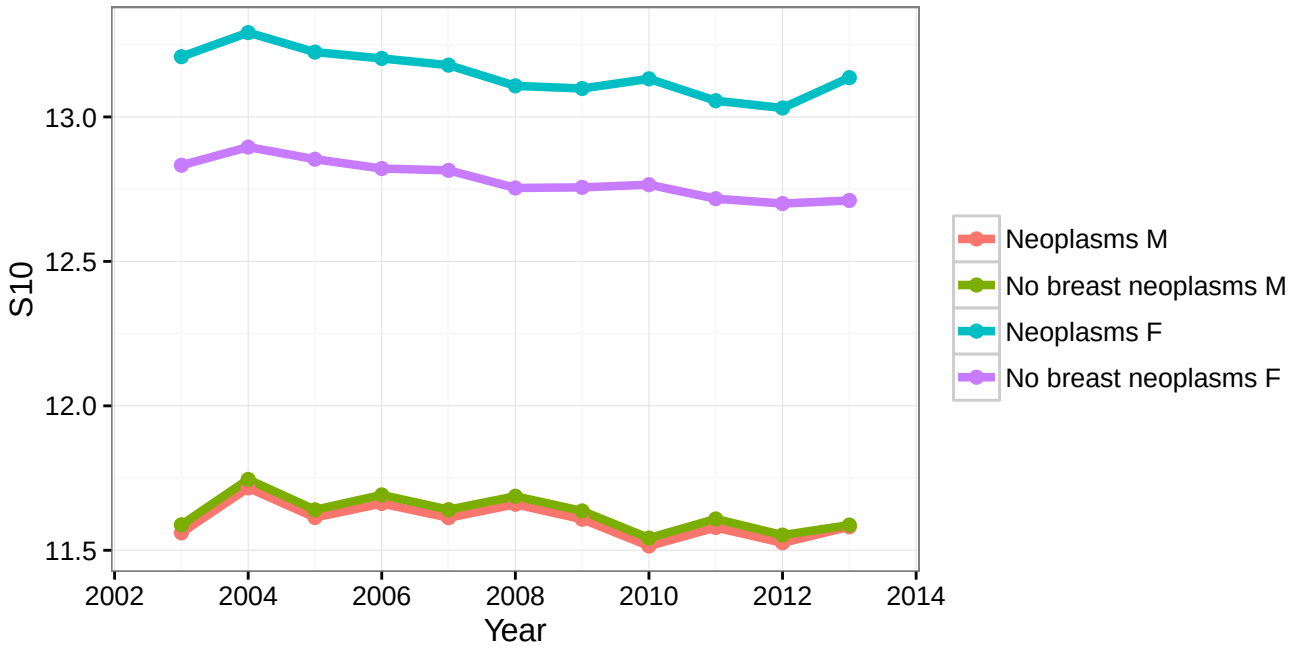


FIGURE 5.7: Neoplasms and neoplasms after removing brest's cancer males and females S_{10} values from 2003 to 2013.

The results evidence that, even if removing neoplasm of breast has the effect of lowering neoplasms S_{10} for females of about 0.5 years, while for males it produces basically no changes, a considerable gap between the two sexes inequality persists. Thus, neoplasm of breast surely contribute to the observed male advantage regarding lifespan variability but, on the other hand, it cannot be addressed as the only cause.

As a consequence, the gender gap is also attributable to other types of neoplasms. Table 5.1 reports, for males and females, S_{10} values in 2003 and 2013 for 21 different types of neoplasms (the 21 types of neoplasms reported by the HCD intermediate classification). The results are clear: lifespan is more unequal for females than males for almost every type of neoplasms. Thus, the males advantage in S_{10} for neoplasms doesn't come from specific cancer location but is more generalized. In the period covered by the study, indeed, a part from malignant neoplasms of liver and pancreas for which S_{10} values are more or less the same between the sexes, for any other kind of cancer the age-at-death distribution is more compressed for males than for females. The difference is particularly big for malignant neoplasms of oral cavity, oesophagus, rectum, larynx, lung, skin, breast and genital organs.

Type of neoplasms	**Males 2003**	**Males 2013**	**Females 2003**	**Females 2013**
Malignant neoplasms of oral cavity	12.78	12.98	14.98	15.61
Malignant neoplasms of oeshophagus	11.22	11.39	15.23	13.88
Malignant neoplasms of stomach	11.26	11.41	12.83	12.76
Malignant neoplasms of colon	10.85	10.99	13.20	13.12
Malignant neoplasms of rectum	11.35	11.19	12.47	13.24
Malignant neoplasms of liver	10.17	10.77	10.10	10.88
Malignant neoplasms of pancreas	11.21	11.04	11.07	11.40
Other malignant neoplasms of the digestive system	11.01	10.87	12.09	11.48
Malignant neoplasms of larynx	11.09	11.22	13.06	15.08
Malignant neoplasms of lung	10.00	10.04	12.20	12.39
Malignant neoplasms of skin	14.19	14.69	17.31	17.85
Malignant neoplasms of breast	12.97	12.25	14.57	14.94
Malignant neoplasms of genital organs	8.67	8.54	13.26	13.79
Malignant neoplasms of kidney	11.34	11.95	12.31	12.45
Malignant neoplasms of bladder	9.70	9.60	10.08	10.92
Malignant neoplasms of nervous system	14.51	14.76	15.77	15.58
Leucemia	14.97	14.31	15.49	15.80
Other malignant neoplasms of the lymphoid and hematopoietic tissue	14.43	14.12	16.02	16.11
Malignant neoplasms of indipendent multiple sites	12.96	12.44	13.01	12.99
Other neoplasms	11.75	11.18	12.47	12.01
Benign neoplasms and neoplasms of uncertain or unknown behaviour	12.84	11.86	12.93	12.12

TABLE 5.1: S_{10} values for 21 different types of neoplasms in 2003 and 2013 for males and females.

A part from gender differences, the results shown in Table 5.1 enable to find out the variability of the age-at-death distribution for each type of

neoplasms and therefore to know the current differences, in terms of lifespan inequality, between different cancer's location. As we have seen before, from 1980 to 2013, males reported a quite stable trend in S_{10}, a bit above 11.5 years, for the macro-chapter of neoplasms. This level of variability is also reported for many specific types of cancer, like malignant neoplasms of oesophagus, stomach, colon, rectum, pancreas, larynx, kidney and other neoplasms. Well above this average value are, instead, malignant neoplasms of oral cavity, skin, breast, nervous system, other malignant neoplasms and leucemia. On the other hand malignant neoplasms of liver, lung, genital organs and bladder are the types of cancer for whom the age-at-death distribution is more compressed. For females, even if at a higher level of variability, the situation is quite similar. Also for them, indeed, malignant neoplasms of oral cavity, skin, breast, nervous system and leucemia are the one with highest value of S_{10}, while malignant neoplasms of liver, lung and bladder the one with low value of S_{10}.

5.2 Sensitivity analysis of S_{10}

All-cause mortality lifespan variability trend, as measured by S_{10}, has revealed that the age-at-death distribution, of either males or females, is still becoming more and more compressed, highlighting how a scenario of shifting mortality it is still not conceivable for Italy.

A powerful tool to deeper understand the reasons behind the observed reduction in the age-at-death distribution variability is represented by sensitivity analysis. In particular, through this method, it is possible to investigate on the relationship between the age pattern of mortality change and trends in variability of age at death. As illustrated in Paragraph 2.1.2 and 3.2.5, using S_{10} as measure of lifespan variability, improvements in mortality don't automatically translate into a lower variability of the age-at-death distribution, conversely, "young" ages decline (increase) produces compression (expansion) while "old" ages decline (increase) produces expansion (compression).

A full comprehension of this mechanisms is of fundamental importance since it reveals the manner in which variability in length of light evolves over time. Probably, at first sight, such double effect of mortality improvements on lifespan variability could appear nonsense. A better understanding of this scenario was pointed out in Chapter 2 and displayed by Figure 2.4: from the top panels one can see that the effect of reducing mortality rates, m_x, at young ages is to decrease the variability of the density function, d_x , as deaths are pushed out further on the tail of the age-at-death distribution. From the bottom panels, instead, one can see that the effect of reducing mortality rates at old ages is to increase the variability of the density function as deaths of the right-hand tail of the age-at-death distribution are more spread out and concentrated at later ages.

Of course, mortality reduction has to be always seen as advantageous even though at later ages it produces an increase in the variability of the age-at-death distribution. What matters is that the variability's rise produced by "old" age mortality decline is compensated (or even better overtaken) by the "young" age mortality decline. Therefore, the increase in lifespan variability due to a reduction of "old" age death rates can be regarded

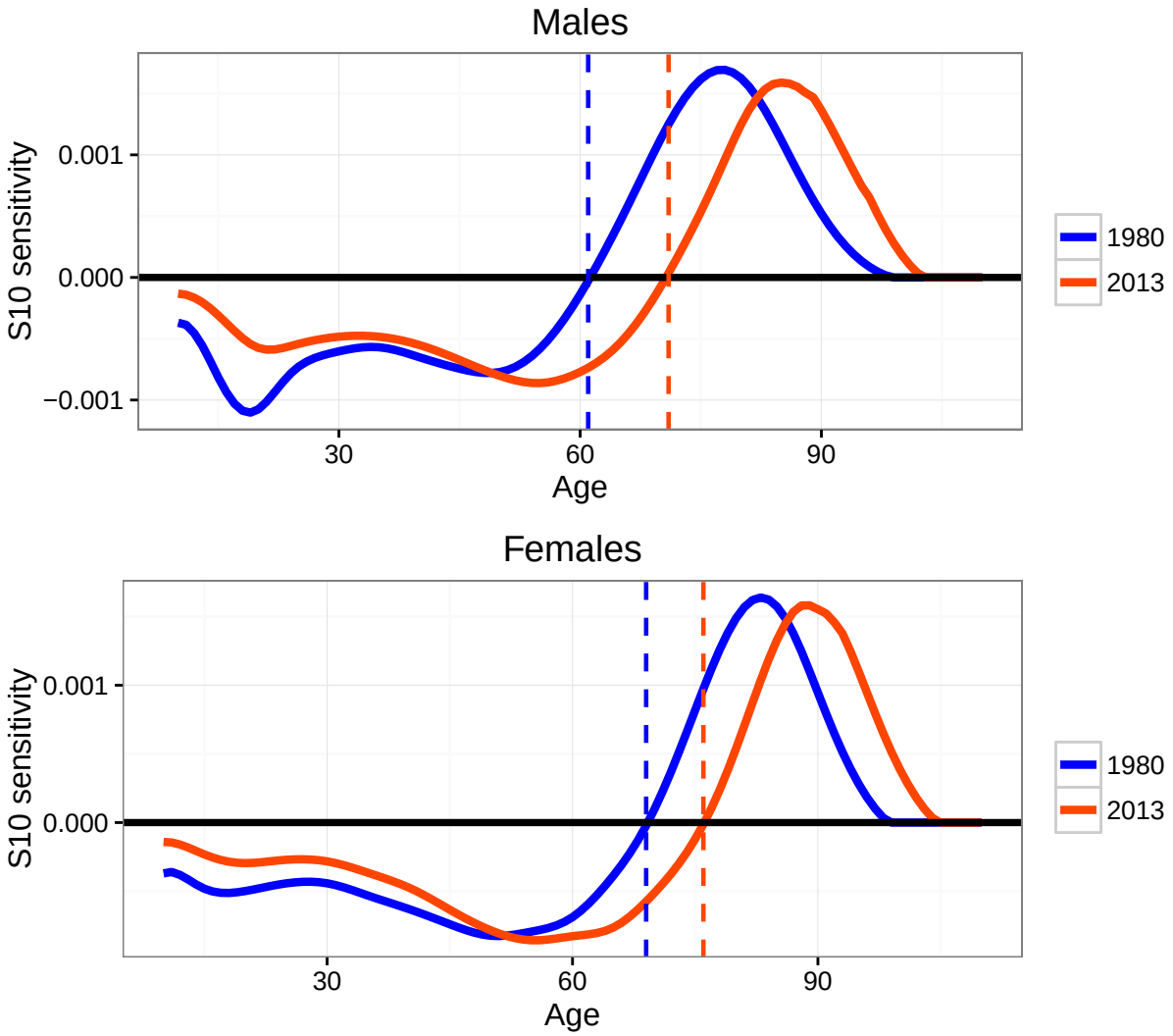


FIGURE 5.8: Sensitivity of S_{10} in 1980 and 2013 to a proportional 1% decrease in mortality rates at all ages.

as something positive since such increase comes from a shift on the right of the age-at-death distribution.

Figure 5.8 shows, for each sex separately, the sensitivity of S_{10}, in 1980 and 2013, to a proportional 1% drop in mortality rates at all ages, i.e. what would be the contribution of each age to a change in S_{10} if mortality decreased by 1% at all ages. As already said, sensitivity analysis' outcome is important in three ways. First, they allow understanding which ages, after a reduction of death rates, would produce a decrease or an increase in S_{10}. Second, they assess the magnitude of this potential decrease or increase. Third, they identify a threshold age T_{10}, which individuate, regarding lifespan variability, the "young" and old "age" range, or in other words, the "early" and "late" deaths.

What emerges from Figure 5.8 is as follow. Before the threshold, i.e. where mortality decline acts reducing S_{10}, males' sensitivity presents two bumps, the first at young ages, more precisely between age 15 and 25, and the second at adult ages, more precisely between age 40 and 50 in 1980 and between age 45 and 60 in 2013. The first bump, considerably reduced in

2013 with respect to 1980, is the result of the rapid increase in mortality produced by external causes of death at those ages. As a consequence, a decline in mortality between 15 and 25 would have a greater impact on S_{10} than a decline at immediately previous ages (10-14) or immediately subsequent ages (26-30). The second bump reflects the fact that at adult ages mortality is no more as low as at young ages and, therefore, improvements would lead to a more considerable decline of S_{10}. After T_{10}, instead, the contributions are bell-shaped, with the maximum positioned about 15 years after the threshold.

Females' pattern looks like quite similar to males' pattern, with the major difference represented by a very contained bump between ages 15-25 as a result of their considerable mortality advantage at those ages due to the minor impact of external causes on mortality rates.

In the period under analysis, the age separating "early" and "late" deaths moved forward considerably, from 61 to 71 years for males and from 69 to 76 for females. Thus, as time passes by, negative contributions have more ages available as T_{10} shifts to the right, but also a slightly lower sensitivity, while positive contributions have fewer ages available and roughly the same sensitivity's levels.

Sensitivity analysis is surely really intriguing, but it only allows understanding how S_{10} would react to a proportional mortality decline. However, in reality, death rates do not decline proportionally at all ages and, of course, in some cases they could even increase. Therefore, in order to understand how the actual age pattern of mortality change from 1980 to 2013 has affected all-cause S_{10} trend, I computed, for each age, the actual contribution to changes in S_{10}. The results obtained are displayed in Figure 5.9 using Lexis' surface. Areas with the same color imply the same contribution's level to changes in variability of age at death. Negative contributions, i.e. the ones that act lowering S_{10}, are green colored, while positive contributions, i.e. the ones that act incrementing S_{10}, are red colored. The deeper is the color the higher is the contribution (in both senses).

The clear partition of negative and positive contributions to changes in S_{10} along ages can be immediately seen. The threshold T_{10} distinctly separates ages which contribute to reduce variability in length of life after a reduction of mortality rates from those which contribute to increase variability in length of life after a reduction of mortality rates. Therefore, red areas before T_{10} and green areas after T_{10} indicate an increase in mortality. In particular, it is possible to note the great impact (especially for males) that AIDS epidemic had in increasing S_{10} between the late 1980s and early 1990s, explaining the increase in uncertainty about the time of death registered in those years.

The most important feature emerging from Figure 5.9 comes from the analysis of how trends in lifespan variability are determined by these two opposing forces. As stated before, the increase in S_{10} attributable to old-age death rates reduction, cannot be seen itself as something negative since such greater variability is the result of the postponement of deaths at older ages. On the other hand, enlarging inequality in lifespan can never be seen as something positive as well. So, it results of basic importance that young-age mortality reduction is strong enough to overtake the increase of uncertainty in the timing of death produced by a shift on the right of the age-at-death distribution at old ages. Thankfully, this is exactly what is happening in

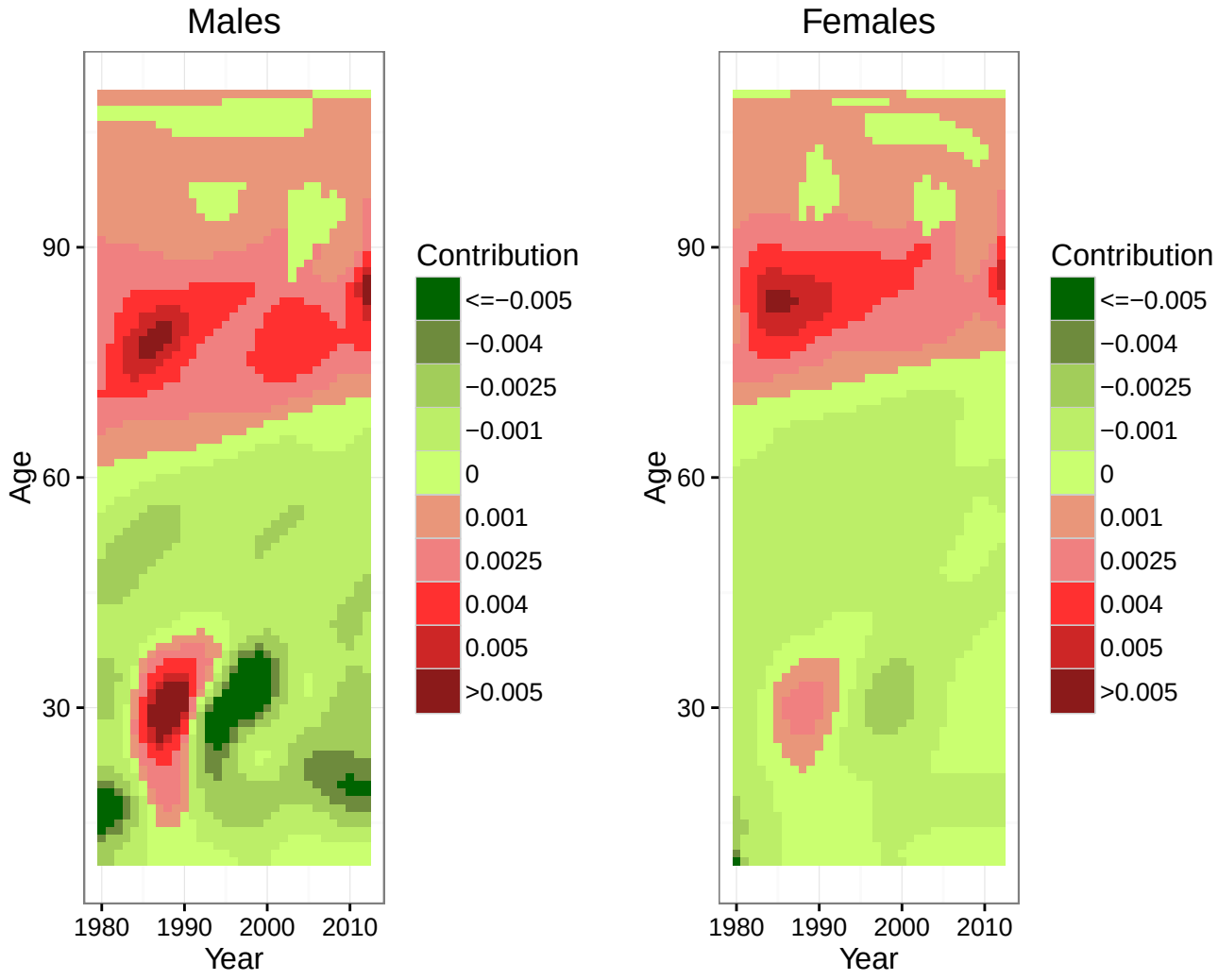


FIGURE 5.9: Age-specific contributions of changes in S_{10} from 1980 to 2013.

the Italian population: as shown in Paragraph 5.1.1, inequality in lifespan continues to go down in both sexes. Such observed reduction in variability of age at death, it is due to the capability of mortality changes before T_{10} to be more influent than mortality changes after T_{10}.

5.3 Cause-of-death decomposition of S_{10}

Which causes are responsible for the observed mortality compression? Which causes are responsible for the gender gap in lifespan variability? In this paragraph, I will answer to these questions exploiting decomposition technique.

With regard to the first question, I have previously shown that from 1980 to 2013 S_{10} has decreased of 1.45 years for males and of 1.02 years for females. In section 5.3.1, these two differences will be decomposed into the contribution of changes in age and cause-specific mortality rates, $m_{x,k}$, from 1980 to 2013 to change in S_{10} over the same period. In this way, for each sex separately, it will be possible to understand the age and cause profiles that have determined the observed mortality compression. The results are displayed in Figure 5.10. The bars under the zero line indicate particular age-cause profiles which contribute to the observed mortality compression measured by S_{10}, while the bars above the zero line indicate particular age-cause profiles which acted oppositely, i.e. expanding mortality.

With regard to the second question, in section 5.3.2, the lifespan variability gender gap in 1980 (1.59 years) and 2013 (1.16 years) will be decomposed into the contribution of differences in age and cause-specific mortality rates, $m_{x,k}$, between sexes to the difference in S_{10} between sexes. The results are displayed in Figure 5.11. In this case, the bars under the zero line indicate particular age-cause profiles which positively contribute to the observed female's advantage in S_{10}, while the bars above the zero line indicate particular age-cause profiles which contribute reducing the lifespan variability gender gap.

5.3.1 Decomposing the change over time of S_{10}

In this section, for each sex separately, the first question stated above, i.e. which causes are responsible for mortality compression? is going to find an answer.

Figure 5.10 shows decomposition's result with males in the top panel and females in the bottom panel. Before entering into details, it is worth noting the shape of the contribution of changes in $m_{x,k}$ from 1980 to 2013 to changes in S_{10} over the same period displayed by both sexes. Negative and positive contribution are not equally distributed along the x-axis, representing age. The first dominate from the first considered age (10) up to adult-old ages, the latter prevail at old and very old ages. Why this pattern? As just seen in the previous paragraph, the effect of mortality decline on S_{10} varies by age: up to a certain age mortality reduction acts lowering lifespan variability and after such age acts incrementing lifespan variability. Therefore, the observed mortality decline, which both males and females experimented from 1980 to 2013, did not always produce a decline in S_{10}, as it would be for life expectancy.

Table 5.2 and 5.3 are helpful to better distinguish such contrast between young-adult ages and old ages and thus have a better comprehension of decomposition's outcomes. The tables report, for each cause, the overall contribution to changes in S_{10} split into two categories. The categories are identified using the first age at which a decline in mortality rates would increase S_{10} as the upper limit of the young-adult category and lower limit of

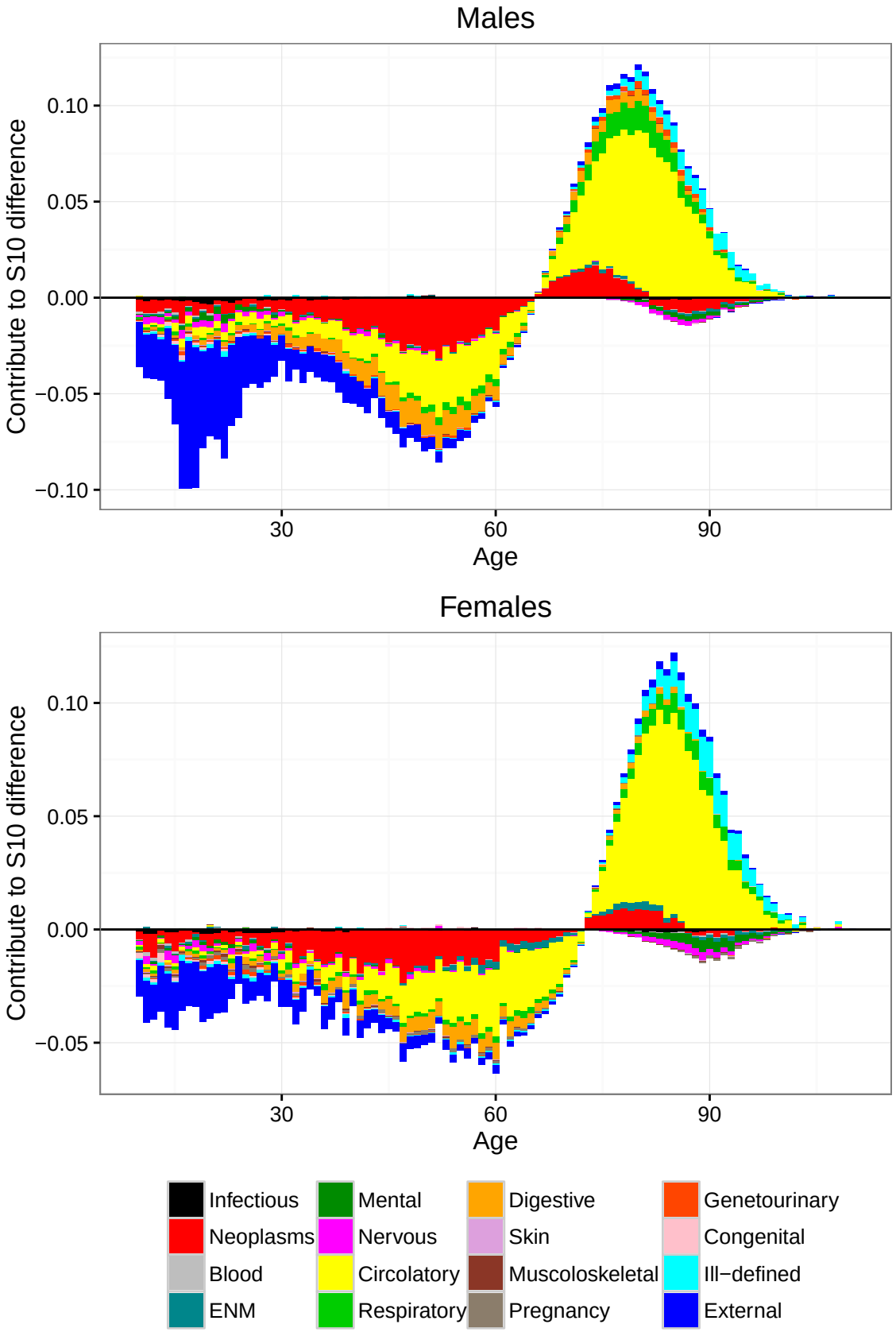


FIGURE 5.10: Contribution of changes in age and cause-specific mortality rates from 1980 to 2013 to change in S_{10} over the same period in Italy. Males (top panel) and females (bottom panel).

Cause of Death	Untill 66	After 66	Total
All causes	-3.25	1.80	-1.45
Infectious diseases	-0.04	-0.01	-0.05
Neoplasms	-0.71	0.02	-0.69
Diseases of the blood	-0.01	0.00	-0.01
Endocrine Nutritional and Metabolic	-0.02	0.00	-0.02
Mental disorders	-0.05	-0.03	-0.08
Diseases of the nervous system	-0.08	-0.03	-0.11
Diseases of the circulatory system	-0.69	1.28	0.59
Diseases of the respiratory system	-0.10	0.22	0.12
Diseases of the digestive system	-0.37	0.10	-0.27
Diseases of the skin	0.00	0.00	0.00
Diseases of the muskoloskeletal system	-0.01	0.00	-0.01
Diseases of the genetourinary system	-0.03	0.05	0.02
Congenital malformations	-0.02	0.00	-0.02
Ill-defined causes	-0.02	0.16	0.14
External causes	-1.10	0.04	-1.06

TABLE 5.2: Contribution of changes in age and cause-specific mortality rates from 1980 to 2013 to change in S_{10} over the same period in Italy. Males

Cause of Death	Until 73	After 73	Total
All causes	-2.54	1.52	-1.02
Infectious diseases	-0.03	-0.02	-0.05
Neoplasms	-0.62	0.05	-0.57
Diseases of the blood	-0.01	0.00	-0.01
Endocrine Nutritional and Metabolic	-0.07	0.02	-0.05
Mental disorders	-0.02	-0.08	-0.10
Diseases of the nervous system	-0.06	-0.04	-0.10
Diseases of the circulatory system	-0.71	1.21	0.50
Diseases of the respiratory system	-0.08	0.13	0.05
Diseases of the digestive system	-0.25	0.03	-0.22
Diseases of the skin	0.00	0.00	0.00
Diseases of the muskoloskeletal system	-0.01	0.00	-0.01
Diseases of the genetourinary system	-0.06	0.00	-0.06
Pregnancy and related conditions	-0.02	0.00	-0.02
Congenital malformation	-0.01	0.00	-0.01
Ill-defined causes	-0.04	0.22	0.18
External causes	-0.55	0.05	-0.50

TABLE 5.3: Contribution of changes in age and cause-specific mortality rates from 1980 to 2013 to change in S_{10} over the same period in Italy. Females

the old category. This age can be placed at 66 for males and 73 for females. It is to be noted that such threshold age is conceptually different from those found in Paragraph 5.2. The thresholds of Paragraph 5.2 tell us at which age the effect of a decline in death rates on S_{10} is reversed from negative (lowering lifespan variability) to positive (increasing lifespan variability) given the age pattern of mortality change registered in *each year*. The threshold found in the current Paragraph, instead, tells us at which age the effect of a decline in death rates on S_{10} is reversed from negative to positive *all over* the period 1980-2013, thus considering the whole age pattern of mortality change observed from 1980 to 2013.

About males' results, the registered mortality compression of 1.45 years has been primarily driven by the reduction of external causes death rates, especially from age 10 to 30. The contribution of this group of causes in lowering S_{10} is really remarkable (-1.06 years) and explain more than the 70% of males reduction in uncertainty in the timing of death. Intriguing is the role played by neoplasms and diseases of the circulatory system. The two major causes of death act in a similar way up to age 66, whereas the sign of their contribution is reversed. Until that age, indeed, neoplasms produce a reduction of S_{10} equal to 0.71 years and diseases of the circulatory system of 0.69 years. Their contribution in lowering lifespan variability is particularly important at adult ages, from age 45 to 60. At old ages, where a decline in mortality rates produce an increase in lifespan variability, instead, they behave differently. With regard to diseases of the circulatory system the dramatic decrease in mortality rates acts rising S_{10}, with a noticeable effect of 1.32 years. For neoplasms the effect is really close to 0: up to age 82 they produce a small increase of S_{10} which is however offset by a decline of S_{10} from age 83, meaning that from this age, mortality rates have been growing in the period under study. Summing up all ages, the overall contribution to changes in lifespan variability of the two major causes is negative (-0.69) for neoplasms and positive (0.59) for diseases of the circulatory system. Another group of causes with a considerable effect is diseases of the digestive system, whose contribution in reducing S_{10} is of 0.27. To be noted is also the role played by diseases of the respiratory system (lowering lifespan variability by 0.10 years at young-adult ages and incrementing lifespan variability by 0.22 years at old ages), which is comparable to the one of diseases of circulatory system but at a lower level of intensity, and the role played by ill-defined causes that, due to a consistent reduction in old ages mortality rates, produced an increase in S_{10} equals to 0.14 years.

Females' results are quite similar to males' results, either regarding the age or cause profile contribution pattern. Also in this case neoplasms play a major role in lowering lifespan variability (-0.57) while diseases of the circulatory system act incrementing S_{10} (0.50). The major difference regards external causes of death, whose impact on mortality compression is much less remarkable (0.52 years). This outcome is not surprising. As seen in Paragraph 4.2.3, external causes of death mortality rates decline at young ages has been noticeable in both sexes. Males' decline, however, has led to a greater compression of the age-at-death distribution because their mortality rates at young ages are much higher than the one of the females. In other words, males' S_{10} is more sensitive to a reduction in mortality at those ages than females' S_{10} (see Paragraph 5.2). Less remarkable is also the role of

diseases of the digestive system producing a decline of S_{10} equal to 0.22 years and diseases of the respiratory system (-0.08 at young-adult ages and 0.13 at old ages).

5.3.2 Decomposing differences between sexes in S_{10}

Italian females have been experiencing a more certain length of life than italian males. An epidemiological explanation of the female advantage in lifespan variability is given by figuring out which causes of death are determining such advantage through the decomposition of the S_{10} gender gap. Since the decompositions are performed for 1980 and 2013 it will also be possible to evaluate potential changes in the causes of death profiles generating the gap over the course of the considered period.

The shape of the contribution of differences in $m_{x,k}$ between sexes to difference in S_{10} between sexes presents a similar pattern of the previously analyzed decomposition over time of S_{10}. The reason is clearly the same: lower mortality rates are an advantage in terms of lifespan variability only up to a certain age, thus the fact that males have higher rates than females at any age leads to a situation in which young and adult ages produce the gender gap while old and very old ages reduce it. So, in this case, the thresholds founded (65 in 1980 and 74 in 2013) tell us at which age the effect of having lower death rates, *between sexes*, on the S_{10} difference, *between sexes*, is reversed from negative to positive.

In 1980, the 1.59 years of difference in S_{10} between males and females was primarily driven by external causes at young and adult ages: 1.19 years, of which most between ages 15 and 30. At those ages external causes of death basically mean transport accidents: females less risky behavior substantially explain more than half of the lifespan variability gender gap in 1980. The role played by neoplasms and diseases of the circulatory system is ambiguous: if on one hand they are surely important (red and yellow bars stand out clearly in the graph) on the other hand their contribution to the S_{10} gender gap is quite small. Their young-adult ages contribution generating the males-females difference in lifespan variability (0.36 years for neoplasms and 0.55 years for diseases of the circulatory system) is, indeed, basically offset by their old and very old ages contribution which acts in the opposite way (0.32 years for neoplasms and 0.45 years for diseases of the circulatory system). About the remaining causes of death, what is worth mentioning, it is the role played by diseases of the digestive system which positively contribute to the gender gap by 0.19 years and the role played by diseases of the respiratory system which, instead, reduce female advantage by 0.15 years.

The 2013's decomposition reveals a similar pattern: lifespan variability gender gap is generated by females advantage in mortality rates up to age 74 to be then reduced at older ages for the same reasoning. Neoplasms and diseases of the circulatory system contribution to the sex difference in S_{10} are a bit more pronounced than in 1980. Neoplasms young-adult ages produce an advantage for females of 0.24 years while old ages advantage males of 0.36 years. Thus, overall, neoplasms are not responsible for the gender gap in lifespan variability but to the contrary, they actually behave oppositely since the total effect is of 0.12 in favor of males. This is in line with what observed in Paragraph 5.1.2 and 5.1.3 about higher females neoplasms

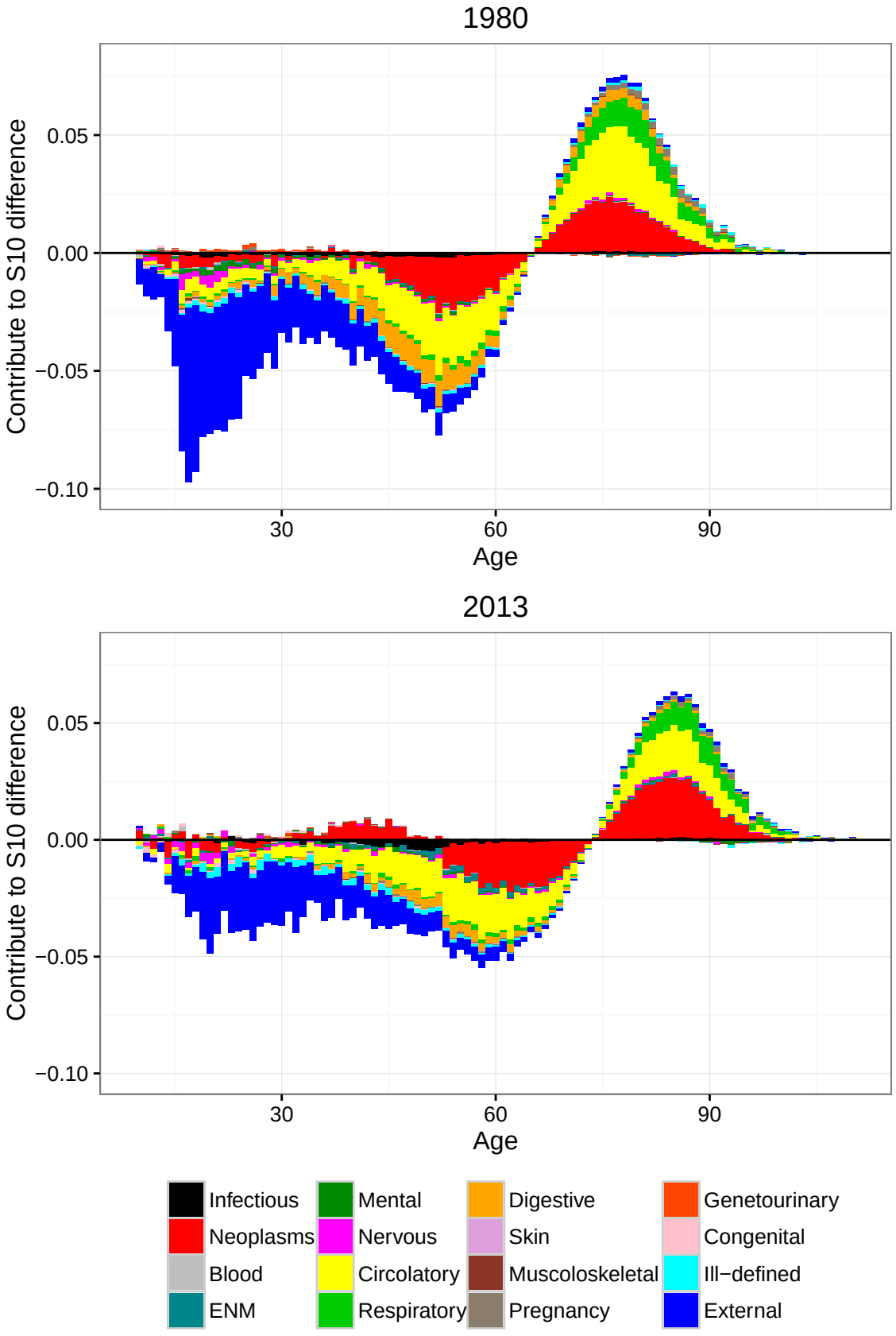


FIGURE 5.11: Contribution of differences in age and cause-specific mortality rates between sexes to the difference in S_{10} between sexes in Italy. 1980 (top panel) and 2013 (bottom panel).

Cause of Death	Until 65	After 65	Total
All causes	-2.74	1.15	-1.59
Infectious diseases	-0.06	0.01	-0.05
Neoplasms	-0.36	0.32	-0.04
Diseases of the blood	-0.01	0.00	-0.01
Endocrine Nutritional and Metabolic	-0.02	-0.02	-0.04
Mental disorders	-0.06	0.01	-0.05
Diseases of the nervous system	-0.06	0.02	-0.04
Diseases of the circulatory system	-0.55	0.45	-0.10
Diseases of the respiratory system	-0.07	0.19	0.12
Diseases of the digestive system	-0.27	0.08	-0.19
Diseases of the skin	0.00	0.00	0.00
Diseases of the muskoloskeletal system	0.00	0.00	0.00
Diseases of the genetourinary system	0.00	0.05	0.05
Pregnancy and related conditions	0.00	0.00	0.00
Ill-defined causes	-0.07	0.02	-0.15
External causes	-1.21	0.02	-1.19

TABLE 5.4: Contribution of differences in ages and cause-specific mortality rates between sexes to the difference in S_{10} between sexes in Italy. 1980.

Cause of Death	Until 74	After 74	Total
All causes	-2.07	0.91	-1.16
Infectious diseases	-0.08	0.01	-0.07
Neoplasms	-0.24	0.36	0.12
Diseases of the blood	0.00	0.00	0.00
Endocrine Nutritional and Metabolic	-0.06	0.01	-0.05
Mental disorders	-0.02	-0.01	-0.03
Diseases of the nervous system	-0.05	0.01	-0.04
Diseases of the circulatory system	-0.53	0.27	-0.26
Diseases of the respiratory system	-0.05	0.17	0.12
Diseases of the digestive system	-0.11	0.02	-0.09
Diseases of the skin	0.00	0.00	0.00
Diseases of the muskoloskeletal system	0.00	0.00	0.00
Diseases of the genetourinary system	-0.01	0.03	0.02
Pregnancy and related conditions	0.00	0.00	0.00
Ill-defined causes	-0.10	0.00	-0.10
External causes	-0.78	0.03	-0.75

TABLE 5.5: Contribution of differences in ages and cause-specific mortality rates between sexes to the difference in S_{10} between sexes in Italy. 1980.

S_{10} trend, evidencing one more time how the neoplasms age-at-death distribution is more spread out among females than males. For diseases of the circulatory system, instead, the young-adult ages contribution to the lifespan variability gender gap (0.53) is higher than males advantage at old-ages (0.27).

What really differentiates 1980 and 2013 is the role played by external causes of death. In 2013, indeed, external causes' contribution to the lifespan variability males-females difference was equal to 0.75 years, a considerable less amount than in 1980. Males are thus recovering their young-age mortality disadvantage for this group of causes and this mostly explains why the gender gap in S_{10} have been reduced over the time-period considered here (1.59 years in 1980 vs 1.16 years in 2013).

Chapter 6

Old-age lifespan variability and its relationship with the age pattern of mortality change

In Chapter 5 the dispersion of deaths over a broad range of ages (10+) has been examined. However, in the era of longevity extension, particular attention has to be given to changes in the death distribution within old ages. Thus, to tackle this issue, the current Chapter is entirely devoted to the investigation of variability in length of life at old ages.

Here, the standard deviation of ages at death above the mode, $SD(M+)$, is the indicator used to monitor the evolution of lifespan variability at old ages. In particular, the study is carried out using a simplified version of the logistic model - the Kannisto model - exploiting its well-known property of linking the compression of deaths above the mode with the rate of increase in mortality, i.e. the rate of aging. Moreover, to arrive to a deeper understanding of lifespan variability at old ages, causes of death are taken into consideration. Indeed, exploiting the cause-of-death decomposition of the changes in lifespan variability above the modal age at death proposed by Horiuchi and his colleagues (Horiuchi et al, 2012) new epidemiological insights about the observed mortality compression will be given.

6.1 The steepening of the right-hand tail

In the last decades of the XXth century and the first decade of the XXIst century a considerable decline in old-age mortality was registered in developed countries (Janssen, 2007; Kannisto, 1984; Staetsky, 2009). With regard to Italy, in Paragraph 4.2.3, using the Rates Of Mortality Improvements (ROMI), I have shown that death rates among the old have been decreasing constantly from 1980 to nowadays in both sexes, with few exceptions registered at extremely old ages.

Given that, a relevant question that demographers have to take into account is whether the remarkable improvements registered in old-age mortality are associated with a steepening of the right-hand tail of the age-at-death distribution or not. The importance of this question arises from the consideration that, if deaths are becoming more and more compressed at old ages, it may be an indication of increasing resistance to further longevity extension, i.e. human's beings lifespan is approaching a biological limit (Horiuchi et al., 2012).

To monitor the changes of the right-hand tail of the age-at-death distribution, *SD(M+)* has been used firstly by Kannisto and then in various studies (Cheung et al., 2005, 2008, 2009; Cheung and Robine, 2007; Kannisto, 1984; Kannisto, 2000; Kannisto 2001; Ouellette and Barbeau, 2011; Robine et al., 2006). These researches have demonstrated that compression of ages at death above the mode is occurring in several developed countries even though the pace of compression has been pretty slow.

A first evaluation of how the distribution of deaths above the modal age at death is behaving can be simply done comparing its shape in two different moments in time. Figure 6.1 helps visualizing such comparison. It displays, for each sex separately, the shape of the right-hand tail of the age-at-distribution in Italy for 1980 and in 2013. The comparison makes sense because having age-M on the x-axis guarantees that both distributions have the same starting point (age 0) and having d_x / d_M on the y-axis guarantees that both distributions have the same height (1) at that starting point. What emerges from Figure 6.1 is that deaths above the mode have been increasingly concentrated in a narrower age range over the considered period 1980-2013. Although old-age compression is occurring in both sexes, it is much more pronounced among males than females, i.e. the steepening of the right-hand tail of the distribution is much more evident for males than females.

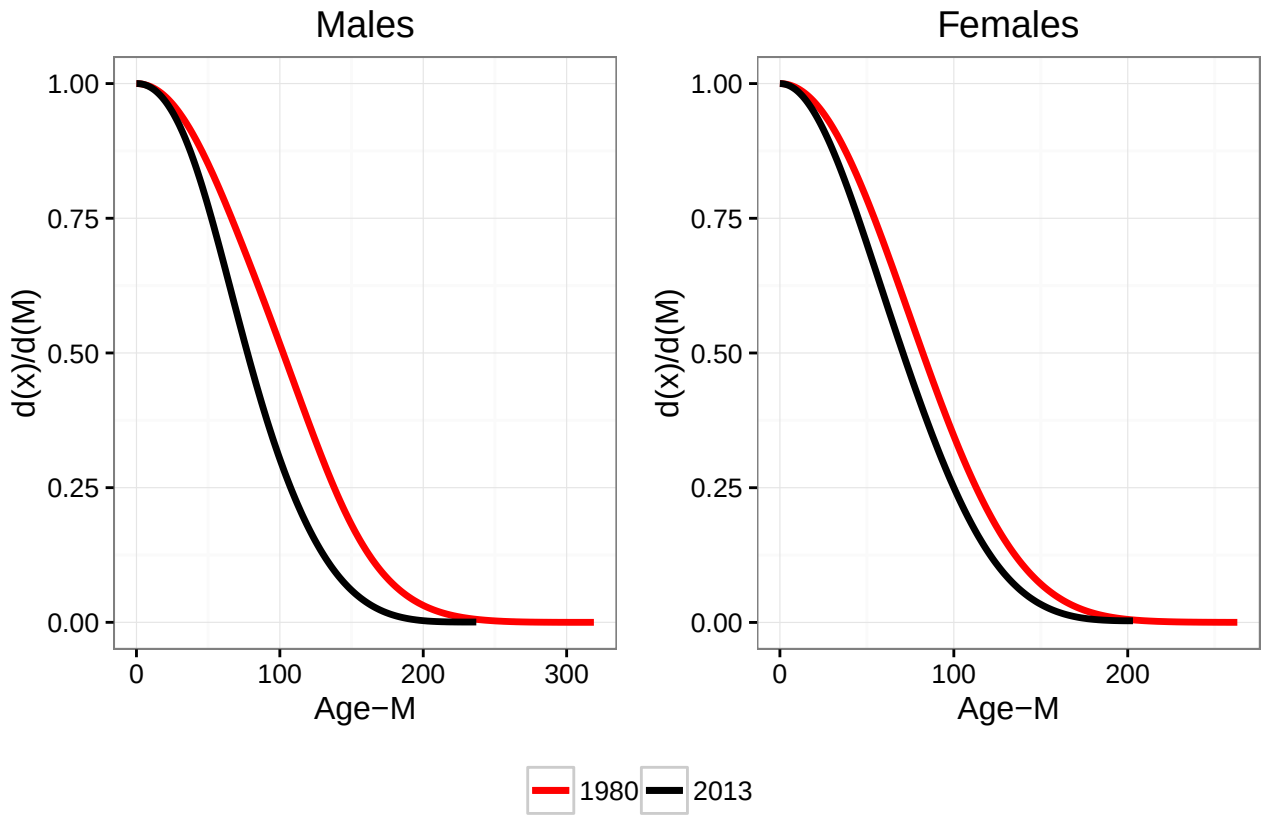


FIGURE 6.1: Distribution of deaths above the modal age at death M for Italian males and females in 1980 and 2013.

Even if quite useful, the information resulting from Figure 6.1 are neither able to precisely quantify the extent of the registered compression over time nor the exact level of variability. An accurate assessment about the dispersion of deaths above the mode can be only done computing *SD(M+)*. As already highlighted in Paragraph 2.1.2 and 3.2.6, Thatcher et al. have demonstrated that, using the simplified logistic model with only two parameters (as known as the Kannisto model), it exists a relationship between *SD(M+)* and the accelerating pace of age-related mortality increase, denoted by the parameter b in the model and measured by the logistic rate of mortality rise. The exact relationship between *SD(M+)* and b is reported by Equation 3.35 of Paragraph 3.2.6, and it holds that if mortality rises more steeply with age (i.e. if the parameter b increases) then compression above the mode will occur (*SD(M+)* will decrease). As a consequence, the parameter b itself can be used as a measure of the steepness of the right-hand tail of the age-at-death distribution.

The appealing property of linking the rate of aging and mortality compression combined with really good fit of the model at old ages for modern populations, makes Kannisto model a suitable choice to investigate on the evolution of lifespan variability of deaths above the mode. Thus, I adopted the model to assess, quantificate and understand better the old-age compression's scenario previously revealed by Figure 6.1.

The model has been applied to males and females mortality rates between age 70 and 90 in the time interval from 1980 to 2013[1]. Figure 6.2 displays the logit of m_x (dots) and the estimated logit lines from the Kannisto model for males and females in Italy for 1980 and 2013. First of all, it can be clearly seen that the observed point, i.e. the logit of m_x, look really straight making the fit of the estimated logit lines almost perfect. This justifies the application of the model, confirming the right choice of the logistic equation to model mortality data at old ages. Secondly, it shows that as time passes by death rates noticeably decreased at old ages and that the pattern of mortality decline has not been proportional across ages. Indeed, the logit lines in 1980 and 2013 do not have the same slope since death rates have been fallen faster at age 70 than at age 90: the rate of aging, measured by the parameter b of the model, has increased. Moreover, the increase in the logit's line slope has been more pronounced for males than females. This basically confirms what was previously illustrated in Figure 6.1: from 1980 to 2013 compression above the mode has occurred, i.e. the right-hand tail of the distribution has become steeper, and such compression results to be more evident for males than for females.

Now, to precisely assess compression's extent and gender differences it is necessary to look at the trends in b parameter estimated from the Kannisto model. The results are shown in Figure 6.3. Three essential features emerge. Firstly, all over the considered period 1980-2013 females show higher b's values than males. This means that the rate of aging is higher among females than males and as a consequence lifespan above the mode are less unequal for females than males. Secondly, as expected, b's values have been growing up over time for both sexes: the rate of aging is increasing and therefore compression above the modal age at death is occurring.

[1] For further methodological details on the model's application see Paragraph 3.2.6.

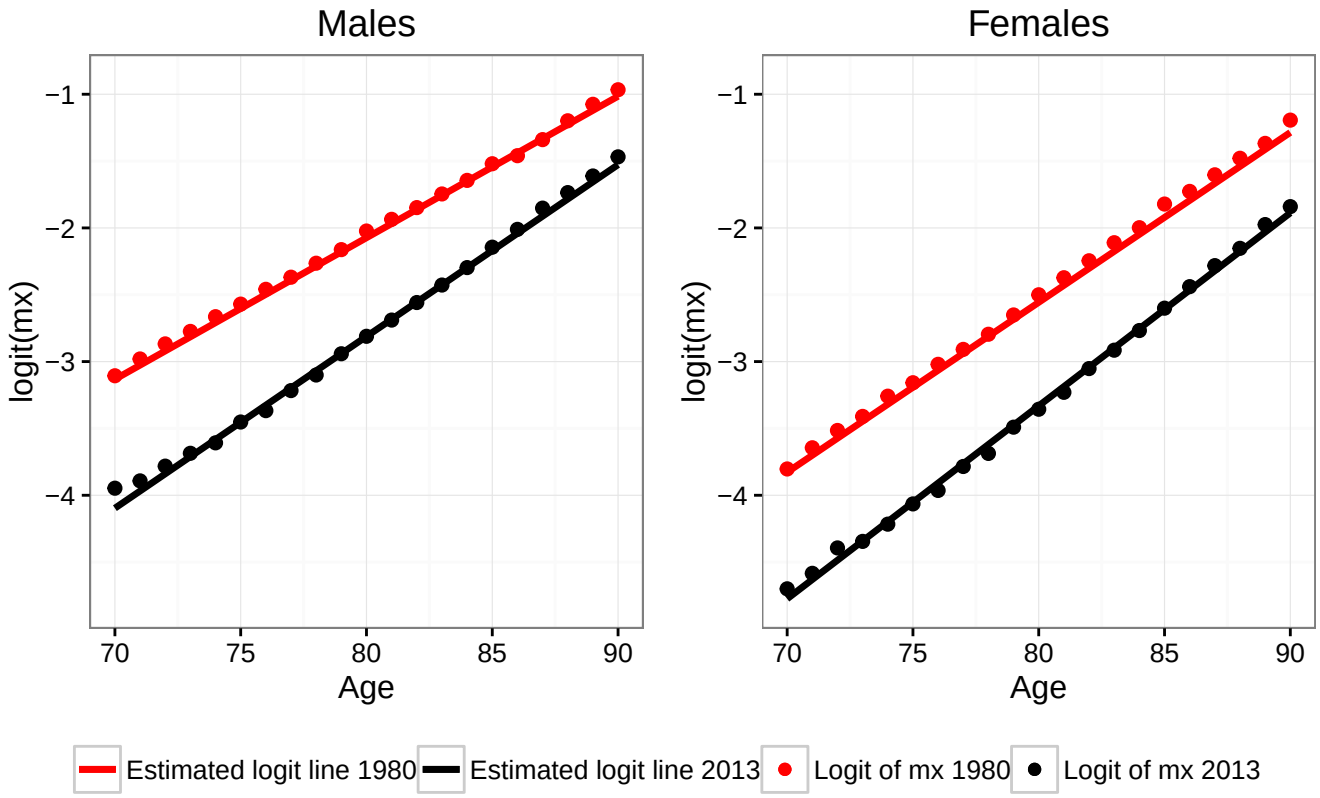


FIGURE 6.2: Dots: logit of m_x for males and females in Italy in 1980 and 2013. Lines: estimated logit line for males and females in Italy in 1980 and 2013 from the Kannisto model.

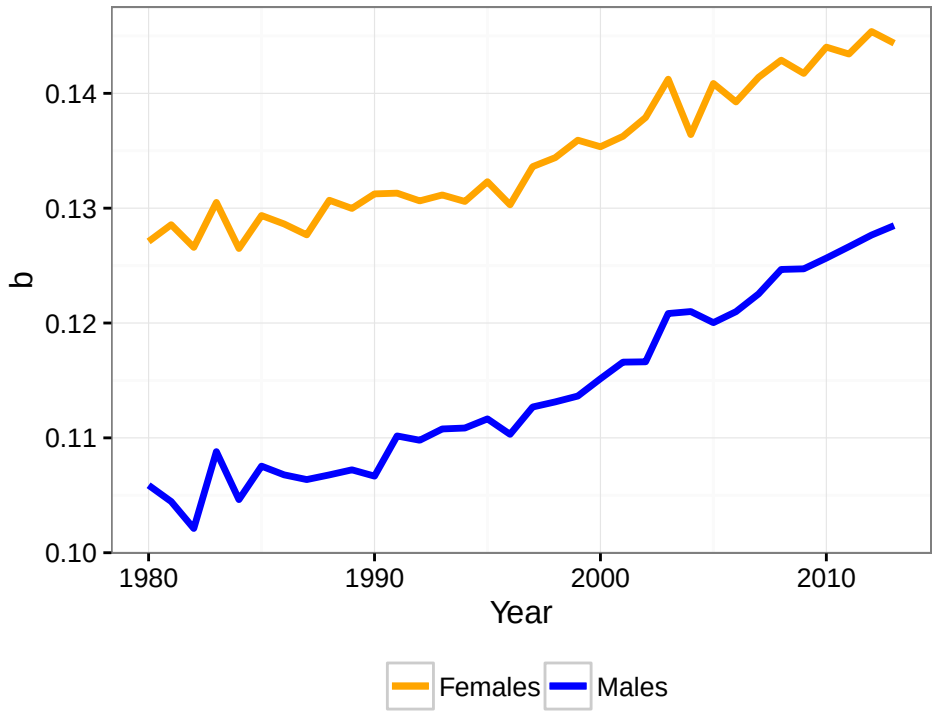


FIGURE 6.3: Trends in the b parameter of the Kannisto model for Italian males and females from 1980 to 2013.

Thirdly, as expected again, b's values increased more for males than females, confirming what was already clear from Figure 6.1 and 6.2. Precisely, b's values rose of 0.023 for males (from 0.105 in 1980 to 0.128 in 2013) and of 0.017 for females (from 0.127 in 1980 to 0.144 in 2013). Thus, if on one hand lifespan variability above the mode is smaller among females, on the other hand males are recovering their disadvantage.

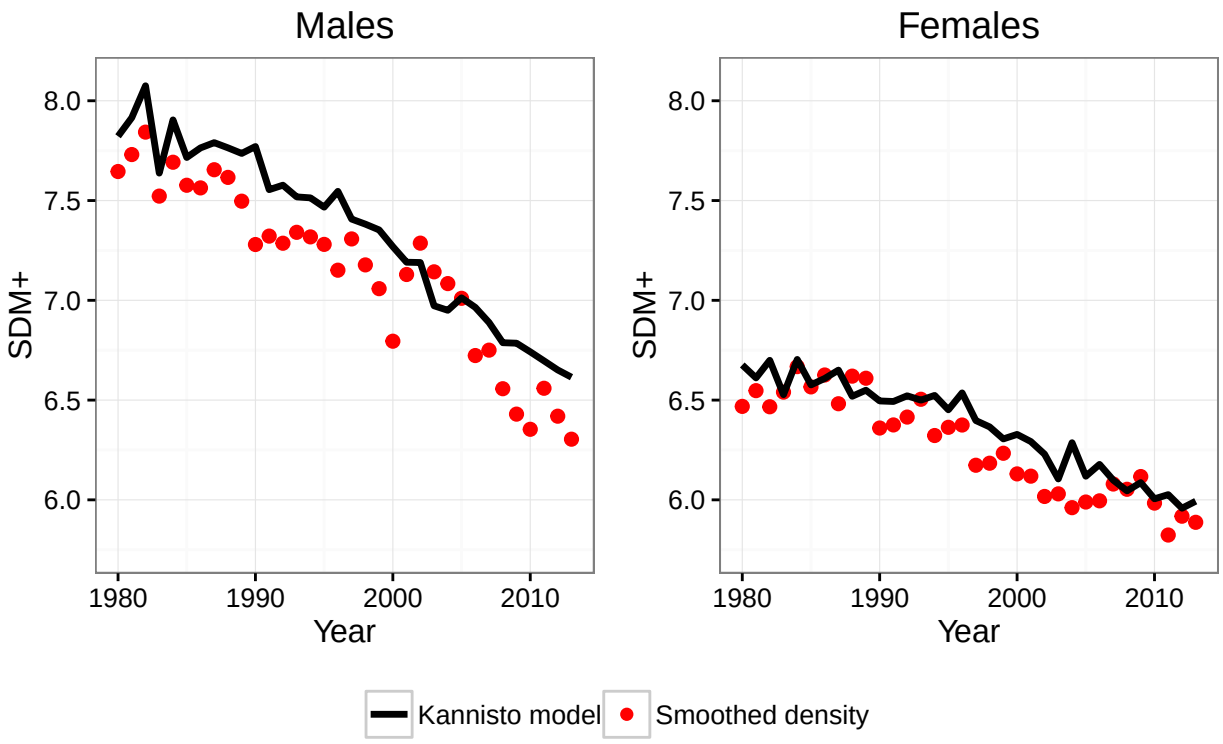


FIGURE 6.4: Standard deviation of ages at death above the mode, *SDM+*, for Italian males and females from 1980 to 2013.

Even though b's values are themselves a measure of compression above the modal age at death, in practice it is hard to evaluate lifespan variability extent only looking at them. To do that, it is better to monitor *SD(M+)* trends since their unit of measurement is in years. Figure 6.4 illustrate *SD(M+)* trends for Italian males and females from 1980 to 2013. The lines represent the estimated *SD(M+)* values obtained by Kannisto model through the application of formula 3.35. Dots, instead, represent the estimated *SD(M+)* values obtained by non-parametric approach of p-slines (i.e. smoothed density) already extensively used in Chapter 5 where lifespan variability over a wider age range was studied. *SD(M+)* estimated values do not differ a lot between the two methodologies, confirming one more time the accuracy of the simple logistic equation in modeling the data. In particular, *SD(M+)* estimated values through smoothed densities appear to be systematically lower than those estimated through the Kannisto model, but the differences are, at most, in the order of few decimal of years. The results allow to assess better what already emerged by looking at the trends of parameter b. Females higher b's values translate into smaller values of *SD(M+)* and the greater increase of b among males in the considered time interval translate into a reduction of males disadvantage as time passes by. Indeed, in 1980

SD(M+) was equal to 7.82 years for males and to 6.67 years for females, a difference of 1.15 years, while in 2013 the gender gap was reduced to 0.62 years with males having a *SD(M+)* of 6.61 years and females of 5.99 years.

6.2 Epidemiological assessment of the compression of deaths above the mode

Until this moment, old-age lifespan variability investigation has relied on all-cause mortality data. Now, to find an epidemiological interpretation of the observed old-age mortality compression, causes of death are taken into consideration. In particular, the following sections have two aims: a) Understand *which types of causes* are significantly contributing to old-age mortality compression b) Understand *in what manners* such causes are contributing to old-age mortality compression.

6.2.1 Level and slope effect hypothesis

All-cause *SD(M+)* is declining as a consequence of the rise in the rate of aging which is measured by the *b* parameter in the Kannisto model. What are the effects of cause-specific mortality on the steepening of deaths above the mode? To answer this question, one must firstly consider that causes of death differ considerably in the rate of aging. Such differentiation in the pace of mortality increase with age has been shown in some studies in the past (Horiuchi 2006; Horiuchi and Wilmoth 1997; Horiuchi et al. 2003; Horiuchi et al. 2012). In particular, causes of death such as congestive heart failure, infarctive stroke, pneumonia, influenza, septicemia, renal failure, accidental falls and ingestion accidents have reported elevated values in the rate of aging being highly related with senescent processes. On the other hand, the rate of aging seems to be much more modest in causes such as neoplasms, multiple sclerosis, acute myocardial infarction, hemorrhagic stroke, emphysema, and liver cirrhosis which are instead highly associated with some risk factors and tend to develop selectively and prematurely (Horiuchi et al., 2012).

Based on this consideration, Horiuchi and his collaborators have hypothesized two possible reasons for the observed old-age mortality compression, i.e. two possible reasons for the increase of the all-cause *b* parameter:

- Cause-specific death rates decrease proportionally at old ages, but the decrease in death rates for causes of death reporting high *b* values is more moderate than the decrease in death rates for causes of death reporting low *b* values. If so, at old ages, the proportion of deaths due to high *b* causes of death will increase. This will result into a rise of all-cause *b* even though cause-specific *b* remains constant.
- Cause-specific *b* is rising for both high *b* and low *b* causes of death, i.e. cause-specific death rates don't decrease proportionally. Of course this will result into a rise of all-cause *b*.

In order to visualize these two mechanisms, they are reported in Figure 6.5. The first it is named level effect hypothesis since it focuses on differential *levels* of decrease of cause-specific death rates. An explanation for

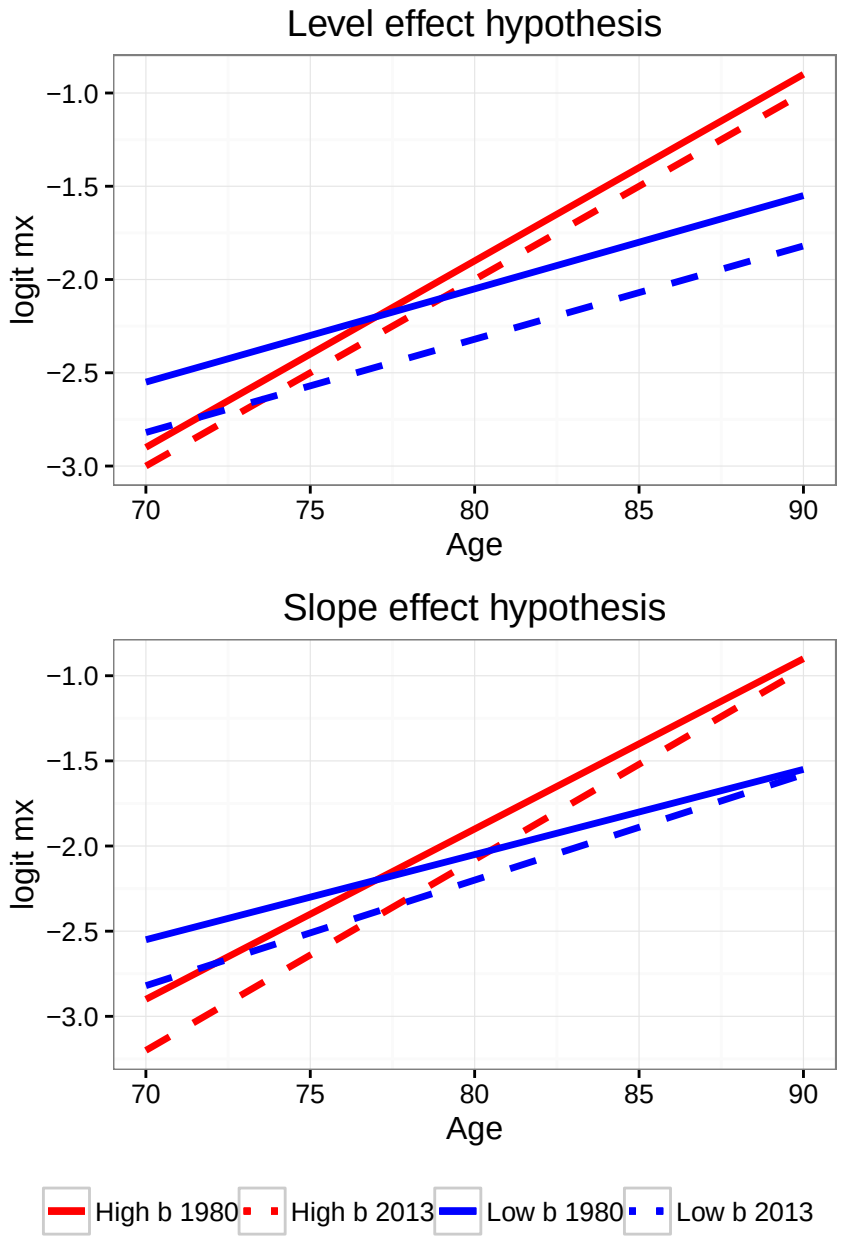


FIGURE 6.5: Cause-specific explanation of the rise in the rate of aging. Solid lines that portray an earlier period and dashed lines that portray a later period, here conventionally represented by 1980 and 2013 respectively since this is the time interval considered in the analysis of this disseratation. Adapted from Horiuchi et al., 2012.

this conjecture is that it could be difficult to slow down senescence and thus mortality of high *b* causes may decrease slower than mortality of low *b* causes. The second it is named slope effect hypothesis since it focuses on the increase of death rates with age which is measured by the logit-mortality *slope*.

In order to test these hypothesis, the changes over time of the *b* parameter of Kannisto model needs to be distinguished into cause-specific level and slope effect. As seen in Paragraph 3.2.6, the parameter b is estimated using the formula:

$$b = \frac{logit(m(90)) - logit(m(70))}{90 - 70} \tag{6.1}$$

Taking causes of death into consideration equation 6.1 can be rewritten as follow:

$$b = \frac{logit \sum_i (m_i(90)) - logit \sum_i (m_i(70))}{90 - 70} \tag{6.2}$$

Where m_i is the death rate of cause *i*.

Now, to separate the effects of changes in the level of cause-specific mortality and effects of changes in the slope of cause-specific mortality, equation 6.2 must be expressed in terms of the sum and difference of cause-specific death rates for the two considered ages:

$$s_i = logit(m_i(90)) + logit(m_i(70)) \tag{6.3}$$

$$d_i = logit(m_i(90)) - logit(m_i(70)) \tag{6.4}$$

Indeed, since $s_i/2$ is the mean logit of death rates of cause *i* in the age range, s_i can be regarded as a measure of cause-specific mortality level in the age range; and since $d_i/(90-70)$ is the logistic rate of mortality increase of cause *i* in the age range, d_i can be regarded as a measure of cause-specific rate of mortality increase. Equation 6.2 can be rewritten as:

$$b = \frac{logit \sum_i \frac{e^{(s_i+d_i)/2}}{1+e^{(s_i+d_i)/2}} - logit \sum_i \frac{e^{(s_i-d_i)/2}}{1+e^{(s_i-d_i)/2}}}{90 - 70} \tag{6.5}$$

So, *b* is a function of s_i and d_i:

$$b = f(s_1, ..., s_i, ..., s_n, d_1, ..., d_i, ..., d_n) \tag{6.6}$$

and the changes in *b* over time, using the concept of line integral (Horiuchi et al., 2008), can be decomposed as:

$$\begin{aligned} b(2013) - b(1980) = &\int_{1980}^{2013} \frac{\delta b}{\delta s_1}\frac{\delta s_1}{\delta t} + ... + \int_{1980}^{2013} \frac{\delta b}{\delta s_i}\frac{\delta s_i}{\delta t} + ... + \int_{1980}^{2013} \frac{\delta b}{\delta s_n}\frac{\delta s_n}{\delta t} + \\ &+ \int_{1980}^{2013} \frac{\delta b}{\delta d_1}\frac{\delta d_1}{\delta t} + ... + \int_{1980}^{2013} \frac{\delta b}{\delta d_i}\frac{\delta d_i}{\delta t} + ... + \int_{1980}^{2013} \frac{\delta b}{\delta d_n}\frac{\delta d_n}{\delta t} \end{aligned} \tag{6.7}$$

If the level effect hypothesis holds, the changes in b will be mainly attributable to $\int_{1980}^{2013} \frac{\delta b}{\delta s_i} \frac{\delta s_i}{\delta t}$ while if the slope effect hypothesis holds, the changes in b will be mainly attributable to $\int_{1980}^{2013} \frac{\delta b}{\delta d_i} \frac{\delta d_i}{\delta t}$. It has to be mentioned that the level and slope mechanisms can operate simultaneously, so the aim is to figure out which of the two prevails and to which extent.

6.2.2 Cause-specific rate of aging

Before presenting decomposition's results of changes in b over time, it is worth looking at cause-specific b's values and their evolution over time to assess how much causes of death differ with respect to the rate of aging and start understanding which of the two hypothesis seems to be more reasonable.

Cause-specific b's values are reported in Table 6.1, 6.2, 6.3, and 6.4. Table 6.1 and 6.2 show males' results in 1980 and 2013 respectively while Table 6.3 and 6.4 show females' results in 1980 and 2013 respectively. Inside tables, causes of death are ranked according to their rate of aging.

As expected, the rate of increase in mortality with age greatly varies from cause to cause and the results are rather coherent with previous studies. Of course, since all-cause b is higher for females than males, also cause-specific b are generally higher among females. From 1980 to 2013 b rose for all causes, a first clue suggesting that the level effect hypothesis is a wrong prediction. In either both sexes or period ill-defined causes are, by far, the one reporting the highest b. This particular group of causes gathers together all deaths for which a specific underlying cause was not found. This situation is usually typical of extremely old ages whereas death is due to several factors and no specific cause can be reported as underlying. It is not surprising, thus, that ill-defined causes have the highest rate of mortality increase with age since older is the age at death higher is the probability for a given death to be recorded as ill-defined. At the bottom of the ranking, instead, there are neoplasms. Their b's values are extremely lower than those of other causes even though a significant rise was registered as time passes by: from 0.036 in 1980 to 0.064 in 2013 for males and from 0.042 in 1980 to 0.058 in 2013 for females. Also in this case the outcome is not surprising. As it was shown in Paragraph 5.1.2 neoplasms are the cause of death reporting the smaller modal age at death a signal that this group of diseases are not highly associated with senescence process but rather related to some specific risk factors indipendent from age. Another important group of causes showing low b's values in both sexes are diseases of the digestive system, although, as for neoplasms, their rate of aging increased noticeably along the considered time interval: from 0.040 in 1980 to 0.100 in 2013 for males and from 0.065 in 1980 to 0.119 in 2013 for females. Particularly interesting is the tremendous growth of b for mental disorders in both sexes, a result in line with what found in Paragraph 4.2.3 (a considerable growth of mortality rates at old ages) and 5.1.2 (a considerable growth of the modal age at death). Finally, as expected, diseases of the circulatory and respiratory system display a quite high rate of aging being strongly related with senescent processes.

Cause of death	b
1 Ill-defined causes	0.203
2 Diseases of the skin	0.122
3 Diseases of the genetourinary system	0.114
4 Diseases of the circulatory system	0.112
5 Diseases of the respiratory system	0.104
6 External causes	0.092
7 Mental and behavioural disorders	0.084
8 Diseases of the muskoloskeletal system	0.082
9 Congenital malformations	0.073
10 Diseases of the blood	0.070
11 Infectious diseases	0.068
12 Diseases of the nervous system	0.061
13 Endocrine Nutritional and Metabolic diseases	0.048
14 Diseases of the digestive system	0.040
15 Neoplasms	0.036

TABLE 6.1: Cause-specific b for Italian males in 1980.

Cause of death	b
1 Ill-defined causes	0.193
2 Mental and behavioural disorders	0.185
3 Diseases of the skin	0.167
4 Diseases of the genetourinary system	0.158
5 Diseases of the respiratory system	0.155
6 Diseases of the circulatory system	0.151
7 Diseases of the blood	0.141
8 External causes	0.128
9 Diseases of the muskoloskeletal system	0.127
10 Endocrine Nutritional and Metabolic diseases	0.112
11 Diseases of the nervous system	0.110
12 Infectious diseases	0.100
13 Diseases of the digestive system	0.100
14 Congenital malformations	0.070
15 Neoplasms	0.064

TABLE 6.2: Cause-specific b for Italian males in 2013.

Cause of death	**b**
1 Ill-defined causes	0.219
2 Diseases of the respiratory system	0.139
3 Diseases of the skin	0.135
4 Diseases of the circulatory system	0.133
5 External causes	0.126
6 Mental and behavioural disorders	0.103
7 Infectious diseases	0.101
8 Diseases of the nervous system	0.088
9 Diseases of the blood	0.081
10 Diseases of the genetourinary system	0.078
11 Congenital malformations	0.074
12 Diseases of the muskoloskeletal system	0.071
13 Diseases of the digestive system	0.065
14 Endocrine Nutritional and Metabolic diseases	0.047
15 Neoplasms	0.042

TABLE 6.3: Cause-specific b for Italian females in 1980.

Cause of death	**b**
1 Ill-defined causes	0.217
2 Mental and behavioural disorders	0.216
3 Diseases of the skin	0.190
4 Diseases of the circulatory system	0.177
5 Diseases of the genetourinary system	0.161
6 Diseases of the respiratory system	0.157
7 External causes	0.156
8 Diseases of the blood	0.156
9 Diseases of the nervous system	0.129
10 Endocrine Nutritional and Metabolic diseases	0.127
11 Diseases of the digestive system	0.119
12 Diseases of the muskoloskeletal system	0.116
13 Infectious diseases	0.115
14 Congenital malformations	0.061
15 Neoplasms	0.058

TABLE 6.4: Cause-specific b for Italian females in 2013.

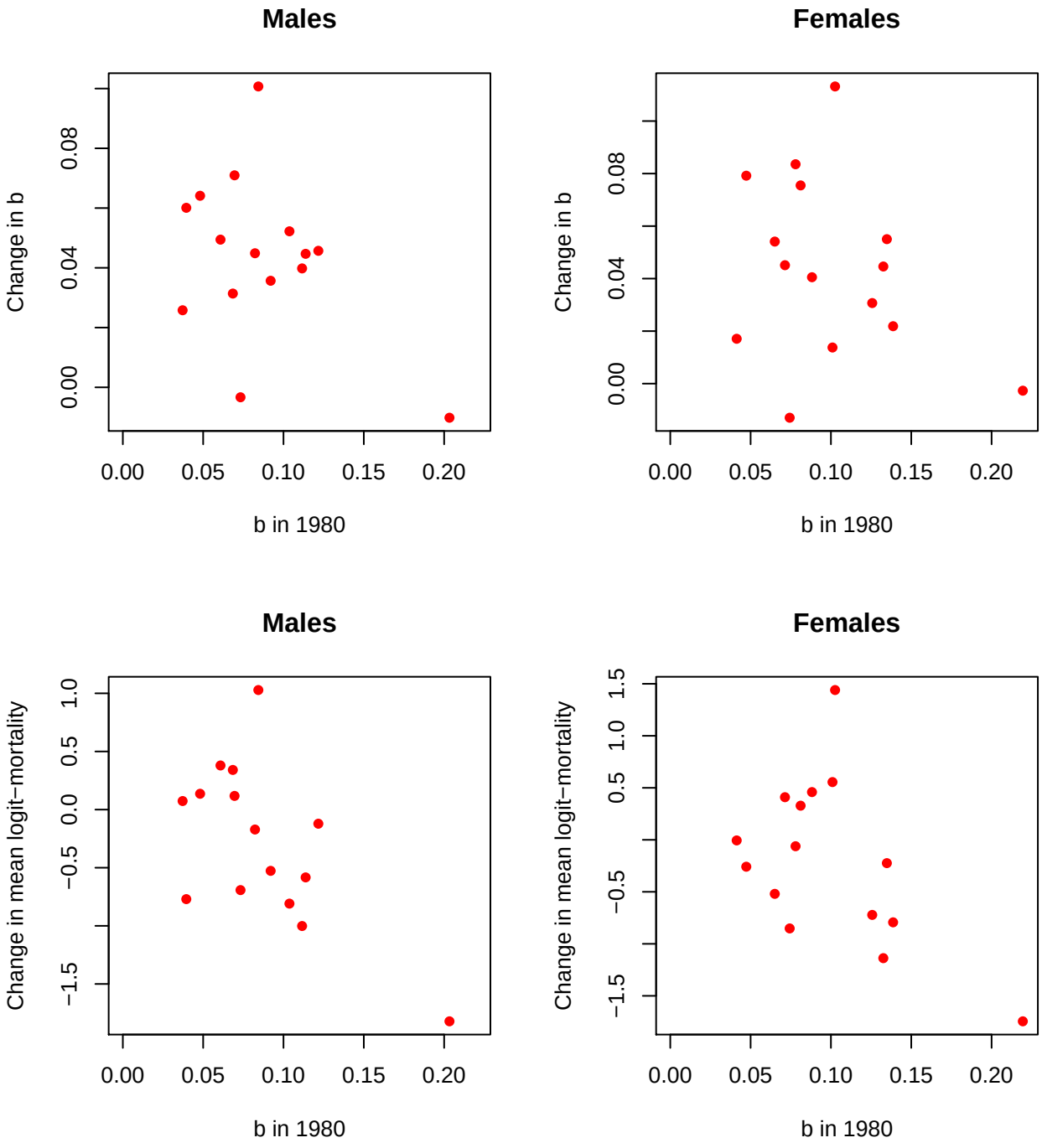


FIGURE 6.6: Top panel: relationship between cause-specific b in 1980 and change in b from 1980 to 2013 for Italian males and females. Bottom panel: relationship between cause-specific b and change in the mean logit death rate for ages 70 and 90 from 1980 to 2013 for Italian males and females.

Now, in order to have a better understanding of decomposition's result of changes in b that will be presented in next section, Figure 6.6 shows, in the top panel, the relationship between cause-specific b in 1980 and change in cause-specific b from 1980 to 2013 for males and females and, in the bottom panel, the relationship between cause-specific b and change in the mean logit death rate for ages 70 and 90 from 1980 to 2013 for males and females in the bottom panel.

In the case of the relationship between cause-specific b in 1980 and change in cause-specific b from 1980 to 2013, if the level effect hypothesis was correct, the changes in b over time should be close to 0 or anyway moderate, while in the case the relationship between cause-specific b and change in the mean logit death rate, if the level effect hypothesis was correct, there should appear a positive correlation in the scatterplot, i.e. the logit death rates for causes of death with higher b in 1980 must decrease less between 1980 and 2013 than those for causes of death with lower b in 1980. In both cases, the above described situations don't show up. In fact, the rate of aging increased considerably for almost any cause regardless of its value in 1980 and there is no positive correlation between changes in logit death rates and cause-specific b in 1980, on the contrary it seems that exactly the causes of death reporting highest rate of aging are those experiencing the major mortality improvements, a clear hint of the failure of the level effect hypothesis.

6.2.3 Cause-of-death decomposition of the rise in the rate of aging

As seen in Paragraph 6.1, estimated all-cause b increased from 0.105 in 1980 to 0.128 in 2013 for males and from 0.127 in 1980 to 0.144 in 2013 for females. In this section, in order to understand *which* type of causes are responsible for the observed compression of deaths above the mode and in *what manner*, the increment of 0.023 for males and of 0.017 for females has been decomposed into seventeen cause-specific level effects and seventeen cause-specific slope effects. The cause-of-death decomposition results confirm what found in the previous section, that is the failure of the level effect hypothesis in favor of the slope effect hypothesis.

The outcomes, multiplied by 1000 in order to be more readable, are shown in Table 6.5 (males) and 6.6 (females). In both sexes, the total level effect has negative sign (-24.337 for males and -24.578 for females), meaning that, overall, the levels of mortality of causes of death reporting high b values declined more than those of causes reporting low b values, having as a consequence the effect of lowering all-cause b. The level effect is canceled and overtaken by the total slope effect (47.303 for males and 41.894 for females) which actually led to to the registered rise in the rate of aging.

For males, causes of death with the strongest slope effect are diseases of the circulatory system and neoplasms, whose preeminence is, obviously, partially due to their considerable proportion of deaths at old ages which enable them to have a greater impact than other causes of death. Noticeable is also the slope effects of endocrine, nutritional and metabolic diseases, diseases of the respiratory system, diseases of the digestive system and external causes of death. Females show similar results. Diseases of the circulatory system are abundantly the cause with the strongest slope effect,

Cause of death	Level effect	Slope effect	Total effect
Infectious diseases	-0.023	0.425	0.401
Neoplasms	-0.902	11.114	10.211
Diseases of the blood	0.006	0.216	0.222
Endocrine Nutritional and Metabolic diseases	-0.035	2.192	2.157
Mental and behavioural disorders	0.633	1.262	1.896
Diseases of the nervous system	-0.024	1.218	1.194
Diseases of the circulatory system	-13.077	19.832	6.755
Diseases of the respiratory system	-3.236	6.019	2.783
Diseases of the digestive system	0.710	2.741	3.451
Diseases of the skin	-0.008	0.062	0.053
Diseases of the muskoloskeletal system	-0.009	0.149	0.140
Diseases of the genetourinary system	-0.718	1.270	0.552
Congenital malformations	0.010	-0.003	0.007
Ill-defined causes	-7.281	-0.504	-7.786
External causes	-0.383	1.307	0.924
Total	-24.337	47.303	22.966

TABLE 6.5: Cause-of-death decomposition of change in the logistic rate of aging (x1000) between age 70 and 90 from 1980 to 2013. Males

Cause of death	Level effect	Slope effect	Total effect
Infectious diseases	-0.047	0.188	0.141
Neoplasms	0.075	4.777	4.852
Diseases of the blood	0.012	0.291	0.303
Endocrine Nutritional and Metabolic diseases	0.308	3.880	4.188
Mental and behavioural disorders	1.183	1.746	2.929
Diseases of the nervous system	-0.045	1.159	1.114
Diseases of the circulatory system	-16.314	23.265	6.951
Diseases of the respiratory system	-2.080	1.685	-0.394
Diseases of the digestive system	0.442	2.249	2.691
Diseases of the skin	-0.021	0.109	0.088
Diseases of the muskoloskeletal system	-0.039	0.251	0.212
Diseases of the genetourinary system	-0.010	1.439	1.429
Congenital malformations	0.021	-0.046	-0.025
Ill-defined causes	-7.400	-0.145	-7.544
External causes	-0.665	1.044	0.379
Total	-24.578	41.894	17.316

TABLE 6.6: Cause-of-death decomposition of change in the logistic rate of aging (x1000) between age 70 and 90 from 1980 to 2013. Females

followed by neoplasms, endocrine nutritional and metabolic diseases, diseases of the digestive system, mental and behavioral disorders and diseases of the respiratory system. It is worth nothing, however, the reduced impact of neoplasms and diseases of the respiratory system with respect to the one of males. These results are in line with what previously found analyzing cause-specific increase in b over time. The increase in the rate of aging for neoplasms and diseases of the respiratory system from 1980 to 2013, indeed, has been much more remarkable among males than females (see Table 6.1, 6.2, 6.3 and 6.4).

Overall, cause-of-death decomposition's outcome presented in this section are pretty similar with what found by Horiuchi and his colleagues who decomposed the increase in the logistic rate of aging from 1979 to 1994 for French males and females (Horiuchi et al., 2012). Also in that case the results gave strong support to the slope effect hypothesis while the level effect appeared to act in the opposite direction, i.e. lowering all-cause b. The implications of such results are not fully clear, since they can be read in two different ways. On one hand, indeed, the results claim that all-cause b increased over time because cause-specific b increased as well. The increase in cause-specific b was observed for all types of causes of death either with high or average or low b values. It means that the logit death rate of all the considered causes has decreased more at old ages (70 in this case) than at older old ages (90 in this case). This result suggests that the observed old-age mortality compression may be an indication of an increasing resistance to further longevity extension. On the other hand, however, it also appears that the old-age mortality decline of causes of death with high b values, i.e. those causes which are considered to be more senescence related and so less preventable, is stronger than the decline registered for causes of death with medium or low b values. This finding advocates a positive prospect for further reduction of senescent mortality and consequently for further longevity extension.

6.3 Further considerations on longevity extension

A comprehensive analysis of whether humans lifespan is approaching or not a biological limit goes beyond the purpose of my dissertation. However, in the light of what found in the previous section, I would like to clarify some points about this issue.

Table 6.7 shows average death rate registered in twenty developed countries in 1950, 1980 and 2014 at ages above 95 and their rate of change from 1950 to 1980 and from 1980 to 2014. The countries included in the analysis are all those of the HMD which have available deaths count data from 1950 to 2014 [2]. Merging data from various countries gives a global understanding of mortality changes at extremely old ages and thus of how populations are aging.

As before, the outcomes highlight a situation that can be read in two ways. On one hand, the rate of mortality decrease slows down with age, i.e. the rate of aging is increasing. Undoubtedly, human mortality is finding

[2]The countries are: Australia, Austria, Belgium, Czech Republic, Denmark, England and Wales, Finland, France, Hungary, Ireland, Italy, Japan, Netherlands, Norway, Portugal, Slovakia, Spain, Sweden, Switzerland and USA.

Age	m_x 1950	m_x 1980	m_x 2014	% change 1950-1980	% change 1980-2014
95	0.399	0.351	0.270	-11.92	-23.12
96	0.425	0.378	0.298	-10.98	-21.05
97	0.451	0.405	0.328	-10.02	-18.99
98	0.477	0.433	0.360	-9.07	-16.96
99	0.503	0.461	0.393	-8.13	-14.97
100	0.529	0.490	0.427	-7.22	-13.04
101	0.554	0.519	0.461	-6.33	-11.21
102	0.580	0.548	0.496	-5.48	-9.47
103	0.605	0.576	0.531	-4.67	-7.84
104	0.629	0.604	0.566	-3.92	-6.33
105	0.653	0.632	0.601	-3.21	-4.97
106	0.676	0.658	0.634	-2.55	-3.73
107	0.698	0.684	0.666	-1.94	-2.63
108	0.719	0.709	0.697	-1.40	-1.68
109	0.739	0.732	0.726	-0.91	-0.86
110+	0.758	0.755	0.753	-0.47	-0.26

TABLE 6.7: Average death rates of twenty countries of the HDM at ages 95+ in 1950, 1980 and 2014 and their rate of change over time.

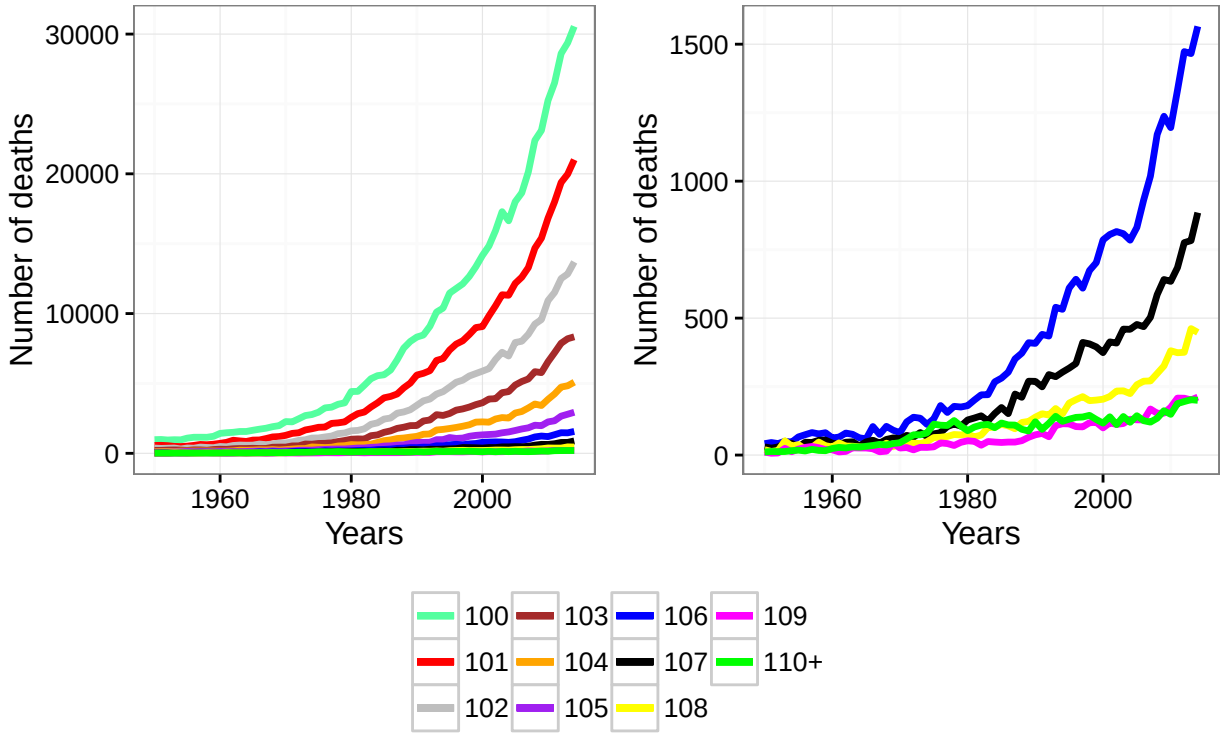


FIGURE 6.7: Number of deaths above age 100 in twenty countries of the HMD.

a kind of resistance or however a new barrier to tear down. On the other hand, even at extremely old ages, until at least age 100, the reduction of death rates in the last decades has been impressive. Moreover, the reduction was stronger from 1980 to 2013 than to 1950 to 1980 endorsing the reasoning that noticeable improvements against senescence keep going on.

Such great decline reflects into the emergence of centenerians and supercentenarians as shown in Figure 6.7. Thanks to remarkable old-age mortality decline, the cumulative number of people died at age 100 or more in the twenty countries is exponentially increasing from 1950 to 2014. Until the 1970s the number of centenarians was small and the number of supercentenarians nonexistent. Nowadays over those twenty countries, more than 30'000 deaths occur at age 100, 5'000 at age 105 and about 250 above age 110. In slightly more than 60 years longevity has been extended enormously.

At the state of the knowledge, it seems that evidences are mixed. Progress against senescence are remarkable but it is also undeniable that improvements are facing some kind of resistence with age. The number of supercentenarians is still too low and the variability of death rates above age 110 consequently too high. Thus, surely, to comprehend if further extensions are feasible it is going to be important to better assess how death rates above age 110 will behave and if new barriers may be overcome.

Chapter 7

Conclusion

The originality and innovation of my thesis does not lie into the methods used to carry out my analysis but into the analysis itself through which I obtained results that highlight new findings in a research field still scarcely explored. Indeed, while the study of variability in human longevity has a quite long tradition, the epidemiological understanding of the observed reduction in the uncertainty about the timing of death has not been pointed out yet. Only a few studies, in recent years, have started to tackle this issue analyzing the impact of the various causes of death on changes in lifespan variability.

In particular, the main purpose of my research was threefold. First, I aimed to revisit cause-specific mortality in Italy in the last decades in the light of reconstruction of coherent cause-of-death time series. Second, I aimed to point out the great existing difference between causes of death in terms of lifespan variability and I sought to explain, from a cause-of-death perspective, why variability in length of life in Italy is declining and what are the determinants of females advantage. Third, focusing the attention on old ages, I tried to understand the role played by causes of death on the steepening of the right-hand tail of the age-at-death distribution and if the observed compression of deaths above the mode may be a sign of a biological limit to human lifespan.

In this chapter I summarize my results and their implications, I outline the limitations of my analysis and describe avenues for further research.

7.1 Revisiting cause-specific mortality in Italy

The biggest problem in studying long-term series of cause-specific mortality it is represented by continuous revisitation of the classification of diseases due to the greater and greater knowledge in the field of medicine. From a purely statistical point of view, this means that a correct analysis of cause-specific mortality strictly depends on the reconstruction of coherent time series. Thus, in order to get an epidemiological understanding of mortality compression in Italy over the last decades, the first issue I had to face was to eliminate the statistical disruption in the cause-of-death series brought by the introduction of ICD-10 in 2003. Once solved this problem thanks to bridge-coding data provided by ISTAT, thus having in hands coherent data for the 1980-2013 time interval, I exploited reconstruction's results not only to study trends in lifespan variability but also to revisit cause-specific mortality in Italy.

First of all, using the modal age at death, M, as a longevity indicator, I demonstrated that lifespan is lengthening for all types of causes of death

here considered. Any cause-specific modal age at death, indeed, increased remarkably from 1980 to 2013 in either males or females and, moreover, that rise occurred at a similar pace among most causes of death. The analysis also highlights a noticeable diversity in terms of longevity extension between causes of death. In both sexes neoplasms resulted to be the cause with lower M and ill-defined causes the cause with higher M. The finding that neoplasms' age-at-death distribution reaches its peak earlier than any other is particularly intriguing and demonstrates how cancer is not a disease typical of extremely old ages tending to affect people sooner than other causes of death.

The considerable extension of longevity registered for any cause of death is the consequence of the constant decline of cause-specific mortality as revealed by the computation of Rates Of Mortality Improvements (ROMI) in Paragraph 4.2.3. Death rates decrease over time has been particularly strong at young and adult ages for all causes of death. At old and extremely old ages, although mortality declined noticeably as well, the pace of the reduction was generally slower and, furthermore, it did not regard all causes of death. Indeed, diseases like neoplasms, mental disorders, ENM diseases, diseases of the nervous system and infectious diseases has shown an increase in death rates at old and very old ages in time-period of my analysis. This demonstrates that not all causes of death are following the same age pattern of mortality change.

Diseases of the circulatory system are still the major cause of death even though the percentage of deaths attributable to them has been lowered either for males or females as a consequence of a strong reduction of their death rates at any age and in any time-period from 1980 to 2013. This situation perfectly fits the fourth stage of the health transition theory proposed by Frenk (Frenk, 1991). Computing ROMI, however, it also emerged, in the last few years, a slow down in the pace of mortality reduction for circulatory system related deaths at adult ages. In Paragraph 4.2.3 I hypothesize that such deceleration may be a sign of the economic crisis impact on health since the worsening in the economic conditions could have lead the lower class people to reduce prevention which, as well-known, plays a fundamental role in avoiding some types of circulatory diseases. A deceleration in the pace of mortality improvements at adult ages from the late 2000s, or even an increase in death rates at those ages, it is also registered for diseases of the nervous and respiratory system, external causes of death and infectious diseases. Even in this cases the reason could lie in the impact of the recent economic crisis on health. However, this is just a speculation and more years and more data are needed to evaluate better how the economic crisis is influencing the health status of the Italian population.

The considerable lowering in the percentage of deaths attributable to diseases of the circulatory system is leaving space to the emergence of others causes of death. In particular, degenerative related diseases like neoplasms, mental disorders, ENM diseases, diseases of the nervous system and infectious diseases are nowadays responsible for a greater proportion of deaths than few decades ago. The emergence of such causes of death has not been necessarily accompanied by a general worsening in their death rates. Overall, indeed, significant reduction of mortality rates have been registered for these causes of death in the time interval here considered. Such reductions, however, as mentioned above, are principally reported at

young and adult ages while old and very old ages are actually experiencing an increase in death rates. This is particularly true for diseases of the nervous system and mental disorders. Computing ROMIs for these causes, in fact, revealed, a long and often really noticeable (greater than 3%) rise in their death rates, in both sexes, for ages above 70. Thus, if on one hand young and adult ages improvements have lead to an extension of longevity also for these causes of death highly related with degenerative processes (as highlighted by the increase of their modal age at death) it is also true that a considerable rise in their old-ages death rates is registered, likely as a consequence of the population aging.

Among the above-analyzed diseases related with the degenerative process, neoplasms certainly represent the most important cause of death. As for the other, neoplasms overall burden on Italian mortality is increasing and their pattern of mortality change shows a reduction of death rates at young and adult ages (although that reduction has not been very pronounced, usually between 1% and 3%) while an increase at old and extremely old ages which stopped in the middle 1990s for females and middle 2000s for males. Such age pattern of mortality change seems to suggest that, thanks to improvements in medical knowledge and therapies, we are now able to postpone neoplasms related deaths although not to really reduce them.

7.2 Variability in human longevity and causes of death

Monitoring trends in lifespan variability is important for several factors. First of all, as every measure of variability, it implies a form of inequality. Measures of variability in length of life, indeed, can be recast as measures of inequality in the length of life: the time we live is not the same for everybody and that can be surely regarded as one of the most important forms of inequality for humans. Thus, the observed reduction of lifespan inequality (i.e. reduction of the variability of the age-at-death distribution) in developed countries is advantageous because it leads to a greater certainty in timing of death and also to a more fair distribution of deaths. Moreover, as widely highlighted in Paragraph 2.1.5 the compression of deaths in a narrower age range has many implications on other areas of human life besides death such as psychology, demography, economy and biology.

Despite the importance of studying the evolution of variability in length of life is nowadays largely recognized, there is still a conspicuous lack of understanding the epidemiological reasons behind the reduction of the uncertainty in timing of death. Exploiting data on causes of death in Italy from 1980 to 2013 my dissertation had the aim to fill this lack and get a better comprehension of the observed changes in lifespan variability. In particular, my study of variability in length of life using a cause-of-death approach have focused its attention on two distinct features of this phenomenon: the dispersion of deaths over the entire range of adult ages (measured by the standard deviation of ages at death above age 10, S_{10}) and the dispersion of deaths within old ages (measured by the standard deviation of ages at death above the mode, *SD(M+)*). In this paragraph I draw my conclusion with regard to the first feature while in the next paragraph with regard to the latter feature.

During the analyzed period, S_{10} considerably decreased for both sexes, highlighting how mortality compression is still ongoing and the hypothesis of shifting mortality not yet reasonable for Italy. Although females are currently enjoying a greater certainty in the timing of death than males, their advantage has been reduced from 1.59 years in 1980 to 1.16 years in 2013 as a result of a faster pace of compression among males (-1.45 years from 1980 to 2013) than females (-1.02 years from 1980 to 2013).

I began the epidemiological understanding of such scenario computing cause-specific S_{10} values for each sex. To my knowledge this is the first study to do that, thus allowing for the first time to assess differences between causes of death in terms of lifespan variability. The results demonstrate that cause-specific S_{10} considerably differ each other, i.e. there is a great diversity between causes of death in the way in which they distribute deaths over ages. External causes and congenital malformations results to be the group of causes with the highest lifespan variability while diseases of the respiratory and circulatory system the group of causes with the lowest lifespan variability. The difference between these two extremes is huge, about 11 years on average. With regard to gender differences, the results demonstrated that, in line with all-cause's outcome, even at the cause-specific level duration of life is less variable among females than males. The only two exceptions concern congenital malformations and neoplasms. The fact that neoplasms' S_{10} results to be higher among females than males is surely one of the most surprising and interesting findings of my dissertation. Performing a deeper analysis at the cause-of-death level [1] I also demonstrated that the gap in favour of males does not originate from a specific neoplasm, as for instance neoplasm of breast, but it is attributable to all types of cancer. In fact, lifespan variability has resulted to be higher among females for all the 21 different typologies of neoplasms for which S_{10} was computed. This outcome proves that cancers tends to generally affect females much more unequally than males. Another intriguing finding revealed by the computation of cause-specific S_{10} is that while the reduction of variability of age at death is occurring for the great majority of causes (12 out of 15) the pace of the compression differs considerably between causes. Minor causes of death such as infectious diseases, diseases of the blood, mental disorders, diseases of the nervous system and diseases of the muskoloskeletal system are those showing the greater reduction of lifespan variability while in the case of diseases of the circulatory system and neoplasms, which are the two major causes of death, it seems that mortality compression might be arrived at the end, since they are actually showing a very slow decline in variability of age at death.

My epidemiological understanding of S_{10} trends continued seeking the reasons behind 1) the observed compression of mortality registered from 1980 to 2013 by either males and females and 2) the variability gender gap in favor of females. In order to get insight on these issues, I exploited decomposition techniques that allowed me to figure out which causes and which ages are responsible for the lowering of S_{10} over time and which causes and which ages are responsible for the females advantage in terms of S_{10}.

[1] Performed only for the ICD-10 time interval (2003-2013), since the reconstruction of coherent time series was carried out exclusively for the whole group of neoplasms

Regarding the first issue, two patterns has to be distinguished, the *age* pattern and the *cause-of-death* pattern. From the point of view of the age pattern, the decomposition revealed that, in both sexes, changes in S_{10} over time are shaped by two contrasting forces. On one hand, mortality improvements registered from 1980 to 2013 before the threshold age T_{10} acted lowering lifespan variability, while mortality improvements registered from 1980 to 2013 after the threshold age T_{10} acted incrementing lifespan variability. Overall, mortality compression is occurring because the contributions before the cross-over age (-3.25 years for males and -2.54 years for females) are bigger than the contributions after the cross-over age (1.80 years for males and 1.52 years for females). This means that, as time passes by, either males or females age-at-death distribution show a greater dispersion of deaths at old ages which, however, it is overtaken by the compression of mortality occurred at young and adult ages. With regard to the cause-of-death pattern, the decomposition demonstrates that changes in S_{10} over time are mainly driven by 3 causes of death: neoplasms, diseases of the circulatory system and external causes. In particular, the decline in external causes mortality, especially between age 10-30, accounts for a wide proportion (73% for males and 49% for females) of the compression of mortality occurred in Italy from 1980 to 2013. In other words, the observed decline in lifespan variability is largely due to a reduction of transport accident related deaths among the youngs. Also neoplasms play an important role in lowering either males or females S_{10}. As seen computing ROMIs, major improvements in neoplasms mortality are registered at young and adult ages while at old ages neoplasms death rates increased until the 1990s to then start to moderately declining. Such age pattern of mortality change explains the significant contribution of neoplasms in lowering S_{10} before the threshold age and the small contribution of neoplasms in incrementing S_{10} after the threshold age. Diseases of the circulatory system, instead, produced for both sexes, an increase of the variability in the length of life in the period 1980-2013. This is due to the dramatic decrease of circulatory diseases death rates at old ages which have lead to a remarkable greater dispersion of the age-at-distribution in his latter part.

Regarding to the second issue, also here the *age* pattern and the *cause-of-death* pattern has to be distinguished. For what concern the age pattern, it looks like the one of the decompositions over time of S_{10}. In this case, females mortality advantage before the cross-over age produced the gender gap in their favor while females mortality advantage after the cross-over age allowed males to make it smaller. Thus, the gender gap exists because the contributions before T_{10} (-2.74 years in 1980 and -2.07 years in 2013) are bigger than the contributions after T_{10} (1.15 years in 1980 and 0.91 years in 2013), although the difference has been reduced from 1980 to 2013. This means that females age-at-death distribution, with respect to males age-at-death distribution, results to be more compressed at young and adult ages but also more dispersed at old ages. With regard to the cause-of-death pattern, the decomposition revealed that the gender gap in S_{10} is basically produced by external causes while the contribution of chronic and degenerative diseases, in most cases, almost balance out on either side of the threshold age. This outcome demonstrates that the reason why females are actually enjoying a greater certainty in the timing of death than males fundamentally lies in their less risky behavior during their youth years.

7.3 The compression of deaths above the mode and its theoretical implications

There are many different ways of analyzing trends in lifespan variability. Looking at how the distribution of ages at death is changing as whole, as I did in Chapter 5, is surely important because it allows understanding if, overall, inequality in the length of life has been reduced or not. However, in developed countries an higher and higher proportion of deaths is nowadays occurring at advanced ages and thus the study of variation in lifespan among survivors to older ages has become increasingly important and interesting. I investigated on time trends in old-age lifespan variability in the last substantive Chapter of my dissertation, Chapter 6, where using *SD(M+)* I examined the changes over time of the right-hand tail of the age-at-death distribution. As advocated by some scholars (Kannisto, 2001; Thatcher et al., 2010), looking at trends in *SD(M+)* not only allow to figure out if variability of deaths above the modal age is increasing or declining but also to get insights on a fascinating question: is there a limit to human lifespan? Indeed, since the reduction of lifespan variability above the mode is associated with an increase in the rate of aging, observing compression of deaths above the mode could be a sign of increasing resistance to further improvements against senescence, and thus to further longevity extension, meaning that human lifespan is approaching a biological limit. Although the major goal of the analysis conducted in Chapter 6 was to monitor changes in *SD(M+)* from 1980 to 2013 and to figure out which types causes of death and in what manners are determining those changes, the theoretical implications behind the evolution over time of *SD(M+)* also permitted me to draw some conclusions about the open question regarding the pre-determined biological limit to human lifespan.

The debate over whether there is or not a boundary to how long people can live has a long tradition and came to the fore again very recently. In 2016 Dong and his collaborators, showed that gains in survival appear to decline after age 100, and that the age at death of the world's oldest person (the French woman Jeanne Calment who died at 122) has not augmented since 1997 (Dong et al., 2016). Based on this results, the authors stated that human length of life is likely subject to natural constraints, with a maximum fixed around age 115. In response to Dong's article five groups of researchers have published, in 2017, a series of rebuttals (Brown et al, 2017; de Beer et al., 2017; Hughes and Hekimi, 2017; Lenart and Vaupel, 2017; Rozing et al., 2017). According to the authors of these rebuttals, the data are actually consistent with many different patterns and thus with many different lifespan trajectories, including either a later plateau for maximum lifespan or no plateau at all. Other criticism arose because Dongs and his colleagues included in their analysis only the oldest to die in any year which produced a small sample of deaths and thus a lot of randomnesses.

My results prove that, from 1980 to 2013, *SD(M+)* have been declining for either Italian males or females. In particular, the compression of deaths above the mode has been more pronounced for males, whose *SD(M+)* decreased of 1.21 years, than females, whose *SD(M+)* decreased of 0.68 years. The roots of such trends have to be found in the increase of the rate of aging registered by both sexes during the analyzed period and measured by the

b parameter of the Kannisto model. Exploiting the cause-of-death decomposition of b over time proposed by Horiuchi and his collaborators, I was then able to assess, from a cause-of-death perspective, the reasons behind the observed increase in the rate of aging. The result gives strong support to the slope-effect hypothesis, i.e. all-cause b is increasing because cause-specific b are increasing, and also suggests that the level-effect hypothesis was actually a wrong prediction, i.e. the causes of death with higher b, thus the causes of death more related with the senescence process and so less preventable, are those reporting more conspicuous mortality decline. This outcome indicates that longevity extension may be actually facing some resistance since for all causes of death there is a slow down in the decline of death rates with ages but also suggests that the maximum attainable lifespan may further increase since old-age mortality is still considerably declining and, moreover, the decline is more marked for senescence related causes.

At the moment, data on the age pattern of mortality change at extremely old ages seem to be readable in various way. Probably humans' lifespan has not reached its maximum yet, but likely further longevity extension will occur more slowly than in the past.

Generally speaking, opinions among gerontologist, biologist and demographer greatly differ about whether human lifespan will ever face barriers or not and whether such hypothetical barriers are due mainly to biological causes (Comfort, 1979; Fries, 1980; Fries et al., 1989; Harman, 1991) or to practical impediments (Olshansky et al., 1990). Moreover, if a limit exists, we currently do not know if it can be relaxed by unforesseable breakthroughs in slowing the process of aging itself. However, it is not just important if we will continue to live longer, but also if longevity extension would be accompanied by an extension of the period of healthy life.

7.4 Limitations and further study

In my dissertation I have benefited from having access to cause-of-death data for Italy in the time interval 1980-2013. As already said, the statistical disruptions present in the database were fixed through the reconstruction of coherent cause-specific time series, exploting the cross-classification of deaths between ICD-9 and ICD-10 performed by ISTAT. Bridge coding data, however, were available only for 17 groups of causes of death and thus only for these 17 causes it was possible to reconstruct the series. That represents a considerable limitation to the epidemiological findings of my analysis. Inside any of these 17 groups of causes of death, indeed, several types of diseases, sometimes really diverse each other, are grouped together. For instance, cerebrovascular diseases and ischaemic heart diseases, both belonging to diseases of the circulatory system, result to be very different either from a medical point of view or from an epidemiological point of view. The same reasoning can be done for neoplasms which gather together all type of cancer. Another important example of such diversity inside the same group of causes regards external causes of death which, for instance, gather together transport accident (a cause of death really typical of young ages) and accidental falls (a cause of death really typical of old

ages). Thus, in order to get a more accurate understanding of the epidemiological changes which are leading to a greater certainty about the timing of death, more detailed analysis at the cause-of-death level is needed. To improve the knowledge in this field, therefore, researcher must take into account a larger and more comprehensive set of causes of death. In my thesis a tried to overcome this problem conducting a few more detailed analysis at the cause-of-death level for the time interval 2003-2013, in which causes of death are classified according to ICD-10 all over the period and thus no statistical disruption are reported. However, this time interval is too short to get a complete epidemiological understanding of the observed mortality compression.

With regard to data limitation, I must also evidenciate that my analysis is restricted to Italian data. My reliance on the Italian cause-of-death series does not allow to make broad inferences about the epidemiological nature of changes in lifespan variability because it is not possible to assume that the Italian case it is generalizable. Thus, it would be useful if further studies would analyze additional countries in order to highlight differences and similarity across nations.

With regard to limitations in terms of measures, my study of lifespan variability relies on S_{10} and *SD(M+)*. As already widely explained throughout the dissertation, however, there are several indicators that can be used to study variability in length of life, each of them evidencing a different aspect of this phenomenon. Even though I have chosen these two measures because they fit well the aim of my work, it would have been useful to take into account additional indicators of lifespan variability, since they would have highlighted different pattern of the changes in the age-at-death distribution over time.

Finally, with regard to limitations in terms of methods, I must point out that in the cause-of-death decomposition of changes in variability of age at death between 1980-2013 I don't incorporate data from intermediate years. Taking into account data for these intermediate years would have lead to a more accurate and precise cause-of-death explanation of sex-specific trends in variability in the length of life. Moreover, in my cause-of-death analysis of old-age lifespan variability, I assume that Kannisto model holds for cause-specific mortality rates. Since cause-specific mortality at old ages generally do behave similarly respect to all-cause mortality, the assumption is reasonable. However, for some causes of death the fit of the model might be less accurate. Thus, cause-specific rate of aging estimates may result less accurate as well as the results of the cause-of-death decomposition of b over time.

Chapter 8

Bibliography

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

Barbi E., Caselli G., Yashin A. (2000), Age and time patterns of mortality by cause in Italy: A mortality surface approach, Proceedings of the XLII Scientific Meeting of the Italian Society of Statistics, Bari 9-11 June, 2004.

Beard, R.E. (1971). Some aspects of theories of mortality, cause of death analysis, forecasting and stochastic processes. In: Brass, W. (ed.). Biological Aspects of Demography. London: Taylor and Francis: 57-68.

Bergeron-Boucher M., Ebeling M., Canudas-Romo V. (2015). Decomposing changes in life expectancy: Compression versus shifting mortality. Demogr Res. 2015;33(14):391–424.

Blacklow, R.S. (2007). Actuarially speaking: an overview of life expectancy. What can we anticipate? Am J Clin Nutr 86, 1560S–1562S.

Bohk C. and Rau R. (2017). Probabilistic mortality forecasting with varying age-specific mortality improvements. Genus (2017) 73:1

Bongaarts, J. (2005). Long-range trends in adult mortality: Models and projection methods. Demography, 42(1):23–49

Bongaarts J. (2009) Trends in senescent life expectancy. Population Studies. 2009;63:203–213.

Bossier de Laroix F. (1763). Nosologia methodica sistens morborum classes, genera et species, juxta Sydenhami mentem et Botanicorum ordinem, Amsterdam, Frères De Tournes, 5 volumes

Brown, D. C., Hayward, M., Montez, J. K., Hummer, R. A., Chiu, C.-T., and Hidajat, M. M. (2012). The significance of education for mortality compression in the United States. Demography, 49, 819–840.

Brown, D.C., Stephen L. and Amy Cotton. (2003). Risk-Mitigating Beliefs, Risk Estimates and Self-Reported Speeding, Journal of Safety Research 34(2), 183–188.

Caldwell, J. C. (2001). Demographers and the study of mortality: scope, perspectives, and theory. Ann N Y Acad Sci, 954:19–34.

Camarda C.G., Pechholdova M. (2014). Assessing the presence of disruptions in cause-specific mortality series: A statistical approach. Modicod/Dimocha workshop, INED, Paris (France), October 2014. European Population Conference. Budapest (Hungary), June 2014 79th Annual Meeting of the Population Association of America. Boston (USA), May 2014.

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

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

Canudas-Romo, V. (2008). The modal age at death and the shifting mortality hypothesis. Demographic Research, 19(30):1179–1204.

Canudas-Romo, V. (2010). Three measures of longevity: Time trends and record value. Demography 47(2)299312

Caselli G. and Egidi V., (1991). A new insight into morbidity and mortality transition in Italy, Genus47(3-4) 1-29.

Caselli G. (1991), Health Transition and the Cause-Specific Mortality, in R. Schofield, D. Reher, A. Bideau (editors), The Decline of Mortality in Europe, Oxford, Clarendon Press, 68-96.

Caselli G, Meslé F, Vallin J. (2002) Epidemiologic transition theory exceptions. Genus. 2002;9:9–51.

Caswell H. (1978). A general formula for the sensitivity of population growth rate to changes in life history parameters. Theoretical Population Biology.

Caswell, H. (2006). Applications of markov chains in demography. In: Langville, A.; Stewart, W., editors. MAM 2006: An International Conference to Celebrate the 150th Anniversary of the Birth of AA Markov. Raleigh, North Carolina; Boson Books: 2006. p. 319-334.

Caswell H. (2008). Perturbation analysis of nonlinear matrix population models. Demographic Research. 2008; 18(3):59–116.

Caswell H. (2009). Stage, age and individual stochasticity in demography. Oikos. 2009; 118(12)

Caswell, H. (2010). Perturbation analysis of longevity using matrix calculus. Paper presented at the 2010 meeting of the Population Association of America; Dallas, Texas. 2010.

Caswell H. (2010). Sensitivity analysis of discrete markov chains via matrix calculus. Linear Algebra and its Applications. 2013;

Caswell H., Ouellette N. (2015). Mortality and causes of death: matrix formulation and sensitivity analysis. Presented at the European Population Conference, Mainz, August 2016.

Cheung S. L. K., Robine J.-M. and Caselli, G. (2005). Three dimensions of the survival curve: horizontalization, verticalization, and longevity extension. Demography, 42(2):243–258.

Cheung S.L.K., Robine J.-M., and Caselli, G. (2008). The use of cohort and period data to explore changes in adult longevity in low mortality countries. Genus 64(1–2) 101129.

Cheung, S. L. K. and Robine, J.-M. (2007). Increase in common longevity and the compression of mortality: the case of Japan. Popul Stud (Camb), 61(1):85–97.

Cheung S.L.K., Robine J.-M., Paccaud F., and Marazzi, A. (2009). Dissecting the compression of mortality in Switzerland, 18762005. Demographic Research 21(19) 569598

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

Comfort, A. (1979). The biology of senescence. Elsevier, New York.

Corsini, C. and Viazzo P. (1997) The decline of infant and child mortality: the European experience, 1750-1990.

Currie I.D., Durban M., and Eilers P.H.C. (2004). Smoothing and forecasting mortality rates. Statistical Modelling 4(4) 279298

de Boor, C. (1978). A practical guide to splines. Berlin: Springer.

Désesquelles A., Salvatore M. A., Frova L., Pace M., Pappagallo M., Meslé M., Egidi V. (2010). Revisiting the mortality of France and Italy with the multiple-cause-of-death approach. Demographic research, 23, 71-806.

Désesquelles A., Salvatore M. A., Pappagallo M., Frova L., Pace M., Meslé M., Egidi V. (2012). Analysing multiple causes of death: Which methods for which data? An application to the cancer-related mortality in France and Italy. European Journal of Population/Revue Européenne de Démographie, 28, 467-498.

Desesquelles A., Gamboni A., Demuru E., and the multi-cause network (2016). We die only once, but from how many causes? Population and societies, June 2016, Number 354.

Diaconu V., Ouellette N., Camarda C.G., Bourbeau R. (2016). Insight typical longevity: An analysis of the modal lifespan by leading causes of death in Canada

Dong, X, Brandon, M., Vijg, J. (2016). Evidence for a limit to human lifespan. Nature 538, 257–259 (13 October 2016)

Dorn, H.F. and Moriyama, I.M. (1964). Uses and significance of multiple cause tabulations for mortality statistics. American Journal of Public Health Nations Health 54 400-406.

Dosman, D. M., Adamowicz, W. L., and Hrudey, S. E. (2001). Socioeconomic determinants of health- and food safety-related risk perceptions. Risk Analysis, 21(2), 307–318

Dye, N. S. and Smith, D. B. (Sep., 1986). Mother love and infant death, 1750-1920. The Journal of American History, 73(2)329–353.

Eakin, T. and M. Witten. (1995). How Square Is the Survival Curve of a given Species? Experimental Gerontology, 30 (1):33-64.

Edwards, R. D. and Tuljapurkar, S. (2005). Inequality in life spans and a new perspective on mortality convergence across industrialized countries. Population and Development Review, 31(4):645–674.

Edwards, R. D. (2009). The cost of uncertain life span. Dowloaded from author's personal web page.

Edwards, R. D. (2009). Examining variance in world life spans since 1970. Presented at the Population Association of America meeting 2008.

Eilers P.H.C., Marx B.D. (1996). Flexible smoothing with B-splines and penality. Statistical science 11(2), 89-121.

Eubank, R.L. (1988). Spline smoothing and nonparametric regression. New York: Dekker.

Farr W. (1839). Letter to the Registrar General. In First annual report of the Registrar General, London.

Finch, C.E. and Kirkwood, T.B.L. (2000). Chance, development, and aging. Oxford University Press.

Frenk J, Bobadilla JL, Stern C, Frejka T, Lozano R. (1991). Elements for a theory of the health transition. Health Transit Rev. 1991 Apr;1(1)21–38.

Fries, J.F. (1980). Aging, Natural Death, and the Compression of Morbidity. New England Journal of Medicine 303(3):130-35.

Fries, J.F., Green, L.W., Levine S. (1989). Health promotion and the compression of morbidity. Lancet. 1989 Mar 4;1(8636):481-3.

Frova L., Pace M., Pappagallo M., Grippo F. (2010). Analisi del bridge coding ICD-9 ICD-10 per le statistiche di mortalità per causa in Italia. ISTAT

Gakidou, E. E., Murray, C. J., and Frenk, J. (2000). Defining and measuring health inequality: an approach based on the distribution of health expectancy. Bull World Health Organ, 78(1):42–54.

Gavrilov, L.A. and Gavrilova, N.S. (1991). The biology of life span: a quantitative approach. Chur: Harwood Academic Publishers.

Geddes M., Franceschi S., Barchielli A., Falcini F., Carli S., Cocconi G., Conti E., Crosignani P., Gaf L., Giarelli L., Vercelli M. and Zanetti R.

(1994). Kaposi's sarcoma in Italy before and after the AIDS epidemic. British Journal of Cancer 69, 333–336

Gillespie, D. O., Trotter, M. V., and Tuljapurkar, S. D. (2014). Divergence in age patterns of mortality change drives international divergence in lifespan inequality. Demography, 51, 1003–1017.

Go G., Brustrom E., Lynch F., Aldwin M. (1995). Ethnic Trends in Survival Curves and Mortality. The Gerontologist. 1995;35:318–326

Greenspan D. and Greenspan JS (1996). HIV-related oral diseases. The Lancet, Volume 348, Issue 9029, 14 September 1996, Pages 729-733

Harman, D (1991). The aging process: major risk factor for diseases and health. Proc Natl Acad Sci U S A. 1991 Jun 15; 88(12): 5360–5363.

Hill, G. (1993). The entropy of the survival curve: An alternative measure. Canadian Studies in Population 20(1): 43–57.

Horiuchi, S., Coale, A.J. (1990). Age patterns of mortality for older women: an analysis using age-specific rate of mortality change with age. Mathematical Population Studies 2(4):245-267.

Horiuchi, S., Wilmoth, J.R. (1997). Age patterns of the life-table aging rate for major causes of death in Japan, 1951-1990. Journal of Gerontology: Biological Sciences 52A:B67-B77.

Horiuchi, S. (2003a). Interspecies differences in the life span distribution: humans versus invertebrates. In: Carey, J.R. and Tuljapurkar, S. (eds.). Population and Development Review 29(Supplement): 127-151.

Horiuchi, S., Finch, C., Meslé, F., and Vallin, J. (2003b). Differential patterns of age-related mortality increase in middle age and old age. Journal of Gerontology: Biological Sciences, 58A(6):495-507.

Horiuchi, S. (2006). Causes of death among the oldest-old: Distributions and age variations. In: Jean-Marie Robine, Eileen Crimmins, Shiro Horiuchi and Yi Zeng, eds., Human Longevity, Individual Life Duration, and the Growth of the Oldest-Old Population. Springer, pp.215-235.

Horiuchi S., Wilmouth J.R., Pletcher S.D. (2008). A decomposition method based on a model of continuous change, Demography , 2008, vol. 45 (pg. 785-801)

Horiuchi, S., Cheung S., Robine JM. (2012). Cause-of-death decomposition of old-age mortality compression in France, 1979-1994.

Huges, B.G., Hekimi, S. (2017). Many possible lifespan trajectories. Nature Communications Arising, 2017

Human Cause-Of-Death Database. (2017). Institut National d'Estudes démographiques (INED), Paris (France) and Max Planck Institute for Demographic Research (MPIDR), Rostock, (Germany).
Available at www.causesofdeath.org

Human Mortality Database. (2017). University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (MPIDR), Rostock (Germany). Available at www.mortality.org

Hurd, M. D. and McGarry, K. (2002). The predictive validity of subjective probabilities of survival. The Economic Journal, 112(482)966–985.

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

Jung A.C. and Douglas S. (1998). Diagnosing HIV-related diseases. Journal of general internal medicine. Volume 13, Issue 2 February 1998 Pages 131–136

Kannisto, V., Lauritsen, J., Thatcher, A. R., and Vaupel, J. W. (1994). Reductions in mortality at advanced ages: Several decades of evidence from 27 countries. Population and Development Review, 20(4) 793–810.

Kannisto, V. (1996). The advancing frontier of survival: life tables for old age. Odense University Press, Odense.

Kannisto, V. (2000). Measuring the compression of mortality. Demogr Res, 3[24]

Kannisto, V. (2001). Mode and dispersion of the length of life. Population: An English Selection, 13(1):159–171

Keyfitz, N. (1977). Applied Mathematical Demography. New York, NY: John Wiley and Sons.

Lee, R. (2003). The demographic transition: Three centuries of fundamental change. Journal of Economic Perspectives, 17(4):167–190.

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

Lenart, A., Vaupel, J.W. (2017). Questionable evidence for a limit to human lifespan. Nature Communications Arising, 2017.

Lexis, W. (1878). Zur Theorie der Massenerscheinungen in der menschlichen Gesellschaft. as quoted by Porter T.M., 1986, The Rise of Statistical Thinking, 1820-1900.

Liu, Jin-Tan and Chee-Ruey Hsieh. (1995). Risk Perception and Smoking Behaviour, Empirical Evidence from Taiwan, Journal of Risk and Uncertainty 11(2), 139–157.

Longo B., Camoni L., Boros S., Suligoi B. (2008). Increasing proportion of AIDS diagnoses among older adults in Italy AIDS Patient Care STDS, 22 (2008), pp. 365-371

Lundborg, Petter and Henrik Andersson. (2006). Gender, Risk Perceptions and Smoking Behaviour, Mimeo, Lund University Centre for Health Economics (LUCHE), Lund, Sweden.

Lundborg, Petter and Björn Lindgren. (2004). Do They Know What They are Doing? Risk Perceptions and Smoking Behavior Among Swedish Teenagers, Journal of Risk and Uncertainty 28(3), 261–286.

Lynch, S.M. and Brown, J.S. (2001). Reconsidering mortality compression and deceleration: An alternative model of mortality rates. Demography 38(1): 79–95.

Lynch S, Brown S, Harmsen G. (2003). Black-white differences in Mortality Compression and Deceleration and the Mortality Crossover Reconsidered. Research on Aging. 2003;25(5):456–483

Manton, K.G. and Tolley, H.D. (1991). Rectangularization of the survival curve. Journal of Aging and Health 3(2): 172–193.

Mesle, F. (1999). Classifying causes of death according to an aetiological axis. Population Studies, 53(1):97–105.

Mesle, F. and Vallin, J. (1981). Methodologies for the collection and analysis of mortality data, chapter The problem of studying mortality patterns by cause over a long period of time: an example from France, 1925 to 1978, pages 449–492. Ordina Editions. Proceedings of a Seminar at Dakar, Senegal July 7-10, 1981.

Mesle, F. and Vallin, J. (1996). Reconstructing long-term series of causes of death. Historical Methods, 29(2)72

Mesle, F., and J. Vallin. (2003). « Increase in life expectancy and concentration of ages at death », dans J.-M. Robine et al. Determining Health Expectancies. Chichester, John Wiley and Sons 13-33.

Mesle F, Vallin J. (2008). The effect of ICD-10 on continuity in cause-of-death statistics. The example of France. Population 2008, 63 383-396.

Mirowsky, J. (1999). Subjective life expectancy in the US: correspondence to actuarial estimates by age, sex and race. Social Science and Medicine, 49(7)967–979.

Myers, G.C and K.G. Menton. (1984a). Compression of Mortality Myth or Reality? The Gerontologist 24(4):346-53.

Myers, G. C. and Manton, K. G. (1984b). Recent changes in the U.S. age at death distribution: further observations. Gerontologist, 24(6):572–575.

Murray, C.J., Kulkarni, S.C., Michaud, C., Tomijima, N., Bulzacchelli, M.T., Iandiorio, T.J., and Ezzati, M. (2006). Eight Americas: Investigating mortality disparities across races, counties, and race-counties in the United States. PLoS Medicine 3(9): 1513–1524.

Nagnur, D. (1986). Rectangularization of the survival curve and entropy: The Canadian experience, 1921-1981. Canadian Studies in Population 13(1): 83–102.

Nau, C., and Firebaugh, G. (2012). A new method for determining why length of life is more unequal in some populations than in others. Demography, 49, 1207–1230.

Nau, C., and Firebaugh, G. (2014). Why lifespan are more variable among blacks than among whites in the United States. Demography. 2014 Dec; 51(6): 2025–2045.

Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalized linear models. Journal of the Royal Statistical Society. Series A 135(3): 370–384.

Notestein, F. (1945). Food for the World, chapter Population–The Long View., pages 36–57. University of Chicago Press.

Nusselder, W. J. and Mackenbach, J. P. (1996). Rectangularization of the survival curve in the Netherlands, 1950-1992. Gerontologist, 36(6):773–782.

Olshansky, S.J., Carnes, B.A., Cassel, C. (1990). In search of Methuselah: estimating the upper limits to human longevity. Science. 1990 Nov 2;250(4981):634-40.

Oeppen, J. and Vaupel, J. W. (2002). Broken limits to life expectancy. Science, 296(5570):1029–1031.

Omran, A. R. (1971). The epidemiologic transition: A theory of the epidemiology of population change. The Milbank Memorial Fund Quarterly, 49(4)509–538.

Olshansky, S. J. and Ault, A. B. (1986). The fourth stage of the epidemiologic transition: The age of delayed degenerative diseases. The Milbank Quarterly, 64(3):355–391.

Ouellette, N. and Bourbeau, R. (2011). Changes in the age-at-death distribution in four low mortality countries: A nonparametric approach. Demographic Research, 25, article 19, 595–628.

Paccaud, F., Pinto, C.S., Marazzi, A., and Mili, J. (1998). Age at death and rectangularization of the survival curve: Trends in Switzerland, 1969-1994. Journal of Epidemiology and Community Health 52(7): 412–415.

Pampel, F. C. (2002). Declining sex differences in mortality from lung cancer in high-income nations. Demography, 40(1):45–65.

Pappas, G., Queen, S., Hadden, W., and Fisher, G. (1993). The increasing disparity in mortality between socioeconomic groups in the United States, 1960 and 1986. N Engl J Med, 329(2):103–109.

Pearl, R. ([1923] 1940). Introduction to Medical Biometry and Statistics. Saunders, Philadelphia, 3rd edition.

Peltzman, S. (2009). Mortality inequality. Journal of Economic Perspectives, 23(4):175–90.

Phillips, W. (1954). A basic curve of death. Journal of the Institute of Actuaries 80: 289-325.

Pollard, J.H. (1991). Fun with Gompertz. Genus XLVII(1-2): 1-20.

Preston P.H., Heuveline P., and Guillot M. (2010). Demography: Measuring and Modeling Population Processes

Riley, J. C. (2005). The timing and pace of health transitions around the world. Population and Development Review, 31(4):741–764.

Riffe T. (2011). R package DecompHoriuchi

Robine, J.-M. (2001). Redefining the stages of the epidemiological transition by a study of the dispersion of life spans: The case of France. Population: An English Selection, 13(1):173–193.

Robine, J.M., Saito, Y., and Jagger, C. (2003a). The emergence of extremely old people: the case of Japan. Experimental Gerontology 37(7): 739-743.

Robine, J.M. (2003b). The Encyclopedia of Population, volume 2, chapter Epidemiological Transition, pages 307–310. Macmillan Reference USA, New York.

Robine, J.-M., Romieu, I., and Michel, J.-P. (2003c). Determining Health Expectancies, chapter 4. Trends in Health Expectancies, pages 75–101. John Wiley and Sons, Hoboken, NJ.

Robine, J.M., Cheung S.L.K., Thatcher, A.R., and Horiuchi, S. (2006). What can be learnt by studying the adult modal age at death? PAA paper, Population Association of America Annual Meeting, Los Angeles, California. U.S.A. March 30 – April 1, 2006.

Robine, J.-M. (2008). Between compression and shifting mortality the longevity revolution. In European Papers on the New Welfare, number 9, The Turin Conference on the New Welfare. The Risk Institute.

Rothenberg, R., H.R. et al (1991). "Population Aging Patterns: The Expansion of Mortality." Journal of Gerontology: Social Sciences 46(2) 866-70

Savage, Ian. (1993). Demographic Influences on Risk Perceptions, Risk Analysis 13(4), 413–420.

Scheper-Hughes, N. (1992). Death Without Weeping: The Violence of Everyday Life in Brazil. University of California Press, Berkeley, CA.

Shkolnikov V., Mesle F., Vallin J. (1996). Health crisis in Russia. II. Changes in causes of death: a comparison with France and England and Wales (1970 to 1993).Population8155189

Shkolnikov, V., Andreev, E., and Begun, A. Z. (2003). Gini coefficient as a life table function: Computation from discrete data, decomposition of differences and empirical examples. Demographic Research, 8(11):305–358.

Silverman, B.W. (1986). Density estimation for statistics and data analysis. London: Chapman and Hall.

Staetsky, L. (2009). Diverging trends in female old-age mortality: A reappraisal. Demographic Research 21(30):885-914. doi:10.4054/DemRes .2009.21.30.

Strehler, B.L. and Mildvan, A.S. (1960). General theory of mortality and aging. Science 132(3418): 14-21. doi:10.1126/science.132.3418.14.

Sunstein C. (2002). The perception of risk. University of Chicago Law School.

Thatcher, A.R. (1999a). The demography of centenarians in England and Wales. Population Trends 162 (1): 5-43.

Thatcher, A.R. (1999b). The long-term pattern of adult mortality and the highest attained age. Journal of the Royal Statistical Society (A) 162 (Part 1): 5-43.

Thatcher, A. R., Cheung, S. L. K., Horiuchi, S., and Robine, J.-M. (2010). The compression of deaths above the mode. Population Association of America. Presented at the Population Association of America meeting 2008.

Thatcher, A. R., Kannisto, V., and Vaupel, J. W. (1998). The Force of Mortality at Ages 80 to 120. Odense University Press, Odense, Denmark.

Thompson, W.S. (1929) : Population, American Journal of Sociology, 34(6), 959-75.

Tschachler E., Bergstresser P, Stingl G. (1996). HIV-related skin diseases. The Lancet, Volume 348, Issue 9028, 7 September 1996, Pages 659-663

Tuljapurkar, S. and Edwards, R. D. (2009). Variance in death and its implications for modeling and forecasting mortality. Technical Report 15288, National Bureau of Economic Research.

Upton, A. (1977). Handbook of the Biology of Aging, chapter Pathology, pages 513–35. Van Nostrand Reinhold, New York.

Vallin, J. and Mesle, F. (1988). Les causes de deces en France de 1925 a 1978. INED, PUF, Paris. (Travaux et Documents, Cahier 115).

Vallin, J. and Meslé, F. (2001) Trends in mortality in Europe since 1950: Age-, sex- and cause-specific mortality. In: Vallin J, Meslé F, Valkonen T, editors. Trends in Mortality and Differential Mortality. Strasbourg, France: Council of Europe; 2001. pp. 131–186.

Van Raalte, A. A. (2011). Lifespan variation: Methods, trends and the role of socioeconomic inequality.

Van Raalte, A. A., A. E. Kunst, P. Deboosere, M. Leinsalu, O. Lundberg, P. Martikainen, B. H. Strand, B. Artnik, B. Wojtyniak and J. P. Mackenbach (2011). More variation in lifespan in lower educated groups: evidence from 10 European countries. International Journal of Epidemiology

Van Raalte, A. A., Martikainen, P., and Myrskylä, M. (2014). Lifespan variation by occupational class: Compression or stagnation over time? Demography, 51, 73–95

Van Raalte, A. A., Caswell H. (2012). Perturbation analysis of indices of lifespan. MPIDR working paper WP 2012-004 January 2012

Vaupel, J.W., Manton, K.G., and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography 16(3): 439-54.

Vaupel J.W. (1997). The remarkable improvements in survuval at older ages. Philosophical transaction of the Royal Society B, 352(1363), 1799-1804.

Wilmoth, J. R. and Horiuchi, S. (1999). Rectangularization revisited: variability of age at death within human populations. Demography, 36(4) 475–495.

Wilmoth, J. R. and Lundstrom, H. (1996). Extreme longevity in five countries. European Journal of Population/Revue europeenne de Demographie, 12(1):63–93.

World Health Organization (2007). Case definition of HIV for surveillance and revised clinical staging immunological classification of HIV-related diseases in adult and children.

Yashin, A.I., Begun, A.S., Boiko, S.I., Ukrainseva, S.V., and Oeppen, J. (2001). The New Trends in Survival Improvement Require a Revision of Traditional Gerontological Concepts. Experimental Gerontology 37(1): 157-167.

Zhang, Z. and Vaupel, J. (2009). The age separating early deaths from late deaths. Demographic Research, 20(29):721–730.

Zureick, S. (2010). Certainty in timing of death: A new analysis of shifting mortality and life span disparity. University of California at Berkeley.

Appendix A

A.1 Intermediate classification of causes of death

The intermediate classification of causes of death (104 causes), proposed by the Human Cause-of-Death database has been used in Paragraph 4.2.2 and 5.1.3 for a more detailed analysis at the cause-of-death level than those carried out using the 17 groups of causes classification.

Cause of death	**ICD-10 code**
1 Other specified intestinal infections	A00-A08
2 Diarrhoea and gastroenteritis of presumed infectious origin	A09
3 Tubercolosis	A15-A19, B90
4 Septicemia	A40-A41
5 Other bacterial diseases	A20-A28, A30-A39 A42-A44, A46, A48-A49
6 HIV diseases	B20-B24
7 Viral hepatitis	B15-B19
8 Other viral diseases	A80-A89, B00-B09, B25-B34
9 Other and unspecified infectious diseases	A50-A75, A77-A79, A90-A99, B35-B60, B64-B89, B91, B92, B94-B97, B99
10 Malignant neoplasms of oral cavity	C00-C14
11 Malignant neoplasms of oeshophagus	C15
12 Malignant neoplasms of stomach	C16
13 Malignant neoplasms of colon	C18
14 Malignant neoplasms of rectum	C19-C21
15 Malignant neoplasms of liver	C22
16 Malignant neoplasms of pancreas	C25
17 Other malignant neoplasms of the digestive system	C17,C23-C24,C26
18 Malignant neoplasms of larynx	C32
19 Malignant neoplasms of lung	C33-C34
20 Malignant neoplasms of skin	C43-C44
21 Malignant neoplasms of breast	C50
22 Malignant neoplasms of cervix uteri	C53
23 Malignant neoplasms of uterus	C54-C55
24 Malignant neoplasms of ovary	C56

TABLE A.1: Cause-of-death intermediate classification.

Cause of death	**ICD-10 code**
25 Malignant neoplasms of prostate	C61
26 Malignant neoplasms of other genital organs	C51-C52,C57-C58,C60
27 Malignant neoplasms of kidney	C67
28 Malignant neoplasms of bladder	C64-C66, C68
29 Malignant neoplasms of nervous system	C70-C72
30 Leucemia	C91-C95
31 Other malignant neoplasms of lymphoid, hematopoietic and related tissue	C81-90,C96
32 Malignant neoplasms of indipendent multiple sites	C97
33 Other neoplasms	C30-C31, C37-C41, C45-C49,C73-C80
34 Benign neoplasms and neoplasms of uncertain or unknown behavior	D00-D48
35 Diabetes mellitus	E10-E14
36 Malnutrition	E40-E46
37 Other endocrinologic and metabolic diseases	E00-E07,E15-E16, E20-E35,E50-E68 E70-E90
38 Blood diseases	D50-D59
39 Dementia	F01,F03
40 Alcohol abuse	F10
41 Drug abuse	F11-F19
42 Other mental disorders	F04-F09,F20-F99
43 Systemic atrophies and demyelinating diseases of the central nervous system	G10-G12, G35-G37
44 Parkinson's diseases	G20-G25
45 Alzheimer	G30-G31
46 Epilepsy	G40-G41
47 Other diseases of the nervous system	G00-G09,G43-G44, G47-G83,G90-G99 H00-H95
48 Rheumatic heart diseases	I00-I09
49 Essential hypertension	I10
50 Hypertensive diseases	I11-I15
51 Acute myocardial infarction	I21-I23
52 Atherosclerotic cardiovascular and heart diseases	I25.0,I25.1
53 Other IHD	I20,I24,I25.2-I25.9
54 Pulmonary heart diseases	I26-I28
55 Non rheumatic valve disorders	I34-I38
56 Cardiac arrest	I46
57 Heart failure	I50
58 Other heart diseases	I30-I33,I40-I45 I47-I49,I51
59 Intracranial haemorrhage	I60-I62
60 Cerebral infarction, occlusion, and stenosis	I63, I65-I66
61 Other cerebrovascular diseases	G45,I64,I67
62 Sequelae of cerebrovascular disease	I69

TABLE A.2: Cause-of-death intermediate classification. (Continued)

Cause of death	ICD-10 code
63 Diseases of arteries, arterioles and capillaries	I70-I78
64 Other circulatory diseases	I80-I99
65 Influenza	J09-J11
66 Pneumonia	J12-J18
67 Other acute respiratory infection	J00-J06, J20-J22, U04
68 Asthma	J45-J46
69 Other chronic obstructive pulmonary disease	J40-J44, J47
70 Pneumonitis due to solids and liquids	J69
71 Pneumoconioses and chemical effects	J60-J68, J70
72 Other respiratory diseases, principally affecting the interstitium	J80-J84
73 Other diseases of the respiratory system	J30-J39, J85-J98
74 Gastric and duodenal ulcer	K25-K28
75 Hernia	K40-K46
76 Enteritis, colitis and other intestinal diseases	K35-K38, K50-K63
77 Alcoholic cirrhosis of liver	K70
78 Other cirrhoses of liver	K74
79 Other diseases of liver	K71-K73, K75, K76
80 Cholelithiasis and other disorders of biliary tracts	K80-K83
81 Diseases of pancreas	K85-K86
82 Other digestive diseases	K00-K22, K29-K31, K65-K66, K90-K92
83 Diseases of skin and subcutaneous tissue	L00-L98
84 Diseases of the musculoskeletal system	M00-M99
85 Renal tubulo-interstitial diseases	N00-N15
86 Renal failure	N17-N19
87 Other diseases of urinary system	N20-N36, N39
88 Diseases of genital organs	N40-N99
89 Complications of pregnancy, childbirth, and puerperium	O00-O99
90 Certain conditions originating in the perinatal period	P00-P96
91 Congenital malformations, deformations, and chromosomal abnormalities	Q00-Q99
92 Sudden infant death syndrome	R95
93 Transport accident	V01-V99
94 Accidental falls	W00-W19
95 Accidental drowning and submersion	W65-W74
96 Accidental exposure to smoke, fire and flames	X00-X09
97 Accidental poisoning by alcohol	X45
98 Accidental poisoning by other substance	X40-X44, X46-X49
99 Other accidental threats to breathing	W75-W84
100 Suicide and self-inflicted injury	X60-X84
101 Assault	X85-Y09, Y35-Y36
102 Event of undetermined intent	Y10-Y34
103 Complications of medical and surgical care	Y40-Y84
104 Other accidents and late effects of accidents	W20-W64, W85-W99 X10-X39, X50-X59 Y85-Y91, Y95-Y98

TABLE A.3: Cause-of-death intermediate classification. (Continued)

A.2 Correspondence table

Correspondence table contains the cross-classification of deaths between ICD-9 and ICD-10 and represents the starting point for the reconstruction of coherent cause-specific time series. Two correspondence tables have been used in this dissertation, one for deaths occurred in the first year of life and one for deaths above age 1.

A.2.1 Correspondence table of deaths in the first year of life

	Inf.	Neo.	Blood	ENM	Ment.	Nerv.	Circ.	Resp	Dig.
Infectious	23	0	0	0	0	0	0	0	0
Neoplasms	0	3	0	0	0	0	0	0	0
Blood	0	0	5	0	0	0	0	0	0
ENM	0	0	0	21	0	0	0	0	0
Mental	0	0	0	0	0	0	0	0	0
Nervous	0	0	0	0	0	33	0	0	0
Circolatory	0	0	0	0	0	0	31	0	0
Respiratory	0	0	0	0	0	0	2	16	0
Digestive	0	0	0	0	0	0	0	0	9
Skin	0	0	0	0	0	0	0	0	0
Muskoloskeletal	0	0	0	0	0	0	0	0	0
Genetourinary	0	0	0	0	0	0	0	0	0
Perinatal	7	0	1	1	0	4	13	5	8
Congenital	0	1	1	1	0	3	12	3	7
Ill-defined	0	0	0	0	0	0	6	3	0
External	0	0	0	0	0	0	0	0	0
Total-10	30	11	7	23	0	40	64	27	24

TABLE A.4: Cross-classification of deaths between ICD-9 and ICD-10 at age 0.

	Skin	Mus.	Gen.	Per.	Cong.	Ill	Ext.	Tot-9
Infectious	0	0	0	6	1	0	0	30
Neoplasms	0	0	0	1	0	0	0	11
Blood	0	0	0	2	0	0	0	7
ENM	0	0	0	4	0	0	0	25
Mental	0	0	0	0	0	0	0	0
Nervous	0	0	0	2	3	0	0	38
Circolatory	0	0	0	8	1	0	0	40
Respiratory	0	0	0	11	1	0	0	30
Digestive	0	0	0	0	1	0	0	10
Skin	0	0	0	1	0	0	0	1
Muskoloskeletal	0	1	0	0	0	0	0	1
Genetourinary	0	4	0	5	1	0	0	10
Perinatal	0	0	0	1072	29	12	0	1152
Congenital	0	3	0	61	576	0	0	668
Ill-defined	0	0	0	12	0	40	0	61
External	0	0	0	0	0	0	40	40
Total-10	0	1	7	1185	613	52	40	2124

TABLE A.5: Cross-classification of deaths between ICD-9 and ICD-10 at age 0.
Continued

A.2.2 Correspondence table of deaths above age 1

	Inf.	Neo.	Blood	ENM	Ment.	Nerv.	Circ.	Resp
Infectious	3192	101	47	24	16	30	76	49
Neoplasms	97	127626	55	63	12	53	234	101
Blood	41	335	1339	14	7	3	54	19
ENM	42	188	34	16850	55	43	306	201
Mental	26	79	15	225	5482	1058	1721	189
Nervous	42	67	9	79	56	12069	563	147
Circolatory	397	1643	92	1181	674	1217	190989	2077
Respiratory	182	560	35	143	510	421	1994	29988
Digestive	402	231	25	66	31	29	313	91
Skin	4	2	1	4	3	1	15	12
Muskoloskeletal	14	19	1	14	6	35	46	12
Genetourinary	66	53	11	42	10	27	231	58
Pregnancy	0	0	0	0	0	0	0	0
Congenital	5	13	3	9	0	9	99	4
Ill-defined	112	5	6	22	14	9	77	307
External	13	37	3	43	30	41	353	66
Total-10	4634	130982	1675	18780	6906	15049	197079	33339

TABLE A.6: Cross-classification of deaths between ICD-9 and ICD-10 for ages above 1.

	Dig	Skin	Mus.	Gen	Pre	Cong.	Ill	Ext.	Tot-9
Infectious	107	25	15	19	0	4	4	4	3712
Neoplasms	69	2	13	34	0	120	6	24	128510
Blood	59	0	28	7	1	142	1	3	2053
ENM	32	5	30	164	0	7	10	13	17980
Mental	54	25	60	37	0	3	46	16	9036
Nervous	56	13	58	27	0	13	40	17	13258
Circolatory	339	56	365	375	3	58	3396	232	203094
Respiratory	161	12	127	73	0	40	34	42	34322
Digestive	17478	5	18	30	0	8	14	27	18769
Skin	5	432	11	0	0	0	0	2	491
Muskoloskeletal	2	6	1328	5	0	2	5	36	1531
Genetourinary	28	10	9	6125	0	10	8	7	6695
Pregnancy	0	0	0	0	6	0	0	0	6
Congenital	10	1	2	6	0	274	2	2	529
Ill-defined	14	6	9	10	0	5	5942	20	6584
External	48	2	21	33	0	1	6	5466	6164
Total-10	18468	600	2095	6948	11	780	9513	5912	452773

TABLE A.7: Cross-classification of deaths between ICD-9 and ICD-10 for ages above 1.
Continued

A.3 Transition matrix

Transition matrix contains coefficient of redistribution matrix that are applied to cause-specific deaths count in order to avoid statistical disrpution and reconstruct cause-specific time series. Two transition matrices have been used, one for deaths occurred in the first year of life and one for deaths above age 1.

A.3.1 Transition matrix of deaths in the first year of life

	Inf.	Neo.	Blood	ENM	Ment.	Nerv.	Circ.	Resp
Infectious	0.767	0	0	0	0	0	0	0
Neoplasms	0	0.909	0	0	0	0	0	0
Blood	0	0	0.714	0	0	0	0	0
ENM	0	0	0	0.840	0	0	0	0
Mental	0	0	0	0	0	0	0	0
Nervous	0	0	0	0	0	0.868	0	0
Circolatory	0	0	0	0	0	0	0.775	0
Respiratory	0	0	0	0	0	0	0.067	0.533
Digestive	0	0	0	0	0	0	0	0
Skin	0	0	0	0	0	0	0	0
Muskoloskeletal	0	0	0	0	0	0	0	0
Genetourinary	0	0	0	0	0	0	0	0
Perinatal	0.006	0	0.001	0.001	0	0.003	0.011	0.004
Congenital	0	0.001	0.001	0.001	0	0.004	0.018	0.004
Ill-defined	0	0	0	0	0	0	0.098	0.050
External	0	0	0	0	0	0	0	0

TABLE A.8: Coefficient of redistributions for deaths at age 0.

	Dig.	Skin	Mus.	Gen.	Per.	Cong.	Ill	Ext.
Infectious	0	0	0	0	0.200	0.033	0	0
Neoplasms	0	0	0	0	0.091	0	0	0
Blood	0	0	0	0	0.286	0	0	0
ENM	0	0	0	0	0.160	0	0	0
Mental	0	0	0	0	0	0	0	0
Nervous	0	0	0	0	0.053	0.079	0	0
Circolatory	0	0	0	0	0.200	0.250	0	0
Respiratory	0	0	0	0	0.367	0.033	0	0
Digestive	0.900	0	0	0	0	0.100	0	0
Skin	0	0	0	0	0	1	0	0
Muskoloskeletal	0	0	1	0	0	0	0	0
Genetourinary	0	0	0	0.400	0.500	0.100	0	0
Perinatal	0.007	0	0	0	0.931	0.025	0.010	0
Congenital	0.010	0	0	0.005	0.091	0.862	0	0
Ill-defined	0	0	0	0	0.197	0	0.656	0
External	0	0	0	0	0	0	0	1

TABLE A.9: Coefficient of redistributions for deaths at age 0.
Continued

A.3.2 Transition matrix of deaths above age 1

	Inf.	Neo.	Blood	ENM	Ment.	Nerv.	Circ.	Resp
Infectious	0.860	0.027	0.013	0.006	0.004	0.008	0.020	0.013
Neoplasms	0.001	0.993	0.001	0.001	0.001	0.001	0.001	0.002
Blood	0.020	0.163	0.652	0.007	0.003	0.001	0.026	0.009
ENM	0.002	0.010	0.002	0.937	0.003	0.002	0.017	0.011
Mental	0.003	0.009	0.002	0.025	0.607	0.117	0.190	0.021
Nervous	0.003	0.005	0.001	0.006	0.004	0.910	0.042	0.011
Circolatory	0.002	0.008	0.001	0.006	0.003	0.006	0.904	0.010
Respiratory	0.005	0.016	0.001	0.004	0.015	0.012	0.058	0.874
Digestive	0.021	0.012	0.001	0.003	0.002	0.002	0.017	0.005
Skin	0.008	0.004	0.002	0.008	0.006	0.002	0.030	0.024
Muskoloskeletal	0.009	0.012	0.001	0.009	0.004	0.023	0.030	0.008
Genetourinary	0.010	0.008	0.002	0.006	0.001	0.004	0.034	0.009
Pregnancy	0	0	0	0	0	0	0	0
Congenital	0.011	0.030	0.007	0.021	0	0.0020	0.226	0.009
Ill-defined	0.017	0.001	0.001	0.003	0.002	0.001	0.012	0.047
External	0.002	0.006	0.001	0.007	0.005	0.007	0.057	0.011

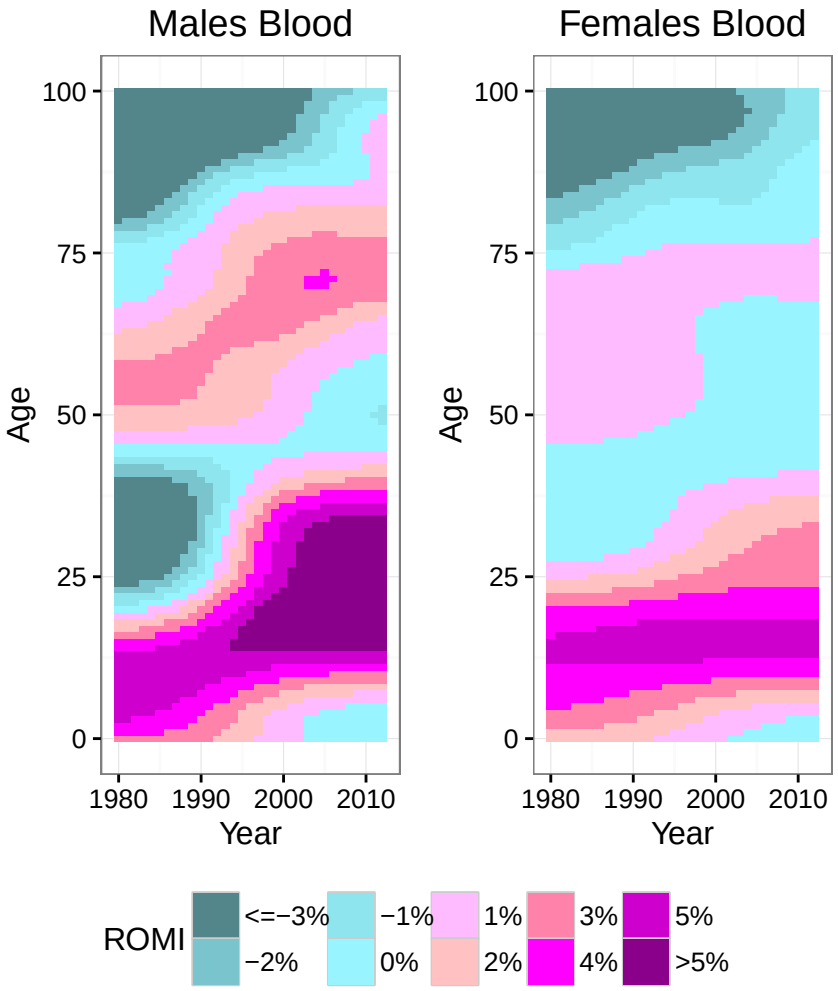
TABLE A.10: Coefficient of redistributions for deaths above age 1.

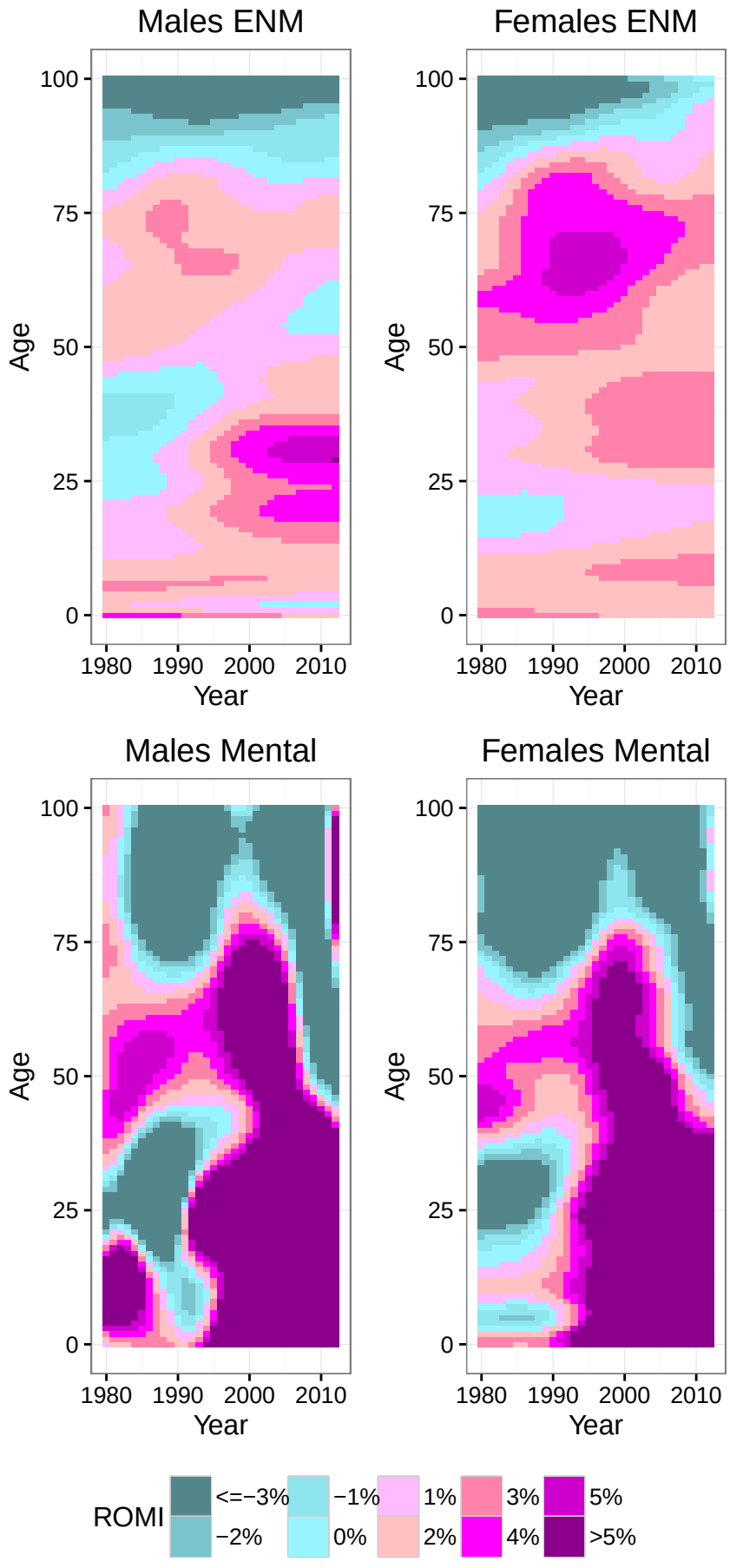
	Dig	Skin	Mus.	Gen	Pre	Cong.	Ill	Ext.
Infectious	0.029	0.007	0.004	0.005	0	0.001	0.001	0.001
Neoplasms	0.001	0	0	0	0	0.001	0	0
Blood	0.029	0	0.014	0.003	0	0.069	0	0.001
ENM	0.002	0.001	0.002	0.009	0	0	0.001	0.001
Mental	0.006	0.003	0.007	0.004	0	0	0.005	0.002
Nervous	0.004	0.001	0.004	0.002	0	0.001	0.003	0.001
Circolatory	0.002	0	0.002	0.002	0	0	0.016	0.001
Respiratory	0.005	0.001	0.004	0.002	0	0.001	0.001	0.001
Digestive	0.931	0.001	0.001	0.002	0	0.001	0.001	0.001
Skin	0.010	0.878	0.022	0	0	0	0	0.004
Muskoloskeletal	0.001	0.003	0.867	0.003	0	0.001	0.003	0.023
Genetourinary	0.004	0.001	0.001	0.914	0	0.002	0.001	0.001
Pregnancy	0	0	0	0	1	0	0	0
Congenital	0.023	0.002	0.005	0.014	0	0.624	0.005	0.005
Ill-defined	0.002	0.001	0.001	0.001	0	0.001	0.906	0.003
External	0.008	0.001	0.003	0.005	0	0	0.001	0.887

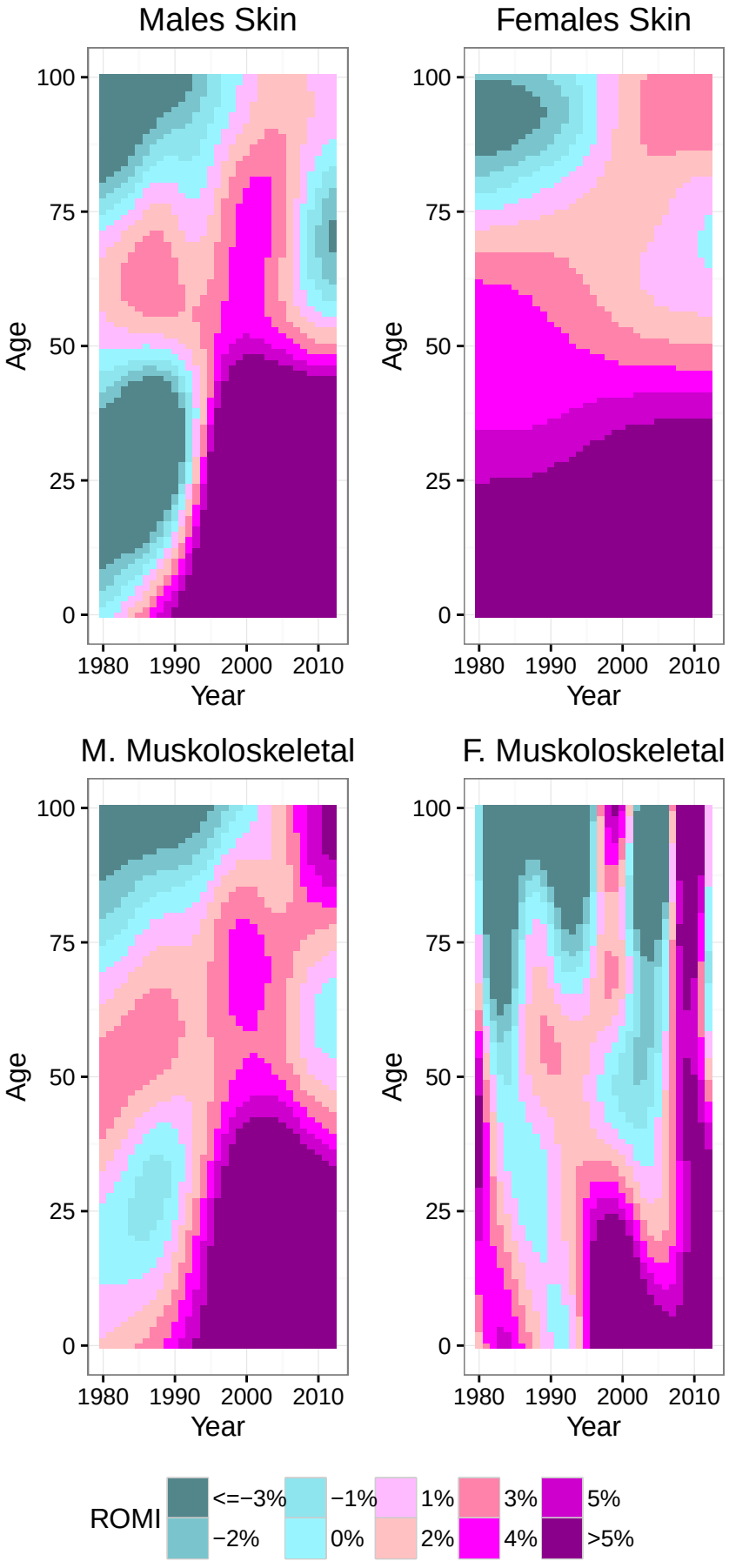
TABLE A.11: Coefficient of redistributions for deaths above age 1.
Continued

A.4 Rates Of Mortality Improvement

Cause-specific ROMIs were analyzed in Paragraph 4.2.3. Here are reported figures of all cause-specific ROMIs not shown in Paragraph 4.2.3.




Males ENM
Females ENM
Males Mental
Females Mental
Age
Year
100
75
50
25
0
1980
1990
2000
2010
ROMI
<=−3%
−2%
−1%
0%
1%
2%
3%
4%
5%
>5%


Males Skin
Females Skin
M. Muskoloskeletal
F. Muskoloskeletal
Age
Year
100
75
50
25
0
1980
1990
2000
2010
ROMI
<=−3%
−2%
−1%
0%
1%
2%
3%
4%
5%
>5%

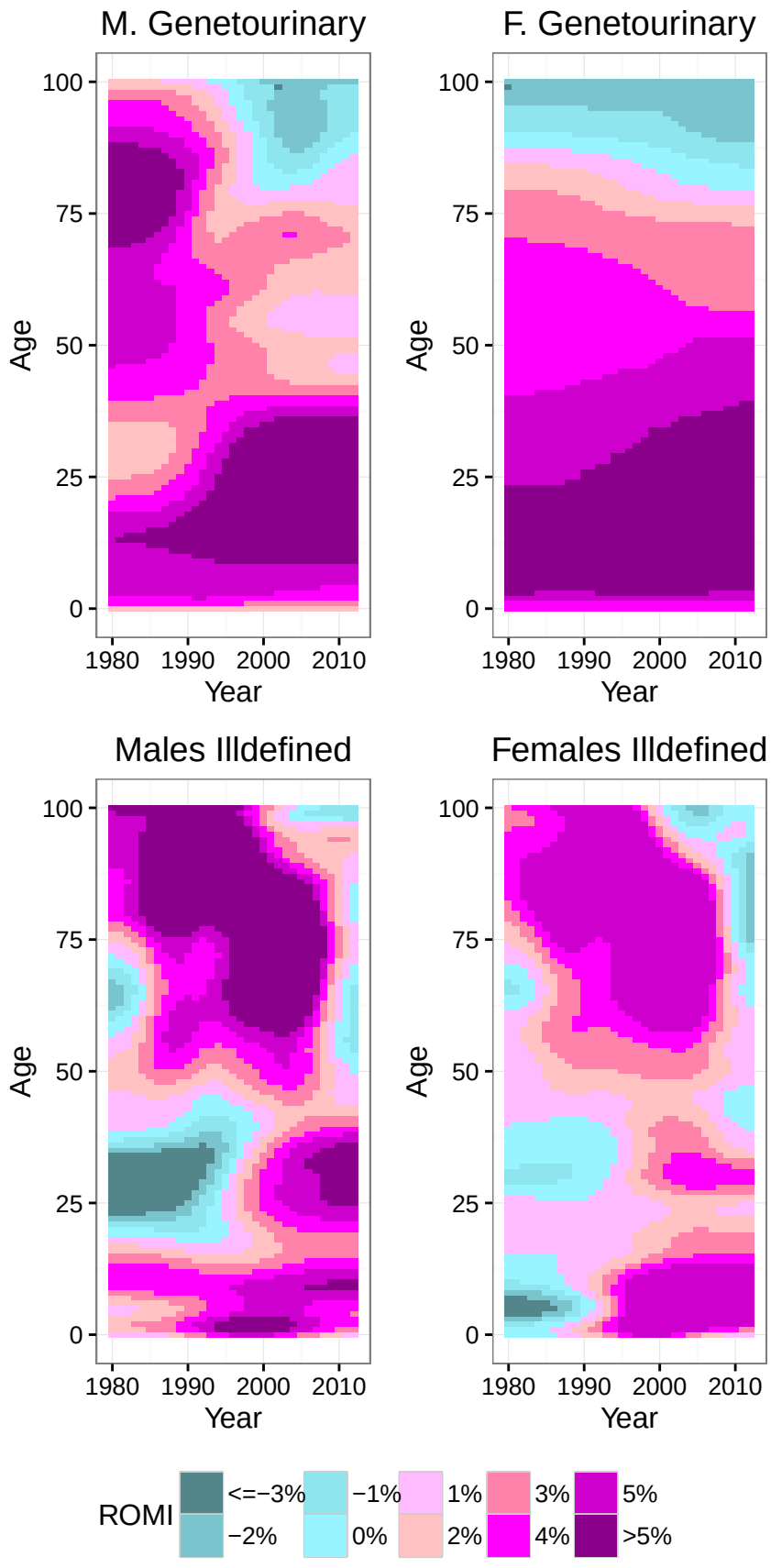


FIGURE A.1: Cause-specific surface of mortality improvements based on smoothed death rates. Italy, 1980-2013.